arXiv:2602.09240v1 [math.ST] 9 Feb 2026

正交不变广义线性模型中的最优估计：谱初始化与近似消息传递

Yihan Zhang* Hong Chang Ji† Ramji Venkataramanan‡ Marco Mondelli§

2026年2月11日

摘要

我们考虑从具有在概率分布上正交不变的随机设计矩阵的广义线性模型中进行参数估计的问题。此类模型允许设计具有任意的奇异值分布，并且仅假设其奇异向量是一般的。它是理论文献中通常考虑的独立同分布高斯设计的大幅推广，其动机在于真实数据往往具有复杂的相关结构，因此依赖独立同分布假设的方法可能高度次优。基于谱初始化迭代优化范式，本文提出最优谱估计量，并将其与近似消息传递（AMP）算法相结合，为这两个算法步骤建立了严格的性能保证。谱初始化与随后的近似消息传递均达到了关于估计的基本极限的现有猜想——前者针对高效弱恢复的最优样本复杂度，后者针对最优误差。数值实验表明，我们的方法有效，且我们的理论在正交不变数据之外仍然准确。

目录

1 引言 **1**

2 相关工作 **4**

3 预备知识 5
3.1 记号与定义 .5
3.2 模型 .6

4 主要结果 **7**
4.1 最优谱估计量 .7
4.2 谱初始化 GVAMP 及其状态演化 .10
4.3 谱初始化的 Bayes-GVAMP 及其状态演化 .14
4.4 猜想的贝叶斯风险 .15

5 启发式方法与证明概述 **18**
5.1 使用线性化 GVAMP 模拟谱估计量 .18
5.2 D 的分析及定理 4.1 的证明概述 .20
5.3 通用 GVAMP 形式体系及定理 4.3 的证明概述 .21
5.4 谱初始化的设计 .21

*布里斯托大学数学学院。电子邮箱：yihan.zhang@bristol.ac.uk。
†成均馆大学数学系。电子邮箱：hcji@skku.edu。
‡剑桥大学工程系。电子邮箱：rv285@cam.ac.uk。§奥地利科学技术研究所。电子邮箱：marco.mondelli@ist.ac.at。

6 示例与数值实验 216.1 合成数据 .226.2 真实数据 .24

7 讨论 24

A 自由累积量和矩 32

B 定理 4.1 的证明 32

C 定理 B.1 的证明 34
C.1 定理 C.1 的证明 .36C.2 定理 C.2 的证明 .40C.3 定理 C.3 的证明 .40C.4 定理 C.4 的证明 .44C.5 定理 B.1 的证明 .44

D 定理 B.2 的证明 44
D.1 设置 .46
D.2 谱体边缘的收敛 .47D.2.1 Haar 正交矩阵与确定性矩阵的强渐近自由性 .47D.2.2 对 δ 的不敏感性. 52D.2.3 定理 D.2 的证明 .53D.3 自由乘法卷积的谱边缘 .54D.3.1 设置 .56

D.3.2 解析从属.59D.3.3 解析从属的推论 .59D.3.4 定理 D.9 的证明 .60D.4 对我们情形的特化 .61

E 定理 B.3 的证明 62
E.1 定理 E.1 的证明 65
E.2 定理 E.2 的证明 67

F 定理 4.2 的证明 68

G 定理 4.3 的证明 70
G.1 定理 G.1 的证明 .71G.2 定理 G.2 的证明 .95G.3 定理 4.3 的证明.100

H 定理 4.4 的证明 107

I 定理 4.6 的证明 114

J 示例和数值实验的详细信息 116
J.1 副本鞍点方程 .116J.2 合成数据 .118J.3 真实数据 .120

K 辅助引理 120

1 引言

我们研究从广义线性模型（GLM）中进行参数估计这一经典问题 [78]。考虑一个数据集 $\mathcal{D} = (X,\ y) \in \mathbb{R}^{n\times d} \times \mathbb{R}^n$，其中设计矩阵 $X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top$ 由 n 个协变量向量组成，每个向量的维数为 d，且 y 是一个响应变量向量，其中每个响应变量均由以下方式给出

$$y_i = q(\langle x_i, \beta_* \rangle, \varepsilon_i), \qquad 1 \le i \le n.$$

这里，$q : \mathbb{R}^2 \to \mathbb{R}$ 是链接函数，ε_i 是噪声变量，$\beta_* \in \mathbb{R}^d$ 是未知的回归系数向量，而 $\langle \cdot\ ,\ \cdot \rangle$ 表示欧几里得内积。任务是根据 $\mathcal{D}$估计 β_*。非线性链接函数q 推广了线性回归，并涵盖工程学和自然科学中最常用的统计模型，包括逻辑回归 [100], 、相位恢复 [93, 47] （$y_i = |\langle x_i,\ \beta_* \rangle| + \varepsilon_i$）以及1比特压缩感知 [12] （$y_i = \mathrm{sgn}(\langle x_i,\ \beta_* \rangle + \varepsilon_i)$ ）。

在样本量 n 远大于参数维度 d的经典情形下，β_* 中的回归系数可以得到一致估计，置信区间和基于似然的检验等标准推断程序也能提供准确的不确定性量化。然而，在现代高维情形下，n 和 d 都很大且大小相近（$n/d \to \delta \in (0,\ \infty)$），这些经典结果不再成立 [38, 95]。因此，大量研究致力于确定这一高维情形下的统计极限并构造高效估计器，其重点大多放在设计矩阵X 无结构的情形，即其元素为独立同分布高斯 [5, 8, 38, 98, 95]。

高维情形下的估计可分为两个步骤：*(i)* 生成一个与 β_*具有非平凡相关性的初步估计 $\widehat{\beta} \in \mathbb{R}^d$，以及 *(ii)* 通过迭代程序改进初始化 $\widehat{\beta}$。更准确地说，在步骤*(i)* 中，我们旨在生成一个与 β_*具有非消失渐近重叠的估计：

$$\liminf_{d\to\infty} \frac{|\langle \widehat{\beta}, \beta_* \rangle|}{\|\widehat{\beta}\|_2 \|\beta_*\|_2} > 0. \tag{1.1}$$

文献中通常将这一准则称为“弱恢复” [71, 67, 62]。谱方法 [31]提供了一种简单且广受欢迎的方法，可用于获得满足 β (1.1)的谱估计量。广义线性模型的谱估计量由以下依赖于数据的矩阵的主特征向量给出：

$$D = X^\top \operatorname{diag}(\mathcal{T}(y_1), \cdots, \mathcal{T}(y_n)) X = \sum_{i=1}^n \mathcal{T}(y_i) x_i x_i^\top \in \mathbb{R}^{d\times d}, \tag{1.2}$$

其中 $\mathcal{T}$：$\mathbb{R} \to \mathbb{R}$ 是用户定义的预处理函数。通过式(1.2) 中的矩阵得到的谱估计量，其性能已得到深入研究 [79, 18, 30, 60, 71, 61]；具体而言，对于独立同分布高斯设计和先验分布，文献 [71] 证明，适当选择 $\mathcal{T}$，即可针对一类链接函数实现达到信息论极限的弱恢复 q，其中包括相位恢复。接下来讨论步骤*(ii)*，人们提出了多种迭代程序，从通过梯度下降或交替最小化优化适当的损失函数（例如惩罚对数似然） [18, 79, 106, 26] 到 Kaczmarz方法 [104]和高阶方法 [20]。

在贝叶斯设定中，若 β_* 按照与 X独立的先验分布生成，则后验均值 $\mathbb{E}[\beta_* \mid X, y]$ 使均方误差最小。尽管这通常在计算上不可行，渐近贝叶斯风险$\lim_{d\to\infty} \frac{1}{d}\mathbb{E}\left[\|\beta_* - \mathbb{E}[\beta_* \mid X, y]\|_2^2\right]$ 已在适当的模型假设 [5]下，针对独立同分布高斯设计得到精确刻画。对于该设定，已证明 [73] ，（1.2）中的谱估计量可通过广义近似消息传递（GAMP）算法加以改进[87]。在某些情形下，这一组合方法可达到贝叶斯风险 [5], ；更广泛地，GAMP估计量已被证明在广义一阶方法（GFOMs） [24, 75]中是最优的。

上述结果清晰地说明了，当设计矩阵 X 具有独立同分布高斯元素时，如何在广义线性模型中进行最优估计。其中一些渐近结果，例如谱方法以及某些凸惩罚估计量的结果，可以推广到更广泛的高斯设计设定，其具有

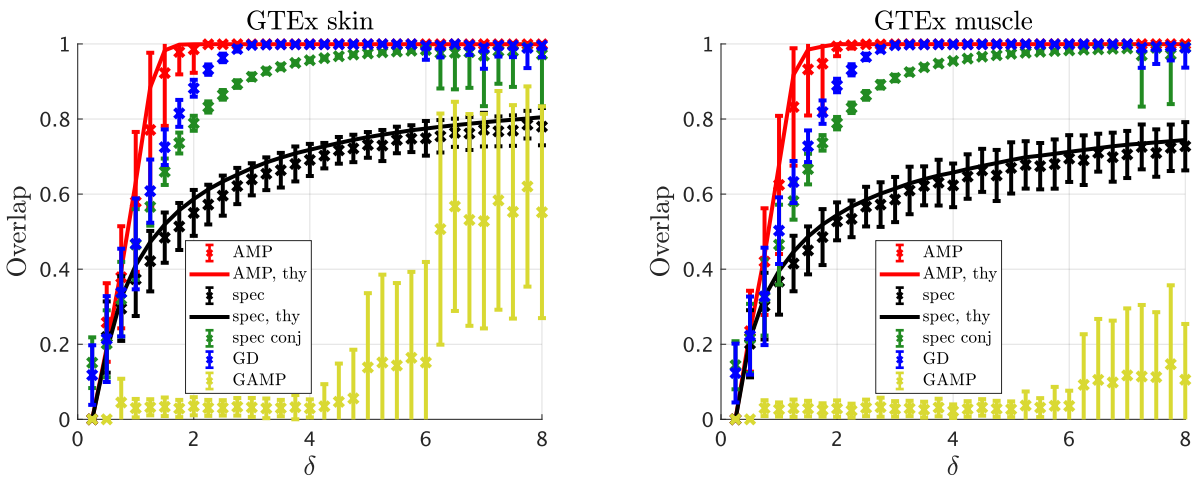

图1：性能对比5 种用于相位恢复的算法，其设计矩阵取自 GTEx 的两个真实数据集之一，详情见 第6节 。GAMP算法 [73] （黄色）在高斯数据上具有可证明的最优性，但在此处不再那么有效。本文提出的 Bayes-GVAMP（红色的“AMP”）采用谱初始化（黑色的“spec”），表现突出；正交不变性假设下的理论预测与此吻合。梯度下降（蓝色的“GD”）和 GAMP 均使用 [61, 公式（20）] 中的谱估计量（绿色的“spec conj”）进行初始化。它们达到与谱初始化的 Bayes-GVAMP 相同的弱恢复阈值，但重叠量较低。

任意协方差 [107, 109, 23]。然而，在许多应用中，数据具有复杂结构，且 X 各元素之间的相关性（包括行间依赖）无法被高斯设计充分刻画。因此，依赖于设计的高斯假设的方法可能高度次优。这在图1中得以体现：当 X 具有独立同分布高斯元素 [5, 73] 时被证明最优的算法，在 X 取自计算基因组学 [59]数据集的情况下，达不到理论所承诺的性能。

捕捉数据结构并保持数学上的可处理性的一种常见方法，是将设计矩阵 X 建模为服从正交不变的概率分布。这相当于X 的奇异向量是通用的（更形式化地说，是在正交群上服从 Haar 分布），而其谱可以是任意的，因此设计中既可以存在行列依赖，也可以包含其他多种结构。经验上已观察到这种设计选择具有普适性 [83, 13, 62]，并且对于一大类知名系综也已得到形式化证明 [39, 42,29, 43, 103] ，其中包括由离散余弦/正弦变换和 Hadamard-Walsh变换得到的系综；参见 [42,第1节3]。

由于具有灵活性，正交不变设计近来在统计学文献中受到研究，例如在线性回归 [58]和主成分分析 [44, 4, 41]领域。针对采用正交不变设计的广义线性模型中的估计，已有研究提出了多种 AMP 算法 [91, 101], ，并证明其不动点与借助统计物理工具所推测的贝叶斯风险相吻合 [53, 96]。然而，这些 AMP 算法要求初始化满足 (1.1) ，且与设计矩阵相互独立。对于若干常见的非线性链接函数（ q 例如相位恢复），假设能够获得这样的初始化并不现实。谱估计量是自然的初始化选择，但现有结果要么对设计矩阵的谱有诸多限制（例如 [40], 相关高斯设计中的单位奇异值 [107]） ，要么仍停留在猜想层面 [66]。简言之，要严格证明初始化方法的性能保证一直十分困难，更不用说将初始化与 AMP 等迭代方法结合起来。

贡献。 本文探讨了正交不变广义线性模型中最优估计的若干未解问题：我们提出了最优谱估计量，并展示了如何将其与近似消息传递（AMP）

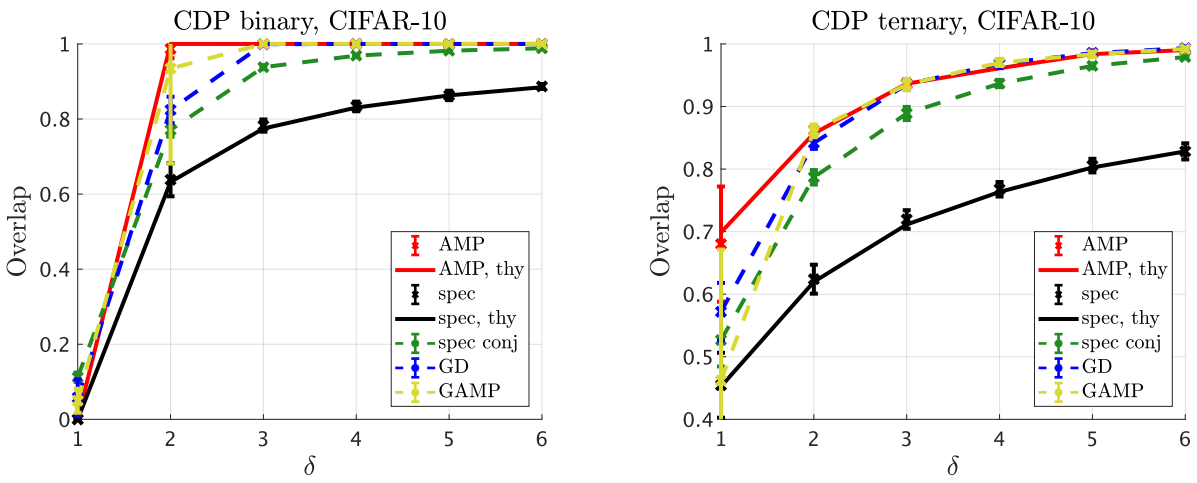

图 2：采用二元或三元编码衍射图样（CDP）作为设计矩阵的相位恢复问题中5 种算法的性能，详见第6节。回归系数是通过对一张 32×32 CIFAR-10 数据集中的卡车图像进行标准预处理得到的。即使设计矩阵并非正交不变，先验分布也并非独立同分布，我们的理论保证（“thy”，实线）仍然准确，这有力地体现了普适性。

算法结合起来，并为二者建立了严格的渐近保证。通过近似消息传递进行的初始化和迭代细化，均符合关于估计基本极限的现有猜想——前者涉及高效弱恢复的最优样本复杂度， [67],后者涉及最优误差 [53, 96]。我们的贡献总结如下。

1. 我们建立了一类预处理函数对应的谱估计量的渐近重叠（定理 4.1）。更具体地说，我们揭示了一个临界条件(4.13) ，在该条件下， D 在(1.2) 中的最大特征值会脱离谱的主体，而由 D 的最大特征向量给出的谱估计量能够按照(1.1)的定义实现弱恢复。这一现象类似于著名的 BBP 相变 [3]。

2. 利用定理 4.1的刻画，我们确定了一个能够以尽可能低的样本复杂度实现弱恢复的预处理函数（定理 4.2）。值得注意的是，谱估计量实现非零渐近重叠所需的临界样本复杂度，与 [67] 中猜想的、在所有高效估计器中达到最优的样本复杂度一致。

3. 我们设计了一类通用的近似消息传递迭代，按照 [14], 中的术语，我们称之为广义向量近似消息传递（GVAMP）。GVAMP 通过对设计矩阵的奇异值以及迭代量的各个分量应用非线性变换，迭代地改进谱初始化。我们证明，这类算法的动态过程（因此也包括其性能）可以通过状态演化递推来跟踪；状态演化递推是一组简洁的低维确定性方程（定理 4.3）。这解决了相关研究中常见的问题：要求初始化既能实现弱恢复，又独立于设计矩阵。

4. 最后，通过对定理 4.3中的通用 GVAMP 作适当特化，我们得到一种算法——谱初始化的 Bayes-GVAMP——其不动点与贝叶斯风险相符（定理 4.6），正如 [53, 96]中所猜想的那样。这为广义线性回归问题提供了一种端到端的算法解决方案，并对其性能作出了严格刻画；同时，该算法被猜想在一大类高效算法中具有最优性。

鉴于正交不变设计具有上述普适性，我们预计研究结果适用于应用中常见的设计矩阵。图 1 和 2展示了这一点：实证结果与

我们的理论预测相符，即使 X 并非根据正交不变分布生成，而是取自计算成像领域的 [17]或定量基因组学领域的 [59] 数据集；我们提出的算法始终优于基于设计矩阵独立同分布假设的方法。

2 相关工作

谱方法。 一系列研究表明，在采用独立同分布高斯设计 [60, 71, 61, 39], 相关高斯设计 [107] 和部分 Haar设计 [62, 40]的广义线性模型中，谱方法对于弱恢复是有效的，且在某些情况下可达到最优。此外，对于独立同分布高斯设计，谱方法已扩展至混合回归 [108], 多指标模型 [55, 35] 和高阶张量[33, 34]。线性估计量与谱估计量的最优组合见 [72]。本文进一步推进了这方面的研究，考察正交不变设计：我们提出并分析了谱估计量，其有效范围可达推测的计算上最优弱恢复阈值 δ_* [67]，即弱恢复能够通过高效算法实现的最小 δ。谱方法的计算最优性猜想，以其与线性化 AMP [66]之间广为人知的启发式联系为依据。具体而言，基于统计物理的考量提出，Bayes-GVAMP 在所有多项式时间算法中是最优的。将谱估计量视为幂迭代时，它们可看作 Bayes-GVAMP 在其平凡不动点附近的线性化。于是， δ_* 可通过启发式方法确定为平凡不动点具有排斥性的最小 δ。我们关于谱估计量的结果建立在这一联系之上，并使其得到严格论证。

近似消息传递。我们的通用 GVAMP 形式体系借鉴了许多用于贝叶斯线性模型或广义线性模型的其他迭代算法，包括自适应 Thouless–Anderson–Palmer（AdaTAP） [80, 82], 期望传播（EP） [70, 92, 62, 64], 期望一致近似（EC） [81, 54], S-AMP [16, 15], RI-GAMP [101], 正交 AMP [63, 39, 42, 43, 41], 向量 AMP [88, 91, 97]。文献 [65] 通过一种非严格的 Plefka 展开方法指出，AdaTAP、EC 和 VAMP 的推导是等价的。上述所有算法都对某些特定的预解式型矩阵进行运算 X （例如线性最小均方误差估计），并对迭代量应用最小均方误差去噪器。这一特点根本上源于这些算法的推导方式。我们的通用 GVAMP 形式体系并未预先嵌入最小均方误差估计，因此允许对 X进行任意无迹谱变换，并对迭代量应用散度为零的逐元素非线性函数。预计这种普适性在贝叶斯最优设置之外也会发挥作用。采用最小均方误差去噪器的 VAMP 型算法也已用于比广义线性模型更复杂的模型 [85, 84, 105]。预计我们的通用 GVAMP 形式体系及其相关的状态演化结果可以扩展到这些设置。针对独立同分布设计开发的 [7, 11] 近似消息传递算法可作为算法基元和证明技术，应用于各种问题[76, 74, 21, 19, 51]。本文提出的 GVAMP 在针对独立同分布高斯设计进行特化后，与文献 [7, 11]中的近似消息传递算法有所不同：经验上观察到，GVAMP 与近似消息传递算法具有相同的不动点（图 8左图），但收敛速度更快。

贝叶斯风险。在独立同分布高斯设计下，广义线性模型的贝叶斯风险已通过自适应插值方法在 [5]中建立，[6]。据我们所知，该方法尚未成功应用于正交不变设计。然而，最近 [57] 证明了在附加技术条件下，正交不变设计下线性模型的贝叶斯风险；该条件限制了 X 的极限谱分布支撑集，使其足够窄。在 [53, 96, 67]中提出的正交不变设计下广义线性模型的贝叶斯风险猜想，尚未在其完整的一般性下得到严格证明。 [89,50] 将这类关于贝叶斯风险的猜想推广到了更复杂的架构，例如采用正交不变权重矩阵的多层网络。

3 预备知识

3.1 符号与定义

所有向量均为列向量。对于对称矩阵 $M \in \mathbb{R}^{d\times d}$，我们将 $\lambda_1(M) \geq \cdots \geq \lambda_d(M)$ 记为其按非递增顺序排列的（实）特征值，将 $v_1(M) \cdots, v_d(M) \in \mathbb{S}^{d-1}$ 记为相应的特征向量（其范数为单位范数）。对于矩阵 $V = \begin{bmatrix} v_1 & \cdots & v_k \end{bmatrix} \in \mathbb{R}^{n\times k}$，我们用 $\Pi_V \in \mathbb{R}^{n\times n}$ 表示到张成空间 $\operatorname{span}\{(v_i)_{i\in[k]}\}$的正交投影，并令 $\Pi_{V^\perp} = I_n - \Pi_V$。

除非另有说明，所有 lim、lim inf、lim sup 均几乎必然成立。我们用 $\delta_x(\cdot) = \delta(\cdot - x)$ 表示位于 $x \in \mathbb{C}$ 处的点质量。维数为 n 的正交群定义为

$$\mathbb{O}(n) \coloneqq \left\{O \in \mathbb{R}^{n\times n} : OO^\top = O^\top O = I_n\right\}.$$

其上唯一的双不变 Haar 测度（归一化为概率测度）记为 $\mathrm{Haar}(\mathbb{O}(n))$。矩阵 M 的经验谱分布定义为：若 $M \in \mathbb{R}^{d\times d}$ 为对称矩阵，则其特征值的经验分布为：

$$\frac{1}{d}\sum_{i=1}^{d} \delta_{\lambda_i(M)},$$

若 $M \in \mathbb{R}^{n\times d}$ 为矩形矩阵，则其奇异值的经验分布为：

$$\frac{1}{n\wedge d}\sum_{i=1}^{n\wedge d} \delta_{\sigma_i(M)}.$$

通常，我们用 p_{A}、$p_{\mathsf{A},\mathsf{B}}$、$p_{\mathsf{A}\mid\mathsf{B}}$ 表示随机变量 A、B 的边缘密度、联合密度和条件密度。

定义 3.1（伪-Lipschitz函数）。对于 $p \geq 1$，函数 $f\colon \mathbb{R}^k \to \mathbb{R}$ 称为 p 阶伪利普希茨函数，如果存在 $C > 0$，使得对所有 $x,\ y \in \mathbb{R}^k$，

$$|f(x) - f(y)| \leq C\|x - y\|_2\Big(1 + \|x\|_2^{p-1} + \|y\|_2^{p-1}\Big).$$

记$\mathrm{PL}_{k\to 1}^p$ 为所有此类函数的集合，且$\mathrm{PL}_{k\to 1} = \bigcup_{p\geq 1}\mathrm{PL}_{k\to 1}^p$。

定义 3.2（多项式增长）。对于 $p \geq 1$，函数 $f\colon \mathbb{R}^k \to \mathbb{R}^\ell$ 称为具有 p 阶多项式增长，如果存在 $C > 0$ 使得

$$\|f(x)\|_2 \leq C(1 + \|x\|_2^p), \tag{3.1}$$

for 所有 $x \in \mathbb{R}^k$。记$\mathrm{PG}_{k\to\ell}^p$ 为所有此类函数的集合，且$\mathrm{PG}_{k\to\ell} = \bigcup_{p\geq 1}\mathrm{PG}_{k\to\ell}^p$。

注意，任何伪利普希茨函数都具有多项式增长，即$\mathrm{PL}_{k\to 1}^p \subset \mathrm{PG}_{k\to 1}^p$。

定义 3.3（Wasserstein 收敛，[102, 第 6] 章）。对于 $p \geq 1$，我们称（可能是随机）矩阵 $\begin{bmatrix} v_1 & \cdots & v_k \end{bmatrix} = \begin{bmatrix} v_{i,j} \end{bmatrix}_{(i,\ j)\ \in[n]\times[p]} \in \mathbb{R}^{n\times k}$ 的行的经验分布在 Wasserstein-p意义下收敛到随机向量 $\begin{bmatrix} \mathsf{V}_1 & \cdots \mathsf{V}_k \end{bmatrix} \in \mathbb{R}^{1\times k}$，且其 $\mathbb{E}\big[\big\|\begin{bmatrix} \mathsf{V}_1 & \cdots \mathsf{V}_k \end{bmatrix}\big\|_2^p\big]$ 有限，记作

$$\begin{bmatrix} v_1 & \cdots & v_k \end{bmatrix} \xrightarrow{W_p} \begin{bmatrix} \mathsf{V}_1 & \cdots & \mathsf{V}_k \end{bmatrix}, \tag{3.2}$$

若对任意 $f \in \mathrm{PL}_{k\to 1}^p$，几乎必然有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} f(v_{i,1}, \cdots, v_{i,k}) = \mathbb{E}[f(\mathsf{V}_1, \cdots, \mathsf{V}_k)]. \tag{3.3}$$

我们用 $\begin{bmatrix} v_1 & \cdots & v_k \end{bmatrix} \xrightarrow{W} \begin{bmatrix} \mathsf{V}_1 & \cdots \mathsf{V}_k \end{bmatrix}$ 表示对每个固定的 $p \geq 1$ 都成立的 W_p 收敛，其中右侧具有所有阶的有限混合矩。

要验证 (3.2)，只需验证 (3.3) 对于以下任一情形成立：(a) 任意 $f \in \mathrm{PG}_{k\to 1}^p$ 连续；或 (b) $f(v_1, \cdots, v_k) = \big\|\begin{bmatrix} v_1 & \cdots & v_k \end{bmatrix}\big\|_2^p$ 以及任意有界利普希茨 f；参见 [44, 附录 E]。

3.2 模型

设 $X \in \mathbb{R}^{n\times d}$ 为设计矩阵，$\beta_* \in \mathbb{R}^d$ 为回归系数向量，$\varepsilon \in \mathbb{R}^n$ 为噪声变量向量。考虑一个广义线性模型

$$y = q(X\beta_*, \varepsilon). \tag{3.4}$$

在上述表达式中，函数 q: $\mathbb{R}^2 \to \mathbb{R}$ 逐分量作用于其向量输入，即 $q(z,\ \varepsilon) = \big[q(z_1,\ \varepsilon_1)\cdots\ \ q(z_n,\ \varepsilon_n)\big]^\top$，其中 $z := X\beta_*$。等价地，

$$y \,|\, X\beta_* \sim Q(\cdot \,|\, X\beta_*), \tag{3.5}$$

对于逐分量作用的条件密度 $Q(\cdot\,|\,\cdot)$，即 $\mathbb{P}(y\,|\,z) = \prod_{i=1}^n Q(y_i\,|\,z_i)$。

我们假设 β_*，X，ε 相互独立，且各自满足以下分布假设。

(A1) $\beta_* \sim P_{\mathsf{B}_*}^{\otimes d}$，其中 P_{B_*} 是定义在 $\mathbb{R}$ 上的一个分布，其均值为 0，方差为 $\rho \in (0,\ \infty)$。

(A2) X 在概率分布意义下是双正交不变的，即 $X \overset{d}{=} \widetilde{O}X\widetilde{Q}^\top$ 对于任意确定的 $\widetilde{O} \in \mathbb{O}(n)$ $\widetilde{Q} \in \mathbb{O}(d)$。等价地，$X$ 可表示为其奇异值分解 $X = O\Lambda Q^\top$，其中 $(O,\ Q)$ $\sim$Haar($\mathbb{O}(n)$)$\otimes$Haar($\mathbb{O}(d)$) 与矩形对角矩阵 $\Lambda \in \mathbb{R}^{n\times d}$ 相互独立。令 $\lambda \in \mathbb{R}^{n\wedge d}$ 为包含 Λ 对角元素的向量，我们进一步假设，几乎必然地，λ 的经验分布在 Wasserstein（所有阶）意义下收敛于随机变量 Λ 的概率分布 μ_Λ；该随机变量的支撑集是 $(0,\ \infty)$ 上的紧集，且具有正方差。此外，$\max_{1\le i\le n\wedge d}\lambda_i$，$\max_{1\le i\le n\wedge d}\lambda_i^{-1}$ 在 $n,\ d$ 上有一致上界。几乎必然地，对于每个 $\varsigma > 0$，存在 $N(\varsigma)$，使得对于每个$n \wedge d \ge N(\varsigma)$，

$$\{\lambda_i\}_{1\le i\le n\wedge d} \subset \operatorname{supp}\mu_\Lambda + (-\varsigma, \varsigma). \tag{3.6}$$

以下两个极限成立

$$\lim_{z \nearrow \inf\operatorname{supp}\mu_\Lambda} \mathbb{E}\left[\frac{\Lambda}{\Lambda - z}\right] = \lim_{z \searrow \sup\operatorname{supp}\mu_\Lambda} \mathbb{E}\left[\frac{\Lambda}{z - \Lambda}\right] = \infty. \tag{3.7}$$

(A3) $\varepsilon \in \mathbb{R}^n$ 满足 $\varepsilon \xrightarrow{W} \mathsf{E}$，其中 E 是一个具有有限方差的实值随机变量。记 P_{E} 为 E 的概率分布。

(A4) $n,\ d \to \infty$ 且 $n/d \to \delta \in (0,\ \infty)$。除非另有说明，所有极限均是在此尺度设定下取的。

我们对假设（A2）作几点评论。该假设中的条件(3.6)意味着 X 的经验谱分布强收敛到 μ_Λ，其含义是 X 的最大和最小特征值分别收敛到 Λ 支撑集的上边缘和下边缘，即

$$0 < \lim_{n\wedge d\to\infty} \min_{1\le i\le n\wedge d} \lambda_i = \inf\operatorname{supp}\Lambda < \sup\operatorname{supp}\Lambda = \lim_{n\wedge d\to\infty} \max_{1\le i\le n\wedge d} \lambda_i < \infty.$$

这尤其排除了位于 X谱左侧或右侧的离群奇异值。否则，这些离群值可能导致按 D 构造的(1.2) 谱估计量 [10, 定理 25], 的谱中出现离群值，从而与由 β_*产生的信息性离群值混淆。

式 (3.7) 中的极限要求 Λ 的支撑集在最左侧和最右侧边界处具有硬边。这类假设在高维统计学文献中很常见；例如，参见 [37, 假设 6“主体边缘厚度”] 以及 [107, (A6), (A7)]。我们认为，这纯粹是一个技术性假设，可以通过扰动论证予以去除 [107, 附录 C]。事实上，第6节中的模拟数值验证了我们的理论结果，其中 Λ 分别与 Beta(3,1) 和 Beta(1,2) 成比例时，式 (3.7) 中的第一个和第二个极限分别有限。类似地，由于 $\inf\operatorname{supp}\mathrm{Beta}(3,1) = \inf\operatorname{supp}\mathrm{Beta}(1,2) = 0$,

关于 Λ 的支撑集左边缘与 0 保持正距离的假设是技术性的，或可去除。另一方面，关于 Λ 的有界支撑集假设可能是必要的；否则， D 在(1.2) 中的谱将无界 [9, 定理 59]。

对 X 的概率分布的双正交不变性假设刻画了如下情形： X 的左、右奇异向量是一般的，并且彼此独立，也与奇异值独立。这使得所考虑的模型具有解析可处理性。然而，第 6.2 节中的数值实验表明，对于某些不具有正交不变性、甚至是确定性的设计系综，我们的理论预测仍然准确。识别 X 的合适普适类，并在其中论证我们理论的有效性，留待未来工作。这一方向的进展包括 [42, 43, 103]，尽管它们并不直接适用于我们所提出的算法。

假设 q， P_{B_*}，概率分布(Λ) P_{E} 已知（参见第 7 节中关于这一要求的进一步讨论），我们的兴趣在于构造 $\beta_* \in \mathbb{R}^d$ 的估计量，所用数据为$(X,\ y)$，而估计量$\widehat{\beta} \equiv \widehat{\beta}(X,\ y) \in \mathbb{R}^d$ 的质量由其与 β_* 的渐近重叠来衡量，该渐近重叠定义于(1.1)。

我们引入本文其余部分经常使用的一些记号。给定随机变量 Λ 于假设（A2）中，我们还定义 Λ_n、 Λ_d 如下

$$\mathrm{law}(\Lambda_n) \coloneqq \begin{cases} \mathrm{law}(\Lambda), & \text{if } \delta \le 1 \\ \delta^{-1}\,\mathrm{law}(\Lambda) + (1-\delta^{-1})\delta_0, & \text{if } \delta > 1 \end{cases}, \quad \mathrm{law}(\Lambda_d) \coloneqq \begin{cases} \delta\,\mathrm{law}(\Lambda) + (1-\delta)\delta_0, & \text{if } \delta \le 1 \\ \mathrm{law}(\Lambda), & \text{if } \delta > 1 \end{cases}. \tag{3.8}$$

根据定义， Λ_n与 Λ_d 分别服从 X的极限谱分布，并分别计入 n个与 d个奇异值。此外， $XX^\top$与 $X^\top X$ 的极限谱分布分别由 Λ_n^2与Λ_d^2的概率分布给出。令 $(\kappa_{2k})_{k\ge 0} \subset \mathbb{R}$ 与 $(m_{2k})_{k\ge 0} \subset \mathbb{R}_{\ge 0}$ 分别为 Λ_n^2；见附录 A 的相关背景。令

$$\sigma^2 \coloneqq \overline{\kappa}_2 \rho, \tag{3.9}$$

其中 ρ 是 P_{B_*} （来自假设（A1））的方差。注意， $\sigma^2 > 0$ ，因为假设（A2）蕴含$\kappa_2 = \mathbb{E}\left[\Lambda_n^2\right] > 0$。令

$$(\mathsf{B}_*, \mathsf{Z}, \mathsf{E}) \sim P_{\mathsf{B}_*} \otimes \mathcal{N}(0, \sigma^2) \otimes P_{\mathsf{E}}, \qquad \mathsf{Y} = q(\mathsf{Z}, \mathsf{E}). \tag{3.10}$$

我们还定义 $\mathsf{Z} \coloneqq \sigma^{-1}\mathsf{Z} \sim \mathcal{N}(0,\ 1)$。容易验证 $\beta_* \xrightarrow{W} \mathsf{B}_*$， $\begin{bmatrix} z & \varepsilon & y \end{bmatrix} \xrightarrow{W} \begin{bmatrix} \mathsf{Z} & \mathsf{E} & \mathsf{Y} \end{bmatrix}$。

4 主要结果

本节给出我们的主要形式化结果。 第 4.1 节 精确刻画矩阵 D 在 (1.2) 中的谱性质（特别是其主特征向量的重叠（1.1））以及预处理函数的最优选择。 第 4.2 节 介绍通用 GVAMP 形式体系，并刻画其在采用第 4.1 节的谱估计量初始化时由状态演化递推描述的动态。 第 4.3 节介绍采用最优去噪器的 GVAMP，由此得到本文的核心算法成果——谱初始化的 Bayes-GVAMP，并提供精确的渐近保证。 第 4.4 节 通过将其性能与猜想中的贝叶斯风险联系起来，为其计算最优性提供证据。

4.1 最优谱估计量

任取预处理函数 $\mathcal{T}\colon \mathbb{R} \to \mathbb{R}$。令

$$T \coloneqq \mathrm{diag}(\mathcal{T}(y)), \qquad D \coloneqq X^\top T X, \tag{4.1a}$$

其中， $\mathcal{T}$逐分量作用于 y。该谱估计量输出

$$\widehat{\beta}_{\mathrm{s}} \coloneqq v_1(D). \tag{4.1b}$$

我们将考虑满足以下条件的预处理函数 $\mathcal{T}$。

(A5) 预处理函数 $\mathcal{T}\colon \mathbb{R} \to \mathbb{R}$ 具有如下形式

$$\mathcal{T}(y) = \frac{\overline{g}(y)}{\overline{g}(y) + \gamma}, \qquad \text{where } \gamma \coloneqq \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right], \tag{4.2}$$

对于某个有上界且满足以下条件的函数 g

$$\inf_{y \in \operatorname{supp} \mathsf{Y}} \overline{g}(y) > -\gamma, \qquad \mathbb{E}[\overline{g}(\mathsf{Y})] = 0, \qquad \mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right] > 0. \tag{4.3}$$

此外，

$$\lim_{z \searrow \sup \operatorname{supp} \mu_{\mathsf{T}}} \mathbb{E}\left[\frac{\mathsf{T}}{z - \mathsf{T}}\right] = \infty, \tag{4.4}$$

其中 $\mathsf{T} = \mathcal{T}(\mathsf{Y})$。

式 (4.3) 中的第一个不等式确保 $\mathcal{T}(y)$ 在 (4.2) 中对任意 $y \in \operatorname{supp}\mathsf{Y}$ 均有定义，因为分母 $\overline{g}(y) + \gamma$ 为正。注意，对于 $\mathbb{E}\left[g(\mathsf{Y})\overline{\mathsf{Z}}^2\right]$ 的正性假设意味着 $\gamma > 0$。这是因为根据矩-累积量关系（参见 附录 A），

$$\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta = \frac{\overline{m}_4 - \overline{m}_2^2}{\overline{m}_2^2} > 0, \tag{4.5}$$

其中，该不等式由回顾 m_{2k} 的定义并应用柯西–施瓦茨不等式 $\mathbb{E}[\Lambda_n]^2 \le \mathbb{E}[\Lambda_n^2]$ 得到。$\mathbb{E}\left[g(\mathsf{Y})\overline{\mathsf{Z}}^2\right]$ 的正性不失一般性。若 $\mathbb{E}\left[g(\mathsf{Y})\overline{\mathsf{Z}}^2\right] < 0$，则可以改为考虑 $-g(y)$。此时，在映射 $\overline{g}(y) \mapsto -g(y)$ 下，$\gamma \mapsto -\gamma$ 和函数 $\mathcal{T}$ 保持不变。我们排除 $\mathbb{E}\left[g(\mathsf{Y})\overline{\mathsf{Z}}^2\right] = 0$ 的情形，因为这会导致对每个 y 都成立的平凡函数 $\mathcal{T}(y) = 1$。关于 $\operatorname{supp} \mu_{\mathsf{T}}$ 上边缘的条件 (4.4) 类似于对 $\operatorname{supp} \mu_{\Lambda}$ 施加的条件 (3.7)，且有可能可以去除。

假设 (A5) 较为宽松，因为存在满足该假设的预处理函数，能够达到推测上最优的弱恢复阈值；见 定理 4.2。由假设 (A5) 定义的函数类主要是受我们基于线性化 GVAMP 的证明技巧启发；见 第 5 节。尽管采用更大的函数类可以获得更高的渐近重叠（见 第 6 节和 第 7 节），但这些预处理函数无法使弱恢复阈值低于定理 4.2 中的阈值，而且后续对 Bayes-GVAMP 的改进（定理 4.4）所带来的优势也超过了它们在重叠方面的优势。另一方面，满足假设 (A5) 的预处理函数所对应的谱估计量更便于通过 GVAMP 进行改进，因为它们源自 GVAMP 的线性化。

下面的定理 4.1 刻画了 D 的最大的两个特征值的极限，以及满足临界条件时谱估计量的渐近重叠。为陈述这些刻画结果，我们需要先给出一些定义。回忆 $\mathsf{T} = \mathcal{T}(\mathsf{Y})$，并为简便起见记 $\mathsf{S} = \Lambda_n^2$。定义函数 $\psi\colon (\sup \operatorname{supp} \mathsf{T}\ \infty) \to \mathbb{R}$ 如下：

$$\psi(a) = a\mathbb{E}\left[\frac{\mathsf{S}}{\omega(a) - \mathsf{S}}\right]\mathbb{E}\left[\frac{1}{\omega(a) - \mathsf{S}}\right]^{-1} = \omega(a)\mathbb{E}\left[\frac{\mathsf{T}}{a - \mathsf{T}}\right]\mathbb{E}\left[\frac{1}{a - \mathsf{T}}\right]^{-1}, \tag{4.6}$$

wh其中 $\omega(a)$ 由下式隐式定义为唯一解：$\omega \in \mathbb{R} \setminus (\inf \operatorname{supp} \Lambda^2,\ \operatorname{supsupp}\ \Lambda^2)$ t o

$$\mathbb{E}\left[\frac{\mathsf{T}}{a - \mathsf{T}}\right] = \mathbb{E}\left[\frac{\mathsf{S}}{\omega - \mathsf{S}}\right], \tag{4.7}$$

或者等价地（在等式两边加上(4.7)），

$$\mathbb{E}\left[\frac{a}{a - \mathsf{T}}\right] = \mathbb{E}\left[\frac{\omega}{\omega - \mathsf{S}}\right]. \tag{4.8}$$

定义

$$\lambda^\circ = \psi(a^\circ), \tag{4.9}$$

其中，$a^\circ \in (\sup \operatorname{supp} \mathsf{T}\ \infty)$ 是 ψ 的最大临界点，即在 $(\sup \operatorname{supp} \mathsf{T}\ \infty)$ 中满足 $\psi'(a^\circ) = 0$ 的最大解，或者更明确地说，

$$\frac{\mathbb{E}\left[\frac{\mathsf{T}}{a^\circ - \mathsf{T}}\right]\mathbb{E}\left[\frac{1}{a^\circ - \mathsf{T}}\right]}{\mathbb{E}\left[\frac{\mathsf{T}}{(a^\circ - \mathsf{T})^2}\right]} + \frac{\mathbb{E}\left[\frac{\mathsf{S}}{\omega(a^\circ) - \mathsf{S}}\right]\mathbb{E}\left[\frac{1}{\omega(a^\circ) - \mathsf{S}}\right]}{\mathbb{E}\left[\frac{\mathsf{S}}{(\omega(a^\circ) - \mathsf{S})^2}\right]} = 1. \tag{4.10}$$

最后，令

$$\eta := \delta \frac{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)^2 \mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2 - \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\right]}{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)^3 \mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2 + \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)^2 \mathbb{E}\left[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\right] + \left(\frac{\overline{\kappa}_6}{\overline{\kappa}_2^3} - 2\frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^4} - \delta\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)\right)\mathbb{E}[\overline{g}(\mathsf{Y})^2]}. \tag{4.11}$$

定理 4.1（谱估计量）。考虑通过（4.1）定义的谱估计量，其中预处理函数满足假设（A5）。则几乎必然有：

$$\lim_{d\to\infty} \lambda_2(D) = \lambda^\circ, \tag{4.12}$$

其中 λ° 定义见（4.9）。此外，如果

$$\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2 > \mathbb{E}\left[\overline{g}(\mathsf{Y})^2\right], \tag{4.13}$$

那么我们有如下结果：

$$\lim_{d\to\infty} \lambda_1(D) = \overline{\kappa}_2 > \lambda^\circ, \qquad \lim_{d\to\infty} \frac{|\langle \beta_*, v_1(D)\rangle|^2}{\|\beta_*\|_2^2} = \eta \in (0,1), \tag{4.14}$$

其中 η 定义见（4.11）。

定理 4.1 由附录 B中的一系列结果推出，其证明见附录 C至E。详见第 5.1 节和第 5.2 节，其中对证明进行了详细概述。

注 4.1（普适性）。渐近重叠的值 η 在(4.11) 中具有一项令人惊讶的普适性：它与 X的极限谱分布有关。具体而言，表达式 Λ_n 中对(4.11) 的唯一依赖是通过 κ_2、κ_4、κ_6（或者等价地，根据矩-累积量关系，通过 Λ_n^2 的前三阶矩）。这表明，对于满足假设 (A5)的谱估计量而言，设计的谱律细节对估计量的性能影响极小。

以下结果刻画了假设 (A5)所给出的预处理函数族中的最优谱阈值（用于弱恢复），即满足以下条件的最小 δ：存在一个预处理函数 $\mathcal{T}$，与 g 相对应，且满足(4.13)。

定理 4.2 (最优谱阈值和预处理)。

1. 假设

$$\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)\int \frac{\mathbb{E}\left[Q(y \mid \mathsf{Z})\left(\overline{\mathsf{Z}}^2 - 1\right)\right]^2}{\mathbb{E}[Q(y \mid \mathsf{Z})]}\,\mathrm{d}y > 1. \tag{4.15}$$

然后，通过取

$$\overline{g}(y)=\frac{\mathbb{E}\Big[\Big(\overline{\mathsf{Z}}^2-1\Big)Q(y\,|\,\mathsf{Z})\Big]}{\mathbb{E}[Q(y\,|\,\mathsf{Z})]}=\mathbb{E}\Big[\overline{\mathsf{Z}}^2-1\,\Big|\,\mathsf{Y}=y\Big],\tag{4.16}$$

通过以下方式定义的相应预处理函数 假设(A5) 能够实现弱恢复，具体而言， (4.13) 成立。此外， g 在 (4.16) 中满足 (4.3) 的条件是：

$$\mathrm{Var}\Big[\mathbb{E}\Big[\overline{\mathsf{Z}}^2\,\Big|\,\mathsf{Y}\Big]\Big]=\mathbb{E}\Big[\mathbb{E}\Big[\overline{\mathsf{Z}}^2-1\,\Big|\,\mathsf{Y}\Big]^2\Big]>0.\tag{4.17}$$

2. 反之，如果存在一个满足 假设(A5) 且能够实现弱恢复（即 (4.13) 成立）的预处理函数，那么 (4.15) 也必须成立。

定理 4.2 的证明见 附录 F。条目 2 由对临界条件 (4.13)应用柯西–施瓦茨论证得到。条目 1 (4.16) 中 g 的最优选择 可以从柯西–施瓦茨论证的等号情形中直接读出。

注 4.2（猜想的计算上最优弱恢复阈值）。 利用矩-累积量关系（见 附录 A）

$$\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}+\delta=\delta\frac{\overline{m}_4'}{\overline{m}_2'^2}-1,$$

我们可以将条件(4.15) 等价地写成

$$\delta>\frac{\overline{m}_2'^2}{\overline{m}_4'}\left[1+\left(\int\frac{\mathbb{E}\Big[\Big(\overline{\mathsf{Z}}^2-1\Big)Q(y\,|\,\mathsf{Z})\Big]^2}{\mathbb{E}[Q(y\,|\,\mathsf{Z})]}\,\mathrm{d}y\right)^{-1}\right].\tag{4.18}$$

上述条件与 [67, (11)] 一致；后者被猜想为计算上最优弱恢复阈值，即人们认为，如果违反（4.18），则不存在其他多项式时间算法能够实现正的渐近重叠。我们注意到，（4.18)（或等价地（4.15））给出的阈值相对于 X 的极限谱分布表现出显著的普适性，因为它仅通过 Λ_d^2 的前两个矩（或等价地，其前两个自由累积量）依赖于其概率分布。这意味着，对于极限谱分布的前两个矩相匹配的所有设计，该阈值（4.18）都保持不变。注 4.3（非高斯设计对谱阈值的影响）。回顾 [44, 注 5.1]，当 $X_{i,j}\overset{\text{i.i.d.}}{\sim}\mathcal{N}(0,1/d)$ 时，Λ_n^2 服从 Marchenko–Pastur 定律，其矩形自由累积量为 $\kappa_2=1$ 和 $\kappa_{2k}=0$ （$k\geq 2$）。将此代入（4.15），即可得到 i.i.d. 高斯设计中通过[71, 公式（20）] 确定的最优谱阈值。因此可以清楚地看出，非高斯设计对最优谱阈值的影响体现在（4.15）左侧的累积量比值 κ_4/κ_2^2 上。

4.2 谱初始化 GVAMP 及其状态演化

本节介绍谱初始化 GVAMP，并利用确定性的状态演化递推刻画其极限动力学。通用 GVAMP 迭代按照以下规则更新其迭代量：对于每个 $t\geq 0$,

$$r^{t+1}=\Phi_t(X^\top X)\widetilde{r}^t+\widetilde{\Phi}_t(X)^\top\widetilde{p}^t,\qquad \widetilde{r}^{t+1}=f_{t+1}(r^{t+1};\beta_*,\Theta),\tag{4.19a}$$

$$p^{t+1}=\Psi_t(XX^\top)\widetilde{p}^t+\widetilde{\Psi}_t(X)\widetilde{r}^t,\qquad \widetilde{p}^{t+1}=g_{t+1}(p^{t+1};z,\varepsilon,\Xi),\tag{4.19b}$$

采用初始化 $\widetilde{r}^0\in\mathbb{R}^d$， $\widetilde{p}^0\in\mathbb{R}^n$。矩阵 $\Theta\in\mathbb{R}^{d\times d'}$， $\Xi\in\mathbb{R}^{n\times n'}$ 是满足假设（A6）的侧信息。

(A6) 矩阵 $\Theta \in \mathbb{R}^{d\times d'}$，$\Xi \in \mathbb{R}^{n\times n'}$ 彼此独立，且独立于 O，Q 中的 假设（A2）。它们满足以下 Wasserstein 极限：

$$\begin{bmatrix}\beta_* & \Theta\end{bmatrix} \xrightarrow{W} \begin{bmatrix}\mathsf{B}_* & \Theta^\top\end{bmatrix}, \qquad \begin{bmatrix}z & \varepsilon & \Xi\end{bmatrix} \xrightarrow{W} \begin{bmatrix}\mathsf{Z} & \mathsf{E} & \Xi^\top\end{bmatrix},$$

因为 d，$n \to \infty$ 和 d'，n' 是固定的。

函数 Φ_t, Ψ_t：$\mathbb{R} \to \mathbb{R}$ 分别作用于其对称输入矩阵的每个特征值，并保留~ ~特征向量；而函数 Φ_t, Ψ_t：$\mathbb{R} \to \mathbb{R}$ 分别作用于其矩形输入矩阵的每个奇异值，并保留奇异向量。例如，对于对称矩阵 A, $\Phi_t(A)$ 可理解为 $\Phi_t(A) = \sum_i \Phi_t(\lambda_i(A))v_i(A)v_i(A)^\top$。函数 f_{t+1}：$\mathbb{R}^{2+d'} \to \mathbb{R}, g_{t+1}$：$\mathbb{R}^{3+n'} \to \mathbb{R}$ 逐行作用于其输入，例如，$f_{t+1}(r^{t+1}; \beta_*, \Theta) = \begin{bmatrix}f_{t+1}(r_1^{t+1}; \beta_*, 1, \Theta_{1,:}) \cdots & f_{t+1}(r_d^{t+1}; \beta_*, d, \Theta_{d,:})\end{bmatrix}^\top$。关于（4.19）形式化表述的动机，参见第 5.3 节。

初始化量 $\tilde{r}^0$、$\tilde{p}^0$ 取自第 4.1 节所述的谱估计量族。第 4.1 节具体而言，考虑一个矩阵 D，如 (4.1a) 所示，它由如下形式的预处理函数构造而成：(4.2)，该函数对应于某个 g。具体来说，令 γ 通过使用此 (4.2) 定义。 g我们将推导以主特征向量 (4.19) （的函数） $v_1(D)$初始化的 GVAMP 迭代的状态演化结果。我们考虑初始化量 $\tilde{r}_0 = f_0(r^0)$ 以及 $\tilde{p}^0 = g_0(p^0;\ y)$，其中

$$r^0 = s\, c_r \sqrt{d}\, v_1(D), \qquad p^0 = c_p\, \gamma(\gamma I_n + \overline{G})^{-1} X \sqrt{d}\, v_1(D), \tag{4.20}$$

with

$$\overline{G} = \mathrm{diag}(\overline{g}(y)), \qquad s = \mathrm{sgn}(\langle v_1(D), \beta_* \rangle), \qquad c_r = \frac{\sqrt{\rho + w_2}}{w_2}, \qquad c_p = \frac{\sqrt{\rho + w_2}}{w_1 w_3}. \tag{4.21}$$

常数 w_1、w_2、w_3 在 (4.21) 以及 w_4 如下给出：

$$w_1 := \rho \frac{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\overline{\kappa}_2\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\right] + \left(\frac{\overline{\kappa}_6}{\overline{\kappa}_2^2} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3} + \delta\frac{\overline{\kappa}_4}{\overline{\kappa}_2}\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2}{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)^2\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2 - \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)\mathbb{E}[\overline{g}(\mathsf{Y})^2]}, \tag{4.22}$$

$$w_2 := \frac{\rho}{\delta} \frac{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)^2\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\right] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)^2\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2 + \left(\frac{\overline{\kappa}_6}{\overline{\kappa}_2^3} - 2\frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^4}\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\right]}{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)^2\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2 - \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)\mathbb{E}[\overline{g}(\mathsf{Y})^2]}, \tag{4.23}$$

$$w_3 = 1 + \frac{\overline{\kappa}_4}{\delta\overline{\kappa}_2^2}, \qquad w_4 = \delta\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]. \tag{4.24}$$

注意，根据 (4.2) t hat $w_3 w_4 = \gamma$。此外，只要 w_1、w_2 满足 (4.13)，其分母均为正 s.

状态演化。(4.19) 的动态过程在高维极限下可由确定性的状态演化递推准确预测。事实上，下面的定理 4.3 表明，对于 $t \geq 1$，r^t 的经验分布收敛到 $b_t\mathsf{B}_* + \mathsf{K}_t$ 的概率分布，其中 b_t 是一个确定性标量，K_t 是一个与 B_* 独立、均值为零的高斯分布变量。类似地，p^t 的经验分布收敛到 $a_t\mathsf{Z} + \mathsf{J}_t$ 的概率分布，其中标量为 a_t，J_t 是一个与 Z 独立、均值为零的高斯分布变量。我们递归地定义这些状态演化参数如下。

Let

$$a_0 = \frac{1}{w_1}, \qquad b_0 = \frac{1}{w_2}. \tag{4.25}$$

定义

$$\mathsf{R}_0 = b_0\mathsf{B}_* + \mathsf{K}_0, \qquad \mathsf{P}_0 = a_0\mathsf{Z} + \mathsf{J}_0, \qquad \widetilde{\mathsf{R}}_0 = f_0(\mathsf{R}_0), \qquad \widetilde{\mathsf{P}}_0 = g_0(\mathsf{P}_0; \mathsf{Y}), \tag{4.26}$$

其中

$$(\mathsf{B}_*, \mathsf{K}_0) \sim P_{\mathsf{B}_*} \otimes \mathcal{N}(0, b_0), \qquad (\mathsf{Z}, \mathsf{J}_0) \sim \mathcal{N}(0, \sigma^2) \otimes \mathcal{N}(0, a_0), \qquad \mathsf{Y} \sim Q(\cdot|\mathsf{Z})。 \tag{4.27}$$

此外，令

$$\widetilde{a}_0 = \sigma^{-2}\mathbb{E}\Big[\widetilde{\mathsf{P}}_0\mathsf{Z}\Big], \qquad \widetilde{b}_0 = \rho^{-1}\mathbb{E}\Big[\widetilde{\mathsf{R}}_0\mathsf{B}_*\Big], \tag{4.28}$$

且 $\widetilde{\mathsf{P}}_{-1} = \gamma^{-1}\underline{g}(\mathsf{Y})\mathsf{P}_0$，$\widetilde{a}_{-1} = \gamma^{-1}a_0\mathbb{E}\Big[\underline{g}(\mathsf{Y})\mathsf{Z}^2\Big]$。

令 $(\mathsf{J}_r)_{0\le r\le t}$ 和 $(\mathsf{K}_r)_{0\le r\le t}$ 为两个彼此独立，且与 $\widetilde{\ }\ \widetilde{\ }(\mathsf{R}_0, \mathsf{P}_0, \mathsf{B}_*, \mathsf{Z}, \mathsf{E}\ \Lambda_n, \Lambda_d)$相互独立的中心化高斯过程，其协方差结构如下所示。$\mathsf{J}_0, \mathsf{K}_0$的概率分布见 (4.27)。对于任意 $t \ge 0$，J_{t+1}（分别为 K_{t+1}）与 J_0（分别为 K_0）之间的相关性为:

$$\begin{aligned}
\mathbb{E}[\mathsf{J}_{t+1}\mathsf{J}_0] &= \overline{\kappa}_2^{-1}\mathbb{E}\big[\Psi_t(\Lambda_n^2)\Lambda_n^2\big]\mathbb{E}\Big[\widetilde{\mathsf{P}}_t\widetilde{\mathsf{P}}_{-1}\Big] - 2\mathbb{E}\big[\Psi_t(\Lambda_n^2)\Lambda_n^2\big]\widetilde{a}_t\widetilde{a}_{-1}\rho + \overline{\kappa}_2^{-1}\mathbb{E}\big[\Psi_t(\Lambda_n^2)\Lambda_n^4\big]\widetilde{a}_t\widetilde{a}_{-1}\rho \\
&\quad + \mathbb{E}\Big[(\overline{\kappa}_2^{-1}\Lambda_n^2 - 1)\widetilde{\Psi}_t(\Lambda_n)\Lambda_n\Big]\widetilde{a}_{-1}\widetilde{b}_t\rho - a_{t+1}a_0\sigma^2, \qquad (4.29a) \\
\mathbb{E}[\mathsf{K}_{t+1}\mathsf{K}_0] &= \overline{\kappa}_2^{-1}\mathbb{E}\Big[\widetilde{\Phi}_t(\Lambda_d)\Lambda_d\Big]\mathbb{E}\Big[\widetilde{\mathsf{P}}_t\widetilde{\mathsf{P}}_{-1}\Big] + \mathbb{E}\Big[(\overline{\kappa}_2^{-1}\Lambda_d^2 - 1)\widetilde{\Phi}_t(\Lambda_d)\Lambda_d\Big]\widetilde{a}_t\widetilde{a}_{-1}\rho \\
&\quad + \overline{\kappa}_2^{-1}\mathbb{E}\big[\Phi_t(\Lambda_d^2)\Lambda_d^2\big]\widetilde{a}_{-1}\widetilde{b}_t\rho - b_{t+1}b_0\rho. \qquad (4.29b)
\end{aligned}$$

对于每个 $0 \le r,\ s \le t$,

$$\begin{aligned}
\mathbb{E}[\mathsf{J}_{r+1}\mathsf{J}_{s+1}] &= \mathbb{E}\big[\Psi_r(\Lambda_n^2)\Psi_s(\Lambda_n^2)\big]\Big(\mathbb{E}\Big[\widetilde{\mathsf{P}}_r\widetilde{\mathsf{P}}_s\Big] - \widetilde{a}_r\widetilde{a}_s\sigma^2\Big) + \mathbb{E}\big[\Psi_r(\Lambda_n^2)\Psi_s(\Lambda_n^2)\Lambda_n^2\big]\widetilde{a}_r\widetilde{a}_s\rho \\
&\quad + \mathbb{E}\Big[\widetilde{\Psi}_r(\Lambda_n)\widetilde{\Psi}_s(\Lambda_n)\Big]\mathbb{E}\Big[\widetilde{\mathsf{R}}_r\widetilde{\mathsf{R}}_s\Big] - a_{r+1}a_{s+1}\sigma^2 \\
&\quad + \mathbb{E}\Big[\Psi_r(\Lambda_n^2)\widetilde{\Psi}_s(\Lambda_n)\Lambda_n\Big]\widetilde{a}_r\widetilde{b}_s\rho + \mathbb{E}\Big[\Psi_s(\Lambda_n^2)\widetilde{\Psi}_r(\Lambda_n)\Lambda_n\Big]\widetilde{a}_s\widetilde{b}_r\rho, \qquad (4.29c) \\
\mathbb{E}[\mathsf{K}_{r+1}\mathsf{K}_{s+1}] &= \mathbb{E}\big[\Phi_r(\Lambda_d^2)\Phi_s(\Lambda_d^2)\big]\mathbb{E}\Big[\widetilde{\mathsf{R}}_r\widetilde{\mathsf{R}}_s\Big] - b_{r+1}b_{s+1}\rho \\
&\quad + \mathbb{E}\Big[\widetilde{\Phi}_r(\Lambda_d)\widetilde{\Phi}_s(\Lambda_d)\Big]\Big(\mathbb{E}\Big[\widetilde{\mathsf{P}}_r\widetilde{\mathsf{P}}_s\Big] - \widetilde{a}_r\widetilde{a}_s\sigma^2\Big) + \mathbb{E}\Big[\widetilde{\Phi}_r(\Lambda_d)\widetilde{\Phi}_s(\Lambda_d)\Lambda_d^2\Big]\widetilde{a}_r\widetilde{a}_s\rho \\
&\quad + \mathbb{E}\Big[\Phi_r(\Lambda_d^2)\widetilde{\Phi}_s(\Lambda_d)\Lambda_d\Big]\widetilde{a}_s\widetilde{b}_r\rho + \mathbb{E}\Big[\Phi_s(\Lambda_d^2)\widetilde{\Phi}_r(\Lambda_d)\Lambda_d\Big]\widetilde{a}_r\widetilde{b}_s\rho. \qquad (4.29d)
\end{aligned}$$

在 (4.29a) 至 (4.29d)中，标量 a_{r+1}，b_{r+1} 定义如下

$$\widetilde{a}_r = \frac{1}{\sigma^2}\mathbb{E}\Big[\widetilde{\mathsf{P}}_r\mathsf{Z}\Big], \qquad \widetilde{b}_r = \frac{1}{\rho}\mathbb{E}\Big[\widetilde{\mathsf{R}}_r\mathsf{B}_*\Big], \tag{4.29e}$$

$$a_{r+1} = \frac{1}{\overline{\kappa}_2}\Big(\mathbb{E}\big[\Psi_r(\Lambda_n^2)\Lambda_n^2\big]\widetilde{a}_r + \mathbb{E}\Big[\widetilde{\Psi}_r(\Lambda_n)\Lambda_n\Big]\widetilde{b}_r\Big), \qquad b_{r+1} = \mathbb{E}\Big[\widetilde{\Phi}_r(\Lambda_d)\Lambda_d\Big]\widetilde{a}_r, \tag{4.29f}$$

$\widetilde{\ }\qquad\qquad\widetilde{\ }$以及随机变量 $(\mathsf{P}_{r+1})_{0\le r\le t}$，$(\mathsf{R}_{r+1})_{0\le r\le t}$ 定义为

$$\widetilde{\mathsf{P}}_{r+1} = g_{r+1}(\mathsf{P}_{r+1}; \mathsf{Z}, \mathsf{E}, \Xi), \qquad \widetilde{\mathsf{R}}_{r+1} = f_{r+1}(\mathsf{R}_{r+1}; \mathsf{B}_*, \Theta), \tag{4.29g}$$

其中

$$\mathsf{P}_{r+1} = a_{r+1}\mathsf{Z} + \mathsf{J}_{r+1}, \qquad \mathsf{R}_{r+1} = b_{r+1}\mathsf{B}_* + \mathsf{K}_{r+1}. \tag{4.29h}$$

这就完全规定了高斯过程 $(\mathsf{J}_t)_{t\ge0}$，$(\mathsf{K}_t)_{t\ge0}$ 的概率分布。

现在，我们可以给出谱初始化 GVAMP 的状态演化结果。该结果断言，上文定义的随机变量确定了 (4.19) 的迭代量的 Wasserstein 极限。要使该结果成立，矩阵去噪器 Φ_t，Ψ_t 以及向量去噪器 f_{t+1}，g_{t+1} 必须满足以下条件。

(A7) Φ_t, Ψ_t 对每个 $t \geq 0$ 都是迹为零的，即 $\mathbb{E}\big[\Phi_t(\Lambda_d^2)\big] = \mathbb{E}\big[\Psi_t(\Lambda_n^2)\big] = 0$。此外，在 Λ_d 的概率分布下，Φ_t, $\widetilde{\Phi}_t \in \mathrm{PG}_{1\to 1}$ 以概率 1 连续；在 Λ_n 下，Ψ_t, $\Psi_t \in \mathrm{PG}_{1\to 1}$ 以概率 1 连续。

(A8) f_{t+1}, g_{t+1} 对于每个 $t \geq 0$，相对于其第一个输入均为散度为零，即 i s,

$$\mathbb{E}\big[f'_{t+1}(\mathsf{R}_{t+1}; \mathsf{B}_*, \Theta)\big] = 0, \qquad \mathbb{E}\big[g'_{t+1}(\mathsf{P}_{t+1}; \mathsf{Z}, \mathsf{E}, \Xi)\big] = 0,$$

其中 $f'_t{}_{+1}$, g'_{t+1} 表示 f_{t+1}, g_{t+1} 关于第一个参数的偏导数。

回顾 假设（A2）并利用 定理 K.3，假设（A7）可知，几乎必然有：

$$\lim_{d\to\infty} \frac{1}{d} \operatorname{Tr}(\Phi_t(X^\top X)) = \lim_{n\to\infty} \frac{1}{n} \operatorname{Tr}(\Psi_t(XX^\top)) = 0.$$

最后，我们提出 关于矩阵去噪器和向量去噪器正则性的两个技术假设 .

（A9）对于每个 $t \geq 0$,

$$\limsup_{d\to\infty} \big\|\Phi_t(\Lambda^\top \Lambda)\big\|_2, \qquad \limsup_{d\to\infty} \Big\|\widetilde{\Phi}_t(\Lambda)\Big\|_2, \qquad \limsup_{n\to\infty} \big\|\Psi_t(\Lambda\Lambda^\top)\big\|_2, \qquad \limsup_{n\to\infty} \Big\|\widetilde{\Psi}_t(\Lambda)\Big\|_2$$

都几乎必然是有限的。

（A10）对于每个 $t \geq 0$、f_{t+1}, g_{t+1} 关于所有变量均为 Lipschitz连续。此外，$r \mapsto f_{t+1}(r;\, b,\, \theta)$ 和 $p \mapsto g_{t+1}(p;\, z,\, e,\, \xi)$ 的偏导数相对于 $(\mathsf{R}_{t+1},\ \mathsf{B}_*,\ \Theta)$ 和 $(\mathsf{P}_{t+1},\ \mathsf{Z},\ \mathsf{E}\ \Xi)$ 的概率分布分别以概率 1 连续。

定理 4.3（谱初始化 GVAMP 的状态演化）。考虑 *GVAMP* 迭代 （4.19），其初始化为 $\widetilde{r}^0$, $\widetilde{p}^0$，给定于 （4.20）。若 假设 （A5） 至 （A10） 以及 （4.13）成立，则对于任意固定的 $t \geq 0$,

$$\begin{bmatrix} r^0 & \cdots & r^t & \widetilde{r}^0 & \cdots & \widetilde{r}^t & \beta_* & \Theta \end{bmatrix} \xrightarrow{W_2} \begin{bmatrix} \mathsf{R}_0 & \cdots & \mathsf{R}_t & \widetilde{\mathsf{R}}_0 & \cdots & \widetilde{\mathsf{R}}_t & \mathsf{B}_* & \Theta^\top \end{bmatrix}, \tag{4.30}$$

$$\begin{bmatrix} p^0 & \cdots & p^t & \widetilde{p}^0 & \cdots & \widetilde{p}^t & z & \varepsilon & \Xi \end{bmatrix} \xrightarrow{W_2} \begin{bmatrix} \mathsf{P}_0 & \cdots & \mathsf{P}_t & \widetilde{\mathsf{P}}_0 & \cdots & \widetilde{\mathsf{P}}_t & \mathsf{Z} & \mathsf{E} & \Xi^\top \end{bmatrix}. \tag{4.31}$$

定理 4.3 的完整证明推迟至 G 节。证明的简要概述见第 5.3 节。注 4.4（初始化器的选择）。式（4.20）中的初始化器 r^0 由谱估计量乘以一个缩放因子 c_r给出，而 p^0 则受到 5.4 节中基于线性化 GVAMP 的启发式方法启发。式（4.20）中的缩放因子 c_r 确保当 $d \to \infty$ 时，$r^0 \xrightarrow{W_2} \mathsf{R}_0 = b_0\mathsf{B}_* + \sqrt{b_0}\mathsf{G}$，其中 $\mathsf{G} \sim \mathcal{N}(0,1)$，且与 B_* 独立；见（4.26）。也就是说，信号强度和噪声方差均等于 b_0。p^0 中的缩放因子c_p 也有类似作用。对于采用谱初始化的一般 GVAMP 算法，这些特殊性质并不重要。然而，对于在下一节中获得谱初始化的 Bayes-GVAMP 的简洁状态演化，它们至关重要。

注 4.5（符号校准）。式（4.20）中的谱初始化器 r^0 包含式（4.21）中定义的符号因子 s，从而确保特征向量的符号满足 $\langle r^0, \beta_* \rangle \geq 0$。然而，$s$ 统计学家无法计算，因为其中涉及未知的 β_*。实际上，在没有符号校准的情况下，可以分别使用 $s = 1$ 和 $s = -1$ 运行两次状态演化，其中一次试验必然能够准确跟踪谱初始化 GVAMP 的动态。受 [28, 附录 H], 的启发，我们认为，在链接函数 q 和先验分布P_{B_*}具有额外结构（例如非对称性）时，s可通过矩方法等在统计上加以识别；否则则无法识别（另见 [48, 注 3.6]）。我们不再详细讨论这一问题，但要指出的是，对于理解以绝对值定义的重叠（1.1）而言，符号校准无关紧要；无论 $s \in \{-1, 1\}$取何值，我们的状态演化结果对渐近重叠给出的理论预测都依然准确。

4.3 谱初始化的 Bayes-GVAMP 及其状态演化

如果链接函数 q 以及 B_*、Λ、E 在 假设 (A1) 至 (A3) 中均已知，则可以选择定义 GVAMP 的矩阵去噪器和向量去噪器，以最大化 r^t 与 β_*之间的渐近重叠。现在我们介绍 GVAMP 的这一特殊情形，称为 Bayes-GVAMP；它使用 q、P_{B_*}、概率分布(Λ) P_{E} 来定义每次迭代中使用的去噪器。

定义两个函数 $g_{x1}\colon\ \mathbb{R}^2 \to \mathbb{R}$ 和 $g_{z1}\colon\ \mathbb{R}^3 \to \mathbb{R}$ 如下

$$g_{x1}(r_0;\gamma) = \frac{\int x P_{\mathsf{B}_*}(x)\exp\left(-\frac{\gamma}{2}x^2 + r_0 x\right)\mathrm{d}x}{\int P_{\mathsf{B}_*}(x)\exp\left(-\frac{\gamma}{2}x^2 + r_0 x\right)\mathrm{d}x} = \mathbb{E}\left[\mathsf{B}_* \,\middle|\, \mathsf{B}_* + \frac{1}{\sqrt{\gamma}}\mathsf{J} = \frac{r_0}{\gamma}\right], \tag{4.32}$$

$$g_{z1}(p_0, y_0;\tau) = \frac{\int z Q(y_0 \mid z)\exp\left(-\frac{\tau}{2}\left(z - \frac{p_0}{\tau}\right)^2\right)\mathrm{d}z}{\int Q(y_0 \mid z)\exp\left(-\frac{\tau}{2}\left(z - \frac{p_0}{\tau}\right)^2\right)\mathrm{d}z} = \mathbb{E}[\mathsf{Z} \mid \mathsf{P} = p_0, \mathsf{Y} = y_0], \tag{4.33}$$

对于任意 r_0、γ、p_0、y_0、$\tau \in \mathbb{R}$。(4.32) 和 (4.33) 中的期望是对概率分布如下所指定的随机变量取的。在 (4.32) 中，$(\mathsf{B}_*, \mathsf{J}) \sim P_{\mathsf{B}_*} \otimes \mathcal{N}(0,1)$。在 (4.33) 中，$\mathsf{Z} = \frac{1}{\tau}\mathsf{P} + \frac{1}{\sqrt{\tau}}\mathsf{K}$，其中 $(\mathsf{P}, \mathsf{K}) \sim \mathcal{N}(0, (\tau - \tau^{-2})\tau\sigma^2) \otimes \mathcal{N}(0, 1)$。此外，$\mathsf{Y} \sim Q(\cdot \mid \mathsf{Z})$。我们用 g'_{x1}、g'_{z1} 表示 g_{x1}、g_{z1}关于其第一个自变量的导数。经过初等计算可得

$$g'_{x1}(r_0;\gamma) = \mathrm{Var}\left[\mathsf{B}_* \,\middle|\, \mathsf{B}_* + \frac{1}{\sqrt{\gamma}}\mathsf{J} = \frac{r_0}{\gamma}\right], \qquad g'_{z1}(p_0, y_0;\tau) = \mathrm{Var}[\mathsf{Z} \mid \mathsf{P} = p_0, \mathsf{Y} = y_0]. \tag{4.34}$$

我们还定义函数 $g_{x2}\colon\ \mathbb{R}^d \times \mathbb{R}^n \times \mathbb{R}^2 \to \mathbb{R}^d$ 和 $g_{z2}\colon\ \mathbb{R}^d \times \mathbb{R}^n \times \mathbb{R}^2 \to \mathbb{R}^n$ 如下

$$g_{x2}(r, p;\gamma,\tau) := \frac{1}{\gamma}r + X^\top\left(\frac{\gamma}{\tau}I_n + XX^\top\right)^{-1}\left(\frac{1}{\tau}p - \frac{1}{\gamma}Xr\right), \tag{4.35a}$$

$$g_{z2}(r,\ p;\ \gamma,\ \tau) := X g_{x2}(r,\ p,\ \gamma,\ \tau), \tag{4.35b}$$

对于任意 $r \in \mathbb{R}^d$，$p \in \mathbb{R}^n$ 和 γ，$\tau \in \mathbb{R}$。

Bayes-GVAMP 算法维护一组迭代量，并在每个 $t \geq 1$ 时按如下方式更新：

$$\widehat{x}_1^t = g_{x1}(r^t;\gamma_1^t) \in \mathbb{R}^d, \qquad \widehat{z}_1^t = g_{z1}(p^t, y;\tau_1^t) \in \mathbb{R}^n, \tag{4.36a}$$

$$v_1^t = \mathbb{E}\left[g'_{x1}(\mathsf{R}_t;\gamma_1^t)\right] \in \mathbb{R}, \qquad c_1^t = \mathbb{E}\left[g'_{z1}(\mathsf{P}_t, \mathsf{Y};\tau_1^t)\right] \in \mathbb{R}, \tag{4.36b}$$

$$\widetilde{r}^t = \frac{1}{v_1^t}\widehat{x}_1^t - r^t \in \mathbb{R}^d, \qquad \widetilde{p}^t = \frac{1}{c_1^t}\widehat{z}_1^t - p^t \in \mathbb{R}^n, \tag{4.36c}$$

$$\gamma_2^t = \frac{1}{v_1^t} - \gamma_1^t \in \mathbb{R}, \qquad \tau_2^t = \frac{1}{c_1^t} - \tau_1^t \in \mathbb{R}, \tag{4.36d}$$

$$\widehat{x}_2^t = g_{x2}\ (\widetilde{r}^t,\ \widetilde{p}^t;\ \gamma_2^t,\ \tau_2^t)\ \in \mathbb{R}^d \qquad \widehat{z}_2^t = g_{z2}(\widetilde{r}^t, \widetilde{p}^t;\gamma_2^t, \tau_2^t) \in \mathbb{R}^n \tag{4.36e}$$

$$v_2^t = \mathbb{E}\left[\frac{1}{\tau_2^t\Lambda_d^2 + \gamma_2^t}\right] \in \mathbb{R}, \qquad c_2^t = \mathbb{E}\left[\frac{\Lambda_n^2}{\tau_2^t\Lambda_n^2 + \gamma_2^t}\right] \in \mathbb{R}, \tag{4.36f}$$

$$r^{t+1} = \frac{1}{v_2^t}\widehat{x}_2^t - \widetilde{r}^t \in \mathbb{R}^d, \qquad p^{t+1} = \frac{1}{c_2^t}\widehat{z}_2^t - \widetilde{p}^t \in \mathbb{R}^n, \tag{4.36g}$$

$$\gamma_1^{t+1} = \frac{1}{v_2^t} - \gamma_2^t \in \mathbb{R}, \qquad \tau_1^{t+1} = \frac{1}{c_2^t} - \tau_2^t \in \mathbb{R}, \tag{4.36h}$$

其中

$$\mathsf{R}_t = \gamma_1^t\mathsf{B}_* + \sqrt{\gamma_1^t}\mathsf{J}_t, \qquad \mathsf{P}_t = \left(\tau_1^t - \frac{1}{\sigma^2}\right)\mathsf{Z} + \sqrt{\tau_1^t - \frac{1}{\sigma^2}}\mathsf{K}_t, \tag{4.37}$$

$\overline{\mathsf{J}}_t \sim \mathcal{N}(0,1)$ 与 B_*相互独立，且 $\mathsf{K}_t \sim \mathcal{N}(0,1)$ 与 Z、E相互独立。上述迭代以 r^0、p^0 给定，见 （4.20）和

$$\gamma_1^0 = \frac{1}{w_2}, \qquad \tau_1^0 = \frac{1}{w_1} + \frac{1}{\sigma^2}. \tag{4.38}$$

Bayes-GVAMP 迭代 （4.36）可以证明是（4.19）中的通用 GVAMP 迭代的一个特殊情形。利用 g_{x2} 的定义，g_{z2} 并对 （4.36）进行基本运算，可得

$$\begin{aligned} r^{t+1} &= \left[\left(\frac{1}{v_2^t\gamma_2^t} - 1\right) I_d - \frac{1}{v_2^t\gamma_2^t} X^\top \left(\frac{\gamma_2^t}{\tau_2^t} I_n + XX^\top\right)^{-1} X\right] \left(\frac{1}{v_1^t}\widehat{x}_1^t - r^t\right) \\ &\quad + \frac{1}{v_2^t\tau_2^t} X^\top \left(\frac{\gamma_2^t}{\tau_2^t} I_n + XX^\top\right)^{-1} \left(\frac{1}{c_1^t}\widehat{z}_1^t - p^t\right), \end{aligned} \tag{4.39a}$$

$$\begin{aligned} p^{t+1} &= \left[\frac{1}{c_2^t\tau_2^t} XX^\top \left(\frac{\gamma_2^t}{\tau_2^t} I_n + XX^\top\right)^{-1} - I_n\right] \left(\frac{1}{c_1^t}\widehat{z}_1^t - p^t\right) \\ &\quad + \frac{1}{c_2^t\gamma_2^t} X \left[I_d - X^\top \left(\frac{\gamma_2^t}{\tau_2^t} I_n + XX^\top\right)^{-1} X\right] \left(\frac{1}{v_1^t}\widehat{x}_1^t - r^t\right). \end{aligned} \tag{4.39b}$$

这相当于令（4.19）中的去噪器为

$$\begin{aligned} \Phi_t(x) &= \left(\frac{1}{v_2^t\gamma_2^t} - 1\right) - \frac{1}{v_2^t\gamma_2^t} \cdot \frac{x}{\frac{\gamma_2^t}{\tau_2^t} + x}, & \widetilde{\Phi}_t(x) &= \frac{1}{v_2^t\tau_2^t} \cdot \frac{x}{\frac{\gamma_2^t}{\tau_2^t} + x^2}, \\ &= \frac{1}{v_2^t} \cdot \frac{1}{\gamma_2^t + \tau_2^t x} - 1, & &= \frac{1}{v_2^t} \cdot \frac{x}{\gamma_2^t + \tau_2^t x^2}, \end{aligned} \tag{4.39c}$$

$$\begin{aligned} \Psi_t(x) &= \frac{1}{c_2^t\tau_2^t} \cdot \frac{x}{\frac{\gamma_2^t}{\tau_2^t} + x} - 1, & \widetilde{\Psi}_t(x) &= \frac{1}{c_2^t\gamma_2^t} \cdot x\left(1 - \frac{x^2}{\frac{\gamma_2^t}{\tau_2^t} + x^2}\right), \\ &= \frac{1}{c_2^t} \cdot \frac{x}{\gamma_2^t + \tau_2^t x} - 1, & &= \frac{1}{c_2^t} \cdot \frac{x}{\gamma_2^t + \tau_2^t x^2}, \end{aligned} \tag{4.39d}$$

$$f_t(r) = \frac{1}{v_1^t} g_{x1}(r;\gamma_1^t) - r, \qquad g_t(p;y) = \frac{1}{c_1^t} g_{z1}(p,y;\tau_1^t) - p. \tag{4.39e}$$

上述矩阵去噪器 Φ_t，Ψ_t 以及向量去噪器 f_t，g_t 确实满足假设（A7）和（A8）所要求的迹为零和散度为零条件，定理 H.1对此作了证明。

定理 4.4 （谱初始化的 Bayes-GVAMP 的状态演化）。 考虑 （4.36）中的 *Bayes-GVAMP* 算法，并用 （4.20）进行初始化。那么，对于任意固定的 $t \geq 0$，

$$\begin{bmatrix} r^t & \beta_* \end{bmatrix} \xrightarrow{W_2} \begin{bmatrix} \mathsf{R}_t & \mathsf{B}_* \end{bmatrix}, \qquad \begin{bmatrix} p^t & z & \varepsilon \end{bmatrix} \xrightarrow{W_2} \begin{bmatrix} \mathsf{P}_t & \mathsf{Z} & \mathsf{E} \end{bmatrix},$$

其中右侧的随机变量定义于（4.37）中。

定理 4.4 在H 节中通过对 t 进行归纳证明，并在（4.39）的配置下将状态演化递推 （4.36）简化为（4.29）。（4.39）。 图 3 通过数值结果验证，谱初始化的 Bayes-GVAMP （4.36）所实现的重叠量轨迹可由其状态演化准确追踪。可以观察到，即使在 10 步的中等 δ 下，Bayes-GVAMP 通常也能快速收敛。

4.4 推测的贝叶斯风险

本节回顾 [67] 中使用非严格的复制子方法推导出的贝叶斯风险，并证明 Bayes-GVAMP 具有相同的不动点。定义两个函数 $\mathcal{Z}_0$： $\mathbb{R}^2 \to \mathbb{R}_{\geq 0}$和 $\mathcal{Z}_{\text{out}}$： $\mathbb{R}^2 \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$

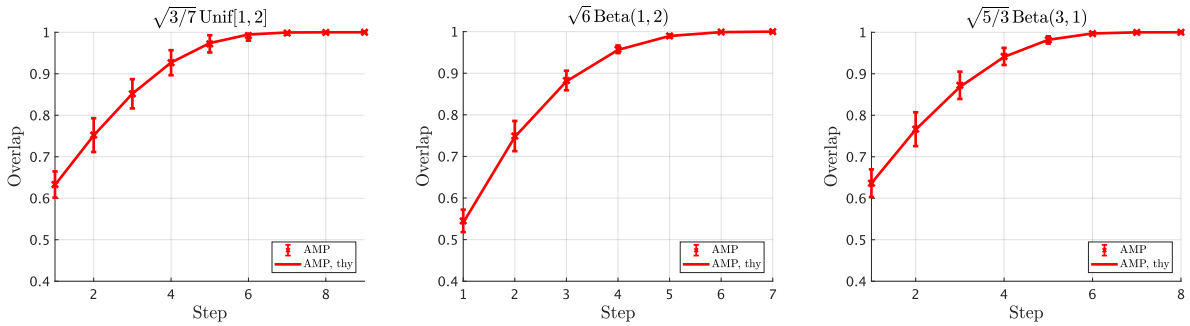

图 3：谱初始化的 Bayes-GVAMP 在相位恢复问题中的(4.36) 轨迹可由其状态演化准确追踪。图中绘制了3 种谱分布在(6.3) 处的轨迹 $\delta = 2$。

as

$$\mathcal{Z}_0(b, a) = \int P_{\mathsf{B}_*}(x) \exp\left(-\frac{1}{2}ax^2 + bx\right) \mathrm{d}x, \qquad \mathcal{Z}_{\mathrm{out}}(y; \xi, v) = \mathbb{E}\big[Q\big(y \,|\, \sqrt{v}\,\Omega + \xi\big)\big], \tag{4.40}$$

其中，期望是对 $\Omega \sim \mathcal{N}(0, 1)$ 取的。分别用 $\partial_b \log \mathcal{Z}_0(b, \cdot)$ 和 $\partial_\xi \log \mathcal{Z}_{\mathrm{out}}(\cdot; \xi, \cdot)$ 表示映射 $b \mapsto \log \mathcal{Z}_0(b, \cdot)$ 和 $\xi \mapsto \log \mathcal{Z}_{\mathrm{out}}(\cdot; \xi, \cdot)$ 的导数。

考虑以下关于变量 q_x、q_y、$\widehat{q}_x$、$\widehat{q}_y$、γ_x、γ_y 的方程组：

$$q_x = \mathbb{E}\left[\mathcal{Z}_0\left(\sqrt{\widehat{q}_x}\,\Xi, \widehat{q}_x\right)\left(\partial_b \log \mathcal{Z}_0\left(\sqrt{\widehat{q}_x}\,\Xi, \widehat{q}_x\right)\right)^2\right], \tag{4.41a}$$

$$q_z = \frac{1}{\widehat{Q}_z + \widehat{q}_z}\left(\frac{\widehat{q}_z}{\widehat{Q}_z} + \mathbb{E}\left[\int \mathcal{Z}_{\mathrm{out}}\left(y; \sqrt{\frac{\widehat{q}_z}{\widehat{Q}_z(\widehat{Q}_z + \widehat{q}_z)}}\,\Xi, \frac{1}{\widehat{Q}_z + \widehat{q}_z}\right)\left(\partial_\xi \log \mathcal{Z}_{\mathrm{out}}\left(y; \sqrt{\frac{\widehat{q}_z}{\widehat{Q}_z(\widehat{Q}_z + \widehat{q}_z)}}\,\Xi, \frac{1}{\widehat{Q}_z + \widehat{q}_z}\right)\right)^2 \mathrm{d}y\right]\right), \tag{4.41b}$$

$$\widehat{q}_x = \frac{q_x}{\rho(\rho - q_x)} - \gamma_x, \tag{4.41c}$$

$$\widehat{q}_z = \frac{q_z}{Q_z(Q_z - q_z)} - \gamma_z, \tag{4.41d}$$

$$\rho - q_x = \mathbb{E}\left[\frac{1}{\rho^{-1} + \gamma_x + \gamma_z \Lambda_d^2}\right], \tag{4.41e}$$

$$Q_z - q_z = \mathbb{E}\left[\frac{\Lambda_n^2}{\rho^{-1} + \gamma_x + \gamma_z \Lambda_n^2}\right], \tag{4.41f}$$

~其中，前两个方程中的期望是对 $\Xi \sim \mathcal{N}(0, 1)$ 取的，常数 Q_z、Q_z 定义为

$$Q_z = \frac{\rho \mathbb{E}\big[\Lambda_d^2\big]}{\delta}, \qquad \widehat{Q}_z = \frac{1}{Q_z}. \tag{4.42}$$

我们还记得，ρ 是 P_{B_*} (假设（A1）) 的方差。文献中将方程组 (4.41)称为副本鞍点方程 [67]。

Define $\widehat{\beta}_{\mathrm{B}} \equiv \widehat{\beta}_{\mathrm{B}}(X, y) \in \mathbb{R}^d$ （下标 B 表示“贝叶斯”）作为 β_* 的后验均值 ：

$$\widehat{\beta}_{\mathrm{B}} = \mathbb{E}[\beta_* \,|\, X, y] = \frac{\int_{\mathbb{R}^d} \beta P_{\mathsf{B}_*}^{\otimes d}(\beta) Q^{\otimes d}(y \,|\, X\beta)\, \mathrm{d}\beta}{\int_{\mathbb{R}^d} P_{\mathsf{B}_*}^{\otimes d}(\beta) Q^{\otimes d}(y \,|\, X\beta)\, \mathrm{d}\beta}. \tag{4.43}$$

^众所周知，β_{B} 对于广义线性模型(3.5)，β_* 在给定(X, y) 时，β_{B} 是估计 β_* 的均方误差最小的估计量。

猜想 4.5 ([67, 猜想 2.1])。 存在一个元组

$$(\widehat{q}_x^*, \widehat{q}_z^*, \gamma_x^*, \gamma_z^*, q_x^*, q_z^*) \in [0,\infty) \times [0,\infty) \times [0,\infty) \times [0,\infty) \times [0,\rho] \times [0, Q_z]$$

求解(4.41)使得

$$\lim_{d\to\infty} \frac{1}{d}\mathbb{E}\left[\left\|\beta_* - \widehat{\beta}_{\mathrm{B}}\right\|_2^2\right] = \rho - q_x^*,$$

其中期望是关于 β_*、 X、 y的。

当(4.41) 的解不唯一时， q_x^* 通过最大化某个自由熵泛函得到。由于其复杂性，我们不正式定义该泛函，而是请读者参阅 [67, 猜想2。1]。接下来，我们验证(4.41) 的解集与(4.36)中Bayes-GVAMP的状态演化递推的不动点集一一对应。

定理 4.6 不动点集

$$(\gamma_1, \tau_1, \gamma_2, \tau_2, v_1, v_2, c_1, c_2) \in [0,\infty) \times \left[\frac{1}{\sigma^2}, \infty\right) \times \left[\frac{1}{\rho}, \infty\right) \times [0,\infty) \times [0,\rho] \times [0,\rho] \times [0,\sigma^2] \times [0,\sigma^2] \quad (4.44)$$

关于$(\gamma_1^t,\ \tau_1^t,\ \gamma_2^t,\ \tau_2^t,\ v_1^t,\ v_2^t,\ c_1^t,\ c_2^t)$ 在(4.36) 中的递推，与元组集合相同

$$\left(\widehat{q}_x, \widehat{q}_z + \frac{1}{\sigma^2}, \gamma_x + \frac{1}{\rho}, \gamma_z, \rho - q_x, \rho - q_x, \sigma^2 - q_z, \sigma^2 - q_z\right) \quad (4.45)$$

使得

$$(\widehat{q}_x, \widehat{q}_z, \gamma_x, \gamma_z, q_x, q_z) \in [0,\infty) \times [0,\infty) \times [0,\infty) \times [0,\infty) \times [0,\rho] \times [0, Q_z]$$

是副本鞍点方程 (4.41)的解。

定理 4.6 在第 I 节 中证明，方法是验证(4.36) 与(4.41) 在(4.45)所示变量替换下的等价性。

定理 4.6 中的不动点元组可能并不唯一。由于 (4.44) 与 (4.45) 等价，我们只考察 (4.44)。例如，考虑带有高斯先验分布 $P_{\mathrm{B}_*} = \mathcal{N}(0, 1)$ 的（无噪声）相位恢复模型 $q(z, \varepsilon) = |z|$。将定理 4.5 专门应用于这一情形，猜想渐近贝叶斯风险为 $v \in [0, 1]$，其满足方程 $\mathcal{F}(v) = v$，其中 $\mathcal{F}$ 是某个函数；关于 $\mathcal{F}$ 的精确定义及这一结果的推导，见第 J.1 节。图 4 展示了在 3 种不同的谱分布 Λ（其定义见第 6.1 节）以及某些 δ 取值下，解 v 的非唯一性。

在上述相同设定下，我们还在图 5 中绘出了谱初始化的 Bayes-GVAMP 在收敛时达到的重叠度（即其状态演化的重叠度）、猜想的贝叶斯最优重叠度 $\sqrt{1 - v/\rho}$（可通过对定理 4.5应用条件期望的塔式性质得到），以及作为对比，由定理 4.2给出的谱估计量的渐近重叠。可以观察到，当解(4.36) 唯一时，谱初始化的 Bayes-GVAMP 达到了猜想的贝叶斯最优重叠度；否则，它达到的重叠度对应于一个较差的解 v。 v

对于非结构化的独立同分布高斯设计，已有充分观察表明 [5]，存在多个状态演化不动点，并由此导致最优 AMP 算法 [87, 73] 与贝叶斯最优后验均值（4.43）之间出现性能差距，这种情况可能发生在存在结构化先验分布时。这里，图 4和5 均仅针对非结构化高斯先验。因而，我们的发现补充了上述观察，指出了多个不动点和性能差距的另一种成因：结构化设计。我们注意到，这种性能差距可能非常显著。事实上，在图 5最左侧和最右侧的图中，存在一段 δ 范围，在此范围内，后验均值能够实现完美恢复，而 Bayes-GVAMP 则表现不佳。

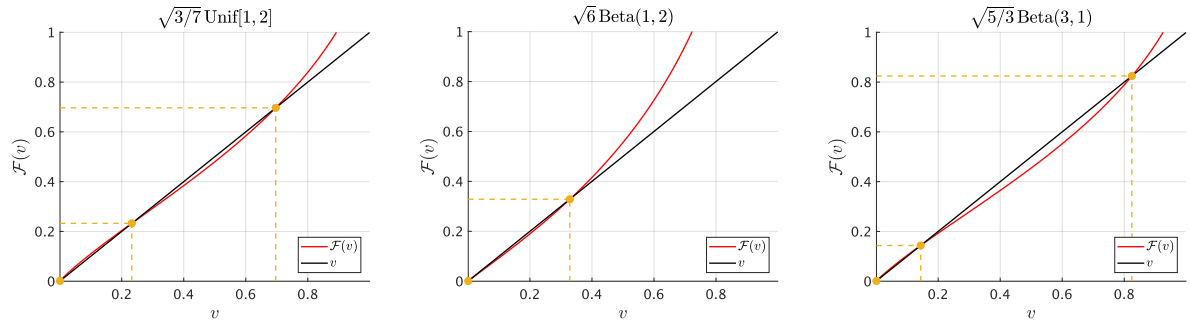

图 4：函数 $\mathcal{F}$ 及其不动点（即方程 $v \in [0,\rho]\quad v = \mathcal{F}(v)$的解），对应于3 个谱分布(6.3)。从最左侧到最右侧的面板中，δ 分别取值为1.44、1、1.34。

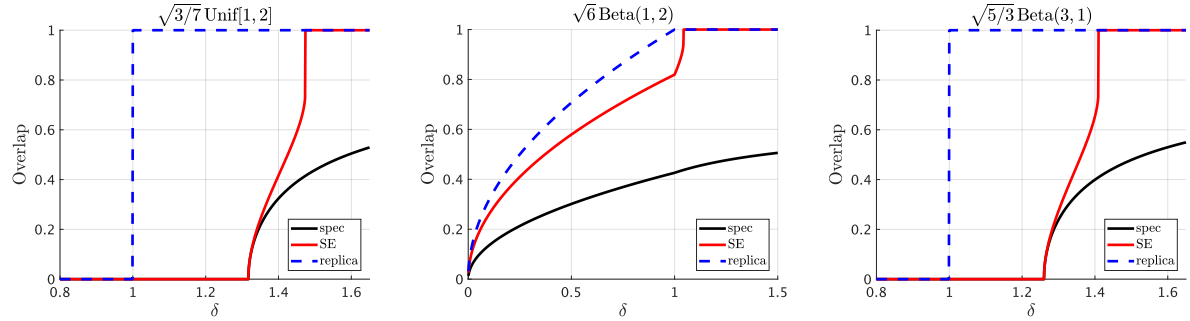

图 5：定理 4.2 (4.11)中的谱估计量所达到的渐近重叠（黑色的“spec”），谱初始化的 Bayes-GVAMP 的状态演化(4.36) 迭代至收敛后所达到的重叠（红色的“SE”），以及用重叠表示的推测贝叶斯风险(4.41)（蓝色的“replica”）。所有曲线均针对3 个(6.3)中给出的谱分布绘制。

5 启发式方法与证明概述

在第 5.1 节中，我们探讨线性化 GVAMP 与谱估计量之间的启发式联系，从而为如下形式的谱估计量提供动机：（4.1），并满足假设 (A5)。在第 5.2 节中，我们重点介绍随机矩阵 D 谱分析中的关键难点，并概述定理 4.1的证明。第 5.3 节为通用 GVAMP 形式体系提供动机。在5.4 节中，我们进一步阐释 GVAMP 中（4.20）所用谱初始化的设计思路。

5.1 使用线性化 GVAMP 模拟谱估计量

考虑如下采用线性去噪函数的 GVAMP 算法：

$$u^t = \big(\overline{\kappa}_2^{-1} XX^\top - I_n\big) g(u^{t-1}), \qquad v^t = \overline{\kappa}_2^{-1} X^\top g(u^{t-1}), \tag{5.1}$$

其中

$$g(u) = \left(\frac{G}{\mathrm{Tr}(G)/n} - I_n\right) u,$$

且 $G = \mathrm{diag}(\hat{g}(y))$ 是通过对 $\hat{g}\colon \mathbb{R} \to \mathbb{R}$ 逐分量应用任意函数 y 得到的对角矩阵。如C.1 节所示，(5.1) 是(4.19) 中 GVAMP 算法的一个特殊情形，也是 [101]中的 RI-GAMP 算法的一个特殊情形。根据 [66, 方程（16）], 对 G作特定选择后，迭代(5.1) 可视为 Bayes-GVAMP 算法在其平凡不动点附近的线性化结果 [91]（另见(4.36)）。在

在采用部分 Haar 设计的广义线性模型背景下，可通过期望传播推导出类似的迭代 [62, 方程（21a）]，并利用它对谱估计量进行启发式分析。

请注意，上述 u^t 迭代作用于迹趋于零的随机矩阵 ：

$$\lim_{n\to\infty}\frac{1}{n}\operatorname{Tr}\left(\kappa_2^{-1}XX^\top - I_n\right) = 0.$$

此外，去噪函数 g 的散度为零，因为

$$\operatorname{Tr}\left(\frac{G}{\operatorname{Tr}(G)/n} - I_n\right) = 0.$$

对于较大的 t，我们预期 $u^t/\|u^t\|_2$ 会按以下意义收敛：

$$\lim_{t\to\infty}\lim_{n\to\infty}\frac{\left|\langle u^t, u^{t-1}\rangle\right|}{\|u^t\|_2\|u^{t-1}\|_2} = 1. \tag{5.2}$$

换言之，假设极限

$$\gamma = \lim_{t\to\infty}\lim_{d\to\infty}\frac{\langle u^t, u^{t-1}\rangle}{\|u^{t-1}\|_2^2} \tag{5.3}$$

存在，那么我们预期对于较大的 t，相邻的两个迭代量共线，即

$$\lim_{t\to\infty}\lim_{n\to\infty}\frac{\left\|u^t - \gamma u^{t-1}\right\|_2}{\|u^{t-1}\|_2} = 0. \tag{5.4}$$

请注意，$\gamma \equiv \gamma(\hat{g})$ 取决于函数 $\hat{g}$。需要注意的是，$\gamma = 1$ 未必成立；在这种情况下，(5.2)或 (5.4) 并不能推出 $\lim_{t\to\infty}\lim_{n\to\infty} n^{-1}\left\|u^t - u^{t-1}\right\|_2^2 = 0$。更多细节请参见 定理 B.1 之后的讨论。迭代 (5.1) 的不动点于是可写为：

$$\gamma\hat{u} = \left(\overline{\kappa}_2^{-1}XX^\top - I_n\right)\left(\frac{G}{\operatorname{Tr}(G)/n} - I_n\right)\hat{u} \tag{5.5}$$

对于某个单位向量 $\hat{u}\in\mathbb{S}^{n-1}$。该方程与 D 的一个特征方程密切相关，其形式为 (4.1a)(相对于某个 $\mathcal{T}$，具体形式将在下文给出)。事实上，重新整理各项可得

$$\left[(\gamma-1)I_n + \frac{G}{\operatorname{Tr}(G)/n}\right]\hat{u} = \left(\overline{\kappa}_2^{-1}XX^\top\right)\left(\frac{G}{\operatorname{Tr}(G)/n} - I_n\right)\hat{u}.$$

对两边左乘

$$X^\top\left(\frac{G}{\operatorname{Tr}(G)/n} - I_n\right)\left[(\gamma-1)I_n + \frac{G}{\operatorname{Tr}(G)/n}\right]^{-1},$$

可得

$$X^\top\left(\frac{G}{\operatorname{Tr}(G)/n} - I_n\right)\hat{u} = \overline{\kappa}_2^{-1}X^\top\left(\frac{G}{\operatorname{Tr}(G)/n} - I_n\right)\left[(\gamma-1)I_n + \frac{G}{\operatorname{Tr}(G)/n}\right]^{-1}XX^\top\left(\frac{G}{\operatorname{Tr}(G)/n} - I_n\right)\hat{u}.$$

我们注意到，这是该矩阵的特征方程

$$X^\top\left(\frac{G}{\operatorname{Tr}(G)/n} - I_n\right)\left[(\gamma-1)I_n + \frac{G}{\operatorname{Tr}(G)/n}\right]^{-1}X \tag{5.6}$$

对应于特征向量（允许重新缩放）

$$X^\top\left(\frac{G}{\mathrm{Tr}(G)/n}-I_n\right)\widehat{u}. \tag{5.7}$$

矩阵(5.6) 渐近等价于 D 中(4.1a) 的 $\mathcal{T}$取(4.2) 的形式，且 g 由下式给出

$$\overline{g}(y)=\frac{\widehat{g}(y)}{\mathbb{E}[\widehat{g}(\mathsf{Y})]}-1,$$

这显然满足 $\mathbb{E}[g(\mathsf{Y})]=0$ 所要求的条件(4.3)。我们指出，γ 在(5.3) 中的值可通过状态演化推导得出（参见附录 B，更具体地说，参见定理 E.1），从而得到(4.2)中的显式表达式。这启发我们针对满足(4.1) 的预处理函数构造谱估计量，这些函数满足假设 (A5)。

5.2 D 的分析与定理 4.1的证明概述

我们对矩阵 D （见式 (4.1a)）的分析不同于标准的随机矩阵理论方法。回顾假设（A2），可将该矩阵写为 $D=Q\Lambda^\top O^\top TO\Lambda Q^\top$，其形式上类似于尖峰乘法模型 [36]。然而，两者有三点重要区别。在标准尖峰乘法模型中，通常假设：(a) Λ、T与 Haar 矩阵 O、Q相互独立；(b) T为半正定矩阵，这一性质带来便利，使得 $Q\Lambda^\top O^\top TO\Lambda Q^\top$ 和 $T^{1/2}O\Lambda\Lambda^\top O^\top T^{1/2}$的谱仅在 0 处不同；(c) Λ、T在固定位置存在谱离群值。最后一个假设也可等价地理解为：Λ、T 上的低秩尖峰与对应于给定离群特征值的特征向量对齐。在我们的设定中，(a)、(b)、(c) 均不成立：T 涉及来自O、Q 的 Haar 随机性，且可以取负值（见定理 4.2）；尖峰则通过沿未知参数向量 β_*的一个一维投影$X\beta_*$ 进入 T。因此，关于尖峰乘法模型的现有结果并不适用。相反，我们结合近似消息传递、随机矩阵理论和自由概率中的工具，利用并严格论证第 5.1 节中谱估计量与线性化 GVAMP 之间的联系。

在 D 的定理 4.1中，对渐近谱性质进行精确刻画分为三个步骤。首先，受第 5 节中的线性化启发式方法启发，我们考虑(5.1) 中的一个 GVAMP 实例(4.19)，推导其状态演化并研究其不动点。这些结果总结于定理 B.1 (证明见第 C 节）。其次，借助随机矩阵理论和自由概率中的工具，我们确定 D的第二个（非离群）特征值的极限；见定理 B.2 (证明见D 节）。以上述两个要素为基础，论证的关键在于定理 B.3 (证明见第 E 节），该定理表明，线性化 GVAMP 迭代量渐近地与感兴趣的 $v_1(D)$ 对齐，并且相应的离群特征值 $\lambda_1(D)$ 可以被准确定位。由此，调用第一步中的状态演化结果，即可计算谱估计量达到的渐近重叠。

我们将更详细地讨论如何刻画 D (定理 B.2) 中谱主体右边缘的证明策略，因为这一问题在随机矩阵理论和自由概率中本身就很有研究价值。本部分又分为三个步骤。 D.1 节 为研究 D 的“零”对应情形奠定基础；对于这一情形，第 3.2 节 和 第 4.1 节 中的一些假设可以放宽。粗略地说，X 和 T 在零矩阵中相互独立。D.2 节 证明，零矩阵的前 k 个特征值（对于任意固定的 $k\geq 1$ 相对于 n，d）都收敛到其极限谱分布支撑集的上确界。这类结果称为强收敛 [52]。根据特征值交错（定理 D.1），这立即意味着 D 的第二大特征值具有相同的极限。事实上，这一步的一个关键部分建立了关于 Haar 正交矩阵与确定性矩阵强渐近自由性的一个更一般的结论，该结论蕴含涉及此类矩阵的矩阵乘积强收敛。我们通过函数演算，将问题化归到已知的高斯正交系综与确定性矩阵强渐近自由性结果 [46, 定理 4。2]。极限谱分布可用自由乘法卷积表示，但目前似乎还没有对其右边缘的刻画。在D.3 节中，我们借助复分析和自由概率的工具，对更一般的无限维算子自由卷积的右边缘给出一种显式刻画

，其形式是某个泛函的最大临界点。粗略而言，这个泛函可以理解为自由卷积的斯蒂尔切斯逆变换。最后，D.4 节 将这一显式刻画专门应用于 第 3.2 节 和第 4.1 节 的情形，在该情形下还可以进一步简化，最终得到 定理 B.2中的刻画。

5.3 通用GVAMP形式体系及定理 4.3的证明概述

我们的通用 GVAMP 形式体系见 (4.19) ，其灵感来自多个来源（参见第2节），其中在思路上最接近的是 [43]。对于数据矩阵 $X \in \mathbb{R}^{d\times d}$ 为对称矩阵的情形， $r^{t+1} = \Phi_t(X)\, f_t(r^t)$ 中引入了 VAMP 迭代的一种抽象形式， [43], 其中 Φ_t 迹为零，且 f_t 在与假设 (A7) 和(A8)相同的意义下分别满足散度为零。这两个条件是在不损失最优性的前提下获得易处理的状态演化的关键。在广义线性模型等 $X \in \mathbb{R}^{n\times d}$ 通常为矩形矩阵的情形中，我们的 GVAMP 形式体系可理解为作用于增广分块矩阵 $Y_t \in \mathbb{R}^{(n+d)\times(n+d)}$:

$$\begin{bmatrix} r^{t+1} \\ p^{t+1} \end{bmatrix} = \underbrace{\begin{bmatrix} \Phi_t(X^\top X) & \widetilde{\Phi}_t(X)^\top \\ \widetilde{\Psi}_t(X) & \Psi_t(XX^\top) \end{bmatrix}}_{Y_t} \begin{bmatrix} f_t(r^t) \\ g_t(p^t) \end{bmatrix}, \tag{5.8}$$

同时仍要求向量去噪器满足散度为零条件，但只要求 Y_t的对角块满足迹为零条件。注意，(5.8)不能化为 [43] 中研究的对称版本，原因有二。首先， Y_t 在 $\mathbb{O}(n+d)$ 的共轭作用下并不保持不变，而是在$\mathbb{O}(n)$ 的共轭作用下保持分块不变$\times\mathbb{O}(d)$。其次， Y_t 不是对称的，因为 Φ_t和 Ψ_t 可能不同。这种灵活性在广义线性模型的应用中特别有用，因为 Φ_t和 Ψ_t 的最优选择确实不同。粗略地说，这是因为统计模型 (3.4) 只揭示了 y ，却没有揭示 β_*，最终打破了 (5.8) 中n维与 d维迭代量之间的对称性。GVAMP 的状态演化结果（定理 4.3）的证明遵循 [44] 的迭代式Haar条件化策略，并仔细处理特殊向量β_* 和 y （或更一般地， z、ε），它们可作为 f_t 和 g_t 的输入（尽管在 (5.8) 的记号中省略了这些向量）。形式体系 (5.8) 很容易看出与 (4.19) 等价，并且足够一般，能够涵盖最初在 [91] 中通过启发式方法推导出的 Bayes-GVAMP (4.36)。本文建立的一般理论能够严谨地追踪采用谱初始化的 GVAMP 的动态。

5.4 设计谱初始化

为了使用谱估计量正确初始化 GVAMP， d 维和 n 维初始化量 r^0、p^0 需要满足某种代数关系，以确保状态演化的正确性。再次考虑带有某个 $\hat{g}$ 的（5.1）中的线性化 GVAMP 算法。令 $G = {}^{G}\,\mathrm{Tr}(G)/n - I_n$。将该迭代的不动点记为去掉时间指标后的相同记号。令 $\widehat{u}$、γ 分别如（5.3）和（5.5）所示。根据（5.7），$v_1(D)$ 与 $\widehat{v} = \kappa_2^{-1}X^\top G\widehat{u}$–对齐，而对于较大的 d、t，后者又与 v^t 对齐。于是根据（5.1）的更新规则，

$$\gamma\widehat{u} = (\overline{\kappa}_2^{-1}XX^\top - I)\overline{G}\widehat{u} = X\widehat{v} - \overline{G}\widehat{u}.$$

对 $\gamma\widehat{u}$ 求解可得

$$\gamma\widehat{u} = \gamma\big(\gamma I_n + \overline{G}\big)^{-1}X\widehat{v},$$

从而促使我们选择 $c_p^{-1}p^0$ 于 （4.20）。

6 示例与数值实验

我们展示了所提出算法（以及文献中的一些现有方法）的数值模拟结果，并利用合成数据和真实数据，对相位恢复模型和泊松回归模型的相关理论进行了实证验证。一些实现细节见J节。

图 6 和7 分别比较了五种算法在相位恢复和泊松回归中的性能，涵盖三种设计谱分布。算法包括：定理 4.2 中的谱估计量（图中以黑色表示的“spec”）、 [61, 公式 (20)] 中的谱估计量（绿色的“spec conj”）；式 (4.36) 中的 Bayes-GVAMP（红色的“AMP”）、梯度下降（蓝色的“GD”），以及针对独立同分布高斯变量开发的广义近似消息传递算法 X [73, 节 3] （黄色的“GAMP”）。

Bayes-GVAMP 使用‘spec’初始化，而梯度下降和 GVAMP 使用‘spec conj’初始化。来自 [61] （‘spec conj’）的谱估计量与 (4.1) 形式相同，但使用预处理函数

$$\mathcal{T}^{\circ}(y) = \frac{\overline{g}(y)}{\overline{g}(y)+1}, \tag{6.1}$$

其中 $\bar{g}$由(4.16)给出。对于独立同分布高斯 [61], 相关高斯 [107] 以及部分Haar设计 [40], , 同一个 $\mathcal{T}^{\circ}$ 在 (6.1) 中被证明是最优选择，它在一个更大的函数类内同时最小化弱恢复阈值并最大化重叠，该函数类超越了由 假设 (A5)所定义的函数类。利用统计物理中的非严格方法， [66, 猜想 2] 预测 $\mathcal{T}^{\circ}$的这种最优性对任意正交不变设计成立。上述 $\mathcal{T}^{\circ}$ 不满足 假设 (A5)，因此我们在 定理 4.1 和 4.4 中的结果并不适用。在此，我们对其性能进行数值检验，并观察到仅针对 $\mathcal{T}^{\circ}$ 的谱估计量性能较差，不及使用由 (4.16) 和 假设 (A5)定义的我们的谱估计量进行初始化的 Bayes-GVAMP。

对于梯度下降，我们采用标准的全批量版本

$$\theta^{t+1} = \theta^{t} - \eta \nabla L(\theta^{t}),$$

其中，学习率 $\eta > 0$ 固定，经验损失

$$L(\theta) = \sum_{i=1}^{n} \ell(\hat{q}(\langle x_i, \theta \rangle), y_i)$$

定义在假定模型 $\hat{q}\colon \mathbb{R} \to \mathbb{R}$ 和适当的损失函数 $\ell\colon \mathbb{R}^2 \to \mathbb{R}$ 下（两者将在下文详细介绍）。

我们注意到，对于 [87, 73], 独立同分布高斯设计下被认为在计算上最优的 GAMP， [5, 75], 在图6和7中的非高斯设计下明显次优。事实上，在某些情况下，例如当 X 具有Beta(1,2)分布的奇异值时（图6和7中间面板），或数据来自 GTEx 数据集（图1）时，GAMP 的表现甚至不及其谱初始化器(6.1)，也不如梯度下降等通用算法。

我们注意到，在所有图中，定理 4.2中的谱估计量以及以其初始化的 Bayes-GVAMP 所达到的重叠量都能被我们在定理 4.1和4.4 中的理论准确预测（图中分别以对应颜色的实线绘出“spec, thy”和“AMP, thy”）。

6.1 合成数据

所有使用合成数据的数值实验中，设计矩阵均按照假设 (A2)构造，其中 Λ 取为

$$\Lambda = \begin{cases} \widetilde{\Lambda}, & \delta \le 1 \\ \sqrt{\delta}\,\widetilde{\Lambda}, & \delta > 1 \end{cases}, \tag{6.2}$$

其中， $\widetilde{\Lambda}$ 是一个具有紧支撑集、二阶矩为 1 且方差为正的正随机变量。容易验证，上述构造确保了 $\kappa_2 = m_2 = \mathbb{E}\big[\Lambda_n^2\big] = 1$，这是一种自然的归一化，与标准高斯设计一致

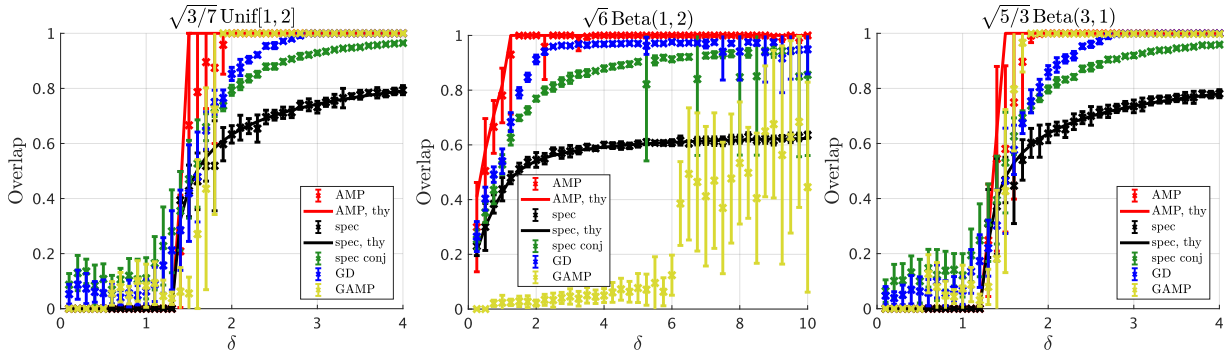

图6：在相位恢复模型 5 上各算法达到的重叠量 (6.4)。对设计矩阵的3 种谱分布（见(6.3)）重复进行相同的实验 X。

其中 $X_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/d)$。所有合成实验均针对 Λ 的 3 种选择重复进行：

$$\sqrt{3/6}\,\mathrm{Unif}([1,2]), \qquad \sqrt{6}\,\mathrm{Beta}(1,2), \qquad \sqrt{5/3}\,\mathrm{Beta}(3,1), \tag{6.3}$$

其密度在各自的支撑集上分别为常数、线性函数和二次函数：

$$p_{\mathrm{Unif}([1,2])}(x) = \mathbb{1}_{[1,2]}(x), \qquad p_{\mathrm{Beta}(1,2)}(x) = 2(1-x)\mathbb{1}_{[0,1]}(x), \qquad p_{\mathrm{Beta}(3,1)}(x) = 3x^2\mathbb{1}_{[0,1]}(x).$$

在本节中，我们固定 $\rho = 1$ 并考虑高斯先验 $P_{\mathsf{B}_*} = \mathcal{N}(0,\ 1)$。根据 $\sigma^2 = 1$定义 (3.9)，这意味着 $\sigma^2 = 1$。我们固定回归系数的数量为 $d = 1000$，并通过相应地改变 n来改变 δ。对于 δ的每个取值，我们按照不同算法的 (1.1) 报告重叠量，结果取 10次独立同分布试验的平均值，误差条表示 1 个标准差。

相位恢复。考虑无噪声的相位恢复模型

$$q(z, \varepsilon) = |z| \tag{6.4}$$

对于该模型，$Q(y \mid z) = \delta_{|z|}(y)$。对于梯度下降，我们令假定模型 $\hat{q}(z) = |z|$ 与真实模型 q（6.4）相同，并使用平方损失 $\ell(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$。对于（6.3）中的 3 种谱分布，学习率 η分别取 0.1、0.01 和 0.1。

泊松回归。考虑如下泊松回归：

$$y_i \sim \mathrm{Pois}(\langle x_i, \beta_* \rangle^2). \tag{6.5}$$

也就是说，条件密度 Q由下式给出：

$$Q(y \mid z) = \frac{z^{2y} e^{-z^2}}{y!}, \qquad \text{for } y \in \mathbb{Z}_{\geq 0}.$$

对于梯度下降，我们将假定模型 $\hat{q}(z) = z^2$ 设为真实模型的无噪声类比(6.5)，并采用绝对损失 $\ell(\hat{y}, y) = |\hat{y} - y|$。对 $\eta = 0.001$ 所有3 个谱分布(6.3)，学习率均固定为该值。

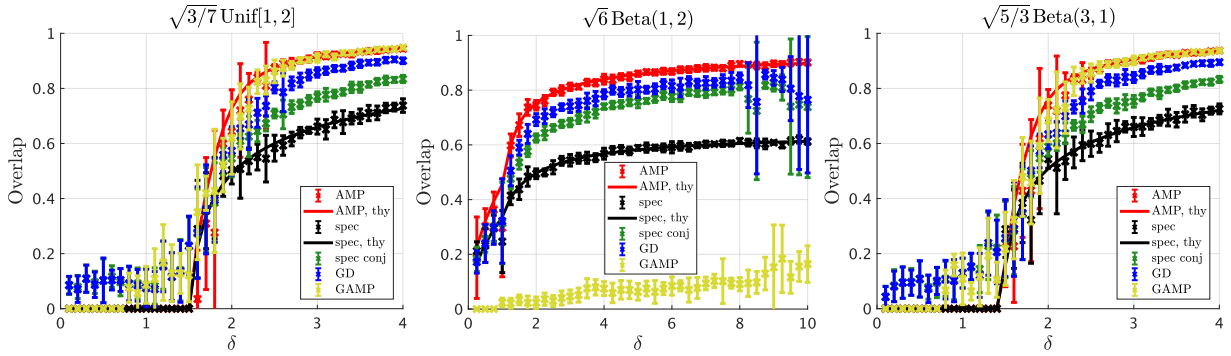

图7：由5 种算法在泊松回归模型(6.5)上达到的重叠量。对3种谱分布重复相同实验，这些谱分布由(6.3) 给出，设计矩阵为 X。

6.2 真实数据

对于相位恢复模型(6.4)，我们还进行了数值实验，其中设计矩阵和/或回归系数来自真实数据。在所有此类实验中，采用谱初始化的 Bayes-GVAMP 在(4.36) 中的表现均优于4 种其他方法（图 1 和2）。与 Bayes-GVAMP 不同，GAMP 等方法 [87, 73]是在独立同分布假设下提出的，因此无法很好地适应数据分布。此外，定理 4.1 和定理 4.1以及4.4 分别针对定理 4.2 中的谱估计量，以及定理 4.2 和以其初始化的 Bayes-GVAMP 所给出的理论保证，对于不满足第 3.2 节中分布假设的真实数据仍然准确。这表明正交不变性假设具有灵活性，而且我们的理论对于数据分布具有普适性。

编码衍射图样。 编码衍射图样（CDP）是一类常用的设计集合，用于刻画相位恢复中的物理测量。我们考虑 [17, 第2节2]中定义的二元和三元复数编码衍射图样的实值对应形式。更多细节见J.3节。回归系数向量通过对3 个通道（R、G、B）进行向量化、中心化与归一化得到，这些通道来自 $d = 32 \times 32$CIFAR-10 数据集中一辆卡车的图像 [56]。在Bayes-GVAMP(4.36)和GAMP [73],中，向量去噪器均在独立同分布高斯先验假设下构造。数值结果见图2。图中每个数据点由15 次试验的平均值得到，其中对R、G、B每个通道各进行5 次试验。

GTEx。 我们考虑两个 GTEx 数据集“皮肤－日晒部位（小腿）”和“肌肉－骨骼肌”（在相应图表标题中分别简称为“皮肤”和“肌肉”） [59]。经过标准预处理（详见 J.3 节 后，它们分别成为大小为 6461×701 和 7036×803 的矩阵。设计矩阵 X 由这些矩阵的前 n 行构成。对于 Bayes-GVAMP，涉及 Λ 的期望以相对于相应经验谱分布的平均值取代 X。回归系数按照 $P_{\mathsf{B}_*} = \mathcal{N}(0,1)$独立同分布生成。数值结果见 图1 ，其中每个数据点均由 100 次独立同分布试验取平均得到。

7 讨论

链接函数和先验分布的估计。 我们关于谱估计量（定理 4.1）和 GVAMP（定理 4.3）的一般结果并不假设已知链接函数 q 或先验分布P_{B_*}。然而，谱估计量的预处理函数（定理 4.2）和 GVAMP 的去噪器（定理 4.4）的最优选择确实需要这些知识。当 q 和 P_{B_*} 由相对于 n， d固定的（相对）参数数量进行参数化时，可以在比例渐近区间（假设（A4）中通过矩方法等方法对其进行一致估计。于是，我们的所有结果在代入这些一致

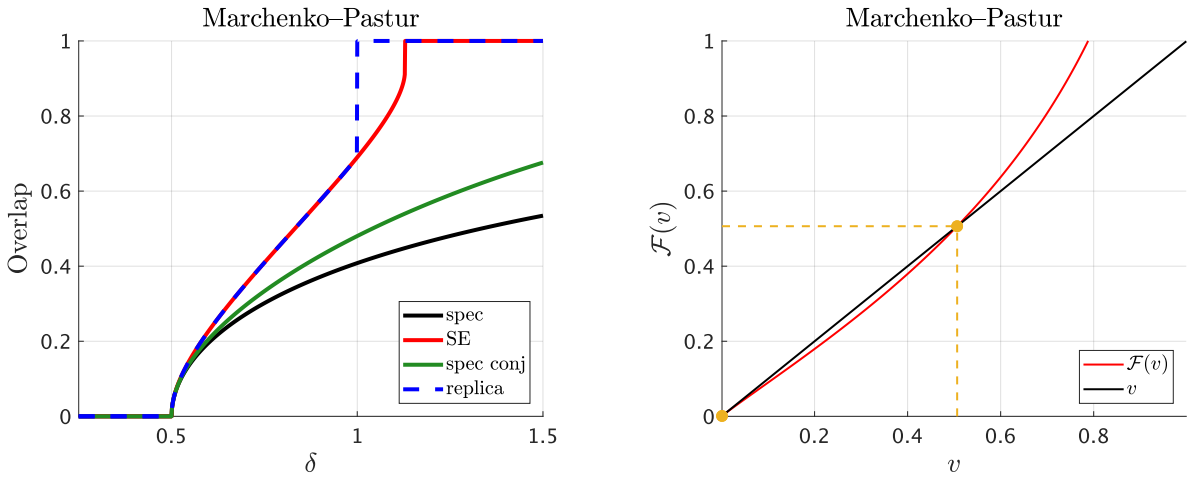

图 8：标准高斯设计下相位恢复的统计—计算权衡，其谱分布由 Marchenko–Pastur 定律给出。左图展示 4 条曲线，绘制方式与图 5 相同，另增加一条曲线，对应于谱估计量 [61, 公式 (20)]所达到的渐近重叠。右图绘制函数 $\mathcal{F}$ 在 $\delta = 1.01$处的值。

估计量后仍然成立。在更一般的（半参数或非参数）情形下，诸如 q 和 P_{B_*} 之类的干扰泛函不再被期望能够得到一致估计。在先验分布和链接函数设定错误的情况下，后验均方误差的退化程度已在 [53, 96] 中使用非严格的复制子方法进行了量化。经验贝叶斯方法 [90, 77, 45] 已针对若干具有独立同分布设计的贝叶斯统计模型提出并进行了分析。我们相信，将类似思想迁移到正交不变广义线性模型的设定中将会富有成效。这些方法的一个常见缺点是，它们要求样本量足够大，或信噪比足够低。开发能够保留本文所确立的预言机对应方法最优性的数据驱动估计和去偏程序[110, 58, 22, 25] ，是未来研究的一个有前景方向。

统计与计算最优性。对于一般的正交不变广义线性模型，我们的谱初始化 Bayes-GVAMP 是最先进的高效估计量。在某些区域，其性能达到猜想的贝叶斯风险；而在其他区域则仍存在差距，尤其是状态演化方程具有多个不动点时（参见定理 4.6之后的讨论）。有人猜想[66, 67] Bayes-GVAMP 在所有多项式时间算法中具有最优性。不过需要注意，仅仅存在多个状态演化不动点，以及贝叶斯风险与谱初始化的 Bayes-GVAMP 的性能之间存在差异（见图 8），并不一定意味着统计极限与计算极限之间存在根本差距。例如，对于采用独立同分布高斯设计的无噪声相位恢复(6.4) ，存在能够 [1, 94]达到重叠1 （即完美恢复）的高效算法，只要$n \geq d + 1$ （特别是， $\delta \geq 1$）。相比之下，Bayes-GVAMP 或 GAMP 要达到同样的效果则需要 $\delta \gtrsim 1.13$ 。理解 Bayes-GVAMP 在何时以及为何具有计算最优性，是未来研究的一个有趣方向。

致谢

这项工作部分是在 Y.Z. 于奥地利科学技术研究所任职期间完成的，当时他获得了欧盟（欧洲研究理事会，INF^2,项目编号 101161364）的资助。M.M. 获得了欧盟（欧洲研究理事会，INF^2, 项目编号 101161364）的资助。然而，文中表达的观点和意见仅代表作者本人，并不一定反映欧盟或欧洲研究理事会的观点。

执行机构。欧盟和资助机构均不对此承担责任。Y.Z. 感谢 Rishabh Dudeja 在 2024年2月访问威斯康星大学麦迪逊分校统计系期间进行的启发性讨论，这次讨论促使作者将注意力从 RI-GAMP 转向 [101] VAMP [42, 43]。Y.Z. 还感谢 Junjie Ma 在 [41]2025年2月于莱苏什物理学校举办的“迈向典型情形算法难度理论”研讨会期间进行的有益讨论。

参考文献

[1] Alexandr Andoni、Daniel Hsu、Kevin Shi和Xiaorui Sun。《对应检索》。载于：2017年学习理论会议论文集。第65卷。机器学习研究论文集。2017年7月，第105–126页（引于第25页）。[2] Octavio Arizmendi、Takahiro Hasebe和Yu Kitagawa。与实数轴上任意测度的自由乘法卷积。2025年。arXiv：2503.14992 [math.PR]（引于第59页）。[3] Jinho Baik、Gérard Ben Arous和Sandrine Péché。《非零复样本协方差矩阵最大特征值的相变》。载于：《概率年刊》33.5（2005），第1643–1697页（引于第3页）。[4] Jean Barbier、Francesco Camilli、Marco Mondelli和Manuel Sáenz。《结构化主成分分析的基本极限及其实现方法》。载于：美国国家科学院院刊120.30（2023），e2302028120（引于第2页）。[5] Jean Barbier、Florent Krzakala、Nicolas Macris、Léo Miolane和Lenka Zdeborová。《高维广义线性模型中的最优误差与相变》。载于：美国国家科学院院刊116.12（2019），第5451–5460页（引于第1、2、4、17、22页）。[6] Jean Barbier和Nicolas Macris。《自适应插值方法：证明贝叶斯推断中复制公式的一种简易方案》。载于：《概率论及相关领域》174.3（2019年8月），第1133–1185页（引于第4页）。[7] Mohsen Bayati和Andrea Montanari。《稠密图上的消息传递动力学及其在压缩感知中的应用》。载于：IEEE信息论汇刊57.2（2011），第764–785页（引于第4页）。[8] Mohsen Bayati和Andrea Montanari。《高斯矩阵的LASSO风险》。载于：IEEE信息论汇刊58.4（2012），第1997–2017页（引于第1页）。[9] Serban Belinschi、Hari Bercovici和Ching-Wei Ho。自由卷积的支撑集。2024年。arXiv：2408.06573 [math.OA]（引于第7页）。[10] Serban T. Belinschi、Hari Bercovici、Mireille Capitaine和Maxime Février。《大型变形酉不变模型谱中的离群值》。载于：《概率年刊》45.6A（2017），第3571–3625页（引于第6页）。[11] Erwin Bolthausen。《Sherrington–Kirkpatrick模型TAP方程解的迭代构造》。载于：《数学物理通讯》325.1（2014年1月），第333–366页（引于第4页）。[12] Petros T. Boufounos和Richard G. Baraniuk。《1比特压缩感知》。载于：2008年第42届信息科学与系统年度会议。2008年，第16–21页（引于第1页）。[13] Burak Çakmak和Manfred Opper。《具有任意旋转不变随机耦合矩阵的伊辛模型Thouless-Anderson-Palmer方程的无记忆动力学》。载于：Phys. Rev. E 99（6，2019年6月），第062140页（引于第2页）。[14] Collin Cademartori和Cynthia Rush。《旋转不变设计下广义向量近似消息传递算法的非渐近分析》。载于：IEEE Trans. Inform. Theory 70.8（2024），第5811–5856页（引于第3页）。[15] Burak Çakmak、Ole Winther和Bernard H. Fleury。《用于非线性观测模型的S-AMP》。载于：2015年IEEE国际信息论研讨会（ISIT）。2015年，第2807–2811页（引于第4页）。[16] Burak Çakmak、Ole Winther和Bernard H. Fleury。《S-AMP：适用于一般矩阵系综的近似消息传递》。载于：2014年IEEE信息论研讨会（ITW 2014）。2014年，第192–196页（引于第4页）。[17] Emmanuel J. Candès、Xiaodong Li和Mahdi Soltanolkotabi。《从编码衍射图样中进行相位恢复》。载于：《应用与计算调和分析》39.2（2015），第277–299页（引于第4、24、120页）。

[18] Emmanuel J. Candès、Xiaodong Li和Mahdi Soltanolkotabi。《通过Wirtinger Flow进行相位恢复：理论与算法》。载于：IEEE信息论汇刊61.4（2015），第1985–2007页（引于第1页）。[19] Michael Celentano。《近似消息传递后的Sudakov–Fernique方法，以及TAP自由能局部凸性的新证明》。载于：《概率年刊》52.3（2024），第923–954页（引于第4页）。[20] Michael Celentano、Chen Cheng、Ashwin Pananjady和Kabir Aladin Verchand。超越一阶方法的状态演化 I：严格预测与有限样本保证。2025年。arXiv：2507.19611 [math.ST]（引于第1页）。[21] Michael Celentano、Zhou Fan和Song Mei。《TAP自由能的局部凸性及用于同步的近似消息传递收敛性》。载于：《统计年刊》51.2（2023），第519–546页（引于第4页）。[22] Michael Celentano和Andrea Montanari。《相关性调整去偏Lasso：利用不准确的协变量模型对Lasso去偏》。载于：《皇家统计学会会刊 B辑：统计方法学》86.5（2024年6月），第1455–1482页（引于第25页）。[23] Michael Celentano、Andrea Montanari和Yuting Wei。《具有一般高斯设计的Lasso及其在假设检验中的应用》。载于：《统计年刊》51.5（2023），第2194–2220页（引于第2页）。[24] Michael Celentano、Andrea Montanari和Yuchen Wu。《一般一阶方法的估计误差》。载于：第三十三届学习理论会议论文集。第125卷。2020年，第1078–1141页（引于第1页）。[25] Michael Celentano和Martin J. Wainwright。不一致区域的挑战：缺失数据模型的新型去偏方法。2023年。arXiv：2309.01362 [math.ST]（引于第25页）。[26] Kabir Aladin Chandrasekher、Ashwin Pananjady和Christos Thrampoulidis。《利用随机数据进行迭代非凸优化的尖锐全局收敛保证》。载于：《统计年刊》51.1（2023），第179–210页（引于第1页）。[27] Christopher C Chang、Carson C Chow、Laurent CAM Tellier等。《第二代PLINK：迎接更大、更丰富数据集的挑战》。载于：GigaScience 4.1（2015年2月），s13742-015-0047–8（引于第120页）。[28] Haohua Chen、Songbin Liu和Junjie Ma。适用于矩形尖峰矩阵模型、具有最优谱初始化的正交近似消息传递。2025年。arXiv：2512.19334 [cs.IT]（引于第13、70页）。[29] Wei-Kuo Chen和Wai-Kit Lam。《近似消息传递算法的普适性》。载于：电子概率杂志26.none（2021），第1–44页（引于第2页）。[30] Yuxin Chen和Emmanuel J. Candès。《求解随机二次方程组几乎与求解线性方程组一样容易》。载于：纯粹与应用数学通讯70.5（2017），第822–883页（引于第1页）。[31] Yuxin Chen、Yuejie Chi、Jianqing Fan和Cong Ma。《数据科学的谱方法：统计学视角》。载于：机器学习基础与趋势14.5（2021），第566–806页（引于第1页）。[32] Benoît Collins和Camille Male。《Haar矩阵与确定性矩阵的强渐近自由性》。载于：Ann. Sci. Éc. Norm. Supér. (4) 47.1（2014），第147–163页（引于第48页）。[33] Alex Damian、Jason D. Lee和Joan Bruna。《生成式飞跃：高效学习高斯多指标模型所需的紧致样本复杂度》。载于：第三十九届神经信息处理系统年度会议。2025年（引于第4页）。[34] Alex Damian、Loucas Pillaud-Vivien、Jason Lee和Joan Bruna。《高斯单指标模型中的计算—统计差距（扩展摘要）》。载于：第三十七届学习理论会议论文集。第247卷。机器学习研究论文集。2024年6月30日–7月3日，第1262–1262页（引于第4页）。[35] Leonardo Defilippis、Yatin Dandi、Pierre Mergny、Florent Krzakala和Bruno Loureiro。《高维多指标模型中的最优谱转变》。载于：神经信息处理系统进展。2025年（引于第4页）。[36] Xiucai Ding和Hong Chang Ji。《尖峰乘法随机矩阵与主成分》。载于：随机过程及其应用163（2023），第25–60页（引于第20页）。[37] David Donoho、Matan Gavish和Elad Romanov。《ScreeNOT：相关噪声下均方误差最优奇异值阈值化的精确实现》。载于：《统计年刊》51.1（2023），第122–148页（引于第6页）。[38] David Donoho和Andrea Montanari。《高维稳健M估计：通过近似消息传递求渐近方差》。载于：《概率论及相关领域》166.3（2016年12月），第935–969页（引于第1页）。

[39] Rishabh Dudeja 和 Milad Bakhshizadeh。《结构化感知矩阵相位恢复中线性化消息传递的普适性》。载于：IEEE信息论汇刊 68.11（2022年），第 7545–7574 页（引文见第 2、4 页）。[40] Rishabh Dudeja、Milad Bakhshizadeh、Junjie Ma 和 Arian Maleki。《随机正交矩阵相位恢复的谱方法分析》。载于：IEEE信息论汇刊 66.8（2020），第 5182–5203 页（引文见第 2、4、22 页）。[41] Rishabh Dudeja、Songbin Liu 和 Junjie Ma。《旋转不变噪声下尖峰矩阵模型的近似消息传递算法的最优性》。载于：《统计年刊》（2025年）。即将发表（引文见第 2、4、26 页）。[42] Rishabh Dudeja、Yue M. Lu 和 Subhabrata Sen。《半随机矩阵近似消息传递的普适性》。载于：《概率年刊》51.5（2023年），第 1616–1683 页（引文见第 2、4、7、26 页）。[43] Rishabh Dudeja、Subhabrata Sen 和 Yue M. Lu。《近乎确定性感知矩阵正则化线性回归中的谱普适性》。载于：IEEE信息论汇刊 70.11（2024年），第 7923–7951 页（引文见第 2、4、7、21、26 页）。[44] Zhou Fan。《旋转不变矩阵的近似消息传递算法》。载于：统计年刊 50.1（2022年），第 197–224 页（引文见第 2、5、10、21、32、120、121 页）。[45] Zhou Fan、Leying Guan、Yandi Shen 和 Yihong Wu。《高维线性模型中经验贝叶斯的梯度流》。2023年。arXiv：2312.12708 [math.ST]（引文见第 25 页）。[46] Zhou Fan、Yi Sun 和 Zhichao Wang。《具有一般谱主体的线性混合模型中的主成分》。载于：统计年刊 49.3（2021年），第 1489–1513 页（引文见第 20、48 页）。[47] Albert Fannjiang 和 Thomas Strohmer。《相位恢复的数值方法》。载于：数值学报 29（2020年），第 125–228 页（引文见第 1 页）。[48] Oliver Y. Feng、Ramji Venkataramanan、Cynthia Rush 和 Richard J. Samworth。《近似消息传递统一教程》。载于：机器学习基础与趋势 15.4（2022年5月），第 335–536 页（引文见第 13 页）。[49] Dominique Fourdrinier、William E. Strawderman 和 Martin T. Wells。《收缩估计》。施普林格统计学丛书。施普林格，Cham，2018年，第 xiii 页+333（引文见第 86、89 页）。[50] Marylou Gabrié、Andre Manoel、Clément Luneau 等。《深度神经网络模型中的熵与互信息*》。载于：Journal of Statistical Mechanics: Theory and Experiment 2019.12（2019年12月），第 124014 页（引文见第 4 页）。[51] Qiyang Han。《一般一阶方法的逐元素动力学与普适性》。载于：《统计年刊》53.4（2025年），第 1783–1807 页（引文见第 4 页）。[52] Ramon van Handel。《强收敛现象》。arXiv:2507.00346。2025年（引文见第 20 页）。[53] Y Kabashima。《相关模式下的推断：感知机学习与线性向量信道的统一理论》。载于：物理学杂志：会议系列 95.1（2008年1月），第 012001 页（引文见第 2–4、25 页）。[54] Yoshiyuki Kabashima 和 Mikko Vehkaperä。《使用期望一致近似恢复线性观测信号》。载于：2014年 IEEE国际信息论研讨会。2014年，第 226–230 页（引文见第 4 页）。[55] Filip Kovačević、Zhang Yihan 和 Marco Mondelli。《多指标模型的谱估计量：精确渐近分析与最优弱恢复》。载于：第三十八届学习理论会议论文集。2025年，第 3354–3404 页（引文见第 4 页）。[56] Alex Krizhevsky。《从微小图像中学习多层特征》。技术报告。加拿大安大略省多伦多：多伦多大学，2009年（引文见第 24 页）。[57] Yufan Li、Zhou Fan、Subhabrata Sen 和 Yihong Wu。《旋转不变设计下的随机线性估计：高温渐近分析》。载于：IEEE信息论汇刊 70.3（2024年），第 2118–2153 页（引文见第 4、120 页）。[58] Yufan Li 和 Pragya Sur。《谱感知去偏：一种现代推断框架及其在主成分回归中的应用》。2025年。arXiv：2309.07810 [math.ST]（引文见第 2、25 页）。[59] John Lonsdale、Jeffrey Thomas、Mike Salvatore 等。《基因型-组织表达（GTEx）项目》。载于：自然遗传学 45.6（2013年6月），第 580–585 页（引文见第 2、4、24、120 页）。

[60] Yue M Lu 和 Gen Li。《高维非凸估计中谱初始化的相变》。载于：Information and Inference: A Journal of the IMA 9.3（2019年11月），第 507–541 页（引文见第 1、4 页）。[61] Wangyu Luo、Wael Alghamdi 和 Yue M. Lu。《信号恢复的最优谱初始化及其在相位恢复中的应用》。载于：IEEE Transactions on Signal Processing 67.9（2019年），第 2347–2356 页（引文见第 1、2、4、22、25、118、119 页）。[62] Junjie Ma、Rishabh Dudeja、Ji Xu、Arian Maleki 和 Xiaodong Wang。《相位恢复的谱方法：期望传播视角》。载于：IEEE信息论汇刊 67.2（2021年），第 1332–1355 页（引文见第 1、2、4、19 页）。[63] Junjie Ma 和 Li Ping。《正交 AMP》。载于：IEEE Access 5（2017年），第 2020–2033 页（引文见第 4 页）。[64] Junjie Ma、Ji Xu 和 Arian Maleki。《面向广义线性逆问题的最优感知矩阵设计》。载于：IEEE信息论汇刊 70.1（2024年），第 482–508 页（引文见第 4 页）。[65] Antoine Maillard、Laura Foini、Alejandro Lage Castellanos 等。《高温展开与消息传递算法》。载于：Journal of Statistical Mechanics: Theory and Experiment 2019.11（2019年11月），第 113301 页（引文见第 4 页）。[66] Antoine Maillard、Florent Krzakala、Yue M. Lu 和 Lenka Zdeborova。《构造相位恢复的最优谱方法》。载于：第二届数学与科学机器学习会议论文集。第 145 卷。机器学习研究论文集。2022年，第 693–720 页（引文见第 2、4、18、22、25 页）。[67] Antoine Maillard、Bruno Loureiro、Florent Krzakala 和 Lenka Zdeborová。《高维相位恢复：统计与计算相变》。载于：神经信息处理系统进展。第 33 卷。2020年，第 11071–11082 页（引文见第 1、3、4、10、15–17、25 页）。[68] Camille Male。《大型随机矩阵与确定性矩阵中多项式的范数》。载于：Probab. Theory Related Fields 154.3-4（2012年）。附录由 Dimitri Shlyakhtenko 撰写，第 477–532 页（引文见第 53 页）。[69] Elizabeth S. Meckes。《经典紧群的随机矩阵理论》。第 218 卷。Cambridge Tracts in Mathematics。剑桥大学出版社，剑桥，2019年，第 xi 页+212 （引文见第 48、51 页）。[70] Thomas P. Minka。《近似贝叶斯推断的一类算法》。博士论文，麻省理工学院。ProQuest LLC，密歇根州安阿伯，2001年，（无页码）（引文见第 4 页）。[71] Marco Mondelli 和 Andrea Montanari。《弱恢复的基本极限及其在相位恢复中的应用》。载于：Foundations of Computational Mathematics 19 1, 4, 10, 70 3 2019 703 773 .（June ），第 – 页（引文见第 页）。[72] Marco Mondelli、Christos Thrampoulidis 和 Ramji Venkataramanan。《广义线性模型中线性估计量与谱估计量的最优组合》。载于：Foundations of Computational Mathematics 22.5（2022年10月），第 1513–1566 页（引文见第 4 页）。[73] Marco Mondelli 和 Ramji Venkataramanan。《用于广义线性模型的谱初始化近似消息传递*》。载于：Journal of Statistical Mechanics: Theory and Experiment 2022.11（2022年11月），第 114003 页（引文见第 1、2、17、22、24、120 页）。[74] Andrea Montanari 和 Ramji Venkataramanan。《通过近似消息传递估计低秩矩阵》。载于：《统计年刊》49.1（2021年），第 321–345 页（引文见第 4 页）。[75] Andrea Montanari 和 Yuchen Wu。《统计最优一阶算法：通过正交化给出的证明》。载于：Information and Inference: A Journal of the IMA 13.4（2024年10月），iaae027（引文见第 1、22 页）。[76] Ali Mousavi、Arian Maleki 和 Richard G. Baraniuk。《LASSO 与近似消息传递的一致参数估计》。载于：《统计年刊》46.1（2018年），第 119–148 页（引文见第 4 页）。[77] Sumit Mukherjee、Bodhisattva Sen 和 Subhabrata Sen。《高维线性回归中经验贝叶斯估计的平均场方法》。2023年。arXiv：2309.16843 [math.ST]（引文见第 25 页）。[78] J. A. Nelder 和 R. W. M. Wedderburn。《广义线性模型》。载于：Journal of the Royal Statistical Society. Series A (General) 135.3（1972年），第 370–384 页（引文见第 1 页）。[79] Praneeth Netrapalli、Prateek Jain 和 Sujay Sanghavi。《使用交替最小化进行相位恢复》。载于：IEEE Transactions on Signal Processing 63.18（2015年），第 4814–4826 页（引文见第 1 页）。[80] Manfred Opper 和 Ole Winther。《用于概率建模的自适应、自平均 Thouless-Anderson-Palmer 平均场理论》。载于：Phys. Rev. E 64（5，2001年10月），第 056131 页（引文见第 4 页）。

[81] Manfred Opper 和 Ole Winther。《期望一致近似推断》。载于：《机器学习研究期刊》6.73（2005年），第2177–2204页（引用见第4页）。[82] Manfred Opper 和 Ole Winther。《概率模型的可处理近似：自适应 Thouless-Anderson-Palmer 平均场方法》。载于：物理评论快报 86（17，2001年4月），第3695–3699页（引用见第4页）。[83] Samet Oymak 和 Babak Hassibi。《在线性逆问题中采用正交测量的理由》。载于：2014年 IEEE国际信息论研讨会。2014年，第3175–3179页（引用见第2页）。[84] Parthe Pandit、Mojtaba Sahraee Ardakan、Sundeep Rangan、Philip Schniter 和 Alyson K Fletcher。《多层模型中的矩阵推断与估计》。载于：神经信息处理系统进展。第33卷。Curran Associates, Inc.，2020年，第22456–22467页（引用见第4页）。[85] Parthe Pandit、Mojtaba Sahraee-Ardakan、Sundeep Rangan、Philip Schniter 和 Alyson K. Fletcher。《高维中基于深度生成先验的推断》。载于：IEEE 信息论选定领域期刊 1.1（2020年），第336–347页（引用见第4页）。[86] William H. Press、Saul A. Teukolsky、William T. Vetterling 和 Brian P. Flannery。《数值配方：科学计算艺术》。第3版。第3版。美国：剑桥大学出版社，2007年（引用见第119页）。[87] Sundeep Rangan。《用于随机线性混合估计的广义近似消息传递》。载于：2011 1, 17, 22, 24 2011, 2168 2172 IEEE国际信息论研讨会论文集。第 – 页（引用见第 页）。[88] Sundeep Rangan、Philip Schniter 和 Alyson K. Fletcher。《向量近似消息传递》。载于：IEEE信息论汇刊 65.10（2019年），第6664–6684页（引用见第4、123页）。[89] Galen Reeves。《通过加性高斯噪声变换实现多层网络中的信息可加性》。载于：2017年第55届 Allerton 通信、控制与计算会议（Allerton）。2017年，第1064–1070页（引用见第4页）。[90] Kazuma Sawaya、Yoshimasa Uematsu 和 Masaaki Imaizumi。《高维单指标模型：链接函数估计与边缘推断》。2024年。arXiv：2404。17812 [math.ST]（引用见第25页）。[91] Philip Schniter、Sundeep Rangan 和 Alyson K. Fletcher。《广义线性模型的向量近似消息传递》。载于：2016年第50届阿西洛马信号、系统与计算会议。2016年，第1525–1529页（引用见第2、4、18、21页）。[92] Matthias Seeger。《指数族的期望传播》。2005年（引用见第4页）。[93] Yoav Shechtman、Yonina C. Eldar、Oren Cohen 等。《相位恢复及其在光学成像中的应用：当代综述》。载于：IEEE 信号处理杂志 32.3（2015年），第87–109页（引用见第1页）。[94] Min Jae Song、Ilias Zadik 和 Joan Bruna。《关于学习单周期神经元的密码学困难性》。载于：神经信息处理系统进展。第34卷。Curran Associates, Inc.，2021年，第29602–29615页（引用见第25页）。[95] Pragya Sur 和 Emmanuel J. Candès。《高维逻辑回归的现代极大似然理论》。载于：美国国家科学院院刊 116.29（2019年），第14516–14525页（引用见第1页）。[96] Takashi Takahashi 和 Yoshiyuki Kabashima。《模型失配情形下向量近似消息传递的宏观分析》。载于：IEEE信息论汇刊 68.8（2022年），第5579–5600页（引用见第2–4、25页）。[97] Keigo Takeuchi。《基于期望传播的酉不变测量信号恢复的严格动力学分析》。载于：2017年 IEEE国际信息论研讨会（ISIT）。2017年，第501–505页（引用见第4页）。[98] Christos Thrampoulidis、Ehsan Abbasi 和 Babak Hassibi。《高维中正则化 M -估计量的精确误差分析》。载于：IEEE信息论汇刊 64.8（2018年），第5592–5628页（引用见第1页）。[99] Christos Thrampoulidis 和 Babak Hassibi。《各向同性随机正交矩阵：LASSO 的性能与最小锥奇异值》。载于：2015年 IEEE国际信息论研讨会（ISIT）。2015年，第556–560页（引用见第121页）。[100] Juliana Tolles 和 William J. Meurer。《逻辑回归：将患者特征与结局联系起来》。载于：《美国医学会杂志》316.5（2016年8月），第533–534页（引用见第1页）。

[101] Ramji Venkataramanan、Kevin Kögler 和 Marco Mondelli。《通过近似消息传递估计旋转不变广义线性模型》。载于：第39届国际机器学习大会论文集。第162卷。2022年，第22120–22144页（引用见第2、4、18、26、36–38、102页）。[102] Cédric Villani。《最优传输》。第338卷。数学科学基础丛书 [数学科学基本原理]。新旧论述。施普林格出版社，柏林，2009年，第xxii页+973 （引用见第5页）。[103] Tianhao Wang、Xinyi Zhong 和 Zhou Fan。《近似消息传递算法与张量网络的普适性》。载于：《应用概率年刊》34.4（2024年），第3943–3994页（引用见第2、7页）。[104] Ke Wei。《利用 Kaczmarz 方法求解无相位方程组：概念验证研究》。载于：《反问题》31.12（2015年11月），第125008页（引用见第1页）。[105] Yizhou Xu、TianQi Hou、ShanSuo Liang 和 Marco Mondelli。《旋转不变模型中多层估计的近似消息传递》。载于：2023年 IEEE 信息论研讨会（ITW）。IEEE。2023年，第294–298页（引用见第4页）。[106] Xinyang Yi、Constantine Caramanis 和 Sujay Sanghavi。《混合线性回归的交替最小化》。载于：第31届国际机器学习大会论文集。第32卷。机器学习研究论文集2。中国北京，2014年，第613–621页（引用见第1页）。[107] Yihan Zhang、Hong Chang Ji、Ramji Venkataramanan 和 Marco Mondelli。《通过近似消息传递构造结构化广义线性模型的谱估计量》。载于：《数学统计与学习》8.3-4（2025年），第193–304页（引用见第2、4、6、22、55、60、70页）。[108] Yihan Zhang、Marco Mondelli 和 Ramji Venkataramanan。《混合广义线性模型中谱方法的精确渐近分析》。载于：《SIAM 数据科学数学杂志》（SIMODS）（2026年）。即将发表（引用见第4页）。[109] Qian Zhao、Pragya Sur 和 Emmanuel J. Candès。《高维逻辑回归模型中最大似然估计量的渐近分布：任意协方差》。载于：《伯努利》28.3（2022年），第1835–1861页（引用见第2页）。[110] Xinyi Zhong、Chang Su 和 Zhou Fan。《高维经验贝叶斯主成分分析》。载于：《皇家统计学会会刊 B辑：统计方法学》84.3（2022年1月），第853–878页（引用见第25页）。

自由累积量和矩

对于矩阵 $X \in \mathbb{R}^{n\times d}$，记 m_{2k} 为 $XX^\top$的经验谱分布的第 k 阶（其中 $k \in \mathbb{Z}_{\geq 0}$）矩。于是， $X^\top X$ 的经验谱分布的第 k 阶矩为

$$m'_{2k} = \begin{cases} 1, & k = 0 \\ \widehat{\delta} m_{2k}, & k \geq 1 \end{cases},$$

其中 $\widehat{\delta} = n/d$。记 $(\kappa_{2k})_{k\geq 1}$ 为 $XX^\top$的矩形自由累积量；详见 [44, 第2节4] 中的精确定义。前四个矩形自由累积量明确表示为:

$$\kappa_2 = m_2 = \frac{m'_2}{\widehat{\delta}}, \tag{A.1a}$$

$$\kappa_4 = m_4 - (1+\widehat{\delta}) m_2^2 = \frac{m'_4}{\widehat{\delta}} - \frac{1+\widehat{\delta}}{\widehat{\delta}^2} m_2'^2, \tag{A.1b}$$

$$\kappa_6 = m_6 - (3+3\widehat{\delta}) m_4 m_2 + (2+3\widehat{\delta}+2\widehat{\delta}^2) m_2^3 = \frac{m'_6}{\widehat{\delta}} - \frac{3+3\widehat{\delta}}{\widehat{\delta}^2} m'_4 m'_2 + \frac{2+3\widehat{\delta}+2\widehat{\delta}^2}{\widehat{\delta}^3} m_2'^3, \tag{A.1c}$$

$$\begin{aligned} \kappa_8 &= m_8 - (4+4\widehat{\delta}) m_6 m_2 - (2+2\widehat{\delta}) m_4^2 + (10+16\widehat{\delta}+10\widehat{\delta}^2) m_4 m_2^2 - (5+10\widehat{\delta}+10\widehat{\delta}^2+5\widehat{\delta}^3) m_2^4 \\ &= \frac{m'_8}{\widehat{\delta}} - \frac{4+4\widehat{\delta}}{\widehat{\delta}^2} m'_6 m'_2 - \frac{2+2\widehat{\delta}}{\widehat{\delta}^2} m_4'^2 + \frac{10+16\widehat{\delta}+10\widehat{\delta}^2}{\widehat{\delta}^3} m'_4 m_2'^2 - \frac{5+10\widehat{\delta}+10\widehat{\delta}^2+5\widehat{\delta}^3}{\widehat{\delta}^4} m_2'^4. \end{aligned} \tag{A.1d}$$

Th $X^\top X$ 的矩形自由累积量记作 $(\kappa'_{2k})_{k\geq 1}$，并满足 $\kappa'_{2k} = \widehat{\delta}\kappa_{2k}$。

如果 $XX^\top$ 和 $X^\top X$ 的经验谱分布按照 假设 (A2)的意义收敛，则m_{2k}、m'_{2k}和 κ_{2k}、κ'_{2k} 也收敛。我们将它们的极限分别记为 m_{2k}、m'_{2k}和 κ_{2k}、κ'_{2k}。由于 $\kappa_{2k} = \delta\kappa'_{2k}$之间存在一一对应关系，我们将始终把 κ_{2k}称为该对象。

B 定理4.1的证明

如第 5节所示，我们首先考虑以下线性化 GVAMP 迭代：对于 $t \geq 1$,

$$v^t = \overline{\kappa}_2^{-1} X^\top g^t, \tag{B.1a}$$

$$u^t = \left(\overline{\kappa}_2^{-1} XX^\top - I_n\right) g^t, \qquad g^{t+1} = \overline{G} u^t, \tag{B.1b}$$

其中 $\overline{G} = \mathrm{diag}(g(y)) \in \mathbb{R}^{n\times n}$ 对于任意满足 假设 (A5)的函数 g，初始化器 g^1给定为

$$g^1 = \overline{G}(\nu_0 z + \tau_0 n^0), \tag{B.2}$$

其中 $n^0 \sim \mathcal{N}(0_n, I_n)$ 与其他所有对象独立，且 ν_0、τ_0 定义为

$$\nu_0 = \frac{1}{\sqrt{\rho + w_2}}, \qquad \tau_0 = \sqrt{\frac{w_1}{\rho + w_2}}. \tag{B.3}$$

定理 B.1 （证明见第 C 节）表明，在采用初始化方案(B.2)时，(B.1)中的若干状态演化参数在所有迭代量中都保持在其不动点上。

引理 B.1 (平稳状态演化参数)。 考虑 (B.1) 中经 (B.2) 初始化的 (4.13) 成立。则对于每个 $t \geq 1$，存在 χ_t、$\nu_t \neq 0$ 和 ω_t、$\tau_t > 0$ 使得

$$\begin{bmatrix} v^t & \beta_* \end{bmatrix} \xrightarrow{W_2} \begin{bmatrix} \chi_t \mathsf{B}_* + \omega_t \mathsf{M}_t & \mathsf{B}_* \end{bmatrix}, \qquad \begin{bmatrix} u^t & z & \varepsilon \end{bmatrix} \xrightarrow{W_2} \begin{bmatrix} \nu_t \mathsf{Z} + \tau_t \mathsf{N}_t & \mathsf{Z} & \mathsf{E} \end{bmatrix}, \tag{B.4}$$

其中 $(\mathsf{B}_*, \mathsf{M}_t) \sim P_{\mathsf{B}_*} \otimes \mathcal{N}(0,\ 1)$ 且 $(\mathsf{Z},\ \mathsf{M}_t) \sim \mathcal{N}(0,\ \sigma^2) \otimes \mathcal{N}(0,\ 1)$。此外，

$$\frac{\tau_t^2}{\nu_t^2} = w_1, \qquad \frac{\omega_t^2}{\chi_t^2} = w_2, \qquad \frac{\nu_t}{\chi_t} = w_3, \qquad \frac{\chi_{t+1}}{\nu_t} = w_4, \qquad \frac{\nu_t}{\nu_{t-1}} = \gamma, \tag{B.5}$$

其中上述所有等式的右端均为正数。

根据定理 B.1，我们来强调一下式 (B.1) 中线性化 GVAMP 算法的一个颇为不同寻常的特征。尽管状态演化参数在 (B.3) 中以正的有限值初始化，但当 $t \to \infty$ 时，它们不一定具有正的有限极限。根据 (B.5)，状态演化参数构成一个以 t 为索引、比值为 γ 的几何级数；当 $t \to \infty$ 时，除非 $\gamma = 1$，它们可能收敛到 0 或 ∞。例如，这意味着当 $t \to \infty$ 时，$\lim_{n\to\infty} n^{-1}\left\|u^{t+1}\right\|_2^2$ 可能不收敛到正的有限值，而 $\lim_{n\to\infty} n^{-1}\left\|u^{t+1} - u^t\right\|_2^2$ 可能不收敛到 0。然而，定理 B.1 表明，状态演化参数的某些比值确实具有非平凡极限。利用这一点可以证明，当 $t \to \infty$ 时，$\lim_{n\to\infty}\left\|u^{t+1}\right\|_2 / \|u^t\|_2$ 收敛到 γ（见 (E.5)），而 $\lim_{n\to\infty}\left\|\left\|u^{t+1}\right\|_2^{-1} u^{t+1} - \|u^t\|_2^{-1} u^t\right\|_2$ 收敛到 0（见 (E.23)）。确定其不动点（在大 t 极限下）存在的这类比值，是证明定理 B.3 中 v^{t+1} 与 $v_1(D)$ 之间对齐关系的关键一步。

引理 B.2（谱主体右边缘）。考虑通过以下方式定义的谱估计量：(4.1) 其中使用的预处理函数满足假设(A5)。于是有

$$\lim_{d\to\infty} \lambda_2(D) = \lambda^\circ,$$

其中 λ° 在右侧定义于(4.9)。

定理 B.2 在D 节中证明。

下一个引理表明，当 v^t 与 D 的最大特征向量对足够大的 t 和 d齐次。证明见E 节。

引理 B.3。考虑(B.1) 中的线性化 *GVAMP*，其以(B.2)初始化。设(4.13) 成立。那么几乎必然有，

$$\lim_{t\to\infty}\lim_{d\to\infty} \frac{\left|\left\langle v^{t+1}, v_1(D)\right\rangle\right|}{\left\|v^{t+1}\right\|_2} = 1, \qquad \lim_{d\to\infty} \lambda_1(D) = \overline{\kappa}_2,$$

其中 D 定义于(4.1a)。

现在，定理 4.1可由前面的3 个引理轻松推出。

定理 *4.1* 的证明定理 *4.1*。（4.12）和（4.14）给出了关于 D 的第一大和第二大特征值的刻画，这些结果已在定理 B.2和B.3中得到。接下来需要计算谱估计量的性能。根据定理 B.3，只需计算 β_* 与 v^t之间的归一化平方重叠。利用定理 B.1，可得

$$\lim_{d\to\infty} \frac{\left|\langle \beta_*, v_1(D)\rangle\right|^2}{\|\beta_*\|_2^2} = \lim_{t\to\infty}\lim_{d\to\infty} \frac{\left|\langle \beta_*, v^t\rangle\right|^2}{\|\beta_*\|_2^2\|v^t\|_2^2} = \lim_{t\to\infty} \frac{\chi_t^2\rho}{\chi_t^2\rho + \omega_t^2} = \lim_{t\to\infty} \frac{\rho}{\rho + \frac{\omega_t^2}{\chi_t^2}} = \frac{\rho}{\rho + w_2} = \eta,$$

其中，最后两个等式由 η、w_2 在(C.8) 和(4.11) 中的定义以及初等计算得出。注意，根据 $\eta \in (0,\ 1)$，由 $w_2 \in (0,\ \infty)$ 以及定理 B.1可知。 □

C 定理B.1 证明

考虑以下迭代，它比(B.1)稍具一般性。对于 $t \geq 1$,

$$v^t = \overline{\kappa}_2^{-1} X^\top g^t, \tag{C.1a}$$

$$u^t = \big(\overline{\kappa}_2^{-1} X X^\top - I_n\big) g^t, \qquad g^{t+1} = g_{t+1}(u^t; z, \varepsilon), \tag{C.1b}$$

以 $g^1 = g_1(z, \varepsilon)$ 初始化。(C.1) 通过如下设置，可将其视为 (4.19) 中 GVAMP 的一个实例：

$$\Phi_t(x) = 0, \qquad \widetilde{\Phi}_t(x) = \overline{\kappa}_2^{-1} x, \qquad \Psi_t(x) = \overline{\kappa}_2^{-1} x - 1, \qquad \widetilde{\Psi}_t(x) = 0, \tag{C.2a}$$

并进行如下变量替换

$$r^{t+1} = v^t, \qquad \widetilde{r}^{t+1} = 0_d, \qquad p^{t+1} = u^t, \qquad \widetilde{p}^{t+1} = g^{t+1}. \tag{C.2b}$$

我们定义 (C.1) 的状态演化递推。定义随机变量：对于， $t \geq 1$

$$\begin{bmatrix} \mathsf{B}_* & \mathsf{W}_1 & \cdots & \mathsf{W}_t \end{bmatrix}^\top \sim P_{\mathsf{B}_*} \otimes \mathcal{N}(0_t, \Omega_t), \tag{C.3a}$$

$$\begin{bmatrix} \mathsf{B}_1 & \cdots & \mathsf{B}_t \end{bmatrix}^\top = \begin{bmatrix} \chi_1 & \cdots & \chi_t \end{bmatrix}^\top \mathsf{B}_* + \begin{bmatrix} \mathsf{W}_1 & \cdots & \mathsf{W}_t \end{bmatrix}^\top, \tag{C.3b}$$

$$\begin{bmatrix} \mathsf{E} & \mathsf{Z} & \mathsf{Z}_1 & \cdots & \mathsf{Z}_t \end{bmatrix}^\top \sim P_{\mathsf{E}} \otimes \mathcal{N}(0_t, \Sigma_{t+1}). \tag{C.3c}$$

上面公式中的状态演化参数 $\Omega_t \in \mathbb{R}^{t \times t}$, $\chi_t \in \mathbb{R}$ 和 $\Sigma_{t+1} \in \mathbb{R}^{(t+1)\times(t+1)}$ 定义如下。首先递归地定义半无限数组 $(\Omega_{i,\,j})_{(i,j)\in\mathbb{Z}_{\geq 1}}$、$(\Sigma_{i,\,j})_{(i,j)\in\mathbb{Z}_{\geq 1}}$ 和 $(\chi_i)_{i\in\mathbb{Z}_{\geq 1}}$，并以如下内容初始化

$$\Sigma_{1,1} = \sigma^2, \qquad \Omega_{1,1} = \delta \frac{1}{\overline{\kappa}_2} \mathbb{E}\big[g_1(\mathsf{Z}, \mathsf{E})^2\big] + \delta \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} \rho \mathbb{E}\big[\partial_z g_1(\mathsf{Z}, \mathsf{E})\big]^2, \qquad \chi_1 = \delta \mathbb{E}[\partial_z g_1(\mathsf{Z}, \mathsf{E})]. \tag{C.3d}$$

下文中，我们用 $\partial_z g_t(z_{t-1}; z, e)$ 表示 g_t 关于其第二个参数的偏导数。对于任意 $t \geq 1$、$1 \leq i \leq t$ 和 $0 \leq j \leq t$,

$$\Sigma_{t+1,1} = \overline{\kappa}_2 \rho \chi_t + \frac{\overline{\kappa}_4}{\overline{\kappa}_2} \rho \mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1}; \mathsf{Z}, \mathsf{E})], \tag{C.3e}$$

$$\begin{aligned} \Sigma_{t+1,i+1} = {} & \overline{\kappa}_2 (\chi_t \chi_i \rho + \Omega_{t,i}) + \frac{\overline{\kappa}_4}{\overline{\kappa}_2} \big(\mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1}; \mathsf{Z}, \mathsf{E})] \rho \chi_i + \mathbb{E}[\partial_z g_i(\mathsf{Z}_{i-1}; \mathsf{Z}, \mathsf{E})] \rho \chi_t \big) \\ & + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} \mathbb{E}[g_t(\mathsf{Z}_{t-1}; \mathsf{Z}, \mathsf{E}) g_i(\mathsf{Z}_{i-1}; \mathsf{Z}, \mathsf{E})] + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2} \rho \mathbb{E}[\partial_z g_i(\mathsf{Z}_{i-1}; \mathsf{Z}, \mathsf{E})] \mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1}; \mathsf{Z}, \mathsf{E})], \end{aligned} \tag{C.3f}$$

$$\Omega_{t+1,j+1} = \delta \bigg(\frac{1}{\overline{\kappa}_2} \mathbb{E}[g_{t+1}(\mathsf{Z}_t; \mathsf{Z}, \mathsf{E}) g_{j+1}(\mathsf{Z}_j; \mathsf{Z}, \mathsf{E})] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} \rho \mathbb{E}[\partial_z g_{t+1}(\mathsf{Z}_t; \mathsf{Z}, \mathsf{E})] \mathbb{E}[\partial_z g_{j+1}(\mathsf{Z}_j; \mathsf{Z}, \mathsf{E})] \bigg), \tag{C.3g}$$

$$\chi_{t+1} = \delta \mathbb{E}[\partial_z g_{t+1}(\mathsf{Z}_t; \mathsf{Z}, \mathsf{E})], \tag{C.3h}$$

以及 $\Sigma_{1,\,t+1} = \Sigma_{t+1,\,1}$，$\Sigma_{i+1,\,t+1} = \Sigma_{t+1,\,i+1}$，$\Omega_{j+1,\,t+1} = \Omega_{t+1,\,j+1}$。

最后，通过取上述半无限数组的顺序主子式，可以得到 $\Omega_t = (\Omega_{i,\,j})_{(i,j)\in[t]^2}$ 以及 $\Sigma_{t+1} = (\Sigma_{i,\,j})_{(i,j)\in[t+1]^2}$。

状态演化结果的有效性，下述定理 C.1 在很大程度上依赖于去噪器的散度自由性质（g_{t+1}）$_{t\geq 1}$。

（A11）对于每个 $t \geq 0$，g_t 在其所有参数上都是 Lipschitz连续的。对于每个 $t \geq 1$,

$$\mathbb{E}\big[g'_{t+1}(\mathsf{Z}_t; \mathsf{Z}, \mathsf{E})\big] = 0, \tag{C.4}$$

其中，(Z_t, Z, E) 的分布由 (C.3) 指定。下文中，$g'_{t+1}(z_t; z, e)$ 表示 g_{t+1} 关于其第一个自变量的偏导数。此外，相对于 (Z_t, Z, E) 的概率分布，$g'_{t+1}(z_t; z, \varepsilon)$ 和 $\partial_z g_{t+1}(z_t; z, \varepsilon)$ 以概率 1 连续。

引理 C.1 （状态演化）。 考虑 (C.1) 中的线性化 *GVAMP* 算法，并令 假设 (A11)成立。对于任意 $t \geq 1$,

$$\begin{bmatrix} \varepsilon & z & u^1 & \cdots & u^t & g^1 & g^2 & \cdots & g^{t+1} \end{bmatrix} \begin{bmatrix} \beta_* & v^1 & \cdots & v^t \end{bmatrix} \xrightarrow[W_2]{W_2} \begin{bmatrix} \mathsf{E} & \mathsf{Z} & \mathsf{Z}_1 & \cdots & \mathsf{Z}_t \end{bmatrix} \begin{bmatrix} \mathsf{B}_* & \mathsf{B}_1 & \cdots & \mathsf{B}_t \end{bmatrix},$$
$$\begin{bmatrix} \mathsf{E} & \mathsf{Z} & \mathsf{Z}_1 & \cdots & \mathsf{Z}_t & g_1(\mathsf{Z},\ \mathsf{E}) & g_2(\mathsf{Z}_1;\ \mathsf{Z},\ \mathsf{E}) & \cdots & g_{t+1}(\mathsf{Z}_t;\ \mathsf{Z},\ \mathsf{E}) \end{bmatrix},$$

其中，右侧的随机变量定义于(C.3)。

定理 C.1 的证明见C.1 节。

接下来，我们给出随机变量B_t、Z_t的一个等价表示；证明见C.2 节。

引理 C.2 (替代状态演化)。 随机变量$(\mathsf{B}_i)^{i\in[t]}$、$(\mathsf{Z}_i)^{i\in[t]}$ 定义于(C.3) ， 可等价地写为：

$$\mathsf{B}_i = \chi_i \mathsf{B}_* + \omega_i \mathsf{M}_i, \qquad \mathsf{Z}_i = \nu_i \mathsf{Z} + \tau_i \mathsf{N}_i, \qquad \forall i \in [t].$$

在上面的表达式中，$(\quad \omega_i \quad i \ i\in[t]$ $\quad [t] \quad \tau_i \ i \ i\in[t]$ $\quad {}_* \sim P_{\mathsf{B}_*} \quad {}_i \sim \mathcal{N}\ 0,1 \quad \forall i \in \mathsf{M})$是一个与B独立的中心化高斯向量，且M()；$(\mathsf{N})$是一个与Z(，) 独立的中心化高斯向量，且N()。系数$(\omega_i)^{i\in[t]}$、$(\tau_i)_{i\in[t]} \in \mathbb{R}^t_{>0}$ 以及相关性$(\mathbb{E}[\omega_r \mathsf{M}_r \cdot \omega_s \mathsf{M}_s])^{(r,s)\in[t]^2}$、$(\mathbb{E}[\tau_r \mathsf{N}_r \cdot \nu_s \mathsf{N}_s])^{(r,s)\in[t]^2}$满足以下递推关系。对于 $t = 1$,

$$\chi_1 = \delta \mathbb{E}[\partial_z g_1(\mathsf{Z}, \mathsf{E})], \qquad \omega_1 = \left(\frac{\delta}{\overline{\kappa}_2} \mathbb{E}\big[g_1(\mathsf{Z}, \mathsf{E})^2\big] + \delta \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} \rho \mathbb{E}[\partial_z g_1(\mathsf{Z}, \mathsf{E})]^2 \right)^{1/2}. \tag{C.5}$$

对于 $t \geq 1$,

$$\begin{aligned}
\nu_t &= \chi_t + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} \mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1}; \mathsf{Z}, \mathsf{E})], \\
\tau_t^2 &= \overline{\kappa}_2 \omega_t^2 + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} \mathbb{E}\big[g_t(\mathsf{Z}_{t-1}; \mathsf{Z}, \mathsf{E})^2\big] + \left(\frac{\overline{\kappa}_6}{\kappa_2^2} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3} \right) \rho \mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1}; \mathsf{Z}, \mathsf{E})]^2, \\
\chi_{t+1} &= \delta \mathbb{E}[\partial_z g_{t+1}(\mathsf{Z}_t; \mathsf{Z}, \mathsf{E})], \\
\omega_{t+1}^2 &= \frac{\delta}{\overline{\kappa}_2} \mathbb{E}\big[g_{t+1}(\mathsf{Z}_t; \mathsf{Z}, \mathsf{E})^2\big] + \delta \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} \rho \mathbb{E}[\partial_z g_{t+1}(\mathsf{Z}_t; \mathsf{Z}, \mathsf{E})]^2.
\end{aligned}$$

对于 $r, s \geq 1$,

$$\begin{aligned}
\mathbb{E}[\tau_r \mathsf{N}_r \cdot \tau_s \mathsf{N}_s] &= \overline{\kappa}_2 \mathbb{E}[\omega_r \mathsf{M}_r \cdot \omega_s \mathsf{M}_s] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} \mathbb{E}[g_r(\mathsf{Z}_{r-1}; \mathsf{Z}, \mathsf{E}) g_s(\mathsf{Z}_{s-1}; \mathsf{Z}, \mathsf{E})] \\
&\quad + \left(\frac{\overline{\kappa}_6}{\overline{\kappa}_2^2} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3} \right) \rho \mathbb{E}[\partial_z g_r(\mathsf{Z}_{r-1}; \mathsf{Z}, \mathsf{E})] \mathbb{E}[\partial_z g_s(\mathsf{Z}_{s-1}; \mathsf{Z}, \mathsf{E})], \\
\mathbb{E}[\omega_r \mathsf{M}_r \cdot \omega_s \mathsf{M}_s] &= \delta \left(\frac{1}{\overline{\kappa}_2} \mathbb{E}[g_r(\mathsf{Z}_{r-1}; \mathsf{Z}, \mathsf{E}) g_s(\mathsf{Z}_{s-1}; \mathsf{Z}, \mathsf{E})] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} \rho \mathbb{E}[\partial_z g_r(\mathsf{Z}_{r-1}; \mathsf{Z}, \mathsf{E})] \mathbb{E}[\partial_z g_s(\mathsf{Z}_{s-1}; \mathsf{Z}, \mathsf{E})] \right).
\end{aligned}$$

正如第 5 节中的启发式方法所建议的，我们将考虑 g_{t+1} 以下形式且满足假设 (A11)。

(A12) 函数 g_{t+1} 在(C.1) 具有如下形式

$$g_{t+1}(u^t; z, \varepsilon) = \overline{g}_{t+1}(q(z, \varepsilon)) u^t \tag{C.6a}$$

对于满足以下条件的某个函数g_{t+1}：

$$\mathbb{E}\big[\overline{g}_{t+1}(\mathsf{Y})\big] = 0, \qquad \mathbb{E}\Big[\overline{g}_{t+1}(\mathsf{Y}) \overline{\mathsf{Z}}^2\Big] > 0. \tag{C.6b}$$

将定理 C.1 专门应用于此情形，可得以下定理 C.3 其证明见C.3 节。

引理 C.3（状态演化、信噪比）。 考虑 (C.1) 中的线性化 *GVAMP*。设 $\chi_1 \neq 0$，则 ν_t，$\chi_{t+1} \neq 0$ 且 τ_t，$\omega_t > 0$ 对每个 $t \geq 1$ 均成立。此外，可由 (C.3) 中的状态演化推导出以下递推关系。对于任意 $t \geq 1$，$1 \leq i \leq t$ 以及$0 \leq j \leq t$，

$$\frac{\tau_t^2}{\nu_t^2} = \frac{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)\mathbb{E}\left[\overline{g}_t(\mathsf{Y})^2\right]}{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)^2 \mathbb{E}\left[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2} \frac{\tau_{t-1}^2}{\nu_{t-1}^2} + \frac{\rho\left\{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2} + \delta\overline{\kappa}_2\right)\mathbb{E}\left[\overline{g}_t(\mathsf{Y})^2\overline{\mathsf{Z}}^2\right] + \left(\frac{\overline{\kappa}_6}{\overline{\kappa}_2^2} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3} + \delta\frac{\overline{\kappa}_4}{\overline{\kappa}_2}\right)\mathbb{E}\left[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2\right\}}{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)^2 \mathbb{E}\left[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2}. \tag{C.7}$$

$$\frac{\omega_{t+1}^2}{\chi_{t+1}^2} = \frac{\overline{\kappa}_2^{-1}\mathbb{E}\left[\overline{g}_{t+1}(\mathsf{Y})^2\right]}{\delta\mathbb{E}\left[\overline{g}_{t+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2} \frac{\tau_t^2}{\nu_t^2} + \frac{\rho\left\{\mathbb{E}\left[\overline{g}_{t+1}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\right] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\left[\overline{g}_{t+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2\right\}}{\delta\mathbb{E}\left[\overline{g}_{t+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2}, \tag{C.8}$$

$$\begin{aligned}\Sigma_{t+1,i+1} &= \overline{\kappa}_2(\chi_t\chi_i\rho + \Omega_{t,i}) + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\rho\left(\mathbb{E}\left[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\right]\nu_{t-1}\chi_i + \mathbb{E}\left[\overline{g}_i(\mathsf{Y})\overline{\mathsf{Z}}^2\right]\nu_{i-1}\chi_t\right) \\ &\quad + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}[\overline{g}_t(\mathsf{Y})\overline{g}_i(\mathsf{Y})]\mathbb{E}[\tau_{t-1}\mathsf{N}_{t-1}\cdot\tau_{i-1}\mathsf{N}_{i-1}] \\ &\quad + \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2}\mathbb{E}\left[\overline{g}_t(\mathsf{Y})\overline{g}_i(\mathsf{Y})\overline{\mathsf{Z}}^2\right] + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}\mathbb{E}\left[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\right]\mathbb{E}\left[\overline{g}_i(\mathsf{Y})\overline{\mathsf{Z}}^2\right]\right)\rho\nu_{t-1}\nu_{i-1},\end{aligned} \tag{C.9}$$

$$\Omega_{t+1,j+1} = \delta\left\{\rho\left(\mathbb{E}\left[\overline{g}_{t+1}(\mathsf{Y})\overline{g}_{j+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\right] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\left[\overline{g}_{t+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]\mathbb{E}\left[\overline{g}_{j+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]\right)\nu_t\nu_j + \frac{1}{\overline{\kappa}_2}\mathbb{E}[\overline{g}_{t+1}(\mathsf{Y})\overline{g}_{j+1}(\mathsf{Y})]\mathbb{E}[\tau_t\mathsf{N}_t\cdot\tau_j\mathsf{N}_j]\right\}. \tag{C.10}$$

若进一步选取 $g_{t+1} = g$ 为时间不变的，则状态演化具有以下不动点。证明见 C.4 节 。

引理 C.4（状态演化不动点）。 取 g_{t+1} 为 (C.6a) 的形式，其中 $g_{t+1} = g$ 满足(C.6b)。将 (C.7) 和 (C.8) 中状态演化参数的不动点记为去掉时间指标后的相同记号。假设 (4.13) 成立，则这些不动点满足：

$$\frac{\overline{\tau}^2}{\overline{\nu}^2} = w_1 > 0, \qquad \frac{\overline{\omega}^2}{\overline{\chi}^2} = w_2 > 0.$$

最后，上述要素使我们能够在第 C.5 节中证明定理 B.1。

C.1 定理 C.1 的证明

根据(C.2)中的特化，(C.1) 中的线性化 GVAMP 是(4.19)中通用 GVAMP 的一个实例。因此，其状态演化 (C.3) 可通过对(4.29)进行特化推导。不过，为了在技术上更方便地清晰展现 [101, 公式 (4) 和 (5)]中所研究的旋转不变广义近似消息传递（RI-GAMP）算法，我们将从该算法推导其状态演化，以使自由累积量在 (C.3) 中的出现更加直观。

RI-GAMP 迭代如下：对于 $t \geq 1$，

$$b^t = X^\top h^t - \sum_{i=1}^{t-1}\beta_{t,i}f^i, \qquad f^t = f_t(b^1, \cdots, b^t), \tag{C.11a}$$

$$z^t = Xf^t - \sum_{i=1}^{t}\alpha_{t,i}h^i, \qquad h^{t+1} = h_{t+1}(z^1, \cdots, z^t; z, \varepsilon), \tag{C.11b}$$

以 $h^1 = h_1(z, \varepsilon)$ 初始化。

命题 C.5 ([101, 定理 1])。 对于任意固定的 $t \geq 1$，存在随机变量 Z、E、$(\mathsf{B}_i)^{i\in[t]}$，$(\mathsf{Z}_i)^{i\in[t]}$（定义见 *[101, 方程 (11) 至 (22)]*），使得若满足以下条件

(A13) 对于每个 $1 \leq i \leq t$，$f_i(b_1, \cdots, b_i)$ 以及 $h_i(z_1, \cdots, z_{i-1};\ z, \varepsilon)$，它们关于各自的每个自变量都是 *Lipschitz*连续的；

(A14) 对于每个 $1 \leq i \leq t$、$1 \leq j \leq i$ 和 $1 \leq k \leq i-1$，偏导数 $\partial_{x_j} f_i(x_1, \cdots, x_i)$、$\partial_{z_k} h_i(z_1, \cdots, z_{i-1}, z, \varepsilon)$ 以及 $\partial_z h_i(z_1, \cdots, z_{i-1}, z, \varepsilon)$ 关于 Z、E、$(\mathsf{B}_i)^{i\in[t]}$、$(\mathsf{Z}_i)^{i\in[t]}$ 的联合分布，除一个测度为 0 的集合外，均连续，

then

$$\begin{bmatrix}\beta_* & v^1 & \cdots & v^t & f^1 & \cdots & f^t\end{bmatrix} \xrightarrow{W_2} \begin{bmatrix}\mathsf{B}_* & \mathsf{B}_1 & \cdots & \mathsf{B}_t & \mathsf{F}_1 & \cdots & \mathsf{F}_t\end{bmatrix},$$
$$\begin{bmatrix}\varepsilon & z & u^1 & \cdots & u^t & h^1 & \cdots & h^t\end{bmatrix} \xrightarrow{W_2} \begin{bmatrix}\mathsf{E} & \mathsf{Z} & \mathsf{Z}_1 & \cdots & \mathsf{Z}_t & \mathsf{H}_1 & \cdots & \mathsf{H}_t\end{bmatrix},$$

其中 $\mathsf{F}_i := f_i(\mathsf{B}_1, \cdots, \mathsf{B}_i)$，以及 $\mathsf{H}_i := h_i(\mathsf{Z}_1, \cdots, \mathsf{Z}_{i-1}; \mathsf{Z}, \mathsf{E})$。

考虑上述迭代的一个特殊情形，其中

$$f_t(b^t) = b^t, \qquad h_{t+1}(z^t; z, \varepsilon) = \overline{\kappa}_2^{-1} g_{t+1}(z^t; z, \varepsilon), \qquad h_1(z, \varepsilon) = \overline{\kappa}_2^{-1} g_1(z, \varepsilon). \tag{C.12}$$

假设 (A11) 保证 (C.12) 中的函数 (C.12) 满足 假设 (A13) 和 (A14)。我们可以将(C.11) 写成

$$b^t = X^\top h^t - \varepsilon^t, \tag{C.13a}$$
$$z^t = X b^t - \overline{\kappa}_2 h^t - \delta^t, \qquad h^{t+1} = h_{t+1}(z^t; z, \varepsilon), \tag{C.13b}$$

其中

$$\varepsilon^t := \sum_{i=1}^{t-1} \beta_{t,i} b^i, \qquad \delta^t := \sum_{i=1}^{t-1} \alpha_{t,i} h^i + (\alpha_{t,t} - \overline{\kappa}_2) h^t.$$

利用 h_{t+1} 在 (C.12) 中的选取，并结合 (C.13a) 和 (C.13b)，可得

$$\begin{aligned} b^t &= \overline{\kappa}_2^{-1} X^\top g^t - \varepsilon^t, \\ z^t &= \big(XX^\top - \overline{\kappa}_2 I_n\big) h^t - X\varepsilon^t - \delta^t \qquad\qquad g^{t+1} = g_{t+1}(z^t; z, \varepsilon). \\ &= \big(\overline{\kappa}_2^{-1} XX^\top - I_n\big) g^t - X\varepsilon^t - \delta^t, \end{aligned}$$

将上述迭代与 (C.1)进行比较，我们发现

$$v^t = b^t + \varepsilon^t, \qquad u^t = z^t + X\varepsilon^t + \delta^t. \tag{C.14}$$

此外，根据$(\alpha_{t,\ i})^{i\in[t]}$，$(\beta_{t,\ i})^{i\in[t-1]}$ 评估去偏系数 [101, 公式 (6) 至 (10)]，我们得到：

$$\lim_{n\to\infty} \alpha_{t,t} = \overline{\kappa}_2, \qquad \lim_{n\to\infty} \alpha_{t,i} = \lim_{n\to\infty} \beta_{t,i} = 0, \qquad \forall i \in [t-1].$$

这一步关键依赖于 g_{t+1} 在 假设 (A11) 下具有散度自由性质，这使得 (C.11)中的所有记忆项系数均为零，从而使得到的状态演化递推更易处理。由于 $\lim_{n\to\infty} \|X\|_2 < \infty$ 根据 假设（A2），

$$\lim_{n\to\infty} n^{-1/2} \big\|X\varepsilon^t\big\|_2 = \lim_{d\to\infty} d^{-1/2} \big\|\varepsilon^t\big\|_2 = \lim_{n\to\infty} n^{-1/2} \big\|\delta^t\big\|_2 = 0, \tag{C.15}$$

因此也有 $\lim_{n\to\infty} n^{-1/2}\|X\varepsilon^t+\delta^t\|_2=0$。我们已经证明，迭代 (C.1) 除迭代量中的消失扰动 b^t之外，可以表示为 z^t 的一个实例。我们声称，[101, 定理 1] 中的相同状态演化结果适用于迭代 (C.1) 和 (C.11)。为此，任取一个测试函数 $f\in \mathrm{PL}^2_{(t+1)\to 1^\circ}$ 那么存在 $C>0$ 使得

$$\left|\frac{1}{d}\sum_{i=1}^d f(\beta_{*,i},v_i^1,\cdots,v_i^t)-\frac{1}{d}\sum_{i=1}^d f(\beta_{*,i},b_i^1,\cdots,b_i^t)\right|\le \frac{1}{d}\sum_{i=1}^d\left|f(\beta_{*,i},v_i^1,\cdots,v_i^t)-f(\beta_{*,i},b_i^1,\cdots,b_i^t)\right| \tag{C.16}$$

$$\le \frac{C}{d}\sum_{i=1}^d\left\|(v_i^j)_{j\in[t]}-(b_i^j)_{j\in[t]}\right\|_2\left(1+\left\|(v_i^j)_{j\in[t]}\right\|_2+\left\|(b_i^j)_{j\in[t]}\right\|_2\right) \tag{C.17}$$

$$=\frac{C}{d}\sum_{i=1}^d\left\|(\varepsilon_i^j)_{j\in[t]}\right\|_2\left(1+\left\|(v_i^j)_{j\in[t]}\right\|_2+\left\|(b_i^j)_{j\in[t]}\right\|_2\right) \tag{C.18}$$

$$\le \frac{C}{d}\sqrt{\sum_{i=1}^d\left\|(\varepsilon_i^j)_{j\in[t]}\right\|_2^2}\sqrt{\sum_{i=1}^d\left(1+\left\|(v_i^j)_{j\in[t]}\right\|_2+\left\|(b_i^j)_{j\in[t]}\right\|_2\right)^2} \tag{C.19}$$

$$\le \frac{\sqrt{3}C}{d}\sqrt{\sum_{i=1}^d\left\|(\varepsilon_i^j)_{j\in[t]}\right\|_2^2}\sqrt{\sum_{i=1}^d\left(1+\left\|(v_i^j)_{j\in[t]}\right\|_2^2+\left\|(b_i^j)_{j\in[t]}\right\|_2^2\right)} \tag{C.20}$$

$$=\sqrt{3}C\sqrt{\sum_{j=1}^t d^{-1}\|\varepsilon^j\|_2^2}\sqrt{1+\sum_{j=1}^t d^{-1}\|v^j\|_2^2+\sum_{j=1}^t d^{-1}\|b^j\|_2^2}$$

$$\le \sqrt{3}C\sqrt{\sum_{j=1}^t d^{-1}\|\varepsilon^j\|_2^2}\sqrt{1+2\sum_{j=1}^t d^{-1}\|\varepsilon^j\|_2^2+3\sum_{j=1}^t d^{-1}\|b^j\|_2^2}. \tag{C.21}$$

在上述（不）等式链中，(C.17) 由 f 的 PL^2 性质得出（见定义 3.1）；(C.18) 由 (C.14) 得到；(C.16) 与 (C.19) 均由柯西–施瓦茨不等式得到；(C.20) 使用初等不等式 $(a+b+c)^2\le 3(a^2+b^2+c^2)$，对任意 a、b、$c\in\mathbb{R}$ 成立；(C.21) 成立是因为 $\|u+v\|_2^2\le 2(\|u\|_2^2+\|v\|_2^2)$。状态演化结果 [101, 定理 1] 适用于 $(b^j)_{j\in[t]}$，并特别蕴含对每个 $j\in[t]$ 有 $\lim_{d\to\infty} d^{-1/2}\|b^j\|_2<\infty$。由于$t$ 是固定的，这与 (C.15) 相结合表明

$$\lim_{d\to\infty}\left|\frac{1}{d}\sum_{i=1}^d f(\beta_{*,i},v_i^1,\cdots,v_i^t)-\frac{1}{d}\sum_{i=1}^d f(\beta_{*,i},b_i^1,\cdots,b_i^t)\right|=0.$$

类似的论证表明，

$$\lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^n g(\varepsilon_i,z_i,u_i^1,\cdots,u_i^t,g_i^1,\cdots,g_i^t)-\frac{1}{n}\sum_{i=1}^n g(\varepsilon_i,z_i,z_i^1,\cdots,z_i^t,\overline{\kappa}_2 h_i^1,\cdots,\overline{\kappa}_2 h_i^t)\right|=0,$$

对于任意 $g\in\mathrm{PL}^2_{2(t+1)\to 1^\circ}$。因此，Wasserstein-2 极限 $\begin{bmatrix}\beta_* & v^1 & \cdots & v^t\end{bmatrix}$ 和 $\begin{bmatrix}\varepsilon & z & u^1 & \cdots & u^t & g^1 & \cdots g^t\end{bmatrix}$分别与 $\begin{bmatrix}\beta_* & b^1 & \cdots & b^t\end{bmatrix}$ 和 $\begin{bmatrix}\varepsilon & z & z^1 & \cdots & z^t & \kappa_2 h^1 & \cdots & \kappa_2 h^t\end{bmatrix}$相同，而状态演化结果 [101, 定理 1]适用于后者。

接下来，我们将 [101, 方程（11）至（22）] 中的状态演化递推专门化为（C.13），它对（C.1）同样成立。定义随机变量：

$$\begin{bmatrix}\mathsf{E} & \mathsf{Z} & \mathsf{Z}_1 & \cdots & \mathsf{Z}_{t-1}\end{bmatrix}^\top\sim P_{\mathsf{E}}\otimes\mathcal{N}(0_t,\Sigma_t), \tag{C.22a}$$

$$\widetilde{\mathsf{Z}}_t=h_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E}), \tag{C.22b}$$

$$\begin{bmatrix}\mathsf{B}_* & \mathsf{W}_1 & \cdots & \mathsf{W}_t\end{bmatrix}^\top\sim P_{\mathsf{B}_*}\otimes\mathcal{N}(0_t,\Omega_t), \tag{C.22c}$$

$$\begin{bmatrix}\mathsf{B}_1 & \cdots & \mathsf{B}_t\end{bmatrix}^\top = \begin{bmatrix}\chi_1 & \cdots & \chi_t\end{bmatrix}^\top \mathsf{B}_* + \begin{bmatrix}\mathsf{W}_1 & \cdots & \mathsf{W}_t\end{bmatrix}^\top. \tag{C.22d}$$

状态演化　　参数 $\Sigma_t \in \mathbb{R}^{t\times t}$、$\Omega_t \in \mathbb{R}^{t\times t}$、$\chi_t \in \mathbb{R}$ 递归定义如下。对于 $t = 1$,

$$\Sigma_1 = \sigma^2, \qquad \Omega_1 = \delta\overline{\kappa}_2\mathbb{E}\big[h_1(\mathsf{Z},\mathsf{E})^2\big] + \delta\overline{\kappa}_4\rho\mathbb{E}[\partial_z h_1(\mathsf{Z},\mathsf{E})]^2, \qquad \mu_1 = \delta\overline{\kappa}_2\mathbb{E}[\partial_z h_1(\mathsf{Z},\mathsf{E})], \tag{C.22e}$$

其中（Z、E）$\sim \mathcal{N}(0,\ \sigma^2)\otimes P_{\mathsf{E}}$。

对于 $t \geq 1$，现在说明如何从 Σ_t、Ω_t、$(\chi_i)^{i\in[t]}$ 计算出 Σ_{t+1}，Ω^{t+1}，χ_{t+1}。定义

$$\begin{aligned}
\phi_t &\coloneqq \begin{bmatrix}\mathbb{E}[\partial_z h_1(\mathsf{Z},\mathsf{E})] & \mathbb{E}[\partial_z h_2(\mathsf{Z}_1;\mathsf{Z},\mathsf{E})] & \cdots & \mathbb{E}[\partial_z h_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})]\end{bmatrix}^\top \in \mathbb{R}^t,\\
\gamma_t &\coloneqq \begin{bmatrix}\mathbb{E}[\mathsf{B}_*\mathsf{B}_1] & \cdots & \mathbb{E}[\mathsf{B}_*\mathsf{B}_t]\end{bmatrix}^\top \in \mathbb{R}^t,\\
\Gamma_t &\coloneqq \big[\mathbb{E}[\mathsf{B}_i\mathsf{B}_j]\big]_{(i,j)\in[t]^2} \in \mathbb{R}^{t\times t},\\
\Delta_t &\coloneqq \Big[\mathbb{E}\Big[\widetilde{\mathsf{Z}}_i\widetilde{\mathsf{Z}}_j\Big]\Big]_{(i,j)\in[t]^2} \in \mathbb{R}^{t\times t}.
\end{aligned}$$

随后将 Σ_{t+1} 更新为：

$$\Sigma_{t+1} = \overline{\kappa}_2\begin{bmatrix}\rho & \gamma_t^\top\\ \gamma_t & \Gamma_t\end{bmatrix} + \overline{\kappa}_4\left(\begin{bmatrix}0 & \rho\phi_t^\top\\ \rho\phi_t & \phi_t\gamma_t^\top + \gamma_t\phi_t^\top\end{bmatrix} + \begin{bmatrix}0 & 0_t^\top\\ 0_t & \Delta_t\end{bmatrix}\right) + \overline{\kappa}_6\begin{bmatrix}0 & 0_t^\top\\ 0_t & \rho\phi_t\phi_t^\top\end{bmatrix}.$$

矩阵 Ω_{t+1} 更新为：

$$\Omega_{t+1} \coloneqq \delta\big(\overline{\kappa}_2\Delta_{t+1} + \overline{\kappa}_4\rho\phi_{t+1}\phi_{t+1}^\top\big).$$

标量 χ_{t+1} 更新为：

$$\chi_{t+1} = \delta\overline{\kappa}_2\phi_{t+1,t+1},$$

其中 $\phi_{t+1,\ t+1}$ 表示向量 $\phi_{t+1} \in \mathbb{R}^{t+1}$ 的最后一个元素。

注意，Σ_t，Ω_t 分别是 Σ_{t+1}，Ω^{t+1}的顺序主子式（阶数为 t）。因此，我们明确写出 Σ_{t+1}，Ω^{t+1} 和 χ_{t+1}的最后一行和列的更新规则。对于 $1 \leq i \leq t$ 和 $0 \leq j \leq t$,

$$\Sigma_{t+1,1} = \overline{\kappa}_2\mathbb{E}[\mathsf{B}_*\mathsf{B}_t] + \overline{\kappa}_4\rho\mathbb{E}[\partial_z h_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})] = \overline{\kappa}_2\rho\chi_t + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\rho\mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})], \tag{C.22f}$$

$$\begin{aligned}
\Sigma_{t+1,i+1} &= \overline{\kappa}_2\mathbb{E}[\mathsf{B}_t\mathsf{B}_i] + \overline{\kappa}_4(\mathbb{E}[\partial_z h_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})]\mathbb{E}[\mathsf{B}_*\mathsf{B}_i] + \mathbb{E}[\partial_z h_i(\mathsf{Z}_{i-1};\mathsf{Z},\mathsf{E})]\mathbb{E}[\mathsf{B}_*\mathsf{B}_t])\\
&\quad + \overline{\kappa}_4\mathbb{E}\Big[\widetilde{\mathsf{Z}}_t\widetilde{\mathsf{Z}}_i\Big] + \overline{\kappa}_6\rho\mathbb{E}[\partial_z h_i(\mathsf{Z}_{i-1};\mathsf{Z},\mathsf{E})]\mathbb{E}[\partial_z h_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})]\\
&= \overline{\kappa}_2(\chi_t\chi_i\rho + \Omega_{t,i}) + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}(\mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})]\rho\chi_i + \mathbb{E}[\partial_z g_i(\mathsf{Z}_{i-1};\mathsf{Z},\mathsf{E})]\rho\chi_t)\\
&\quad + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}[g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})g_i(\mathsf{Z}_{i-1};\mathsf{Z},\mathsf{E})] + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}\rho\mathbb{E}[\partial_z g_i(\mathsf{Z}_{i-1};\mathsf{Z},\mathsf{E})]\mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})],
\end{aligned} \tag{C.22g}$$

$$\begin{aligned}
\Omega_{t+1,j+1} &= \delta\Big(\overline{\kappa}_2\mathbb{E}\Big[\widetilde{\mathsf{Z}}_{t+1}\widetilde{\mathsf{Z}}_{j+1}\Big] + \overline{\kappa}_4\rho\mathbb{E}[\partial_z h_{t+1}(\mathsf{Z}_t;\mathsf{Z},\mathsf{E})]\mathbb{E}[\partial_z h_{j+1}(\mathsf{Z}_j;\mathsf{Z},\mathsf{E})]\Big)\\
&= \delta\left(\frac{1}{\overline{\kappa}_2}\mathbb{E}[g_{t+1}(\mathsf{Z}_t;\mathsf{Z},\mathsf{E})g_{j+1}(\mathsf{Z}_j;\mathsf{Z},\mathsf{E})] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\rho\mathbb{E}[\partial_z g_{t+1}(\mathsf{Z}_t;\mathsf{Z},\mathsf{E})]\mathbb{E}[\partial_z g_{j+1}(\mathsf{Z}_j;\mathsf{Z},\mathsf{E})]\right),
\end{aligned} \tag{C.22h}$$

$$\chi_{t+1} = \delta\overline{\kappa}_2\mathbb{E}[\partial_z h_{t+1}(\mathsf{Z}_t;\mathsf{Z},\mathsf{E})] = \delta\mathbb{E}[\partial_z g_{t+1}(\mathsf{Z}_t;\mathsf{Z},\mathsf{E})], \tag{C.22i}$$

其中我们回顾 h_t 在 (C.12)中的选取。

我们注意到，状态演化递推 (C.22) 与 (C.3) 相同，从而完成证明。

C.2 定理C.2 的证明

根据 (C.3)，$(\mathsf{Z}、\mathsf{Z}_1、\cdots、\mathsf{Z}_t)$ 是一个中心化高斯向量。因此，我们可以等价地写成

$$\mathsf{Z}_i = \nu_i \mathsf{Z} + \tau_i \mathsf{N}_i, \qquad i \in [t]$$

其中，$(\mathsf{N}_i)_{i\in[t]} \sim \mathcal{N}(0_t,\ I_t)$是一个与 Z 独立的标准高斯向量。系数 $(\nu_i)^{i\in[t]}$，$(\tau_i)^{i\in[t]}$以及相关性 $(\mathbb{E}[\tau_r \mathsf{N}_r \cdot \tau_s \mathsf{N}_s])_{(r,s)\in[t]^2}$ 可由协方差矩阵 Σ_{t+1}推导得出。

为此，我们首先注意到，$(\mathsf{Z},\ \mathsf{Z}_t)$ 的协方差结构为

$$\begin{aligned}
&\mathbb{E}\big[\mathsf{Z}^2\big] = \sigma^2,\\
&\mathbb{E}[\mathsf{Z}\mathsf{Z}_t] = \overline{\kappa}_2 \rho \chi_t + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\rho \mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})],\\
&\mathbb{E}\big[\mathsf{Z}_t^2\big] = \overline{\kappa}_2\big(\chi_t^2\rho + \Omega_{t,t}\big) + 2\frac{\overline{\kappa}_4}{\overline{\kappa}_2}\mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})]\rho\chi_t + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\big[g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})^2\big] + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}\rho\mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})]^2.
\end{aligned}$$

因此，系数 ν_t，τ_t 为

$$\begin{aligned}
\nu_t &= \frac{\mathbb{E}[\mathsf{Z}\mathsf{Z}_t]}{\sigma^2} = \chi_t + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})],\\
\tau_t^2 &= \mathbb{E}\big[\mathsf{Z}_t^2\big] - \nu_t^2\sigma^2 = \overline{\kappa}_2\big(\chi_t^2\rho + \Omega_{t,t}\big) + 2\frac{\overline{\kappa}_4}{\overline{\kappa}_2}\mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})]\rho\chi_t + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\big[g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})^2\big] + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}\rho\mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})]^2\\
&\quad - \left(\chi_t + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})]\right)^2\sigma^2,\\
&= \overline{\kappa}_2\Omega_{t,t} + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\big[g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})^2\big] + \left(\frac{\overline{\kappa}_6}{\overline{\kappa}_2^2} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3}\right)\rho\mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})]^2. \qquad \text{(C.23)}
\end{aligned}$$

类似地，相关性 $\mathbb{E}[\tau_r \mathsf{N}_r \cdot \tau_s \mathsf{N}_s]$ 为

$$\begin{aligned}
&\mathbb{E}[\tau_r \mathsf{N}_r \cdot \tau_s \mathsf{N}_s] = \Sigma_{r+1,s+1} - \nu_r\nu_s\sigma^2\\
&= \overline{\kappa}_2(\chi_r\chi_s\rho + \Omega_{r,s}) + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}(\mathbb{E}[\partial_z g_r(\mathsf{Z}_{r-1};\mathsf{Z},\mathsf{E})]\rho\chi_s + \mathbb{E}[\partial_z g_s(\mathsf{Z}_{s-1};\mathsf{Z},\mathsf{E})]\rho\chi_r)\\
&\quad + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}[g_r(\mathsf{Z}_{r-1};\mathsf{Z},\mathsf{E})g_s(\mathsf{Z}_{s-1};\mathsf{Z},\mathsf{E})] + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}\rho\mathbb{E}[\partial_z g_s(\mathsf{Z}_{s-1};\mathsf{Z},\mathsf{E})]\mathbb{E}[\partial_z g_r(\mathsf{Z}_{r-1};\mathsf{Z},\mathsf{E})]\\
&\quad - \left(\chi_r + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}[\partial_z g_r(\mathsf{Z}_{r-1};\mathsf{Z},\mathsf{E})]\right)\left(\chi_s + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}[\partial_z g_s(\mathsf{Z}_{s-1};\mathsf{Z},\mathsf{E})]\right)\sigma^2\\
&= \overline{\kappa}_2\Omega_{r,s} + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}[g_r(\mathsf{Z}_{r-1};\mathsf{Z},\mathsf{E})g_s(\mathsf{Z}_{s-1};\mathsf{Z},\mathsf{E})] + \left(\frac{\overline{\kappa}_6}{\overline{\kappa}_2^2} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3}\right)\rho\mathbb{E}[\partial_z g_r(\mathsf{Z}_{r-1};\mathsf{Z},\mathsf{E})]\mathbb{E}[\partial_z g_s(\mathsf{Z}_{s-1};\mathsf{Z},\mathsf{E})].
\end{aligned}$$

注意，对于 $r = s = t$，上述表达式与 τ_t^2 中的 (C.23) 相同。

转到 B_{t+1}，表示式 $\mathsf{B}_{t+1} = \chi_{t+1}\mathsf{B}_* + \omega_{t+1}\mathsf{M}_{t+1}$（其中 $(\mathsf{B}_*,\ \mathsf{M}_{t+1}) \sim P_{\mathsf{B}_*} \otimes \mathcal{N}(0,\ 1)$）只需重新定义 $\mathsf{M}_{t+1} = \mathsf{W}_{t+1}/\sqrt{\Omega_{t+1,\ t+1}}$ 以及 $\omega_{t+1} = \sqrt{\Omega_{t+1,\ t+1}}$ 其中右端如(C.3)所示。

C.3 定理C.3的证明

首先，我们运用 Stein 引理和导数链式法则，推导出 $\mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1};\ \mathsf{Z},\ \mathsf{E})]$ (它用于 定理 C.2)：

$$\mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})] = \sigma^{-2}\mathbb{E}[\mathsf{Z} g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})] - \nu_{t-1}\mathbb{E}[g_t'(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})].$$

当 g_t 具有(C.6)的形式时，这个恒等式变为

$$\mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})] = \sigma^{-2}\mathbb{E}[\mathsf{Z}\overline{g}_t(\mathsf{Y})\mathsf{Z}_{t-1}] - \nu_{t-1}\mathbb{E}[\overline{g}_t(\mathsf{Y})] = \sigma^{-2}\nu_{t-1}\mathbb{E}\big[\overline{g}_t(\mathsf{Y})\mathsf{Z}^2\big] = \nu_{t-1}\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\Big], \quad \text{(C.24)}$$

这里回顾一下 $\mathsf{Z}\overline{=}\sigma^{-1}\mathsf{Z}$，它在 第 3.2 节 中定义。

现在，将定理 C.2 应用于(C.6)这一情形，可得

$$\nu_t = \left(1 + \frac{\overline{\kappa}_4}{\delta\overline{\kappa}_2^2}\right)\chi_t, \quad \text{(C.25)}$$

$$\begin{aligned}\tau_t^2 &= \overline{\kappa}_2\omega_t^2 + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\big[\overline{g}_t(\mathsf{Y})^2\mathsf{Z}_{t-1}^2\big] + \left(\frac{\overline{\kappa}_6}{\kappa_2^2} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3}\right)\rho\mathbb{E}[\partial_z g_t(\mathsf{Z}_{t-1};\mathsf{Z},\mathsf{E})]^2 \\ &= \overline{\kappa}_2\omega_t^2 + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\big(\nu_{t-1}^2\mathbb{E}\big[\overline{g}_t(\mathsf{Y})^2\mathsf{Z}^2\big] + \tau_{t-1}^2\mathbb{E}\big[\overline{g}_t(\mathsf{Y})^2\big]\big) + \left(\frac{\overline{\kappa}_6}{\kappa_2^2} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3}\right)\rho\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2\nu_{t-1}^2 \quad \text{(C.26)} \\ &= \overline{\kappa}_2\omega_t^2 + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\big[\overline{g}_t(\mathsf{Y})^2\big]\tau_{t-1}^2 + \rho\left\{\frac{\overline{\kappa}_4}{\overline{\kappa}_2}\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})^2\overline{\mathsf{Z}}^2\Big] + \left(\frac{\overline{\kappa}_6}{\kappa_2^2} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3}\right)\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2\right\}\nu_{t-1}^2, \quad \text{(C.27)}\end{aligned}$$

$$\chi_{t+1} = \delta\mathbb{E}\Big[\overline{g}_{t+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\nu_t, \quad \text{(C.28)}$$

$$\begin{aligned}\omega_{t+1}^2 &= \frac{\delta}{\overline{\kappa}_2}\mathbb{E}\big[\overline{g}_{t+1}(\mathsf{Y})^2\mathsf{Z}_t^2\big] + \delta\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\rho\mathbb{E}[\partial_z g_{t+1}(\mathsf{Z}_t;\mathsf{Z},\mathsf{E})]^2 \\ &= \frac{\delta}{\overline{\kappa}_2}\big(\nu_t^2\mathbb{E}\big[\overline{g}_{t+1}(\mathsf{Y})^2\mathsf{Z}^2\big] + \tau_t^2\mathbb{E}\big[\overline{g}_{t+1}(\mathsf{Y})^2\big]\big) + \delta\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\rho\mathbb{E}\Big[\overline{g}_{t+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2\nu_t^2 \quad \text{(C.29)} \\ &= \delta\overline{\kappa}_2^{-1}\mathbb{E}\big[\overline{g}_{t+1}(\mathsf{Y})^2\big]\tau_t^2 + \delta\rho\left\{\mathbb{E}\Big[\overline{g}_{t+1}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\Big] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\Big[\overline{g}_{t+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2\right\}\nu_t^2. \quad \text{(C.30)}\end{aligned}$$

等式(C.26)、(C.28) 和(C.29) 均可由(C.24)推出。

根据假设 $\chi_1 \neq 0$、递归关系(C.25) 以及观察(4.5)，可得 $\nu_1 \neq 0$。从(C.25) 和(C.28)中消去χ_t，得到关于 ν_t的以下递归关系：

$$\nu_t = \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\nu_{t-1}. \quad \text{(C.31)}$$

在(C.31) 中结合(4.5) 与假设 (A12)，并对 t进行归纳，可进一步得到 $\nu_t \neq 0$ 对每个$t \geq 1$均成立。再次在(C.28)中使用假设 (A12)，进而可推出 $\chi_{t+1} \neq 0$ 对每个 $t \geq 1$均成立。

接下来，让我们重新 将 ω_1 从(C.5) 中的表达式称为，并作出两个观察。Stein 引理蕴含 s

$$\mathbb{E}[\partial_z g_1(\mathsf{Z},\mathsf{E})] = \sigma^{-2}\mathbb{E}[\mathsf{Z}g_1(\mathsf{Z},\mathsf{E})] = \sigma^{-1}\mathbb{E}\big[\overline{\mathsf{Z}}g_1(\mathsf{Z},\mathsf{E})\big]. \quad \text{(C.32)}$$

柯西–施瓦茨不等式蕴含

$$\mathbb{E}\big[g_1(\mathsf{Z},\mathsf{E})^2\big] \geq \mathbb{E}\big[\overline{\mathsf{Z}}g_1(\mathsf{Z},\mathsf{E})\big]^2. \quad \text{(C.33)}$$

根据上述两个观察，并回顾(3.9) 中 σ^2的定义，

$$\omega_1^2 \geq \frac{\delta}{\overline{\kappa}_2}\sigma^2\mathbb{E}[\partial_z g_1(\mathsf{Z},\mathsf{E})]^2 + \delta\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\rho\mathbb{E}[\partial_z g_1(\mathsf{Z},\mathsf{E})]^2 = \delta\rho\mathbb{E}[\partial_z g_1(\mathsf{Z},\mathsf{E})]^2\left(1 + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\right). \quad \text{(C.34)}$$

由 $\chi_1 \neq 0$ 和 假设（A1）可知，(C.34) 中的第一项因子为正。我们声称，(C.34) 中括号内的第二项因子也为正。根据矩-累积量关系（见 附录 A），

$$1 + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} = \frac{\overline{m}_4}{\overline{m}_2^2} - 1 - \delta.$$

根据定义， $m_{\underline{2k}} = \mathbb{E}\big[\Lambda_n^{2k}\big]$ 其中 Λ_n 由 (3.8) 给出。因此

$$1 + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} = \frac{\mathbb{E}\big[\Lambda_n^4\big]}{\mathbb{E}[\Lambda_n^2]^2} - \delta = \begin{cases} \frac{\mathbb{E}[\Lambda^4]}{\mathbb{E}[\Lambda^2]^2} - \delta, & \delta \le 1 \\ \delta \frac{\mathbb{E}[\Lambda^4]}{\mathbb{E}[\Lambda^2]^2} - \delta, & \delta > 1 \end{cases}. \tag{C.35}$$

在这两种情况下，由柯西–施瓦茨不等式和假设（A2）可知，右侧为正。这证实了 ω_1的正性。

接下来证明 $\tau_1^2 > 0$。由定理 C.2,

$$\begin{aligned} \tau_1^2 &= \overline{\kappa}_2 \omega_1^2 + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} \mathbb{E}\big[g_1(\mathsf{Z}, \mathsf{E})^2\big] + \left(\frac{\overline{\kappa}_6}{\kappa_2^2} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3}\right) \rho \mathbb{E}[\partial_z g_1(\mathsf{Z}, \mathsf{E})]^2 \\ &= \left(\delta + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\right) \mathbb{E}\big[g_1(\mathsf{Z}, \mathsf{E})^2\big] + \left(\delta \frac{\overline{\kappa}_4}{\overline{\kappa}_2} + \frac{\overline{\kappa}_6}{\kappa_2^2} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3}\right) \rho \mathbb{E}[\partial_z g_1(\mathsf{Z}, \mathsf{E})]^2, \end{aligned} \tag{C.36}$$

其中第二步用到了 ω_1的表达式。将(C.32) 和(C.33) 代入(C.36)，得到

$$\tau_1^2 \ge \left(\delta + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \frac{\overline{\kappa}_6}{\kappa_2^3} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^4}\right) \mathbb{E}\big[\overline{\mathsf{Z}} g_1(\mathsf{Z}, \mathsf{E})\big]^2.$$

根据矩-累积量关系（见附录 A），

$$\delta + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \frac{\overline{\kappa}_6}{\kappa_2^3} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^4} = \frac{\overline{m}_2 \overline{m}_6 - \overline{m}_4^2}{\overline{m}_2^4}.$$

根据 m_{2k} 的定义（见附录 A），上面右侧的分子为

$$\overline{m}_2 \overline{m}_6 - \overline{m}_4^2 = \mathbb{E}\big[\Lambda_n^2\big] \mathbb{E}\big[\Lambda_n^6\big] - \mathbb{E}\big[\Lambda_n^4\big]^2 > 0, \tag{C.37}$$

由霍尔德不等式（应用于非负随机变量 $\Lambda_n \cdot \Lambda_n^3$ ；见假设（A2）） 可得。因此$\tau_1^2 > 0$。

消去 ω_t 后，由(C.27) 和 (C.30)可得关于 τ_t 的以下递归:

$$\tau_t^2 = \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right) \mathbb{E}\big[\overline{g}_t(\mathsf{Y})^2\big] \tau_{t-1}^2 + \rho \left\{ \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2} + \delta \overline{\kappa}_2\right) \mathbb{E}\Big[\overline{g}_t(\mathsf{Y})^2 \overline{\mathsf{Z}}^2\Big] + \left(\frac{\overline{\kappa}_6}{\overline{\kappa}_2^2} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3} + \delta \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\right) \mathbb{E}\Big[\overline{g}_t(\mathsf{Y}) \overline{\mathsf{Z}}^2\Big]^2 \right\} \nu_{t-1}^2. \tag{C.38}$$

注意,

$$\mathbb{E}\big[\overline{g}_t(\mathsf{Y})^2\big] > 0 \tag{C.39}$$

否则，根据假设 (A12)、柯西–施瓦茨不等式以及 $\mathsf{Z} \sim \mathcal{N}(0, \overline{1})$,

$$0 < \mathbb{E}\Big[\overline{g}_t(\mathsf{Y}) \overline{\mathsf{Z}}^2\Big] \le \sqrt{\mathbb{E}[\overline{g}_t(\mathsf{Y})^2]} \cdot \sqrt{3} = 0,$$

这就产生了矛盾。我们声称， ν_{t-1}^2 在(C.38) 右侧前面的乘法系数也为正。事实上，根据柯西–施瓦茨不等式,

$$\mathbb{E}\Big[\overline{g}_t(\mathsf{Y}) \overline{\mathsf{Z}}^2\Big]^2 \le \mathbb{E}\Big[\overline{g}_t(\mathsf{Y})^2 \overline{\mathsf{Z}}^2\Big] \mathbb{E}\Big[\overline{\mathsf{Z}}^2\Big] = \mathbb{E}\Big[\overline{g}_t(\mathsf{Y})^2 \overline{\mathsf{Z}}^2\Big], \tag{C.40}$$

这可用于对感兴趣的系数给出如下下界

$$\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2} + \delta \overline{\kappa}_2\right) \mathbb{E}\Big[\overline{g}_t(\mathsf{Y})^2 \overline{\mathsf{Z}}^2\Big] + \left(\frac{\overline{\kappa}_6}{\overline{\kappa}_2^2} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3} + \delta \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\right) \mathbb{E}\Big[\overline{g}_t(\mathsf{Y}) \overline{\mathsf{Z}}^2\Big]^2 \ge \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2} + \delta \overline{\kappa}_2 + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3} + \delta \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\right) \mathbb{E}\Big[\overline{g}_t(\mathsf{Y}) \overline{\mathsf{Z}}^2\Big]^2. \tag{C.41}$$

右边的第二个因子根据假设 (A12) 为正，且第一个因子也为正，正如我们已在(C.37)中所见。将上述结果与(4.5)相结合，并回顾$(\nu_t^2)_{t\geq 1}$ （见下方(C.31) ）的正性，并对 t 在(C.38)中进行归纳，我们得出 τ_t 对每一个 $t \geq 1$的正性。

接下来， ω_t我们考虑递归关系(C.30)。可以类似地看出， ν_t^2 在(C.30) 右侧前面的系数为正：

$$\mathbb{E}\Big[\overline{g}_{t+1}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\Big] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\Big[\overline{g}_{t+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2 \geq \mathbb{E}\Big[\overline{g}_{t+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2\left(1+\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\right) > 0, \tag{C.42}$$

这里我们用到了(C.35) 和(C.40)。结合(C.39) 以及刚刚证明的$(\tau_t^2)_{t\geq 1}$、$(\nu_t^2)_{t\geq 1}$ 的正性，并对 t进行归纳，我们得出 ω_t 对每个 $t \geq 1$都为正。

考虑(C.31) 和(C.38) 关于 ν_t、 τ_t的递推关系。将(C.38) 两边除以 ν_t^2 ，并使用(C.31)，可得到比值 τ_t^2/ν_t^2的递归关系：

$$\begin{aligned}
\frac{\tau_t^2}{\nu_t^2} &= \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}+\delta\right)\mathbb{E}\big[\overline{g}_t(\mathsf{Y})^2\big]\frac{\tau_{t-1}^2}{\nu_{t-1}^2}\frac{\nu_{t-1}^2}{\nu_t^2} + \rho\left\{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2}+\delta\overline{\kappa}_2\right)\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})^2\overline{\mathsf{Z}}^2\Big] + \left(\frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}-\frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3}+\delta\frac{\overline{\kappa}_4}{\overline{\kappa}_2}\right)\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2\right\}\frac{\nu_{t-1}^2}{\nu_t^2} \\
&= \frac{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}+\delta\right)\mathbb{E}\big[\overline{g}_t(\mathsf{Y})^2\big]}{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}+\delta\right)^2\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2}\frac{\tau_{t-1}^2}{\nu_{t-1}^2} + \frac{\rho\left\{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2}+\delta\overline{\kappa}_2\right)\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})^2\overline{\mathsf{Z}}^2\Big] + \left(\frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}-\frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3}+\delta\frac{\overline{\kappa}_4}{\overline{\kappa}_2}\right)\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2\right\}}{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}+\delta\right)^2\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2},
\end{aligned}$$

与(C.7)一致，正如所需。同样，将(C.30) 两边除以 χ_{t+1}^2 ，并使用(C.28)，可得(C.8)。最后，我们还将(C.3e) 特化为(C.3g) 在满足假设 (A12)的情形。再次利用(C.24)，对任意 $1 \leq i \leq t$， $0 \leq j \leq t$，

$$\begin{aligned}
\Sigma_{t+1,i+1} &= \overline{\kappa}_2(\chi_t\chi_i\rho+\Omega_{t,i}) + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\rho\Big(\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\nu_{t-1}\chi_i + \mathbb{E}\Big[\overline{g}_i(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\nu_{i-1}\chi_t\Big) \\
&\quad + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}[\overline{g}_t(\mathsf{Y})\overline{g}_i(\mathsf{Y})\mathsf{Z}_{t-1}\mathsf{Z}_{i-1}] + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}\rho\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\mathbb{E}\Big[\overline{g}_i(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\nu_{t-1}\nu_{i-1} \\
&= \overline{\kappa}_2(\chi_t\chi_i\rho+\Omega_{t,i}) + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\rho\Big(\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\nu_{t-1}\chi_i + \mathbb{E}\Big[\overline{g}_i(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\nu_{i-1}\chi_t\Big) \\
&\quad + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\Big(\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{g}_i(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\sigma^2\nu_{t-1}\nu_{i-1} + \mathbb{E}[\overline{g}_t(\mathsf{Y})\overline{g}_i(\mathsf{Y})]\mathbb{E}[\tau_{t-1}\mathsf{N}_{t-1}\cdot\tau_{i-1}\mathsf{N}_{i-1}]\Big) \\
&\quad + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}\rho\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\mathbb{E}\Big[\overline{g}_i(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\nu_{t-1}\nu_{i-1} \\
&= \overline{\kappa}_2(\chi_t\chi_i\rho+\Omega_{t,i}) + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\rho\Big(\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\nu_{t-1}\chi_i + \mathbb{E}\Big[\overline{g}_i(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\nu_{i-1}\chi_t\Big) \\
&\quad + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}[\overline{g}_t(\mathsf{Y})\overline{g}_i(\mathsf{Y})]\mathbb{E}[\tau_{t-1}\mathsf{N}_{t-1}\cdot\tau_{i-1}\mathsf{N}_{i-1}] \\
&\quad + \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2}\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{g}_i(\mathsf{Y})\overline{\mathsf{Z}}^2\Big] + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}\mathbb{E}\Big[\overline{g}_t(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\mathbb{E}\Big[\overline{g}_i(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\right)\rho\nu_{t-1}\nu_{i-1}, \\
\Omega_{t+1,j+1} &= \delta\left(\frac{1}{\overline{\kappa}_2}\mathbb{E}[\overline{g}_{t+1}(\mathsf{Y})\overline{g}_{j+1}(\mathsf{Y})\mathsf{Z}_t\mathsf{Z}_j] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\rho\mathbb{E}\Big[\overline{g}_{t+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\mathbb{E}\Big[\overline{g}_{j+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\nu_t\nu_j\right) \\
&= \delta\frac{1}{\overline{\kappa}_2}\Big(\mathbb{E}\Big[\overline{g}_{t+1}(\mathsf{Y})\overline{g}_{j+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\sigma^2\nu_t\nu_j + \mathbb{E}[\overline{g}_{t+1}(\mathsf{Y})\overline{g}_{j+1}(\mathsf{Y})]\mathbb{E}[\tau_t\mathsf{N}_t\cdot\tau_j\mathsf{N}_j]\Big) \\
&\quad + \delta\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\rho\mathbb{E}\Big[\overline{g}_{t+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\mathbb{E}\Big[\overline{g}_{j+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\nu_t\nu_j \\
&= \delta\left\{\rho\left(\mathbb{E}\Big[\overline{g}_{t+1}(\mathsf{Y})\overline{g}_{j+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\Big[\overline{g}_{t+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\mathbb{E}\Big[\overline{g}_{j+1}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]\right)\nu_t\nu_j + \frac{1}{\overline{\kappa}_2}\mathbb{E}[\overline{g}_{t+1}(\mathsf{Y})\overline{g}_{j+1}(\mathsf{Y})]\mathbb{E}[\tau_t\mathsf{N}_t\cdot\tau_j\mathsf{N}_j]\right\},
\end{aligned}$$

如(C.9) 和(C.10)中所述。

C.4 定理TheoremC.4的证明

对(4.22)中的 τ^2/ν^2 不动点方程，通过从(C.7)中移除时间指标并求解 τ^2/ν^2即可得到。对(4.23) 中的 ω^2/χ^2 不动点方程，将(4.22) 中 τ^2/ν^2 的表达式代入(C.8)（移除时间指标后），并进行初等代数运算即可得到。

接下来，我们证明 τ^2/ν^2和 ω^2/χ^2均为正。首先考虑 τ^2/ν^2。根据假设(4.22)(4.13)，分子为正已在(C.41)中证明。因此，$\tau^2/\nu^2 > 0$。结合假设 (A12)、(C.39) 和(C.42) 在(C.8)中，可得 $\omega^2/\chi^2 > 0$。_

C.5 定理TheoremB.1的证明

收敛结果 (B.4) 已由定理 C.1保证。接下来需要验证 (B.5)中比值的平稳性。为此，首先注意到，根据 ν_0的选择，τ_0 在(B.3)中，

$$\frac{\tau_0^2}{\nu_0^2} = w_1 = \frac{\overline{\tau}^2}{\overline{\nu}^2}, \tag{C.43}$$

其中，最后一个等式由定理 C.4得出。

根据(B.2) 对 g_1 的选择以及 Stein 引理，

$$\begin{aligned}
\mathbb{E}[\partial_z g_1(\mathsf{Z},\mathsf{E})] &= \sigma^{-2}\mathbb{E}[\mathsf{Z}\cdot\overline{g}(\mathsf{Y})(\nu_0\mathsf{Z}+\tau_0\mathsf{N}_0)] = \nu_0\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right],\\
\mathbb{E}\left[g_1(\mathsf{Z},\mathsf{E})^2\right] &= \mathbb{E}\left[\overline{g}(\mathsf{Y})^2(\nu_0\mathsf{Z}+\tau_0\mathsf{N}_0)^2\right] = \sigma^2\nu_0^2\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\right] + \tau_0^2\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\right],
\end{aligned}$$

其中 $\mathsf{N}_0 \sim \mathcal{N}(0,\ 1)$ 与 Z、E 相互独立。利用这些公式，我们根据 χ_1，ω_1 按照 (C.3d)：

$$\begin{aligned}
\chi_1 &= \delta\nu_0\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right],\\
\omega_1^2 &= \delta\frac{1}{\overline{\kappa}_2}\left(\sigma^2\nu_0^2\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\right] + \tau_0^2\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\right]\right) + \delta\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\rho\nu_0^2\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2\\
&= \delta\rho\nu_0^2\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\right] + \delta\frac{\tau_0^2}{\overline{\kappa}_2}\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\right] + \delta\rho\nu_0^2\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2.
\end{aligned}$$

由于 ν_0 为正，因此 χ_1 也为正。于是

$$\begin{aligned}
\frac{\omega_1^2}{\chi_1^2} &= \frac{\rho}{\delta}\frac{\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\right] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2}{\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2} + \frac{\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\right]}{\delta\overline{\kappa}_2\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2}\frac{\tau_0^2}{\nu_0^2}\\
&\overset{(C.43)}{=} \frac{\rho}{\delta}\frac{\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\right] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2}{\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2} + \frac{\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\right]}{\delta\overline{\kappa}_2\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2}\frac{\overline{\tau}^2}{\overline{\nu}^2} \overset{(C.8)}{=} \frac{\overline{\omega}^2}{\overline{\chi}^2} = w_2,
\end{aligned}$$

其中，最后一个等式由 定理 C.4 得出。由此，通过对 t 进行归纳，可得 (B.5) 中的前两个结论。(B.5) 中的其余结论可直接由 (C.25)、(C.28) 和 (C.31) 得出。

D 定理 B.2 的证明

回忆 $z = X\beta_* = O\Lambda Q^\top\beta_*$，其中第二个等式由 X 的奇异值分解得出（见 假设（A2））。令$w := \Lambda Q^\top\beta_*$，使得 $z = Ow$。定义

$$\check{D} := Q\Lambda^\top\Pi_{w^\perp}\widetilde{O}^\top\Pi_{z^\perp}^\top T\Pi_{z^\perp}\widetilde{O}\Pi_{w^\perp}^\top\Lambda Q^\top, \tag{D.1}$$

~其中 $O \sim \mathrm{Haar}(\mathbb{O}(n-1))$ 与其他所有变量独立。

引理 D.1。 以下结论几乎必然成立：

$$\lambda_3(\breve{D}) \le \lambda_2(D) \le \lambda_1(\breve{D}). \tag{D.2}$$

证明。 我们将矩阵 D 写作

$$D = X^\top T X = Q\Lambda^\top O^\top T O \Lambda Q^\top.$$

只需研究 $\Lambda^\top O^\top T O \Lambda$ 的特征值，因为正交矩阵的共轭变换会保持其谱。根据定理 K.6，我们有

$$O \,|\, \{Ow = z\} \overset{d}{=} \frac{zw^\top}{\|z\|_2^2} + \Pi_{z^\perp} \widetilde{O} \Pi_{w^\perp}^\top,$$

其中 $\widetilde{O} \sim \mathrm{Haar}(\mathbb{O}(n-1))$ 与其他所有变量相互独立。于是

$$\Lambda^\top O^\top T O \Lambda \,|\, \{Ow = z\} \overset{d}{=} \Lambda^\top \left(\frac{wz^\top}{\|z\|_2^2} + \Pi_{w^\perp} \widetilde{O}^\top \Pi_{z^\perp}^\top \right) T \left(\frac{zw^\top}{\|z\|_2^2} + \Pi_{z^\perp} \widetilde{O} \Pi_{w^\perp}^\top \right) \Lambda.$$

根据 Courant–Fischer 定理，

$$\begin{aligned}
\lambda_2\big(\Lambda^\top O^\top T O \Lambda\big) \,|\, \{Ow = z\} &\overset{d}{=} \min_{\substack{\mathcal{V} \subset \mathbb{R}^d \\ \dim(\mathcal{V}) = d-1}} \max_{v \in \mathcal{V} \cap \mathbb{S}^{d-1}} v^\top \left(\frac{\Lambda^\top w z^\top}{\|z\|_2^2} + \Lambda^\top \Pi_{w^\perp} \widetilde{O}^\top \Pi_{z^\perp}^\top \right) T \left(\frac{z w^\top \Lambda}{\|z\|_2^2} + \Pi_{z^\perp} \widetilde{O} \Pi_{w^\perp}^\top \Lambda \right) v \\
&\le \max_{\substack{v \in \mathbb{S}^{d-1} \\ \left\langle v, \frac{\Lambda^\top w}{\|z\|_2} \right\rangle = 0}} v^\top \left(\frac{\Lambda^\top w z^\top}{\|z\|_2^2} + \Lambda^\top \Pi_{w^\perp} \widetilde{O}^\top \Pi_{z^\perp}^\top \right) T \left(\frac{z w^\top \Lambda}{\|z\|_2^2} + \Pi_{z^\perp} \widetilde{O} \Pi_{w^\perp}^\top \Lambda \right) v \\
&\le \max_{v \in \mathbb{S}^{d-1}} v^\top \Lambda^\top \Pi_{w^\perp} \widetilde{O}^\top \Pi_{z^\perp}^\top T \Pi_{z^\perp} \widetilde{O} \Pi_{w^\perp}^\top \Lambda v \\
&= \lambda_1\Big(\Lambda^\top \Pi_{w^\perp} \widetilde{O}^\top \Pi_{z^\perp}^\top T \Pi_{z^\perp} \widetilde{O} \Pi_{w^\perp}^\top \Lambda \Big).
\end{aligned}$$

这就证明了(D.2)中的最左侧不等式。

再次由 Courant–Fischer 定理，

$$\begin{aligned}
\lambda_2\big(\Lambda^\top O^\top T O \Lambda\big) \,|\, \{Ow = z\} &\overset{d}{=} \min_{\substack{\mathcal{V} \subset \mathbb{R}^d \\ \dim(\mathcal{V}) = d-1}} \max_{v \in \mathcal{V} \cap \mathbb{S}^{d-1}} v^\top \left(\frac{\Lambda^\top w z^\top}{\|z\|_2^2} + \Lambda^\top \Pi_{w^\perp} \widetilde{O}^\top \Pi_{z^\perp}^\top \right) T \left(\frac{z w^\top \Lambda}{\|z\|_2^2} + \Pi_{z^\perp} \widetilde{O} \Pi_{w^\perp}^\top \Lambda \right) v \\
&= \max_{\substack{v \in \mathbb{S}^{d-1} \\ \langle v, v_* \rangle = 0}} v^\top \left(\frac{\Lambda^\top w z^\top}{\|z\|_2^2} + \Lambda^\top \Pi_{w^\perp} \widetilde{O}^\top \Pi_{z^\perp}^\top \right) T \left(\frac{z w^\top \Lambda}{\|z\|_2^2} + \Pi_{z^\perp} \widetilde{O} \Pi_{w^\perp}^\top \Lambda \right) v \\
&\ge \max_{\substack{v \in \mathbb{S}^{d-1} \\ \langle v, v_* \rangle = 0 \\ \left\langle v, \frac{\Lambda^\top w}{\|z\|_2} \right\rangle = 0}} v^\top \left(\frac{\Lambda^\top w z^\top}{\|z\|_2^2} + \Lambda^\top \Pi_{w^\perp} \widetilde{O}^\top \Pi_{z^\perp}^\top \right) T \left(\frac{z w^\top \Lambda}{\|z\|_2^2} + \Pi_{z^\perp} \widetilde{O} \Pi_{w^\perp}^\top \Lambda \right) v \\
&= \max_{\substack{v \in \mathbb{S}^{d-1} \\ \langle v, v_* \rangle = 0 \\ \left\langle v, \frac{\Lambda^\top w}{\|z\|_2} \right\rangle = 0}} v^\top \Lambda^\top \Pi_{w^\perp} \widetilde{O}^\top \Pi_{z^\perp}^\top T \Pi_{z^\perp} \widetilde{O} \Pi_{w^\perp}^\top \Lambda v \\
&\ge \min_{\substack{\mathcal{U} \subset \mathbb{R}^d \\ \dim(\mathcal{U}) = d-2}} \max_{v \in \mathcal{U} \cap \mathbb{S}^{d-1}} v^\top \Lambda^\top \Pi_{w^\perp} \widetilde{O}^\top \Pi_{z^\perp}^\top T \Pi_{z^\perp} \widetilde{O} \Pi_{w^\perp}^\top \Lambda v \\
&= \lambda_3\Big(\Lambda^\top \Pi_{w^\perp} \widetilde{O}^\top \Pi_{z^\perp}^\top T \Pi_{z^\perp} \widetilde{O} \Pi_{w^\perp}^\top \Lambda \Big),
\end{aligned}$$

我们将维数为 $\mathcal{V}_*$ 的最小化子空间 $d-1$ 与其法向量 $v_* \in \mathbb{S}^{d-1}$对应，即$\mathcal{V}_* = \{v \in \mathbb{R}^d : \ \langle v, \ v_* \rangle = 0\}$。这证明了 (D.2) 中最右侧的不等式。 □

接下来研究矩阵 D 在(D.1)中的极端特征值。鉴于从随机矩阵理论的角度来看，这一问题可能颇有意义，我们在D.1 节至D.4 节中给出稍更一般版本的 D 自(D.3)出发的完备结果。

D.1 设置

设 $\Sigma \in \mathbb{R}^{(d\wedge n)\times(d\wedge n)}$ 和 $T \in \mathbb{R}^{n\times n}$ 为对称矩阵，且 Σ 严格正定。考虑如下对称随机矩阵：

$$\widetilde{D} \coloneqq \Lambda^\top O^\top T O \Lambda \in \mathbb{R}^{d\times d},$$

其中 $O \sim \mathrm{Haar}(\mathbb{O}(n))$，而 $\Lambda \in \mathbb{R}^{d\times n}$ 定义为

$$\Lambda \coloneqq \begin{cases} \begin{bmatrix} \Sigma^{1/2} \\ 0_{(n-d)\times d} \end{bmatrix} & \text{if } n \geq d, \\ \begin{bmatrix} \Sigma^{1/2} & 0_{n\times(d-n)} \end{bmatrix} & \text{if } n < d. \end{cases}$$

我们对 Σ 和 T作如下假设。

(A15) $n/d \to \delta \in (0, \ \infty)$ 当 $n \to \infty$ 时。

(A16) $\|\Sigma\|_2$、 $\left\|\Sigma^{-1}\right\|_2$ 和 $\|T\|_2$ 在 n 上一致有界。

(A17) 当 $n \to \infty$ 时， T 和 Σ 的经验谱分布 μ_T 和 μ_Σ 分别收敛于 μ_T 和 μ_Σ，且 μ_T， $\mu_\Sigma \neq \delta_0$。

(A18) 对每个固定的 $\varepsilon > 0$，我们有

$$\operatorname{spec}\Sigma \subset \operatorname{supp}\mu_\Sigma + (-\varepsilon, \varepsilon), \qquad \operatorname{spec}T \subset \operatorname{supp}\mu_\mathsf{T} + (-\varepsilon, \varepsilon)$$

当 n 足够大时，其中 spec(·) 表示特征值的集合。

(A19) 测度 μ_Σ 和 μ_T 在边缘附近不会衰减得过快，具体来说 :

$$\lim_{z \nearrow \inf\operatorname{supp}\mu_\Sigma} \int_{\mathbb{R}} \frac{x}{x-z} \mathrm{d}\mu_\Sigma(x) = \infty = \lim_{z \searrow \sup\operatorname{supp}\mu_\Sigma} \int_{\mathbb{R}} \frac{x}{z-x} \mathrm{d}\mu_\Sigma(x),$$
$$\lim_{z \searrow \sup\operatorname{supp}\mu_\mathsf{T}} \int_{\mathbb{R}} \frac{x}{z-x} \mu_\mathsf{T}(x) = \infty.$$

(A20) 上确界 $\operatorname{supp}\mu_\mathsf{T} > 0$。

我们对上述假设作如下说明。假设（A15）是对假设（A4）的重新表述。注意，在第 3.2 节的设定中，$\Sigma^{1/2}$ 对应于 diag(λ)，其中 λ 在假设（A2）中定义。因此，可以验证，其他所有假设均已包含在假设 (A2) 和 (A5) 中。特别地，假设（A16）中 $\|T\|_2$的一致有界性成立，因为由假设 (A5) 给出的 γ 和 $g(y) + \gamma$ 均具有正性，所以 $\mathcal{T}(y) = 1 - {}_g\, y^{\gamma}(\,) + \gamma$有上界。此外，由于 (4.3)， g(Y) 不会恒等于 0，这意味着假设（A16）中的 $\mu_\mathsf{T} \neq \delta_0$ 成立。最后，由于 g(Y) 大于−γ、均值为 0，且不可能恒等于 0，因此随机变量 T = ^{g}(Y) $_g$(Y)+γ 在 $\mathbb{R}_{>0}$上赋予正测度，从而假设（A20）成立。

D.2 体边缘的收敛

考虑 D 的以下变体:

$$\widehat{D} = \widetilde{\Lambda}^\top \widetilde{O}^\top \widetilde{T} \widetilde{O} \widetilde{\Lambda} \in \mathbb{R}^{d\times d}, \tag{D.3}$$

其中, O 服从 $\mathbb{O}(n-1)$ 上的 Haar分布, 且与 T 和 Λ 独立; T是 T 的一个随机子矩阵, 而 Λ满足

$$\widetilde{\Lambda} = \begin{cases} \begin{bmatrix} \widetilde{\Sigma}^{1/2} & 0_{d-1} \\ 0_{(n-d)\times(d-1)} & 0 \end{bmatrix} & \text{if } n \geq d, \\ \begin{bmatrix} \widetilde{\Sigma}^{1/2} & 0_{(n-1)\times(d-n+1)} \end{bmatrix} & \text{if } n < d, \end{cases}$$

对于大小为 $n \wedge d - 1$ 的 Σ 的随机子矩阵 Σ。

在继续之前, 我们验证 D 在(D.1) 是 D 在 (D.3)中的特殊情形。因此, 所有关于 D 的结果都适用于 D。事实上, D 的谱与以下矩阵的谱相同

$$\Lambda^\top \Pi_{w^\perp} \widetilde{O}^\top \Pi_{z^\perp}^\top T \Pi_{z^\perp} \widetilde{O} \Pi_{w^\perp}^\top \Lambda. \tag{D.4}$$

令

$$\widetilde{T} = \Pi_{z^\perp}^\top T \Pi_{z^\perp} \in \mathbb{R}^{(n-1)\times(n-1)}, \qquad \widetilde{\Lambda} = \Pi_{w^\perp}^\top \Lambda \in \mathbb{R}^{(n-1)\times d},$$

我们看到, 矩阵 (D.4) 是 D的特殊情形。

对于矩阵 $\widehat{D}$, 我们的目标是证明如下结论。

引理 D.2。 几乎必然地, 当 $n \to \infty$ 时, 有

$$\lambda_1(\widehat{D}), \lambda_3(\widehat{D}) \to \sup \operatorname{supp}(\mu_{\Sigma_n} \boxtimes \mu_{\mathsf{T}}),$$

其中

$$\mu_{\Sigma_n} = \begin{cases} \frac{1}{\delta}\mu_\Sigma + \frac{1-\delta}{\delta}\delta_0 & \text{if } \delta > 1, \\ \mu_\Sigma & \text{otherwise.} \end{cases} \tag{D.5}$$

注意, Σ_n 在(D.5) 与 Λ_n^2 在(3.8)相同。我们在本节末证明定理 D.2。

D.2.1 Haar 正交矩阵与确定性矩阵的强渐近自由性

证明定理 D.2 的第一个要素是 Haar 正交矩阵与确定性矩阵的强渐近自由性, 这一结果本身也很有意义。

命题 D.3。 设 $U_1, \ldots, U_p \in \mathbb{R}^{N\times N}$是一组在 $\mathbb{SO}(N)$ 上 *Haar* 分布的独立随机矩阵, 且 $Y_1, \ldots, Y_q \in \mathbb{R}^{N\times N}$为确定性矩阵。假设 $\max_{j=1,\ldots,q}\|Y_j\|_2 \leq C$对于某个常数 $C > 0$成立。那么, 对于任意关于 $(p+q)$ 变量的自伴 $*$-多项式 Q 以及任意常数 $\varepsilon > 0$, 当 N 充分大时, 我们有

$$\operatorname{spec}(Q(U_1, \ldots, U_p, Y_1, \ldots, Y_p)) \subset \operatorname{spec}(Q(u_1, \ldots, u_p, Y_1, \ldots, Y_p)) + (-\varepsilon, \varepsilon),$$

其中, 右侧的二元运算为闵可夫斯基和, 而$(u_1, \ldots, u_p)$ 是一组 $*$-自由的 *Haar*酉元[1], 它们与$(Y_1, \ldots, Y_q)$是 $*$-自由的。

[1]Haar 酉元是迹 von Neumann 代数 $(\mathcal{M}, \langle\cdot\rangle)$ 中的酉算子 u, 其谱分布在 $\mathbb{S}^1$ 上均匀分布。

注意，定理 D.3 直接适用于我们的模型 D，因为我们可以取 O 在 $\mathbb{SO}(n-1)$上服从 Haar 分布，而不改变 $\mathbb{O}(n-1)$的概率分布 D。

总体而言，我们的证明遵循 [32, 节 3]；我们的目标是将 U_i表示为某个已知具有强渐近自由性的其他矩阵的函数演算。我们以高斯正交系综的强渐近自由性开始证明[2] 以及确定性矩阵。

引理 D.4 ([46, 定理 4.2])。 定理 D.3 成立，只要将$(U_1, \ldots, U_p)$ 替换为相互独立的高斯正交系综$(W_1, \ldots, W_p)$，并将$(u_1, \ldots, u_p)$ 替换为自由半圆元素$(s_1, \ldots, s_p)$。

在下一部分中，我们将 Haar 正交矩阵表示为标准形 $U\Theta U^\top$ 及一个 Haar 正交 U，从而可以将其表示为某个正交不变矩阵的函数演算。在下文中，我们记

$$R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \in \mathbb{R}^{2\times 2}, \qquad \theta \in \mathbb{R}.$$

引理 D. 5 ([69, 定理 3.5])。 设 $\Theta \in \mathbb{R}^{N\times N}$ 是一个服从分布的随机分块对角矩阵 as

$$\Theta = \begin{cases} \operatorname{diag}(R(\theta_1), \ldots, R(\theta_m)) & N = 2m, \\ \operatorname{diag}(R(\theta_1), \ldots, R(\theta_m), 1) & N = 2m+1, \end{cases}$$

其中 $0 \le \theta_1 \le \cdots \le \theta_m \le \pi$ 是从联合密度中采样得到的（需对 Z_N 和 Z'_N 作适当选择）

$$\frac{1}{Z_N} \prod_{1\le j<k\le m} (\cos\theta_j - \cos\theta_k)^2, \qquad if\ N = 2m,$$

$$\frac{1}{Z'_N} \prod_{1\le j\le m} \sin^2\left(\frac{\theta_j}{2}\right) \prod_{1\le k<\ell\le m} (\cos\theta_j - \cos\theta_k)^2, \qquad if\ N = 2m+1.$$

设 V 是一个与 Θ 独立的 *Haar* 正交矩阵。那么，矩阵 $V\Theta V^\top$ 在 $\mathbb{SO}(N)$ 上服从 *Haar* 分布。

如我们所见，与通常的对角化不同，我们发现矩阵 Θ 是一个由（2×2）个块组成的分块对角矩阵。实际上，Haar 正交矩阵 U 的真正特征向量应当是具有复共轭对称性的复向量。这使得难以将 U 表示为高斯正交系综的函数演算，因为高斯正交系综的特征向量是实数。

相反，我们将把 U 表示为一个反对称矩阵的函数演算；反对称矩阵已经具有相同的标准形，由（2×2）个块组成，唯一的区别在于，这些块具有 xJ 的形式，其中 $x \in \mathbb{R}$，而非 $R(\theta)$，其中

$$J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

注意，反对称矩阵的谱纯为虚数，J 的特征值为 –i，i。更根本地说，使用反对称矩阵是很自然的，因为与群 $\mathbb{SO}(2m)$ 关联的李代数 $\mathfrak{so}(2m)$恰好就是反对称矩阵空间。在复数情形下，与 $\mathbb{U}(n)$ 关联的李代数$\mathfrak{u}(n)$是反厄米矩阵空间，而反厄米矩阵不过是乘以因子 i 后旋转得到的厄米矩阵。

因此，我们使用最简单的正交不变反对称矩阵，它可以写作$X - X^\top$其中 $X \in \mathbb{R}^{N\times N}$具有独立的 $\mathcal{N}(0, 1/N)$ 个元素，即 Ginibre 系综。为此，接下来我们证明，在 W_i定理 D.4 中的 Theorem D.4 可以取为实 Ginibre 系综:

引理 D.6。 定理 D.3 若将 $(U_1, \ldots, U_p)$ 替换为独立实 *Ginibre* 系综$(X_1, \ldots, X_p) \in \mathbb{R}^{N\times N}$，并将 $(u_1, \ldots, u_p)$ 替换为 $*$- 自由圆元素 $(x_1, \ldots, x_p)$，则仍成立。

[2]高斯正交系综 GOE(d) 是定义在 $d\times d$ 实对称矩阵 G 上的概率测度，满足$G_{i,\,i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\ 2/d)$，并且 $G_{i,\,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\ 1/d)$对所有 $1 \le i < j \le d$ 均成立。

证明。 在下文中，我们使用

$$A \otimes B = \begin{bmatrix} B_{1,1}A & \cdots & B_{1,q}A \\ \vdots & \ddots & \vdots \\ B_{p,1}A & \cdots & B_{p,q}A \end{bmatrix}$$

用来表示 $m \times n$ 矩阵 A 与 $p \times q$ 矩阵 B的克罗内克积。令 $W_1, \ldots, W_p \in \mathbb{R}^{2N \times 2N}$ 为相互独立的高斯正交系综，并取投影矩阵

$$P_1 = \begin{bmatrix} I_N & 0_{N \times N} \end{bmatrix}, \qquad P_2 = \begin{bmatrix} 0_{N \times N} & I_N \end{bmatrix}, \qquad \Pi_{k\ell} = P_k^\top P_\ell = I_N \otimes E_{k\ell} \in \mathbb{R}^{2N \times 2N},$$

其中 $E_{k\ell} = e_k e_\ell^\top \in \mathbb{R}^{2\times 2}$ 当 $k, \ell \in \{1, 2\}$时。于是我们发现，这个集合

$$X_i = \sqrt{2} P_1 W_i P_2^\top$$

由相互独立的 Ginibre 系综组成。现在我们发现

$$\begin{aligned} Q(X_1, \ldots, X_p, Y_1, \ldots, Y_q) \otimes I_2 &= Q(X_1 \otimes I_2, \ldots, X_p \otimes I_2, Y_1 \otimes I_2, \ldots, Y_q \otimes I_2) \\ &= \widetilde{Q}(W_1, \ldots, W_p, \Pi_{11}, \Pi_{12}, \Pi_{21}, \Pi_{22}, Y_1 \otimes I_2, \ldots, Y_q \otimes I_2) \end{aligned}$$

$\widetilde{}$对于另一个 $*$-多项式 Q，它由 Q 通过以下方式得到

$$X_i \otimes I_2 = \sqrt{2} \sum_{k=1}^{2} P_k^\top P_1 W_i P_2^\top P_k = \sqrt{2} \sum_{k=1}^{2} \Pi_{k1} W_i \Pi_{2k}.$$

现在应用 定理 D.4 可得，在极限 $N \to \infty$下，

$$\begin{aligned} \operatorname{spec}(Q(X_1, \ldots, X_p, Y_1, \ldots, Y_q)) &= \operatorname{spec}(Q(X_1, \ldots, X_p, Y_1, \ldots, Y_q) \otimes I_2) \\ &\subset \operatorname{spec}(\widetilde{Q}(s_1, \ldots, s_p, \Pi_{11}, \Pi_{12}, \Pi_{21}, \Pi_{22}, Y_1 \otimes I_2, \ldots, Y_q \otimes I_2)) + (-\varepsilon, \varepsilon) \\ &= \operatorname{spec}(Q(\widehat{x}_1, \ldots, \widehat{x}_p, Y_1 \otimes I_2, \ldots, Y_q \otimes I_2) + (-\varepsilon, \varepsilon), \end{aligned} \tag{D.6}$$

其中 $(s_i)_{1 \le i \le p}$ 是自由半圆元素，且与 $\{Y_j \otimes I_2, \Pi_{k\ell}\}$ 相对于 $*$ 自由；二者均在自由积冯·诺伊曼代数 $\mathcal{M} * \mathbb{C}^{2N \times 2N}$ 中实现。最后一行中，我们将

$$\widehat{x}_i := \sqrt{2} \sum_{k=1}^{2} \Pi_{k1} s_i \Pi_{2k}.$$

接下来我们证明 $*$-分布恒等式

$$(\widehat{x}_1, \ldots, \widehat{x}_p, Y_1 \otimes I_2, \ldots, Y_q \otimes I_2) \overset{d}{=} (x_1, \ldots, x_p, Y_1, \ldots, Y_q), \tag{D.7}$$

其中 x_1、……、x_p 是一组圆元素，它们彼此之间以及与 Y_1、……、Y_q之间均满足 $*$-自由性。我们通过所谓的元论证来实现这一点。根据高斯正交系综的标准渐近自由性，对于任意 $*$-多项式P，我们有

$$\langle P(\widehat{x}_1, \ldots, \widehat{x}_p, Y_1 \otimes I_2, \ldots, Y_q \otimes I_2) \rangle = \lim_{M \to \infty} \left\langle P\left(\widehat{X_1}, \ldots, \widehat{X_p}, I_M \otimes Y_1 \otimes I_2, \ldots, I_M \otimes Y_q \otimes I_2\right) \right\rangle,$$

其中 $\left\{\widehat{X_i} \in \mathbb{R}^{(2NM \times 2NM)} : \ 1 \le i \le p\right\}$ 由下式给出

$$\widehat{X_i} = \sqrt{2} \sum_{k=1}^{2} (I_M \otimes \Pi_{k1}) \widetilde{W_i} (I_M \otimes \Pi_{2k})$$

以及一个集合（ ~ 每个 X_i 都可以写成 $\sim W_i$）$_{1\le i\le p}$ 个大小为（$2NM\times 2NM$）的高斯正交系综。接下来注意到，as

$$\widehat{X}_i=\begin{bmatrix}\widetilde{X}_i & 0\\ 0 & \widetilde{X}_i\end{bmatrix}=\widetilde{X}_i\otimes I_2,$$

~其中（X_i）$_{1\le i\le p}$ 是一组大小为（$NM\times NM$）的独立实 Ginibre 系综。因此

$$\left\langle P\left(\widehat{X}_1,\dots,\widehat{X}_p,I_M\otimes Y_1\otimes I_2,\dots,I_M\otimes Y_q\otimes I_2\right)\right\rangle=\left\langle P\left(\widetilde{X}_1,\dots,\widetilde{X}_p,I_M\otimes Y_1,\dots,I_M\otimes Y_q\right)\right\rangle,$$

并令 $M\to\infty$ 取极限，得到

$$\langle P\left(\widehat{x}_1,\dots,\widehat{x}_p,Y_1\otimes I_2,\dots,Y_q\otimes I_2\right)\rangle=\langle P\left(x_1,\dots,x_p,Y_1,\dots,Y_q\right)\rangle.$$

由于 P 是一般的，这表明(D.7) 因此

$$\operatorname{spec}(Q(\widehat{x}_1,\dots,\widehat{x}_p,Y_1\otimes I_2,\dots,Y_q\otimes I_2))=\operatorname{spec}(Q(x_1,\dots,x_p,Y_1,\dots,Y_q)).$$

将其代入(D.6) 即可证明该引理。 □

证明定理 *em D.3.* 我们旨在近似耦合一个阶数为 N 的 Haar 正交矩阵与

$$F((X-X^\top)/\sqrt{2})$$

对于某个函数 F，通过函数演算应用该函数，其中 $X\in\mathbb{R}^{N\times N}$ 是一个实 Ginibre 系综。事实上，由于 $X-X^\top$ 在正交共轭下不变且为反对称矩阵，我们可以将其标准形写成定理 D.5 中的形式：

$$\frac{X-X^\top}{\sqrt{2}}=V\widetilde{\Theta}V^\top,\qquad \widetilde{\Theta}=\begin{cases}\operatorname{diag}(\lambda_1 J,\dots,\lambda_m J), & N=2m,\\ \operatorname{diag}(\lambda_1 J,\dots,\lambda_m J,0), & N=2m+1,\end{cases}\qquad J=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix},\tag{D.8}$$

~其中 V 是一个与 Θ 独立的 Haar 正交矩阵，并且

$$-\lambda_m\le\cdots\le-\lambda_1\le\lambda_1\le\cdots\le\lambda_m$$

是（$X-X^\top$）/（$\sqrt{2}$i）的有序特征值；当 N~为奇数时，还需计入一个额外的零特征值（回忆一下，奇数阶反对称矩阵是奇异的）。根据通常的函数演算，我们得到

$$F((X-X^\top)/\sqrt{2})=VF(\widetilde{\Theta})V^\top,$$

并取 F（·）$=$exp（f（·））为奇函数 f：i$\mathbb{R}\to$i$\mathbb{R}$，则得到

$$F(\widetilde{\Theta})=\begin{cases}\operatorname{diag}(F(\lambda_1 J),\dots,F(\lambda_m J)), & N=2m,\\ \operatorname{diag}(F(\lambda_1 J),\dots,F(\lambda_m J),\exp(f(0))), & N=2m+1,\end{cases}$$

根据谱定理，我们可以将每个块计算为

$$\begin{aligned}F(\lambda_i J)&=F\left(\widetilde{U}\begin{bmatrix}\mathrm{i}\lambda_i & 0\\ 0 & -\mathrm{i}\lambda_i\end{bmatrix}\widetilde{U}^*\right)=\exp\left(\widetilde{U}\begin{bmatrix}f(\mathrm{i}\lambda_i) & 0\\ 0 & -f(\mathrm{i}\lambda_i)\end{bmatrix}\widetilde{U}^*\right)\\&=\exp(-\mathrm{i}f(\mathrm{i}\lambda_i)J)=\begin{bmatrix}\cos(\mathrm{i}f(\mathrm{i}\lambda_i)) & \sin(\mathrm{i}f(\mathrm{i}\lambda_i))\\ -\sin(\mathrm{i}f(\mathrm{i}\lambda_i)) & \cos(\mathrm{i}f(\mathrm{i}\lambda_i))\end{bmatrix}=R(-\mathrm{i}f(\mathrm{i}\lambda_i)),\end{aligned}$$

使用适当的酉矩阵 $\widetilde{U}\in\mathbb{C}^{2\times 2}$ 对 J进行对角化。

~接下来，我们选择一个合适的奇函数 f，使得对于独立采样的 Θ 和 Θ（回想一下，$(\lambda_i)_{1\le i\le m}$ 和 $(\theta_i)_{1\le i\le m}$ 按递增顺序排列）

$$\max_{1\le i\le m}\|F(\lambda_i J)-R(\theta_i)\|\le\max_{1\le i\le m}|-\mathrm{i}f(\mathrm{i}\lambda_i)-\theta_i|=o(1),$$

从而得到 $F((X-X^\top)/\sqrt{2})$ 与 $\mathbb{SO}(N)$-Haar 矩阵 U 之间的耦合

$$\left\|F((X-X^\top)/\sqrt{2})-U\right\|=\left\|VF(\widetilde{\Theta})V^\top-V\Theta V^\top\right\|=\max_{1\le i\le m}\|F(\lambda_i J)-R(\theta_i)\|=o(1),$$

~其中 V、Θ 和 Θ 均独立抽样。

为了构造适当的函数 f，取 $F_{\mathrm{qc}}:\ \mathbb{R}\to[0,1]$ 为四分之一圆分布的累积分布函数的奇延拓，即

$$F_{\mathrm{qc}}(x)\coloneqq\int_0^x\mathbb{1}_{[-2,2]}(t)\frac{\sqrt{4-t^2}}{\pi}\mathrm{d}t.$$

现在，由于 $\mathrm{i}(X-X^\top)/\sqrt{2}$ 是一个复 Wigner 矩阵，λ_i 的经验分布收敛于 $[0,2]$，上的四分之一圆分布，这给出了在 $\mathbb{R}$ 上累积分布函数的一致收敛：

$$\sup_{x\in[0,2]}\left|F_{\mathrm{qc}}(x)-\frac{|\{i\in[m]:\lambda_i\le x\}|}{\lceil N/2\rceil}\right|\to0,\qquad\text{almost surely.}$$

特别地，限制到 $x=\lambda_i$ 可得

$$\max_{1\le i\le N/2}\left|F_{\mathrm{qc}}(\lambda_i)-\frac{i+\mathbb{1}(N\text{ is odd})}{\lceil N/2\rceil}\right|\to0,\qquad\text{almost surely.}$$

同样地，取 μ_θ 为 θ_i（在 $[0,\pi]$上）的经验累积分布函数，根据 定理 D.5，有 μ_θ 收敛于 $[0,\pi]$ 上的均匀分布（例如参见 [69, 定理 4。13]），因此

$$\max_{1\le i\le N/2}\left|\mu_\theta(\theta_i)-\frac{\theta_i}{\pi}\right|=\max_{1\le i\le N/2}\left|\frac{i+\mathbb{1}(N\text{ is odd})}{\lceil N/2\rceil}-\frac{\theta_i}{\pi}\right|\to0.$$

因此我们得到

$$\max_{1\le i\le N/2}|\pi F_{\mathrm{qc}}(\lambda_i)-\theta_i|\to0\qquad\text{almost surely.}$$

因此我们将奇函数 f 选取为

$$f(\mathrm{i}x)=\mathrm{i}\pi F_{\mathrm{qc}}(x),$$

从而保证，对于函数 $F=\mathrm{e}^f$，

$$\left\|VF(\widetilde{\Theta})V^\top-V\Theta V^\top\right\|\to0,\qquad\text{almost surely.}$$

利用 Stone–Weierstrass 定理的复版本，我们接下来取一列多项式 $(P_n)_{n\in\mathbb{N}}$ 使得

$$\lim_{n\to\infty}\sup_{x\in[-4,4]}|P_n(x)-\exp(\mathrm{i}\pi F_{\mathrm{qc}}(x))|=0.$$

于是我们有

$$\left\|Q(U_1,\dots,U_p,Y_1,\dots,Y_q)-Q(F((X_1-X_1^\top)/\sqrt{2}),\dots,F((X_p-X_p^\top)/\sqrt{2}),Y_1,\dots,Y_q)\right\|$$

$$= \left\| Q(V_1\Theta_1 V_1^\top, \dots, V_p\Theta_p V_p^\top, Y_1, \dots, Y_q) - Q(V_1 F(\widetilde{\Theta}_1) V_1^\top, \dots, V_p F(\widetilde{\Theta}_p) V_p^\top, Y_1, \dots, Y_q) \right\|$$
$$\lesssim \max_{1\le j\le p} \left\| \Theta_i - F(\widetilde{\Theta}_i) \right\| \to 0, \qquad \text{almost surely,} \tag{D.9}$$

其中隐含常数仅取决于 Q 以及上界 C 对于 $\|Y_j\|$ 的，$(V_i,\ \Theta_i,\ \Theta_i)_{1\le i\le p}$ 均相互独立，V_i 服从 Haar 分布于 $\mathbb{O}(N)$，Θ_i 的分布如 定理 D.5 所述，而 Θ_i 服从 (D.8) 所给出的分布。现在我们使用多项式 P_n 得到

$$\left\| Q(V_1 F(\widetilde{\Theta}_1) V_1^\top, \dots, V_p F(\widetilde{\Theta}_p) V_p^\top, Y_1, \dots, Y_q) - Q(V_1 P_n(-\mathrm{i}\widetilde{\Theta}_1) V_1^\top, \dots, V_p P_n(-\mathrm{i}\widetilde{\Theta}_p) V_p^\top, Y_1, \dots, Y_q) \right\|$$
$$\lesssim \sup_{\substack{x\in\mathbb{R} \\ |x|\le \max_{j=1,\dots,p} \|\widetilde{\Theta}_j\|}} |P_n(x) - F(\mathrm{i}x)|. \tag{D.10}$$

最后一个量在依次取极限 $\lim_{n\to\infty}\lim_{N\to\infty}$ 时收敛到 0，因为 $\left\|\widetilde{\Theta}_j\right\| \le 2\|X_j\|$ 几乎必然有界于 4，而 $N\to\infty$（因为 $X^\top X$ 是 Wishart系综），并且

$$\sup_{|x|\le 4} |P_n(x) - F(\mathrm{i}x)| = \sup_{|x|\le 4} |P_n(x) - \exp(\mathrm{i}\pi F_{\mathrm{qc}}(x))| \to 0 \qquad \text{as } n\to\infty.$$

因此我们得出结论，在极限 $\lim_{n\to\infty}\lim_{N\to\infty}$ 下，

$$\begin{aligned}
&\operatorname{spec}(Q(U_1, \dots, U_p, Y_1, \dots, Y_q)) \\
&\subset \operatorname{spec}(Q(V_1 P_n(-\mathrm{i}\widetilde{\Theta}_1) V_1^\top, \dots, V_p P_n(-\mathrm{i}\widetilde{\Theta}_p) V_p^\top, Y_1, \dots, Y_q)) + (-\varepsilon/3, \varepsilon/3) \\
&= \operatorname{spec}(Q(P_n(-\mathrm{i}(X_1 - X_1^\top)/\sqrt{2}), \dots, P_n(-\mathrm{i}(X_p - X_p^\top)/\sqrt{2}), Y_1, \dots, Y_q)) + (-\varepsilon/3, \varepsilon/3) \\
&\subset \operatorname{spec}(Q(P_n(-\mathrm{i}(x_1 - x_1^*)/\sqrt{2}), \dots, P_n(-\mathrm{i}(x_p - x_p^*)/\sqrt{2}), Y_1, \dots, Y_q)) + (-2\varepsilon/3, 2\varepsilon/3) \\
&\subset \operatorname{spec}(Q(F((x_1 - x_1^*)/\sqrt{2}), \dots, F((x_p - x_p^*)/\sqrt{2}), Y_1, \dots, Y_q)) + (-\varepsilon, \varepsilon),
\end{aligned}$$

其中第二行由 (D.9) 和 (D.10) 得出，第四行由 定理 D.6 得出，最后一行再次由 (D.10) 得出。利用 $(F((x_i - x_i^*)/\sqrt{2}))_{1\le i\le p}$ 是 $*$-自由的 Haar酉元，我们得出结论：它们的 $*$-自由性继承自 x_i 的自由性，而它们各自都是 Haar 酉元，因为 $\mathrm{i}(x_i - x_i^*)/\sqrt{2} = \sqrt{2}\mathrm{Re}[\mathrm{i}x_i]$ 是一个半圆元，其在 $x \mapsto F(-\mathrm{i}x)$ 下的推前分布是在 $\mathbb{S}^1$ 上的均匀谱分布。 □

D.2.2 对 δ 的不敏感性

以下结果是我们的第二个关键工具，用于处理当 $\delta = 1$ 时 Λ 与 Λ 之间的差异。

引理 D.7. 设 μ 和 ν 分别是支撑于 $(0,\infty)$ 和 $\mathbb{R}$ 上的紧支撑博雷尔概率测度。假设 $\operatorname{supsupp}\nu > 0$。那么，对于所有 $\varepsilon \in (0,1],$，我们有

$$\sup\operatorname{supp}\left(((1-\varepsilon)\mu + \varepsilon\delta_0) \boxtimes \nu\right) \le \sup\operatorname{supp}\left(\mu \boxtimes \nu\right).$$

证明。设 $\boldsymbol{s}$ 和 $\boldsymbol{t}$ 分别是在 W^*-概率空间 $\mathcal{M}_1$ 和 $\mathcal{M}_2$ 上 μ 和 ν 的实现。那么我们容易发现，测度 $(1-\varepsilon)\mu + \varepsilon\delta_0$ 是 (冯·诺依曼代数) 张量积 $\mathcal{M}_1 \otimes L^\infty([0,1])$ 中 $\boldsymbol{s} \otimes (1 - \boldsymbol{p}_\varepsilon)$ 的谱测度，其中 $\boldsymbol{p}_\varepsilon = \mathbb{1}_{[0,\varepsilon]} \in L^\infty([0,1])$ 是投影。现在我们有自由卷积

$$((1-\varepsilon)\mu + \varepsilon\delta_0) \boxtimes \nu$$

是冯·诺依曼代数自由积 $(\mathcal{M}_1 \otimes L^\infty([0,1])) * \mathcal{M}_2$ 中 $(\sqrt{\boldsymbol{s}} \otimes (1 - \boldsymbol{p}_\varepsilon))\boldsymbol{t}(\sqrt{\boldsymbol{s}} \otimes (1 - \boldsymbol{p}_\varepsilon))$ 的谱测度。

然后我们得到

$$\begin{aligned}
\sup\operatorname{supp}\left(((1-\varepsilon)\mu+\varepsilon\delta_0)\boxtimes\nu\right) &= \sup_{v\in\mathcal{H},\|v\|=1}\langle v,(\sqrt{\boldsymbol{s}}\otimes(1-\boldsymbol{p}_\varepsilon))\boldsymbol{t}(\sqrt{\boldsymbol{s}}\otimes(1-\boldsymbol{p}_\varepsilon))v\rangle \\
&= \sup_{v\in\mathcal{H},\|v\|=1}\langle(1\otimes(1-\boldsymbol{p}_\varepsilon))v,(\sqrt{\boldsymbol{s}}\otimes 1)\boldsymbol{t}(\sqrt{\boldsymbol{s}}\otimes 1)(1\otimes(1-\boldsymbol{p}_\varepsilon))v\rangle \\
&\le \sup_{w\in\mathrm{Range}(1\otimes(1-\boldsymbol{p}_\varepsilon)),\|w\|\le 1}\langle w,(\sqrt{\boldsymbol{s}}\otimes 1)\boldsymbol{t}(\sqrt{\boldsymbol{s}}\otimes 1)w\rangle \\
&= 0\vee\sup_{w\in\mathrm{Range}(1\otimes(1-\boldsymbol{p}_\varepsilon)),\|w\|=1}\langle w,(\sqrt{\boldsymbol{s}}\otimes 1)\boldsymbol{t}(\sqrt{\boldsymbol{s}}\otimes 1)w\rangle \\
&\le 0\vee\sup\operatorname{supp}(\mu\boxtimes\nu),
\end{aligned}$$

其中，$\mathcal{H}$ 是 $(\mathcal{M}_1\otimes L^\infty([0,1]))*\mathcal{M}_2$的盖尔范德–奈马克–西格尔（GNS）希尔伯特空间。最后，只需注意 $\sup\operatorname{supp}(\mu\boxtimes\nu)\ge 0$：否则，实现 $\sqrt{\boldsymbol{s}}\boldsymbol{t}\sqrt{\boldsymbol{s}}$ 将是负半定的，这与 $\sup\operatorname{supp}\,\nu>0$ 相矛盾，因为 $\boldsymbol{s}$ 可逆，这是因为 μ 支撑于 $(0,\infty)$。 □

D.2.3 定理D.2的证明

注意，根据柯西交错定理，$\widetilde{\Lambda}\widetilde{\Lambda}^\top$ 和 $\widetilde{T}$ 的经验特征值分布分别收敛于 μ_{Σ_n} 和 μ_{T}。因此，由通常的渐近自由性可立即推出，D 的经验谱$\widehat{\ }$分布收敛于 $\mu_{\Sigma_n}\boxtimes\mu_{\mathsf{T}}$，进而得到

$$\liminf_{n\to\infty}\lambda_3(\widehat{D})\ge\sup\operatorname{supp}(\mu_{\Sigma_n}\boxtimes\mu_{\mathsf{T}}),\qquad\text{almost surely},$$

因为左侧有下界，且该下界为 $\mu_{\Sigma_n}\boxtimes\mu_{\mathsf{T}}$ 的任意 $O(1)$ 分位数。

对于上界，我们首先断言 $\lambda_1(\widehat{D})=\lambda_1(\sqrt{\widetilde{\Lambda}\widetilde{\Lambda}^\top}\widetilde{O}^\top\widetilde{T}\widetilde{O}\sqrt{\widetilde{\Lambda}\widetilde{\Lambda}^\top})$。为此，注意到

$$\lambda_1(\widehat{D})=\begin{cases}\lambda_1(\sqrt{\widetilde{\Lambda}\widetilde{\Lambda}^\top}\widetilde{O}^\top\widetilde{T}\widetilde{O}\sqrt{\widetilde{\Lambda}\widetilde{\Lambda}^\top}) & \text{if } n\ge d,\\ 0\vee\lambda_1(\sqrt{\widetilde{\Lambda}\widetilde{\Lambda}^\top}\widetilde{O}^\top\widetilde{T}\widetilde{O}\sqrt{\widetilde{\Lambda}\widetilde{\Lambda}^\top})=0\vee\lambda_1(\widetilde{\Sigma}^{1/2}\widetilde{O}^\top\widetilde{T}\widetilde{O}\widetilde{\Sigma}^{1/2}) & \text{if } n<d,\end{cases}$$

并且对于 $n<d$ 矩阵 $\widetilde{\Sigma}^{1/2}\widetilde{O}^\top\widetilde{T}\widetilde{O}^\top\widetilde{\Sigma}$ 由于 假设（A20），因此不可能是负半定的。$\widetilde{\ }$由于我们可以将 O 替换为从 $\mathbb{SO}(n-1)$ 中 Haar 分布的矩阵，因此可以直接应用 定理 D.3 得到

$$\lim_{n\to\infty}\lambda_1(\widehat{D})=\sup\operatorname{supp}(\mu_{\widetilde{\Lambda}\widetilde{\Lambda}^\top}\boxtimes\mu_{\widetilde{T}})$$

几乎必然成立，其中 $n\to\infty$，这里我们用到了

$$\mu_{\sqrt{\widetilde{\Lambda}\widetilde{\Lambda}^\top}\boldsymbol{u}^\top\widetilde{T}\boldsymbol{u}\sqrt{\widetilde{\Lambda}\widetilde{\Lambda}^\top}}=\mu_{\widetilde{\Lambda}\widetilde{\Lambda}^\top}\boxtimes\mu_{\widetilde{T}},$$

对于 Haar 酉元 $\boldsymbol{u}$。

为了完成证明，必须对矩阵 $n\to\infty$ 和 $\widetilde{\Lambda}\widetilde{\Lambda}^\top$ $\widetilde{T}$取极限。实际上，利用 [68,推论 2.1 和定理 A.1], ，我们有

$$\sup\operatorname{supp}\mu_{\widetilde{\Lambda}\widetilde{\Lambda}^\top}\boxtimes\mu_{\widetilde{T}}\to\sup\operatorname{supp}\mu_{\Sigma_n}\boxtimes\mu_{\mathsf{T}}$$

当 $n\to\infty$ 时几乎必然成立，前提是矩阵强收敛，即对于每个固定的 $\varepsilon>0$,

$$\operatorname{spec}(\widetilde{T})\subset\operatorname{supp}\mu_{\mathsf{T}}+(-\varepsilon,\varepsilon),\tag{D.11}$$

$$\operatorname{spec}(\widetilde{\Lambda}\widetilde{\Lambda}^\top)\subset\operatorname{supp}\mu_{\Sigma_n}+(-\varepsilon,\varepsilon).\tag{D.12}$$

事实上，(D.11) 总是由假设（A18）保证。另一方面，对于(D.12)，请注意 t

$$\mu_{\widetilde{\Lambda}\widetilde{\Lambda}^\top} = \begin{cases} \frac{d-1}{n-1}\mu_{\widetilde{\Sigma}} + \frac{n-d}{n-1}\delta_0 & \text{if } n \geq d, \\ \mu_{\widetilde{\Sigma}} & \text{if } n < d. \end{cases}$$

因此，只要 $\delta \neq 1$，测度$\mu_{\widetilde{\Lambda}\widetilde{\Lambda}^\top}$ 和 $\mu_{\mathbf{\Sigma}_n}$ 在原点处的点质量大小相当（或者二者均无点质量）。在这种情况下，假设（A18）同样保证(D.12)。因此，对于 $\delta \neq 1$ 我们已证明

$$\limsup_{n\to\infty} \lambda_1(\widehat{D}) = \limsup_{n\to\infty} \sup \operatorname{supp}(\mu_{\widetilde{\Lambda}\widetilde{\Lambda}^\top} \boxtimes \mu_{\widetilde{T}}) = \sup \operatorname{supp}(\mu_{\mathbf{\Sigma}_n} \boxtimes \mu_{\mathsf{T}}).$$

然而，当(D.12) 可能不成立，因为 $\delta = 1$，$\mu_{\mathbf{\Sigma}_n} = \mu_{\mathbf{\Sigma}}$ 在零点处没有点质量，而 $\widetilde{\Lambda}\widetilde{\Lambda}^\top$（以及 $\Lambda\Lambda^\top$）可能具有 $o(1)$ 零度。为处理这一情况，我们使用定理 D.7得到

$$\sup \operatorname{supp}(\mu_{\widetilde{\Lambda}\widetilde{\Lambda}^\top} \boxtimes \mu_{\widetilde{T}}) \leq \sup \operatorname{supp}(\mu_{\widetilde{\Sigma}} \boxtimes \mu_{\widetilde{T}}).$$

现在，$\mu_{\widetilde{\Sigma}}$ 按照 $\mu_{\mathbf{\Sigma}} = \mu_{\mathbf{\Sigma}_n}$ 的意义，依据(D.12) 通过柯西交错定理「强」收敛于，因此

$$\sup \operatorname{supp}(\mu_{\widetilde{\Sigma}} \boxtimes \mu_{\widetilde{T}}) \to \sup \operatorname{supp}(\mu_{\mathbf{\Sigma}_n} \boxtimes \mu_{\mathsf{T}})$$

当 $n \to \infty$时几乎必然成立。这就完成了定理 D.2的证明。

D.3 自由乘法卷积的谱边缘

本节给出自由乘法卷积的刻画。由于其本身具有独立研究价值，我们将 Σ 和 T 替换为一般算子 $\boldsymbol{s}$ 和 $\boldsymbol{t}$，它们是在 W^*-概率空间 $(\mathcal{M}, \langle\cdot\rangle)$ 中的自伴自由元素，满足 $\boldsymbol{s} \geq 0$。对谱测度 $\mu_{\boldsymbol{s}}$ 和 $\mu_{\boldsymbol{t}}$ 作如下假设。

(A21) $\mu_{\boldsymbol{s}}$ 和 $\mu_{\boldsymbol{t}}$ 是非退化的，即不是点质量。

(A22) 映射

$$\mathbb{C} \setminus [0,\infty) \ni z \mapsto \int_0^\infty \frac{s}{s-z} \mathrm{d}\mu_{\boldsymbol{s}}(s) = \left\langle \frac{\boldsymbol{s}}{\boldsymbol{s}-z} \right\rangle \in \mathbb{C}$$

可解析地延拓穿过原点，即谱测度 $\mu_{\boldsymbol{s}}$ 支撑于一个严格正区间上，但在原点处可能存在一个点质量。记 $\widetilde{\mu}_{\boldsymbol{s}} = \mu_{\boldsymbol{s}} - \mu_{\boldsymbol{s}}(\{0\})\delta_0$。

(A23) $\sup \operatorname{supp} \mu_{\boldsymbol{t}} > 0$。

(A24) 我们有

$$\lim_{z \searrow \sup \operatorname{supp} \mu_{\boldsymbol{s}}} \left\langle \frac{\boldsymbol{s}}{\boldsymbol{s}-z} \right\rangle = -\infty = \lim_{z \nearrow \inf \operatorname{supp} \widetilde{\mu}_{\boldsymbol{s}}} \left\langle \frac{\boldsymbol{s}}{z-\boldsymbol{s}} \right\rangle,$$
$$\lim_{z \searrow \sup \operatorname{supp} \mu_{\boldsymbol{t}}} \left\langle \frac{\boldsymbol{t}}{\boldsymbol{t}-z} \right\rangle = -\infty.$$

引理 D.8。

1. 定义 $\mathcal{S}_{\boldsymbol{t}}$、$\mathcal{S}'_{\boldsymbol{t}} \subset \mathbb{R}$ 为

$$\mathcal{S}_{\boldsymbol{t}} \coloneqq \left\{ w \in \mathbb{R} \setminus \operatorname{supp} \mu_{\boldsymbol{t}} : \left\langle \frac{\boldsymbol{t}}{\boldsymbol{t}-w} \right\rangle = 0 \right\},$$
$$\mathcal{S}'_{\boldsymbol{t}} \coloneqq \left\{ w \in \mathbb{R} \setminus \operatorname{supp} \mu_{\boldsymbol{t}} : \left\langle \frac{\boldsymbol{t}}{\boldsymbol{t}-w} \right\rangle = 1 - \mu_{\boldsymbol{s}}(\{0\}) \right\}.$$

那么 $\mathcal{S}_{\boldsymbol{t}} \cap (\sup \operatorname{supp} \mu_{\boldsymbol{t}}, \infty)$ 和 $\mathcal{S}'_{\boldsymbol{t}} \cap (\sup \operatorname{supp} \mu_{\boldsymbol{t}}, \infty)$ 都是有限集。

2. 对于每个 $w \in (\sup \operatorname{supp} \mu_{\boldsymbol{t}}, \infty) \setminus \mathcal{S}$, 存在唯一的 $\omega(w) \in \mathbb{R} \setminus (\inf \operatorname{supp} \widetilde{\mu}_{\boldsymbol{s}}, \sup \operatorname{supp} \widetilde{\mu}_{\boldsymbol{s}})$ 使得 t

$$\left\langle \frac{\boldsymbol{t}}{\boldsymbol{t}-w} \right\rangle = \left\langle \frac{\boldsymbol{s}}{\boldsymbol{s}-\omega(w)} \right\rangle. \tag{D.13}$$

3. 映射 $\omega : (\sup \operatorname{supp} \mu_{\boldsymbol{t}}, \infty) \setminus \mathcal{S}_{\boldsymbol{t}} \to \mathbb{R}$ 在 条目 *2* 中定义，并亚纯地延拓到 $\mathbb{C}$中的一个开集，该开集包含 $(\sup \operatorname{supp} \mu_{\boldsymbol{t}}, \infty)$。该延拓在每个 $w \in (\sup \operatorname{supp} \mu_{\boldsymbol{t}}, \infty) \setminus \mathcal{S}_{\boldsymbol{t}}$处解析，在每个 $w \in \mathcal{S}_{\boldsymbol{t}}$处有一个极点，并在每个 $w \in \mathcal{S}'_{\boldsymbol{t}}$处有一个零点。

4. 将函数 $\psi : (\operatorname{supsupp} \mu_{\boldsymbol{t}}, \infty) \setminus \mathcal{S}_{\boldsymbol{t}} \to \mathbb{R}$ 定义为

$$\psi(w) = w\omega(w)\left(1 - \left\langle \frac{\boldsymbol{s}}{\boldsymbol{s}-\omega(w)} \right\rangle^{-1}\right)^{-1} \equiv w\omega(w)\left(1 - \left\langle \frac{\boldsymbol{t}}{\boldsymbol{t}-w} \right\rangle^{-1}\right)^{-1}. \tag{D.14}$$

于是 ψ 解析延拓 到 $\mathbb{C}$ 中包含 $(\sup \operatorname{supp} \mu_{\boldsymbol{t}}, \infty)$的一个开集，且 $\lim_{w \to +\infty} \psi'(w) = \langle \boldsymbol{s} \rangle$.

证明。 关于 条目 1 至 3的证明，见 [107, 引理 E 的条目 1 至 3。2]。对于条目 4，注意我们可以将 ψ写为

$$\psi(w) = w\left\langle \frac{\boldsymbol{s}}{\boldsymbol{s}-\omega(w)} \right\rangle \left\langle \frac{1}{\boldsymbol{s}-\omega(w)} \right\rangle^{-1}.$$

因此 ψ由$(\sup \operatorname{supp} \mu_{\boldsymbol{t}}, \infty)$ 是其邻域上的一个亚纯函数，这由条目 2可知，可能的极点位于$\mathcal{S}_{\boldsymbol{t}}$。因此，我们只需检查每个 $w \in \mathcal{S}_{\boldsymbol{t}}$ 对 ψ而言是否为可去奇点。回忆一下， $|\omega(z)| \to \infty$当 $z \to w \in \mathcal{S}_{\boldsymbol{t}}$时，由控制收敛可得

$$\psi(z) = z\left\langle \frac{\boldsymbol{s}\omega(z)}{\boldsymbol{s}-\omega(z)} \right\rangle \left\langle \frac{\omega(z)}{\boldsymbol{s}-\omega(z)} \right\rangle^{-1} \to \langle \boldsymbol{s} \rangle w, \qquad \text{as } z \to w. \tag{D.15}$$

这表明， w 是 ψ的可去奇点。对于关于 ψ'的最后一部分，直接计算可得，只要 $w > \operatorname{supsupp} \mu_{\boldsymbol{t}}$,

$$\begin{aligned}
\psi'(w) &= \frac{\mathrm{d}}{\mathrm{d}w}\left(\omega(w)\left\langle \frac{\boldsymbol{t}}{\boldsymbol{t}-w} \right\rangle \left\langle \frac{1}{\boldsymbol{t}-w} \right\rangle^{-1}\right) \\
&= \left(\omega'(w)\left\langle \frac{\boldsymbol{t}}{\boldsymbol{t}-w} \right\rangle + \omega(w)\left\langle \frac{\boldsymbol{t}}{(\boldsymbol{t}-w)^2} \right\rangle\right)\left\langle \frac{1}{\boldsymbol{t}-w} \right\rangle^{-1} - \omega(w)\left\langle \frac{\boldsymbol{t}}{\boldsymbol{t}-w} \right\rangle \left\langle \frac{1}{(\boldsymbol{t}-w)^2} \right\rangle \left\langle \frac{1}{\boldsymbol{t}-w} \right\rangle^{-2} \\
&= \left(\left\langle \frac{\boldsymbol{t}}{(\boldsymbol{t}-w)^2} \right\rangle \left\langle \frac{\boldsymbol{s}}{(\boldsymbol{s}-\omega(w))^2} \right\rangle^{-1} \left\langle \frac{\boldsymbol{t}}{\boldsymbol{t}-w} \right\rangle + \omega(w)\left\langle \frac{\boldsymbol{t}}{(\boldsymbol{t}-w)^2} \right\rangle\right)\left\langle \frac{1}{\boldsymbol{t}-w} \right\rangle^{-1} \\
&\quad - \omega(w)\left\langle \frac{\boldsymbol{t}}{\boldsymbol{t}-w} \right\rangle \left\langle \frac{1}{(\boldsymbol{t}-w)^2} \right\rangle \left\langle \frac{1}{\boldsymbol{t}-w} \right\rangle^{-2} \qquad (D.16) \\
&= \left(\left\langle \frac{\boldsymbol{s}}{(\boldsymbol{s}-\omega(w))^2} \right\rangle^{-1} \left\langle \frac{\boldsymbol{s}}{\boldsymbol{s}-\omega(w)} \right\rangle + \omega(w)\right)\left\langle \frac{\boldsymbol{t}}{(\boldsymbol{t}-w)^2} \right\rangle \left\langle \frac{1}{\boldsymbol{t}-w} \right\rangle^{-1} \\
&\quad - \omega(w)\left\langle \frac{\boldsymbol{t}}{\boldsymbol{t}-w} \right\rangle \left\langle \frac{1}{(\boldsymbol{t}-w)^2} \right\rangle \left\langle \frac{1}{\boldsymbol{t}-w} \right\rangle^{-2} \\
&= \left\langle \frac{\boldsymbol{t}}{(\boldsymbol{t}-w)^2} \right\rangle \left\langle \frac{\boldsymbol{s}}{(\boldsymbol{s}-\omega(w))^2} \right\rangle^{-1} \left\langle \frac{\boldsymbol{s}^2}{(\boldsymbol{s}-\omega(w))^2} \right\rangle \left\langle \frac{1}{\boldsymbol{t}-w} \right\rangle^{-1} \\
&\quad - \left\langle \frac{\omega(w)\boldsymbol{s}}{\boldsymbol{s}-\omega(w)} \right\rangle \left\langle \frac{1}{\boldsymbol{t}-w} \right\rangle^{-2} \left\langle \frac{1}{(\boldsymbol{t}-w)^2} \right\rangle, \qquad (D.17)
\end{aligned}$$

其中在 (D.16)中我们使用以下由定义方程 $\omega'(w)$ 推导出的表达式 (D.13) 以及隐函数定理:

$$\omega'(w) = \left\langle \frac{\boldsymbol{t}}{(\boldsymbol{t}-w)^2} \right\rangle \left\langle \frac{\boldsymbol{s}}{(\boldsymbol{s}-\omega(w))^2} \right\rangle^{-1}.$$

利用 $|\omega(w)| \to \infty$ 作为 $w \to +\infty$，我们很容易验证 (D.17) 的第一项趋于零，而第二项在极限下收敛到 $\langle \boldsymbol{s} \rangle$ 。 □

定理 D.9. 我们有

$$\sup \operatorname{supp} \mu_{\sqrt{\boldsymbol{s}}\boldsymbol{t}\sqrt{\boldsymbol{s}}} = \psi(a^\circ),$$

其中 $a^\circ > \operatorname{supsupp}\, \mu_{\boldsymbol{t}}$ 是解析函数 $\psi : (\sup \operatorname{supp}\, \mu_{\boldsymbol{t}}, \infty) \to \mathbb{R}$ 在上文 条目 *4* 中定义的定理 *D.8*的最大临界点。

我们在汇集所需结果后，于本节末尾证明 定理 D.9。

D.3.1 设置

记 $\mathbb{C}_+ := \{z \in \mathbb{C} : \operatorname{Im}\, z > 0\}$ 和 $\mathbb{C}_- := \{z \in \mathbb{C} : \operatorname{Im} z < 0\}$。

定义 D。 **1.** 设 $\boldsymbol{a} \in \mathcal{M}$ 为一个非零自伴元素。在 $\mathbb{C}_+$ 上定义两个解析映射如下:

$$K_{\boldsymbol{a}}(z) := 1 - \left\langle \frac{\boldsymbol{a}}{\boldsymbol{a}-z} \right\rangle^{-1},$$
$$L_{\boldsymbol{a}}(z) := \frac{K_{\boldsymbol{a}}(z)}{z} \equiv \left\langle \frac{1}{\boldsymbol{a}-z} \right\rangle \left\langle \frac{\boldsymbol{a}}{\boldsymbol{a}-z} \right\rangle^{-1}.$$

注 D.1。 注意， $K_{\boldsymbol{a}}$ 确实定义良好，即映射 $\langle \boldsymbol{a}/(\boldsymbol{a}-z) \rangle$ 在 $\mathbb{C}_+$中没有零点。为此，反设存在一个零点 $z \in \mathbb{C}_+$, 则有

$$\int_{\mathbb{R}} \frac{x}{x-z} \mathrm{d}\mu_{\boldsymbol{a}}(x) = \left\langle \frac{\boldsymbol{a}}{\boldsymbol{a}-z} \right\rangle = 0.$$

于是我们得到

$$\operatorname{Im} \int_{\mathbb{R}} \frac{x}{x-z} \mathrm{d}\mu_{\boldsymbol{a}}(x) = \operatorname{Im} z \int_{\mathbb{R}} \frac{x}{|x-z|^2} \mathrm{d}\mu_{\boldsymbol{a}}(x) = 0$$

and

$$\operatorname{Re} \int_{\mathbb{R}} \frac{x}{x-z} \mathrm{d}\mu_{\boldsymbol{a}}(x) = \int_{\mathbb{R}} \frac{x^2 - x \operatorname{Re} z}{|x-z|^2} \mathrm{d}\mu_{\boldsymbol{a}}(x) = 0.$$

结合这两个方程可得

$$\int_{\mathbb{R}} \frac{x^2}{|x-z|^2} \mathrm{d}\mu_{\boldsymbol{a}}(x) = 0,$$

这显然是一个矛盾。

引理 D.10. 设 $\boldsymbol{a} \in \mathcal{M}$ 是自伴的，且具有非退化谱测度。设 $K_{\boldsymbol{a}}$ 和 $L_{\boldsymbol{a}}$ 由 定义 *D.1*给出。

1. $K_{\boldsymbol{a}}$ 和 $L_{\boldsymbol{a}}$ 亚纯延拓至 $\mathbb{C}\setminus$ 支撑 $\mu_{\boldsymbol{a}}$，其中 $K_{\boldsymbol{a}}(z)=\overline{K_{\boldsymbol{a}}(z)}$ 以及 $L_{\boldsymbol{a}}(z)=\overline{L_{\boldsymbol{a}}(z)}$。这两个亚纯延拓在 $\mathbb{C}\setminus$ 支撑 $\mu_{\boldsymbol{a}}$中具有相同的极点集合，其表达式为

$$\mathcal{S}_{\boldsymbol{a}}=\left\{z\in\mathbb{R}\setminus\operatorname{supp}\mu_{\boldsymbol{a}}:\left\langle\frac{\boldsymbol{a}}{\boldsymbol{a}-z}\right\rangle=0\right\}. \tag{D.18}$$

2. $L_{\boldsymbol{a}}(\mathbb{C}_+)\subset\mathbb{C}_+$。此外，若 $\boldsymbol{a}\geq 0$，则 $K_{\boldsymbol{a}}(\mathbb{C}_+)\subset\mathbb{C}_+$。

证明。 条目 1 是由 $K_{\boldsymbol{a}}$ 和 $L_{\boldsymbol{a}}$ 的定义以及映射

$$\mathbb{C}\setminus\operatorname{supp}\mu_{\boldsymbol{a}}\ni z\mapsto\left\langle\frac{\boldsymbol{a}}{\boldsymbol{a}-z}\right\rangle\ni\mathbb{C} \tag{D.19}$$

是解析的。 $K_{\boldsymbol{a}}$ 和 $L_{\boldsymbol{a}}$ 的极点恰好是该函数的零点；注意，这两个映射

$$\left\langle\frac{1}{\boldsymbol{a}-z}\right\rangle,\qquad\qquad\left\langle\frac{\boldsymbol{a}}{\boldsymbol{a}-z}\right\rangle$$

不可能有公共零点。最后，(D.19) 中的函数的零点全为实数，这由 注 D.1可知，进而 (D.18)得证。

对于 条目 2，为验证 $L_{\boldsymbol{a}}(\mathbb{C}_+)\subset\mathbb{C}_+$，令 $z\in\mathbb{C}_+$ 并写成

$$\begin{aligned}\operatorname{Im}L_{\boldsymbol{a}}(z)&=\operatorname{Im}\left\langle\frac{1}{\boldsymbol{a}-z}\right\rangle\left\langle\frac{\boldsymbol{a}}{\boldsymbol{a}-z}\right\rangle^{-1}=\left|\left\langle\frac{\boldsymbol{a}}{\boldsymbol{a}-z}\right\rangle\right|^{-2}\operatorname{Im}\left[\left\langle\frac{1}{\boldsymbol{a}-z}\right\rangle\left(1+\bar{z}\overline{\left\langle\frac{1}{\boldsymbol{a}-z}\right\rangle}\right)\right]\\&=\operatorname{Im}z\left|\int_{\mathbb{R}}\frac{x}{x-z}\mathrm{d}\mu_{\boldsymbol{a}}(x)\right|^{-2}\left(\int_{\mathbb{R}}\frac{1}{|x-z|^2}\mathrm{d}\mu_{\boldsymbol{a}}(x)-\left|\int_{\mathbb{R}}\frac{1}{x-z}\mathrm{d}\mu_{\boldsymbol{a}}(x)\right|^2\right)>0,\end{aligned} \tag{D.20}$$

其中最后一个不等式用到了柯西-施瓦茨不等式。对于 $\boldsymbol{a}\geq 0$，注意到 $K_{\boldsymbol{a}}(\mathbb{C}_+)\subset\mathbb{C}_+$，即可得出这一事实

$$\left\langle\frac{\boldsymbol{a}}{\boldsymbol{a}-z}\right\rangle=\int_{\mathbb{R}}\frac{1}{x-z}(x\mathrm{d}\mu_{\boldsymbol{a}}(x))$$

是正测度 $x\mathrm{d}\mu_{\boldsymbol{a}}(x)$的斯蒂尔切斯变换。 □

引理 D.11. 映射 $K_{\boldsymbol{a}}$ 在 ∞ 附近是亚纯函数，在 ∞处有一个极点。当 $\langle\boldsymbol{a}\rangle\neq 0$ 时，该极点是简单极点；当 $\langle\boldsymbol{a}\rangle=0$ 时，则为二阶极点。更准确地说，存在常数 $c>0$ 和在 $\{z$ 上的解析函数 $g:\ |z|<c\}$ 使得

$$\frac{1}{K_{\boldsymbol{a}}(1/z)}=\langle\boldsymbol{a}\rangle z+(\langle\boldsymbol{a}^2\rangle-\langle\boldsymbol{a}\rangle^2)z^2+z^3g(z),\qquad |z|<c. \tag{D.21}$$

因此，存在常数 $C>0$，使得取以下选择时有

$$\mathcal{N}:=\{z:|z|>C\},\qquad \mathcal{C}:=\{z:|z|>C,\ \operatorname{Re}z>C|\operatorname{Im}z|\}.$$

1. 若 $\langle\boldsymbol{a}\rangle>0$，则存在 $K_{\boldsymbol{a}}$ 在 $\mathcal{N}$上的唯一解析逆函数 $K_{\boldsymbol{a}}^{-1}$。映射 $K_{\boldsymbol{a}}^{-1}$分别将 $\mathcal{N}\cap\mathbb{C}_+$、$\mathcal{N}\cap\mathbb{C}_-$和 $\mathcal{N}\cap\mathbb{R}$ 映至 $\mathbb{C}_+$、$\mathbb{C}_-$和 $\mathbb{R}$。

2. 若 $\langle\boldsymbol{a}\rangle=0$，则存在定义在 $\mathcal{N}$上的唯一解析映射 $h_{\boldsymbol{a}}$，使得

$$K_{\boldsymbol{a}}(h_{\boldsymbol{a}}(\sqrt{z}))=z=K_{\boldsymbol{a}}(h_{\boldsymbol{a}}(-\sqrt{z})),\qquad z\in\mathcal{N}, \tag{D.22}$$

$$h_{\boldsymbol{a}}(\sqrt{K_{\boldsymbol{a}}(w)})=w,\qquad w\in\mathcal{C}, \tag{D.23}$$

其中我们选择 $\sqrt{z}$的分支为$(-\infty,0)$，即对 $\theta\in(-\pi,\pi]$有 $\sqrt{r\mathrm{e}^{\mathrm{i}\theta}}=\sqrt{r}\mathrm{e}^{\mathrm{i}\theta/2}$。映射 $h_{\boldsymbol{a}}$ 分别将$\mathcal{C}\cap\mathbb{C}_+$、$\mathcal{C}\cap\mathbb{C}_-$和 $\mathcal{C}\cap\mathbb{R}$ 映入 $\mathbb{C}_+$、$\mathbb{C}_-$和 $\mathbb{R}$。在此情形下，我们在 ∞的一个适当邻域上记 $K_{\boldsymbol{a}}^{-1}(z):=h_{\boldsymbol{a}}(\sqrt{z})$。

证明。 首先注意，(D.21) 是定义的直接推论：

$$1/K_{\boldsymbol{a}}(1/z) = z\left\langle \frac{\boldsymbol{a}}{1-z\boldsymbol{a}}\right\rangle\left\langle \frac{1}{1-z\boldsymbol{a}}\right\rangle^{-1} = z\left(\langle \boldsymbol{a}\rangle + z\langle \boldsymbol{a}^2\rangle + O(|z|^2)\right)\left(1+\langle \boldsymbol{a}\rangle z + O(|z|^2)\right)^{-1} \\ = \langle \boldsymbol{a}\rangle z + (\langle \boldsymbol{a}^2\rangle - \langle \boldsymbol{a}\rangle^2)z^2 + O(|z|^3). \tag{D.24}$$

对于条目 1，由于(D.21) 中的线性项不为零，因此在原点附近存在唯一的解析逆函数 f ，它是$1/K_{\boldsymbol{a}}($ $1/z)$ 。然后容易验证， $1/f(1/z)$ 是 $K_{\boldsymbol{a}}$的反函数。最后一个断言源于以下事实：当$\operatorname{Im} K_{\boldsymbol{a}}(z)$ 足够大时，其符号与$\operatorname{Im} z$ 相同，而这又是由 $|z|$推出的

$$\operatorname{Im} K_{\boldsymbol{a}}(z) = \operatorname{Im} z\left|\left\langle \frac{\boldsymbol{a}}{\boldsymbol{a}-z}\right\rangle\right|^{-2}\left\langle \frac{\boldsymbol{a}}{|\boldsymbol{a}-z|^2}\right\rangle = \operatorname{Im} z\frac{\langle \boldsymbol{a}\rangle|z|^{-2}+O(|z|^{-3})}{|\langle \boldsymbol{a}\rangle z^{-1}+O(|z|^{-2})|^2} = \operatorname{Im} z\left(\frac{1}{\langle \boldsymbol{a}\rangle}+O(|z|^{-1})\right).$$

对于条目 2，我们可以找到唯一的解析函数 $p_{\boldsymbol{a}}(z)$，它定义在原点的某个邻域内，并满足

$$\frac{1}{K_{\boldsymbol{a}}(1/z)} = p_{\boldsymbol{a}}(z)^2, \qquad p_{\boldsymbol{a}}(0)=0, \qquad p'_{\boldsymbol{a}}(0) = \sqrt{\langle \boldsymbol{a}^2\rangle} > 0. \tag{D.25}$$

根据上述相同的推理，映射 $p_{\boldsymbol{a}}$ 在原点附近有一个逆映射 $p_{\boldsymbol{a}}^{-1}$。然后我们选择$h_{\boldsymbol{a}}(z) = 1/p_{\boldsymbol{a}}^{-1}(1/z)$ ，使其定义在 ∞的某个邻域内。现在，我们可以将(D.22) 的第一个等式写作

$$K_{\boldsymbol{a}}(h_{\boldsymbol{a}}(\sqrt{z})) = K_{\boldsymbol{a}}(1/p_{\boldsymbol{a}}^{-1}(1/\sqrt{z})) = \frac{1}{p_{\boldsymbol{a}}(p_{\boldsymbol{a}}^{-1}(1/\sqrt{z}))^2} = z.$$

同理，我们可以证明(D.22)中的第二个等式。

为证明(D.23)，注意到 $p'_{\boldsymbol{a}}(0) > 0$ 意味着存在常数 $c>0$ ，使得 $z\mapsto p_{\boldsymbol{a}}(z)^2$将锥 $\mathcal{C}' := \{|w| < c, |\operatorname{Im} w| < c\operatorname{Re} w\}$ 映射到裂平面 $\mathbb{C}\setminus(-\infty,0]$。在此锥上，我们有$\sqrt{p_{\boldsymbol{a}}(w)^2} = p_{\boldsymbol{a}}(w)$，因此对于 $w^{-1}\in\mathcal{C}'$ ，有

$$h_{\boldsymbol{a}}(\sqrt{K_{\boldsymbol{a}}(w)}) = \frac{1}{p_{\boldsymbol{a}}^{-1}(1/\sqrt{K_{\boldsymbol{a}}(w)})} = \frac{1}{p_{\boldsymbol{a}}^{-1}(\sqrt{1/K_{\boldsymbol{a}}(w)})} = \frac{1}{p_{\boldsymbol{a}}^{-1}(\sqrt{p_{\boldsymbol{a}}(1/w)^2})} = w.$$

注意到逆锥 $\{w^{-1}: \ w\in\mathcal{C}'\}$ 包含另一个形如 $\{w: \ |w|>C、\ \operatorname{Re} w > C|\operatorname{Im} w|\}$的集合，因此我们证明了 (D.23)。唯一性由 $p_{\boldsymbol{a}}$及其逆映射的唯一性推出。最后，对于关于 $h_{\boldsymbol{a}}(z)$奇偶性的最后一项断言，我们利用 $\langle \boldsymbol{a}\rangle = 0$ 来写出

$$\operatorname{Im} K_{\boldsymbol{a}}(w) = \operatorname{Im}\left[\left\langle \frac{\boldsymbol{a}}{\boldsymbol{a}-w}+\frac{\boldsymbol{a}}{w}\right\rangle^{-1}\right] = \operatorname{Im} w\left|\left\langle \frac{\boldsymbol{a}^2}{w(\boldsymbol{a}-w)}\right\rangle\right|^{-2}\left\langle \frac{\boldsymbol{a}^2(2\operatorname{Re} w-\boldsymbol{a})}{|w(\boldsymbol{a}-w)|^2}\right\rangle. \tag{D.26}$$

代入 $w = K_{\boldsymbol{a}}^{-1}(z)$ 并在一个锥中取 z

$$\{z : |z| > C', \operatorname{Re} z > C'|\operatorname{Im} z|\}, \tag{D.27}$$

我们有

$$\operatorname{Re} K_{\boldsymbol{a}}^{-1}(z) = \operatorname{Re} h_{\boldsymbol{a}}(\sqrt{z}) > \operatorname{Re}[\sqrt{\langle \boldsymbol{a}\rangle^2}\sqrt{z}(1+O(|z|^{-1/2}))] > \|\boldsymbol{a}\|,$$

这里我们用到了，根据 (D.25)， $h_{\boldsymbol{a}}$ 具有展开式

$$h_{\boldsymbol{a}}(z) = \frac{1}{p_{\boldsymbol{a}}^{-1}(1/z)} = \left(\frac{1}{p'_{\boldsymbol{a}}(0)z}+O(|z|^{-2})\right)^{-1} = \sqrt{\langle \boldsymbol{a}^2\rangle}z(1+O(|z|^{-1})).$$

这证明了(D.26) 中的第二个因子对于 $w = K_{\boldsymbol{a}}^{-1}(z)$为正，因此 $\operatorname{Im} K_{\boldsymbol{a}}^{-1}(z) = \operatorname{Im} h_{\boldsymbol{a}}(\sqrt{z})$ 与 $\operatorname{Im} z$同号，因而与 $\operatorname{Im}\sqrt{z}$同号。 □

D.3.2 解析从属

命题 D.12 ([2, 定理 1.3])。 存在唯一一对解析函数 $\Omega_1,\Omega_2:\mathbb{C}_+\to\mathbb{C}$ 满足以下条件，对每个 $z\in\mathbb{C}_+$:

$$\begin{aligned}\Omega_1(z)/z\in\mathbb{C}_+,\qquad \Omega_2(z)\in\mathbb{C}_+,\\ K_{\boldsymbol{s}}(\Omega_1(z))=K_{\boldsymbol{t}}(\Omega_2(z))=\frac{\Omega_1(z)\Omega_2(z)}{z},\\ \lim_{r\to+\infty}|\Omega_1(rz)|=\infty=\lim_{r\to+\infty}|\Omega_2(rz)|.\end{aligned}\tag{D.28}$$

此外，(D.28) 中的三个量均等于

$$K_{\boldsymbol{s}^{1/2}\boldsymbol{t}\boldsymbol{s}^{1/2}}(z).$$

D.3.3 解析从属的推论

接下来，我们记

$$\lambda^\circ:=\sup\operatorname{supp}\mu_{\sqrt{\boldsymbol{s}}\boldsymbol{t}\sqrt{\boldsymbol{s}}},$$

并用 $\mathbb{R}$ 表示 $\mathbb{R}$的单点紧化 $\mathbb{R}\cup\{\infty\}\cong\mathbb{S}^1$。

推论 D.13. 映射 Ω_1 和（分别）Ω_2可亚纯地（分别解析地）延拓穿过半直线 (λ°,∞)的一个邻域。它们对 (λ°,∞) 的像满足 $\Omega_1((\lambda^\circ,\infty))\subset\mathbb{R}\setminus\operatorname{supp}\mu_{\boldsymbol{\Sigma}}$,$\Omega_2((\lambda^\circ,\infty))\subset(\sup\operatorname{supp}\mu_{\mathsf{T}},\infty)$。最后，映射 Ω_2 在 (λ°,∞)上严格递增。

证明。将定理 D.8 的条目 3 中的 $\boldsymbol{t}$ 替换为 $\boldsymbol{s}\boldsymbol{t}$后，我们发现映射 $K_{\boldsymbol{s}}^{-1}\circ K_{\boldsymbol{s}\boldsymbol{t}}:(\lambda^\circ,\infty)\to\mathbb{R}\setminus\operatorname{supp}\mu_{\boldsymbol{\Sigma}}$ 定义良好，并可亚纯地延拓到 (λ°,∞) 的一个邻域。另一方面，定理 D.12 表明，在靠近 ∞ 的一个开集内，$K_{\boldsymbol{s}}^{-1}\circ K_{\boldsymbol{s}\boldsymbol{t}}$ 与 Ω_1 相一致。因此，$K_{\boldsymbol{s}}^{-1}\circ K_{\boldsymbol{s}\boldsymbol{t}}$ 为 Ω_1 提供了一个亚纯延拓，使其延拓到 (λ°,∞) 的一个邻域。接着，我们通过 (D.28) 将 Ω_2 延拓到同一个邻域:

$$\Omega_2(z)=\frac{zK_{\boldsymbol{s}\boldsymbol{t}}(z)}{\Omega_1(z)}.$$

还可立即得出，$\Omega_1((\lambda^\circ,\infty))\subset\mathbb{R}\setminus\operatorname{supp}\mu_{\boldsymbol{\Sigma}}$，且 $\Omega_2((\lambda^\circ,\infty))\subset\mathbb{R}$。

接下来，我们证明 Ω_2 在 (λ°,∞) 内没有极点。反设存在一个极点 $x>\lambda^\circ$。那么我们有

$$\lim_{z\to x,z\in\mathbb{C}_+}|L_{\boldsymbol{s}}(\Omega_1(z))|=\lim_{z\to x,z\in\mathbb{C}_+}\frac{|\Omega_2(z)|}{|z|}=\infty,$$

这就产生了矛盾，因为 $\Omega_1(x)\in\mathbb{R}\setminus\operatorname{supp}\mu_{\boldsymbol{\Sigma}}$。因此，$\Omega_2$ 在 (λ°,∞) 中没有极点，故可解析地延拓穿过直线 (λ°,∞)。

接下来，我们断言存在一个有限的博雷尔测度 $\widetilde{\mu}$，使得

$$\Omega_2(z)=\frac{z}{\langle\boldsymbol{s}\rangle}-\frac{\langle\boldsymbol{s}^2\rangle-\langle\boldsymbol{s}\rangle^2}{\langle\boldsymbol{s}\rangle^2}\langle\boldsymbol{t}\rangle+\int_{\mathbb{R}}\frac{1}{x-z}\mathrm{d}\widetilde{\mu}(x),\qquad z\in\mathbb{C}_+,\tag{D.29}$$

根据 Nevalinna-Pick 表示定理，这又可由下式推出:

$$\limsup_{y\to\infty}y\left|\Omega_2(\mathrm{i}y)-\frac{1}{\langle\boldsymbol{s}\rangle}\mathrm{i}y+\frac{\langle\boldsymbol{s}^2\rangle-\langle\boldsymbol{s}\rangle^2}{\langle\boldsymbol{s}\rangle^2}\langle\boldsymbol{t}\rangle\right|<\infty.\tag{D.30}$$

为了证明 （D.30），当 y 足够大，使得 $\Omega_1(\mathrm{i}y)=K_{\boldsymbol{s}}^{-1}(K_{\boldsymbol{s}\boldsymbol{t}}(\mathrm{i}y))$ 且 $\Omega_2(\mathrm{i}y)=K_{\boldsymbol{t}}^{-1}(K_{\boldsymbol{s}\boldsymbol{t}}(\mathrm{i}y))$，注意到

$$\Omega_2(\mathrm{i}y)-\frac{1}{\langle\boldsymbol{s}\rangle}\mathrm{i}y=\mathrm{i}y\left(L_{\boldsymbol{s}}(\Omega_1(\mathrm{i}y))-\frac{1}{\langle\boldsymbol{s}\rangle}\right).\tag{D.31}$$

接着，我们使用 （D.21）得到

$$L_{\boldsymbol{s}}(\Omega_1(\mathrm{i}y)) - \frac{1}{\langle \boldsymbol{s}\rangle} = \frac{1}{\langle \boldsymbol{s}\rangle + (\langle \boldsymbol{s}^2\rangle - \langle \boldsymbol{s}\rangle^2)\Omega_1(\mathrm{i}y)^{-1} + O(|\Omega_1(\mathrm{i}y)|^{-2})} - \frac{1}{\langle \boldsymbol{s}\rangle} = -\frac{\langle \boldsymbol{s}^2\rangle - \langle \boldsymbol{s}\rangle^2}{\langle \boldsymbol{s}\rangle^2\Omega_1(\mathrm{i}y)} + O(|\Omega_1(\mathrm{i}y)|^{-2}),$$

因此

$$\begin{aligned} y\left|\Omega_2(\mathrm{i}y) - \frac{1}{\langle \boldsymbol{s}\rangle}\mathrm{i}y + \frac{\langle \boldsymbol{s}^2\rangle - \langle \boldsymbol{s}\rangle^2}{\langle \boldsymbol{s}\rangle^2}\langle \boldsymbol{t}\rangle\right| &\le \frac{y(\langle \boldsymbol{s}^2\rangle - \langle \boldsymbol{s}\rangle^2)}{\langle \boldsymbol{s}\rangle^2}\left|-\frac{\mathrm{i}y}{\Omega_1(\mathrm{i}y)} + \langle \boldsymbol{t}\rangle\right| + O\left(\frac{y^2}{|\Omega_1(\mathrm{i}y)|^2}\right) \\ &\lesssim y\left|-\frac{1}{L_{\boldsymbol{t}}(\Omega_2(\mathrm{i}y))} + \langle \boldsymbol{t}\rangle\right| + |L_{\boldsymbol{t}}(\Omega_2(\mathrm{i}y))|^{-2} \\ &\lesssim \frac{y}{|\Omega_2(\mathrm{i}y)|} + |L_{\boldsymbol{t}}(\Omega_2(\mathrm{i}y))|^{-2} \\ &= |L_{\boldsymbol{s}}(\Omega_1(\mathrm{i}y)|^{-1} + |L_{\boldsymbol{t}}(\Omega_2(\mathrm{i}y))|^{-2}, \end{aligned}$$

其中第三行由 （D.21）将 $\boldsymbol{t}$ 代入 $\boldsymbol{a}$得到。这证明了 （D.30），因为

$$\lim_{y\to\infty} L_{\boldsymbol{s}}(\Omega_1(\mathrm{i}y))^{-1} = \lim_{|\omega|\to\infty} L_{\boldsymbol{s}}(\omega)^{-1} = \langle \boldsymbol{s}\rangle, \quad \text{and} \quad \lim_{y\to\infty} L_{\boldsymbol{t}}(\Omega_2(\mathrm{i}y))^{-1} = \lim_{|\omega_2|\to\infty} L_{\boldsymbol{t}}(\omega_2)^{-1} = \langle \boldsymbol{t}\rangle.$$

对 （D.29）应用 Stieltjes 反演。由于 Ω_2 可解析地延拓穿过 （λ°， ∞）且在其中取实值，我们得到 $\operatorname{supp}\widetilde{\mu} \subset$ （$-\infty$， λ°]。由此还可推出 Ω_2 在 （λ°， ∞）上严格递增。如果Ω_2（x） $=\operatorname{supsupp}\mu_{\mathsf{T}}$ 对某个 $x > \lambda^\circ$成立，则有

$$-\infty = \lim_{w\searrow \sup\operatorname{supp}\mu_{\mathsf{T}}} \left\langle \frac{\boldsymbol{t}}{\boldsymbol{t}-w}\right\rangle = \lim_{z\searrow x}\left\langle \frac{\boldsymbol{t}}{\boldsymbol{t}-\Omega_2(z)}\right\rangle = \lim_{z\searrow x}\left\langle \frac{\boldsymbol{s}}{\boldsymbol{s}-\Omega_1(z)}\right\rangle,$$

这又意味着 $\lim_{z\searrow x}\Omega_1(z) = \sup\operatorname{supp}\mu_{\mathsf{\Sigma}}$。但这就产生了矛盾，因为

$$0 = \lim_{z\searrow x} K_{\boldsymbol{st}}(z) = \lim_{z\searrow x}\frac{\Omega_1(z)\Omega_2(z)}{z} = \frac{(\sup\operatorname{supp}\mu_{\mathsf{T}})(\sup\operatorname{supp}\mu_{\mathsf{\Sigma}})}{x}. \qquad \square$$

D.3.4 定理D.9的证明

定理 D.9 由以下引理得出。

引理 D.14。 设 $w > \sup\operatorname{supp}\mu_{\mathsf{T}}$。回忆一下在 ψ 中定义的映射 条目 *4*， 见 定理 *D.8*。

1. 若 $\psi(\widetilde{w}) > \lambda^\circ$ 对所有 $\widetilde{w} \ge w$均成立，则 $\psi'(w) > 0$。
2. 若 $\psi'(\widetilde{w}) > 0$ 对所有 $\widetilde{w} \ge w$均成立，则 $\psi(\widetilde{w}) \ge \lambda^\circ$。

我们省略 定理 D.9 的证明 定理 D.14， 因为它与 [107, 引理 5.5 的证明（见引理 E）完全相同1]。

定理 *D.14* 的证明。 回忆 定理 D.8 中的三元组 $(\psi(\widetilde{w})\ \omega(\widetilde{w})\ \widetilde{w})$， 它由亚纯函数组成，定义在$\widetilde{w} \in (\sup\operatorname{supp}\mu_{\mathsf{T}}, \infty)$上。由 ψ和 ω的定义还可知，该三元组满足与 (D.28)中相同的从属方程，即

$$K_{\boldsymbol{s}}(\omega(\widetilde{w})) = K_{\boldsymbol{t}}(\widetilde{w}) = \frac{\widetilde{w}\omega(\widetilde{w})}{\psi(\widetilde{w})}, \qquad \widetilde{w} > \sup\operatorname{supp}\mu_{\mathsf{T}}. \tag{D.32}$$

回顾定理 D.8 的证明可知， $\psi'(\widetilde{w})$ 在足够大的 $\widetilde{w} > 0$时为正。接下来我们断言，对于 $(\omega/\psi)'$，结论同样成立。首先注意到

$$\frac{\omega(\widetilde{w})}{\psi(\widetilde{w})} = \frac{\omega(\widetilde{w})}{\widetilde{w}}L_{\boldsymbol{s}}(\omega(\widetilde{w})) = \frac{1}{\widetilde{w}}K_{\boldsymbol{s}}(\omega(\widetilde{w})) = \frac{1}{\widetilde{w}}K_{\boldsymbol{t}}(\widetilde{w}) = L_{\boldsymbol{t}}(\widetilde{w}),$$

其中最后一个等号处用到了 ω 的定义。接着，我们可以利用(D.20)通过 Nevanlinna 导数轻松计算出导数:

$$\frac{\mathrm{d}}{\mathrm{d}\widetilde{w}} L_{\boldsymbol{t}}(\widetilde{w}) = \lim_{\eta \searrow 0} \frac{\operatorname{Im} L_{\boldsymbol{t}}(\widetilde{w} + \mathrm{i}\eta)}{\eta} = \left\langle \frac{\boldsymbol{t}}{\boldsymbol{t} - \widetilde{w}} \right\rangle^{-2} \left(\left\langle \frac{1}{(\boldsymbol{t} - \widetilde{w})^2} \right\rangle - \left\langle \frac{1}{\boldsymbol{t} - \widetilde{w}} \right\rangle^2 \right) > 0.$$

现在取足够大的 $\widetilde{w} > 0$，使得 ψ 和 ω/ψ 的导数均为正。由此可知，在 $\mathbb{C}_+$ 以 $\widetilde{w}$ 为基点的一个小开扇形区域内，所有 $(\psi(\widetilde{w})\ \omega(\widetilde{w})/\psi(\widetilde{w})\ \widetilde{w})$ 都属于 $\mathbb{C}_+$。由解析延拓，从属方程 (D.32) 在该扇形区域内成立，因此根据 定理 D.12 的唯一性，我们有

$$\begin{aligned} \Omega_2(\psi(\widetilde{w})) &= \widetilde{w}, \\ \Omega_1(\psi(\widetilde{w})) &= \omega(\widetilde{w}), \end{aligned} \tag{D.33}$$

在 $(\sup \operatorname{supp}\ \mu_{\mathsf{T}},\ \infty)$ 的某个邻域内。

我们先证明 条目 1。对 (D.33) 求导，并注意到根据 定理 D.13，Ω_2 在 $(\lambda^\circ,\ \infty)$ 上严格递增，$\psi'(\widetilde{w}) > 0$ 对所有 $\widetilde{w} \geq w$ 均成立。

对于条目 2，我们考虑解析反函数 $\psi^{-1}\colon [\psi(w)\ \infty) \to (w, \infty)$。根据 (D.33)，此反函数必须与 Ω_2，相符，因此 Ω_2 可解析地延拓到 $[\psi(w)\ \infty)$ 的一个邻域，并且在该区间上取实值。若反过来假设 $\lambda^\circ > \psi(w)$，那么对于 $x \in (\psi(w)\ \lambda^\circ) \setminus \{0\}$，我们将有

$$\lim_{\eta \searrow 0} \operatorname{Im} \left\langle \frac{1}{\boldsymbol{st} - x - \mathrm{i}\eta} \right\rangle = \frac{1}{x} \lim_{\eta \searrow 0} \operatorname{Im} \left\langle \frac{\boldsymbol{st}}{\boldsymbol{st} - x - \mathrm{i}\eta} \right\rangle = \frac{1}{x} \lim_{\eta \searrow 0} \operatorname{Im} \left\langle \frac{\boldsymbol{t}}{\boldsymbol{t} - \Omega_2(x + \mathrm{i}\eta)} \right\rangle = \frac{1}{x} \operatorname{Im} \left\langle \frac{\boldsymbol{t}}{\boldsymbol{t} - \Omega_2(x)} \right\rangle = 0,$$

其中最后两个等式由 定理 D.13 得出。由 Stieltjes 反演可知，这意味着$(\psi(w)\ \lambda^\circ) \setminus \{0\}$ 与 支撑集 $\mu_{\boldsymbol{st}}$不相交，这与定义 $\lambda^\circ =$ 支撑集 $\mu_{\boldsymbol{st}}$相矛盾。 □

D.4 针对我们设定的特化

我们对 T 的选择将由 (4.1a) 给出，其中的函数 $\mathcal{T}$ 满足 假设 (A5)。这意味着支撑集 $\mu_{\mathsf{T}} < 1$，此外还有

$$\mathbb{E}\left[\frac{\mathsf{T}}{\mathsf{T} - 1}\right] = \frac{1}{\gamma}\mathbb{E}[\overline{g}(\mathsf{Y})] = 0, \tag{D.34}$$

其中 $\mathsf{T} = \mathcal{T}(\mathsf{Y})$ (回忆一下 Y 见 (3.10))。下文中，根据 定理 D.2，我们采用以下选择应用 定理 D.9:

$$\mu_{\boldsymbol{s}} = \mu_{\Sigma_n}, \qquad \mu_{\boldsymbol{t}} = \mu_{\mathsf{T}}. \tag{D.35}$$

注意，在这些选择下，假设 (A21)至(A24)在D.3 节 中要求的条件均已由 假设 (A2) 和 (A5)涵盖。特别地，回忆 Σ_n 来自 (D.5) 及其与 Λ_n 在 (3.8)中的关系，可将 $\widetilde{\mu}_{\boldsymbol{s}}$ 在 假设 (A22) 识别为

$$\widetilde{\mu}_{\boldsymbol{s}} = \begin{cases} \operatorname{law}(\Lambda^2), & \delta \leq 1 \\ \delta^{-1} \operatorname{law}(\Lambda^2), & \delta > 1 \end{cases}.$$

定理 D.15。设 $\overline{\psi}(a) = \psi$(最大$\{a, a^\circ\}$)。则以下三个条件等价 *t.*

1. $a^\circ < 1$; *2.* $\psi(a^\circ) < \psi(1)$，或者更明确地说，$\lambda^\circ < \kappa_{\overline{2}}$;

3. $\overline{\psi}'(1) > 0$，或者更明确地说，

$$\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)\mathbb{E}\left[\frac{\mathsf{T}^2}{(\mathsf{T} - 1)^2}\right] < 1. \tag{D.36}$$

证明。 注意， ψ 根据条目 4 和定理 D.8最终是递增的。另一方面，由于 a° 是最大临界点，因此 ψ 在 $[\underline{a^\circ}, \infty)$ 中递增。因此， $a^\circ < 1$、 $\psi(a^\circ) < \psi(1)$和 $\psi'(1) > 0$ 等价。

对于条目 2的显式形式，我们注意到当 $a^\circ < 1$时，根据 ψ的定义，有 $\psi(1) = \psi(1)$。接下来我们计算 $\psi(1)$的值。由于

$$\left\langle \frac{\boldsymbol{t}}{\boldsymbol{t}-1} \right\rangle = (1 \wedge \delta)\mathbb{E}\left[\frac{\mathsf{T}}{\mathsf{T}-1}\right] = 0,$$

我们有 $1 \in \mathcal{S}_{\boldsymbol{t}}$ ，因此 ω 在1处有一个极点。因此，由(D.15) ，我们有 $\psi(1) = \langle \boldsymbol{s} \rangle = \mathbb{E}[\Sigma_n] = \mathbb{E}\big[\Lambda_n^2\big] = m_{\underline{2}} = \kappa_{\underline{2}}$。

对于条目 3的显式形式，注意当 $a^\circ < 1$时，有 $\psi'(1) = \psi'(1)$。我们最后计算$\psi'(1)$ ，方法是取极限 $a \to 1$ ，见(D.17)。对于第一项，我们利用 $|\omega(a)| \to \infty$ 得到

$$\begin{aligned}
&\lim_{a\to 1} \left\langle \frac{\boldsymbol{t}}{(\boldsymbol{t}-a)^2} \right\rangle \left\langle \frac{\boldsymbol{s}}{(\boldsymbol{s}-\omega(a))^2} \right\rangle^{-1} \left\langle \frac{\boldsymbol{s}^2}{(\boldsymbol{s}-\omega(a))^2} \right\rangle \left\langle \frac{1}{\boldsymbol{t}-a} \right\rangle^{-1} \\
&= \left\langle \frac{\boldsymbol{t}}{(\boldsymbol{t}-1)^2} \right\rangle \left\langle \frac{1}{\boldsymbol{t}-1} \right\rangle^{-1} \lim_{a\to 1} \left\langle \frac{\boldsymbol{s}\omega(a)^2}{(\boldsymbol{s}-\omega(a))^2} \right\rangle^{-1} \left\langle \frac{\boldsymbol{s}^2\omega(a)^2}{(\boldsymbol{s}-\omega(a))^2} \right\rangle = \left\langle \frac{\boldsymbol{t}}{(\boldsymbol{t}-1)^2} \right\rangle \left\langle \frac{1}{\boldsymbol{t}-1} \right\rangle^{-1} \frac{\langle \boldsymbol{s}^2 \rangle}{\langle \boldsymbol{s} \rangle}.
\end{aligned}$$

对于第二项，我们有

$$\lim_{a\to 1} \left\langle \frac{\omega(a)\boldsymbol{s}}{\boldsymbol{s}-\omega(a)} \right\rangle \left\langle \frac{1}{\boldsymbol{t}-a} \right\rangle^{-2} \left\langle \frac{1}{(\boldsymbol{t}-a)^2} \right\rangle = -\langle \boldsymbol{s} \rangle \left\langle \frac{1}{\boldsymbol{t}-1} \right\rangle^{-2} \left\langle \frac{1}{(\boldsymbol{t}-1)^2} \right\rangle.$$

因此我们得出

$$\psi'(1) = \langle \boldsymbol{s} \rangle \left\langle \frac{1}{\boldsymbol{t}-1} \right\rangle^{-2} \left(\frac{\langle \boldsymbol{s}^2 \rangle}{\langle \boldsymbol{s} \rangle^2} \left\langle \frac{\boldsymbol{t}}{(\boldsymbol{t}-1)^2} \right\rangle \left\langle \frac{1}{\boldsymbol{t}-1} \right\rangle + \left\langle \frac{1}{(\boldsymbol{t}-1)^2} \right\rangle \right).$$

注意， $\langle \boldsymbol{t}/(\boldsymbol{t}-1) \rangle = 0$ 蕴含

$$\left\langle \frac{1}{\boldsymbol{t}-1} \right\rangle = -1, \qquad \left\langle \frac{1}{(\boldsymbol{t}-1)^2} \right\rangle = 1 + \left\langle \frac{\boldsymbol{t}}{(\boldsymbol{t}-1)^2} \right\rangle = 1 + \left\langle \frac{\boldsymbol{t}^2}{(\boldsymbol{t}-1)^2} \right\rangle,$$

由此，我们可以进一步将最后一个表达式简化为

$$\psi'(1) = \langle \boldsymbol{s} \rangle \left(\left(-\frac{\langle \boldsymbol{s}^2 \rangle}{\langle \boldsymbol{s} \rangle^2} + 1 \right) \left\langle \frac{\boldsymbol{t}^2}{(\boldsymbol{t}-1)^2} \right\rangle + 1 \right).$$

回忆 $\boldsymbol{s}$ 和 $\boldsymbol{t}$ 在(D.35)中的定义，我们计算得

$$\begin{aligned}
\psi'(1) &= \mathbb{E}[\Sigma_n] \left(\left(-\frac{\mathbb{E}\big[\Sigma_n^2\big]}{\mathbb{E}[\Sigma_n]^2} + 1 \right) \mathbb{E}\left[\frac{\mathsf{T}^2}{(\mathsf{T}-1)^2}\right] + 1 \right) \\
&= \overline{m}_2 \left(1 - \left(\frac{\overline{m}_4 - \overline{m}_2^2}{\overline{m}_2^2} \right) \mathbb{E}\left[\frac{\mathsf{T}^2}{(\mathsf{T}-1)^2}\right] \right) \\
&= \overline{\kappa}_2 \left(1 - \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta \right) \mathbb{E}\left[\frac{\mathsf{T}^2}{(\mathsf{T}-1)^2}\right] \right).
\end{aligned}$$

因此， $\psi'(1) > 0$ 等价于(D.36)。 □

E 定理B.3的证明

回顾 (C.1) 中的线性化 GVAMP 迭代

$$u^t = \left(\overline{\kappa}_2^{-1} X X^\top - I_n\right)\left(\widetilde{G} - I_n\right) u^{t-1}, \tag{E.1}$$

其中

$$\widetilde{G} \coloneqq \operatorname{diag}(\overline{g}(y)) + I_n. \tag{E.2}$$

采用该记号

$$\beta_t \coloneqq \left\|u^t\right\|_2, \qquad \widehat{u}^t \coloneqq \frac{u^t}{\beta_t}, \tag{E.3}$$

我们可以将(E.1) 写成

$$\frac{\beta_t}{\beta_{t-1}}\widehat{u}^t = \left(\overline{\kappa}_2^{-1} XX^\top - I_n\right)\left(\widetilde{G} - I_n\right)\widehat{u}^{t-1}. \tag{E.4}$$

引理 E.1. 设 β_t 定义于 (E.3)。下列极限成立：

$$\lim_{t\to\infty}\lim_{n\to\infty}\frac{\left\langle u^t, u^{t-1}\right\rangle}{\|u^{t-1}\|_2^2} = \lim_{t\to\infty}\lim_{n\to\infty}\frac{\beta_t}{\beta_{t-1}} = \gamma, \tag{E.5}$$

其中 γ 定义于 (4.2)。

关于定理 E.1 的证明推迟至第 E.1 节。

加上 $(\widetilde{G} - I_n)\widehat{u}^t$ 在 (E.4) 两边，得到

$$\left[\left(\frac{\beta_t}{\beta_{t-1}} - 1\right)I_n + \widetilde{G}\right]\widehat{u}^t = \overline{\kappa}_2^{-1} XX^\top\left(\widetilde{G} - I_n\right)\widehat{u}^{t-1} + \left(\widetilde{G} - I_n\right)(\widehat{u}^t - \widehat{u}^{t-1}).$$

继续之前，回顾定义 (E.2) 中关于 G的结论

$$\left(\frac{\beta_t}{\beta_{t-1}} - 1\right)I_n + \widetilde{G} = \frac{\beta_t}{\beta_{t-1}}I_n + \operatorname{diag}(\overline{g}(y)). \tag{E.6}$$

根据 (E.5) 中的第二个恒等式以及 g 在 (4.3) 中的第一个假设，上述矩阵以概率 1 在所有充分大的 n下均为严格正定矩阵。现在，将两边乘以 $X^\top\left(\widetilde{G} - I_n\right)\left[\left(\frac{\beta_t}{\beta_{t-1}} - 1\right)I_n + \widetilde{G}\right]^{-1}$：

$$\begin{aligned} X^\top\left(\widetilde{G} - I_n\right)\widehat{u}^t = \overline{\kappa}_2^{-1} X^\top\left(\widetilde{G} - I_n\right)\left[\left(\frac{\beta_t}{\beta_{t-1}} - 1\right)I_n + \widetilde{G}\right]^{-1} XX^\top\left(\widetilde{G} - I_n\right)\widehat{u}^{t-1} \\ + X^\top\left(\widetilde{G} - I_n\right)^2\left[\left(\frac{\beta_t}{\beta_{t-1}} - 1\right)I_n + \widetilde{G}\right]^{-1}(\widehat{u}^t - \widehat{u}^{t-1}). \end{aligned}$$

进一步定义

$$\widehat{v}^{t+1} \coloneqq X^\top\left(\widetilde{G} - I_n\right)\widehat{u}^t, \tag{E.7}$$

并将上述迭代改写为：

$$\begin{aligned} \widehat{v}^{t+1} &= \overline{\kappa}_2^{-1} X^\top\left(\widetilde{G} - I_n\right)\left[(\gamma-1)I_n + \widetilde{G}\right]^{-1} X\widehat{v}^t \\ &\quad + \underbrace{\overline{\kappa}_2^{-1} X^\top\left(\widetilde{G} - I_n\right)\left\{\left[\left(\frac{\beta_t}{\beta_{t-1}} - 1\right)I_n + \widetilde{G}\right]^{-1} - \left[(\gamma-1)I_n + \widetilde{G}\right]^{-1}\right\} X\widehat{v}^t}_{\widehat{e}_2^t} \qquad \text{(E.8)} \\ &\quad + \underbrace{X^\top\left(\widetilde{G} - I_n\right)^2\left[\left(\frac{\beta_t}{\beta_{t-1}} - 1\right)I_n + \widetilde{G}\right]^{-1}(\widehat{u}^t - \widehat{u}^{t-1})}_{\widehat{e}_1^t}. \qquad \text{(E.9)} \end{aligned}$$

回顾定义 $\widetilde{D}$ 的定义 D 在 (4.1a)中的形式 $\mathcal{T}$ 在 (4.2) 以及记号 G在 (E.2)中，我们认出

$$\overline{\kappa}_2^{-1} X^\top \left(\widetilde{G} - I_n\right) \left[(\gamma-1) I_n + \widetilde{G}\right]^{-1} X = \overline{\kappa}_2^{-1} D.$$

于是我们得到

$$\widehat{v}^{t+1} = \overline{\kappa}_2^{-1} D \widehat{v}^t + \widehat{e}_2^t + \widehat{e}_1^t.$$

我们将 $\kappa_2^{-1} D$ 的谱向右平移，使所得矩阵成为正定的（对所有充分大的 n，这一结论以概率 1 成立）。取 $\ell > 0$ 为一个充分大的常数，使得对所有充分大的 d， $\mathbb{P}\big(\lambda_d(\kappa_2^{-1} D + \ell I_d) > 0\big) = 1$。于是，迭代可以改写为

$$\widehat{v}^{t+1} = \underbrace{\frac{\overline{\kappa}_2^{-1} D + \ell I_d}{1+\ell}}_{\widetilde{D}} \widehat{v}^t + \underbrace{\frac{\ell}{1+\ell}\big(\widehat{v}^{t+1} - \widehat{v}^t\big) + \frac{1}{1+\ell}\big(\widehat{e}_2^t + \widehat{e}_1^t\big)}_{\widehat{e}_3^t}. \tag{E.10}$$

注意到 $\widetilde{D}$ 和 D 具有相同的特征向量。将上述迭代展开 s 步，可得

$$\widehat{v}^{t+s} = \widetilde{D}^s \widehat{v}^t + \widehat{e}^{t,s}, \tag{E.11}$$

其中

$$\widehat{e}^{t,s} := \sum_{i=0}^{s-1} \widetilde{D}^{s-1-i} \widehat{e}_3^{t+i}. \tag{E.12}$$

让我们考察等式两边平方范数的极限(E.11)。对于左边，我们有

$$\begin{aligned}
\lim_{s\to\infty}\lim_{t\to\infty}\lim_{d\to\infty} \big\|\widehat{v}^{t+s}\big\|_2^2 &= \lim_{t\to\infty}\lim_{d\to\infty} \big\|\widehat{v}^{t+1}\big\|_2^2 = \lim_{t\to\infty}\lim_{d\to\infty} \frac{\big\|X^\top\big(\widetilde{G} - I_n\big)u^t\big\|_2^2}{\|u^t\|_2^2} \overset{(a)}{=} \lim_{t\to\infty}\lim_{d\to\infty} \frac{\overline{\kappa}_2^2 \|v^{t+1}\|_2^2}{\|u^t\|_2^2} \\
&\overset{(b)}{=} \lim_{t\to\infty} \frac{\overline{\kappa}_2^2\big(\chi_{t+1}^2\rho + \omega_{t+1}^2\big)}{\nu_t^2\sigma^2 + \tau_t^2} = \lim_{t\to\infty} \frac{\chi_{t+1}^2}{\nu_t^2} \frac{\overline{\kappa}_2^2\Big(\rho + \frac{\omega_{t+1}^2}{\chi_{t+1}^2}\Big)}{\sigma^2 + \frac{\tau_t^2}{\nu_t^2}} \overset{(c)}{=} w_4^2 \frac{\overline{\kappa}_2^2(\rho + w_2)}{\sigma^2 + w_1} \in (0,\infty),
\end{aligned} \tag{E.13}$$

其中，等式(a) 是通过回顾线性化 GVAMP 动力学中 v^t 的更新规则(C.1)得到的；(b) 由定理 C.1 和C.2得出；而(c) 则来自 (B.5)。

接下来我们转向 (E.11) 的右边。下面的引理（其证明推迟到E.2 节）表明，误差项 $\widehat{e}^{t,s}$ 渐近可忽略。

引理 E.2。 误差向量 $\widehat{e}^{t,s}$ 定义于(E.12) 满足

$$\lim_{s\to\infty}\lim_{t\to\infty}\lim_{n\to\infty} \big\|\widehat{e}^{t,s}\big\|_2^2 = 0.$$

根据定理 E.2 以及 D的谱分解，我们有

$$\begin{aligned}
&\lim_{s\to\infty}\lim_{t\to\infty}\lim_{d\to\infty} \Big\|\widetilde{D}^s\widehat{v}^t + \widehat{e}^{t,s}\Big\|_2^2 = \lim_{s\to\infty}\lim_{t\to\infty}\lim_{d\to\infty} \Big\|\widetilde{D}^s\widehat{v}^t\Big\|_2^2 \\
&= \lim_{s\to\infty}\lim_{t\to\infty}\lim_{d\to\infty} \lambda_1(\widetilde{D})^{2s}\big\langle v_1(D), \widehat{v}^t\big\rangle^2 + \sum_{i=2}^{d} \lambda_i(\widetilde{D})^{2s}\big\langle v_i(D), \widehat{v}^t\big\rangle^2 \\
&= \lim_{s\to\infty}\lim_{t\to\infty}\lim_{d\to\infty} \Big\|\widetilde{D}^s\Pi\widehat{v}^t\Big\|_2^2 + \Big\|\widetilde{D}^s\Pi^\perp\widehat{v}^t\Big\|_2^2,
\end{aligned} \tag{E.14}$$

其中

$$\Pi \coloneqq v_1(D)v_1(D)^\top, \qquad \Pi^\perp \coloneqq I_d - \Pi.$$

第二项(E.14) 可以上界为

$$\left\|\widetilde{D}^s\Pi^\perp\widehat{v}^t\right\|_2^2 = \left\|\left(\widetilde{D}\Pi^\perp\right)^s\widehat{v}^t\right\|_2^2 \le \|\widehat{v}^t\|_2^2\sigma_1\left(\left(\widetilde{D}\Pi^\perp\right)^s\right)^2 = \|\widehat{v}^t\|_2^2\lambda_1\left(\widetilde{D}\Pi^\perp\right)^{2s} = \|\widehat{v}^t\|_2^2\lambda_2(\widetilde{D})^{2s}, \tag{E.15}$$

其中，在倒数第二个等号处用到了 $\widetilde{D}\Pi^\perp$ 的正定性。

根据定理 B.2,

$$\lim_{d\to\infty}\lambda_2(D) = \lambda^\circ < \overline{\kappa}_2,$$

~其中，严格不等式在假设(4.13)成立时成立。根据 D 在(E.10)中的定义，这意味着

$$\lim_{d\to\infty}\lambda_2(\widetilde{D}) = \frac{\overline{\kappa}_2^{-1}\lambda^\circ + \ell}{1+\ell} < 1.$$

回到(E.15)，利用上述界以及$\lim_{t\to\infty}\lim_{d\to\infty}\|\widehat{v}^t\|_2$ 在(E.13)中所示的有界性，可得

$$\lim_{s\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}\left\|\widetilde{D}^s\Pi^\perp\widehat{v}^t\right\|_2^2 \le \left(\lim_{t\to\infty}\lim_{d\to\infty}\|\widehat{v}^t\|_2^2\right)\left(\lim_{s\to\infty}\lim_{d\to\infty}\lambda_2(\widetilde{D})^{2s}\right) = 0. \tag{E.16}$$

将(E.16) 代入(E.14)可得

$$\lim_{s\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}\left\|\widetilde{D}^s\widehat{v}^t + \widehat{e}^{t,s}\right\|_2^2 = \lim_{s\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}\lambda_1(\widetilde{D})^{2s}\left\langle v_1(D),\widehat{v}^t\right\rangle^2. \tag{E.17}$$

最后，结合 (E.13)和(E.17)可得

$$1 = \frac{\lim_{s\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}\lambda_1(\widetilde{D})^{2s}\left\langle v_1(D),\widehat{v}^t\right\rangle^2}{\lim_{t\to\infty}\lim_{d\to\infty}\left\|\widehat{v}^t\right\|_2^2},$$

这意味着

$$\lim_{d\to\infty}\lambda_1(\widetilde{D}) = 1, \qquad \lim_{t\to\infty}\lim_{d\to\infty}\frac{\left\langle v_1(D),\widehat{v}^t\right\rangle^2}{\|\widehat{v}^t\|_2^2} = 1.$$

Rec ~代入 D 于 (E.10) 中的定义，并由 (E.7) 可知 $\widehat{v}^t = \kappa_2 v^t$，我们得出结论 e:

$$\lim_{d\to\infty}\lambda_1(D) = \overline{\kappa}_2, \qquad \lim_{t\to\infty}\lim_{d\to\infty}\frac{\left\langle v_1(D),v^t\right\rangle^2}{\|v^t\|_2^2} = 1,$$

正如定理 B.3中所述。

E.1 定理E.1的证明

回顾定理 C.2中的状态演化表示，令

$$n_t \coloneqq \mathbb{E}[\mathsf{N}_t\mathsf{N}_{t-1}] = \frac{1}{\tau_t\tau_{t-1}}\left(\Sigma_{t+1,t} - \nu_t\nu_{t-1}\sigma^2\right). \tag{E.18}$$

我们首先证明

$$\lim_{t\to\infty} n_t = 1. \tag{E.19}$$

为此，我们首先推导 n_t的递归。根据状态演化递推式(C.9) 和(C.10),

$$\begin{aligned}
\Sigma_{t+1,t} &= \overline{\kappa}_2(\chi_t\chi_{t-1}\rho + \Omega_{t,t-1}) + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\rho\mathbb{E}\Big[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big](\nu_{t-1}\chi_{t-1} + \nu_{t-2}\chi_t) \\
&\quad + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\big[\overline{g}(\mathsf{Y})^2\big]\mathbb{E}[\tau_{t-1}\mathsf{N}_{t-1}\cdot\tau_{t-2}\mathsf{N}_{t-2}] + \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2}\mathbb{E}\Big[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\Big] + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}\mathbb{E}\Big[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2\right)\rho\nu_{t-1}\nu_{t-2}, \\
\Omega_{t+1,t} &= \delta\left\{\rho\left(\mathbb{E}\Big[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\Big] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\Big[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2\right)\nu_t\nu_{t-1} + \frac{1}{\overline{\kappa}_2}\mathbb{E}\big[\overline{g}(\mathsf{Y})^2\big]\mathbb{E}[\tau_t\mathsf{N}_t\cdot\tau_{t-1}\mathsf{N}_{t-1}]\right\}.
\end{aligned}$$

利用(E.18) 中 n_t的定义，我们得到

$$\begin{aligned}
n_t &= \frac{1}{\tau_t\tau_{t-1}}\left\{\overline{\kappa}_2(\chi_t\chi_{t-1}\rho + \Omega_{t,t-1}) + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\rho\mathbb{E}\Big[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big](\nu_{t-1}\chi_{t-1} + \nu_{t-2}\chi_t)\right. \\
&\quad \left. + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\big[\overline{g}(\mathsf{Y})^2\big]\tau_{t-1}\tau_{t-2}n_{t-1} + \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2}\mathbb{E}\Big[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\Big] + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}\mathbb{E}\Big[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2\right)\rho\nu_{t-1}\nu_{t-2} - \sigma^2\nu_{t-1}\nu_t\right\}, \\
\Omega_{t+1,t} &= \delta\left\{\rho\left(\mathbb{E}\Big[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\Big] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\Big[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2\right)\nu_t\nu_{t-1} + \frac{1}{\overline{\kappa}_2}\mathbb{E}\big[\overline{g}(\mathsf{Y})^2\big]\tau_t\tau_{t-1}n_t\right\}.
\end{aligned}$$

消去 $\Omega_{t,\ t-1}$，得到

$$\begin{aligned}
n_t &= \frac{1}{\tau_t\tau_{t-1}}\left\{\overline{\kappa}_2\chi_t\chi_{t-1}\rho + \delta\overline{\kappa}_2\rho\left(\mathbb{E}\Big[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\Big] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\Big[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2\right)\nu_{t-1}\nu_{t-2} + \delta\mathbb{E}\big[\overline{g}(\mathsf{Y})^2\big]\tau_{t-1}\tau_{t-2}n_{t-1}\right. \\
&\quad + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\rho\mathbb{E}\Big[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big](\nu_{t-1}\chi_{t-1} + \nu_{t-2}\chi_t) + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\big[\overline{g}(\mathsf{Y})^2\big]\tau_{t-1}\tau_{t-2}n_{t-1} \\
&\quad \left. + \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2}\mathbb{E}\Big[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\Big] + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}\mathbb{E}\Big[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2\right)\rho\nu_{t-1}\nu_{t-2} - \sigma^2\nu_{t-1}\nu_t\right\} \\
&= \frac{1}{\tau_t\tau_{t-1}}\left\{\overline{\kappa}_2\chi_t\chi_{t-1}\rho + \rho\left(\left(\delta\overline{\kappa}_2 + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\right)\mathbb{E}\Big[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\Big] + \left(\delta\frac{\overline{\kappa}_4}{\overline{\kappa}_2} + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}\right)\mathbb{E}\Big[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2\right)\nu_{t-1}\nu_{t-2}\right. \\
&\quad + \left(\delta + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\right)\mathbb{E}\big[\overline{g}(\mathsf{Y})^2\big]\tau_{t-1}\tau_{t-2}n_{t-1} + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\rho\mathbb{E}\Big[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big](\nu_{t-1}\chi_{t-1} + \nu_{t-2}\chi_t) - \sigma^2\nu_{t-1}\nu_t\Big\}.
\end{aligned} \tag{E.20}$$

回顾 w_1, w_2, w_3, w_4, 以及 γ 由 定理 B.1 , 并注意到 $w_3w_4 = \gamma$, 对于每个充分大的 t, 我们有以下恒等式:

$$\begin{aligned}
\frac{\chi_{t-1}\chi_t}{\tau_{t-1}\tau_t} &= \frac{\chi_{t-1}\chi_t}{\nu_{t-1}\nu_t}\frac{\nu_{t-1}\nu_t}{\tau_{t-1}\tau_t} = w_3^{-2}w_1^{-1}, \\
\frac{\nu_{t-2}\nu_{t-1}}{\tau_{t-1}\tau_t} &= \frac{\nu_{t-2}\nu_{t-1}}{\nu_{t-1}\nu_t}\frac{\nu_{t-1}\nu_t}{\tau_{t-1}\tau_t} = \gamma^{-2}w_1^{-1}, \\
\frac{\tau_{t-2}}{\tau_t} &= \frac{\tau_{t-2}}{\nu_{t-2}}\frac{\nu_{t-2}}{\nu_{t-1}}\frac{\nu_{t-1}}{\nu_t}\frac{\nu_t}{\tau_t} = \gamma^{-2}, \\
\frac{\chi_{t-1}\nu_{t-1}}{\tau_{t-1}\tau_t} &= \frac{\chi_{t-1}}{\nu_{t-1}}\frac{\nu_{t-1}}{\tau_{t-1}}\frac{\nu_{t-1}}{\nu_t}\frac{\nu_t}{\tau_t} = w_3^{-1}w_1^{-1}\gamma^{-1}, \\
\frac{\chi_t\nu_{t-2}}{\tau_{t-1}\tau_t} &= \frac{\chi_t}{\nu_{t-1}}\frac{\nu_{t-1}}{\tau_{t-1}}\frac{\nu_{t-2}}{\nu_{t-1}}\frac{\nu_{t-1}}{\nu_t}\frac{\nu_t}{\tau_t} = w_4w_1^{-1}\gamma^{-2}, \\
\frac{\nu_{t-1}\nu_t}{\tau_{t-1}\tau_t} &= w_1^{-1}.
\end{aligned}$$

将 (E.20) 两边除以 $\tau_t\tau_{t-1}$ 并使用上述恒等式，可得

$$\begin{aligned} n_t &= \rho\overline{\kappa}_2 w_3^{-2} w_1^{-1} + \rho\left\{\left(\delta\overline{\kappa}_2 + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\right] + \left(\delta\frac{\overline{\kappa}_4}{\overline{\kappa}_2} + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2\right\}\gamma^{-2}w_1^{-1} \\ &\quad + \left(\delta + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\right]\gamma^{-2}n_{t-1} + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\rho\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]\left(w_4 w_1^{-1}\gamma^{-2} + w_3^{-1}w_1^{-1}\gamma^{-1}\right) - \sigma^2 w_1^{-1}, \\ &= An_{t-1} + B, \end{aligned} \tag{E.21}$$

其中

$$\begin{aligned} A &:= \left(\delta + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\right]\gamma^{-2} = \frac{\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\right]}{\left(\delta + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2}, \\ B &:= \rho\overline{\kappa}_2 w_3^{-2} w_1^{-1} + \rho\left\{\left(\delta\overline{\kappa}_2 + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\right] + \left(\delta\frac{\overline{\kappa}_4}{\overline{\kappa}_2} + \frac{\overline{\kappa}_6}{\overline{\kappa}_2^2}\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2\right\}\gamma^{-2}w_1^{-1} \\ &\quad + \frac{\overline{\kappa}_4}{\overline{\kappa}_2}\rho\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]\left(w_4 w_1^{-1}\gamma^{-2} + w_3^{-1}w_1^{-1}\gamma^{-1}\right) - \sigma^2 w_1^{-1}. \end{aligned}$$

注意，由于临界条件 $A \in (0,1)$ 满足(4.13) 。

回顾一下，我们希望证明(E.19)。我们将证明：

$$1 \le \underline{n} := \liminf_{t\to\infty} n_t \le \limsup_{t\to\infty} n_t =: \overline{n} \le 1.$$

为证明最右侧不等式，我们对limsup 两边取(E.21) 的limsup，并应用其次可加性：$n \le An + B$，从而得到 $\overline{n} \le \overline{B}/(1-A)$。利用 $\overline{A}$、B的定义，经过冗长计算可验证 $B/(1-A) = 1$。因此 $n \le 1$。类似地，利用 liminf 的超可加性，可以证明最左侧不等式 $n \ge 1$。因此(E.19) 成立。

现在考虑(E.5)中的第一个极限。利用(E.18) 对 n_t的定义、(E.19)中的极限以及定理 B.1，我们有

$$\lim_{t\to\infty}\lim_{n\to\infty}\frac{\langle u^t, u^{t-1}\rangle}{\|u^{t-1}\|_2^2} = \lim_{t\to\infty}\frac{\Sigma_{t+1,t}}{\Sigma_{t,t}} = \lim_{t\to\infty}\frac{\tau_t\tau_{t-1}n_t + \nu_t\nu_{t-1}\sigma^2}{\tau_{t-1}^2 + \nu_{t-1}^2\sigma^2} = \lim_{t\to\infty}\frac{\frac{\tau_t\tau_{t-1}}{\nu_t\nu_{t-1}}n_t + \sigma^2}{\frac{\nu_{t-1}}{\nu_t}\frac{\tau_{t-1}^2}{\nu_{t-1}^2} + \frac{\nu_{t-1}}{\nu_t}\sigma^2} = \frac{w_1 + \sigma^2}{\gamma^{-1}w_1 + \gamma^{-1}\sigma^2} = \gamma,$$

如所述。

转向 (E.5)中的第二个极限，再次使用 定理 B.1，我们得到

$$\lim_{t\to\infty}\lim_{n\to\infty}\frac{\|u^t\|_2^2}{\|u^{t-1}\|_2^2} = \lim_{t\to\infty}\frac{\nu_t^2\sigma^2 + \tau_t^2}{\nu_{t-1}^2\sigma^2 + \tau_{t-1}^2} = \lim_{t\to\infty}\frac{\nu_t^2}{\nu_{t-1}^2}\frac{\sigma^2 + \frac{\tau_t^2}{\nu_t^2}}{\sigma^2 + \frac{\tau_{t-1}^2}{\nu_{t-1}^2}} = \gamma^2\frac{\sigma^2 + w_1}{\sigma^2 + w_1} = \gamma^2,$$

由 γ 的正性（见 假设 (A5)）可知

$$\lim_{t\to\infty}\lim_{n\to\infty}\frac{\|u^t\|_2}{\|u^{t-1}\|_2} = \gamma,$$

正如 (E.5) 第二个等式中所述。

E.2 定理E.2 的证明

首先，根据定义(E.8)，我们可以将 $\widehat{e}_2^t$ 写作

$$\widehat{e}_2^t = \overline{\kappa}_2^{-1}\left(\gamma - \frac{\beta_t}{\beta_{t-1}}\right)X^\top\left(\widetilde{G} - I_n\right)\left[\left(\frac{\beta_t}{\beta_{t-1}} - 1\right)I_n + \widetilde{G}\right]^{-1}\left[(\gamma-1)I_n + \widetilde{G}\right]^{-1}X\widehat{v}^t.$$

于是我们立即得到

$$\lim_{t\to\infty}\lim_{d\to\infty}\left\|\widehat{e}_2^t\right\|_2 = 0, \tag{E.22}$$

因为根据 定理 E.1,

$$\lim_{t\to\infty}\lim_{d\to\infty}\frac{\beta_t}{\beta_{t-1}} = \gamma,$$

并且根据(E.13) 以及 (E.6) 的可逆性,

$$\lim_{t\to\infty}\lim_{d\to\infty}\left\|\overline{\kappa}_2^{-1}X^\top\left(\widetilde{G}-I_n\right)\left\{\left[\left(\frac{\beta_t}{\beta_{t-1}}-1\right)I_n+\widetilde{G}\right]^{-1}\left[(\gamma-1)I_n+\widetilde{G}\right]^{-1}\right\}X\widehat{v}^t\right\|_2 < \infty.$$

接下来，我们断言

$$\lim_{t\to\infty}\lim_{d\to\infty}\left\|\widehat{u}^t-\widehat{u}^{t-1}\right\|_2^2 = 0. \tag{E.23}$$

展开左侧的平方项，可得

$$\lim_{t\to\infty}\lim_{d\to\infty}\left\|\widehat{u}^t-\widehat{u}^{t-1}\right\|_2^2 = 2-2\lim_{t\to\infty}\lim_{d\to\infty}\frac{\left\langle u^t,u^{t-1}\right\rangle}{\|u^t\|_2\|u^{t-1}\|_2}, \tag{E.24}$$

因为 $\widehat{u}^t\in\mathbb{S}^{n-1}$ 根据定义 (E.3)。利用 定理 E.1，可得

$$\lim_{t\to\infty}\lim_{d\to\infty}\frac{\left\langle u^t,u^{t-1}\right\rangle}{\|u^t\|_2\|u^{t-1}\|_2} = \lim_{t\to\infty}\lim_{d\to\infty}\frac{\beta_{t-1}}{\beta_t}\frac{\left\langle u^t,u^{t-1}\right\rangle}{\|u^{t-1}\|_2^2} = 1,$$

这意味着(E.23)。

根据 $\widehat{e}_1^t$ 在 (E.9) 和 $\widehat{v}^t$ 在 (E.7)的定义，我们有

$$\lim_{t\to\infty}\lim_{d\to\infty}\left\|\widehat{e}_1^t\right\|_2 = 0, \qquad \lim_{t\to\infty}\lim_{d\to\infty}\left\|\widehat{v}^{t+1}-\widehat{v}^t\right\|_2 = 0. \tag{E.25}$$

回顾 $\widehat{e}_3^t$ 在 (E.10)中的定义，根据 (E.22) 和 (E.25)，于是有

$$\lim_{t\to\infty}\lim_{d\to\infty}\left\|\widehat{e}_3^t\right\|_2 = 0. \tag{E.26}$$

最后，考虑 $\widehat{e}^{t,\,s}$ 定义于 (E.12)。利用 (E.26) 和三角不等式,

$$\lim_{s\to\infty}\lim_{t\to\infty}\lim_{d\to\infty}\left\|\widehat{e}^{t,s}\right\|_2 = 0.$$

F 定理 4.2 的证明

证明条目 2。考虑任意一个由函数 g 确定的谱估计量，该函数满足(4.3)，且临界条件(4.13) 成立。根据(4.3) 和柯西–施瓦茨不等式，我们有

$$\begin{aligned}
\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right] &= \mathbb{E}\left[\overline{g}(\mathsf{Y})\left(\overline{\mathsf{Z}}^2-1\right)\right] \\
&= \iint p_{\mathsf{Z}}(z)Q(y\,|\,z)\overline{g}(y)\left(\sigma^{-2}z^2-1\right)\mathrm{d}z\,\mathrm{d}y \\
&= \int\frac{\mathbb{E}\left[Q(y\,|\,\mathsf{Z})\left(\overline{\mathsf{Z}}^2-1\right)\right]}{\sqrt{\mathbb{E}[Q(y\,|\,\mathsf{Z})]}}\overline{g}(y)\sqrt{\mathbb{E}[Q(y\,|\,\mathsf{Z})]}\,\mathrm{d}y
\end{aligned}$$

$$\leq \left(\int \frac{\mathbb{E}\Big[Q(y \mid \mathsf{Z})\Big(\overline{\mathsf{Z}}^2 - 1\Big)\Big]^2}{\mathbb{E}[Q(y \mid \mathsf{Z})]} \, \mathrm{d}y \right)^{1/2} \left(\int \overline{g}(y)^2 \mathbb{E}[Q(y \mid \mathsf{Z})] \, \mathrm{d}y \right)^{1/2}$$
$$= \left(\int \frac{\mathbb{E}\Big[Q(y \mid \mathsf{Z})\Big(\overline{\mathsf{Z}}^2 - 1\Big)\Big]^2}{\mathbb{E}[Q(y \mid \mathsf{Z})]} \, \mathrm{d}y \right)^{1/2} \mathbb{E}\big[\overline{g}(\mathsf{Y})^2\big]^{1/2}.$$

将其代入(4.13)，可推出以下不等式：

$$1 < \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta \right) \frac{\mathbb{E}\Big[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big]^2}{\mathbb{E}[\overline{g}(\mathsf{Y})^2]} \leq \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta \right) \int \frac{\mathbb{E}\Big[Q(y \mid \mathsf{Z})\Big(\overline{\mathsf{Z}}^2 - 1\Big)\Big]^2}{\mathbb{E}[Q(y \mid \mathsf{Z})]} \, \mathrm{d}y,$$

这等价于(4.15)，如所需。

证明 **条目 1的证明**。假设(4.15)成立，并考虑由 g 给出的 (4.16)。我们可以计算 ：

$$\begin{aligned}
\mathbb{E}\Big[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big] &= \iint p_{\mathsf{Z}}(z) Q(y \mid z) \frac{\int p_{\mathsf{Z}}(z')\Big(\frac{z'^2}{\sigma^2} - 1\Big) Q(y \mid z') \, \mathrm{d}z'}{\int p_{\mathsf{Z}}(z'') Q(y \mid z'') \, \mathrm{d}z''} \frac{z^2}{\sigma^2} \, \mathrm{d}y \, \mathrm{d}z \\
&= \int \frac{\int p_{\mathsf{Z}}(z')\Big(\frac{z'^2}{\sigma^2} - 1\Big) Q(y \mid z') \, \mathrm{d}z'}{\int p_{\mathsf{Z}}(z'') Q(y \mid z'') \, \mathrm{d}z''} \left[\int p_{\mathsf{Z}}(z) Q(y \mid z) \frac{z^2}{\sigma^2} \, \mathrm{d}z \right] \mathrm{d}y, \\
&= \int \frac{\int p_{\mathsf{Z}}(z')\Big(\frac{z'^2}{\sigma^2} - 1\Big) Q(y \mid z') \, \mathrm{d}z'}{\int p_{\mathsf{Z}}(z'') Q(y \mid z'') \, \mathrm{d}z''} \left[\int p_{\mathsf{Z}}(z) Q(y \mid z) \left(\frac{z^2}{\sigma^2} - 1 \right) \mathrm{d}z + \int p_{\mathsf{Z}}(z) Q(y \mid z) \, \mathrm{d}z \right] \mathrm{d}y, \\
&= \int \frac{\Big[\int p_{\mathsf{Z}}(z')\Big(\frac{z'^2}{\sigma^2} - 1\Big) Q(y \mid z') \, \mathrm{d}z'\Big]^2}{\int p_{\mathsf{Z}}(z'') Q(y \mid z'') \, \mathrm{d}z''} \mathrm{d}y + \iint p_{\mathsf{Z}}(z') \left(\frac{z'^2}{\sigma^2} - 1 \right) Q(y \mid z') \, \mathrm{d}z' \, \mathrm{d}y \\
&= \int \frac{\mathbb{E}\Big[\Big(\overline{\mathsf{Z}}^2 - 1\Big) Q(y \mid \mathsf{Z})\Big]^2}{\mathbb{E}[Q(y \mid \mathsf{Z})]} \, \mathrm{d}y, \qquad \text{(F.1)}
\end{aligned}$$

其中 (F.1) 成立是因为 $\mathbb{E}\big[\mathsf{Z}^2\big] = \sigma^2$，因此前一步中的第二项为零。我们还有：

$$\begin{aligned}
\mathbb{E}\big[\overline{g}(\mathsf{Y})^2\big] &= \int p_{\mathsf{Y}}(y) \frac{\Big[\int p_{\mathsf{Z}}(z')\Big(\frac{z'^2}{\sigma^2} - 1\Big) Q(y \mid z') \, \mathrm{d}z'\Big]^2}{\Big[\int p_{\mathsf{Z}}(z'') Q(y \mid z'') \, \mathrm{d}z''\Big]^2} \, \mathrm{d}y \\
&= \int \frac{\Big[\int p_{\mathsf{Z}}(z')\Big(\frac{z'^2}{\sigma^2} - 1\Big) Q(y \mid z') \, \mathrm{d}z'\Big]^2}{\int p_{\mathsf{Z}}(z'') Q(y \mid z'') \, \mathrm{d}z''} \, \mathrm{d}y = \int \frac{\mathbb{E}\Big[\Big(\overline{\mathsf{Z}}^2 - 1\Big) Q(y \mid \mathsf{Z})\Big]^2}{\mathbb{E}[Q(y \mid \mathsf{Z})]} \, \mathrm{d}y. \qquad \text{(F.2)}
\end{aligned}$$

根据上述表达式，条件 (4.15) 等价于 (4.13)。

接下来，我们验证对于 (4.3) 和(4.4) 中 g 取 (4.16)时结论成立。(4.3)中的中间恒等式由条件期望的塔式性质可得：

$$\mathbb{E}[\overline{g}(\mathsf{Y})] = \mathbb{E}\Big[\mathbb{E}\Big[\overline{\mathsf{Z}}^2 - 1 \,\Big|\, \mathsf{Y}\Big]\Big] = \mathbb{E}\Big[\overline{\mathsf{Z}}^2\Big] - 1 = 0.$$

对于(4.3)中的第三个不等式，根据(F.1) 可知

$$\mathbb{E}\Big[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\overline{\mathsf{Z}}^2 - 1 \,\Big|\, \mathsf{Y}\Big]^2\Big],$$

根据假设(4.17)，其为正。

对于(4.3)中的第一个不等式，注意到 $g(y) = \mathbb{E}\left[Z^2 \,\middle|\, Y = y\right] - 1 \geq -1$ 对任意 y 均成立。由于 $\mathbb{E}\left[g(Y)Z^2\right] = \mathbb{E}\left[g(Y)^2\right]$ 根据 (F.1) 和(F.2)，因此有

$$\overline{g}(y) \geq -\frac{\mathbb{E}\left[\overline{g}(Y)^2\right]}{\mathbb{E}\left[\overline{g}(Y)\overline{Z}^2\right]} > -\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)\mathbb{E}\left[\overline{g}(Y)\overline{Z}^2\right] = -\gamma,$$

其中严格不等式由 (4.13) 得出。

最后，我们转向 (4.4)。我们声称 supp μ_T 是有限的。事实上，对于任意 $y \in \text{supp } Y$，都有 $\mathcal{T}(y) = 1 - g(y)^2 + \gamma \leq 1$，因为第二项为正。如果 T 在 supp μ_T 处有一个点质量，则 (4.4) 成立。否则，我们可以通过在 supp μ_T 处添加一个小的点质量来扰动 T，同时保留假设 (A5) 要求的其他性质。只要扰动足够小，扰动后的预处理函数仍满足 (4.13)，因此可实现正的渐近重叠。这类论证已在多篇先前的研究中使用（例如参见 [71, 定理 2] 的证明或 [107, 第 C 节]），这里省略细节。

G 定理 4.3 的证明

谱初始化 GVAMP 的状态演化结果（定理 4.3）的证明分为 3个主要步骤。首先，下面的 定理 G.1 给出了采用与数据 (X, y)相独立的"预言机"初始化的 GVAMP 的状态演化结果。随后，定理 G.2放宽了该结果所需的一些技术性假设。请注意，上述两个定理都不涵盖与 (X, y)相关的谱初始化。第 G.3 节通过使用两阶段 GVAMP 算法模拟谱初始化 GVAMP，解决了这一问题。

在完成这项工作期间，我们注意到近期有一项独立研究 [28]，其附录 B 给出了一般正交近似消息传递（OAMP）算法的状态演化结果。一般 OAMP 的形式与 (4.19)类似，不同之处在于它不假设任何底层统计模型，例如 (3.4)。另一方面，[28] 中的一般 OAMP 允许向量去噪器 f_t, g_t 依赖于截至时刻 t 的所有迭代量（但须满足相对于所有自变量的散度为零条件）。鉴于这种灵活性，我们承认可以通过将 GVAMP 化约为一般 OAMP 来证明 定理 G.1的可能性。然而，处理定理 4.3所涉及的谱初始化，仍需采用 [28] 中针对尖峰矩阵模型的技术之外的方法。因此，我们选择从基本原理出发进行证明。此外，将结果扩展到允许向量去噪器具有类似 [28]中的记忆是直接的。由于这类扩展对于第 4.3 节并非必要，为使记号更简洁，我们不再探讨这些扩展。

让我们从定理 G.1开始，该定理需要以下列出的附加假设。

(A25) 对于每个 $t \geq 0$，$\max\left\{\mathbb{E}\left[\Phi_t(\Lambda_d^2)^2\right], \text{Var}\left[\widetilde{\Phi}_t(\Lambda_d)\right]\right\} > 0$ 以及 $\max\left\{\mathbb{E}\left[\Psi_t(\Lambda_n^2)^2\right], \text{Var}\left[\widetilde{\Psi}_t(\Lambda_n)\right]\right\} > 0$均成立。

(A26) R_0 几乎必然不等于 $\pm\widetilde{B}_*$，且 P_0 几乎必然不等于 $\pm\widetilde{Z}$。对于每个 $t \geq 0$，不存在常数 $\beta_0, \cdots, \beta_t, \alpha_0, \cdots, \alpha_t \in \mathbb{R}$ 使得

$$\text{either} \quad \widetilde{R}_{t+1} - \widetilde{b}_{t+1}\widetilde{B}_* = \sum_{s=0}^{t} \beta_s\left(\widetilde{R}_s - \widetilde{b}_s\widetilde{B}_*\right) \quad \text{or} \quad \widetilde{P}_{t+1} - \widetilde{a}_{t+1}\widetilde{Z} = \sum_{s=0}^{t} \alpha_s\left(\widetilde{P}_s - \widetilde{a}_s\widetilde{Z}\right)$$

几乎必然成立。

(A27) 对于每个 $t \geq 1$，函数 $f_t \in \text{PG}^{(2+d')\to 1}$ 关于第一个自变量弱可微，关于其他自变量连续。$r \mapsto f_t(r; b, \theta)$ 的弱导数记为 $f_t'(r; b, \theta)$，它也属于 $\text{PG}^{(2+d')\to 1}$，且 $f_t'(R_t; B_*, \Theta)$ 以概率 1 连续。

对于每个 $t \geq 1$，函数 $g_t \in \mathrm{PG}_{(3+n')\to 1}$ 关于第一个自变量弱可微，关于其他自变量连续。$p \mapsto g_t(p;\, z, e, \xi)$ 的弱导数记为 $g_t'(p;\, z, e, \xi)$，它也属于 $\mathrm{PG}_{(3+n')\to 1}$，且 $g_t'(\mathsf{P}_t;\, \mathsf{Z}, \mathsf{E}\ \Xi)$ 以概率 1 连续。

初始化量和侧信息须满足以下假设。

(A28) Θ、Ξ 彼此独立，且独立于 O、Q（来自假设（A2））。初始化量 $\tilde{r}^0$、$\tilde{p}^0$ 取如下形式

$$\tilde{r}^0 = f_0(\beta_*, \Theta), \qquad \tilde{p}^0 = g_0(z, \varepsilon, \Xi)$$

对某些连续函数 $f_0 \in \mathrm{PG}_{(1+d')\to 1}$ 和 $g_0 \in \mathrm{PG}_{(2+n')\to 1}$，并满足

$$\begin{bmatrix} \tilde{r}^0 & \beta_* & \Theta \end{bmatrix} \xrightarrow{W} \begin{bmatrix} \widetilde{\mathsf{R}}_0 & \mathsf{B}_* & \Theta^\top \end{bmatrix}, \qquad \begin{bmatrix} \tilde{p}^0 & z & \varepsilon & \Xi \end{bmatrix} \xrightarrow{W} \begin{bmatrix} \widetilde{\mathsf{P}}_0 & \mathsf{Z} & \mathsf{E} & \Xi^\top \end{bmatrix}, \tag{G.1}$$

其中右侧的随机变量独立于 Λ_n、Λ_d。

定理 G.1. 考虑 *GVAMP* 迭代，即(4.19)，其初始化为(G.1)。设假设(A7)、(A8) 以及(A25) 至(A28) 成立。则对任意固定的 $t \geq 0$,

$$\begin{bmatrix} r^1 & \cdots & r^{t+1} & \tilde{r}^0 & \cdots & \tilde{r}^{t+1} & \beta_* & \Theta \end{bmatrix} \xrightarrow{W} \begin{bmatrix} \mathsf{R}_1 & \cdots & \mathsf{R}_{t+1} & \widetilde{\mathsf{R}}_0 & \cdots & \widetilde{\mathsf{R}}_{t+1} & \mathsf{B}_* & \Theta^\top \end{bmatrix}, \tag{G.2a}$$

$$\begin{bmatrix} p^1 & \cdots & p^{t+1} & \tilde{p}^0 & \cdots & \tilde{p}^{t+1} & z & \varepsilon & \Xi \end{bmatrix} \xrightarrow{W} \begin{bmatrix} \mathsf{P}_1 & \cdots & \mathsf{P}_{t+1} & \widetilde{\mathsf{P}}_0 & \cdots & \widetilde{\mathsf{P}}_{t+1} & \mathsf{Z} & \mathsf{E} & \Xi^\top \end{bmatrix}, \tag{G.2b}$$

其中右侧的随机变量定义于(4.29)。此外，$(\mathsf{J}_{s+1})_{0\leq s\leq t}$、$(\mathsf{K}_{s+1})_{0\leq s\leq t}$ 的协方差矩阵对任意 $t \geq 0$均可逆。

定理 G.1 的完整证明见G.1 节。

下一个结果放宽了假设 (A25)和(A27)，它们是定理 G.1 所要求的，并将其改为假设 (A10)，但代价是只能得出低阶的 Wasserstein 收敛。

定理 G.2. 考虑(4.19)中的 *GVAMP* 迭代，其初始化为(G.1)。设假设(A7)至(A10) 以及(A28) 均成立。那么对于任意固定的 $t \geq 0$,

$$\begin{bmatrix} r^1 & \cdots & r^{t+1} & \tilde{r}^0 & \cdots & \tilde{r}^{t+1} & \beta_* & \Theta \end{bmatrix} \xrightarrow{W_2} \begin{bmatrix} \mathsf{R}_1 & \cdots & \mathsf{R}_{t+1} & \widetilde{\mathsf{R}}_0 & \cdots & \widetilde{\mathsf{R}}_{t+1} & \mathsf{B}_* & \Theta^\top \end{bmatrix}, \tag{G.3a}$$

$$\begin{bmatrix} p^1 & \cdots & p^{t+1} & \tilde{p}^0 & \cdots & \tilde{p}^{t+1} & z & \varepsilon & \Xi \end{bmatrix} \xrightarrow{W_2} \begin{bmatrix} \mathsf{P}_1 & \cdots & \mathsf{P}_{t+1} & \widetilde{\mathsf{P}}_0 & \cdots & \widetilde{\mathsf{P}}_{t+1} & \mathsf{Z} & \mathsf{E} & \Xi^\top \end{bmatrix}, \tag{G.3b}$$

其中，右侧的随机变量定义见(4.29)。

证明见G.2 节。

为完成定理 4.3，还需处理谱初始化，详见G.3 节。

G.1 定理定理 G.1

为了推导定理 G.1中的状态演化结果，我们将(4.19)中的 GVAMP 迭代重新参数化，使其成为下文(G.5)所示的抽象 VAMP 算法形式。随后，我们证明这一抽象 VAMP 实例的状态演化结果，即定理 G.3，由此推出定理 G.1。

引入一个由 $(O, Q) \sim \mathrm{Haar}(\mathbb{O}(n)) \otimes \mathrm{Haar}(\mathbb{O}(d))$ 驱动的抽象迭代。我们很快将配置此迭代，以实现 (4.19) 中的原始 GVAMP 算法。设 $r_{\mathrm{in}}, r_{\mathrm{out}}, s_{\mathrm{in}}, s_{\mathrm{out}}, k_{\mathrm{in}}, k_{\mathrm{out}}, \ell_{\mathrm{in}}, \ell_{\mathrm{out}}$ 是相对于 n, d 固定的非负整数。设 $w_{p,\mathrm{in}} \in \mathbb{R}^{d\times r_{\mathrm{in}}}, w_{p,\mathrm{out}} \in \mathbb{R}^{n\times r_{\mathrm{out}}}, w_{q,\mathrm{in}} \in \mathbb{R}^{d\times s_{\mathrm{in}}}, w_{q,\mathrm{out}} \in \mathbb{R}^{n\times s_{\mathrm{out}}}$ 是与 O, Q 独立的侧信息。假设对于每个 $t \geq 0$，我们给定以下函数:

$$F_{q,\mathrm{in}}^t : \mathbb{R}^{\ell_{\mathrm{in}}} \times \mathbb{R}^{\ell_{\mathrm{out}}} \times \mathbb{R}^{s_{\mathrm{in}}} \to \mathbb{R}^{k_{\mathrm{in}}}, \qquad F_{q,\mathrm{out}}^t : \mathbb{R}^{\ell_{\mathrm{in}}} \times \mathbb{R}^{\ell_{\mathrm{out}}} \times \mathbb{R}^{s_{\mathrm{out}}} \to \mathbb{R}^{k_{\mathrm{out}}}, \tag{G.4a}$$

$$F_{p,\text{in}}^{t+1}\colon \mathbb{R}^{k_{\text{in}}}\times\mathbb{R}^{k_{\text{out}}}\times\mathbb{R}^{r_{\text{in}}}\to\mathbb{R}^{\ell_{\text{in}}},\qquad F_{p,\text{out}}^{t+1}\colon \mathbb{R}^{k_{\text{in}}}\times\mathbb{R}^{k_{\text{out}}}\times\mathbb{R}^{r_{\text{out}}}\to\mathbb{R}^{\ell_{\text{out}}}. \tag{G.4b}$$

考虑以下迭代：对于 $t\ge 0$,

$$q_{\text{in}}^{t}=Q^{\top}f_{p,\text{in}}^{t}\in\mathbb{R}^{d\times\ell_{\text{in}}},\qquad q_{\text{out}}^{t}=O^{\top}f_{p,\text{out}}^{t}\in\mathbb{R}^{n\times\ell_{\text{out}}}, \tag{G.5a}$$

$$f_{q,\text{in}}^{t}=F_{q,\text{在}}^{t}(q_{\text{in}}^{t},q_{\text{out}}^{t},w_{q,\text{在}})\in\mathbb{R}^{d\times k_{\text{in}}},\qquad f_{q,\text{out}}^{t}=F_{q,\text{out}}^{t}(q_{\text{in}}^{t},q_{\text{out}}^{t},w_{q,\text{out}})\in\mathbb{R}^{n\times k_{\text{out}}}, \tag{G.5b}$$

$$p_{\text{在}}^{t+1}=Qf_{q,\text{in}}^{t}\in\mathbb{R}^{d\times k_{\text{in}}},\qquad p_{\text{out}}^{t+1}=Of_{q,\text{out}}^{t}\in\mathbb{R}^{n\times k_{\text{out}}}, \tag{G.5c}$$

$$f_{p,\text{在}}^{t+1}=F_{p,\text{在}}^{t+1}(p_{\text{在}}^{t+1},p_{\text{out}}^{t+1},w_{p,\text{in}})\in\mathbb{R}^{d\times\ell_{\text{in}}},\qquad f_{p,\text{外}}^{t+1}=F_{p,\text{out}}^{t+1}(p_{\text{在}}^{t+1},p_{\text{out}}^{t+1},w_{p,\text{输出}})\in\mathbb{R}^{n\times\ell_{\text{out}}}, \tag{G.5d}$$

以 $f_{p,\text{输入}}^{0}\in\mathbb{R}^{d\times\ell_{\text{in}}}$，$f_{p,\text{out}}^{0}\in\mathbb{R}^{n\times\ell_{\text{out}}}$ 且与 O，Q 独立地初始化。

(G.4) 中的函数(G.4) 在更新(G.5b)和(G.5d)中，均按行作用于其矩阵输入。请注意，这些函数可以接受行数不固定的输入矩阵。如果输出矩阵的行数大于某些输入矩阵的行数，则按照约定，在较短的输入矩阵后附加额外的零行，再对其应用该函数。否则，在逐行应用该函数之前，会截去较高输入矩阵中多出的行。这些操作确保在这两种情况下，所有输入矩阵的列维数都与输出矩阵相同。例如，考虑(G.5b)中的第一个方程。对于任意 $i\in[d]$，根据上述约定，如果 $d>n$,

$$f_{q,\text{in}}^{t}=\begin{bmatrix}F_{q,\text{in}}^{t}((q_{\text{in}}^{t})_{1,:},(q_{\text{out}}^{t})_{1,:},(w_{q,\text{in}})_{1,:})^{\top}\\ \vdots\\ F_{q,\text{in}}^{t}((q_{\text{in}}^{t})_{n,:},(q_{\text{out}}^{t})_{n,:},(w_{q,\text{in}})_{n,:})^{\top}\\ F_{q,\text{in}}^{t}((q_{\text{in}}^{t})_{n+1,:},0_{\ell_{\text{out}}\times 1},(w_{q,\text{in}})_{n+1,:})^{\top}\\ \vdots\\ F_{q,\text{in}}^{t}((q_{\text{in}}^{t})_{d,:},0_{\ell_{\text{out}}\times 1},(w_{q,\text{in}})_{d,:})^{\top}\end{bmatrix}\in\mathbb{R}^{d\times k_{\text{in}}},$$

否则，如果 $d\le n$,

$$f_{q,\text{in}}^{t}=\begin{bmatrix}F_{q,\text{in}}^{t}((q_{\text{in}}^{t})_{1,:},(q_{\text{out}}^{t})_{1,:},(w_{q,\text{in}})_{1,:})^{\top}\\ \vdots\\ F_{q,\text{in}}^{t}((q_{\text{in}}^{t})_{d,:},(q_{\text{out}}^{t})_{d,:},(w_{q,\text{in}})_{d,:})^{\top}\end{bmatrix}\in\mathbb{R}^{d\times k_{\text{in}}}.$$

我们现在设计函数 $F_{q,\text{输入}}^{t}$ $F_{q,\text{out}}^{t}$，$F_{p}^{t+1}{}_{,\text{输入}}$ $F_{p}^{t+1}{}_{,\text{输出}}$ 并作适当选取，使得迭代 (G.5) 实现（一个轻微的推广）(4.19)。为此，我们首先任意固定 $\Gamma\ge 0$ 作为时间索引的上界。由 假设（A2）可知，$O\in\mathbb{R}^{n\times n}$、$Q\in\mathbb{R}^{d\times d}$ 是独立的 Haar 矩阵，它们收集左、右奇异向量，这些向量来自 X，而 $\lambda\in\mathbb{R}^{n\wedge d}$ 是逐项非负的向量，收集奇异值，这些奇异值来自 X。我们定义向量 $\lambda_n\in\mathbb{R}^{n}_{\ge 0}$ 与 $\lambda_d\in\mathbb{R}^{d}_{\ge 0}$ 为:

$$\lambda_n=\begin{cases}\lambda, & n\le d\\ \begin{bmatrix}\lambda^{\top} & 0_{n-d}^{\top}\end{bmatrix}^{\top}, & n>d\end{cases},\qquad \lambda_d=\begin{cases}\begin{bmatrix}\lambda^{\top} & 0_{d-n}^{\top}\end{bmatrix}^{\top}, & n<d\\ \lambda, & n\ge d\end{cases}. \tag{G.6}$$

这些向量具有 Wasserstein 极限 $\lambda_n\xrightarrow{W}\Lambda_n$，$\lambda_d\xrightarrow{W}\Lambda_d$，其中极限随机变量定义于 (3.8)。考虑 4 个向量序列（以 $(n,\ d)$ 为索引）$(\phi_t)_{0\le t\le\Gamma}\subset\mathbb{R}^{d}$，$(\widetilde{\phi}_t)_{0\le t\le\Gamma}\subset\mathbb{R}^{n\wedge d}_{\ge 0}$，$(\psi_t)_{0\le t\le\Gamma}\subset\mathbb{R}^{n}$，$(\widetilde{\psi}_t)_{0\le t\le\Gamma}\subset\mathbb{R}^{n\wedge d}_{\ge 0}$，其中 $\widetilde{\phi}_{t,\ d}$，$\widetilde{\psi}_{t,\ n}$ 分别表示长度为 d、n 的向量，它们通过对 ϕ_t 与 ψ_t 进行截断或~~扩展而得到。这些向量序列服从以下假设。

(A29) $(\phi_t$~~$)_{t\ge 0}$，$(\widetilde{\phi}_t)_{t\ge 0}$，$(\psi_t)_{t\ge 0}$，$(\widetilde{\psi}_t)_{t\ge 0}$ 独立于 O，Q，ε，β_*，Θ，Ξ。对于任意固定的 $\Gamma\ge 0$,

$$\begin{aligned}\begin{bmatrix}\lambda_d & \phi_0 & \cdots & \phi_\Gamma & \widetilde{\phi}_{0,d} & \cdots & \widetilde{\phi}_{\Gamma,d}\end{bmatrix}&\xrightarrow{W}\begin{bmatrix}\Lambda_d & \Phi_0 & \cdots & \Phi_\Gamma & \widetilde{\Phi}_0 & \cdots & \widetilde{\Phi}_\Gamma\end{bmatrix},\\ \begin{bmatrix}\lambda_n & \psi_0 & \cdots & \psi_\Gamma & \widetilde{\psi}_{0,n} & \cdots & \widetilde{\psi}_{\Gamma,n}\end{bmatrix}&\xrightarrow{W}\begin{bmatrix}\Lambda_n & \Psi_0 & \cdots & \Psi_\Gamma & \widetilde{\Psi}_0 & \cdots & \widetilde{\Psi}_\Gamma\end{bmatrix}.\end{aligned} \tag{G.7}$$

此外，对于每个 $t \geq 0$，以下条件成立： $\mathbb{E}[\Phi_t] = \mathbb{E}[\Psi_t] = 0$， $\max\left\{\mathbb{E}\left[\Phi_t^2\right], \operatorname{Var}\left[\widetilde{\Phi}_t\right]\right\} > 0$, $\max\left\{\mathbb{E}\left[\Psi_t^2\right], \operatorname{Var}\left[\widetilde{\Psi}_t\right]\right\} > 0$.

上述向量序列以及这对独立的 Haar 矩阵 $(O,\ Q)$ 可用于定义一种迭代，该迭代对 (4.19) 中的 GVAMP 迭代作了轻微推广：

$$r^{t+1} = Q\,\mathrm{diag}(\phi_t)Q^\top \widetilde{r}^t + Q\,\mathrm{diag}_{d\times n}(\widetilde{\phi}_t)O^\top \widetilde{p}^t, \qquad \widetilde{r}^{t+1} = f_{t+1}(r^{t+1}; \beta_*, \Theta), \tag{G.8a}$$

$$p^{t+1} = O\,\mathrm{diag}(\psi_t)O^\top \widetilde{p}^t + O\,\mathrm{diag}_{n\times d}(\widetilde{\psi}_t)Q^\top \widetilde{r}^t, \qquad \widetilde{p}^{t+1} = g_{t+1}(p^{t+1}; z, \varepsilon, \Xi), \tag{G.8b}$$

使用假设 (A28) 中的 $\widetilde{r}^0, \widetilde{p}^0$ 初始化。上述迭代由 4 个随机矩阵序列驱动（以 $t \geq 0$ 为索引）： $Q\,\mathrm{diag}(\phi_t)Q^\top$、 $O\,\mathrm{diag}_{n\times d}(\widetilde{\phi}_t)Q^\top$、 $O\,\mathrm{diag}(\psi_t)O^\top$、 $O\,\mathrm{diag}_{(n\times d)}(\widetilde{\psi}_t)Q^\top$。这些矩阵共享相同的特征空间／奇异空间，但假设（A29）允许它们的谱在不同时间之间任意相关。显然，(4.19) 中的矩阵 $\Phi_t(X^\top X)$ $\widetilde{\Phi}_t(X)$ $\Psi_t(XX^\top)$ $\widetilde{\Psi}_t(X)$ 是一种特殊情形，因此 (4.19) 中的 GVAMP 迭代是 (G.8) 的特殊情形。事实上，如果我们设定

$$\phi_t = \Phi_t(\lambda_d \circ \lambda_d), \qquad \widetilde{\phi}_t = \widetilde{\Phi}_t(\lambda), \qquad \psi_t = \Psi_t(\lambda_n \circ \lambda_n), \qquad \widetilde{\psi}_t = \widetilde{\Psi}_t(\lambda), \tag{G.9}$$

其中所有函数均按分量应用， （G.8）可化为（4.19）。假设（A29）所提供的额外灵活性将在定理 G.2的技术处理中发挥作用。

在进一步说明（G.5）的具体细节之前，我们对一个序列（由 (n,d) 索引）进行条件化，该序列由 λ, $\widetilde{\phi}_t)_{0\leq t\leq\Gamma}$、$(\phi_t)_{0\leq t\leq\Gamma}$、$(\psi_t)_{0\leq t\leq\Gamma}$、$(\widetilde{\psi}_t)_{0\leq t\leq\Gamma}$ 构成，并在本节余下部分始终满足假设 (A2) 和假设（A29）。下面的定理 G.3 给出了在上述向量条件下执行 Γ 步（G.5）的状态演化结果。根据（G.9）选择这些向量即可得到定理 G.1。

回到抽象 VAMP 迭代（G.5），我们将维数固定为：

$$k_{\mathrm{in}} = 1, \quad k_{\mathrm{out}} = 2, \quad \ell_{\mathrm{in}} = 2, \quad \ell_{\mathrm{out}} = 1, \quad r_{\mathrm{in}} = 1 + d', \quad r_{\mathrm{out}} = 1 + n', \quad s_{\mathrm{in}} = 1 + 2(\Gamma+1), \quad s_{\mathrm{out}} = 1 + 2(\Gamma+1).$$

令侧信息由下式给出

$$\begin{aligned} w_{q,\mathrm{in}} &= \begin{bmatrix} \lambda_d & \phi_0 & \cdots & \phi_\Gamma & \widetilde{\phi}_{0,d} & \cdots & \widetilde{\phi}_{\Gamma,d} \end{bmatrix} \in \mathbb{R}^{d\times(2(\Gamma+1)+1)}, \\ w_{q,\mathrm{out}} &= \begin{bmatrix} \lambda_n & \psi_0 & \cdots & \psi_\Gamma & \widetilde{\psi}_{0,n} & \cdots & \widetilde{\psi}_{\Gamma,n} \end{bmatrix} \in \mathbb{R}^{n\times(2(\Gamma+1)+1)}, \\ w_{p,\mathrm{in}} &= \begin{bmatrix} \beta_* & \Theta \end{bmatrix} \in \mathbb{R}^{d\times(1+d')}, \qquad w_{p,\mathrm{out}} = \begin{bmatrix} \varepsilon & \Xi \end{bmatrix} \in \mathbb{R}^{n\times(1+n')}. \end{aligned} \tag{G.10}$$

对任意 $0 \leq t \leq \Gamma$，我们将用 $w_{q,\mathrm{in},\lambda}, w_{q,\mathrm{in},t}, w_{q,\mathrm{in},\tilde{t}}, w_{q,\mathrm{out},\lambda}, w_{q,\mathrm{out},t}, w_{q,\mathrm{out},\tilde{t}}$ 分别指代 $\lambda_d, \phi_t, \phi_{t,d}$, $\lambda_n, \psi_t, \psi_{t,n}$。我们还将使用记号 $w_{p,\mathrm{in},1}, w_{p,\mathrm{in},-}$ 分别指代 β_*, Θ，并用 $w_{p,\mathrm{out},1}, w_{p,\mathrm{out},-}$ 分别指代 ε, Ξ。迭代量按如下方式具体化。对 $t = 0$，令

$$f^0_{p,\mathrm{in}} = \begin{bmatrix} \widetilde{r}^0 - \widetilde{b}_0\beta_* & \beta_* \end{bmatrix} \in \mathbb{R}^{d\times 2}, \qquad f^0_{p,\mathrm{out}} = \widetilde{p}^0 - \widetilde{a}_0 z \in \mathbb{R}^n, \tag{G.11}$$

其中 $\widetilde{r}^0$， $\widetilde{p}^0$ 是(4.19) 的、由(G.1) 给出的初始化量，且 b_0， $\widetilde{a}_0$ 在 （4.29e） 中定义。对 $t \geq 0$，我们按如下方式定义(G.4) 中的函数。对于输入 $q_{\mathrm{in}} = \begin{bmatrix} q_{\mathrm{in},1} & q_{\mathrm{in},2} \end{bmatrix}^\top \in \mathbb{R}^2$, $q_{\mathrm{out}} \in \mathbb{R}$, $p_{\mathrm{in}} \in \mathbb{R}$, $p_{\mathrm{out}} = \begin{bmatrix} p_{\mathrm{out},1} & p_{\mathrm{out},2} \end{bmatrix}^\top \in \mathbb{R}^2$, $w_{q,\ \mathrm{in}} = \begin{bmatrix} w_{q,\mathrm{in},\lambda} & w_{q,\mathrm{in},0} & \cdots & w_{q,\mathrm{in},\Gamma} & \cdots & w_{q,\mathrm{in},\tilde{0}} & \cdots & w_{q,\mathrm{in},\tilde{\Gamma}} \end{bmatrix}^\top \in \mathbb{R}^{2(\Gamma+1)+1}$, $w_{q,\ \mathrm{out}} = \begin{bmatrix} w_{q,\mathrm{out},\lambda} & w_{q,\mathrm{out},0} & \cdots & w_{q,\mathrm{out},\Gamma} & \cdots & w_{q,\mathrm{out},\tilde{0}} & \cdots & w_{q,\mathrm{out},\tilde{\Gamma}} \end{bmatrix}^\top \in \mathbb{R}^{2(\Gamma+1)+1}$，这些函数的输出由下式给出

$$F^t_{q,\mathrm{in}}(q_{\mathrm{in}}, q_{\mathrm{out}}, w_{q,\mathrm{in}}) = w_{q,\mathrm{in},t}(q_{\mathrm{in},1} + \widetilde{\mathfrak{b}}_t q_{\mathrm{in},2}) + w_{q,\mathrm{in},\tilde{t}}(q_{\mathrm{out}} + \widetilde{\mathfrak{a}}_t w_{q,\mathrm{in},\lambda} q_{\mathrm{in},2}) - \mathfrak{b}_{t+1} q_{\mathrm{in},2} \in \mathbb{R}, \tag{G.12a}$$

$$F^t_{q,\mathrm{out}}(q_{\mathrm{in}}, q_{\mathrm{out}}, w_{q,\mathrm{out}}) = \begin{bmatrix} w_{q,\mathrm{out},t}(q_{\mathrm{out}} + \mathfrak{a}_t w_{q,\mathrm{out},\lambda} q_{\mathrm{in},2}) + w_{q,\mathrm{out},\tilde{t}}(q_{\mathrm{in},1} + \mathfrak{b}_t q_{\mathrm{in},2}) - \mathfrak{a}_{t+1} w_{q,\mathrm{out},\lambda} q_{\mathrm{in},2} \\ w_{q,\mathrm{out},\lambda} q_{\mathrm{in},2} \end{bmatrix}^\top \in \mathbb{R}^2, \tag{G.12b}$$

$$F^{t+1}_{p,\text{in}}(p_\text{in}, p_\text{out}, w_{p,\text{in}}) = \begin{bmatrix} f_{t+1}(p_\text{in} + \mathfrak{b}_{t+1} w_{p,\text{in},1}; w_{p,\text{in}}) - \widetilde{\mathfrak{b}}_{t+1} w_{p,\text{in},1} & w_{p,\text{in},1} \end{bmatrix}^\top \in \mathbb{R}^2, \tag{G.12c}$$

$$F^{t+1}_{p,\text{out}}(p_\text{in}, p_\text{out}, w_{p,\text{out}}) = g_{t+1}(p_{\text{out},1} + \mathfrak{a}_{t+1} p_{\text{out},2}; p_{\text{out},2}, w_{p,\text{out}}) - \widetilde{\mathfrak{a}}_{t+1} p_{\text{out},2} \in \mathbb{R}, \tag{G.12d}$$

其中，系数 $\widetilde{\mathfrak{a}}_t$，$\mathfrak{b}_t$，$\mathfrak{a}_{t+1}$，$\mathfrak{b}_{t+1}$ 定义见 (G.14k) 和 (G.14q)。这些系数的设计方式使得对于每个 $t \geq 0$，

$$\frac{1}{n}\langle f^t_{p,\text{out}}, z\rangle \to 0, \qquad \frac{1}{d}\langle f^t_{p,\text{in},1}, \beta_*\rangle \to 0,$$
$$\frac{1}{n}\langle p^{t+1}_\text{out}, z\rangle \to 0, \qquad \frac{1}{d}\langle p^{t+1}_\text{in}, \beta_*\rangle \to 0,$$

这将在定理 G.3的证明中得到严格论证。

现在，我们说明(4.19)中的 GVAMP 迭代与(G.5) 中的抽象 VAMP 迭代之间的联系，其中初始化为 (G.11)，侧信息为(G.10) ，去噪函数取(G.4) 并特化为(G.12)。从下文定理 G.1 的证明中（尤其是其中的条目 1 ）可以看到，在 (G.9) 中的规定下，有 $\widetilde{\mathfrak{a}}_t = \widetilde{a}_t$，$\mathfrak{b}_t = b_t$，$\mathfrak{a}_{t+1} = a_{t+1}$，$\mathfrak{b}_{t+1} = b_{t+1}$ 对所有 $t \geq 0$ 均成立，其中右侧的量定义于（4.29e）和(4.29f)中。由此，通过对 t作归纳法，可以很容易地验证(G.5)可化为以下迭代：

$$\begin{aligned}
q^t_\text{in} &= \begin{bmatrix} Q^\top(\widetilde{r}^t - \widetilde{b}_t \beta_*) & Q^\top \beta_* \end{bmatrix} \in \mathbb{R}^{d\times 2}, \\
q^t_\text{out} &= O^\top(\widetilde{p}^t - \widetilde{a}_t z) \in \mathbb{R}^n, \\
f^t_{q,\text{in}} &= \Phi_t(\Lambda^\top\Lambda) Q^\top \widetilde{r}^t + \widetilde{\Phi}_t(\Lambda)^\top O^\top \widetilde{p}^t - b_{t+1} Q^\top \beta_* \in \mathbb{R}^d, \\
f^t_{q,\text{out}} &= \begin{bmatrix} \Psi_t(\Lambda\Lambda^\top) O^\top \widetilde{p}^t + \widetilde{\Psi}_t(\Lambda) Q^\top \widetilde{r}^t - a_{t+1} \Lambda Q^\top \beta_* & \Lambda Q^\top \beta_* \end{bmatrix} \in \mathbb{R}^{n\times 2}, \\
p^{t+1}_\text{in} &= r^{t+1} - b_{t+1}\beta_* \in \mathbb{R}^d, \\
p^{t+1}_\text{out} &= \begin{bmatrix} p^{t+1} - a_{t+1} z & z \end{bmatrix} \in \mathbb{R}^{n\times 2}, \\
f^{t+1}_{p,\text{in}} &= \begin{bmatrix} \widetilde{r}^{t+1} - \widetilde{b}_{t+1}\beta_* & \beta_* \end{bmatrix} \in \mathbb{R}^{d\times 2}, \\
f^{t+1}_{p,\text{out}} &= \widetilde{p}^{t+1} - \widetilde{a}_{t+1} z \in \mathbb{R}^n,
\end{aligned} \tag{G.13}$$

由此可立即得出 (G.5)与 (4.19) 等价。

我们用 q^t_in,1, f^t_q,out 1, p^{t+1}_out,1, f^{t+1}_p,in,1 表示 q^t_in, $f^t_{q,\text{out}}$, p^{t+1}_out, f^{t+1}_p,in 的第一列，用 q_in,2, $f_{q,\text{out}}$,2, p_out,2, $f_{p,\text{in}}$,2 表示它们的第二列（因为后者是时间不变的）。

从较高层面来看，定理 G.3 中的状态演化结果表明，迭代量的联合经验分布收敛于一个多元高斯分布，其协方差矩阵可以随时间递归地跟踪。例如，$\begin{bmatrix} q^0_{\text{in},1} & \cdots & q^t_{\text{in},1} \end{bmatrix}$ 的各行的经验分布在 Wasserstein 意义下收敛于 $\mathcal{N}(0_{t+1}, \Sigma^t_{q,\text{in}})$。类似的结果也适用于 $(q^s_\text{out})_{0\leq s\leq t}$、$(p^s_\text{in})_{1\leq s\leq t+1}$、$(p^s_\text{out}, 1)_{1\leq s\leq t+1}$，其对应的协方差矩阵为$\Sigma^t_q$、$\Sigma^{t+1}_p$，位于 Σ^{t+1}_p、out中。现在，我们定义这些协方差矩阵，以及与抽象 VAMP 迭代（G.5）相关的其他状态演化参数。对于 $t = 0$，它们由以下初始条件确定：

$$\begin{aligned}
&\begin{bmatrix} f^0_{p,\text{in},1} & w_{q,\text{in},\lambda} & w_{q,\text{in},0} & \cdots & w_{q,\text{in},\Gamma} & w_{q,\text{in},\widetilde{0}} & \cdots & w_{q,\text{in},\widetilde{\Gamma}} & w_{p,\text{in},1} & w_{p,\text{in},-} \end{bmatrix} \\
&\quad = \begin{bmatrix} \widetilde{r}^0 - \widetilde{b}_0\beta_* & \lambda_d & \phi_0 & \cdots & \phi_\Gamma & \widetilde{\phi}_{0,d} & \cdots & \widetilde{\phi}_{\Gamma,d} & \beta_* & \Theta \end{bmatrix} \\
&\quad \xrightarrow{W} \begin{bmatrix} \widetilde{\mathsf{R}}_0 - \widetilde{b}_0 \mathsf{B}_* & \Lambda_d & \Phi_0 & \cdots & \Phi_\Gamma & \widetilde{\Phi}_0 & \cdots & \widetilde{\Phi}_\Gamma & \mathsf{B}_* & \Theta^\top \end{bmatrix} \\
&\quad =: \begin{bmatrix} \mathsf{F}^0_{p,\text{in},1} & \mathsf{W}_{q,\text{in},\lambda} & \mathsf{W}_{q,\text{in},0} & \cdots & \mathsf{W}_{q,\text{in},\Gamma} & \mathsf{W}_{q,\text{in},\widetilde{0}} & \cdots & \mathsf{W}_{q,\text{in},\widetilde{\Gamma}} & \mathsf{W}_{p,\text{in},1} & \mathsf{W}^\top_{p,\text{in},-} \end{bmatrix}, \\
&\begin{bmatrix} f^0_{p,\text{out}} & w_{q,\text{out},\lambda} & w_{q,\text{out},0} & \cdots & w_{q,\text{out},\Gamma} & w_{q,\text{out},\widetilde{0}}, & \cdots & w_{q,\text{out},\widetilde{\Gamma}} & w_{p,\text{out},1} & w_{p,\text{out},-} \end{bmatrix} \\
&\quad = \begin{bmatrix} \widetilde{p}^0 - \widetilde{a}_0 z & \lambda_n & \psi_0 & \cdots & \psi_\Gamma & \widetilde{\psi}_{0,n} & \cdots & \widetilde{\psi}_{\Gamma,n} & \varepsilon & \Xi \end{bmatrix}
\end{aligned} \tag{G.14a}$$

$$\xrightarrow{W} \left[\widetilde{\mathsf{P}}_0 - \widetilde{a}_0\mathsf{Z} \quad \Lambda_n \quad \Psi_0 \quad \cdots \quad \Psi_\Gamma \quad \widetilde{\Psi}_0 \quad \cdots \quad \widetilde{\Psi}_\Gamma \quad \mathsf{E} \quad \Xi^\top\right]$$
$$=: \left[\mathsf{F}^0_{p,\mathrm{out}} \quad \mathsf{W}_{q,\mathrm{out},\lambda} \quad \mathsf{W}_{q,\mathrm{out},0} \quad \cdots \quad \mathsf{W}_{q,\mathrm{out},\Gamma} \quad \mathsf{W}_{q,\mathrm{out},\widetilde{0}}, \quad \cdots \quad \mathsf{W}_{q,\mathrm{out},\widetilde{\Gamma}} \quad \mathsf{W}_{p,\mathrm{out},1} \quad \mathsf{W}^\top_{p,\mathrm{out},-}\right], \tag{G.14b}$$

其中，右侧极限随机变量的存在性由假设（A28）和（A29）保证，常数 $\widetilde{a}_0$、$\widetilde{b}_0$ 按照（4.29e）定义。我们还将使用记号 $\mathsf{W}_{p,\mathrm{in}} = \left[\mathsf{W}_{p,\mathrm{in},\mathbf{1}}\ \mathsf{W}^\top_{p,\mathrm{in},-}\right]^\top$, $\mathsf{W}_{p,\mathrm{out}} = \left[\mathsf{W}_{p,\mathrm{out},\mathbf{1}}\ \mathsf{W}^\top_{p,\mathrm{out},-}\right]^\top$, $\mathsf{W}_{q,\mathrm{in}} = \left[\mathsf{W}_{q,\mathrm{in},\lambda}\ \mathsf{W}_{q,\mathrm{in},0} \ \cdots\ \mathsf{W}_{q,\mathrm{in},\Gamma}\ \mathsf{W}_{q,\mathrm{in},\widetilde{0}} \ \cdots\ \mathsf{W}_{q,\mathrm{in},\widetilde{\Gamma}}\right]^\top$，以及 $\mathsf{W}_{q,\mathrm{out}} = \left[\mathsf{W}_{q,\mathrm{out},\lambda}\ \mathsf{W}_{q,\mathrm{out},0} \ \cdots\ \mathsf{W}_{q,\mathrm{out},\Gamma}\ \mathsf{W}_{q,\mathrm{out},\widetilde{0}}, \ \cdots\ \mathsf{W}_{q,\mathrm{out},\widetilde{\Gamma}}\right]^\top$。为方便记号，令随机向量 $\mathsf{A} \in \mathbb{R}^{4(\Gamma+1)+3+d'+n'}$ 收集：

$$\mathsf{A} = \left[\Phi_0 \quad \cdots \quad \Phi_\Gamma \quad \widetilde{\Phi}_0 \quad \cdots \quad \widetilde{\Phi}_\Gamma \quad \Psi_0 \quad \cdots \quad \Psi_\Gamma \quad \widetilde{\Psi}_0 \quad \cdots \quad \widetilde{\Psi}_\Gamma \quad \Lambda \quad \mathsf{B}_* \quad \mathsf{E} \quad \Theta^\top \quad \Xi^\top\right]^\top.$$

利用这些随机变量，我们定义

$$\Sigma^0_{q,\mathrm{in}} = \mathbb{E}\left[(\widetilde{\mathsf{R}}_0 - \widetilde{b}_0\mathsf{B}_*)^2\right] = \mathbb{E}\left[(\mathsf{F}^0_{p,\mathrm{in},1})^2\right] \in \mathbb{R}, \quad \Sigma^0_{q,\mathrm{out}} = \mathbb{E}\left[(\widetilde{\mathsf{P}}_0 - \widetilde{a}_0\mathsf{Z})^2\right] = \mathbb{E}\left[(\mathsf{F}^0_{p,\mathrm{out}})^2\right] \in \mathbb{R}. \tag{G.14c}$$

注意

$$\Sigma^0_{q,\mathrm{in}} > 0, \qquad \Sigma^0_{q,\mathrm{out}} > 0, \tag{G.14d}$$

~由于根据 b_0 在（4.29e）中的定义，

$$\mathbb{E}\left[(\widetilde{\mathsf{R}}_0 - \widetilde{b}_0\mathsf{B}_*)^2\right] = \mathbb{E}\left[\widetilde{\mathsf{R}}_0^2\right] - 2\widetilde{b}_0\mathbb{E}\left[\widetilde{\mathsf{R}}_0\mathsf{B}_*\right] + \widetilde{b}_0^2\mathbb{E}\left[\mathsf{B}_*^2\right] = \mathbb{E}\left[\widetilde{\mathsf{R}}_0^2\right] - \widetilde{b}_0^2\rho,$$

类似地，$\mathbb{E}\left[\ (\widetilde{\ }\mathsf{P}_0 - \widetilde{a}_0\mathsf{Z})\ ^2\right] = \mathbb{E}\left[\widetilde{\ }\mathsf{P}_0^2\right] - \widetilde{a}_0\sigma^2$。根据假设 (A28)以及柯西–~~施瓦茨不等式的等号成立条件，随机变量R_0,P_0 严格满足以下不等式：

$$\mathbb{E}\left[\widetilde{\mathsf{R}}_0\mathsf{B}_*\right]^2 < \mathbb{E}\left[\widetilde{\mathsf{R}}_0^2\right]\mathbb{E}\left[\mathsf{B}_*^2\right] = \rho\mathbb{E}\left[\widetilde{\mathsf{R}}_0^2\right], \qquad \mathbb{E}\left[\widetilde{\mathsf{P}}_0\mathsf{Z}\right]^2 < \mathbb{E}\left[\widetilde{\mathsf{P}}_0^2\right]\mathbb{E}\left[\mathsf{Z}^2\right] = \sigma^2\mathbb{E}\left[\widetilde{\mathsf{P}}_0^2\right],$$

这意味着所声称的正性。

现在，我们递归地定义状态演化参数 $\mathfrak{a}_{t+1}$、$\mathfrak{b}_{t+1} \in \mathbb{R}$, Σ_p^{t+1}、在 Σ_p^{t+1} 中，out $\in \mathbb{R}^{(t+1)\times(t+1)}$,$\widetilde{\mathfrak{a}}_{t+1}$、$\widetilde{\mathfrak{b}}_{t+1} \in \mathbb{R}$ 以及 Σ_q^{t+1}、在 Σ_q^{t+1} 中，out $\in \mathbb{R}^{(t+2)\times(t+2)}$。令

$$\widetilde{\mathfrak{a}}_0 = \widetilde{a}_0, \qquad \widetilde{\mathfrak{b}}_0 = \widetilde{b}_0. \tag{G.14e}$$

对于 $t \geq 0$，给定 $\Sigma^t_{q,\ \mathrm{in}}$, $\Sigma^t_{q,\ \mathrm{out}}$，令

$$\left[\mathsf{Q}_{\mathrm{in},2} \quad \mathsf{Q}^0_{\mathrm{in},1} \quad \cdots \quad \mathsf{Q}^t_{\mathrm{in},1}\right]^\top \sim \mathcal{N}(0,\rho) \otimes \mathcal{N}(0_{t+1}, \Sigma^t_{q,\mathrm{in}}), \tag{G.14f}$$
$$\left[\mathsf{Q}^0_{\mathrm{out}} \quad \cdots \quad \mathsf{Q}^t_{\mathrm{out}}\right]^\top \sim \mathcal{N}(0_{t+1}, \Sigma^t_{q,\mathrm{out}}) \tag{G.14g}$$

是两个相互独立且也独立于 A 的高斯向量，$(\mathsf{P}^{s+1}$ 中$)_{0\leq s\leq t-1}$，$(\mathsf{P}^{s+1}_{\mathrm{out}})_{0\leq s\leq t-1}$。对于每个 $0 \leq s \leq t$，记 $\mathsf{Q}^s_{\mathrm{in}} = \left[\mathsf{Q}^s_{\mathrm{in},\ 1}\ \mathsf{Q}_{\mathrm{in},\ 2}\right]^\top$。令

$$\begin{aligned}\mathsf{F}^t_{q,\mathrm{in}} &= F^t_{q,\mathrm{in}}(\mathsf{Q}^t_{\mathrm{in}}, \mathsf{Q}^t_{\mathrm{out}}, \Lambda_d) \\ &= \Phi_t(\mathsf{Q}^t_{\mathrm{in},1} + \widetilde{\mathfrak{b}}_t\mathsf{Q}_{\mathrm{in},2}) + \widetilde{\Phi}_t(\mathsf{Q}^t_{\mathrm{out}} + \widetilde{\mathfrak{a}}_t\Lambda_d\mathsf{Q}_{\mathrm{in},2}) - \mathfrak{b}_{t+1}\mathsf{Q}_{\mathrm{in},2}, \end{aligned}\tag{G.14h}$$
$$\begin{aligned}\mathsf{F}^t_{q,\mathrm{out},1} &= F^t_{q,\mathrm{out},1}(\mathsf{Q}^t_{\mathrm{in}}, \mathsf{Q}^t_{\mathrm{out}}, \Lambda_n) \\ &= \Psi_t(\mathsf{Q}^t_{\mathrm{out}} + \widetilde{\mathfrak{a}}_t\Lambda_n\mathsf{Q}_{\mathrm{in},2}) + \widetilde{\Psi}_t(\mathsf{Q}^t_{\mathrm{in},1} + \widetilde{\mathfrak{b}}_t\mathsf{Q}_{\mathrm{in},2}) - \mathfrak{a}_{t+1}\Lambda_n\mathsf{Q}_{\mathrm{in},2}, \end{aligned}\tag{G.14i}$$

$$\mathsf{F}_{q,\mathrm{out},2} = F^t_{q,\mathrm{out},2}(\mathsf{Q}^t_{\mathrm{in}}, \mathsf{Q}^t_{\mathrm{out}}, \Lambda_n) = \Lambda_n \mathsf{Q}_{\mathrm{in},2}, \tag{G.14j}$$

其中常数 $(\mathfrak{a}_{s+1})_{0\le s\le t}$，$(\mathfrak{b}_{s+1})_{0\le s\le t}$ 定义为

$$\mathfrak{a}_{r+1} = \frac{1}{\kappa_2}\Big(\mathbb{E}\big[\Psi_r \Lambda_n^2\big]\widetilde{\mathfrak{a}}_r + \mathbb{E}\Big[\widetilde{\Psi}_r \Lambda_n\Big]\widetilde{\mathfrak{b}}_r\Big), \qquad \mathfrak{b}_{r+1} = \mathbb{E}\Big[\widetilde{\Phi}_r \Lambda_d\Big]\widetilde{\mathfrak{a}}_r. \tag{G.14k}$$

令 $\mathsf{F}^t_{q,\mathrm{out}} = \big[\mathsf{F}^t_{q,\mathrm{out},1}\ \mathsf{F}_{q,\mathrm{out},2}\big]^\top$。利用这些随机变量，定义协方差矩阵

$$\Sigma^{t+1}_{p,\mathrm{in}} = \mathbb{E}\big[\mathsf{F}^{0:t}_{q,\mathrm{in}}(\mathsf{F}^{0:t}_{q,\mathrm{in}})^\top\big] \in \mathbb{R}^{(t+1)\times(t+1)}, \quad \Sigma^{t+1}_{p,\mathrm{out}} = \mathbb{E}\big[\mathsf{F}^{0:t}_{q,\mathrm{out},1}(\mathsf{F}^{0:t}_{q,\mathrm{out},1})^\top\big] \in \mathbb{R}^{(t+1)\times(t+1)}. \tag{G.14l}$$

接下来，定义两个相互独立、具有上述协方差的高斯向量：

$$\big[\mathsf{P}^1_{\mathrm{in}} \quad \cdots, \mathsf{P}^{t+1}_{\mathrm{in}}\big]^\top \sim \mathcal{N}\Big(0_{t+1}, \Sigma^{t+1}_{p,\mathrm{in}}\Big), \tag{G.14m}$$

$$\big[\mathsf{Z} \quad \mathsf{P}^1_{\mathrm{out},1} \quad \cdots \quad \mathsf{P}^{t+1}_{\mathrm{out},1}\big]^\top \sim \mathcal{N}\big(0, \sigma^2\big) \otimes \mathcal{N}\big(0_{t+1}, \Sigma^{t+1}_{p,\mathrm{out}}\big), \tag{G.14n}$$

它们与 $(\mathsf{Q}^s_{\mathrm{in}})_{0\le s\le t}$,$(\mathsf{Q}^s_{\mathrm{out}})_{0\le s\le t}$, A 独立。我们还将使用记号 $\mathsf{P}_{\mathrm{out},2} = \mathsf{Z}$ 和 $\mathsf{P}^{s+1}_{\mathrm{out}} = \big[\mathsf{P}^{s+1}_{\mathrm{out},1}\ \mathsf{P}_{\mathrm{out},2}\big]^\top$，对每个 $0 \le s \le t$ 均如此。令

$$\mathsf{F}^{t+1}_{p,\mathrm{in},1} = F^{t+1}_{p,\mathrm{in},1}(\mathsf{P}^{t+1}_{\mathrm{in}}, \mathsf{P}^{t+1}_{\mathrm{out}}, \mathsf{B}_*, \Theta) = f_{t+1}(\mathsf{P}^{t+1}_{\mathrm{in}} + \mathfrak{b}_{t+1}\mathsf{B}_*; \mathsf{B}_*, \Theta) - \widetilde{\mathfrak{b}}_{t+1}\mathsf{B}_*, \tag{G.14o}$$

$$\mathsf{F}^{t+1}_{p,\mathrm{out}} = F^{t+1}_{p,\mathrm{out}}(\mathsf{P}^{t+1}_{\mathrm{in}}, \mathsf{P}^{t+1}_{\mathrm{out}}, \mathsf{E}, \Xi) = g_{t+1}(\mathsf{P}^{t+1}_{\mathrm{out},1,} + \mathfrak{a}_{t+1}\mathsf{P}_{\mathrm{out},2}; \mathsf{P}_{\mathrm{out},2}, \mathsf{E}, \Xi) - \widetilde{\mathfrak{a}}_{t+1}\mathsf{P}_{\mathrm{out},2}, \tag{G.14p}$$

~其中常数 $\widetilde{\mathfrak{a}}_{t+1}, \mathfrak{b}_{t+1}$ 定义为

$$\begin{aligned}\widetilde{\mathfrak{a}}_{t+1} &= \frac{1}{\sigma^2}\mathbb{E}\big[g_{t+1}(\mathsf{P}^{t+1}_{\mathrm{out},1} + \mathfrak{a}_{t+1}\mathsf{P}_{\mathrm{out},2}; \mathsf{P}_{\mathrm{out},2}, \mathsf{E}, \Xi)\mathsf{Z}\big],\\ \widetilde{\mathfrak{b}}_{t+1} &= \frac{1}{\rho}\mathbb{E}\big[f_{t+1}(\mathsf{P}^{t+1}_{\mathrm{in}} + \mathfrak{b}_{t+1}\mathsf{B}_*; \mathsf{B}_*, \Theta)\mathsf{B}_*\big].\end{aligned} \tag{G.14q}$$

再令 $\mathsf{F}_{p,\mathrm{in},2} = \mathsf{B}_*$ 和 $\mathsf{F}^{t+1}_{p\ ,\mathrm{in}} = \big[\mathsf{F}^{t+1}_{p\ ,\mathrm{in},1}\ \mathsf{F}_{p,\mathrm{in},2}\big]^\top$。利用这些随机变量定义协方差矩阵：

$$\Sigma^{t+1}_{q,\mathrm{in}} = \mathbb{E}\Big[\mathsf{F}^{0:t+1}_{p,\mathrm{in},1}(\mathsf{F}^{0:t+1}_{p,\mathrm{in},1})^\top\Big] \in \mathbb{R}^{(t+2)\times(t+2)}, \quad \Sigma^{t+1}_{q,\mathrm{out}} = \mathbb{E}\big[\mathsf{F}^{0:t+1}_{p,\mathrm{out}}(\mathsf{F}^{0:t+1}_{p,\mathrm{out}})^\top\big] \in \mathbb{R}^{(t+2)\times(t+2)}. \tag{G.14r}$$

注意，根据 (G.14l) 和 (G.14r) 中的定义，对于每个 $t \ge 0$, $\Sigma^t_{q,\mathrm{in}}$,$\Sigma^t_{q,\mathrm{out}}$, Σ^{t+1}_p，in $\Sigma^{t+1}_{p\ ,\mathrm{out}}$ 分别是 Σ^{t+1}_q，in $\Sigma^{t+1}_{q\ ,\mathrm{out}}$, Σ^{t+2}_p,in $\Sigma^{t+2}_{p\ ,\mathrm{out}}$ 的顺序主子式（阶数为 $t+1$）。按照记号约定，我们用 $(i,j) \in \{0,1,\cdots,t\}^2$ 标记 $\Sigma^t_{q,\mathrm{in}}$, $\Sigma^t_{q,\mathrm{out}}$ 中的元素，并用 $(i,j) \in \{1,2,\cdots,t+1\}^2$ 标记 Σ^{t+1}_p,in $\Sigma^{t+1}_{p\ ,\mathrm{out}}$ 中的元素，从而便于表示，例如，对于任意 $0 \le r,s \le t$，有 $(\Sigma^t_{q,\mathrm{in}})_{r,s} = \mathbb{E}[\mathsf{Q}^r_{\mathrm{in}}\mathsf{Q}^s_{\mathrm{in}}]$ 和 $(\Sigma^{t+1}_{p\ ,\mathrm{in}})_{r+1,s+1} = \mathbb{E}\big[\mathsf{P}^{r+1}\mathrm{in}\,\mathsf{P}^{s+1}\mathrm{in}\quad\big]$。

总而言之，对于任意 $t \ge 0$，(G.14) 定义了随机变量 A、$(\mathsf{Q}^s_{\mathrm{in}})_{0\le s\le t}$、$(\mathsf{Q}^s_{\mathrm{out}})_{0\le s\le t}$、$(\mathsf{F}^s_{q,\mathrm{in}})_{0\le s\le t}$、$(\mathsf{F}^s_{q,\mathrm{out}})_{0\le s\le t}$、$(\mathsf{P}^{s+1}\ \mathrm{in}\ )_{0\le s\le t}$、$(\mathsf{P}^{s+1}_{\mathrm{out}})_{0\le s\le t}$、$(\mathsf{F}^s_{p,\mathrm{in}})_{0\le s\le t+1}$、$(\mathsf{F}^s_{p,\mathrm{out}})_{0\le s\le t+1}$。本节的主要技术结果——下文的定理 G.3——断言，这些随机变量确定了 (G.5) 中迭代量的 Wasserstein 极限。

(A30) 对于每个 $t \ge 0$，不存在常数 β_0，$\cdots$，β_t，α_0，$\cdots$，$\alpha_t \in \mathbb{R}$ 使得

$$\text{或者} \qquad \mathsf{F}^{t+1}_{p,\mathrm{in},1} = \sum_{s=0}^t \beta_s \mathsf{F}^s_{p,\mathrm{in},1} \qquad \text{or} \qquad \mathsf{F}^{t+1}_{p,\mathrm{out}} = \sum_{s=0}^t \alpha_s \mathsf{F}^s_{p,\mathrm{out}}$$

几乎必然成立。

对于任意 $t \geq 0$，令 $\boldsymbol{q}^t_{\text{in},1} := \begin{bmatrix} q^0_{\text{in},1} & \cdots & q^t_{\text{in},1} \end{bmatrix} \in \mathbb{R}^{d\times(t+1)}$, $\mathsf{Q}^t_{\text{in},1} := \begin{bmatrix} \mathsf{Q}^0_{\text{in},1} & \cdots \mathsf{Q}^t_{\text{in},1} \end{bmatrix} \in \mathbb{R}^{1\times(t+1)}$ 以及其他记号，包括 $\boldsymbol{f}^t_{q,\text{in}}$, $\boldsymbol{p}^{t+1}\text{in}$, $\boldsymbol{f}^{t+1}_{p}$,in 1, $\boldsymbol{q}^t_{\text{out}}$, $\boldsymbol{f}^t_{q}$,out 1, $\boldsymbol{p}^{t+1}_{\text{out}}$,1, $\boldsymbol{f}^{t+1}_{p}$,out, $\mathsf{F}^t_{q,\text{in}}$, $\mathsf{P}^{t+1}\text{in}$, F^{t+1}_{p},in,1, $\mathsf{Q}^t_{\text{out}}$, $\mathsf{F}^t_{q,\text{out}}$,1, $\mathsf{P}^{t+1}_{\text{out}}$,1, F^{t+1}_{p},out ，定义方式类似。

定理 G.3（抽象 VAMP 的状态演化）。考虑抽象 *VAMP* 迭代 (G.5)，其初始化为 (G.11)，侧信息为 (G.10)，去噪函数 (G.4) 特化为 (G.12)。设假设 (A8) 和 (A27) 至 (A30) 成立。那么对于任意固定的 $t \geq 0$,

$$\begin{aligned} &\begin{bmatrix} q_{\text{in},2} & \boldsymbol{q}^t_{\text{in},1} & \boldsymbol{f}^t_{q,\text{in}} & \boldsymbol{p}^{t+1}_{\text{in}} & f_{p,\text{in},2} & \boldsymbol{f}^{t+1}_{p,\text{in},1} & w_{p,\text{in}} & w_{q,\text{in}} \end{bmatrix} \\ &\quad \xrightarrow{W} \begin{bmatrix} \mathsf{Q}_{\text{in},2} & \mathbf{Q}^t_{\text{in},1} & \mathbf{F}^t_{q,\text{in}} & \mathbf{P}^{t+1}_{\text{in}} & \mathsf{F}_{p,\text{in},2} & \mathbf{F}^{t+1}_{p,\text{in},1} & \mathsf{W}^\top_{p,\text{in}} & \mathsf{W}^\top_{q,\text{in}} \end{bmatrix}, \qquad \text{(G.15)} \\ &\begin{bmatrix} \boldsymbol{q}^t_{\text{out}} & f_{q,\text{out},2} & \boldsymbol{f}^t_{q,\text{out},1} & p_{\text{out},2} & \boldsymbol{p}^{t+1}_{\text{out},1} & \boldsymbol{f}^{t+1}_{p,\text{out}} & w_{p,\text{out}} & w_{q,\text{out}} \end{bmatrix} \\ &\quad \xrightarrow{W} \begin{bmatrix} \mathbf{Q}^t_{\text{out}} & \mathsf{F}_{q,\text{out},2} & \mathbf{F}^t_{q,\text{out},1} & \mathsf{P}_{\text{out},2} & \mathbf{P}^{t+1}_{\text{out},1} & \mathbf{F}^{t+1}_{p,\text{out}} & \mathsf{W}^\top_{p,\text{out}} & \mathsf{W}^\top_{q,\text{out}} \end{bmatrix}. \qquad \text{(G.16)} \end{aligned}$$

由于 (G.5) 中的抽象 VAMP 被设置为模拟 (4.19) 中的 GVAMP 动力学（参见 (G.13)），因此 定理 G.3 中针对前一种迭代的状态演化结果可以转化为针对后一种迭代的结果，如定理 G.1所述。

~ ~
定理 *G.1* 的证明。我们根据 ϕ_t, ϕ_t, ψ_t, ψ_t（对每个 $t \geq 0$）将向量特化为 (G.9)。由于
~ ~
函数 Φ_t, Φ_t, Ψ_t, Ψ_t 在 假设（A25）的保证下满足正则性，定理 K.2 意味着所需的 Wasserstein 收敛满足假设（A29）。特别地，对于每个 $t \geq 0$,

$$\Phi_t = \Phi_t(\Lambda_d^2), \qquad \widetilde{\Phi}_t = \widetilde{\Phi}_t(\Lambda_d), \qquad \Psi_t = \Psi_t(\Lambda_n^2), \qquad \widetilde{\Psi}_t = \widetilde{\Psi}_t(\Lambda_n).$$

Φ_t、Ψ_t 在假设（A7）中的无迹性质意味着 Φ_t、Ψ_t 满足假设（A29）所要求的零均值条件。因此，定理 G.3适用于当前情形。

通过对 $t \geq 0$进行归纳法，我们将证明以下结论，从而将定理 G.3中的状态演化结果转化为定理 G.1。

1. $\widetilde{\mathfrak{a}}_t = \widetilde{a}_t, \widetilde{\mathfrak{b}}_t = \widetilde{b}_t, \mathfrak{a}_{t+1} = a_{t+1}, \mathfrak{b}_{t+1} = b_{t+1}$.
2. .....的联合分布

$$\begin{bmatrix} \mathsf{R}_1 & \cdots & \mathsf{R}_{t+1} & \widetilde{\mathsf{R}}_0 & \cdots & \widetilde{\mathsf{R}}_{t+1} & \mathsf{B}_* & \Theta^\top \end{bmatrix} \qquad \text{(G.17)}$$

与.....的联合分布相同

$$\begin{bmatrix} \mathfrak{b}_1\mathsf{F}_{p,\text{in},2} + \mathsf{P}^1_{\text{in}} & \cdots & \mathfrak{b}_{t+1}\mathsf{F}_{p,\text{in},2} + \mathsf{P}^{t+1}_{\text{in}} & \mathsf{F}^0_{p,\text{in},1} + \widetilde{\mathfrak{b}}_0\mathsf{F}_{p,\text{in},2} & \cdots & \mathsf{F}^{t+1}_{p,\text{in},1} + \widetilde{\mathfrak{b}}_{t+1}\mathsf{F}_{p,\text{in},2} & \mathsf{W}_{p,\text{in},1} & \mathsf{W}^\top_{p,\text{in},-} \end{bmatrix}. \qquad \text{(G.18)}$$

.....的联合分布

$$\begin{bmatrix} \mathsf{P}_1 & \cdots & \mathsf{P}_{t+1} & \widetilde{\mathsf{P}}_0 & \cdots & \widetilde{\mathsf{P}}_{t+1} & \mathsf{Z} & \mathsf{E} & \Xi^\top \end{bmatrix} \qquad \text{(G.19)}$$

与.....的联合分布相同

$$\begin{bmatrix} \mathfrak{a}_1\mathsf{P}_{\text{out},2} + \mathsf{P}^1_{\text{out},1} & \cdots & \mathfrak{a}_{t+1}\mathsf{P}_{\text{out},2} + \mathsf{P}^{t+1}_{\text{out},1} & \mathsf{F}^0_{p,\text{out}} + \widetilde{\mathfrak{a}}_0\mathsf{P}_{\text{out},2} & \cdots & \mathsf{F}^{t+1}_{p,\text{out}} + \widetilde{\mathfrak{a}}_{t+1}\mathsf{P}_{\text{out},2} & \mathsf{P}_{\text{out},2} & \mathsf{W}_{p,\text{out},1} & \mathsf{W}^\top_{p,\text{out},-} \end{bmatrix}. \qquad \text{(G.20)}$$

回想一下 $\mathsf{P}_{\text{out},2} = \mathsf{Z}, \mathsf{F}_{p,\text{in},2} = \mathsf{W}_{p,\text{in},1} = \mathsf{B}_*, \mathsf{W}_{p,\text{out},1} = \mathsf{E}, \mathsf{W}_{p,\text{in},-} = \Theta, \mathsf{W}_{p,\text{out},-} = \Xi$。我们通过以下变量替换来识别随机变量 $(\mathsf{K}_{s+1})_{0\leq s\leq t}, (\mathsf{J}_{s+1})_{0\leq s\leq t}$：

$$\mathsf{K}_{s+1} = \mathsf{P}^{s+1}_{\text{in}}, \qquad \mathsf{J}_{s+1} = \mathsf{P}^{s+1}_{\text{out},1}. \qquad \text{(G.21)}$$

~ ~利用这一点并比较定义 (4.29g) 与 (G.14o) 和 (G.14p)，我们进一步将 $(\mathsf{R}_s)_{0\le s\le t+1},(\mathsf{P}_s)_{0\le s\le t+1}$识别为

$$\widetilde{\mathsf{R}}_s = \mathsf{F}^s_{p,\mathrm{in},1} + \widetilde{\mathfrak{b}}_s \mathsf{B}_*, \qquad \widetilde{\mathsf{P}}_s = \mathsf{F}^s_{p,\mathrm{out}} + \widetilde{\mathfrak{a}}_s \mathsf{Z}. \tag{G.22}$$

为了得出（G.2）中的收敛结果，还需要验证上述等式左侧随机变量的联合分布是否与（4.29）中的描述一致。

~ ~基例。对于 $t=0$，根据定义（G.14e），我们有 $\widetilde{\mathfrak{a}}_0 = \widetilde{a}_0, \mathfrak{b}_0 = b_0$。比较（G.14k）和（4.29f）可得 $\mathfrak{a}_1 = a_1, \mathfrak{b}_1 = \mathfrak{b}_1$。因此，条目 1 在 $t=0$ 处成立。由条目 1 以及（G.14a）和（G.14b），只需验证 $\mathsf{R}_1 = b_1\mathsf{B}_* + \mathsf{P}^1_{\mathrm{in}}, \mathsf{P}_1 = a_1\mathsf{Z} + \mathsf{P}^1_{\mathrm{out}}$,1；结合（G.22）、（G.14o）和（G.14p），并将它们与（4.29g）比较，即可推出条目 2 对 $t=0$ 成立。为此，由（G.14m）和（G.14n）可知，$\mathsf{P}^1_{\mathrm{in}}, \mathsf{P}^1_{\mathrm{out}}$,1 是与 B_*, Z 独立的中心高斯变量。因此，我们只需检查它们的协方差$\Sigma^1_{p,\mathrm{in}}, \Sigma^1_{p,\mathrm{out}}$ 是否分别等于（4.29c）和（4.29d）中给出的 $\mathbb{E}[\mathsf{K}_1^2]$, $\mathbb{E}[\mathsf{J}_1^2]$。我们根据（G.14l）中的定义计算前述量。对于 $\Sigma^1_{p,\mathrm{in}}$，有

$$\begin{aligned}
\Sigma^1_{p,\mathrm{in}} &= \mathbb{E}\big[(\mathsf{F}^0_{q,\mathrm{in}})^2\big] = \mathbb{E}\Big[\Big\{\Phi_0(\Lambda_d^2)(\mathsf{Q}^0_{\mathrm{in},1} + \widetilde{\mathfrak{b}}_0\mathsf{Q}_{\mathrm{in},2}) + \widetilde{\Phi}_0(\Lambda_d)(\mathsf{Q}^0_{\mathrm{out}} + \widetilde{\mathfrak{a}}_0\Lambda_d\mathsf{Q}_{\mathrm{in},2}) - \mathfrak{b}_1\mathsf{Q}_{\mathrm{in},2}\Big\}^2\Big] \\
&= \mathbb{E}\big[\Phi_0(\Lambda_d^2)^2\big]\Sigma^0_{q,\mathrm{in}} + \mathbb{E}\Big[\Big(\widetilde{\mathfrak{b}}_0\Phi_0(\Lambda_d^2) + \widetilde{\mathfrak{a}}_0\widetilde{\Phi}_0(\Lambda_d)\Lambda_d - \mathfrak{b}_1\Big)^2\Big]\rho + \mathbb{E}\Big[\widetilde{\Phi}_0(\Lambda_d)^2\Big]\Sigma^0_{q,\mathrm{out}} \\
&= \mathbb{E}\big[\Phi_0(\Lambda_d^2)^2\big]\mathbb{E}\Big[\widetilde{\mathsf{R}}_0^2\Big] - b_1^2\rho + \mathbb{E}\Big[\widetilde{\Phi}_0(\Lambda_d)^2\Big]\Big(\mathbb{E}\Big[\widetilde{\mathsf{P}}_0^2\Big] - \widetilde{a}_0^2\sigma^2\Big) + \mathbb{E}\Big[\widetilde{\Phi}_0(\Lambda_d)^2\Lambda_d^2\Big]\widetilde{a}_0^2\rho + 2\mathbb{E}\Big[\Phi_0(\Lambda_d^2)\widetilde{\Phi}_0(\Lambda_d)\Lambda_d\Big]\widetilde{a}_0\widetilde{b}_0\rho.
\end{aligned}$$

倒数第二行取自定理 G.3 证明中的（G.47）。最后一行是通过对刚刚证明的 $t=0$ 使用条目 1，代入（G.14c）中的定义 $\Sigma^0_{q,\mathrm{in}} = \mathbb{E}\Big[\widetilde{\mathsf{R}}_0^2\Big] - \widetilde{a}_0^2\rho$ 和 $\Sigma^0_{q,\mathrm{out}} = \mathbb{E}\Big[\widetilde{\mathsf{P}}_0^2\Big] - \widetilde{b}_0\sigma^2$，并化简得到的。所得结果与（4.29d）中的 $\mathbb{E}[\mathsf{K}_1^2]$ 一致，正如所需。对于 $\Sigma^1_{p,\mathrm{out}}$，从（G.48）开始进行类似计算，可得所声称的结果 $\Sigma^1_{p,\mathrm{out}} = \mathbb{E}[\mathsf{J}_1^2]$。这就证明了 $t=0$ 的条目 2。

归纳步骤。假设条目 1 和条目 2 对截至 $t-1$ 的时间均成立，我们验证它们对时间 t 也成立。利用条目 2 中的归纳假设，并比较定义（G.14q）和（4.29e），我们~ ~立即得到 $\widetilde{\mathfrak{a}}_t = \widetilde{a}_t, \mathfrak{b}_t = b_t$。鉴于（G.14k）和（4.29f），这又意味着 $\mathfrak{a}_{t+1} = a_{t+1}, \mathfrak{b}_{t+1} = b_{b+1}$,~，从而证明了条目 1 对时间 t 成立。因此，在下文中，对于任意 $0\le s\le t$，我们将把所有出现的 $\widetilde{\mathfrak{a}}_s, \mathfrak{b}_s, \mathfrak{a}_{s+1}, \mathfrak{b}_{s+1}$~ 替换为 $\widetilde{a}_s, b_s, a_{s+1}, b_{s+1}$。

根据识别关系（G.21）以及其中右侧的定义（G.14m）和（G.14n），我们知道 $(\mathsf{K}_s)_{1\le s\le t+1}$ 和 $(\mathsf{J}_s)_{1\le s\le t+1}$ 是两个相互独立的中心化高斯过程，并且也独立于 Λ_n、Λ_d、B_*、Z、E Θ、Ξ。因此，将（G.17）和（G.19）的概率分布描述（4.29）与（G.18）和（G.20）的概率分布描述（G.14）进行比较可知，要证明时间 t 的条目 2，我们只需考察高斯过程 $(\mathsf{K}_s)_{1\le s\le t+1}$ 和 $(\mathsf{J}_s)_{1\le s\le t+1}$ 的协方差结构。$(\mathsf{J}_s)_{1\le s\le t+1}$ 的协方差如下所示。对于任意 $0\le r$、$s\le t$，

$$\begin{aligned}
&\mathbb{E}[\mathsf{J}_{r+1}\mathsf{J}_{s+1}] = \mathbb{E}\big[\mathsf{P}^{r+1}_{\mathrm{out},1}\mathsf{P}^{s+1}_{\mathrm{out},1}\big] = (\Sigma^{t+1}_{p,\mathrm{out}})_{r+1,s+1} = \mathbb{E}\big[\mathsf{F}^r_{q,\mathrm{out},1}\mathsf{F}^s_{q,\mathrm{out},1}\big] \\
&= \mathbb{E}\Big[\Big(\Psi_r(\Lambda_n^2)(\mathsf{Q}^r_{\mathrm{out}} + \widetilde{a}_r\Lambda_n\mathsf{Q}_{\mathrm{in},2}) + \widetilde{\Psi}_r(\Lambda_n)(\mathsf{Q}^r_{\mathrm{in},1} + \widetilde{b}_r\mathsf{Q}_{\mathrm{in},2}) - a_{r+1}\Lambda_n\mathsf{Q}_{\mathrm{in},2}\Big) \\
&\qquad\cdot\Big(\Psi_s(\Lambda_n^2)(\mathsf{Q}^s_{\mathrm{out}} + \widetilde{a}_s\Lambda_n\mathsf{Q}_{\mathrm{in},2}) + \widetilde{\Psi}_s(\Lambda_n)(\mathsf{Q}^s_{\mathrm{in},1} + \widetilde{b}_s\mathsf{Q}_{\mathrm{in},2}) - a_{s+1}\Lambda_n\mathsf{Q}_{\mathrm{in},2}\Big)\Big] \\
&= \mathbb{E}\big[\Psi_r(\Lambda_n^2)\Psi_s(\Lambda_n^2)(\mathsf{Q}^r_{\mathrm{out}} + \widetilde{a}_r\Lambda_n\mathsf{Q}_{\mathrm{in},2})(\mathsf{Q}^s_{\mathrm{out}} + \widetilde{a}_s\Lambda_n\mathsf{Q}_{\mathrm{in},2})\big] && \text{(G.23)} \\
&\quad + \mathbb{E}\Big[\widetilde{\Psi}_r(\Lambda_n)\widetilde{\Psi}_s(\Lambda_n)\Big]\mathbb{E}\Big[(\mathsf{Q}^r_{\mathrm{in},1} + \widetilde{b}_r\mathsf{Q}_{\mathrm{in},2})(\mathsf{Q}^s_{\mathrm{in},1} + \widetilde{b}_s\mathsf{Q}_{\mathrm{in},2})\Big] + a_{r+1}a_{s+1}\mathbb{E}\big[\Lambda_n^2\big]\mathbb{E}\big[\mathsf{Q}^2_{\mathrm{in},2}\big] && \text{(G.24)} \\
&\quad + \mathbb{E}\Big[\Psi_r(\Lambda_n^2)\widetilde{\Psi}_s(\Lambda_n)(\mathsf{Q}^r_{\mathrm{out}} + \widetilde{a}_r\Lambda_n\mathsf{Q}_{\mathrm{in},2})(\mathsf{Q}^s_{\mathrm{in},1} + \widetilde{b}_s\mathsf{Q}_{\mathrm{in},2})\Big] && \text{(G.25)} \\
&\quad + \mathbb{E}\Big[\Psi_s(\Lambda_n^2)\widetilde{\Psi}_r(\Lambda_n)(\mathsf{Q}^s_{\mathrm{out}} + \widetilde{a}_s\Lambda_n\mathsf{Q}_{\mathrm{in},2})(\mathsf{Q}^r_{\mathrm{in},1} + \widetilde{b}_r\mathsf{Q}_{\mathrm{in},2})\Big] && \text{(G.26)}
\end{aligned}$$

$$- a_{r+1}\Big(\mathbb{E}\big[\Psi_s(\Lambda_n^2)\Lambda_n(\mathsf{Q}_{\text{out}}^s + \widetilde{a}_s\Lambda_n\mathsf{Q}_{\text{in},2})\mathsf{Q}_{\text{in},2}\big] + \mathbb{E}\Big[\widetilde{\Psi}_s(\Lambda_n)\Lambda_n\Big]\mathbb{E}\Big[(\mathsf{Q}_{\text{in},1}^s + \widetilde{b}_s\mathsf{Q}_{\text{in},2})\mathsf{Q}_{\text{in},2}\Big]\Big) \tag{G.27}$$

$$- a_{s+1}\Big(\mathbb{E}\big[\Psi_r(\Lambda_n^2)\Lambda_n(\mathsf{Q}_{\text{out}}^r + \widetilde{a}_r\Lambda_n\mathsf{Q}_{\text{in},2})\mathsf{Q}_{\text{in},2}\big] + \mathbb{E}\Big[\widetilde{\Psi}_r(\Lambda_n)\Lambda_n\Big]\mathbb{E}\Big[(\mathsf{Q}_{\text{in},1}^r + \widetilde{b}_r\mathsf{Q}_{\text{in},2})\mathsf{Q}_{\text{in},2}\Big]\Big), \tag{G.28}$$

其中，上述展示的第二行由定义（G.14i）得出：$\mathsf{F}_{q,\text{ out},1}^t$。我们将分别考察（G.23）至（G.28）中的各项。

我们首先指出若干将在计算中使用的事实：

$$\mathbb{E}[\mathsf{Q}_{\text{out}}^r\mathsf{Q}_{\text{out}}^s] = (\Sigma_{q,\text{out}}^t)_{r,s} = \mathbb{E}[\mathsf{F}_{p,\text{out}}^r\mathsf{F}_{p,\text{out}}^s] = \mathbb{E}\Big[\Big(\widetilde{\mathsf{P}}_r - \widetilde{a}_r\mathsf{Z}\Big)\Big(\widetilde{\mathsf{P}}_s - \widetilde{a}_s\mathsf{Z}\Big)\Big] = \mathbb{E}\Big[\widetilde{\mathsf{P}}_r\widetilde{\mathsf{P}}_s\Big] - \widetilde{a}_r\widetilde{a}_s\sigma^2, \tag{G.29}$$

$$\mathbb{E}[\mathsf{Q}_{\text{in},1}\mathsf{Q}_{\text{in},1}] = (\Sigma_{q,\text{in}}^t)_{r,s} = \mathbb{E}[\mathsf{F}_{p,\text{in},1}^r\mathsf{F}_{p,\text{in},1}^s] = \mathbb{E}\Big[\Big(\widetilde{\mathsf{R}}_r - \widetilde{b}_r\mathsf{B}_*\Big)\Big(\widetilde{\mathsf{R}}_s - \widetilde{b}_s\mathsf{B}_*\Big)\Big] = \mathbb{E}\Big[\widetilde{\mathsf{R}}_r\widetilde{\mathsf{R}}_s\Big] - \widetilde{b}_r\widetilde{b}_s\rho, \tag{G.30}$$

$$\mathbb{E}[\Lambda_n^2]\mathbb{E}[\mathsf{Q}_{\text{in},2}^2] = \overline{m}_2\mathbb{E}[\mathsf{B}_*^2] = \overline{\kappa}_2\rho = \sigma^2. \tag{G.31}$$

在上述推导中，我们使用了关系（G.22）以及以下定义：$\widetilde{a}_t$，$\widetilde{b}_t$ 见（4.29e）。现在我们将（G.23）化简为：

$$\begin{aligned}
&\mathbb{E}\big[\Psi_r(\Lambda_n^2)\Psi_s(\Lambda_n^2)(\mathsf{Q}_{\text{out}}^r + \widetilde{a}_r\Lambda_n\mathsf{Q}_{\text{in},2})(\mathsf{Q}_{\text{out}}^s + \widetilde{a}_s\Lambda_n\mathsf{Q}_{\text{in},2})\big] \\
&= \mathbb{E}\big[\Psi_r(\Lambda_n^2)\Psi_s(\Lambda_n^2)\big]\mathbb{E}[\mathsf{Q}_{\text{out}}^r\mathsf{Q}_{\text{out}}^s] + \widetilde{a}_r\widetilde{a}_s\mathbb{E}\big[\Psi_r(\Lambda_n^2)\Psi_s(\Lambda_n^2)\Lambda_n^2\big]\mathbb{E}\big[\mathsf{Q}_{\text{in},2}^2\big] \\
&= \mathbb{E}\big[\Psi_r(\Lambda_n^2)\Psi_s(\Lambda_n^2)\big]\Big(\mathbb{E}\Big[\widetilde{\mathsf{P}}_r\widetilde{\mathsf{P}}_s\Big] - \widetilde{a}_r\widetilde{a}_s\sigma^2\Big) + \widetilde{a}_r\widetilde{a}_s\mathbb{E}\big[\Psi_r(\Lambda_n^2)\Psi_s(\Lambda_n^2)\Lambda_n^2\big]\rho,
\end{aligned} \tag{G.32}$$

这里我们利用了 $(\mathsf{Q}_{\text{in}}^s)_{0\le s\le t}$ 与 $(\mathsf{Q}_{\text{out}}^s)_{0\le s\le t}$之间的独立性，以及事实（G.29）。可以将（G.24）中的第二个期望化简如下：

$$\mathbb{E}\Big[(\mathsf{Q}_{\text{in},1}^r + \widetilde{b}_r\mathsf{Q}_{\text{in},2})(\mathsf{Q}_{\text{in},1}^s + \widetilde{b}_s\mathsf{Q}_{\text{in},2})\Big] = \mathbb{E}[\mathsf{Q}_{\text{in},1}^r\mathsf{Q}_{\text{in},1}^s] + \widetilde{b}_r\widetilde{b}_s\mathbb{E}[\mathsf{Q}_{\text{in},2}^2] = \mathbb{E}\Big[\widetilde{\mathsf{R}}_r\widetilde{\mathsf{R}}_s\Big], \tag{G.33}$$

这里我们利用了$\mathsf{Q}_{\text{in},2}$ 与 $(\mathsf{Q}_{\text{in},1}^s)_{0\le s\le t}$（见（G.14f））、以及事实（G.30）。接下来转到（G.25）：

$$\begin{aligned}
&\mathbb{E}\Big[\Psi_r(\Lambda_n^2)\widetilde{\Psi}_s(\Lambda_n)(\mathsf{Q}_{\text{out}}^r + \widetilde{a}_r\Lambda_n\mathsf{Q}_{\text{in},2})(\mathsf{Q}_{\text{in},1}^s + \widetilde{b}_s\mathsf{Q}_{\text{in},2})\Big] \\
&= \widetilde{a}_r\mathbb{E}\Big[\Psi_r(\Lambda_n^2)\widetilde{\Psi}_s(\Lambda_n)\Lambda_n\Big]\mathbb{E}\Big[\mathsf{Q}_{\text{in},2}(\mathsf{Q}_{\text{in},1}^s + \widetilde{b}_s\mathsf{Q}_{\text{in},2})\Big] = \widetilde{a}_r\widetilde{b}_s\mathbb{E}\Big[\Psi_r(\Lambda_n^2)\widetilde{\Psi}_s(\Lambda_n)\Lambda_n\Big]\rho.
\end{aligned} \tag{G.34}$$

同理，（G.26）可化简为

$$\mathbb{E}\Big[\Psi_s(\Lambda_n^2)\widetilde{\Psi}_r(\Lambda_n)(\mathsf{Q}_{\text{out}}^s + \widetilde{a}_s\Lambda_n\mathsf{Q}_{\text{in},2})(\mathsf{Q}_{\text{in},1}^r + \widetilde{b}_r\mathsf{Q}_{\text{in},2})\Big] = \widetilde{a}_s\widetilde{b}_r\mathbb{E}\Big[\Psi_s(\Lambda_n^2)\widetilde{\Psi}_r(\Lambda_n)\Lambda_n\Big]\rho. \tag{G.35}$$

最后，考虑（G.28）中括号内的各项：

$$\begin{aligned}
&\mathbb{E}\big[\Psi_r(\Lambda_n^2)\Lambda_n(\mathsf{Q}_{\text{out}}^r + \widetilde{a}_r\Lambda_n\mathsf{Q}_{\text{in},2})\mathsf{Q}_{\text{in},2}\big] + \mathbb{E}\Big[\widetilde{\Psi}_r(\Lambda_n)\Lambda_n\Big]\mathbb{E}\Big[(\mathsf{Q}_{\text{in},1}^r + \widetilde{b}_r\mathsf{Q}_{\text{in},2})\mathsf{Q}_{\text{in},2}\Big] \\
&= \widetilde{a}_r\mathbb{E}\big[\Psi_r(\Lambda_n^2)\Lambda_n^2\big]\mathbb{E}\big[\mathsf{Q}_{\text{in},2}^2\big] + \widetilde{b}_r\mathbb{E}\Big[\widetilde{\Psi}_r(\Lambda_n)\Lambda_n\Big]\mathbb{E}\big[\mathsf{Q}_{\text{in},2}^2\big] \\
&= \frac{\rho}{\delta}\Big(\widetilde{a}_r\mathbb{E}\big[\Psi_r(\Lambda_d^2)\Lambda_d^2\big] + \widetilde{b}_r\mathbb{E}\Big[\widetilde{\Psi}_r(\Lambda_d)\Lambda_d\Big]\Big) = \overline{\kappa}_2\rho a_{r+1} = \sigma^2 a_{r+1},
\end{aligned}$$

其中，第二行成立是因为$\mathsf{Q}_{\text{out}}^r$ 和$\mathsf{Q}_{\text{in},1}^r$ 均独立于$\mathsf{Q}_{\text{in},2}$。同理，（G.27）中括号内的各项可化简为 $\sigma^2 a_{s+1}$。

将所有项综合起来，我们得出结论：$\mathbb{E}[\mathsf{J}_{r+1}\mathsf{J}_{s+1}]$ 等于（4.29c）。
We u类似地，可以用相同步骤检查 $(\mathsf{K}_{r+1})_{0\le r\le t}$ 的协方差结构。对于任意 $0 \le r,\ s \le t$,

$$\begin{aligned}
\mathbb{E}[\mathsf{K}_{r+1}\mathsf{K}_{s+1}] &= \mathbb{E}\big[\mathsf{P}_{\text{in}}^{r+1}\mathsf{P}_{\text{in}}^{s+1}\big] = (\Sigma_{p,\text{in}}^{t+1})_{r+1,s+1} = \mathbb{E}\big[\mathsf{F}_{q,\text{in}}^r\mathsf{F}_{q,\text{in}}^s\big] \\
&= \mathbb{E}\Big[\Big(\Phi_r(\Lambda_d^2)(\mathsf{Q}_{\text{in},1}^r + \widetilde{b}_r\mathsf{Q}_{\text{in},2}) + \widetilde{\Phi}_r(\Lambda_d)(\mathsf{Q}_{\text{out}}^r + \widetilde{a}_r\Lambda_d\mathsf{Q}_{\text{in},2}) - b_{r+1}\mathsf{Q}_{\text{in},2}\Big)
\end{aligned}$$

$$
\begin{aligned}
&\cdot \Big(\Phi_s(\Lambda_d^2)(\mathsf{Q}_{\text{in},1}^s + \widetilde{b}_s \mathsf{Q}_{\text{in},2}) + \widetilde{\Phi}_s(\Lambda_d)(\mathsf{Q}_{\text{out}}^s + \widetilde{a}_s \Lambda_d \mathsf{Q}_{\text{in},2}) - b_{s+1}\mathsf{Q}_{\text{in},2}\Big)\Big] \\
&= \mathbb{E}\big[\Phi_r(\Lambda_d^2)\Phi_s(\Lambda_d^2)\big]\mathbb{E}\Big[(\mathsf{Q}_{\text{in},1}^r + \widetilde{b}_r \mathsf{Q}_{\text{in},2})(\mathsf{Q}_{\text{in},1}^s + \widetilde{b}_s \mathsf{Q}_{\text{in},2})\Big] && \text{(G.36)} \\
&\quad + \mathbb{E}\Big[\widetilde{\Phi}_r(\Lambda_d)\widetilde{\Phi}_s(\Lambda_d)(\mathsf{Q}_{\text{out}}^r + \widetilde{a}_r \Lambda_d \mathsf{Q}_{\text{in},2})(\mathsf{Q}_{\text{out}}^s + \widetilde{a}_s \Lambda_d \mathsf{Q}_{\text{in},2})\Big] + b_{r+1}b_{s+1}\mathbb{E}\big[\mathsf{Q}_{\text{in},2}^2\big] && \text{(G.37)} \\
&\quad + \mathbb{E}\Big[\Phi_r(\Lambda_d^2)\widetilde{\Phi}_s(\Lambda_d)(\mathsf{Q}_{\text{in},1}^r + \widetilde{b}_r \mathsf{Q}_{\text{in},2})(\mathsf{Q}_{\text{out}}^s + \widetilde{a}_s \Lambda_d \mathsf{Q}_{\text{in},2})\Big] && \text{(G.38)} \\
&\quad + \mathbb{E}\Big[\Phi_s(\Lambda_d^2)\widetilde{\Phi}_r(\Lambda_d)(\mathsf{Q}_{\text{in},1}^s + \widetilde{b}_s \mathsf{Q}_{\text{in},2})(\mathsf{Q}_{\text{out}}^r + \widetilde{a}_r \Lambda_d \mathsf{Q}_{\text{in},2})\Big] && \text{(G.39)} \\
&\quad - b_{r+1}\mathbb{E}\Big[\Phi_s(\Lambda_d)(\mathsf{Q}_{\text{out}}^s + \widetilde{a}_s \Lambda_d \mathsf{Q}_{\text{in},2})\mathsf{Q}_{\text{in},2}\Big] - b_{s+1}\mathbb{E}\Big[\Phi_r(\Lambda_d)(\mathsf{Q}_{\text{out}}^r + \widetilde{a}_r \Lambda_d \mathsf{Q}_{\text{in},2})\mathsf{Q}_{\text{in},2}\Big], && \text{(G.40)}
\end{aligned}
$$

其中我们在 Φ_t （见假设（A7））的(G.40)中用到了无迹性质。

我们分别考察(G.36) 至 (G.40) 中的每一项。(G.36) 中的第二个期望已在(G.33)中求得。(G.37) 中的第一项可以仿照(G.32)计算：

$$
\begin{aligned}
&\mathbb{E}\Big[\widetilde{\Phi}_r(\Lambda_d)\widetilde{\Phi}_s(\Lambda_d)(\mathsf{Q}_{\text{out}}^r + \widetilde{a}_r \Lambda_d \mathsf{Q}_{\text{in},2})(\mathsf{Q}_{\text{out}}^s + \widetilde{a}_s \Lambda_d \mathsf{Q}_{\text{in},2})\Big] \\
&= \mathbb{E}\Big[\widetilde{\Phi}_r(\Lambda_d)\widetilde{\Phi}_s(\Lambda_d)\Big]\Big(\mathbb{E}\Big[\widetilde{\mathsf{P}}_r\widetilde{\mathsf{P}}_s\Big] - \widetilde{a}_r\widetilde{a}_s\sigma^2\Big) + \widetilde{a}_r\widetilde{a}_s\mathbb{E}\Big[\widetilde{\Phi}_r(\Lambda_d)\widetilde{\Phi}_s(\Lambda_d)\Lambda_d^2\Big]\rho.
\end{aligned}
$$

类似于(G.38) 和(G.39) 中的各项可以仿照(G.34) 和(G.35)计算：

$$
\begin{aligned}
\mathbb{E}\Big[\Phi_r(\Lambda_d^2)\widetilde{\Phi}_s(\Lambda_d)(\mathsf{Q}_{\text{in},1}^r + \widetilde{b}_r \mathsf{Q}_{\text{in},2})(\mathsf{Q}_{\text{out}}^s + \widetilde{a}_s \Lambda_d \mathsf{Q}_{\text{in},2})\Big] &= \widetilde{a}_s\widetilde{b}_r\mathbb{E}\Big[\Phi_r(\Lambda_d^2)\widetilde{\Phi}_s(\Lambda_d)\Lambda_d\Big]\rho, \\
\mathbb{E}\Big[\Phi_s(\Lambda_d^2)\widetilde{\Phi}_r(\Lambda_d)(\mathsf{Q}_{\text{in},1}^s + \widetilde{b}_s \mathsf{Q}_{\text{in},2})(\mathsf{Q}_{\text{out}}^r + \widetilde{a}_r \Lambda_d \mathsf{Q}_{\text{in},2})\Big] &= \widetilde{a}_r\widetilde{b}_s\mathbb{E}\Big[\Phi_s(\Lambda_d^2)\widetilde{\Phi}_r(\Lambda_d)\Lambda_d\Big]\rho.
\end{aligned}
$$

(G.40) 中的第一个期望计算结果为：

$$
\mathbb{E}\Big[\widetilde{\Phi}_s(\Lambda_d)(\mathsf{Q}_{\text{out}}^s + \widetilde{a}_s \Lambda_d \mathsf{Q}_{\text{in},2})\mathsf{Q}_{\text{in},2}\Big] = \widetilde{a}_s\mathbb{E}\Big[\widetilde{\Phi}_s(\Lambda_d)\Lambda_d\Big]\rho = b_{s+1}\rho,
$$

最后一步根据 b_{s+1} 在(4.29f)中的定义得到。类似地，(G.40) 中的第二个期望为：

$$
\mathbb{E}\Big[\widetilde{\Phi}_r(\Lambda_d)(\mathsf{Q}_{\text{out}}^r + \widetilde{a}_r \Lambda_d \mathsf{Q}_{\text{in},2})\mathsf{Q}_{\text{in},2}\Big] = b_{r+1}\rho.
$$

将所有项代回后，我们得到(4.29d)，这就是 $\mathbb{E}[\mathsf{K}_{r+1}\mathsf{K}_{s+1}]$的所求表达式。 □

定理定理 *G.3*的证明。该证明对 $t \geq 0$采用归纳法。我们将通过归纳法证明以下命题，从而推出定理 G.3：

1. 对每个 $t \geq 0$, $\mathbb{E}\big[\mathsf{F}_{p,\text{ in},2}\mathsf{F}_{p,\text{ in},1}^t\big] = \mathbb{E}\big[\mathsf{F}_{q,\text{ out},2}\mathsf{F}_{q,\text{ out},1}^t\big] = 0$。

2. 对每个 $t \geq 0$, $\Sigma_{q,\text{ in}}^t$, $\Sigma_{q,\text{ out}}^t$, Σ_p^{t+1}，属于 $\Sigma_{p,\text{ out}}^{t+1}$ 均为非奇异的。

3. 对于每个 $t \geq 0$，都存在高斯分布随机变量 $\mathsf{U}_{q,\text{ in}}^t$，$\mathsf{U}_{q,\text{ out}}^t$，它们均值为零且方差严格为正，彼此独立，并且与其他所有变量独立，使得

$$
\begin{aligned}
\mathsf{Q}_{\text{in},1}^t &= \sum_{s=0}^{t-1} \mathsf{Q}_{\text{in},1}^s \Big((\Sigma_{q,\text{in}}^{t-1})^{-1}(\Sigma_{q,\text{in}}^t)_{0:t-1,t}\Big)_s + \mathsf{U}_{q,\text{in}}^t, \\
\mathsf{Q}_{\text{out}}^t &= \sum_{s=0}^{t-1} \mathsf{Q}_{\text{out}}^s \big((\Sigma_{q,\text{out}}^{t-1})^{-1}(\Sigma_{q,\text{out}}^t)_{0:t-1,t}\big)_s + \mathsf{U}_{q,\text{out}}^t.
\end{aligned} \tag{G.41}
$$

存在高斯分布随机变量 U_p^{t+1}，$_{\text{in}}\mathsf{U}_p^{t+1}$，$_{\text{out}}$，它们均值为零且方差严格为正，彼此独立，并且与其他所有变量独立，使得

$$\begin{aligned}\mathsf{P}_{\text{in}}^{t+1} &= \sum_{s=0}^{t-1} \mathsf{P}_{\text{in}}^{s+1}\Big((\Sigma_{p,\text{in}}^{t})^{-1}(\Sigma_{p,\text{in}}^{t+1})_{1:t,t+1}\Big)_{s+1} + \mathsf{U}_{p,\text{in}}^{t+1},\\ \mathsf{P}_{\text{out},1}^{t+1} &= \sum_{s=0}^{t-1} \mathsf{P}_{\text{out},1}^{s+1}\big((\Sigma_{p,\text{out}}^{t})^{-1}(\Sigma_{p,\text{out}}^{t+1})_{1:t,t+1}\big)_{s+1} + \mathsf{U}_{p,\text{out}}^{t+1}.\end{aligned} \tag{G.42}$$

4. (G.15)和 (G.16) 成立。

基例。 我们取 $t=0$ 作为基例，并验证 条目 1 至 4 对于 $t=0$的有效性。

根据 $\mathfrak{b}_0$ 在 (G.14e)中的选取，我们有 $\mathbb{E}\big[\mathsf{F}_{p,\text{in},2}\mathsf{F}_{p,\text{in},1}\big] = \mathbb{E}\big[(\mathsf{R}_0 - \mathfrak{b}_0\mathsf{B}_*)\mathsf{B}_*\big] = 0$ 这证明了条目 1中的第一项陈述。

条目 2 中$\Sigma_{q,\text{in}}^0$、$\Sigma_{q,\text{out}}^0 \in \mathbb{R}$ 的 非奇异性 可归结为 $t=0$处的正性，这由 假设 (A28)推出，正如 (G.14d)中所述。

现在我们验证 $t=0$的(G.15) 和(G.16) 条目 4 。回顾一下，$f_{p,\text{in},2} = w_{p,\text{in},1} = \beta_*$。联合收敛

$$\begin{bmatrix} f_{p,\text{in},2} & f_{p,\text{in},1}^0 & w_{p,\text{in}} & w_{q,\text{in}}\end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} f_{p,\text{out}}^0 & w_{p,\text{out}} & w_{q,\text{out}}\end{bmatrix} \tag{G.43}$$

由初始条件 (G.14)成立。接下来，我们通过依次考察更新 (G.5a) 至 (G.5d) （对 $t=0$），证明 (G.5a) 至 (G.5d) 中涉及的迭代量联合收敛。第一个更新(G.5a) 生成

$$q_{\text{in}}^0 = \begin{bmatrix} Q^\top(\tilde{r}^0 - \tilde{\mathfrak{b}}_0\beta_*) & Q^\top\beta_*\end{bmatrix} \qquad \text{and} \qquad q_{\text{out}}^0 = O^\top(\tilde{p}^0 - \tilde{\mathfrak{a}}_0 z) = O^\top\tilde{p}^0 - \tilde{\mathfrak{a}}_0\Lambda Q^\top\beta_*$$

其与 (G.43) 中的向量联合收敛，这由 定理 K.5保证：

$$\begin{aligned}\begin{bmatrix} f_{p,\text{in},2} & f_{p,\text{in},1}^0 & w_{p,\text{in}} & w_{q,\text{in}} & q_{\text{in},2} & q_{\text{in},1}^0\end{bmatrix} &\xrightarrow{W} \begin{bmatrix} \mathsf{F}_{p,\text{in},2} & \mathsf{F}_{p,\text{in},1}^0 & \mathsf{W}_{p,\text{in}}^\top & \mathsf{W}_{q,\text{in}}^\top & \mathsf{Q}_{\text{in},2} & \mathsf{Q}_{\text{in},1}^0\end{bmatrix},\\ \begin{bmatrix} f_{p,\text{out}}^0 & w_{p,\text{out}} & w_{q,\text{out}} & q_{\text{out}}^0\end{bmatrix} &\xrightarrow{W} \begin{bmatrix} \mathsf{F}_{p,\text{out}}^0 & \mathsf{W}_{p,\text{out}}^\top & \mathsf{W}_{q,\text{out}}^\top & \mathsf{Q}_{\text{out}}^0\end{bmatrix},\end{aligned} \tag{G.44}$$

其中

$$\begin{bmatrix}\mathsf{Q}_{\text{in},2}\\ \mathsf{Q}_{\text{in},1}^0\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\rho & 0\\ 0 & \Sigma_{q,\text{in}}^0\end{bmatrix}\right) \tag{G.45}$$

与 $\mathsf{F}_{p,\text{in},2}$、$\mathsf{F}_{p,\text{in},1}^0$、$\mathsf{W}_{p,\text{in}}$、$\mathsf{W}_{q,\text{in}}$ 独立，而 $\mathsf{Q}_{\text{out}}^0 \sim \mathcal{N}(0, \Sigma_{q,\text{out}}^0)$ 与 $\mathsf{F}_{p,\text{out}}^0$、$\mathsf{W}_{p,\text{out}}$、$\mathsf{W}_{q,\text{out}}$ 独立。由 $\Sigma_{q,\text{in}}^0$ 和 ρ 的正性（见 (G.14d) 和假设（A1））， $\begin{bmatrix}\mathsf{Q}_{\text{in},2} & \mathsf{Q}_{\text{in},1}^0\end{bmatrix}^\top$ 的协方差矩阵严格正定。注意， $\begin{bmatrix}\mathsf{Q}_{\text{in},2} & \mathsf{Q}_{\text{in},1}^0\end{bmatrix}^\top$ 与 $\mathsf{Q}_{\text{out}}^0$ 也彼此独立。这就证明了 $t=0$ 的条目 3 中的 (G.41)。

第二次更新 (G.5b) 生成

$$\begin{aligned} f_{q,\text{in}}^0 &= \text{diag}(\phi_0)Q^\top\tilde{r}^0 + \text{diag}_{n\times d}(\tilde{\phi}_0)^\top O^\top\tilde{p}^0 - \mathfrak{b}_1 Q^\top\beta_*,\\ f_{q,\text{out}}^0 &= \begin{bmatrix}\text{diag}(\psi_0)O^\top\tilde{p}^0 + \text{diag}_{n\times d}(\tilde{\psi}_0)Q^\top\tilde{r}^0 - \mathfrak{a}_1\Lambda Q^\top\beta_* & \Lambda Q^\top\beta_*\end{bmatrix},\end{aligned}$$

其与 (G.44) 中的向量的联合收敛由 定理 K.2 和 K. 5,

$$\begin{aligned}&\begin{bmatrix} f_{p,\text{in},2} & f_{p,\text{in},1}^0 & w_{p,\text{in}} & w_{q,\text{in}} & q_{\text{in},2} & q_{\text{in},1}^0 & f_{q,\text{in}}^0\end{bmatrix}\\ &\quad\xrightarrow{W} \begin{bmatrix} \mathsf{F}_{p,\text{in},2} & \mathsf{F}_{p,\text{in},1}^0 & \mathsf{W}_{p,\text{in}}^\top & \mathsf{W}_{q,\text{in}}^\top & \mathsf{Q}_{\text{in},2} & \mathsf{Q}_{\text{in},1}^0 & \mathsf{F}_{q,\text{in}}^0\end{bmatrix},\end{aligned} \tag{G.46a}$$

$$\begin{aligned}&\left[f^0_{p,\text{out}} \quad w_{p,\text{out}} \quad w_{q,\text{out}} \quad q^0_{\text{out}} \quad f_{q,\text{out},2} \quad f^0_{q,\text{out},1}\right]\\ &\xrightarrow{W}\left[\mathsf{F}^0_{p,\text{out}} \quad \mathsf{W}^\top_{p,\text{out}} \quad \mathsf{W}^\top_{q,\text{out}} \quad \mathsf{Q}^0_{\text{out}} \quad \mathsf{F}_{q,\text{out},2} \quad \mathsf{F}^0_{q,\text{out},1}\right],\end{aligned} \tag{G.46b}$$

其中 $\mathsf{F}^0_{q,\ \text{in}}$，$\mathsf{F}^0_{q,\ \text{out},\ 1}$，$\mathsf{F}_{q,\ \text{out},\ 2}$ 分别在 (G.14h) 至 (G.14j) 中定义。现在可以根据 (G.14l) 计算 $\Sigma^1_{p,\ \text{in}}$，$\Sigma^1_{p,\ \text{out}} \in \mathbb{R}$。我们将论证它们必定严格为正，从而证明 $t=0$ 满足 条目 2 。对于 $\Sigma^1_{p,\ \text{in}}$，我们将其写为

$$\begin{aligned}\Sigma^1_{p,\text{in}} &= \mathbb{E}\left[(\mathsf{F}^0_{q,\text{in}})^2\right] = \mathbb{E}\left[\left\{\Phi_0(\mathsf{Q}^0_{\text{in},1} + \widetilde{\mathfrak{b}}_0\mathsf{Q}_{\text{in},2}) + \widetilde{\Phi}_0(\mathsf{Q}^0_{\text{out}} + \widetilde{\mathfrak{a}}_0\Lambda_d\mathsf{Q}_{\text{in},2}) - \mathfrak{b}_1\mathsf{Q}_{\text{in},2}\right\}^2\right]\\ &= \mathbb{E}\left[\left\{\Phi_0\mathsf{Q}^0_{\text{in},1} + \left(\widetilde{\mathfrak{b}}_0\Phi_0 + \widetilde{\mathfrak{a}}_0\widetilde{\Phi}_0\Lambda_d - \mathfrak{b}_1\right)\mathsf{Q}_{\text{in},2} + \widetilde{\Phi}_0\mathsf{Q}^0_{\text{out}}\right\}^2\right]\\ &= \mathbb{E}\left[\Phi_0^2\right]\Sigma^0_{q,\text{in}} + \mathbb{E}\left[\left(\widetilde{\mathfrak{b}}_0\Phi_0 + \widetilde{\mathfrak{a}}_0\widetilde{\Phi}_0\Lambda_d - \widetilde{\mathfrak{a}}_0\mathbb{E}\left[\widetilde{\Phi}_0\Lambda_d\right]\right)^2\right]\rho + \mathbb{E}\left[\widetilde{\Phi}_0^2\right]\Sigma^0_{q,\text{out}},\end{aligned} \tag{G.47}$$

由于上文首项和末项中至少有一项为正，因此可以看出其为正，这由 (G.14d) 和假设（A29）可知。类似地，

$$\Sigma^1_{p,\text{out}} = \mathbb{E}\left[\Psi_0^2\right]\Sigma^0_{q,\text{out}} + \mathbb{E}\left[\left(\widetilde{\mathfrak{a}}_0\Psi_0\Lambda_n + \widetilde{\mathfrak{b}}_0\widetilde{\Psi}_0 - \mathfrak{a}_1\Lambda_n\right)^2\right]\rho + \mathbb{E}\left[\widetilde{\Psi}_0^2\right]\Sigma^0_{q,\text{in}} \tag{G.48}$$

可由 $\Sigma^0_{q、\ \text{out}}$、$\Sigma^0_{q、\ \text{in}}$ 为正，且 $\mathbb{E}\left[\Psi_0^2\right]$、$\mathbb{E}\left[\widetilde{\Psi}_0^2\right]$ 中至少有一个为正推出。

接下来，第三次更新 (G.5c) 生成

$$\begin{aligned}p^1_{\text{in}} &= Q\,\text{diag}(\phi_0)Q^\top\tilde{r}^0 + Q\,\text{diag}_{n\times d}(\widetilde{\phi}_0)^\top O^\top\tilde{p}^0 - \mathfrak{b}_1\beta_* = r^1 - \mathfrak{b}_1\beta_*,\\ p^1_{\text{out}} &= \left[O\,\text{diag}(\psi_0)O^\top\tilde{p}^0 + O\,\text{diag}_{n\times d}(\widetilde{\psi}_0)Q^\top\tilde{r}^0 - \mathfrak{a}_1 O\Lambda Q^\top\beta_* \quad O\Lambda Q^\top\beta_*\right] = \left[p^1 - \mathfrak{a}_1 z \quad z\right].\end{aligned}$$

为了理解它们的联合收敛，我们对 (G.46) 左侧的所有向量进行条件化。于是，Q、O 的概率分布分别以事件 $Qq^0_{\text{in}} = f^0_{p、\ \text{in}}$、$Oq^0_{\text{out}} = f^0_{p、\ \text{out}}$为条件。(G.44) 中 q^0_{in}和 q^0_{out} 的 Wasserstein 收敛意味着 $d^{-1}(q^0_{\text{in}})^\top q^0_{\text{in}} \to \begin{bmatrix}\Sigma^0_{q,\ \text{in}} & 0\\ 0 & \rho\end{bmatrix} \succ 0_{2\times 2}$（见 (G.45)）以及 $n^{-1}(q^0_{\text{out}})^\top q^0_{\text{out}} \to \Sigma^0_{q,\text{out}} > 0$（见 (G.14d)）。特别地，当 n, d充分大时，左侧均可逆。利用定理 K.6，当 n, d充分大时，我们可以将 $p^1_{\text{in}}, p^1_{\text{out}}$的条件分布写为

$$\begin{aligned}p^1_{\text{in}} \,\big|\, \{Qq^0_{\text{in}} = f^0_{p,\text{in}}\} &= Qf^0_{q,\text{in}} \,\big|\, \{Qq^0_{\text{in}} = f^0_{p,\text{in}}\} \overset{d}{=} f^0_{p,\text{in}}((q^0_{\text{in}})^\top q^0_{\text{in}})^{-1}(q^0_{\text{in}})^\top f^0_{q,\text{in}} + \Pi_{(f^0_{p,\text{in}})^\perp}\widetilde{Q}\Pi^\top_{(q^0_{\text{in}})^\perp}f^0_{q,\text{in}},\\ p^1_{\text{out}} \,\big|\, \{Oq^0_{\text{out}} = f^0_{p,\text{out}}\} &= Of^0_{q,\text{out}} \,\big|\, \{Oq^0_{\text{out}} = f^0_{p,\text{out}}\} \overset{d}{=} f^0_{p,\text{out}}((q^0_{\text{out}})^\top q^0_{\text{out}})^{-1}(q^0_{\text{out}})^\top f^0_{q,\text{out}} + \Pi_{(f^0_{p,\text{out}})^\perp}\widetilde{O}\Pi^\top_{(q^0_{\text{out}})^\perp}f^0_{q,\text{out}},\end{aligned} \tag{G.49}$$

其中 $(Q,O) \sim \text{Haar}(\mathbb{O}(d-2)) \otimes \text{Haar}(\mathbb{O}(n-1))$ 与其他所有变量相互独立。我们声称

$$\frac{1}{d}(q^0_{\text{in}})^\top f^0_{q,\text{in}} \to \begin{bmatrix}0\\ 0\end{bmatrix}, \qquad \frac{1}{n}(q^0_{\text{out}})^\top f^0_{q,\text{out}} \to \begin{bmatrix}0 & 0\end{bmatrix}, \tag{G.50}$$

逐项。为证明(G.50)中的第一个结论，利用联合收敛结果(G.46)、Φ_0 的零均值以及 $\mathfrak{b}_1$ 在(G.14k)中的定义，我们有：

$$\begin{aligned}\frac{1}{d}q^\top_{\text{in},2}f^0_{q,\text{in}} &\to \mathbb{E}\left[\mathsf{Q}_{\text{in},2}\left\{\Phi_0(\mathsf{Q}^0_{\text{in},1} + \widetilde{\mathfrak{b}}_0\mathsf{Q}_{\text{in},2}) + \widetilde{\Phi}_0(\mathsf{Q}^0_{\text{out}} + \widetilde{\mathfrak{a}}_0\Lambda_d\mathsf{Q}_{\text{in},2}) - \mathfrak{b}_1\mathsf{Q}_{\text{in},2}\right\}\right]\\ &= \widetilde{\mathfrak{b}}_0\mathbb{E}[\Phi_0]\mathbb{E}\left[\mathsf{Q}^2_{\text{in},2}\right] + \widetilde{\mathfrak{a}}_0\mathbb{E}\left[\widetilde{\Phi}_0\Lambda_d\right]\mathbb{E}\left[\mathsf{Q}^2_{\text{in},2}\right] - \mathfrak{b}_1\mathbb{E}\left[\mathsf{Q}^2_{\text{in},2}\right]\\ &= \rho\left\{\widetilde{\mathfrak{a}}_0\mathbb{E}\left[\widetilde{\Phi}_0\Lambda_d\right] - \mathfrak{b}_1\right\} = 0,\end{aligned} \tag{G.51}$$

$$\frac{1}{d}(q^0_{\text{in},1})^\top f^0_{q,\text{in}} \to \mathbb{E}\Big[\mathsf{Q}^0_{\text{in},1}\Big\{\Phi_0(\mathsf{Q}^0_{\text{in},1}+\widetilde{\mathfrak{b}}_0\mathsf{Q}_{\text{in},2})+\widetilde{\Phi}_0(\mathsf{Q}^0_{\text{out}}+\widetilde{\mathfrak{a}}_0\Lambda_d\mathsf{Q}_{\text{in},2})-\mathfrak{b}_1\mathsf{Q}_{\text{in},2}\Big\}\Big]$$
$$= \mathbb{E}[\Phi_0]\mathbb{E}\big[(\mathsf{Q}^0_{\text{in},1})^2\big] = 0. \tag{G.52}$$

利用联合收敛结果(G.46)和 Ψ_0 的零均值，可以类似地证明(G.50) 中的第二个结论。因此，根据定理 K.4 和K.5，由于(G.49)右侧的第二项，$p^1_{\text{in}} \,|\, \{Qq^0_{\text{in}} = f^0_{p,\text{在}}\}$ 和 $p^1_{\text{out}} \,|\, \{Oq^0_{\text{out}} = f^0_{p,\text{out}}\}$具有高斯极限，分别记为 $\mathsf{U}^1_{p,\ \text{in}} \in \mathbb{R}$ 和 $(\mathsf{U}^1_{p,\ \text{out}})^\top = \big[\mathsf{U}^1_{p,\ \text{out},\ 1}\ \mathsf{U}_{p,\ \text{out},\ 2}\big] \in \mathbb{R}^2$，且它们与其他所有项独立。用公式表示，

$$\begin{aligned}
&\big[f_{p,\text{in},2}\quad f^0_{p,\text{in},1}\quad w_{p,\text{in}}\quad w_{q,\text{in}}\quad q_{\text{in},2}\quad q^0_{\text{in},1}\quad f^0_{q,\text{in}}\quad p^1_{\text{in}}\big]\\
&\quad\xrightarrow{W}\big[\mathsf{F}_{p,\text{in},2}\quad \mathsf{F}^0_{p,\text{in},1}\quad \mathsf{W}^\top_{p,\text{in}}\quad \mathsf{W}^\top_{q,\text{in}}\quad \mathsf{Q}_{\text{in},2}\quad \mathsf{Q}^0_{\text{in},1}\quad \mathsf{F}^0_{q,\text{in}}\quad \mathsf{U}^1_{p,\text{in}}\big],\\
&\big[f^0_{p,\text{out}}\quad w_{p,\text{out}}\quad w_{q,\text{out}}\quad q^0_{\text{out}}\quad f_{q,\text{out},2}\quad f^0_{q,\text{out},1}\quad p^1_{\text{out},1}\quad p_{\text{out},2}\big]\\
&\quad\xrightarrow{W}\big[\mathsf{F}^0_{p,\text{out}}\quad \mathsf{W}^\top_{p,\text{out}}\quad \mathsf{W}^\top_{q,\text{out}}\quad \mathsf{Q}^0_{\text{out}}\quad \mathsf{F}_{q,\text{out},2}\quad \mathsf{F}^0_{q,\text{out},1}\quad \mathsf{U}^1_{p,\text{out},1}\quad \mathsf{U}_{p,\text{out},2}\big].
\end{aligned} \tag{G.53}$$

接下来，我们通过验证前述随机变量与(G.53) 右侧其他随机变量的联合分布符合(G.14m) 和(G.14n)中的定义，来说明 $\mathsf{U}^1_{p,\text{in}},\mathsf{U}^1_{p,\text{out},1},\mathsf{U}_{p,\text{out},2}$ 恰好就是 $\mathsf{P}^1_{\text{in}},\mathsf{P}^1_{\text{out},1},\mathsf{P}_{\text{out},2}$。根据(G.50)中的第一个断言，$\mathsf{U}^1_{p,\text{in}}$ 服从方差为下式的高斯分布：

$$\lim_{d\to\infty}\frac{1}{d}\|p^1_{\text{in}}\|_2^2 = \lim_{d\to\infty}\frac{1}{d}\|f^0_{q,\text{in}}\|_2^2 = \Sigma^1_{p,\text{in}},$$

它与(G.53)右侧的所有其他随机变量相互独立。类似地，根据(G.50)中的第二个断言以及(G.46)的收敛结果，$(\mathsf{U}^1_{p,\text{out},1},\mathsf{U}_{p,\text{out},2})$ 服从协方差如下的联合高斯分布：

$$\lim_{n\to\infty}\frac{1}{n}(p^1_{\text{out}})^\top p^1_{\text{out}} = \lim_{n\to\infty}\frac{1}{n}(f^0_{q,\text{out}})^\top f^0_{q,\text{out}} = \begin{bmatrix}\Sigma^1_{p,\text{out}} & 0\\ 0 & \sigma^2\end{bmatrix},$$

它与(G.53)右侧的所有其他随机变量相互独立。上方协方差矩阵右下角的元素是由 $\mathbb{E}\big[(\mathsf{F}_{q,\text{out},2})^2\big] = \mathbb{E}\big[\Lambda_n^2\big]\mathbb{E}\big[\mathsf{Q}^2_{\text{in},2}\big] = \kappa_2\rho = \sigma^2$ 得出的。非对角元素则由下式得出：

$$\begin{aligned}
\mathbb{E}\big[\mathsf{F}_{q,\text{out},2}\mathsf{F}^0_{q,\text{out},1}\big] &= \mathbb{E}\big[\Psi_0\Lambda_n\mathsf{Q}_{\text{in},2}(\mathsf{Q}^0_{\text{out}}+\widetilde{\mathfrak{a}}_0\Lambda_n\mathsf{Q}_{\text{in},2})\big] + \mathbb{E}\Big[\widetilde{\Psi}_0\Lambda_n\mathsf{Q}_{\text{in},2}(\mathsf{Q}^0_{\text{in},1}+\widetilde{\mathfrak{b}}_0\mathsf{Q}_{\text{in},2})\Big] - \mathfrak{a}_1\mathbb{E}\big[\Lambda_n^2\mathsf{Q}^2_{\text{in},2}\big]\\
&= \widetilde{\mathfrak{a}}_0\mathbb{E}\big[\Psi_0\Lambda_n^2\big]\rho + \widetilde{\mathfrak{b}}_0\mathbb{E}\Big[\widetilde{\Psi}_0\Lambda_n\Big]\rho - \mathfrak{a}_1\overline{\kappa}_2\rho = 0,
\end{aligned} \tag{G.54}$$

$\widetilde{\ }$以及 $\widetilde{\mathfrak{a}}_0$、$\mathfrak{b}_0$ 在 (G.14e) 中的取值，以及 $\mathfrak{a}_1$ 在 (G.14k)中的取值。这证明了 (G.42) 在 条目 3 中对于 $t=0$成立。此外，上文的 (G.54)证明了 条目 1 中对于 $t=0$的第二个结论。

最后，第四次更新 (G.5d) 生成

$$f^1_{p,\text{in}} = \Big[f_1(r^1;\beta_*,\Theta) - \widetilde{\mathfrak{b}}_1\beta_* \quad \beta_*\Big] \qquad \text{and} \qquad f^1_{p,\text{out}} = g_1(p^1;z,\varepsilon,\Xi) - \widetilde{\mathfrak{a}}_1 z,$$

其与此前所有迭代量的联合收敛性由定理 K.2 和K.5保证。用公式表示，我们得到所需结果：

$$\begin{aligned}
&\big[q_{\text{in},2}\quad q^0_{\text{in},1}\quad f^0_{q,\text{in}}\quad p^1_{\text{in}}\quad f_{p,\text{in},2}\quad f^0_{p,\text{in},1}\quad f^1_{p,\text{in},1}\quad w_{p,\text{in}}\quad w_{q,\text{in}}\big]\\
&\quad\xrightarrow{W}\big[\mathsf{Q}_{\text{in},2}\quad \mathsf{Q}^0_{\text{in},1}\quad \mathsf{F}^0_{q,\text{in}}\quad \mathsf{P}^1_{\text{in}}\quad \mathsf{F}_{p,\text{in},2}\quad \mathsf{F}^0_{p,\text{in},1}\quad \mathsf{F}^1_{p,\text{in},1}\quad \mathsf{W}^\top_{p,\text{in}}\quad \mathsf{W}^\top_{q,\text{in}}\big]\\
&\big[q^0_{\text{out}}\quad f_{q,\text{out},2}\quad f^0_{q,\text{out},1}\quad p_{\text{out},2}\quad p^1_{\text{out},1}\quad f^0_{p,\text{out}}\quad f^1_{p,\text{out}}\quad w_{p,\text{out}}\quad w_{q,\text{out}}\big]\\
&\quad\xrightarrow{W}\big[\mathsf{Q}^0_{\text{out}}\quad \mathsf{F}_{q,\text{out},2}\quad \mathsf{F}^0_{q,\text{out},1}\quad \mathsf{P}_{\text{out},2}\quad \mathsf{P}^1_{\text{out},1}\quad \mathsf{F}^0_{p,\text{out}}\quad \mathsf{F}^1_{p,\text{out}}\quad \mathsf{W}^\top_{p,\text{out}}\quad \mathsf{W}^\top_{q,\text{out}}\big].
\end{aligned} \tag{G.55}$$

归纳步骤。现在假设条目 1 至4 在截至 $t \geq 0$的时刻均成立，我们来验证它们在时刻 $t+1$也成立。

该陈述 $\mathbb{E}\left[\mathsf{F}_{p}^{t+1}{}_{,\mathrm{in},1}\mathsf{F}_{p,\mathrm{in},2}\right]=0$ 在 条目 1 （对于时刻 $t+1$）成立，因为

$$\mathbb{E}\left[\mathsf{F}_{p,\mathrm{in},1}^{t+1}\mathsf{F}_{p,\mathrm{in},2}\right]=\mathbb{E}\left[f_{t+1}(\mathsf{P}_{\mathrm{in}}^{t+1}+\mathfrak{b}_{t+1}\mathsf{B}_{*};\mathsf{B}_{*},\Theta)\mathsf{B}_{*}\right]-\widetilde{\mathfrak{b}}_{t+1}\rho=0, \tag{G.56}$$

根据 $\mathsf{F}_{p}^{t+1}{}_{,\mathrm{in},1}$ 在 (G.14o) 中的定义以及 $\widetilde{\mathfrak{b}}_{t+1}$ 在 (G.14q)中的定义。

在时刻 t 的更新之后，协方差矩阵 $\Sigma_{p}^{t+1}{}_{,\mathrm{in}}$ $\Sigma_{p}^{t+1}{}_{,\mathrm{out}}$, $\Sigma_{q}^{t+1}{}_{,\mathrm{in}}$ $\Sigma_{q}^{t+1}{}_{,\mathrm{out}}$ 已经计算完毕。我们将前两个写成如下分块形式:

$$\Sigma_{q,\mathrm{in}}^{t+1}=\begin{bmatrix}\Sigma_{q,\mathrm{in}}^{t} & \sigma_{q,\mathrm{in}}^{t}\\ (\sigma_{q,\mathrm{in}}^{t})^{\top} & s_{q,\mathrm{in}}^{t+1}\end{bmatrix}\in\mathbb{R}^{(t+2)\times(t+2)},\qquad \Sigma_{q,\mathrm{out}}^{t+1}=\begin{bmatrix}\Sigma_{q,\mathrm{out}}^{t} & \sigma_{q,\mathrm{out}}^{t}\\ (\sigma_{q,\mathrm{out}}^{t})^{\top} & s_{q,\mathrm{out}}^{t+1}\end{bmatrix}\in\mathbb{R}^{(t+2)\times(t+2)}, \tag{G.57}$$

其中我们引入了记号 $\sigma_{q}^{t}{}_{,\mathrm{in}}$、$\sigma_{q}^{t}{}_{,\mathrm{out}}\in\mathbb{R}^{t+1}$ 以及 $s_{q}^{t+1}{}_{,\mathrm{in}}$ $s_{q}^{t+1}{}_{,\mathrm{out}}\in\mathbb{R}$。在记号上，我们用集合 $\{0,1,\cdots,t\}$ 来索引 $\sigma_{q}^{t}{}_{,\mathrm{in}}$、$\sigma_{q}^{t}{}_{,\mathrm{out}}$ 的元素，这与 $\Sigma_{q}^{t+1}{}_{,\mathrm{in}}$ $\Sigma_{q}^{t+1}{}_{,\mathrm{out}}$ 的索引方式一致。我们验证 $\Sigma_{q}^{t+1}{}_{,\mathrm{in}}$ 是严格正定的。回顾定义 (G.14r)，鉴于由条目 2 中的归纳假设可知 $\Sigma_{q}^{t}{}_{,\mathrm{in}}$ 是非奇异的，$\Sigma_{q}^{t+1}{}_{,\mathrm{in}}$ 是奇异的当且仅当存在系数 α_0、$\cdots$、$\alpha_t\in\mathbb{R}$，使得几乎必然，

$$\mathsf{F}_{p,\mathrm{in},1}^{t+1}=\sum_{s=0}^{t}\alpha_{s}\mathsf{F}_{p,\mathrm{in},1}^{s}.$$

然而，根据假设（A30），这不可能发生。因此$\Sigma_{q}^{t+1}{}_{,\mathrm{in}}$的可逆。类似的论证也适用于$\Sigma_{q}^{t+1}{}_{,\mathrm{out}}$。这证明了时间条目 2 的部分内容，其中 $t+1$。

对 q^{t+1} 中的进行归纳。我们将一些过去的迭代量收集成矩阵 U 和 V:

$$U=\begin{bmatrix}q_{\mathrm{in},2} & q_{\mathrm{in},1}^{0} & \cdots & q_{\mathrm{in},1}^{t} & f_{q,\mathrm{in}}^{0} & \cdots & f_{q,\mathrm{in}}^{t}\end{bmatrix}\in\mathbb{R}^{d\times(1+2(t+1))}, \tag{G.58}$$

$$V=\begin{bmatrix}f_{p,\mathrm{in},2} & f_{p,\mathrm{in},1}^{0} & \cdots f_{p,\mathrm{in},1}^{t} & p_{\mathrm{in}}^{1} & \cdots & p_{\mathrm{in}}^{t+1}\end{bmatrix}\in\mathbb{R}^{d\times(1+2(t+1))}. \tag{G.59}$$

根据抽象 VAMP 迭代的更新规则(G.5)，给定 $w_{p,\mathrm{in}}$、$w_{p,\mathrm{out}}$、$w_{q,\mathrm{in}}$、$w_{q,\mathrm{out}}$、$(q_{\mathrm{in}}^{s})_{0\leq s\leq t}$、$(q_{\mathrm{out}}^{s})_{0\leq s\leq t}$、$(f_{q,\mathrm{in}}^{s})_{0\leq s\leq t}$、$(f_{q,\mathrm{out}}^{s})_{0\leq s\leq t}$、$(p^{s+1}$中的$)_{0\leq s\leq t}$、$(p_{\mathrm{out}}^{s+1})_{0\leq s\leq t}$、$(f_{p,}^{s}$中的$)_{0\leq s\leq t+1}$时，$q^{t+1}$中的的分布 $(f_{p,\mathrm{out}}^{s})_{0\leq s\leq t+1}$ 与在事件 $U=Q^{\top}V$条件下 q^{t+1} 中的的概率分布相同。这里我们利用了 Q 的正交性，这意味着由 (G.5c) 中的第一次更新给出的条件 p^{s+1} 中的$=Qf_{q,\mathrm{in}}^{s}$ 与 $Q^{\top}p^{s+1}$ 中的$=f_{q,\mathrm{in}}^{s}$ 相同，对所有 $0\leq s\leq t$均成立。对于所有足够大的 d，后一个条件分布等于

$$q_{\mathrm{in}}^{t+1}\,\big|\,\{U=Q^{\top}V\}=Q^{\top}f_{p,\mathrm{in}}^{t+1}\,\Big|\,\{U=Q^{\top}V\}\overset{d}{=}U(U^{\top}U)^{-1}V^{\top}f_{p,\mathrm{in}}^{t+1}+\Pi_{U^{\perp}}\widetilde{Q}^{\top}\Pi_{V^{\perp}}^{\top}f_{p,\mathrm{in}}^{t+1}\in\mathbb{R}^{d\times 2}, \tag{G.60}$$

其中，最后一步由定理 K.6 得出（其中包含投影算子 $\Pi_{U^{\perp}},\Pi_{V^{\perp}}$ 的定义），且 $\widetilde{Q}\sim\mathrm{Haar}(\mathbb{O}(d-(1+2(t+1))))$ 与其他所有内容均独立。我们声称，上述显示式中的 $U^{\top}U$ 对所有充分大的 d 都可逆，并将在 (G.69) 中证明这一点。（我们指出，在归纳证明的不同阶段，记号 U,V,O,Q 会被重复使用。）为了分析(G.60)右侧的第一项，我们计算 $d^{-1}U^{\top}U$ 和 $d^{-1}V^{\top}f_{p}^{t+1}{}_{,\text{中},1}$ 的极限。

让我们从计算 $d^{-1}U^{\top}U$的极限开始。注意，根据归纳假设(G.15)，对于任意 $0\leq r,\ s\leq t$,

$$\frac{1}{d}\langle q_{\mathrm{in},2},q_{\mathrm{in},2}\rangle=\frac{1}{d}\langle\beta_{*},\beta_{*}\rangle\to\mathbb{E}\left[\mathsf{B}_{*}^{2}\right]=\rho,\qquad \frac{1}{d}\langle q_{\mathrm{in},2},q_{\mathrm{in},1}^{s}\rangle\to\mathbb{E}\left[\mathsf{Q}_{\mathrm{in},2}\mathsf{Q}_{\mathrm{in},1}^{s}\right]=0, \tag{G.61a}$$

$$\frac{1}{d}\langle q_{\mathrm{in},1}^{r},q_{\mathrm{in},1}^{s}\rangle\to\mathbb{E}\left[\mathsf{Q}_{\mathrm{in},1}^{r}\mathsf{Q}_{\mathrm{in},1}^{s}\right]=(\Sigma_{q,\mathrm{in}}^{t})_{r,s},\qquad \frac{1}{d}\langle f_{q,\mathrm{in}}^{r},f_{q,\mathrm{in}}^{s}\rangle\to\mathbb{E}\left[\mathsf{F}_{q,\mathrm{in}}^{r}\mathsf{F}_{q,\mathrm{in}}^{s}\right]=(\Sigma_{p,\mathrm{in}}^{t+1})_{r,s}. \tag{G.61b}$$

然后我们将证明

$$\frac{1}{d}\langle q^r_{\text{in},1}, f^s_{q,\text{in}}\rangle \to 0, \qquad \frac{1}{d}\langle q_{\text{in},2}, f^s_{q,\text{in}}\rangle \to 0. \tag{G.62}$$

第一项为:

$$\begin{aligned}
\frac{1}{d}\langle q^r_{\text{in},1}, f^s_{q,\text{in}}\rangle &= \frac{1}{d}\Big\langle q^r_{\text{in},1}, \operatorname{diag}(\phi_s)(q^s_{\text{in},1} + \widetilde{\mathfrak{b}}_s q_{\text{in},2})\Big\rangle && \text{(G.63)}\\
&\quad + \frac{1}{d}\Big\langle q^r_{\text{in},1}, \operatorname{diag}_{n\times d}(\widetilde{\phi}_s)^\top (q^s_{\text{out}} + \widetilde{\mathfrak{a}}_s \Lambda q_{\text{in},2})\Big\rangle && \text{(G.64)}\\
&\quad - \frac{\mathfrak{b}_{s+1}}{d}\langle q^r_{\text{in},1}, q_{\text{in},2}\rangle, && \text{(G.65)}
\end{aligned}$$

其中，根据归纳假设 (G.15) 和 (G.16) 以及条目 4，右端的极限如下:

$$\begin{aligned}
\text{(G.63)} &\to \mathbb{E}\Big[\mathsf{Q}^r_{\text{in},1}(\mathsf{Q}^s_{\text{in},1} + \widetilde{\mathfrak{b}}_s \mathsf{Q}_{\text{in},2})\Big]\mathbb{E}[\Phi_s] = 0, && \text{(G.66)}\\
\text{(G.64)} &= \frac{1}{d}\Big\langle q^r_{\text{in},1}, \operatorname{diag}_{n\times d}(\widetilde{\phi}_s)^\top q^s_{\text{out}}\Big\rangle + \frac{\widetilde{\mathfrak{a}}_s}{d}\Big\langle q^r_{\text{in},1}, \operatorname{diag}_{n\times d}(\widetilde{\phi}_s)^\top \Lambda q_{\text{in},2}\Big\rangle\\
&\to 0 + \widetilde{\mathfrak{a}}_s \mathbb{E}\big[\mathsf{Q}^r_{\text{in},1}\mathsf{Q}_{\text{in},2}\big]\mathbb{E}\Big[\widetilde{\Phi}_s \Lambda_d\Big] = 0, && \text{(G.67)}\\
\text{(G.65)} &= -\frac{\mathfrak{b}_{s+1}}{d}\langle Q^\top f^1_{p,\text{in}}, Q^\top \beta_*\rangle = -\frac{\mathfrak{b}_{s+1}}{d}\Big\langle f_r(p^r_{\text{in}} + \mathfrak{b}_r\beta_*; \beta_*, \Theta) - \widetilde{\mathfrak{b}}_r\beta_*, \beta_*\Big\rangle \to 0. && \text{(G.68)}
\end{aligned}$$

(G.66) 中的等式源于 $(\mathsf{Q}^s_{\text{in}})_{0\le s\le t}$,$\Phi_s$ 与 Φ_s 在假设 (A29) 下的独立性，且 Φ_s 的均值为零。(G.67) 中的等式源于 $(\mathsf{Q}^s_{\text{in}},1)_{0\le s\le t}$ 与 $\mathsf{Q}_{\text{in}},2$、$(\mathsf{Q}^s_{\text{out}},1)_{0\le s\le t}$.~ 之间的独立性。(G.68) 中的等式源于 (G.14q) 中 $\mathfrak{b}_r$ 的定义。因此，(G.62) 中的第一条陈述成立。

接下来考察 (G.62) 中的第二条陈述，我们有

$$\frac{1}{d}\langle q_{\text{in},2}, f^s_{q,\text{in}}\rangle = \frac{1}{d}\Big\langle q_{\text{in},2}, \operatorname{diag}(\phi_s)(q^s_{\text{in},1} + \widetilde{\mathfrak{b}} q_{\text{in},2})\Big\rangle + \frac{1}{d}\Big\langle q_{\text{in},2}, \operatorname{diag}_{n\times d}(\widetilde{\phi}_s)^\top (q^s_{\text{out}} + \widetilde{\mathfrak{a}}_s \Lambda q_{\text{in},2}) - \mathfrak{b}_{s+1} q_{\text{in},2}\Big\rangle.$$

由于推导 (G.66) 时的类似论证，第一项的极限为 0 。对于第二项，采用与推导 (G.67) 时类似的变换，可得

$$\begin{aligned}
&\frac{1}{d}\Big\langle q_{\text{in},2}, \operatorname{diag}_{n\times d}(\widetilde{\phi}_s)^\top (q^s_{\text{out}} + \widetilde{\mathfrak{a}}_s \Lambda q_{\text{in},2})\Big\rangle - \frac{\mathfrak{b}_{s+1}}{d}\langle q_{\text{in},2}, q_{\text{in},2}\rangle\\
&= \frac{1}{d}\Big\langle q_{\text{in},2}, \operatorname{diag}_{n\times d}(\widetilde{\phi}_s)^\top q^s_{\text{out}}\Big\rangle + \frac{\widetilde{\mathfrak{a}}_s}{d}\Big\langle q_{\text{in},2}, \operatorname{diag}_{n\times d}(\widetilde{\phi}_s)^\top \Lambda q_{\text{in},2}\Big\rangle - \frac{\mathfrak{b}_{s+1}}{d}\langle q_{\text{in},2}, q_{\text{in},2}\rangle\\
&\to 0 + \widetilde{\mathfrak{a}}_s \mathbb{E}\Big[\widetilde{\Phi}_s \Lambda_d\Big]\mathbb{E}\big[\mathsf{Q}^2_{\text{in},2}\big] - \mathfrak{b}_{s+1}\mathbb{E}\big[\mathsf{Q}^2_{\text{in},2}\big]\\
&= \widetilde{\mathfrak{a}}_s \mathbb{E}\Big[\widetilde{\Phi}_s \Lambda_d\Big]\rho - \mathfrak{b}_{s+1}\rho = 0,
\end{aligned}$$

其中最后一步根据 $\mathfrak{b}_{s+1}$ 在(G.14k)中的定义得出。因此，(G.62)中的第二个等式也成立。综合(G.62) 和(G.61)可得

$$\frac{1}{d}U^\top U \to \begin{bmatrix} \rho & 0_{1\times(t+1)} & 0_{1\times(t+1)} \\ 0_{(t+1)\times 1} & \Sigma^t_{q,\text{in}} & 0_{(t+1)\times(t+1)} \\ 0_{(t+1)\times 1} & 0_{(t+1)\times(t+1)} & \Sigma^{t+1}_{p,\text{in}} \end{bmatrix} \in \mathbb{R}^{(1+2(t+1))\times(1+2(t+1))}. \tag{G.69}$$

注意到，右侧的矩阵由 $\rho > 0$ 中的假设（A1）以及条目 2中的归纳假设.

接下来，考虑$d^{-1}V^\top f_p^{t+1}$，在中。根据(G.15)中的归纳假设条目 4，对于任意 $0 \le s \le t$,

$$\frac{1}{d}\left\langle f_{p,\text{in},2}, f_{p,\text{in},1}^{t+1}\right\rangle \to \mathbb{E}\left[\mathsf{F}_{p,\text{in},2}\mathsf{F}_{p,\text{in},1}^{t+1}\right] = 0, \tag{G.70a}$$

$$\frac{1}{d}\left\langle f_{p,\text{in},1}^{s}, f_{p,\text{in},1}^{t+1}\right\rangle \to \mathbb{E}\left[\mathsf{F}_{p,\text{in},1}^{s}\mathsf{F}_{p,\text{in},1}^{t+1}\right] = (\Sigma_{q,\text{in}}^{t+1})_{s,t+1} = (\sigma_{q,\text{in}}^{t})_{s}, \tag{G.70b}$$

$$\frac{1}{d}\langle f_{p,\text{in},2}, f_{p,\text{in},2}\rangle \to \mathbb{E}\left[\mathsf{F}_{p,\text{in},2}^2\right] = \mathbb{E}\left[\mathsf{B}_*^2\right] = \rho, \tag{G.70c}$$

$$\frac{1}{d}\left\langle f_{p,\text{in},1}^{s}, f_{p,\text{in},2}\right\rangle \to \mathbb{E}\left[\mathsf{F}_{p,\text{in},1}^{s}\mathsf{F}_{p,\text{in},2}\right] = 0. \tag{G.70d}$$

(G.70a) 由 (G.56) 得出，且 (G.70d) 由其中的归纳假设得出，见条目 1。此外，我们研究其极限

$$\frac{1}{d}(p_{\text{in}}^{s+1})^\top f_{p,\text{in}}^{t+1} = \left[\tfrac{1}{d}\left\langle p_{\text{in}}^{s+1}, f_{p,\text{in},1}^{t+1}\right\rangle \quad \tfrac{1}{d}\langle p_{\text{in}}^{s+1}, f_{p,\text{in},2}\rangle\right]. \tag{G.71}$$

第二项（即 (G.71) ）容易看出趋于零:

$$\frac{1}{d}\langle p_{\text{in}}^{s+1}, \beta_*\rangle \to \mathbb{E}\left[\mathsf{P}_{\text{in}}^{s+1}\right]\mathbb{E}[\mathsf{B}_*] = 0, \tag{G.72}$$

由于 $\mathsf{P}^{s+1}{}_{\text{in}}$ 与 B_* 相互独立。至于第一项，即 (G.71)，我们有

$$\frac{1}{d}\left\langle p_{\text{in}}^{s+1}, f_{p,\text{in},1}^{t+1}\right\rangle = \frac{1}{d}\langle p_{\text{in}}^{s+1}, f_{t+1}(p_{\text{in}}^{t+1} + \mathfrak{b}_{t+1}\beta_*; \beta_*, \Theta)\rangle - \frac{\widetilde{\mathfrak{b}}_{t+1}}{d}\langle p_{\text{in}}^{s+1}, \beta_*\rangle.$$

我们已在 (G.72) 中证明，右侧第二项趋于零。由 假设 (A27) 与 定理 K.2，第一项收敛到

$$\frac{1}{d}\langle p_{\text{in}}^{s+1}, f_{t+1}(p_{\text{in}}^{t+1} + \mathfrak{b}_{t+1}\beta_*; \beta_*, \Theta)\rangle \to \mathbb{E}\left[\mathsf{P}_{\text{in}}^{s+1} f_{t+1}(\mathsf{P}_{\text{in}}^{t+1} + \mathfrak{b}_{t+1}\mathsf{B}_*; \mathsf{B}_*, \Theta)\right]. \tag{G.73}$$

由于 $(\mathsf{P}^{s+1}{}_{\text{in}}, \mathsf{P}^{t+1}{}_{\text{in}})$ 为联合高斯，且独立于 B_*，上述表达式为零，依据是 Stein 引理[49, 定理 21] 以及 f_{t+1} 的散度为零（见 假设 (A8)）。将 (G.72)与 (G.73)结合起来，我们看到

$$\frac{1}{d}(p_{\text{in}}^{s+1})^\top f_{p,\text{in}}^{t+1} \to \begin{bmatrix}0 & 0\end{bmatrix}. \tag{G.74}$$

结合 (G.74) 与(G.70)，我们得到

$$\frac{1}{d}V^\top f_{p,\text{in}}^{t+1} \to \begin{bmatrix}0 & \rho\\ \sigma_{q,\text{in}}^{t} & 0_{t+1}\\ 0_{t+1} & 0_{t+1}\end{bmatrix} \in \mathbb{R}^{(1+2(t+1))\times 2}. \tag{G.75}$$

现在我们可以 计算 (G.60) 右侧的值。利用 (G.69) 和 (G.75)，我们有 ve

$$(U^\top U)^{-1}V^\top f_{p,\text{in}}^{t+1} = (d^{-1}U^\top U)^{-1}\cdot d^{-1}V^\top f_{p,\text{in}}^{t+1} \to \begin{bmatrix}\rho^{-1} & 0_{1\times(t+1)} & 0_{1\times(t+1)}\\ 0_{(t+1)\times 1} & (\Sigma_{q,\text{in}}^{t})^{-1} & 0_{(t+1)\times(t+1)}\\ 0_{(t+1)\times 1} & 0_{(t+1)\times(t+1)} & (\Sigma_{p,\text{in}}^{t+1})^{-1}\end{bmatrix}\begin{bmatrix}0 & \rho\\ \sigma_{q,\text{in}}^{t} & 0_{t+1}\\ 0_{t+1} & 0_{t+1}\end{bmatrix}$$

$$= \begin{bmatrix}0 & 1\\ (\Sigma_{q,\text{in}}^{t})^{-1}\sigma_{q,\text{in}}^{t} & 0_{t+1}\\ 0_{t+1} & 0_{t+1}\end{bmatrix} \in \mathbb{R}^{(1+2(t+1))\times 2}.$$

根据定理K.5，(G.60)中的第二项在 Wasserstein 距离下具有高斯分布极限，记为$[\mathsf{U}_q^{t+1}$，属于$0] \in \mathbb{R}^{1\times 2}$，且与 A、$(\mathsf{Q}_{\text{in}}^s)_{0\le s\le t}$、$(\mathsf{Q}_{\text{out}}^s)_{0\le s\le t}$、$(\mathsf{F}_{q,\text{ in}}^s)_{0\le s\le t}$、$(\mathsf{F}_{q,\text{ out}}^s)_{0\le s\le t}$、$(\mathsf{P}^{s+1}$ 属于$)_{0\le s\le t}$独立，

$(\mathsf{P}^{s+1}_{\text{out}})_{0\le s\le t}$、$(\mathsf{F}^{s}_{p,\ \text{in}})_{0\le s\le t+1}$、$(\mathsf{F}^{s}_{p,\ \text{out}})_{0\le s\le t+1}$。注意，第二个元素为零。这是因为 $\Pi_{V^\perp}$ 的每一列都与 V 的第一列正交，而后者等于 $f_{p,\ \text{in},\ 2}$，因此 $\Pi_{V^\perp}^\top f_{p,\ \text{in},\ 2}$ 为零。结合上述公式，并回顾

$$U \xrightarrow{W} \begin{bmatrix} \mathsf{Q}_{\text{in},2} & \mathsf{Q}^0_{\text{in},1} & \cdots & \mathsf{Q}^t_{\text{in},1} & \mathsf{F}^0_{q,\text{in}} & \cdots & \mathsf{F}^t_{q,\text{in}} \end{bmatrix}$$

根据归纳假设（G.15）在条目 4中，并应用K.4 定理，我们得出

$$q^{t+1}_{\text{in},1} \,\Big|\, \{U = Q^\top V\} \xrightarrow{W} \begin{bmatrix} \mathsf{Q}^0_{\text{in},1} & \cdots & \mathsf{Q}^t_{\text{in},1} \end{bmatrix} (\Sigma^t_{q,\text{in}})^{-1}\sigma^t_{q,\text{in}} + \mathsf{U}^{t+1}_{q,\text{in}} = \sum_{s=0}^{t} \mathsf{Q}^s_{\text{in},1}\big((\Sigma^t_{q,\text{in}})^{-1}\sigma^t_{q,\text{in}}\big)_s + \mathsf{U}^{t+1}_{q,\text{in}}, \tag{G.76a}$$

$$q_{\text{in},2} \,\Big|\, \{U = Q^\top V\} \xrightarrow{W} \mathsf{Q}_{\text{in},2}. \tag{G.76b}$$

我们声称，上述两个方程右侧恰好是 Q^{t+1} 在$= \begin{bmatrix} \mathsf{Q}^{t+1} \text{ 中},\ 1 & \mathsf{Q}_{\text{in},\ 2} \end{bmatrix}^\top$。为说明这一点，注意到 $(\mathsf{Q}^s_{\text{in}})_{0\le s\le t+1}$ 是联合高斯的，并且与 A，$(\mathsf{P}^{s+1}$ 在$)_{0\le s\le t}$，$(\mathsf{P}^{s+1}_{\text{out}})_{0\le s\le t}$独立。因此，根据（G.14f）的定义，只需验证（G.76a）右侧（记为 $\widetilde{\mathsf{Q}}^{t+1}$ 在）与 $(\mathsf{Q}^s_{\text{in}})_{0\le s\le t}$具有所需的协方差结构。对于任意 $0 \le r \le t$，有

$$\begin{aligned} \mathbb{E}\Big[\widetilde{\mathsf{Q}}^{t+1}_{\text{in},1}\mathsf{Q}^r_{\text{in},1}\Big] &= \sum_{s=0}^{t} \mathbb{E}\big[\mathsf{Q}^s_{\text{in},1}\mathsf{Q}^r_{\text{in},1}\big]\big((\Sigma^t_{q,\text{in}})^{-1}\sigma^t_{q,\text{in}}\big)_s + \mathbb{E}\Big[\mathsf{U}^{t+1}_{q,\text{in}}\mathsf{Q}^r_{\text{in},1}\Big] \\ &= \sum_{s=0}^{t} (\Sigma^t_{q,\text{in}})_{r,s}\big((\Sigma^t_{q,\text{in}})^{-1}\sigma^t_{q,\text{in}}\big)_s + 0 = \big(\Sigma^t_{q,\text{in}}(\Sigma^t_{q,\text{in}})^{-1}\sigma^t_{q,\text{in}}\big)_r = (\sigma^t_{q,\text{in}})_r. \end{aligned} \tag{G.77a}$$

另请注意，

$$\mathbb{E}\Big[(\widetilde{\mathsf{Q}}^{t+1}_{\text{in},1})^2\Big] = \lim_{d\to\infty}\frac{1}{d}\Big\|q^{t+1}_{\text{in},1}\Big\|_2^2 = \lim_{d\to\infty}\frac{1}{d}\Big\|f^{t+1}_{p,\text{in},1}\Big\|_2^2 = \mathbb{E}\Big[(\mathsf{F}^{t+1}_{p,\text{in},1})^2\Big] = s^{t+1}_{q,\text{in}}. \tag{G.77b}$$

此外，$\widetilde{\mathsf{Q}}^{t+1}$ 在，1 与 $\mathsf{Q}_{\text{in},\ 2}$ 独立，因为 $(\mathsf{Q}^s_{\text{in},\ 1})_{0\le s\le t}$ 和 U^{t+1}_{q}，在 都与 $\mathsf{Q}_{\text{in},\ 2}$独立。由此并结合（G.77），可知 $\begin{bmatrix} \widetilde{\mathsf{Q}}^{t+1} \text{ 在},\ 1 & \mathsf{Q}_{\text{in},\ 2} \end{bmatrix}^\top$ 与 $(\mathsf{Q}^s_{\text{in}})_{0\le s\le t}$ 具有所需的协方差结构，我们将其记为 Q^{t+1} 在，因此

$$\begin{aligned} &\begin{bmatrix} q_{\text{in},2} & \boldsymbol{q}^{t+1}_{\text{in},1} & \boldsymbol{f}^t_{q,\text{in}} & \boldsymbol{p}^{t+1}_{\text{in}} & f_{p,\text{in},2} & \boldsymbol{f}^{t+1}_{p,\text{in},1} & w_{p,\text{in}} & w_{q,\text{in}} \end{bmatrix} \\ &\quad\xrightarrow{W} \begin{bmatrix} \mathsf{Q}_{\text{in},2} & \mathbf{Q}^{t+1}_{\text{in},1} & \mathbf{F}^t_{q,\text{in}} & \mathbf{P}^{t+1}_{\text{in}} & \mathsf{F}_{p,\text{in},2} & \mathbf{F}^{t+1}_{p,\text{in},1} & \mathsf{W}^\top_{p,\text{in}} & \mathsf{W}^\top_{q,\text{in}} \end{bmatrix}. \end{aligned} \tag{G.78}$$

我们检查 U^{t+1}_{q}，在 的方差严格为正。由于 $(\mathsf{Q}^s_{\text{in},\ 1})_{0\le s\le t} \sim \mathcal{N}(0_{t+1},\ \Sigma^t_{q,\ \text{in}})$，我们有

$$\begin{aligned} \sum_{s=0}^{t} \mathsf{Q}^s_{\text{in},1}\big((\Sigma^t_{q,\text{in}})^{-1}\sigma^t_{q,\text{in}}\big)_s &\sim \mathcal{N}\Big(0, \big((\Sigma^t_{q,\text{in}})^{-1}\sigma^t_{q,\text{in}}\big)^\top \cdot \Sigma^t_{q,\text{in}} \cdot \big((\Sigma^t_{q,\text{in}})^{-1}\sigma^t_{q,\text{in}}\big)\Big) \\ &= \mathcal{N}\big(0, (\sigma^t_{q,\text{in}})^\top(\Sigma^t_{q,\text{in}})^{-1}\sigma^t_{q,\text{in}}\big). \end{aligned} \tag{G.79}$$

根据 (G.76)，U^{t+1}_{q}，在 的方差为

$$\mathbb{E}\Big[(\mathsf{U}^{t+1}_{q,\text{in}})^2\Big] = s^{t+1}_{q,\text{in}} - (\sigma^t_{q,\text{in}})^\top(\Sigma^t_{q,\text{in}})^{-1}\sigma^t_{q,\text{in}},$$

其中这两个　右侧的项分别来自(G.79) 和(G.77b)。根据舒尔补公式　，

$$\mathbb{E}\Big[(\mathsf{U}^{t+1}_{q,\text{in}})^2\Big] = \frac{1}{\Big[(\Sigma^{t+1}_{q,\text{in}})^{-1}\Big]_{t+1,t+1}} = \Big(e_{t+1}^\top(\Sigma^{t+1}_{q,\text{in}})^{-1}e_{t+1}\Big)^{-1} > 0, \tag{G.80}$$

其中，严格不等式源于Σ^{t+1}_q 的严格正定性，该性质在(G.57)附近给出。这证明了(G.41) 中条目 3的第一个结论。

关于q_{out}^{t+1}的归纳步骤。重新定义矩阵 U、V如下：

$$U = \begin{bmatrix} q_{\text{out}}^0 & \cdots & q_{\text{out}}^t & f_{q,\text{out},2} & f_{q,\text{out},1}^0 & \cdots & f_{q,\text{out},1}^t \end{bmatrix} \in \mathbb{R}^{n\times(1+2(t+1))}, \tag{G.81}$$

$$V = \begin{bmatrix} f_{p,\text{out}}^0 & \cdots & f_{p,\text{out}}^t & p_{\text{out},2} & p_{\text{out},1}^1 & \cdots & p_{\text{out},1}^{t+1} \end{bmatrix} \in \mathbb{R}^{n\times(1+2(t+1))}. \tag{G.82}$$

在给定 q_{out}^{t+1}及所有先前迭代量（包括初始化量和侧信息）的条件下，其概率分布与在条件 q^{t+1}下给定 $U = O^\top V$ 时的概率分布相同。根据定理 K.6，后一个条件分布为

$$q_{\text{out}}^{t+1} \,\big|\, \{U = O^\top V\} \;=\; O^\top f_{\text{out}}^{t+1} \,\big|\, \{U = O^\top V\} \overset{d}{=} U(U^\top U)^{-1}V^\top f_{p,\text{out}}^{t+1} + \Pi_{U^\top}\widetilde{O}^\top \Pi_{V^\perp}^\top f_{p,\text{out}}^{t+1}, \tag{G.83}$$

其中，$O \sim \text{Haar}(\mathbb{O}(n-(1+2(t+1))))$ 与其他所有量相互独立。为了分析右侧的第一项，我们计算 $n^{-1}U^\top U$和$n^{-1}V^\top f_p^{t+1}$的out极限。特别地，(G.87)表明，$U^\top U$在(G.83) 右侧是可逆的，只要 n足够大。

从 $n^{-1}U^\top U$出发，对于任意 $0 \le r,\ s \le t$根据归纳假设条目 1 和4，我们有

$$\frac{1}{n}\langle q_{\text{out}}^r, q_{\text{out}}^s\rangle \to (\Sigma_{q,\text{out}}^t)_{r,s}, \qquad \frac{1}{n}\langle f_{q,\text{out},2}, f_{q,\text{out},2}\rangle = \frac{1}{d}\langle z, z\rangle \to \mathbb{E}\big[\mathsf{Z}^2\big] = \sigma^2, \tag{G.84a}$$

$$\frac{1}{n}\big\langle f_{q,\text{out},2}, f_{q,\text{out},1}^s\big\rangle \to 0, \qquad \frac{1}{n}\langle f_{q,\text{输出}\,1}^r,\ f_{q,\text{输出}\,1}^s\rangle \to\ (\Sigma_p^{t+1},\text{out})_{\ r+1,s+1}. \tag{G.84b}$$

注意，(G.84a) 中的第二个方程意味着 $\mathbb{E}\big[\mathrm{F}_{q,\ \text{out},\ 2}^2\big] = \sigma^2$。接下来我们考察

$$\frac{1}{n}\big\langle q_{\text{out}}^r, f_{q,\text{out},1}^s\big\rangle, \qquad \frac{1}{n}\langle q_{\text{out}}^r, f_{q,\text{out},2}\rangle. \tag{G.85}$$

容易看出第二项趋于零：

$$\frac{1}{n}\langle q_{\text{out}}^r, f_{q,\text{out},2}\rangle = \frac{1}{n}\big\langle O^\top f_{p,\text{out}}^r, \Lambda q_{\text{in},2}\big\rangle = \frac{1}{n}\big\langle g_r(p_{\text{out},1}^r + \mathfrak{a}_r z; z, \varepsilon, \Xi) - \widetilde{\mathfrak{a}}_r z, z\big\rangle \to 0, \tag{G.86}$$

根据 $\widetilde{\mathfrak{a}}_r$ 在 (G.14q)中的定义。接下来考察 (G.85)中的第一项，

$$\begin{aligned}\frac{1}{n}\big\langle q_{\text{out}}^r, f_{q,\text{out},1}^s\big\rangle &= \frac{1}{n}\Big\langle q_{\text{out}}^r, \text{diag}(\psi_s)(q_{\text{out}}^s + \widetilde{\mathfrak{a}}_s\Lambda q_{\text{in},2}) + \text{diag}_{n\times d}(\widetilde{\psi}_s)(q_{\text{in},1}^s + \widetilde{\mathfrak{b}}_s q_{\text{in},2}) - \mathfrak{a}_{s+1}\Lambda q_{\text{in},2}\Big\rangle \\ &= \frac{1}{n}\langle q_{\text{out}}^r, \text{diag}(\psi_s) q_{\text{out}}^s\rangle + \frac{\widetilde{\mathfrak{a}}_s}{n}\langle q_{\text{out}}^r, \text{diag}(\psi_s)\Lambda q_{\text{in},2}\rangle + \frac{1}{n}\Big\langle q_{\text{out}}^r, \text{diag}_{n\times d}(\widetilde{\psi}_s)(q_{\text{in},1}^s + \widetilde{\mathfrak{b}}_s q_{\text{in},2})\Big\rangle - \frac{\mathfrak{a}_{s+1}}{n}\big\langle O^\top f_{p,\text{out}}^r, \Lambda q_{\text{in},2}\big\rangle.\end{aligned}$$

第一项的极限为零，因为 $(\mathsf{Q}_{\text{out}}^r,\ \mathsf{Q}_{\text{out}}^s)$ 与 Ψ_s 独立，且其均值为零（见假设 (A29)）。第二项的极限也为零，因为 $q_{\text{out}}^r,\ q_{\text{in},\ 2}$的 Wasserstein 极限 $\mathsf{Q}_{\text{out}}^r,\ \mathsf{Q}_{\text{in},\ 2}$彼此独立。第三项同样由于 $\mathsf{Q}_{\text{out}}^r$ 与 Q_{in}^s 之间的独立性而趋于零。我们已在(G.86) 中证明第四项趋于零。因此，$\frac{1}{n}\langle q_{\text{out}}^r,\ f_{q,\ \text{out},\ 1}^s\rangle \to 0$。

我们已经证明(G.85)中的两项极限均为零。由此以及(G.84)，

$$\frac{1}{n}U^\top U \to \begin{bmatrix} \Sigma_{q,\text{out}}^t & 0_{(t+1)\times 1} & 0_{(t+1)\times(t+1)} \\ 0_{1\times(t+1)} & \sigma^2 & 0_{1\times(t+1)} \\ 0_{(t+1)\times(t+1)} & 0_{(t+1)\times 1} & \Sigma_{p,\text{out}}^{t+1} \end{bmatrix} \in \mathbb{R}^{(1+2(t+1))\times(1+2(t+1))}. \tag{G.87}$$

位于 r 的矩阵 ight 是可逆的，这由归纳假设（位于条目 2 ）以及正性所保证，该量是 $\sigma^2 = \kappa_2 \quad - \rho$ 所蕴含，依据Assumpti ons(A1) 和 (A2) 。

接下来，我们计算 $n^{-1}V^\top f_{p\ ,\ \text{out}}^{t+1}$ 的极限。根据 (G.16) 中 条目 4 的归纳假设，对于任意$0 \le s \le t$，我们有

$$\frac{1}{n}\big\langle f_{p,\text{out}}^s, f_{p,\text{out}}^{t+1}\big\rangle \to (\Sigma_{q,\text{out}}^{t+1})_{s,t+1} = (\sigma_{q,\text{out}}^t)_s. \tag{G.88}$$

然后我们考察

$$\frac{1}{n}\langle p_{\text{out},2}, f_{p,\text{out}}^{t+1}\rangle, \qquad \frac{1}{n}\langle p_{\text{out},1}^{s+1}, f_{p,\text{out}}^{t+1}\rangle. \tag{G.89}$$

第一项等于 $n^{-1}\langle z,\ g_{t+1}(p_{\text{out},\,1}^{t+1}+\mathfrak{a}_{t+1}z;\ z,\ \varepsilon)-\widetilde{\mathfrak{a}}_{t+1}z\rangle$ 根据$\widetilde{\mathfrak{a}}_{t+1}$在(G.14q)中的定义，它收敛于零。至于(G.89)中的第二项，

$$\frac{1}{n}\langle p_{\text{out},1}^{s+1}, f_{p,\text{out}}^{t+1}\rangle = \frac{1}{n}\langle p_{\text{out},1}^{s+1}, g_{t+1}(p_{\text{out},1}^{t+1}+\mathfrak{a}_{t+1}p_{\text{out},2}; p_{\text{out},2}, w_{p,\text{out}})\rangle - \frac{\widetilde{\mathfrak{a}}_{t+1}}{n}\langle p_{\text{out},1}^{s+1}, p_{\text{out},2}\rangle. \tag{G.90}$$

应用假设 (A27)和定理 K.2，注意到$(\mathsf{P}_{\text{out},\,1}^{s+1},\ \mathsf{P}_{\text{out},\,1}^{t+1},\ \mathsf{P}_{\text{out},\,2})$与$\mathsf{W}_{p,\ \text{out}}$联合高斯且相互独立，利用 Stein 引理 [49, 定理 2。1] 以及 g_{t+1} (见假设 (A8))的散度为零性质，第一项收敛于零。我们断言，(G.90) 中的第二项也趋于零。为证明这一点，注意到

$$\frac{1}{n}\langle p_{\text{out},1}^{s+1}, p_{\text{out},2}\rangle = \frac{1}{n}\langle f_{q,\text{out},1}^{s}, f_{q,\text{out},2}\rangle \to \mathbb{E}\big[\mathsf{F}_{q,\text{out},1}^{s}\mathsf{F}_{q,\text{out},2}\big] = 0,$$

根据条目 1中的归纳假设，如所声称。因此，由(G.90)可知，(G.89)中的第二项趋于零。

We h我们已经证明(G.89) 中的两项都收敛到零。利用这一点以及 (G.88)，可得

$$\frac{1}{n}V^{\top}f_{p,\text{out}}^{t+1} \to \begin{bmatrix}\sigma_{q,\text{out}}^{t}\\ 0\\ 0_{t+1}\end{bmatrix} \in \mathbb{R}^{1+2(t+1)}.$$

进一步结合上述结果与 (G.87)，并回顾

$$U \xrightarrow{W} \big[\mathsf{Q}_{\text{out}}^{0} \quad \cdots \quad \mathsf{Q}_{\text{out}}^{t} \quad \mathsf{F}_{q,\text{out},2} \quad \mathsf{F}_{q,\text{out},1}^{0} \quad \cdots \quad \mathsf{F}_{q,\text{out}.1}^{1}\big]$$

根据(G.16)中的归纳假设，在 条目 4中，我们得到 (G.83)右端第一项的 Wasserstein 极限：

$$\begin{aligned}
&U\cdot(n^{-1}U^{\top}U)^{-1}\cdot n^{-1}V^{\top}f_{p,\text{out}}^{t+1}\\
&\xrightarrow{W}\big[\mathsf{Q}_{\text{out}}^{0} \quad \cdots \quad \mathsf{Q}_{\text{out}}^{t} \quad \mathsf{F}_{q,\text{out},2} \quad \mathsf{F}_{q,\text{out},1}^{0} \quad \cdots \quad \mathsf{F}_{q,\text{out},1}^{t}\big]\begin{bmatrix}\Sigma_{q,\text{out}}^{t} & 0_{(t+1)\times 1} & 0_{(t+1)\times(t+1)}\\ 0_{1\times(t+1)} & \sigma^{2} & 0_{1\times(t+1)}\\ 0_{(t+1)\times(t+1)} & 0_{(t+1)\times 1} & \Sigma_{p,\text{out}}^{t+1}\end{bmatrix}^{-1}\begin{bmatrix}\sigma_{q,\text{out}}^{t}\\ 0\\ 0_{t+1}\end{bmatrix}\\
&=\sum_{s=0}^{t}\mathsf{Q}_{\text{out}}^{s}\big((\Sigma_{q,\text{out}}^{t})^{-1}\sigma_{q,\text{out}}^{t}\big)_{s}.
\end{aligned}$$

此外，根据定理 K.5 ，(G.83) 右端的第二项存在一个高斯分布极限，记为 $\mathsf{U}_{q}^{t+1}{}_{,\ \text{out}}\in\mathbb{R}$，且它与其他所有变量独立。因此，总体而言，$q_{\text{out}}^{t+1}$ 的条件经验分布收敛于随机变量

$$\widetilde{\mathsf{Q}}_{\text{out}}^{t+1} = \sum_{s=0}^{t}\mathsf{Q}_{\text{out}}^{s}\big((\Sigma_{q,\text{out}}^{t})^{-1}\sigma_{q,\text{out}}^{t}\big)_{s} + \mathsf{U}_{q,\text{out}}^{t+1} \tag{G.91}$$

它与 A 和 $(\mathsf{P}^{s+1}$ 中的$)_{0\le s\le t}$，$(\mathsf{P}_{\text{out}}^{s+1})_{0\le s\le t}$独立。接下来，我们将通过指出 $\mathsf{Q}_{\text{out}}^{0}$，$\cdots$、$\mathsf{Q}_{\text{out}}^{t}$，$\widetilde{\mathsf{Q}}_{\text{out}}^{t+1}$联合高斯，并检验它们的协方差结构，来验证随机变量 $\widetilde{\mathsf{Q}}_{\text{out}}^{t+1}$恰好就是 $\mathsf{Q}_{\text{out}}^{t+1}$。

对于任意 $0\le s\le t$，都很容易计算：

$$\mathbb{E}\Big[\widetilde{\mathsf{Q}}_{\text{out}}^{t+1}\mathsf{Q}_{\text{out}}^{r}\Big] = \sum_{s=0}^{t}(\Sigma_{q,\text{out}}^{t})_{r,s}\big((\Sigma_{q,\text{out}}^{t})^{-1}\sigma_{q,\text{out}}^{t}\big)_{s} = (\sigma_{q,\text{out}}^{t})_{r}, \tag{G.92a}$$

$$\mathbb{E}\Big[(\widetilde{\mathsf{Q}}_{\text{out}}^{t+1})^2\Big] = \lim_{n\to\infty}\frac{1}{n}\big\|q_{\text{out}}^{t+1}\big\|_2^2 = \lim_{n\to\infty}\frac{1}{n}\big\|f_{p,\text{out}}^{t+1}\big\|_2^2 = \mathbb{E}\big[(\mathsf{F}_{p,\text{out}}^{t+1})^2\big] = s_{q,\text{out}}^{t+1}, \tag{G.92b}$$

从而验证了 $\mathsf{Q}_{\text{out}}^0$, $\cdots$, $\mathsf{Q}_{\text{out}}^t$, $\widetilde{\mathsf{Q}}_{\text{out}}^{t+1}$ 所具有的预期协方差结构。这使我们能够将 $\widetilde{\mathsf{Q}}_{\text{out}}^{t+1}$ 与 $\mathsf{Q}_{\text{out}}^{t+1}$ 识别为同一对象，并得出

$$\begin{aligned}&\big[\boldsymbol{q}_{\text{out}}^{t+1} \quad f_{q,\text{out},2} \quad \boldsymbol{f}_{q,\text{out},1}^t \quad p_{\text{out},2} \quad \boldsymbol{p}_{\text{out},1}^{t+1} \quad \boldsymbol{f}_{p,\text{out}}^{t+1} \quad w_{p,\text{out}} \quad w_{q,\text{out}}\big]\\ &\quad\xrightarrow{W}\big[\mathbf{Q}_{\text{out}}^{t+1} \quad \mathsf{F}_{q,\text{out},2} \quad \mathbf{F}_{q,\text{out},1}^t \quad \mathsf{P}_{\text{out},2} \quad \mathbf{P}_{\text{out},1}^{t+1} \quad \mathbf{F}_{p,\text{out}}^{t+1} \quad \mathsf{W}_{p,\text{out}}^\top \quad \mathsf{W}_{q,\text{out}}^\top\big].\end{aligned} \tag{G.93}$$

我们还验证了 $\mathsf{U}_{q,\text{out}}^{t+1}$ 的方差严格为正：

$$\begin{aligned}\mathbb{E}\big[(\mathsf{U}_{q,\text{out}}^{t+1})^2\big] &= \mathbb{E}\big[(\mathsf{Q}_{\text{out}}^{t+1})^2\big] - \mathbb{E}\left[\left(\sum_{s=0}^{t}\mathsf{Q}_{\text{out}}^s\big((\Sigma_{q,\text{out}}^t)^{-1}\sigma_{q,\text{out}}^t\big)\right)^2\right]\\ &= s_{q,\text{out}}^{t+1} - \big((\Sigma_{q,\text{out}}^t)^{-1}\sigma_{q,\text{out}}^t\big)^\top\cdot\Sigma_{q,\text{out}}^t\cdot\big((\Sigma_{q,\text{out}}^t)^{-1}\sigma_{q,\text{out}}^t\big)\\ &= s_{q,\text{out}}^{t+1} - (\sigma_{q,\text{out}}^t)^\top(\Sigma_{q,\text{out}}^t)^{-1}\sigma_{q,\text{out}}^t = \frac{1}{e_{t+1}^\top(\Sigma_{q,\text{out}}^{t+1})^{-1}e_{t+1}} > 0,\end{aligned} \tag{G.94}$$

根据舒尔补公式以及$\Sigma_{q}^{t+1}{}_{,\text{out}}$ 在 (G.57)附近所证明的严格正定性。这证明了 (G.41)中条目 3的第二个结论。

对$f_q^{t+1}{}_{,\text{在}}$ $f_q^{t+1}{}_{,\text{out}}$中的归纳步骤。 由前两个步骤中证明的 (G.78)和 (G.93)直接可得，结合抽象 VAMP 算法的 (G.5b)更新规则以及定理 K.2，我们得到

$$\begin{aligned}&\big[q_{\text{in},2} \quad \boldsymbol{q}_{\text{in},1}^{t+1} \quad \boldsymbol{f}_{q,\text{in}}^{t+1} \quad \boldsymbol{p}_{\text{in}}^{t+1} \quad f_{p,\text{in},2} \quad \boldsymbol{f}_{p,\text{in},1}^{t+1} \quad w_{p,\text{in}} \quad w_{q,\text{in}}\big]\\ &\quad\xrightarrow{W}\big[\mathsf{Q}_{\text{in},2} \quad \mathbf{Q}_{\text{in},1}^{t+1} \quad \mathbf{F}_{q,\text{in}}^{t+1} \quad \mathbf{P}_{\text{in}}^{t+1} \quad \mathsf{F}_{p,\text{in},2} \quad \mathbf{F}_{p,\text{in},1}^{t+1} \quad \mathsf{W}_{p,\text{in}}^\top \quad \mathsf{W}_{q,\text{in}}^\top\big],\end{aligned} \tag{G.95}$$

$$\begin{aligned}&\big[\boldsymbol{q}_{\text{out}}^{t+1} \quad f_{q,\text{out},2} \quad \boldsymbol{f}_{q,\text{out},1}^{t+1} \quad p_{\text{out},2} \quad \boldsymbol{p}_{\text{out},1}^{t+1} \quad \boldsymbol{f}_{p,\text{out}}^{t+1} \quad w_{p,\text{out}} \quad w_{q,\text{out}}\big]\\ &\quad\xrightarrow{W}\big[\mathbf{Q}_{\text{out}}^{t+1} \quad \mathsf{F}_{q,\text{out},2} \quad \mathbf{F}_{q,\text{out},1}^{t+1} \quad \mathsf{P}_{\text{out},2} \quad \mathbf{P}_{\text{out},1}^{t+1} \quad \mathbf{F}_{p,\text{out}}^{t+1} \quad \mathsf{W}_{p,\text{out}}^\top \quad \mathsf{W}_{q,\text{out}}^\top\big].\end{aligned} \tag{G.96}$$

我们验证条目 1 中的归纳假设对$\mathsf{F}_q^{t+1}{}_{,\text{out},1}$仍然成立。根据上面的收敛结果(G.96) ，

$$\begin{aligned}&\mathbb{E}\big[\mathsf{F}_{q,\text{out},1}^{t+1}\mathsf{F}_{q,\text{out},2}\big] = \lim_{n\to\infty}\frac{1}{n}\big\langle f_{q,\text{out},1}^{t+1}, f_{q,\text{out},2}\big\rangle\\ &= \lim_{n\to\infty}\frac{1}{n}\Big\langle\operatorname{diag}(\psi_{t+1})(q_{\text{out}}^{t+1} + \widetilde{\mathfrak{a}}_{t+1}\Lambda q_{\text{in},2}) + \operatorname{diag}_{n\times d}(\widetilde{\psi}_{t+1})(q_{\text{in},1}^{t+1} + \widetilde{\mathfrak{b}}_{t+1}q_{\text{in},2}) - \mathfrak{a}_{t+2}\Lambda q_{\text{in},2}, \Lambda q_{\text{in},2}\Big\rangle\\ &= \lim_{n\to\infty}\frac{1}{n}\big\langle\operatorname{diag}(\psi_{t+1})(q_{\text{out}}^{t+1} + \widetilde{\mathfrak{a}}_{t+1}\Lambda q_{\text{in},2}), \Lambda q_{\text{in},2}\big\rangle + \lim_{n\to\infty}\frac{1}{n}\Big\langle\operatorname{diag}_{n\times d}(\widetilde{\psi}_{t+1})(q_{\text{in},1}^{t+1} + \widetilde{\mathfrak{b}}_{t+1}q_{\text{in},2}), \Lambda q_{\text{in},2}\Big\rangle - \mathfrak{a}_{t+2}\lim_{n\to\infty}\frac{1}{n}\langle z, z\rangle.\end{aligned} \tag{G.97}$$

上面的最后一项等于 $-\mathfrak{a}_{t+1}\sigma^2$。第一项可按如下方式计算：

$$\begin{aligned}&\frac{1}{n}\big\langle\operatorname{diag}(\psi_{t+1})(q_{\text{out}}^{t+1} + \widetilde{\mathfrak{a}}_{t+1}\Lambda q_{\text{in},2}), \Lambda q_{\text{in},2}\big\rangle\\ &= \frac{1}{n}\big\langle\Lambda^\top\operatorname{diag}(\psi_{t+1})q_{\text{out}}^{t+1}, q_{\text{in},2}\big\rangle + \frac{\widetilde{\mathfrak{a}}_{t+1}}{n}\big\langle\Lambda^\top\operatorname{diag}(\psi_{t+1})\Lambda q_{\text{in},2}, q_{\text{in},2}\big\rangle\\ &\to 0 + \frac{\widetilde{\mathfrak{a}}_{t+1}}{\delta}\mathbb{E}\big[\Psi_{t+1}\Lambda_d^2\big]\mathbb{E}\big[\mathsf{Q}_{\text{in},2}^2\big] = \frac{\rho}{\delta}\mathbb{E}\big[\Psi_t\Lambda_d^2\big]\widetilde{\mathfrak{a}}_{t+1},\end{aligned} \tag{G.98a}$$

其中，最后一行是由$\mathsf{Q}_{\text{out}}^{t+1}$、$\mathsf{Q}_{\text{in}}$和2的独立性推出的。最后，我们来看 (G.97) 中的第二项：

$$\frac{1}{n}\Big\langle\operatorname{diag}_{n\times d}(\widetilde{\psi}_{t+1})(q_{\text{in},1}^{t+1} + \widetilde{\mathfrak{b}}_{t+1}q_{\text{in},2}), \Lambda q_{\text{in},2}\Big\rangle$$

$$= \frac{1}{n}\Big\langle \Lambda^\top \operatorname{diag}_{n\times d}(\widetilde{\psi}_{t+1}) q_{\text{in},1}^{t+1}, q_{\text{in},2}\Big\rangle + \frac{\widetilde{\mathfrak{b}}_{t+1}}{n}\Big\langle \Lambda^\top \operatorname{diag}_{n\times d}(\widetilde{\psi}_{t+1}) q_{\text{in},2}, q_{\text{in},2}\Big\rangle$$
$$\to 0 + \frac{\widetilde{\mathfrak{b}}_{t+1}}{\delta}\mathbb{E}\Big[\Lambda_d \widetilde{\Psi}_{t+1}\Big]\mathbb{E}\big[\mathsf{Q}_{\text{in},2}^2\big] = \frac{\rho}{\delta}\mathbb{E}\Big[\widetilde{\Psi}_{t+1}\Lambda_d\Big]\widetilde{\mathfrak{b}}_{t+1}, \tag{G.98b}$$

因为$\mathsf{Q}_{\text{in},2}$和Q^{t+1}在$,1$相互独立。由(G.98)，我们得出

$$\frac{1}{n}\langle f_{q,\text{out},1}^{t+1}, f_{q,\text{out},2}\rangle \to \frac{\rho}{\delta}\Big(\mathbb{E}\big[\Psi_{t+1}\Lambda_d^2\big]\widetilde{\mathfrak{a}}_{t+1} + \mathbb{E}\Big[\widetilde{\Psi}_{t+1}\Lambda_d\Big]\widetilde{\mathfrak{b}}_{t+1}\Big) - \mathfrak{a}_{t+2}\sigma^2 = 0, \tag{G.99}$$

根据$\mathfrak{a}_{t+2}$在(G.14k)的定义。这里可以回顾一下Λ_n，Λ_d来自(3.8)。这就证明了$\mathbb{E}\big[\mathsf{F}_q^{t+1}{}_{,\text{ out},1}\mathsf{F}_{q,\text{ out},2}\big]=0$；结合(G.56)，即可确认条目 1 在时刻$t+1$的有效性。

给定新的随机变量$\mathsf{F}_q^{t+1}{}_{,\text{在}}$，$\mathsf{F}_q^{t+1}{}_{,\text{ out}}$，协方差矩阵$\Sigma_p^{t+2}{}_{,\text{在}}$，$\Sigma_p^{t+2}{}_{,\text{ out}}$可根据(G.14l)计算。将这些矩阵写成分块形式：

$$\Sigma_{p,\text{in}}^{t+2} = \begin{bmatrix} \Sigma_{p,\text{in}}^{t+1} & \sigma_{p,\text{in}}^{t+1} \\ (\sigma_{p,\text{in}}^{t+1})^\top & s_{p,\text{in}}^{t+2} \end{bmatrix} \in \mathbb{R}^{(t+2)\times(t+2)}, \quad \Sigma_{p,\text{out}}^{t+2} = \begin{bmatrix} \Sigma_{p,\text{out}}^{t+1} & \sigma_{p,\text{out}}^{t+1} \\ (\sigma_{p,\text{out}}^{t+1})^\top & s_{p,\text{out}}^{t+2} \end{bmatrix} \in \mathbb{R}^{(t+2)\times(t+2)}, \tag{G.100}$$

其中引入了记号$\sigma_p^{t+1}{}_{,\text{in}}$ $\sigma_p^{t+1}{}_{,\text{out}} \in \mathbb{R}^{t+1}$，$s_p^{t+2}{}_{,\text{in}}$ $s_p^{t+2}{}_{,\text{out}} \in \mathbb{R}$。在记号上，我们使用集合$\{1,2,\cdots,t+1\}$对$\sigma_p^{t+1}{}_{,\text{in}}$ $\sigma_p^{t+1}{}_{,\text{out}}$的元素进行索引，这与$\Sigma_p^{t+2}{}_{,\text{in}}$ $\Sigma_p^{t+2}{}_{,\text{out}}$的索引方式一致。现在验证$\Sigma_p^{t+2}{}_{,\text{in}}$的可逆性。回顾定义 (G.14l)，由于归纳假设中条目 2 所保证的$\Sigma_p^{t+1}{}_{,\text{in}}$可逆，$\Sigma_p^{t+2}{}_{,\text{in}}$为奇异矩阵当且仅当存在$\alpha_0, \cdots, \alpha_t \in \mathbb{R}$，使得几乎必然有，

$$\mathsf{F}_{q,\text{in}}^{t+1} = \sum_{s=0}^{t} \alpha_s \mathsf{F}_{q,\text{in}}^{s}.$$

根据F_q^{t+1}(G.14h)的定义$_{,\text{in}}$，上述条件等价于

$$\Phi_{t+1}\mathsf{Q}_{\text{in},1}^{t+1} + \widetilde{\Phi}_{t+1}\mathsf{Q}_{\text{out}}^{t+1} = \sum_{s=0}^{t} \alpha_s \mathsf{F}_{q,\text{in}}^{s} - \widetilde{\mathfrak{b}}_{t+1}\Phi_{t+1}\mathsf{Q}_{\text{in},2} - \widetilde{\mathfrak{a}}_{t+1}\widetilde{\Phi}_{t+1}\Lambda_d\mathsf{Q}_{\text{in},2} + \mathfrak{b}_{t+2}\mathsf{Q}_{\text{in},2}. \tag{G.101}$$

~假设（A29）确保随机变量Φ_{t+1}、Φ_{t+1}中至少有一个不是几乎必然为常数。设Φ_{t+1}满足这一条件。利用$\mathsf{Q}^{t+1}{}_{\text{in},1}$在(G.76a)中给出的表示，还可以将(G.101)进一步写为

$$\begin{aligned}\Phi_{t+1}\mathsf{U}_{q,\text{in}}^{t+1} = &\sum_{s=0}^{t} \alpha_s \mathsf{F}_{q,\text{in}}^{s} - \widetilde{\mathfrak{b}}_{t+1}\Phi_{t+1}\mathsf{Q}_{\text{in},2} - \widetilde{\mathfrak{a}}_{t+1}\widetilde{\Phi}_{t+1}\Lambda_d\mathsf{Q}_{\text{in},2} + \mathfrak{b}_{t+2}\mathsf{Q}_{\text{in},2} \\ &- \widetilde{\Phi}_{t+1}\mathsf{Q}_{\text{out}}^{t+1} - \Phi_{t+1}\sum_{s=0}^{t} \mathsf{Q}_{\text{in},1}^{s}\big((\Sigma_{q,\text{in}}^{t})^{-1}\sigma_{q,\text{in}}^{t}\big)_s.\end{aligned} \tag{G.102}$$

然后，我们对Φ_{t+1}的一个实现(ϕ, f)以及上述方程右侧进行条件化，使得$\phi \neq 0$。由于我们假设Φ_{t+1}不是常数，因此这样的ϕ存在。由于U^{t+1} in,1 与 (G.102) 中的其他随机变量独立，因此条件化不会改变其分布。如 (G.80) 所示，U_q^{t+1},in 服从方差严格为正的高斯分布，因此我们有$\mathbb{P}\Big(\mathsf{U}_q^{t+1}{}_{,\text{in}} \neq f/\phi\Big) > 0$。这意味着 (G.102) 不可能几乎必然成立，因此$\Sigma_p^{t+2}{}_{,\text{in}}$必须是非奇异的。另一方面，如果$\Phi_{t+1}$是常数，~假设（A29）可确保$\Phi_{t+1}$不是常数。我们可以采用相同的论证，但此时使用 (G.91) 中确定的$\mathsf{Q}_{\text{out}}^{t+1}$表示，并将$\widetilde{\Phi}_{t+1}\mathsf{Q}_{\text{out}}^{t+1}$留在 (G.102) 的左侧。由 (G.94) 所示，$\mathsf{U}_q^{t+1}{}_{,\text{out}}$的方差为正，因此我们可以得出相同的结论：$\Sigma_p^{t+2}{}_{,\text{in}}$可逆。对于$\Sigma_p^{t+2}{}_{,\text{out}}$，也可采用类似的证明，此处略去细节。

针对 p^{t+2} 的归纳**步骤**。我们重新定义 U， V 为

$$U = \begin{bmatrix} q_{\text{in},2} & q_{\text{in},1}^0 & \cdots & q_{\text{in},1}^{t+1} & f_{q,\text{in}}^0 & \cdots & f_{q,\text{in}}^t \end{bmatrix} \in \mathbb{R}^{d\times 2(t+2)}, \tag{G.103}$$

$$V = \begin{bmatrix} f_{p,\text{in},2} & f_{p,\text{in},1}^0 & \cdots & f_{p,\text{in},1}^{t+1} & p_{\text{in}}^1 & \cdots & p_{\text{in}}^{t+1} \end{bmatrix} \in \mathbb{R}^{d\times 2(t+2)}. \tag{G.104}$$

在给定所有过去迭代和侧信息的条件下， p^{t+2} 在中的条件分布与 p^{t+2} 在中在条件 $V = QU$下的条件分布相同。由定理 K.6，对每一个充分大的 d，后一条件分布可写为

$$p_{\text{in}}^{t+2} \,\big|\, \{V = QU\} \;=\; Q f_{q,\text{in}}^{t+1} \,\Big|\, \{V = QU\} \overset{d}{=} V(U^\top U)^{-1} U^\top f_{q,\text{in}}^{t+1} + \Pi_{V^\perp} \widetilde{Q} \Pi_{U^\perp}^\top f_{q,\text{in}}^{t+1}, \tag{G.105}$$

~其中 $Q \sim \text{Haar}(\mathbb{O}(d-2(t+2)))$ 与其余一切独立。接下来，我们计算$d^{-1}U^\top U$与 $d^{-1}U^\top f_q^{t+1}$ 的极限，在中。特别地，(G.106) 下文表明 $U^\top U$在 (G.105) 右端对每一个充分大的 d都是可逆的。

计算 $d^{-1}U^\top U$ 的极限的步骤与 (G.69)的步骤类似。与 (G.58)相比， U在上文 (G.103) 中所定义的量多了一列 q^{t+1} 在，1中。因此我们只需检验q^{t+1} 在，1 与 U的其他列之间的内积。由上文所示的收敛结果 (G.78) ，对 $0 \le s \le t$，我们有

$$\frac{1}{d}\left\langle q_{\text{in},1}^{t+1}, q_{\text{in},2} \right\rangle \to 0, \qquad \frac{1}{d}\left\langle q_{\text{in},1}^{t+1}, q_{\text{in},1}^{s} \right\rangle \to (\sigma_{q,\text{in}}^t)_s, \qquad \frac{1}{d}\left\langle q_{\text{in},1}^{t+1}, q_{\text{in},1}^{t+1} \right\rangle \to s_{q,\text{in}}^{t+1}.$$

此外，通过与(G.62)中第一项类似的分析，我们有

$$\frac{1}{d}\left\langle q_{\text{in},1}^{t+1}, f_{q,\text{in}}^{s} \right\rangle \to 0.$$

结合这些事实与(G.69)，我们得到

$$\frac{1}{d}U^\top U \to \begin{bmatrix} \rho & 0_{1\times(t+2)} & 0_{1\times(t+1)} \\ 0_{(t+2)\times 1} & \Sigma_{q,\text{in}}^{t+1} & 0_{(t+2)\times(t+1)} \\ 0_{(t+1)\times 1} & 0_{(t+1)\times(t+2)} & \Sigma_{p,\text{in}}^{t+1} \end{bmatrix} \in \mathbb{R}^{2(t+2)\times 2(t+2)}. \tag{G.106}$$

右侧的矩阵是严格正定的，这由以下性质保证： ρ 由假设 (A1)保证为正， Σ_q^{t+1} 的正定性，在于(G.57)附近所示，以及Σ_p^{t+1} 的正定性，该性质由条目 2中的归纳假设保证。

转到$d^{-1}U^\top f_q^{t+1}$，对于，由上文给出的收敛结果(G.95) 以及与(G.62)中的两项类似的分析，对于每个 $0 \le s \le t+1$和 $0 \le r \le t$，

$$\frac{1}{d}\left\langle q_{\text{in},2}, f_{q,\text{in}}^{t+1} \right\rangle \to 0, \qquad \frac{1}{d}\left\langle q_{\text{in},1}^{s}, f_{q,\text{in}}^{t+1} \right\rangle \to 0, \qquad \frac{1}{d}\left\langle f_{q,\text{in}}^{r}, f_{q,\text{in}}^{t+1} \right\rangle \to (\sigma_{p,\text{in}}^{t+1})_r.$$

因此，

$$\frac{1}{d}U^\top f_{q,\text{in}}^{t+1} \to \begin{bmatrix} 0 \\ 0_{t+2} \\ \sigma_{p,\text{in}}^{t+1} \end{bmatrix} \in \mathbb{R}^{2(t+2)\times 1}. \tag{G.107}$$

将(G.106) 和(G.107) 代入(G.105)的右端，利用(G.105) 中的第二项由于定理 K.5而具有高斯极限，记为 U_p^{t+2}，在，且独立于其他一切，并回顾归纳假设(G.15) 于条目 4中，即

$$V \xrightarrow{W} \begin{bmatrix} \mathsf{F}_{p,\text{in},2} & \mathsf{F}_{p,\text{in},1}^0 & \cdots & \mathsf{F}_{p,\text{in},1}^{t+1} & \mathsf{P}_{\text{in}}^1 & \cdots & \mathsf{P}_{\text{in}}^{t+1} \end{bmatrix},$$

并应用定理 em K.4，我们有 p^{t+2} 在中的条件分布依Wasserstein收敛到概率分布o f

$$\begin{bmatrix}\mathsf{F}_{p,\mathrm{in},2} & \mathsf{F}^0_{p,\mathrm{in},1} & \cdots & \mathsf{F}^{t+1}_{p,\mathrm{in},1} & \mathsf{P}^1_{\mathrm{in}} & \cdots & \mathsf{P}^{t+1}_{\mathrm{in}}\end{bmatrix}\begin{bmatrix}\rho^{-1} & 0_{1\times(t+2)} & 0_{1\times(t+1)}\\ 0_{(t+2)\times 1} & (\Sigma^{t+1}_{q,\mathrm{in}})^{-1} & 0_{(t+2)\times(t+1)}\\ 0_{(t+1)\times 1} & 0_{(t+1)\times(t+2)} & (\Sigma^{t+1}_{p,\mathrm{in}})^{-1}\end{bmatrix}\begin{bmatrix}0\\ 0_{t+2}\\ \sigma^{t+1}_{p,\mathrm{in}}\end{bmatrix}+\mathsf{U}^{t+2}_{p,\mathrm{in}}$$
$$=\sum_{s=1}^{t+1}\mathsf{P}^s_{\mathrm{in}}\left((\Sigma^{t+1}_{p,\mathrm{in}})^{-1}\sigma^{t+1}_{p,\mathrm{in}}\right)_s+\mathsf{U}^{t+2}_{p,\mathrm{in}}. \tag{G.108}$$

上述随机变量与 $(\mathsf{P}^{s+1}\ \text{in}\ )_{0\le s\le t}$ 联合高斯，并且它们与 A 以及 $(\mathsf{Q}^s_{\mathrm{in}})_{0\le s\le t+1}$、$(\mathsf{Q}^s_{\mathrm{out}})_{0\le s\le t+1}$独立。可以以类似于 (G.77) 的方式验证，上述随机变量与 $(\mathsf{P}^{s+1}\ \text{in}\ )_{0\le s\le t}$ 具有所需的协方差矩阵 Σ_p^{t+2} ,in，因此它正是随机变量 P^{t+2} in。此外，利用舒尔补公式、(G.100) 附近所示 Σ_p^{t+2} ,in 的正定性、(G.108) 中两项之间的独立性，并应用类似于 (G.79) 的论证，可知 U_p^{t+2} ,in 的方差严格为正。为避免重复，我们省略这些细节。这证明了条目 3 中 (G.42) 的第一个陈述。

上述推理确立了如下收敛结果：

$$\begin{aligned}&\begin{bmatrix}q_{\mathrm{in},2} & \boldsymbol{q}^{t+1}_{\mathrm{in},1} & \boldsymbol{f}^{t+1}_{q,\mathrm{in}} & \boldsymbol{p}^{t+2}_{\mathrm{in}} & f_{p,\mathrm{in},2} & \boldsymbol{f}^{t+1}_{p,\mathrm{in},1} & w_{p,\mathrm{in}} & w_{q,\mathrm{in}}\end{bmatrix}\\ &\quad\xrightarrow{W}\begin{bmatrix}\mathsf{Q}_{\mathrm{in},2} & \mathbf{Q}^{t+1}_{\mathrm{in},1} & \mathbf{F}^{t+1}_{q,\mathrm{in}} & \mathbf{P}^{t+2}_{\mathrm{in}} & \mathsf{F}_{p,\mathrm{in},2} & \mathbf{F}^{t+1}_{p,\mathrm{in},1} & \mathsf{W}^\top_{p,\mathrm{in}} & \mathsf{W}^\top_{q,\mathrm{in}}\end{bmatrix}.\end{aligned} \tag{G.109}$$

针对 p^{t+2}_{out} 的归纳步骤。重新定义 U，V为

$$U=\begin{bmatrix}q^0_{\mathrm{out}} & \cdots & q^{t+1}_{\mathrm{out}} & f_{q,\mathrm{out},2} & f^0_{q,\mathrm{out},1} & \cdots & f^t_{q,\mathrm{out},1}\end{bmatrix}\in\mathbb{R}^{n\times 2(t+2)}, \tag{G.110}$$
$$V=\begin{bmatrix}f^0_{p,\mathrm{out}} & \cdots & f^{t+1}_{p,\mathrm{out}} & p_{\mathrm{out},2} & p^1_{\mathrm{out},1} & \cdots & p^{t+1}_{\mathrm{out},1}\end{bmatrix}\in\mathbb{R}^{n\times 2(t+2)}. \tag{G.111}$$

在给定所有过去迭代和侧信息的条件下，p^{t+2}_{out} 的条件分布与 p^{t+2}_{out} 在条件 $V=OU$下的条件分布相同。随后我们使用定理 K.6 来刻画后一条件分布：对所有充分大的 n,

$$p^{t+2}_{\mathrm{out}}\,\big|\,\{V=OU\}=Of^{t+1}_{q,\mathrm{out}}\,\big|\,\{V=OU\}\overset{d}{=}V(U^\top U)^{-1}U^\top f^{t+1}_{q,\mathrm{out}}+\Pi_{V^\perp}\widetilde{O}\Pi^\top_{U^\perp}f^{t+1}_{q,\mathrm{out}}, \tag{G.112}$$

其中 $\widetilde{O}\sim$ 服从 Haar($\mathbb{O}(n-2(t+2))$) 且与其余一切独立。我们再次计算极限 $n^{-1}U^\top U$与 $n^{-1}U^\top f^{t+1}_{q,\ \mathrm{out}}$。特别地，(G.113) 在下文中表明，对所有充分大的 n，$U^\top U$的可逆性，而它出现在 (G.112) 的右侧。

首先考虑 $n^{-1}U^\top U$。注意到 U在 (G.110) 中定义，且多出一列 q^{t+1}_{out}，相较于 U在 (G.81)中的定义。因此，为计算 $n^{-1}U^\top U$的极限，我们只需考察 q^{t+1}_{out}与 U的其他列之间的内积。根据已经证明的收敛结果 (G.93)，对任意 $0\le s\le t$,

$$\frac{1}{n}\langle q^{t+1}_{\mathrm{out}},q^s_{\mathrm{out}}\rangle\to(\sigma^t_{q,\mathrm{out}})_s,\qquad \frac{1}{n}\langle q^{t+1}_{\mathrm{out}},q^{t+1}_{\mathrm{out}}\rangle\to s^{t+1}_{q,\mathrm{out}},$$
$$\frac{1}{n}\langle q^{t+1}_{\mathrm{out}},f_{q,\mathrm{out},2}\rangle\to 0,\qquad \frac{1}{n}\langle q^{t+1}_{\mathrm{out}},f^s_{q,\mathrm{out},1}\rangle\to 0,$$

其中，最后两个极限可用与 (G.85) 类似的方法处理。将这些极限与 (G.87) 结合可得

$$\frac{1}{n}U^\top U\to\begin{bmatrix}\Sigma^{t+1}_{q,\mathrm{out}} & 0_{(t+2)\times 1} & 0_{(t+2)\times(t+1)}\\ 0_{1\times(t+2)} & \sigma^2 & 0_{1\times(t+1)}\\ 0_{(t+1)\times(t+2)} & 0_{(t+1)\times 1} & \Sigma^{t+1}_{p,\mathrm{out}}\end{bmatrix}\in\mathbb{R}^{2(t+2)\times 2(t+2)}. \tag{G.113}$$

右侧的矩阵由$\Sigma^{t+1}_{q\ \text{的正定性（见out}}$ 约(G.57)） σ^2 在(3.9) 中指出的正性，以及$\Sigma^{t+1}_{p,\ \mathrm{out}}$ 由条目 2中的归纳假设可知的正定性，严格正定。

接下来考虑 $n^{-1}U^\top f_q^{t+1}{}_{,\,\mathrm{out}}$。对于任意 $0 \le s \le t+1$ 和 $0 \le r \le t$,

$$\frac{1}{n}\langle f_{q,\mathrm{out},1}^{t+1}, q_{\mathrm{out}}^{s}\rangle \to 0, \qquad \frac{1}{n}\langle f_{q,\mathrm{out},1}^{t+1}, f_{q,\mathrm{out},2}\rangle \to 0, \qquad \frac{1}{n}\langle f_{q,\mathrm{out},1}^{t+1}, f_{q,\mathrm{out},1}^{r}\rangle \to (\sigma_{p,\mathrm{out}}^{t+1})_r, \tag{G.114}$$

其中，第一个极限可用与(G.85)中的第一项类似的方法处理。第二个极限由$\mathbb{E}\big[\mathsf{F}_q^{t+1}{}_{,\mathrm{out},1}\mathsf{F}_{q,\mathrm{out},2}\big] = 0$ （见(G.99)）得到；第三个极限则直接由先前给出的收敛结果(G.96)得出。此外，类似于(G.85)中的第二项，可以证明

$$\frac{1}{n}\langle f_{q,\mathrm{out},2}, q_{\mathrm{out}}^{s}\rangle \to 0. \tag{G.115}$$

另请回顾(G.84a)中的第二个等式。由此结合(G.114) 和(G.115)，可得

$$\frac{1}{n}U^\top f_{q,\mathrm{out}}^{t+1} \to \begin{bmatrix} 0_{t+2} & 0_{t+2} \\ 0 & \sigma^2 \\ \sigma_{p,\mathrm{out}}^{t+1} & 0_{t+2} \end{bmatrix} \in \mathbb{R}^{2(t+2)\times 2}. \tag{G.116}$$

根据定理 K.5，(G.112) 中的第二项存在一个高斯分布极限，记为 $\begin{bmatrix}\mathsf{U}_p^{t+2}{}_{,\mathrm{out}} & 0\end{bmatrix} \in \mathbb{R}^{1\times 2}$。第二个分量为零，因为 $\Pi_{U^\perp}$ 的每一列都与列$f_{q,\mathrm{out},2}$ 在 U 中正交，因此$\Pi_{U^\perp}^\top f_{q,\mathrm{out},2} = 0_{n-2(t+2)}$。结合这一观察与(G.113) 和(G.116)，并回顾归纳假设(G.16) 在条目 4 中指出

$$V \xrightarrow{W} \begin{bmatrix}\mathsf{F}_{p,\mathrm{out}}^0 & \cdots & \mathsf{F}_{p,\mathrm{out}}^{t+1} & \mathsf{P}_{\mathrm{out},2} & \mathsf{P}_{\mathrm{out},1}^1 & \cdots & \mathsf{P}_{\mathrm{out},1}^{t+1}\end{bmatrix},$$

并应用 K.4 定理，得到 (G.112) 右侧的 Wasserstein 极限:

$$\begin{aligned}
&p_{\mathrm{out}}^{t+2} \,\big|\, \{V = OU\} \\
&\xrightarrow{W} \begin{bmatrix}\mathsf{F}_{p,\mathrm{out}}^0 & \cdots & \mathsf{F}_{p,\mathrm{out}}^{t+1} & \mathsf{P}_{\mathrm{out},2} & \mathsf{P}_{\mathrm{out},1}^1 & \cdots & \mathsf{P}_{\mathrm{out},1}^{t+1}\end{bmatrix} \begin{bmatrix} \Sigma_{q,\mathrm{out}}^{t+1} & 0_{(t+2)\times 1} & 0_{(t+2)\times(t+1)} \\ 0_{1\times(t+2)} & \sigma^2 & 0_{1\times(t+1)} \\ 0_{(t+1)\times(t+2)} & 0_{(t+1)\times 1} & \Sigma_{p,\mathrm{out}}^{t+1} \end{bmatrix}^{-1} \\
&\qquad \times \begin{bmatrix} 0_{t+2} & 0_{t+2} \\ 0 & \sigma^2 \\ \sigma_{p,\mathrm{out}}^{t+1} & 0_{t+2} \end{bmatrix} + \begin{bmatrix}\mathsf{U}_{p,\mathrm{out}}^{t+2} & 0\end{bmatrix} \\
&= \begin{bmatrix} \sum_{s=0}^{t} \mathsf{P}_{\mathrm{out},1}^{s+1}\big((\Sigma_{p,\mathrm{out}}^{t+1})^{-1}\sigma_{p,\mathrm{out}}^{t+1}\big)_{s+1} + \mathsf{U}_{p,\mathrm{out}}^{t+2} & \mathsf{P}_{\mathrm{out},2} \end{bmatrix}.
\end{aligned}$$

我们声称，右侧的第一项恰好是 $\mathsf{P}_{\mathrm{out}}^{t+2}{}_{,1}$。为说明这一点，我们注意到以下事实。第一项中的随机变量与 $(\mathsf{P}_{\mathrm{out}}^{s+1}{}_{,1})_{0\le s\le t}$ 联合高斯，其协方差矩阵为 $\Sigma_p^{t+2}{}_{,\mathrm{out}}$ （计算过程与 (G.92) 类似）。这些联合高斯随机变量与 A、$(\mathsf{Q}_{\mathrm{in}}^s)_{0\le s\le t+1}$ 和 $(\mathsf{Q}_{\mathrm{out}}^s)_{0\le s\le t+1}$ 相互独立。此外，它们也与 $\mathsf{P}_{\mathrm{out}}{}_{,2}$ 独立，因为 $(\mathsf{P}_{\mathrm{out}}^{s+1}{}_{,1})_{0\le s\le t}$ 和 $\mathsf{U}_p^{t+2}{}_{,\mathrm{out}}$ 都与 $\mathsf{P}_{\mathrm{out}}{}_{,2}$ 独立。与前面一样，$\mathsf{U}_p^{t+2}{}_{,\mathrm{out}}$ 的方差严格为正；这里用到了 (G.100) 附近所示的 $\Sigma_p^{t+2}{}_{,\mathrm{out}}$ 的正定性。这就证明了条目 3 中 (G.42) 的第二个陈述。

因此，我们已经证明

$$\begin{aligned}
&\begin{bmatrix}\boldsymbol{q}_{\mathrm{out}}^{t+1} & f_{q,\mathrm{out},2} & \boldsymbol{f}_{q,\mathrm{out},1}^{t} & p_{\mathrm{out},2} & \boldsymbol{p}_{\mathrm{out},1}^{t+2} & \boldsymbol{f}_{p,\mathrm{out}}^{t+1} & w_{p,\mathrm{out}} & w_{q,\mathrm{out}}\end{bmatrix} \\
&\quad \xrightarrow{W} \begin{bmatrix}\mathbf{Q}_{\mathrm{out}}^{t+1} & \mathsf{F}_{q,\mathrm{out},2} & \mathbf{F}_{q,\mathrm{out},1}^{t} & \mathsf{P}_{\mathrm{out},2} & \mathbf{P}_{\mathrm{out},1}^{t+2} & \mathbf{F}_{p,\mathrm{out}}^{t+1} & \mathsf{W}_{p,\mathrm{out}}^\top & \mathsf{W}_{q,\mathrm{out}}^\top\end{bmatrix}.
\end{aligned} \tag{G.117}$$

关于f_p^{t+2}的归纳步骤，f_p^{t+2}中out。根据更新规则(G.5d)、前两步证明的收敛结果(G.109)和(G.117)，以及定理K.2，我们立即得到

$$\begin{aligned}
&\begin{bmatrix} q_{\mathrm{in},2} & \boldsymbol{q}_{\mathrm{in},1}^{t+1} & \boldsymbol{f}_{q,\mathrm{in}}^{t+1} & \boldsymbol{p}_{\mathrm{in}}^{t+2} & f_{p,\mathrm{in},2} & \boldsymbol{f}_{p,\mathrm{in},1}^{t+2} & w_{p,\mathrm{in}} & w_{q,\mathrm{in}} \end{bmatrix} \\
&\quad \xrightarrow{W} \begin{bmatrix} \mathsf{Q}_{\mathrm{in},2} & \mathbf{Q}_{\mathrm{in},1}^{t+1} & \mathbf{F}_{q,\mathrm{in}}^{t+1} & \mathbf{P}_{\mathrm{in}}^{t+2} & \mathsf{F}_{p,\mathrm{in},2} & \mathbf{F}_{p,\mathrm{in},1}^{t+2} & \mathsf{W}_{p,\mathrm{in}}^{\top} & \mathsf{W}_{q,\mathrm{in}}^{\top} \end{bmatrix}, \\
&\begin{bmatrix} \boldsymbol{q}_{\mathrm{out}}^{t+1} & f_{q,\mathrm{out},2} & \boldsymbol{f}_{q,\mathrm{out},1}^{t+1} & p_{\mathrm{out},2} & \boldsymbol{p}_{\mathrm{out},1}^{t+2} & \boldsymbol{f}_{p,\mathrm{out}}^{t+2} & w_{p,\mathrm{out}} & w_{q,\mathrm{out}} \end{bmatrix} \\
&\quad \xrightarrow{W} \begin{bmatrix} \mathbf{Q}_{\mathrm{out}}^{t+1} & \mathsf{F}_{q,\mathrm{out},2} & \mathbf{F}_{q,\mathrm{out},1}^{t+1} & \mathsf{P}_{\mathrm{out},2} & \mathbf{P}_{\mathrm{out},1}^{t+2} & \mathbf{F}_{p,\mathrm{out}}^{t+2} & \mathsf{W}_{p,\mathrm{out}}^{\top} & \mathsf{W}_{q,\mathrm{out}}^{\top} \end{bmatrix},
\end{aligned}$$

从而最终完成整个归纳论证。 □

G.2 定理 G.2 的证明

我们对原始 GVAMP 迭代 $\Phi_t(X^\top X)$ $\widetilde{\Phi}_t(X)$ $\Psi_t(XX^\top)$ $\widetilde{\Psi}_t(X)$ 中的随机矩阵和去噪函数 f_{t+1}, g_{t+1}进行扰动，并证明扰动后的版本满足 (4.19) 假设 (A27),(A29) 和 (A30)，因此定理 G.3 适用。随后，我们证明这两个迭代及其对应的状态演化参数彼此接近，并令扰动趋于零，从而完成证明。

固定一个小常数 $\zeta > 0$和任意整数 $\Gamma \in \mathbb{Z}_{\geq 0}$。令 $M_t \overset{\text{i.i.d}}{\sim}$ 高斯正交系综(d)、$N_t \overset{\text{i.i.d}}{\sim}$ 高斯正交系综(n)、$\widetilde{M}_t \overset{\text{i.i.d}}{\sim}$ Ginibre 系综(n,d)、$\widetilde{N}_t \overset{\text{i.i.d}}{\sim}$ Ginibre 系综(n,d)、$s^t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0_d, I_d)$、$q^t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0_n, I_n)$（对所有 $t \in \{0,1,\cdots,\Gamma+1\}$）彼此独立，且与其他所有对象独立。[3] 定义增广侧信息矩阵 $\Theta = \begin{bmatrix}\Theta & s^0 & \cdots & s^{\Gamma+1}\end{bmatrix} \in \mathbb{R}^{d\times(d'+\Gamma+2)}$ 和 $\Xi = \begin{bmatrix}\Xi & q^0 & \cdots & q^{\Gamma+1}\end{bmatrix} \in \mathbb{R}^{n\times(n'+\Gamma+2)}$。令扰动初始化量如下

$$\widetilde{r}_\zeta^0 = \widetilde{r}^0 + \zeta s^0, \qquad \widetilde{p}_\zeta^0 = \widetilde{p}^0 + \zeta q^0, \tag{G.118}$$

其中 $\widetilde{r}^0, \widetilde{p}^0$ 是 (4.19)的初始化量。定义

$$f_0^\zeta(b,\theta,\begin{bmatrix}s_0 & \cdots & s_{\Gamma+1}\end{bmatrix}) = f_0(b,\theta) + \zeta s_0, \qquad g_0^\zeta(z,e,\xi,\begin{bmatrix}q_0 & \cdots & q_{\Gamma+1}\end{bmatrix}) = g_0(z,e,\xi) + \zeta q_0,$$

显然，扰动初始化满足 假设 (A28)，其中 $\widetilde{r}_\zeta^0, \widetilde{p}_\zeta^0$的 Wasserstein 极限为

$$\widetilde{\mathsf{R}}_0^\zeta = \widetilde{\mathsf{R}}_0 + \zeta \mathsf{S}_0, \qquad \widetilde{\mathsf{P}}_0^\zeta = \widetilde{\mathsf{P}}_0 + \zeta \mathsf{Q}_0, \tag{G.119}$$

其中 $\mathsf{S}_0, \mathsf{Q}_0$ 是独立于 $\mathsf{B}_*, \mathsf{Z}, \mathsf{E}$ Θ, Ξ的独立同分布标准高斯变量。对于每个 $0 \leq t \leq \Gamma$，定义扰动去噪函数 $f_{t+1}^\zeta\colon \mathbb{R}^{2+d'+\Gamma+2} \to \mathbb{R}$ 和 $g_{t+1}^\zeta\colon \mathbb{R}^{3+n'+\Gamma+2} \to \mathbb{R}$ 如下

$$\begin{aligned}
f_{t+1}^\zeta(r_\zeta; b, \begin{bmatrix}\theta & s_0 & \cdots & s_{\Gamma+1}\end{bmatrix}) &= f_{t+1}(r_\zeta; b, \theta) + \zeta s_{t+1} - \mathbb{E}\Big[f_{t+1}'(\mathsf{R}_{t+1}^\zeta; \mathsf{B}_*, \Theta)\Big] r_\zeta, \\
g_{t+1}^\zeta(p_\zeta; z, e, \begin{bmatrix}\xi & q_0 & \cdots & q_{\Gamma+1}\end{bmatrix}) &= g_{t+1}(p_\zeta; z, e, \xi) + \zeta q_{t+1} - \mathbb{E}\Big[g_{t+1}'(\mathsf{P}_{t+1}^\zeta; \mathsf{Z}, \mathsf{E}, \Xi)\Big] p_\zeta,
\end{aligned} \tag{G.120}$$

其中，$r_\zeta, p_\zeta, b, z, e, s_0, \cdots, s_{\Gamma+1}, q_0, \cdots, q_{\Gamma+1} \in \mathbb{R}$, $\theta \in \mathbb{R}^{d'}$, $\xi \in \mathbb{R}^{n'}$和 f_{t+1}, g_{t+1}是式 (4.19) 中的去噪函数。随机变量 $\mathsf{R}_{t+1}^\zeta, \mathsf{P}_{t+1}^\zeta$ 在下文的式 (G.126) 中给出。特别地，定义 (G.120) 中的负修正项确保，当分别根据 R_{t+1}^ζ, P_{t+1}^ζ 的概率分布计算 $f_{t+1}^\zeta, g_{t+1}^\zeta$ 关于 r_ζ, p_ζ 的导数的期望时，满足假设 (A8) 所要求的散度为零条件。由于 f_{t+1}, g_{t+1} 满足假设（A10），因此 $f_{t+1}^\zeta, g_{t+1}^\zeta$ 满足假设 (A27)。对于任意的 $\beta_0, \cdots, \beta_t \in \mathbb{R}$，设

$$\widetilde{\mathsf{R}}_{t+1}^\zeta - \widetilde{b}_{t+1}^\zeta \mathsf{B}_* = \sum_{s=0}^{t} \beta_s \Big(\widetilde{\mathsf{R}}_s^\zeta - \widetilde{b}_s^\zeta \mathsf{B}_*\Big). \tag{G.121}$$

[3]稍微滥用一下 D.2.1 节中的术语，我们用 Gin(n,d) 表示（矩形）Ginibre 系综，它是定义在 $n \times d$实矩形矩阵 M上的概率测度，满足 $M_{i,j} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, 1/d)$，对所有 $(i,j) \in [n] \times [d]$均成立。

重新整理各项后，这等价于

$$\zeta \mathsf{S}_{t+1} = -f_{t+1}(\mathsf{R}^{\zeta}_{t+1}; \mathsf{B}_*, \Theta) + \mathbb{E}\Big[f'_{t+1}(\mathsf{R}^{\zeta}_{t+1}; \mathsf{B}_*, \Theta)\Big] + \widetilde{b}^{\zeta}_{t+1}\mathsf{B}_* + \sum_{s=0}^{t} \beta_s \Big(\widetilde{\mathsf{R}}^{\zeta}_s - \widetilde{b}^{\zeta}_s \mathsf{B}_*\Big). \tag{G.122}$$

由于 $\zeta > 0$ 和 $\mathsf{S}_{t+1} \sim \mathcal{N}(0,\ 1)$ 与(G.122)右侧的所有项相互独立，因此(G.122)，(G.121) 不可能几乎必然成立。类似的论证也适用于 $\widetilde{\mathsf{P}}^{\zeta}_{t+1}$。因此 f^{ζ}_{t+1}，g^{ζ}_{t+1} 满足假设（A30）。

利用上述量，我们构造一个扰动 GVAMP 迭代。回顾假设（A2），$X = O\Lambda Q^\top$。以 $\widetilde{r}^0_\zeta$ 初始化，$\widetilde{p}^0_\zeta$ 在(G.118)中，对于每个 $0 \le t \le \Gamma$，扰动 GVAMP 迭代量按如下方式更新：

$$r^{t+1}_\zeta = Q(\Phi_t(\Lambda^\top\Lambda) + \zeta M_t)Q^\top \widetilde{r}^t_\zeta + Q(\widetilde{\Phi}_t(\Lambda) + \zeta\widetilde{M}_t)^\top O^\top \widetilde{p}^t_\zeta, \quad \widetilde{r}^{t+1}_\zeta = f^{\zeta}_{t+1}(r^{t+1}_\zeta; \beta_*, \underline{\Theta}), \tag{G.123a}$$

$$p^{t+1}_\zeta = O(\Psi_t(\Lambda\Lambda^\top) + \zeta N_t)O^\top \widetilde{p}^t_\zeta + O(\widetilde{\Psi}_t(\Lambda) + \zeta\widetilde{N}_t)Q^\top \widetilde{r}^t_\zeta, \quad \widetilde{p}^{t+1}_\zeta = g^{\zeta}_{t+1}(p^{t+1}_\zeta; z, \varepsilon, \underline{\Xi}). \tag{G.123b}$$

令 $m_t \in \mathbb{R}^d$、$n_t \in \mathbb{R}^n$ 分别为 M_t、N_t的特征值向量。令 $\widetilde{m}_t \in \mathbb{R}^d$、$\widetilde{n}_t \in \mathbb{R}^n$ 分别为$\widetilde{\ }$ $\widetilde{\ }$的奇异值向量，M_t、N_t的奇异值向量；必要时如(G.6)所示补上零。于是我们定义

$$\phi^{\zeta}_t = \Phi_t(\lambda_d \circ \lambda_d) + \zeta m_t, \quad \widetilde{\phi}^{\zeta}_t = \widetilde{\Phi}_t(\lambda_d) + \zeta\widetilde{m}_t, \quad \psi^{\zeta}_t = \Psi_t(\lambda_n \circ \lambda_n) + \zeta n_t, \quad \widetilde{\psi}^{\zeta}_t = \widetilde{\Psi}_t(\lambda_n) + \zeta\widetilde{n}_t. \tag{G.124}$$

利用正交不变性以及高斯正交系综和 Ginibre 系综经验谱分布的收敛，假设（A7）和定理 K.2，我们得到(G.7)，其中极限随机变量定义如下：

$$\Phi^{\zeta}_t = \Phi_t(\Lambda^2_d) + \zeta\mathsf{M}_t, \quad \widetilde{\Phi}^{\zeta}_t = \widetilde{\Phi}_t(\Lambda_d) + \zeta\widetilde{\mathsf{M}}_t, \quad \Psi^{\zeta}_t = \Psi_t(\Lambda^2_n) + \zeta\mathsf{N}_t, \quad \widetilde{\Psi}^{\zeta}_t = \widetilde{\Psi}_t(\Lambda_n) + \zeta\widetilde{\mathsf{N}}_t, \tag{G.125}$$

$\widetilde{\ }$ $\widetilde{\ }$其中，M_t、N_t服从半圆分布，M_t、N_t服从具有纵横比 δ 的（平方根推前）Marchenko–Pastur 分布；此外，它们彼此独立，并且对所有 $t \ge 0$ 都与 Λ 独立。由于半圆分布的均值为零，因此 Φ^{ζ}_t、Ψ^{ζ}_t 的均值也为零。由于 $\zeta > 0$ 以及 M_t、$\widetilde{\mathsf{M}}_t$、N_t、$\widetilde{\mathsf{N}}_t$ 的方差均为正，因此有 $\min\Big\{\mathbb{E}\Big[(\Phi^{\zeta}_t)^2\Big]$、$\mathrm{Var}\Big[\widetilde{\Phi}^{\zeta}_t\Big]\Big\} > 0$、$\min\Big\{\mathbb{E}\Big[(\Psi^{\zeta}_t)^2\Big]$、$\mathrm{Var}\Big[\widetilde{\Psi}^{\zeta}_t\Big]\Big\} > 0$。因此，(G.124) 中的向量及 (G.125) 中相应的随机变量满足假设（A29）。

现在 定理 G.3 适用于扰动 GVAMP 迭代 (G.123)。其状态演化参数

$$a^{\zeta}_{t+1},\quad b^{\zeta}_{t+1},\quad \widetilde{a}^{\zeta}_t,\quad \widetilde{b}^{\zeta}_t,\quad \widetilde{\mathsf{P}}^{\zeta}_{t+1},\quad \widetilde{\mathsf{R}}^{\zeta}_{t+1},\quad \mathsf{J}^{\zeta}_{t+1},\quad \mathsf{K}^{\zeta}_{t+1},\quad \mathsf{R}^{\zeta}_{t+1},\quad \mathsf{P}^{\zeta}_{t+1},\quad \forall 0 \le t \le \Gamma, \tag{G.126}$$

可通过 (4.29) 定义，但需作如下修改：将 $\Phi_t(\Lambda^2_d)$ $\widetilde{\Phi}_t(\Lambda_d)$ $\Psi_t(\Lambda^2_n)$ $\widetilde{\Psi}_t(\Lambda_n)$、$f_{t+1}(\mathsf{R}_{t+1}; \mathsf{B}_*, \Theta)$、$g_{t+1}(\mathsf{P}_{t+1}; \mathsf{Z}, \mathsf{E}\ \Xi)$ 替换为 (G.120) 和 (G.125) 中引入的扰动版本 Φ^{ζ}_t、$\widetilde{\Phi}^{\zeta}_t$、$\Psi^{\zeta}_t$、$\widetilde{\Psi}^{\zeta}_t$、$f^{\zeta}_{t+1}(\mathsf{R}^{\zeta}_{t+1}; \mathsf{B}_*, \Theta)$、$g^{\zeta}_{t+1}(\mathsf{P}^{\zeta}_{t+1}; \mathsf{Z}, \mathsf{E}\ \Xi)$。

We将归纳证明以下陈述。对于每个 $t \ge 0$ 以及每个 $1 \le r,\ s \le t$ ，

$$\limsup_{d\to\infty} d^{-1}\big\|r^t\big\|_2^2 < \infty, \qquad \limsup_{n\to\infty} n^{-1}\big\|p^t\big\|_2^2 < \infty, \tag{G.127}$$

$$\lim_{\zeta\to 0}\limsup_{n\to\infty} n^{-1}\big\|p^t - p^t_\zeta\big\|_2^2 = 0, \tag{G.128}$$

$$\lim_{\zeta\to 0} a^{\zeta}_t = a_t, \qquad \lim_{\zeta\to 0} b^{\zeta}_t = b_t, \tag{G.129}$$

$$\lim_{\zeta\to 0} \mathbb{E}\big[\mathsf{J}^{\zeta}_r \mathsf{J}^{\zeta}_s\big] = \mathbb{E}[\mathsf{J}_r \mathsf{J}_s], \qquad \lim_{\zeta\to 0} \mathbb{E}\big[\mathsf{K}^{\zeta}_r \mathsf{K}^{\zeta}_s\big] = \mathbb{E}[\mathsf{K}_r \mathsf{K}_s], \tag{G.130}$$

$$\lim_{\zeta\to 0} \mathbb{E}\Big[f'_t(\mathsf{R}^{\zeta}_t; \mathsf{B}_*, \Theta)\Big] = 0, \qquad \lim_{\zeta\to 0} \mathbb{E}\Big[g'_t(\mathsf{P}^{\zeta}_t; \mathsf{Z}, \mathsf{E}, \Xi)\Big] = 0, \tag{G.131}$$

$$\text{上极限}_{d\to\infty}\ d^{-1}\big\|\widetilde{r}^t\big\|_2^2 < \infty, \qquad \limsup_{n\to\infty} n^{-1}\big\|\widetilde{p}^t\big\|_2^2 < \infty, \tag{G.132}$$

$$\lim_{\zeta\to 0}\limsup_{d\to\infty} d^{-1}\left\|\widetilde{r}^t-\widetilde{r}_\zeta^t\right\|_2^2=0,\qquad \lim_{\zeta\to 0}\limsup_{n\to\infty} n^{-1}\left\|\widetilde{p}^t-\widetilde{p}_\zeta^t\right\|_2^2=0, \tag{G.133}$$

$$\lim_{\zeta\to 0}\widetilde{a}_t^\zeta=\widetilde{a}_t,\qquad \lim_{\zeta\to 0}\widetilde{b}_t^\zeta=\widetilde{b}_t, \tag{G.134}$$

其中(G.127) 至(G.131) 对于 $t=0$为空成立。此外，对于每个 $t\geq 0$,

$$\begin{bmatrix} r^1 & \cdots & r^t & \widetilde{r}^0 & \cdots & \widetilde{r}^t & \beta_* & \Theta\end{bmatrix}\xrightarrow{W_2}\left[\mathsf{R}_1 \quad \cdots \quad \mathsf{R}_t \quad \widetilde{\mathsf{R}}_0 \quad \cdots \quad \widetilde{\mathsf{R}}_t \quad \mathsf{B}_* \quad \Theta^\top\right], \tag{G.135}$$

$$\begin{bmatrix} p^1 & \cdots & p^t & \widetilde{p}^0 & \cdots & \widetilde{p}^t & z & \varepsilon & \Xi\end{bmatrix}\xrightarrow{W_2}\left[\mathsf{P}_1 \quad \cdots \quad \mathsf{P}_t \quad \widetilde{\mathsf{P}}_0 \quad \cdots \quad \widetilde{\mathsf{P}}_t \quad \mathsf{Z} \quad \mathsf{E} \quad \Xi^\top\right], \tag{G.136}$$

其中 $t=0$，这些只是 $\begin{bmatrix}\widetilde{r}^0 & \beta_* & \Theta\end{bmatrix}\xrightarrow{W_2}\left[\widetilde{\mathsf{R}}_0\ \mathsf{B}_* \quad \Theta^\top\right]$，$\begin{bmatrix}\widetilde{p}^0 & z & \varepsilon & \Xi\end{bmatrix}\xrightarrow{W_2}\left[\widetilde{\mathsf{P}}_0\ \mathsf{Z}\ \mathsf{E}\ \Xi^\top\right]$。这将推出 定理 G.2。

基例。对于 $t=0$，（G.135）中的 Wasserstein 收敛由假设 (A28) 给出。这也意味着 $d^{-1}\|\widetilde{r}^0\|_2^2\to\mathbb{E}\left[\widetilde{\mathsf{R}}_0^2\right]<\infty$ 几乎必然成立。根据扰动初始化 (G.118) 的选择，当 $d\to\infty$ 时，$d^{-1}\left\|\widetilde{r}_\zeta^0-\widetilde{r}^0\right\|_2^2=\zeta^2 d^{-1}\|s^0\|_2^2\to 0$ 几乎必然按此顺序成立。根据定义 (4.29e) 以及（G.119）中对 $\widetilde{\mathsf{R}}_0^\zeta$ 的识别，$\widetilde{b}_0^\zeta=\rho^{-1}\mathbb{E}\left[(\widetilde{\mathsf{R}}_0+\zeta\mathsf{S}_0)\mathsf{B}_*\right]=\rho^{-1}\mathbb{E}\left[\widetilde{\mathsf{R}}_0\mathsf{B}_*\right]=\widetilde{b}_0$。同样的论证也适用于 $\widetilde{p}^0$、$\widetilde{p}_\zeta^0$、$\widetilde{a}_0^\zeta$。这证明了（G.132）至（G.134）。

归纳步骤。假设(G.127)至(G.133)在时刻 t之前成立，我们验证它们在时刻$t+1$仍成立。

回顾更新规则(4.19a) 并利用三角不等式,

$$\|r^{t+1}\|_2\leq\|\Phi_t(X^\top X)\|_2\|\widetilde{r}^t\|_2+\left\|\widetilde{\Phi}_t(X)\right\|_2\|\widetilde{p}^t\|_2.$$

根据假设（A9）和归纳假设(G.132)，可得

$$\limsup_{d\to\infty} d^{-1}\left\|r^{t+1}\right\|_2^2<\infty. \tag{G.137}$$

相同 的论证可得上极限$_{n\to\infty}\, n^{-1}\|p^{t+1}\|_2^2<\infty$，从而证明(G.127) 对 $t+1$.

比较 r^{t+1}，r_ζ^{t+1} 在(4.19a) 和(G.123a) 中的更新规则，并应用三角不等式,

$$\left\|r^{t+1}-r_\zeta^{t+1}\right\|_2\leq\|\Phi_t(\Lambda^\top\Lambda)\|_2\|\widetilde{r}^t-\widetilde{r}_\zeta^t\|_2+\zeta\|M_t\|_2\|\widetilde{r}_\zeta^t\|_2+\left\|\widetilde{\Phi}_t(\Lambda)\right\|_2\|\widetilde{p}^t-\widetilde{p}_\zeta^t\|_2+\zeta\left\|\widetilde{M}_t\right\|_2\|\widetilde{p}_\zeta^t\|_2.$$

结合假设（A9）、归纳假设(G.132) 和(G.133)以及 $\|M_t\|_2\to 2$、$\left\|\widetilde{M}_t\right\|_2\to 1+\sqrt{\delta}$，我们得到

$$\lim_{\zeta\to 0}\limsup_{d\to\infty} d^{-1}\left\|r^{t+1}-r_\zeta^{t+1}\right\|_2^2<\infty. \tag{G.138}$$

类似的论证表明$\lim_{\zeta\to 0}$上极限$_{n\to\infty}\, n^{-1}\left\|p^{t+1}-p_\zeta^{t+1}\right\|_2^2<\infty$，从而证明(G.128) 对$t+1$成立。

根据定义 (4.29f) 以及 Φ_t 在 (G.125)中的对应关系,

$$b_{t+1}^\zeta=\mathbb{E}\left[(\widetilde{\Phi}_t(\Lambda_d)+\zeta\widetilde{\mathsf{M}}_t)\Lambda_d\right]\widetilde{a}_t^\zeta=\mathbb{E}\left[\widetilde{\Phi}_t(\Lambda_d)\Lambda_d\right]\widetilde{a}_t^\zeta\to b_{t+1}, \tag{G.139}$$

其中 $\zeta \to 0$ 极限处的收敛由归纳假设 (G.134)得出。类似的论证表明$a_{t+1}^{\zeta} \to a_{t+1}$ 当 $\zeta \to 0$时成立，从而证明了 (G.129) 对 $t+1$成立。

对于任意 $0 \le s \le t$，回顾 (4.29d) 并使用三角不等式，

$$
\begin{aligned}
&\left|\mathbb{E}\left[\mathsf{K}_{s+1}^{\zeta}\mathsf{K}_{t+1}^{\zeta}\right] - \mathbb{E}[\mathsf{K}_{s+1}\mathsf{K}_{t+1}]\right| \\
&\le \left|\mathbb{E}\left[\Phi_s^{\zeta}\Phi_t^{\zeta}\right]\mathbb{E}\left[\widetilde{\mathsf{R}}_s^{\zeta}\widetilde{\mathsf{R}}_t^{\zeta}\right] - \mathbb{E}\left[\Phi_s(\Lambda_d^2)\Phi_t(\Lambda_d^2)\right]\mathbb{E}\left[\widetilde{\mathsf{R}}_s\widetilde{\mathsf{R}}_t\right]\right| + \rho\left|b_{s+1}^{\zeta}b_{t+1}^{\zeta} - b_{s+1}b_{t+1}\right| \\
&+ \left|\mathbb{E}\left[\widetilde{\Phi}_s^{\zeta}\widetilde{\Phi}_t^{\zeta}\right]\left(\mathbb{E}\left[\widetilde{\mathsf{P}}_s^{\zeta}\widetilde{\mathsf{P}}_t^{\zeta}\right] - \widetilde{a}_s^{\zeta}\widetilde{a}_t^{\zeta}\sigma^2\right) - \mathbb{E}\left[\widetilde{\Phi}_s(\Lambda_d)\widetilde{\Phi}_t(\Lambda_d)\right]\left(\mathbb{E}\left[\widetilde{\mathsf{P}}_s\widetilde{\mathsf{P}}_t\right] - \widetilde{a}_s\widetilde{a}_t\sigma^2\right)\right| \\
&+ \rho\left|\mathbb{E}\left[\widetilde{\Phi}_s^{\zeta}\widetilde{\Phi}_t^{\zeta}\Lambda_d^2\right]\widetilde{a}_s^{\zeta}\widetilde{a}_t^{\zeta} - \mathbb{E}\left[\widetilde{\Phi}_s(\Lambda_d)\widetilde{\Phi}_t(\Lambda_d)\Lambda_d^2\right]\widetilde{a}_s\widetilde{a}_t\right| \\
&+ \rho\left|\mathbb{E}\left[\Phi_s^{\zeta}\widetilde{\Phi}_t^{\zeta}\Lambda_d\right]\widetilde{a}_s^{\zeta}\widetilde{b}_t^{\zeta} - \mathbb{E}\left[\Phi_s(\Lambda_d^2)\widetilde{\Phi}_t(\Lambda_d)\Lambda_d\right]\widetilde{a}_s\widetilde{b}_t\right| + \rho\left|\mathbb{E}\left[\Phi_t^{\zeta}\widetilde{\Phi}_s^{\zeta}\Lambda_d\right]\widetilde{a}_t^{\zeta}\widetilde{b}_s^{\zeta} - \mathbb{E}\left[\Phi_t(\Lambda_d^2)\widetilde{\Phi}_s(\Lambda_d)\Lambda_d\right]\widetilde{a}_t\widetilde{b}_s\right|.
\end{aligned}
\tag{G.140}
$$

根据 (G.125)，我们有

$$
\lim_{\zeta\to 0}\left|\mathbb{E}\left[\Phi_s^{\zeta}\Phi_t^{\zeta}\right] - \mathbb{E}\left[\Phi_s(\Lambda_d^2)\Phi_t(\Lambda_d^2)\right]\right| \le \lim_{\zeta\to 0}\zeta^2\mathbb{E}\left[\mathsf{M}_t^2\right] = 0,
$$

其中，当且仅当 $s=t$ 时，不等式取等号；最后一个等号成立是因为半圆分布具有有限方差。类似地，$\mathbb{E}\left[\widetilde{\Phi}_s^{\zeta}\widetilde{\Phi}_t^{\zeta}\right] \to \mathbb{E}\left[\widetilde{\Phi}_s(\Lambda_d)\widetilde{\Phi}_t(\Lambda_d)\right]$, $\mathbb{E}\left[\widetilde{\Phi}_s^{\zeta}\widetilde{\Phi}_t^{\zeta}\Lambda_d^2\right] \to \mathbb{E}\left[\widetilde{\Phi}_s(\Lambda_d)\widetilde{\Phi}_t(\Lambda_d)\Lambda_d^2\right]$,$\mathbb{E}\left[\Phi_s^{\zeta}\widetilde{\Phi}_t^{\zeta}\Lambda_d\right] \to \mathbb{E}\left[\Phi_s(\Lambda_d^2)\widetilde{\Phi}_t(\Lambda_d)\Lambda_d\right]$, $\mathbb{E}\left[\Phi_t^{\zeta}\widetilde{\Phi}_s^{\zeta}\Lambda_d\right] \to \mathbb{E}\left[\Phi_t(\Lambda_d^2)\widetilde{\Phi}_s(\Lambda_d)\Lambda_d\right]$，因为 $\zeta \to 0$。此外，回顾 (G.120) 并使用三角不等式，我们有

$$
\begin{aligned}
&\left|\mathbb{E}\left[\widetilde{\mathsf{R}}_s^{\zeta}\widetilde{\mathsf{R}}_t^{\zeta}\right] - \mathbb{E}\left[\widetilde{\mathsf{R}}_s\widetilde{\mathsf{R}}_t\right]\right| \\
&\le \left|\mathbb{E}\left[f_s(b_s^{\zeta}\mathsf{B}_* + \mathsf{K}_s^{\zeta};\mathsf{B}_*,\Theta)f_t(b_t^{\zeta}\mathsf{B}_* + \mathsf{K}_t^{\zeta};\mathsf{B}_*,\Theta)\right] - \mathbb{E}[f_s(b_s\mathsf{B}_* + \mathsf{K}_s;\mathsf{B}_*,\Theta)f_t(b_t\mathsf{B}_* + \mathsf{K}_t;\mathsf{B}_*,\Theta)]\right| \\
&\quad + \zeta^2 + \left|\mathbb{E}\left[f_s'(\mathsf{R}_s^{\zeta};\mathsf{B}_*,\Theta)\right]\right| \cdot \left|\mathbb{E}\left[f_t'(\mathsf{R}_t^{\zeta};\mathsf{B}_*,\Theta)\right]\right| \cdot \left|\mathbb{E}\left[\mathsf{R}_s^{\zeta}\mathsf{R}_t^{\zeta}\right]\right| \\
&\quad + \left|\mathbb{E}\left[f_s'(\mathsf{R}_s^{\zeta};\mathsf{B}_*,\Theta)\right]\right| \cdot \left|\mathbb{E}\left[f_t(\mathsf{R}_t^{\zeta};\mathsf{B}_*,\Theta)\mathsf{R}_s^{\zeta}\right]\right| + \left|\mathbb{E}\left[f_t'(\mathsf{R}_t^{\zeta};\mathsf{B}_*,\Theta)\right]\right| \cdot \left|\mathbb{E}\left[f_s(\mathsf{R}_s^{\zeta};\mathsf{B}_*,\Theta)\mathsf{R}_t^{\zeta}\right]\right|.
\end{aligned}
\tag{G.141}
$$

根据 f_t （来自假设（A10））的利普希茨性质，

$$
\begin{aligned}
&\left|\mathbb{E}\left[f_t(b_t^{\zeta}\mathsf{B}_* + \mathsf{K}_t^{\zeta};\mathsf{B}_*,\Theta) - f_t(b_t\mathsf{B}_* + \mathsf{K}_t;\mathsf{B}_*,\Theta)\right]\right| \\
&\le \mathbb{E}\left[\left|f_t\left(b_t^{\zeta}\mathsf{B}_* + \mathbb{E}\left[(\mathsf{K}_t^{\zeta})^2\right]^{1/2}\mathsf{G};\mathsf{B}_*,\Theta\right) - f_t\left(b_t\mathsf{B}_* + \mathbb{E}\left[\mathsf{K}_t^2\right]^{1/2}\mathsf{G};\mathsf{B}_*,\Theta\right)\right|\right] \\
&\le C\mathbb{E}\left[\left|(b_t^{\zeta} - b_t)\mathsf{B}_* + \left(\mathbb{E}\left[(\mathsf{K}_t^{\zeta})^2\right]^{1/2} - \mathbb{E}\left[\mathsf{K}_t^2\right]^{1/2}\right)\mathsf{G}\right|\right] \\
&\le C\left(\left|b_t^{\zeta} - b_t\right|\rho^{1/2} + \left|\mathbb{E}\left[(\mathsf{K}_t^{\zeta})^2\right] - \mathbb{E}\left[\mathsf{K}_t^2\right]\right|^{1/2}\right) \xrightarrow[\zeta\to 0]{} 0,
\end{aligned}
\tag{G.142}
$$

其中$\mathsf{G} \sim \mathcal{N}(0,\ 1)$ 独立于B_*，且最后一行由归纳假设（G.129）与（G.130）得出。结合（G.142）以及假设（G.129）至（G.131），我们有（G.141）收敛到0，当 $\zeta \to 0$时。将上述所有观察用于（G.140），我们得到

$$
\mathbb{E}\left[\mathsf{K}_{s+1}^{\zeta}\mathsf{K}_{t+1}^{\zeta}\right] \to \mathbb{E}[\mathsf{K}_{s+1}\mathsf{K}_{t+1}].
\tag{G.143}
$$

类似的论证表明 $\mathbb{E}\left[\mathsf{J}_{s+1}^{\zeta}\mathsf{J}_{t+1}^{\zeta}\right] \to \mathbb{E}[\mathsf{J}_{s+1}\mathsf{J}_{t+1}]$ 对于 $\zeta \to 0$ 每个 $0 \le s \le t$。现在 (G.130) 已对 $t+1$得证。

接下来，根据 (G.139) 和 (G.143) （上文已证明，其中 $s=t$），$\mathbb{E}\Big[f'_{t+1}(\mathsf{R}^{\zeta}_{t+1};\mathsf{B}_*,\Theta)\Big]-\mathbb{E}\big[f'_{t+1}(\mathsf{R}_{t+1};\mathsf{B}_*,\Theta)\big]=\mathbb{E}\Big[f'_{t+1}(b^{\zeta}_{t+1}\mathsf{B}_*+\mathsf{K}^{\zeta}_{t+1};\mathsf{B}_*,\Theta)\Big]-\mathbb{E}\big[f'_{t+1}(b_{t+1}\mathsf{B}_*+\mathsf{K}_{t+1};\mathsf{B}_*,\Theta)\big]\to 0$，因为 $\zeta\to 0$。由于根据假设 (A8)，$\mathbb{E}\big[f'_{t+1}(\mathsf{R}_{t+1};\mathsf{B}_*,\Theta)\big]=0$，我们有

$$\lim_{\zeta\to 0}\mathbb{E}\Big[f'_{t+1}(\mathsf{R}^{\zeta}_{t+1};\mathsf{B}_*,\Theta)\Big]=0. \tag{G.144}$$

类似的论证使我们能够证明 (G.131) 中的另一个极限，其中 $t+1$。

对于 $\widetilde{r}^{t+1}$，根据假设（A10）中 f_{t+1} 的 Lipschitz连续性质，

$$\sum_{i=1}^{d}\big(f_{t+1}(r_i^{t+1};\beta_{*,i},\Theta_{i,:})-f_{t+1}(0;\beta_{*,i},\Theta_{i,:})\big)^2\le C^2\|r^{t+1}\|_2^2.$$

结合（G.137）和（G.138）如上所示，我们得到

$$\limsup_{d\to\infty}\frac{1}{\sqrt{d}}\|\widetilde{r}^{t+1}\|_2=\limsup_{d\to\infty}\frac{1}{\sqrt{d}}\Big(\|f_{t+1}(r^{t+1};\beta_*,\Theta)\|_2+\zeta\|s^{t+1}\|_2+\Big|\mathbb{E}\Big[f'_{t+1}(\mathsf{R}^{\zeta}_{t+1};\mathsf{B}_*,\Theta)\Big]\Big|\Big\|r^{t+1}_{\zeta}\Big\|_2\Big)<\infty. \tag{G.145}$$

同样的论证推出上极限$_{n\to\infty}\, n^{-1}\|\widetilde{p}^{t+1}\|_2^2<\infty$，从而证明了（G.132）。回顾

扰动去噪器（G.120）的设计，并应用三角不等式，

$$\begin{aligned}\Big\|\widetilde{r}^{t+1}-\widetilde{r}^{t+1}_{\zeta}\Big\|_2&\le\Big\|f_{t+1}(r^{t+1};\beta_*,\Theta)-f_{t+1}(r^{t+1}_{\zeta};\beta_*,\Theta)\Big\|_2+\zeta\|s^{t+1}\|_2+\Big|\mathbb{E}\Big[f'_{t+1}(\mathsf{R}^{\zeta}_{t+1};\mathsf{B}_*;\Theta)\Big]\Big|\Big\|r^{t+1}_{\zeta}\Big\|_2\\&\le C\Big\|r^{t+1}-r^{t+1}_{\zeta}\Big\|_2+\zeta\|s^{t+1}\|_2+\Big|\mathbb{E}\Big[f'_{t+1}(\mathsf{R}^{\zeta}_{t+1};\mathsf{B}_*;\Theta)\Big]\Big|\Big(\|r^{t+1}\|_2+\Big\|r^{t+1}-r^{t+1}_{\zeta}\Big\|_2\Big),\end{aligned}$$

其中最后一行依据 f_{t+1} 在假设（A10）中的 Lipschitz连续（G.137）、（G.138）和（G.144）如上所示，$d^{-1}\|s^{t+1}\|_2^2\to 1$我们得到

$$\lim_{\zeta\to 0}\limsup_{d\to\infty}d^{-1}\Big\|\widetilde{r}^{t+1}-\widetilde{r}^{t+1}_{\zeta}\Big\|_2^2=0. \tag{G.146}$$

类似的论证推出极限$_{\zeta\to 0}$ 上极限$_{n\to\infty}\, n^{-1}\Big\|\widetilde{p}^{t+1}-\widetilde{p}^{t+1}_{\zeta}\Big\|_2^2=0$，从而验证了（G.133）对 $t+1$。根据（4.29e）的定义，

$$\widetilde{b}^{\zeta}_{t+1}=\rho^{-1}\mathbb{E}\Big[\widetilde{\mathsf{R}}^{\zeta}_{t+1}\mathsf{B}_*\Big]=\rho^{-1}\mathbb{E}\Big[f_{t+1}(b^{\zeta}_{t+1}\mathsf{B}_*+\mathsf{K}^{\zeta}_{t+1};\mathsf{B}_*,\Theta)\mathsf{B}_*\Big]-\rho^{-1}\mathbb{E}\Big[f'_{t+1}(\mathsf{R}^{\zeta}_{t+1};\mathsf{B}_*,\Theta)\Big]\mathbb{E}\Big[\mathsf{R}^{\zeta}_{t+1}\mathsf{B}_*\Big]\to\widetilde{b}_{t+1}, \tag{G.147}$$

其中收敛由(G.139)、(G.143) 以及(G.144)得到。类似地，$\widetilde{a}^{\zeta}_{t+1}\to\widetilde{a}_{t+1}$，从而证明(G.134) 对 $t+1$成立。

接下来，考虑(G.135) 对于 $t+1$的情形。记 $R^{t+1}\in\mathbb{R}^{d\times(2(t+2)+d')}$ 与$\mathbf{R}_{t+1}\in\mathbb{R}^{1\times(2(t+2)+d')}$ 分别为(G.135) 在 $t+1$下的左端与右端。并将其扰动版本记为

$$\begin{aligned}R^{t+1}_{\zeta}&:=\begin{bmatrix}r^1_{\zeta}&\cdots&r^{t+1}_{\zeta}&\widetilde{r}^0_{\zeta}&\cdots&\widetilde{r}^{t+1}_{\zeta}&\beta_*&\Theta\end{bmatrix}\in\mathbb{R}^{d\times(2(t+2)+d')},\\\mathbf{R}^{\zeta}_{t+1}&:=\begin{bmatrix}\mathsf{R}^{\zeta}_1&\cdots&\mathsf{R}^{\zeta}_{t+1}&\widetilde{\mathsf{R}}^{\zeta}_0&\cdots&\widetilde{\mathsf{R}}^{\zeta}_{t+1}&\mathsf{B}_*&\Theta\end{bmatrix}\in\mathbb{R}^{1\times(2(t+2)+d')}.\end{aligned}$$

对于任意 $h\in\mathrm{PL}^2_{(2(t+2)+d')\to 1}$，我们应用三角不等式来给出上界

$$\limsup_{d\to\infty}\Big|\frac{1}{d}\sum_{i=1}^{d}h(R^{t+1}_{i,:})-h(\mathbf{R}_{t+1})\Big|\le\lim_{\zeta\to 0}\limsup_{d\to\infty}\Big|\frac{1}{d}\sum_{i=1}^{d}\Big[h(R^{t+1}_{i,:})-h((R^{t+1}_{\zeta})_{i,:})\Big]\Big| \tag{G.148}$$

$$+\limsup_{d\to\infty}\left|\frac{1}{d}\sum_{i=1}^{d}h((R_\zeta^{t+1})_{i,:})-\mathbb{E}\Big[h(\mathbf{R}_{t+1}^{\zeta})\Big]\right| \tag{G.149}$$

$$+\lim_{\zeta\to 0}\Big|\mathbb{E}\Big[h(\mathbf{R}_{t+1}^{\zeta})\Big]-\mathbb{E}[h(\mathbf{R}_{t+1})]\Big|. \tag{G.150}$$

利用 h 的伪 Lipschitz 性质、三角不等式和柯西–施瓦茨不等式，我们对(G.148) 中的第一项给出上界

$$\left|\frac{1}{d}\sum_{i=1}^{d}\Big[h(R_{i,:}^{t+1})-h((R_\zeta^{t+1})_{i,:})\Big]\right| \le \frac{C}{d}\sum_{i=1}^{d}\Big(1+\|R_{i,:}^{t+1}\|_2+\Big\|(R_\zeta^{t+1})_{i,:}\Big\|_2\Big)\Big\|R_{i,:}^{t+1}-(R_\zeta^{t+1})_{i,:}\Big\|_2$$

$$\le C\sqrt{\frac{3}{d}\sum_{i=1}^{d}\Big(1+4\|R_{i,:}^{t+1}\|_2^2+\Big\|R_{i,:}^{t+1}-(R_\zeta^{t+1})_{i,:}\Big\|_2^2\Big)}\sqrt{\frac{1}{d}\sum_{i=1}^{d}\Big\|R_{i,:}^{t+1}-(R_\zeta^{t+1})_{i,:}\Big\|_2^2},$$

按照该顺序令 $d\to\infty,\zeta\to 0$ 趋于零；利用归纳假设 (G.127)、(G.128)、(G.132) 和 (G.133)，以及上文针对 $t+1$ 给出的对应结果 (G.137)、(G.138)、(G.145) 和 (G.146)，即可得到这一结论。由定理 G.3 中的 Wasserstein 收敛结果可知，对任意固定的 $\zeta>0$，(G.149) 中的第二项在将 R_ζ^{t+1} 代入后等于零。(G.150) 中的第三项为零，这是因为归纳假设 (G.129) 至 (G.131) 和 (G.134)，以及针对 $t+1$的对应结果 (G.139)、(G.143)、(G.144) 和 (G.147)，保证了扰动状态演化参数收敛到原始参数。综合 (G.148) 至 (G.150)，可得 $t+1$满足 (G.135)。通过类似的论证可得 $t+1$ 满足收敛结果 (G.136)，此处省略细节。至此，归纳过程完成，定理 G.2 得证。

G.3 定理4.3 的证明

固定一个较大的常数 $\Gamma\ge 1$。考虑辅助 GVAMP 迭代:

$$\underline{r}^{t+1}=\underline{\Phi}_t(X^\top X)\widetilde{\underline{r}}^t+\widetilde{\underline{\Phi}}_t(X)^\top\widetilde{\underline{p}}^t,\qquad \widetilde{\underline{r}}^{t+1}=\underline{f}_{t+1}(\underline{r}^{t+1}), \tag{G.151a}$$

$$\underline{p}^{t+1}=\underline{\Psi}_t(XX^\top)\widetilde{\underline{p}}^t+\widetilde{\underline{\Psi}}_t(X)\widetilde{\underline{r}}^t,\qquad \widetilde{\underline{p}}^{t+1}=\underline{g}_{t+1}(\underline{p}^{t+1};y). \tag{G.151b}$$

对于 $0\le t\le\Gamma-1$，我们选择

$$\underline{\Phi}_t(x)=0,\quad \widetilde{\underline{\Phi}}_t(x)=\overline{\kappa}_2^{-1}x,\quad \underline{\Psi}_t(x)=\overline{\kappa}_2^{-1}x-1,\quad \widetilde{\underline{\Psi}}_t(x)=0,\quad \underline{f}_t(\underline{r})=0,\quad \underline{g}_t(\underline{p};y)=\overline{g}(y)\underline{p}, \tag{G.151c}$$

使得

$$\begin{aligned}\underline{r}^{t+1}&=\overline{\kappa}_2^{-1}X^\top\widetilde{\underline{p}}^t,\\ \underline{p}^{t+1}&=(\overline{\kappa}_2^{-1}XX^\top-I_n)\widetilde{\underline{p}}^t,\qquad \widetilde{\underline{p}}^{t+1}=\overline{G}\underline{p}^{t+1},\end{aligned} \tag{G.151d}$$

其中 $\overline{G}$=diag($g(\overline{y})$)。对于 $t\ge\Gamma$，我们选择

$$\underline{\Phi}_t=\Phi_{t-\Gamma},\qquad \widetilde{\underline{\Phi}}_t=\widetilde{\Phi}_{t-\Gamma},\qquad \underline{\Psi}_t=\Psi_{t-\Gamma},\qquad \widetilde{\underline{\Psi}}_t=\widetilde{\Psi}_{t-\Gamma}, \tag{G.151e}$$

and

$$\underline{f}_t(\underline{r})=\begin{cases}f_0(c_r\,\underline{r}), & t=\Gamma\\ f_{t-\Gamma}(\underline{r}), & t>\Gamma\end{cases},\qquad \underline{g}_t(\underline{p};y)=\begin{cases}g_0(c_p\,\underline{p};y), & t=\Gamma\\ g_{t-\Gamma}(\underline{p};y), & t>\Gamma\end{cases}, \tag{G.151f}$$

其中常数 c_r，c_p 定义见 (4.21)。我们用以下方式初始化 (G.151d)：

$$\widetilde{\underline{p}}^0=\gamma^{-\Gamma}w_3\,\overline{G}\Big(\frac{1}{\sqrt{\rho+w_2}}z+\underline{j}^0\Big), \tag{G.151g}$$

其中 $j^0 \sim \mathcal{N}\left(0_n, \frac{w_1}{\rho+w_2} I_n\right)$ 与其他所有量相互独立。注意， $\widetilde{r}^0$ 无需用于初始化 (G.151d)。下面的定理 G.4 给出了辅助 GVAMP 迭代的状态演化结果，它是一般性结果定理 G.2的直接推论。

回忆一下Z，Y 来自(3.10)。令

$$\widetilde{\mathsf{P}}_0 = \gamma^{-\Gamma} w_3 \, \overline{g}(\mathsf{Y}) \left(\frac{1}{\sqrt{\rho + w_2}} \mathsf{Z} + \underline{\mathsf{J}}_0 \right), \tag{G.152a}$$

其中 $\mathsf{J}_0 \sim \mathcal{N}\left(0, \frac{w_1}{\rho+w_2}\right)$ 与 Z、Y 独立。令 $(\mathsf{J}_{t+1})_{t\geq 0}$、$(\mathsf{K}_{t+1})_{t\geq 0}$ 为两个彼此独立，且与 $(\mathsf{B}_*$、Z、E $\Lambda)$ 独立的中心化高斯过程，其协方差结构如下所示。

1. 对于 $0 \leq r$、$s \leq \Gamma - 1$ 和 $1 \leq t \leq \Gamma$,

$$\underline{a}_t = \gamma^{-(\Gamma-t)} \frac{w_3}{\sqrt{\rho + w_2}}, \qquad \mathbb{E}\left[\underline{\mathsf{J}}_t^2\right] = \gamma^{-2(\Gamma-t)} \frac{w_3^2 w_1}{\rho + w_2}, \tag{G.152b}$$

$$\underline{b}_t = \gamma^{-(\Gamma-t)} \frac{1}{\sqrt{\rho + w_2}}, \qquad \mathbb{E}\left[\underline{\mathsf{K}}_t^2\right] = \gamma^{-2(\Gamma-t)} \frac{w_2}{\rho + w_2}, \tag{G.152c}$$

$$\mathbb{E}\left[\underline{\mathsf{K}}_{r+1}\underline{\mathsf{K}}_{s+1}\right] = \frac{\delta}{\overline{\kappa}_2}\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\right]\mathbb{E}\left[\underline{\mathsf{J}}_r\underline{\mathsf{J}}_s\right] + \delta\rho\left\{\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\right] + \frac{\overline{\kappa}_4}{\overline{\kappa}_2^2}\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2\right\}\underline{a}_r\underline{a}_s, \tag{G.152d}$$

$$\begin{aligned}\mathbb{E}\left[\underline{\mathsf{J}}_{r+1}\underline{\mathsf{J}}_{s+1}\right] &= \left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2^2} + \delta\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\right]\mathbb{E}\left[\underline{\mathsf{J}}_r\underline{\mathsf{J}}_s\right] \\ &\quad + \rho\left\{\left(\frac{\overline{\kappa}_4}{\overline{\kappa}_2} + \delta\overline{\kappa}_2\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})^2\overline{\mathsf{Z}}^2\right] + \left(\frac{\overline{\kappa}_6}{\overline{\kappa}_2^2} - \frac{\overline{\kappa}_4^2}{\overline{\kappa}_2^3} + \delta\frac{\overline{\kappa}_4}{\overline{\kappa}_2}\right)\mathbb{E}\left[\overline{g}(\mathsf{Y})\overline{\mathsf{Z}}^2\right]^2\right\}\underline{a}_r\underline{a}_s.\end{aligned} \tag{G.152e}$$

我们还对每个 $1 \leq t \leq \Gamma$ 定义 $\mathsf{R}_t = b_t\mathsf{B}_* + \mathsf{K}_t$、$\mathsf{P}_t = a_t\mathsf{Z} + \mathsf{J}_t$，并对每个 $1 \leq t \leq \Gamma - 1$ 定义 $\widetilde{\mathsf{P}}_t = g(\mathsf{Y})\mathsf{P}_t, \widetilde{a}_t = \mathbb{E}\left[g(\mathsf{Y})\mathsf{Z}^2\right]a_t$。注意，由于第一阶段中 (G.151c) 的去噪器选择，我们不需要对 $0 \leq t \leq \Gamma - 1$ 定义 R_t, b_t。

2. 对于 $r, s \geq \Gamma$，相关性 $\mathbb{E}\left[\mathsf{J}_{r+1}\mathsf{J}_{s+1}\right], \mathbb{E}\left[\mathsf{K}_{r+1}\mathsf{K}_{s+1}\right]$ 以及系数 $\widetilde{a}_t, \widetilde{b}_t, a_{t+1}, b_{t+1}$ 和随机变量 $\mathsf{R}_{t+1}, \mathsf{P}_{t+1}, \mathsf{R}_t, \mathsf{P}_t$（对于 $t \geq \Gamma$），形式上遵循相同的递归 (4.29)，其中使用了 (G.151e) 和 (G.151f) 中的去噪器。

3. 对于 $0 \leq r \leq \Gamma - 1 < s$,

$$\begin{aligned}\mathbb{E}\left[\underline{\mathsf{J}}_{r+1}\underline{\mathsf{J}}_{s+1}\right] &= \overline{\kappa}_2^{-1}\mathbb{E}\left[\underline{\Psi}_s(\Lambda_n^2)\Lambda_n^2\right]\mathbb{E}\left[\widetilde{\underline{\mathsf{P}}}_r\widetilde{\underline{\mathsf{P}}}_s\right] - 2\mathbb{E}\left[\underline{\Psi}_s(\Lambda_n^2)\Lambda_n^2\right]\widetilde{\underline{a}}_r\widetilde{\underline{a}}_s\rho + \overline{\kappa}_2^{-1}\mathbb{E}\left[\underline{\Psi}_s(\Lambda_n^2)\Lambda_n^4\right]\widetilde{\underline{a}}_r\widetilde{\underline{a}}_s\rho \\ &\quad + \mathbb{E}\left[(\overline{\kappa}_2^{-1}\Lambda_n^2 - 1)\widetilde{\underline{\Psi}}_s(\Lambda_n)\Lambda_n\right]\widetilde{\underline{a}}_r\widetilde{\underline{b}}_s\rho - \underline{a}_{r+1}\underline{a}_{s+1}\sigma^2,\end{aligned} \tag{G.152f}$$

$$\begin{aligned}\mathbb{E}\left[\underline{\mathsf{K}}_{r+1}\underline{\mathsf{K}}_{s+1}\right] &= \overline{\kappa}_2^{-1}\mathbb{E}\left[\widetilde{\underline{\Phi}}_s(\Lambda_d)\Lambda_d\right]\mathbb{E}\left[\widetilde{\underline{\mathsf{P}}}_r\widetilde{\underline{\mathsf{P}}}_s\right] + \mathbb{E}\left[(\overline{\kappa}_2^{-1} - 1)\widetilde{\underline{\Phi}}_s(\Lambda_d)\Lambda_d\right]\widetilde{\underline{a}}_r\widetilde{\underline{a}}_s\rho \\ &\quad + \overline{\kappa}_2^{-1}\mathbb{E}\left[\underline{\Phi}_s(\Lambda_d^2)\Lambda_d^2\right]\widetilde{\underline{a}}_r\widetilde{\underline{b}}_s\rho - \underline{b}_{r+1}\underline{b}_{s+1}\rho.\end{aligned} \tag{G.152g}$$

Lemma G.4. 考虑辅助 *GVAMP* 迭代 (G.151)。那么对于任意固定的 $t \geq 0$,

$$\begin{bmatrix}\underline{r}^1 & \cdots & \underline{r}^{t+1} & \widetilde{\underline{r}}^\Gamma & \cdots & \widetilde{\underline{r}}^{t+1} & \beta_*\end{bmatrix} \xrightarrow{W_2} \begin{bmatrix}\underline{\mathsf{R}}_1 & \cdots & \underline{\mathsf{R}}_{t+1} & \widetilde{\underline{\mathsf{R}}}_\Gamma & \cdots & \widetilde{\underline{\mathsf{R}}}_{t+1} & \mathsf{B}_*\end{bmatrix},$$

$$\begin{bmatrix}\underline{p}^1 & \cdots & \underline{p}^{t+1} & \widetilde{\underline{p}}^0 & \cdots & \widetilde{\underline{p}}^{t+1} & z & \varepsilon\end{bmatrix} \xrightarrow{W_2} \begin{bmatrix}\underline{\mathsf{P}}_1 & \cdots & \underline{\mathsf{P}}_{t+1} & \widetilde{\underline{\mathsf{P}}}_0 & \cdots & \widetilde{\underline{\mathsf{P}}}_{t+1} & \mathsf{Z} & \mathsf{E}\end{bmatrix},$$

其中右侧的随机变量定义于(G.152)。

证明。该引理直接由一般结果定理 G.2得出。具体而言，将状态演化递推(4.29) 特化到第一阶段(G.151c) 并使用附录 A中的矩-累积量关系，可得(G.152b) 至(G.152e)，它们与通过约化到 [101, 定理 1]得到的(C.28)、(C.30)、(C.31) 和(C.38) 一致。我们省略这些繁琐的计算，仅澄清一点：对于 $1 \le t \le \Gamma$，记号

$$\underline{p}^t, \qquad \widetilde{\underline{p}}^t, \qquad \underline{a}_t, \qquad \underline{\mathsf{J}}_t, \qquad \underline{\mathsf{P}}_t, \qquad \underline{r}^t, \qquad \underline{b}_t, \qquad \underline{\mathsf{K}}_t, \qquad \underline{\mathsf{R}}_t$$

在定理 G.4中对应于记号

$$u^t, \qquad g^t, \qquad \nu_t, \qquad \tau_t \mathsf{N}_t, \qquad \mathsf{Z}_t, \qquad v^t, \qquad \chi_t, \qquad \omega_t \mathsf{M}_t, \qquad \mathsf{B}_t$$

在定理 C.1中。最后，我们说明如何得到表达式(G.152b) 和(G.152c)。注意，初始条件(B.2) 与(G.151g)之间唯一的区别在于额外的 $\gamma^{-\Gamma} w_3$ 因子出现在(G.151g)中。检查定理 B.1的证明可知，这不会改变以下状态演化参数性质的有效性：对任意 $2 \le t \le \Gamma$,

$$\frac{\underline{a}_t}{\underline{a}_{t-1}} = \frac{\underline{b}_t}{\underline{b}_{t-1}} = \frac{\mathbb{E}\big[\underline{\mathsf{J}}_t^2\big]^{1/2}}{\mathbb{E}\big[\underline{\mathsf{J}}_{t-1}^2\big]^{1/2}} = \frac{\mathbb{E}\big[\underline{\mathsf{K}}_t^2\big]^{1/2}}{\mathbb{E}\big[\underline{\mathsf{K}}_{t-1}^2\big]^{1/2}} = \gamma,$$

且对任意 $1 \le t \le \Gamma$,

$$\frac{\underline{a}_t}{\underline{b}_t} = w_3, \qquad\qquad \frac{\mathbb{E}\big[\underline{\mathsf{K}}_t^2\big]^{1/2}}{\underline{b}_t} = \sqrt{w_2}. \tag{G.153}$$

根据 $1 \le t \le \Gamma$ 作归纳法，可得

$$\Big(\underline{a}_t, \underline{b}_t, \mathbb{E}\big[\underline{\mathsf{J}}_t^2\big]^{1/2}, \mathbb{E}\big[\underline{\mathsf{K}}_t^2\big]^{1/2}\Big) = \gamma^t \Big(\underline{a}_0, \underline{b}_0, \mathbb{E}\big[\underline{\mathsf{J}}_0^2\big]^{1/2}, \mathbb{E}\big[\underline{\mathsf{K}}_0^2\big]^{1/2}\Big). \tag{G.154}$$

根据初始条件(G.152a),

$$\underline{a}_0 = \gamma^{-\Gamma} w_3 \frac{1}{\sqrt{\rho + w_2}}, \qquad\qquad \mathbb{E}\big[\underline{\mathsf{J}}_0^2\big]^{1/2} = \gamma^{-\Gamma} w_3 \sqrt{\frac{w_1}{\rho + w_2}}.$$

将其代入(G.154) 并使用(G.153)，我们得出(G.152b) 和(G.152c) 成立。 □

定理 4.3 可由下方引理中的 (G.159) 和 (G.160) 与 定理 G.4结合后立即得到。

引理 G.5. 对于每个 $\tau \ge 0$,

$$\lim_{\Gamma\to\infty} \Big| \underline{a}_\Gamma - \frac{a_0}{c_p} \Big| = \lim_{\Gamma\to\infty} \big| \underline{a}_{\Gamma+\tau+1} - a_{\tau+1} \big| = 0, \quad \lim_{\Gamma\to\infty} \Big| \underline{b}_\Gamma - \frac{b_0}{c_r} \Big| = \lim_{\Gamma\to\infty} \big| \underline{b}_{\Gamma+\tau+1} - a_{\tau+1} \big| = 0, \tag{G.155}$$

$$\lim_{\Gamma\to\infty} \big| \widetilde{\underline{a}}_{\Gamma+\tau} - \widetilde{a}_\tau \big| = 0, \qquad\qquad \lim_{\Gamma\to\infty} \Big| \widetilde{\underline{b}}_{\Gamma+\tau} - \widetilde{b}_\tau \Big| = 0. \tag{G.156}$$

对于每个 ς、$\tau \ge 1$,

$$\lim_{\Gamma\to\infty} \Big| \mathbb{E}\big[\underline{\mathsf{J}}_\Gamma^2\big] - \frac{1}{c_p^2} \mathbb{E}\big[\mathsf{J}_0^2\big] \Big| = \lim_{\Gamma\to\infty} \Big| \mathbb{E}\big[\underline{\mathsf{J}}_\Gamma \underline{\mathsf{J}}_{\Gamma+\tau}\big] - \frac{1}{c_p} \mathbb{E}[\mathsf{J}_0 \mathsf{J}_\tau] \Big| = \lim_{\Gamma\to\infty} \big| \mathbb{E}\big[\underline{\mathsf{J}}_{\Gamma+\varsigma} \underline{\mathsf{J}}_{\Gamma+\tau}\big] - \mathbb{E}[\mathsf{J}_\varsigma \mathsf{J}_\tau] \big| = 0, \tag{G.157}$$

$$\lim_{\Gamma\to\infty} \Big| \mathbb{E}\big[\underline{\mathsf{K}}_\Gamma^2\big] - \frac{1}{c_r^2} \mathbb{E}\big[\mathsf{K}_0^2\big] \Big| = \lim_{\Gamma\to\infty} \Big| \mathbb{E}\big[\underline{\mathsf{K}}_\Gamma \underline{\mathsf{K}}_{\Gamma+\tau}\big] - \frac{1}{c_r} \mathbb{E}[\mathsf{K}_0 \mathsf{K}_\tau] \Big| = \lim_{\Gamma\to\infty} \big| \mathbb{E}\big[\underline{\mathsf{K}}_{\Gamma+\varsigma} \underline{\mathsf{K}}_{\Gamma+\tau}\big] - \mathbb{E}[\mathsf{K}_\varsigma \mathsf{K}_\tau] \big| = 0. \tag{G.158}$$

对于任意固定的 $\tau \geq 0$ 和任意 $\phi_r \in \mathrm{PL}^2_{2(\tau+1)+1\to 1}$、$\phi_p \in \mathrm{PL}^2_{2(\tau+2)\to 1}$，几乎必然

$$\lim_{\Gamma\to\infty}\lim_{d\to\infty}\left|\frac{1}{d}\sum_{i=1}^{d}\Big(\phi_r(c_r\underline{r}_i^{\Gamma},\cdots,\underline{r}_i^{\Gamma+\tau},\widetilde{\underline{r}}_i^{\Gamma},\cdots,\widetilde{\underline{r}}_i^{\Gamma+\tau},\beta_{*,i})-\phi_r(r_i^0,\cdots,r_i^{\tau},\widetilde{r}_i^0,\cdots,\widetilde{r}_i^{\tau},\beta_{*,i})\Big)\right|=0, \quad \text{(G.159)}$$

$$\lim_{\Gamma\to\infty}\lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n}\Big(\phi_p(c_p\underline{p}_i^{\Gamma},\cdots,\underline{p}_i^{\Gamma+\tau},\widetilde{\underline{p}}_i^{\Gamma},\cdots,\widetilde{\underline{p}}_i^{\Gamma+\tau},z_i,\varepsilon_i)-\phi_p(p_i^0,\cdots,p_i^{\tau},\widetilde{p}_i^0,\cdots,\widetilde{p}_i^{\tau},z_i,\varepsilon_i)\Big)\right|=0. \quad \text{(G.160)}$$

证明。 通过对时间进行归纳法证明。

基例。 我们将辅助 GVAMP 迭代 Γ 在时间 (G.151) 作为基例。由 (G.152b) 以及 (4.21) 中对 c_p的定义，

$$c_p\,\underline{a}_{\Gamma}=\frac{\sqrt{\rho+w_2}}{w_1w_3}\frac{w_3}{\sqrt{\rho+w_2}}=\frac{1}{w_1}\overset{(4.25)}{=}a_0. \quad \text{(G.161)}$$

类似地可以证明

$$c_r\,\underline{b}_{\Gamma}=b_0. \quad \text{(G.162)}$$

这就证明了(G.155)中的第一个等式。再由 (G.152b)，

$$c_p^2\mathbb{E}\big[\mathsf{J}_{\Gamma}^2\big]=\frac{\rho+w_2}{(w_1w_3)^2}\frac{w_3^2w_1}{\rho+w_2}=\frac{1}{w_1}\overset{(4.25)}{=}a_0. \quad \text{(G.163)}$$

类似的计算可得

$$c_r^2\mathbb{E}\big[\mathsf{K}_{\Gamma}^2\big]=b_0, \quad \text{(G.164)}$$

从而确立了(G.157)和 (G.158)中的第一个等式。根据 g_Γ 在 (G.151f) 中的设计，以及 (G.161) 和 (G.163) 如上所示，

$$\widetilde{\underline{a}}_{\Gamma}=\sigma^{-2}\mathbb{E}\Big[\underline{g}_{\Gamma}\Big(c_p\Big(\underline{a}_{\Gamma}\mathsf{Z}+\mathbb{E}\big[\mathsf{J}_{\Gamma}^2\big]^{1/2}\mathsf{G}\Big);\mathsf{Y}\Big)\mathsf{Z}\Big]=\sigma^{-2}\mathbb{E}[g_0(a_0\mathsf{Z}+\sqrt{a_0}\mathsf{G};\mathsf{Y})\mathsf{Z}]\overset{(4.28)}{=}\widetilde{a}_0,$$

其中 $\mathsf{G}\sim\mathcal{N}(0,\,1)$ 与 Z，Y独立。类似地，$b_\Gamma=b_0$。因此，(G.156) 对 $\tau=0$成立。

令 $\phi_r\in\mathrm{PL}^2_{3\to 1}$。于是我们有

$$\begin{aligned}
&\left|\frac{1}{d}\sum_{i=1}^{d}\Big(\phi_r(c_r\underline{r}_i^{\Gamma},\widetilde{\underline{r}}_i^{\Gamma},\beta_{*,i})-\phi_r(r_i^0,\widetilde{r}_i^0,\beta_{*,i})\Big)\right|\\
&\leq\frac{C}{d}\sum_{i=1}^{d}\Big(1+\big\|[c_r\underline{r}_i^{\Gamma}\quad\widetilde{\underline{r}}_i^{\Gamma}\quad\beta_{*,i}]\big\|_2+\big\|[r_i^0\quad\widetilde{r}_i^0\quad\beta_{*,i}]\big\|_2\Big)\big\|[c_r\underline{r}_i^{\Gamma}\quad\widetilde{\underline{r}}_i^{\Gamma}]-[r_i^0\quad\widetilde{r}_i^0]\big\|_2\\
&\leq C\sqrt{3}\left[1+d^{-1}\left(\big\|c_r\underline{r}^{\Gamma}\big\|_2^2+\Big\|\widetilde{\underline{r}}^{\Gamma}\Big\|_2^2+\big\|r^0\big\|_2^2+\big\|\widetilde{r}^0\big\|_2^2+2\|\beta_*\|_2^2\right)\right]^{1/2}\left[d^{-1}\left(\big\|c_r\underline{r}^{\Gamma}-r^0\big\|_2^2+\Big\|\widetilde{\underline{r}}^{\Gamma}-\widetilde{r}^0\Big\|_2^2\right)\right]^{1/2},
\end{aligned} \quad \text{(G.165)}$$

其中最后一步根据柯西–施瓦茨不等式。检查定理 B.3的证明后，我们发现，尽管 $\gamma^{-\Gamma}w_3$ 初始化器中存在额外的乘法因子(G.151g)，以下结果仍然成立：

$$\lim_{\Gamma\to\infty}\lim_{d\to\infty}\frac{\big|\langle\underline{r}^{\Gamma},v_1(D)\rangle\big|}{\|\underline{r}^{\Gamma}\|_2}=1,$$

几乎必然成立。此外，由于对初始化器(G.151g)进行了重新缩放，对于任意固定的 Γ，几乎必然有：

$$\lim_{d\to\infty} d^{-1}\|\underline{r}^{\Gamma}\|_2^2 = \underline{b}_{\Gamma}^2\rho + \mathbb{E}[\underline{\mathsf{K}}_{\Gamma}^2] \overset{(G.152c)}{=} \frac{1}{\rho+w_2}\rho + \frac{w_2}{\rho+w_2} = 1. \tag{G.166}$$

结合以上两个结果，我们有几乎必然

$$\lim_{\Gamma\to\infty}\lim_{d\to\infty} d^{-1}\|\underline{r}^{\Gamma} - v_1(D)\|_2^2 = 0. \tag{G.167}$$

鉴于原始 GVAMP 中初始化器的选择 （4.20），这等价于

$$\lim_{\Gamma\to\infty}\lim_{d\to\infty} d^{-1}\|c_r\,\underline{r}^{\Gamma} - r^0\|_2^2 = 0, \qquad \text{a.s.} \tag{G.168}$$

根据 f_{Γ} 在 (G.151f) 中的选择，以及由 假设（A10）所保证的其 Lipschitz连续性质，(G.168) 随即推出

$$d^{-1/2}\left\|\widetilde{\underline{r}}^{\Gamma} - \widetilde{r}^0\right\|_2 = d^{-1/2}\|f_0(c_r\,\underline{r}^{\Gamma}) - f_0(r^0)\|_2 \le Cd^{-1/2}\|c_r\,\underline{r}^{\Gamma} - r^0\|_2 \to 0, \tag{G.169}$$

在极限 $d\to\infty$ 之后接着 $\Gamma\to\infty$时几乎必然成立。 定理 G.4 中的状态演化结果表明，对于任意固定的 Γ，几乎必然有：

$$\lim_{d\to\infty} d^{-1}\left\|\widetilde{\underline{r}}^{\Gamma}\right\|_2^2 = \mathbb{E}\left[\underline{f}_{\Gamma}\left(c_r\left(\underline{b}_{\Gamma}\mathsf{B}_* + \mathbb{E}[\underline{\mathsf{K}}_{\Gamma}^2]^{1/2}\mathsf{G}\right)\right)^2\right] = \mathbb{E}\left[f_0\left(b_0\mathsf{B}_* + \sqrt{b_0}\mathsf{G}\right)^2\right] < \infty, \tag{G.170}$$

其中 $\mathsf{G}\sim\mathcal{N}(0,1)$ 与 B_* 独立，最后一个等号由 (G.162) 和 (G.164)得出。由于上面的右端不依赖于 Γ，因此有限性对 Γ一致成立。结合 (G.169)，我们有 $\lim_{d\to\infty} d^{-1}\|\widetilde{r}^0\|_2^2 < \infty$。结合 (G.166) 和 (G.168)，我们也有 $\lim_{d\to\infty} d^{-1}\|r^0\|_2^2 < \infty$。综合以上所有结果以及 假设（A1）在 (G.165) 中的条件可得 (G.159) 对于 $\tau = 0$成立。

令 $\phi_p \in \mathrm{PL}_{4\to 1}^2$。与 (G.165) 类似，我们有

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\Big(\phi_p(c_p\underline{p}_i^{\Gamma}, \underline{\widetilde{p}}_i^{\Gamma}, z_i, \varepsilon_i) - \phi_p(p_i^0, \widetilde{p}_i^0, z_i, \varepsilon_i)\Big)\right| \\
&\le C\sqrt{3}\left[1 + n^{-1}\left(\|c_p\underline{p}^{\Gamma}\|_2^2 + \left\|\underline{\widetilde{p}}^{\Gamma}\right\|_2^2 + \|p^0\|_2^2 + \|\widetilde{p}^0\|_2^2 + 2\|z\|_2^2 + 2\|\varepsilon\|_2^2\right)\right]^{1/2}\left[1 + n^{-1}\left(\|c_p\underline{p}^{\Gamma} - p^0\|_2^2 + \left\|\underline{\widetilde{p}}^{\Gamma} - \widetilde{p}^0\right\|_2^2\right)\right]^{1/2}.
\end{aligned} \tag{G.171}$$

我们逐一考察右侧的各项。首先考虑 $n^{-1}\|c_p p^{\Gamma} - p^0\|_2^2$。根据辅助 GVAMP 第一阶段中的更新规则 (G.151d)，我们有

$$\underline{p}^{\Gamma} = (\overline{\kappa}_2^{-1}XX^{\top} - I_n)\overline{G}\underline{p}^{\Gamma-1} = X\underline{r}^{\Gamma} - \overline{G}\underline{p}^{\Gamma-1},$$

或者等价地，

$$\underline{p}^{\Gamma} + \overline{G}\underline{p}^{\Gamma-1} = X\underline{r}^{\Gamma}. \tag{G.172}$$

我们声称

$$\lim_{\Gamma\to\infty}\lim_{n\to\infty} n^{-1}\|\underline{p}^{\Gamma} - \gamma\underline{p}^{\Gamma-1}\|_2^2 = 0. \tag{G.173}$$

为证明这一点，根据定理 G.4中的状态演化结果，几乎必然，

$$\lim_{n\to\infty} n^{-1}\|\underline{p}^{\Gamma} - \gamma\underline{p}^{\Gamma-1}\|_2^2 = \underline{a}_{\Gamma}^2\sigma^2 + \mathbb{E}[\underline{\mathsf{J}}_{\Gamma}^2] + \gamma^2\underline{a}_{\Gamma}^2\sigma^2 + \gamma^2\mathbb{E}[\underline{\mathsf{J}}_{\Gamma-1}^2] - 2\gamma(\underline{a}_{\Gamma}\underline{a}_{\Gamma-1}\sigma^2 + \mathbb{E}[\underline{\mathsf{J}}_{\Gamma}\underline{\mathsf{J}}_{\Gamma-1}])$$

$$\overset{(G.152b)}{=} 2\gamma^2\mathbb{E}\big[\underline{\mathsf{J}}_{\Gamma-1}^2\big] - 2\gamma\mathbb{E}\big[\underline{\mathsf{J}}_{\Gamma}\underline{\mathsf{J}}_{\Gamma-1}\big]. \tag{G.174}$$

在(E.19) 中已经证明

$$\lim_{\Gamma\to\infty}\frac{\mathbb{E}\big[\underline{\mathsf{J}}_{\Gamma}\underline{\mathsf{J}}_{\Gamma-1}\big]}{\mathbb{E}\big[\underline{\mathsf{J}}_{\Gamma}^2\big]^{1/2}\mathbb{E}\big[\underline{\mathsf{J}}_{\Gamma-1}^2\big]^{1/2}} \overset{(G.152b)}{=} \lim_{\Gamma\to\infty}\frac{\mathbb{E}\big[\underline{\mathsf{J}}_{\Gamma}\underline{\mathsf{J}}_{\Gamma-1}\big]}{\gamma\mathbb{E}\big[\underline{\mathsf{J}}_{\Gamma-1}^2\big]} = 1. \tag{G.175}$$

注意到 （E.19）是在初始化方案 （B.2）下得到的。不过，容易验证（G.175）在初始化方案 （G.151g）下仍然成立，该方案与 （B.2）相差一个乘法因子 $\gamma^{-\Gamma}w_3$。利用 （G.175）代入（G.174），即可得出 （G.173）。现在结合 （G.167）、（G.172）和（G.173）可得

$$\lim_{\Gamma\to\infty}\lim_{n\to\infty} n^{-1}\big\|(I_n+\gamma^{-1}\overline{G})\underline{p}^{\Gamma} - Xv_1(D)\big\|_2^2 = 0.$$

根据 p^0 在 （4.20）中的定义，可知

$$\lim_{\Gamma\to\infty}\lim_{n\to\infty} n^{-1}\big\|c_p\,\underline{p}^{\Gamma} - p^0\big\|_2^2 = 0. \tag{G.176}$$

根据定理 G.4中的状态演化结果，

$$\lim_{n\to\infty} n^{-1}\big\|\underline{p}^{\Gamma}\big\|_2^2 = \underline{a}_{\Gamma}^2\sigma^2 + \mathbb{E}\big[\underline{\mathsf{J}}_{\Gamma}^2\big] \overset{(G.152b)}{=} \frac{w_3^2(\sigma^2+w_1)}{\rho+w_2} < \infty. \tag{G.177}$$

此外，关于 Γ 的有界性是统一成立的，因此在 $\Gamma\to\infty$ 极限下也成立。（G.176）和（G.177）共同推出 $\lim_{n\to\infty} n^{-1}\big\|p^0\big\|_2^2 < \infty$。类似于推导 （G.169）和（G.170）时的论证，它们还推出

$$\lim_{\Gamma\to\infty}\lim_{n\to\infty} n^{-1}\Big\|\widetilde{\underline{p}}^{\Gamma} - \widetilde{p}^0\Big\|_2^2 = 0, \qquad \lim_{\Gamma\to\infty}\lim_{n\to\infty} n^{-1}\Big\|\widetilde{\underline{p}}^{\Gamma}\Big\|_2^2 < \infty. \tag{G.178}$$

因此，我们进一步有 $\lim_{n\to\infty} n^{-1}\big\|\widetilde{p}^0\big\|_2^2 < \infty$。此外，根据 假设 (A1) 至 (A3)，

$$n^{-1}\|z\|_2^2 \to \overline{\kappa}_2\rho < \infty, \qquad n^{-1}\|\varepsilon\|_2^2 \to \mathbb{E}\big[\mathsf{E}^2\big] < \infty, \tag{G.179}$$

几乎必然成立，其中 $n\to\infty$。综合 (G.176) 至 (G.179)，回到 (G.171)，我们得出 (G.160) 对 $\tau=0$.

归纳步骤。 现在固定任意一个 $t\geq 1$。我们将以下命题作为归纳假设： (G.155)对每个 $0\leq\tau\leq t-1$, (G.156) 对每个 $0\leq\tau\leq t$, (G.157) 以及 (G.158) 对每个 $1\leq\varsigma,\ \tau\leq t$, (G.159)以及 (G.160) 对每个 $0\leq\tau\leq t$。证明的其余部分就是验证对 $t+1$也有类似命题成立。

首先考虑 $a_{\Gamma+t+1}$:

$$\begin{aligned}\overline{\kappa}_2\big|\underline{a}_{\Gamma+t+1} - a_{t+1}\big| &\overset{(4.29f)}{=} \Big|\mathbb{E}\big[\underline{\Psi}_{\Gamma+t}(\Lambda_n^2)\Lambda_n^2\big]\widetilde{\underline{a}}_{\Gamma+t} - \mathbb{E}\big[\Psi_t(\Lambda_n^2)\Lambda_n^2\big]\widetilde{a}_t + \mathbb{E}\Big[\widetilde{\underline{\Psi}}_{\Gamma+t}(\Lambda_n)\Lambda_n\Big]\widetilde{\underline{b}}_{\Gamma+t} - \mathbb{E}\Big[\widetilde{\Psi}_t(\Lambda_n)\Lambda_n\Big]\widetilde{b}_t\Big| \\ &\leq \big|\mathbb{E}\big[\Psi_t(\Lambda_n^2)\Lambda_n^2\big]\big|\cdot\big|\widetilde{\underline{a}}_{\Gamma+t} - \widetilde{a}_t\big| + \Big|\mathbb{E}\Big[\widetilde{\Psi}_t(\Lambda_n)\Lambda_n\Big]\Big|\cdot\Big|\widetilde{\underline{b}}_{\Gamma+t} - \widetilde{a}_t\Big| \to 0,\end{aligned} \tag{G.180}$$

$\widetilde{\ }$作为 $\Gamma\to\infty$，其中最后一行用到了 $\Psi_{\Gamma+t}$, $\Psi_{\Gamma+t}$ 在 (G.151e)中的三角不等式，以及关于 $\widetilde{a}_{\Gamma+t}$, $\widetilde{b}_{\Gamma+t}$ 在 (G.156)中的归纳假设。类似的论证可证明 $\big|b_{\Gamma+t+1} - b_{t+1}\big| \to 0$ 作为 $\Gamma\to\infty$，从而对 (G.155) 成立，其中 $t+1$。

然后，对于任意 $0\leq\tau\leq t$，我们考虑 $\mathbb{E}\big[\mathsf{J}_{\Gamma+t+1}\mathsf{J}_{\Gamma+\tau+1}\big]$:

$$\big|\mathbb{E}\big[\underline{\mathsf{J}}_{\Gamma+t+1}\underline{\mathsf{J}}_{\Gamma+\tau+1}\big] - \mathbb{E}[\mathsf{J}_{t+1}\mathsf{J}_{\tau+1}]\big| \tag{G.181}$$

$$
\begin{aligned}
&\overset{(4.29c)}{\leq} \left|\Psi_t(\Lambda_n^2)\Psi_\tau(\Lambda_n^2)\right|\Big(\Big|\mathbb{E}\Big[\widetilde{\mathsf{P}}_{\Gamma+t}\widetilde{\mathsf{P}}_{\Gamma+\tau}\Big] - \mathbb{E}\Big[\widetilde{\mathsf{P}}_t\widetilde{\mathsf{P}}_\tau\Big]\Big| + \sigma^2\big|\widetilde{a}_{\Gamma+t}\widetilde{a}_{\Gamma+\tau} - \widetilde{a}_t\widetilde{a}_\tau\big|\Big)\\
&+ \left|\mathbb{E}\left[\Psi_t(\Lambda_n^2)\Psi_\tau(\Lambda_n^2)\Lambda_n^2\right]\right| \cdot \rho \cdot \left|\widetilde{a}_{\Gamma+t}\widetilde{a}_{\Gamma+\tau} - \widetilde{a}_t\widetilde{a}_\tau\right|\\
&+ \left|\mathbb{E}\Big[\widetilde{\Psi}_t(\Lambda_n)\widetilde{\Psi}_\tau(\Lambda_n)\Big]\right| \cdot \Big|\mathbb{E}\Big[\widetilde{\mathsf{R}}_{\Gamma+t}\widetilde{\mathsf{R}}_{\Gamma+\tau}\Big] - \mathbb{E}\Big[\widetilde{\mathsf{R}}_t\widetilde{\mathsf{R}}_\tau\Big]\Big| + \sigma^2\big|\underline{a}_{\Gamma+t+1}\underline{a}_{\Gamma+\tau+1} - a_{t+1}a_{\tau+1}\big|\\
&+ \rho\Big|\Psi_t(\Lambda_n^2)\widetilde{\Psi}_\tau(\Lambda)\Lambda_n\Big|\Big|\widetilde{\underline{a}}_{\Gamma+t}\widetilde{\underline{b}}_{\Gamma+\tau} - \widetilde{a}_t\widetilde{b}_\tau\Big| + \rho\Big|\Psi_\tau(\Lambda_n^2)\widetilde{\Psi}_t(\Lambda)\Lambda_n\Big|\Big|\widetilde{\underline{a}}_{\Gamma+\tau}\widetilde{\underline{b}}_{\Gamma+t} - \widetilde{a}_\tau\widetilde{b}_t\Big|.
\end{aligned}
$$

回顾定义(G.151f) 并使用归纳假设 (G.155) 和 (G.157),

$$
\Big|\mathbb{E}\Big[\widetilde{\underline{\mathsf{P}}}_{\Gamma+t}\widetilde{\underline{\mathsf{P}}}_{\Gamma+\tau}\Big] - \mathbb{E}\Big[\widetilde{\mathsf{P}}_t\widetilde{\mathsf{P}}_\tau\Big]\Big| = \big|\mathbb{E}\big[g_t(\underline{a}_{\Gamma+t}\mathsf{Z} + \underline{\mathsf{J}}_{\Gamma+t};\mathsf{Y})g_\tau(\underline{a}_{\Gamma+\tau}\mathsf{Z} + \underline{\mathsf{J}}_{\Gamma+\tau};\mathsf{Y})\big] - \mathbb{E}[g_t(a_t\mathsf{Z} + \mathsf{J}_t;\mathsf{Y})g_\tau(a_\tau\mathsf{Z} + \mathsf{J}_\tau;\mathsf{Y})]\big| \to 0,
$$

当 $\Gamma \to \infty$ 时。同样，$\mathbb{E}\Big[\widetilde{\mathsf{R}}_{\Gamma+t}\widetilde{\mathsf{R}}_{\Gamma+\tau}\Big] - \mathbb{E}\Big[\widetilde{\mathsf{R}}_t\widetilde{\mathsf{R}}_\tau\Big] \xrightarrow[\Gamma\to\infty]{} 0$。利用上述观察以及从 (G.155) 到 (G.157) 的归纳假设，可得 (G.181) 在 $\Gamma \to \infty$ 时收敛到 0。类似的论证可用于证明 $\big|\mathbb{E}\big[\mathsf{J}_\Gamma\mathsf{J}_{\Gamma+t+1}\big] - c_p^{-1}\mathbb{E}[\mathsf{J}_0\mathsf{J}_{t+1}]\big| \xrightarrow[\Gamma\to\infty]{} 0$，从而验证 $t+1$ 的 (G.157)。对 $t+1$ 的 (G.158) 的验证类似，故略去。

接下来，考虑 $\widetilde{a}_{\Gamma+t+1}$:

$$
\begin{aligned}
\widetilde{a}_{\Gamma+t+1} &\overset{(4.29e)}{=} \sigma^{-2}\mathbb{E}\Big[g_{\Gamma+t+1}(\underline{\mathsf{P}}_{\Gamma+t+1};\mathsf{Y})\mathsf{Z}\Big] = \sigma^{-2}\mathbb{E}\big[g_{t+1}(\underline{a}_{\Gamma+t+1}\mathsf{Z} + \underline{\mathsf{J}}_{\Gamma+t+1};\mathsf{Y})\big]\\
&\xrightarrow[\Gamma\to\infty]{} \sigma^{-2}\mathbb{E}[g_{t+1}(a_{t+1}\mathsf{Z} + \mathsf{J}_{t+1};\mathsf{Y})] = \widetilde{a}_{t+1},
\end{aligned}
$$

其中该收敛由上文所示的(G.180) 和(G.181) 得出。类似的论证表明，$\widetilde{b}_{\Gamma+t+1} \to \widetilde{b}_{t+1}$在 $\Gamma \to \infty$时成立，从而证明了(G.156) 对 $t+1$成立。

最后，我们旨在证明(G.160) 对 $t+1$成立。记

$$
\begin{aligned}
\underline{P}^{\Gamma+t+1} &\coloneqq \begin{bmatrix} c_p\underline{p}^\Gamma & \underline{p}^{\Gamma+1} & \cdots & \underline{p}^{\Gamma+t+1} & \widetilde{\underline{p}}^\Gamma & \cdots & \widetilde{\underline{p}}^{\Gamma+t+1} & z & \varepsilon \end{bmatrix} \in \mathbb{R}^{d\times 2(t+3)},\\
P^{t+1} &\coloneqq \begin{bmatrix} p^0 & p^1 & \cdots & p^{t+1} & \widetilde{p}^0 & \cdots & \widetilde{p}^{t+1} & z & \varepsilon \end{bmatrix} \in \mathbb{R}^{d\times 2(t+3)}.
\end{aligned}
\tag{G.182}
$$

为简便起见，记 $\phi \equiv \phi_p$，利用 ϕ 的伪 Lipschitz 性质并应用柯西–施瓦茨不等式，我们将(G.160) 的左端（对 $t+1$）界定如下:

$$
\left|\frac{1}{n}\sum_{i=1}^n \phi(\underline{P}_{i,:}^{\Gamma+t+1}) - \frac{1}{n}\sum_{i=1}^n \phi(P_{i,:}^{t+1})\right| \leq C\sqrt{\frac{3}{n}\sum_{i=1}^n\Big(1 + \big\|\underline{P}_{i,:}^{\Gamma+t+1}\big\|_2^2 + \big\|P_{i,:}^{t+1}\big\|_2^2\Big)}\sqrt{\frac{1}{n}\sum_{i=1}^n\big\|\underline{P}_{i,:}^{\Gamma+t+1} - P_{i,:}^{t+1}\big\|_2^2}.
\tag{G.183}
$$

由定义(G.182),

$$
\frac{1}{n}\sum_{i=1}^n\big\|\underline{P}_{i,:}^{\Gamma+t+1} - P_{i,:}^{t+1}\big\|_2^2 = \frac{1}{n}\left(\big\|c_p\underline{p}^\Gamma - p^0\big\|_2^2 + \sum_{i=1}^{t+1}\big\|\underline{p}^{\Gamma+i} - p^i\big\|_2^2 + \sum_{i=0}^{t+1}\Big\|\widetilde{\underline{p}}^{\Gamma+i} - \widetilde{p}^i\Big\|_2^2\right).
\tag{G.184}
$$

归纳假设(G.160) 意味着上述右侧的所有项在0 处收敛，极限为$n \to \infty$，随后是 $\Gamma \to \infty$，但以下两项除外：$n^{-1}\big\|p^{\Gamma+t+1} - p^{t+1}\big\|_2^2$，$n^{-1}\Big\|\widetilde{p}^{\Gamma+t+1} - \widetilde{p}^{t+1}\Big\|_2^2$ 我们将在下文分别考察这两项。根据更新规则(4.19b)和(G.151b)，以及(G.151e)和 (G.151f)中去噪器的选取,

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\big\|\underline{p}^{\Gamma+t+1} - p^{t+1}\big\|_2 &\leq \frac{1}{\sqrt{n}}\big\|\Psi_t(XX^\top)\big\|_2\big\|g_t(\underline{p}^{\Gamma+t};y) - g_t(p^t;y)\big\|_2 + \frac{1}{\sqrt{n}}\Big\|\widetilde{\Psi}_t(X)\Big\|_2\big\|f_t(\underline{r}^{\Gamma+t}) - f_t(r^t)\big\|_2\\
&\leq \frac{C}{\sqrt{n}}\big\|\Psi_t(XX^\top)\big\|_2\big\|\underline{p}^{\Gamma+t} - p^t\big\|_2 + \frac{C}{\sqrt{n}}\Big\|\widetilde{\Psi}_t(X)\Big\|_2\big\|\underline{r}^{\Gamma+t} - r^t\big\|_2,
\end{aligned}
$$

这里用到了三角不等式以及 f_t的 g_t （见假设（A10））的 Lipschitz 连续性。然后，假设（A9）和归纳假设 (G.159) 及(G.160) 保证

$$\lim_{\Gamma\to\infty}\lim_{n\to\infty} n^{-1}\left\|\underline{p}^{\Gamma+t+1}-p^{t+1}\right\|_2^2=0. \tag{G.185}$$

根据去噪器的选择 (G.151f) 以及 f_{t+1} 的*Lipschitz*连续性， g_{t+1}可立即推出

$$\frac{1}{\sqrt{n}}\left\|\widetilde{\underline{p}}^{\Gamma+t+1}-\widetilde{p}^{t+1}\right\|_2\le\frac{C}{\sqrt{n}}\left\|\underline{p}^{\Gamma+t+1}-p^{t+1}\right\|_2\to 0, \tag{G.186}$$

依次为 $n\to\infty$、$\Gamma\to\infty$。因此，在相同的序贯极限下，(G.184) 收敛到 0。然后我们证明 $\lim_{\Gamma\to\infty}\lim\sup_{n\to\infty} n^{-1}\sum_{i=1}^n\left\|P_i^{\Gamma+t+1}\right\|_2^2$，即: 是有限的。为此，

$$\frac{1}{n}\sum_{i=1}^n\left\|\underline{P}_{i,:}^{\Gamma+t+1}\right\|_2^2=\frac{1}{n}\left(c_p^2\left\|\underline{p}^{\Gamma}\right\|_2^2+\sum_{i=1}^{t+1}\left\|\underline{p}^{\Gamma+i}\right\|_2^2+\sum_{i=0}^{t+1}\left\|\widetilde{\underline{p}}^{\Gamma+i}\right\|_2^2+\|z\|_2^2+\|\varepsilon\|_2^2\right). \tag{G.187}$$

根据 定理 G.4 中关于辅助 GVAMP 迭代 (G.151)的状态演化结果，对于任意 $\Gamma\ge 0$ 固定的以及任意 $0\le i\le t+1$,

$$\frac{1}{n}\left\|\underline{p}^{\Gamma+i}\right\|_2^2\to\mathbb{E}\left[\underline{\mathsf{P}}_{\Gamma+i}^2\right]<\infty,\qquad\qquad\frac{1}{n}\left\|\widetilde{\underline{p}}^{\Gamma+i}\right\|_2^2\to\mathbb{E}\left[\widetilde{\underline{\mathsf{P}}}_{\Gamma+i}^2\right]<\infty,$$

几乎必然地，当 $n\to\infty$时。根据归纳假设 (G.155) 和 (G.157)，

$$\lim_{\Gamma\to\infty}\lim_{n\to\infty}\frac{1}{n}\left\|\underline{p}^{\Gamma+i}\right\|_2^2=\mathbb{E}\left[\mathsf{P}_i^2\right]<\infty,\qquad\lim_{\Gamma\to\infty}\lim_{n\to\infty}\frac{1}{n}\left\|\widetilde{\underline{p}}^{\Gamma+i}\right\|_2^2=\mathbb{E}\left[\widetilde{\mathsf{P}}_i^2\right]<\infty. \tag{G.188}$$

将其与归纳假设 (G.160) 以及上文给出的 (G.185) 和 (G.186) 相结合，并应用三角不等式，我们还可得到

$$\lim_{\Gamma\to\infty}\lim_{n\to\infty}\frac{1}{n}\left\|p^i\right\|_2^2<\infty,\qquad\qquad\lim_{\Gamma\to\infty}\lim_{n\to\infty}\frac{1}{n}\left\|\widetilde{p}^i\right\|_2^2<\infty, \tag{G.189}$$

对每个 $0\le i\le t+1$。结合 (G.179)、(G.188) 和 (G.189)，可知 $n^{-1}\sum_{i=1}^n\left\|P_i^{\Gamma+t+1}{}_{,:}\right\|_2^2$ 以及 $n^{-1}\sum_{i=1}^n\left\|P_i^{t+1}{}_{,:}\right\|_2^2$ 在双重极限下均有界。最后，结合前面已证明 (G.184)趋于 0 这一事实，可知 (G.183) 的左端也趋于 0。因此，这就证明了 (G.159) 对 $t+1$成立。关于 (G.160) 的证明在形式上完全对称，故略去。至此归纳法完成，定理 G.5 中的所需结论得证。

□

H 定理 4.4 的证明

我们通过验证 (4.39) 中的迹为零和散度为零条件，证明 Bayes-GVAMP 的标准形确实是 (4.19) 中通用 GVAMP 的一种特殊情形，这些条件见 假设 (A7) 和 (A8)。

引理 H.1. 函数 Φ_t、Ψ_t 在 (4.39c) 和 (4.39d) 中满足 假设 (A7)所述的迹为零条件，而函数 f_t、g_t 满足 假设 (A8)所述的散度为零条件。

证明。 根据定义 v_2^t，c_2^t 在 (4.36f)中，

$$\mathbb{E}\left[\Phi_t(\Lambda_d^2)\right]=\frac{1}{v_2^t}\mathbb{E}\left[\frac{1}{\gamma_2^t+\tau_2^t\Lambda_d^2}\right]-1=\frac{v_2^t}{v_2^t}-1=0,$$
$$\mathbb{E}\left[\Psi_t(\Lambda_n^2)\right]=\frac{1}{c_2^t}\mathbb{E}\left[\frac{\Lambda_n^2}{\gamma_2^t+\tau_2^t\Lambda_n^2}\right]-1=\frac{c_2^t}{c_2^t}-1=0.$$

根据定义 v_1^t, c_1^t 在 (4.36b)中,

$$\mathbb{E}[f_t'(\mathsf{R}_t)] = \frac{1}{v_1^t}\mathbb{E}\big[g_{x1}'(\mathsf{R}_t;\gamma_1^t)\big] - 1 = \frac{v_1^t}{v_1^t} - 1 = 0,$$
$$\mathbb{E}[g_t'(\mathsf{P}_t;\mathsf{Y})] = \frac{1}{c_1^t}\mathbb{E}\big[g_{z1}'(\mathsf{P}_t,\mathsf{Y};\tau_1^t)\big] - 1 = \frac{c_1^t}{c_1^t} - 1 = 0.$$

证明完毕。 □

那么，定理 4.4 立即由下面的引理得到，该引理指出，标量 γ_1^t、τ_1^t、v_1^t、c_1^t、γ_2^t、τ_2^t、v_2^t、c_2^t与状态演化参数 $\widetilde{a}_t$、$\widetilde{b}_t$、a_t、b_t、$\mathbb{E}[\mathsf{J}_t^2]$、$\mathbb{E}[\mathsf{K}_t^2]$一一对应。

引理 *H.2*。令 γ_1^t、τ_1^t、γ_2^t、τ_2^t通过 *(4.36)* 定义，并令 $\widetilde{a}_t$、$\widetilde{b}_t$、a_t、b_t、$\mathbb{E}[\mathsf{J}_t^2]$、$\mathbb{E}[\mathsf{K}_t^2]$ 通过 *(4.29)*~ 定义，其中 Φ_t、Φ_t、Ψ_t、Ψ_t、f_t、g_t 由 *(4.39c)* 至 *(4.39e)* 给出。那么，这些量满足

$$\widetilde{a}_t = \tau_2^t, \qquad \widetilde{b}_t = \gamma_2^t - \frac{1}{\rho}, \qquad a_{t+1} = \mathbb{E}[\mathsf{J}_{t+1}^2] = \tau_1^{t+1} - \frac{1}{\sigma^2}, \qquad b_{t+1} = \mathbb{E}[\mathsf{K}_{t+1}^2] = \gamma_1^{t+1}. \tag{H.1}$$

证明。 该证明通过对 t 使用归纳法完成。为进行归纳，我们先作一些对任意 t 都成立的观察。

~**准备工作**。 根据定义 (4.39d) ，有 Ψ_t, Ψ_t, 成立

$$\Psi_t(x^2) = x\widetilde{\Psi}_t(x) - 1. \tag{H.2}$$

的无迹性质 Ψ_t （见 定理 H.1）于是推出

$$\mathbb{E}\Big[\Lambda_n\widetilde{\Psi}_t(\Lambda_n)\Big] = 1.$$

~我们利用这些观察来简化涉及 Ψ_t, Ψ_t 的若干期望表达式:

$$\mathbb{E}\big[\Psi_t(\Lambda_n^2)^2\big] = \mathbb{E}\Big[\Big(\Lambda_n\widetilde{\Psi}_t(\Lambda_n) - 1\Big)^2\Big] = \mathbb{E}\Big[\Lambda_n^2\widetilde{\Psi}_t(\Lambda_n)^2\Big] - 1, \tag{H.3a}$$
$$\mathbb{E}\big[\Psi_t(\Lambda_n^2)^2\Lambda_n^2\big] = \mathbb{E}\Big[\Big(\Lambda_n\widetilde{\Psi}_t(\Lambda_n) - 1\Big)^2\Lambda_n^2\Big] = \mathbb{E}\Big[\Lambda_n^4\widetilde{\Psi}_t(\Lambda_n)^2\Big] + \mathbb{E}\big[\Lambda_n^2\big] - 2\mathbb{E}\Big[\Lambda_n^3\widetilde{\Psi}_t(\Lambda_n)\Big], \tag{H.3b}$$
$$\mathbb{E}\Big[\Psi_t(\Lambda_n^2)\widetilde{\Psi}_t(\Lambda_n)\Lambda_n\Big] = \mathbb{E}\Big[\Big(\Lambda_n\widetilde{\Psi}_t(\Lambda_n) - 1\Big)\widetilde{\Psi}_t(\Lambda_n)\Lambda_n\Big] = \mathbb{E}\Big[\Lambda_n^2\widetilde{\Psi}_t(\Lambda_n)^2\Big] - 1. \tag{H.3c}$$

将 $\widetilde{\Psi}_t$ 在 (4.39d) 中的公式代入，我们将右侧的一些项连同 $\mathbb{E}\Big[\widetilde{\Psi}_t(\Lambda_n^2)\Big]$ 更明确地写为

$$\mathbb{E}\Big[\widetilde{\Psi}_t(\Lambda_n)^2\Big] = \frac{1}{(c_2^t)^2}\mathbb{E}\Big[\frac{\Lambda_n^2}{(\gamma_2^t + \tau_2^t\Lambda_n^2)^2}\Big], \qquad \mathbb{E}\Big[\Lambda_n^2\widetilde{\Psi}_t(\Lambda_n)^2\Big] = \frac{1}{(c_2^t)^2}\mathbb{E}\Big[\frac{\Lambda_n^4}{(\gamma_2^t + \tau_2^t\Lambda_n^2)^2}\Big], \tag{H.4a}$$
$$\mathbb{E}\Big[\Lambda_n^4\widetilde{\Psi}_t(\Lambda_n)^2\Big] = \frac{1}{(c_2^t)^2}\mathbb{E}\Big[\frac{\Lambda_n^6}{(\gamma_2^t + \tau_2^t\Lambda_n^2)^2}\Big], \qquad \mathbb{E}\Big[\Lambda_n^3\widetilde{\Psi}_t(\Lambda_n)\Big] = \frac{1}{c_2^t}\mathbb{E}\Big[\frac{\Lambda_n^4}{\gamma_2^t + \tau_2^t\Lambda_n^2}\Big], \tag{H.4b}$$

然后，我们将上述右侧的期望表示为 c_2^t 和 $\mathbb{E}\Big[\frac{\Lambda_n^4}{(\gamma_2^t+\tau_2^t\Lambda_n^2)^2}\Big]$:

$$\mathbb{E}\Big[\frac{\Lambda_n^4}{\gamma_2^t + \tau_2^t\Lambda_n^2}\Big] = \frac{1}{\tau_2^t}\mathbb{E}\Big[\Lambda_n^2 - \gamma_2^t\frac{\Lambda_n^2}{\gamma_2^t + \tau_2^t\Lambda_n^2}\Big] = \frac{\overline{\kappa}_2}{\tau_2^t} - \frac{\gamma_2^t c_2^t}{\tau_2^t},$$
$$\mathbb{E}\Big[\frac{\Lambda_n^6}{(\gamma_2^t + \tau_2^t\Lambda_n^2)^2}\Big] = \frac{1}{\tau_2^t}\mathbb{E}\Big[\frac{\Lambda_n^4}{\gamma_2^t + \tau_2^t\Lambda_n^2}\Big] - \frac{\gamma_2^t}{\tau_2^t}\mathbb{E}\Big[\frac{\Lambda_n^4}{(\gamma_2^t + \tau_2^t\Lambda_n^2)^2}\Big] = \frac{\overline{\kappa}_2}{(\tau_2^t)^2} - \frac{\gamma_2^t c_2^t}{(\tau_2^t)^2} - \frac{\gamma_2^t}{\tau_2^t}\mathbb{E}\Big[\frac{\Lambda_n^4}{(\gamma_2^t + \tau_2^t\Lambda_n^2)^2}\Big],$$

$$\mathbb{E}\left[\frac{\Lambda_n^2}{(\gamma_2^t+\tau_2^t\Lambda_n^2)^2}\right]=\frac{1}{\gamma_2^t}\mathbb{E}\left[\frac{\Lambda_n^2}{\gamma_2^t+\tau_2^t\Lambda_n^2}-\tau_2^t\frac{\Lambda_n^4}{(\gamma_2^t+\tau_2^t\Lambda_n^2)^2}\right]=\frac{c_2^t}{\gamma_2^t}-\frac{\tau_2^t}{\gamma_2^t}\mathbb{E}\left[\frac{\Lambda_n^4}{(\gamma_2^t+\tau_2^t\Lambda_n^2)^2}\right]. \tag{H.5}$$

将这些结果代回（H.4），可得

$$\mathbb{E}\left[\widetilde{\Psi}_t(\Lambda_n)^2\right]=\frac{1}{\gamma_2^t c_2^t}-\frac{\tau_2^t}{\gamma_2^t(c_2^t)^2}\mathbb{E}\left[\frac{\Lambda_n^4}{(\gamma_2^t+\tau_2^t\Lambda_n^2)^2}\right], \tag{H.6a}$$

$$\mathbb{E}\left[\Lambda_n^2\widetilde{\Psi}_t(\Lambda_n)^2\right]=\frac{1}{(c_2^t)^2}\mathbb{E}\left[\frac{\Lambda_n^4}{(\gamma_2^t+\tau_2^t\Lambda_n^2)^2}\right], \tag{H.6b}$$

$$\mathbb{E}\left[\Lambda_n^4\widetilde{\Psi}_t(\Lambda_n)^2\right]=\frac{\overline{\kappa}_2}{(\tau_2^t c_2^t)^2}-\frac{\gamma_2^t}{(\tau_2^t)^2 c_2^t}-\frac{\gamma_2^t}{\tau_2^t(c_2^t)^2}\mathbb{E}\left[\frac{\Lambda_n^4}{(\gamma_2^t+\tau_2^t\Lambda_n^2)^2}\right], \tag{H.6c}$$

$$\mathbb{E}\left[\Lambda_n^3\widetilde{\Psi}_t(\Lambda_n)\right]=\frac{\overline{\kappa}_2}{\tau_2^t c_2^t}-\frac{\gamma_2^t}{\tau_2^t}. \tag{H.6d}$$

~我们对涉及 Φ_t、Φ_t的若干量进行类似的推导。首先注意，根据(4.39c) 中两个函数的定义以及γ_1^{t+1}在(4.36h)中的定义，以下恒等式成立

$$\gamma_2^t\Phi_t(x^2)+\tau_2^t x\widetilde{\Phi}_t(x)=\frac{1}{v_2^t}-\gamma_2^t=\gamma_1^{t+1}. \tag{H.7}$$

通过重新 整理各项并利用 Φ_t 的无迹性质（参见定理 H.1），可得 s

$$\Phi_t(x^2)=\frac{1}{\gamma_2^t}\Big(\gamma_1^{t+1}-\tau_2^t x\widetilde{\Phi}_t(x)\Big),\qquad \mathbb{E}\Big[\Lambda_d\widetilde{\Phi}_t(\Lambda_d)\Big]=\frac{\gamma_1^{t+1}}{\tau_2^t}.$$

~我们利用上述观察来计算涉及这些 Φ_t、Φ_t的若干期望:

$$\begin{aligned}\mathbb{E}\big[\Phi_t(\Lambda_d^2)^2\big]&=\mathbb{E}\left[\left(\frac{1}{\gamma_2^t}\Big(\gamma_1^{t+1}-\tau_2^t\Lambda_d\widetilde{\Phi}_t(\Lambda_d)\Big)\right)^2\right]\\&=\frac{(\gamma_1^{t+1})^2}{(\gamma_2^t)^2}+\frac{(\tau_2^t)^2}{(\gamma_2^t)^2}\mathbb{E}\Big[\Lambda_d^2\widetilde{\Phi}_t(\Lambda_d)^2\Big]-2\frac{\gamma_1^{t+1}\tau_2^t}{(\gamma_2^t)^2}\cdot\frac{\gamma_1^{t+1}}{\tau_2^t}=\frac{(\tau_2^t)^2}{(\gamma_2^t)^2}\mathbb{E}\Big[\widetilde{\Phi}_t(\Lambda_d)^2\Lambda_d^2\Big]-\frac{(\gamma_1^{t+1})^2}{(\gamma_2^t)^2},\end{aligned} \tag{H.8a}$$

$$\begin{aligned}\mathbb{E}\Big[\Phi_t(\Lambda_d^2)\widetilde{\Phi}_t(\Lambda_d)\Lambda_d\Big]&=\mathbb{E}\left[\frac{1}{\gamma_2^t}\Big(\gamma_1^{t+1}-\tau_2^t\Lambda_d\widetilde{\Phi}_t(\Lambda_d)\Big)\widetilde{\Phi}_t(\Lambda_d)\Lambda_d\right]\\&=\frac{(\gamma_1^{t+1})^2}{\gamma_2^t\tau_2^t}-\frac{\tau_2^t}{\gamma_2^t}\mathbb{E}\Big[\widetilde{\Phi}_t(\Lambda_d)^2\Lambda_d^2\Big].\end{aligned} \tag{H.8b}$$

接着，我们推导出更明确的表达式，分别针对右侧出现的 $\mathbb{E}\Big[\widetilde{\Phi}_t\ (\Lambda_d)\ ^2\Lambda_d^2\Big]$ 和$\mathbb{E}\Big[\widetilde{\Phi}_t\ (\Lambda_d)\ ^2\Big]$。根据(4.39c)中对 $\widetilde{\Phi}_t$的定义，

$$\mathbb{E}\Big[\widetilde{\Phi}_t(\Lambda_d)^2\Big]=\frac{1}{(v_2^t)^2}\mathbb{E}\left[\frac{\Lambda_d^2}{(\gamma_2^t+\tau_2^t\Lambda_d^2)^2}\right],\qquad \mathbb{E}\Big[\widetilde{\Phi}_t(\Lambda_d)^2\Lambda_d^2\Big]=\frac{1}{(v_2^t)^2}\mathbb{E}\left[\frac{\Lambda_d^4}{(\gamma_2^t+\tau_2^t\Lambda_d^2)^2}\right],$$

通过初等代数，右侧分式的期望可以用 v_2^t和 $\mathbb{E}\Big[\quad{}^{1}\quad{}_{(\gamma_2^t+\tau_2^t\Lambda_d^2)\ ^2}\Big]$ 表示。具体而言，我们有

$$\mathbb{E}\left[\frac{\Lambda_d^2}{(\gamma_2^t+\tau_2^t\Lambda_d^2)^2}\right]=\frac{1}{\tau_2^t}\mathbb{E}\left[\frac{1}{\gamma_2^t+\tau_2^t\Lambda_d^2}-\gamma_2^t\frac{1}{(\gamma_2^t+\tau_2^t\Lambda_d^2)^2}\right]=\frac{v_2^t}{\tau_2^t}-\frac{\gamma_2^t}{\tau_2^t}\mathbb{E}\left[\frac{1}{(\gamma_2^t+\tau_2^t\Lambda_d^2)^2}\right], \tag{H.9}$$

and

$$(\gamma_2^t)^2\mathbb{E}\left[\frac{1}{(\gamma_2^t+\tau_2^t\Lambda_d)^2}\right]=\mathbb{E}\left[\left(1-\frac{\tau_2^t\Lambda_d^2}{\gamma_2^t+\tau_2^t\Lambda_d^2}\right)^2\right]=1+(\tau_2^t)^2\mathbb{E}\left[\frac{\Lambda_d^4}{(\gamma_2^t+\tau_2^t\Lambda_d^2)^2}\right]-2\mathbb{E}\left[\frac{\tau_2^t\Lambda_d^2}{\gamma_2^t+\tau_2^t\Lambda_d^2}\right]$$
$$=1+(\tau_2^t)^2\mathbb{E}\left[\frac{\Lambda_d^4}{(\gamma_2^t+\tau_2^t\Lambda_d^2)^2}\right]-2\left(1-\gamma_2^t\mathbb{E}\left[\frac{1}{\gamma_2^t+\tau_2^t\Lambda_d^2}\right]\right)$$
$$=(\tau_2^t)^2\mathbb{E}\left[\frac{\Lambda_d^4}{(\gamma_2^t+\tau_2^t\Lambda_d^2)^2}\right]+2\gamma_2^tv_2^t-1,$$

这意味着

$$\mathbb{E}\left[\frac{\Lambda_d^4}{(\gamma_2^t+\tau_2^t\Lambda_d^2)^2}\right]=\frac{(\gamma_2^t)^2}{(\tau_2^t)^2}\mathbb{E}\left[\frac{1}{(\gamma_2^t+\tau_2^t\Lambda_d)^2}\right]-\frac{2\gamma_2^tv_2^t}{(\tau_2^t)^2}+\frac{1}{(\tau_2^t)^2}. \tag{H.10}$$

将(H.9) 和(H.10) 中的两项都除以$(v_2^t)^2$，便得到$\mathbb{E}\left[\widetilde{\Phi}_t(\Lambda_d)^2\right]$、$\mathbb{E}\left[\widetilde{\Phi}_t(\Lambda_d)^2\Lambda_d^2\right]$的如下表达式:

$$\mathbb{E}\left[\widetilde{\Phi}_t(\Lambda_d)^2\right]=\frac{1}{\tau_2^tv_2^t}-\frac{\gamma_2^t}{\tau_2^t(v_2^t)^2}\mathbb{E}\left[\frac{1}{(\gamma_2^t+\tau_2^t\Lambda_d^2)^2}\right],$$
$$\mathbb{E}\left[\widetilde{\Phi}_t(\Lambda_d)^2\Lambda_d^2\right]=\frac{(\gamma_2^t)^2}{(\tau_2^tv_2^t)^2}\mathbb{E}\left[\frac{1}{(\gamma_2^t+\tau_2^t\Lambda_d^2)^2}\right]-\frac{2\gamma_2^t}{(\tau_2^t)^2v_2^t}+\frac{1}{(\tau_2^tv_2^t)^2}.$$

我们还对 v_1^t、 c_1^t作一观察。根据它们在(4.36b)中的定义、 g'_{x1}、 g'_{z1}在(4.34) 中的表达式以及条件期望的塔式性质，我们有

$$v_1^t=\mathbb{E}[\mathrm{Var}[\mathsf{B}_*\,|\,\mathsf{R}_t]]=\mathbb{E}\left[\mathbb{E}\left[\mathsf{B}_*^2\,\middle|\,\mathsf{R}_t\right]-\mathbb{E}[\mathsf{B}_*\,|\,\mathsf{R}_t]^2\right]=\rho-\mathbb{E}[\mathsf{B}_*\mathbb{E}[\mathsf{B}_*\,|\,\mathsf{R}_t]], \tag{H.12}$$
$$c_1^t=\mathbb{E}[\mathrm{Var}[\mathsf{Z}\,|\,\mathsf{P}_t,\mathsf{Y}]]=\mathbb{E}\left[\mathbb{E}\left[\mathsf{Z}^2\,\middle|\,\mathsf{P}_t,\mathsf{Y}\right]-\mathbb{E}[\mathsf{Z}\,|\,\mathsf{P}_t,\mathsf{Y}]^2\right]=\sigma^2-\mathbb{E}[\mathsf{Z}\mathbb{E}[\mathsf{Z}\,|\,\mathsf{P}_t,\mathsf{Y}]], \tag{H.13}$$

从而

$$\mathbb{E}[\mathsf{B}_*\mathbb{E}[\mathsf{B}_*\,|\,\mathsf{R}_t]]=\rho-v_1^t,\qquad\qquad\mathbb{E}[\mathsf{Z}\mathbb{E}[\mathsf{Z}\,|\,\mathsf{P}_t,\mathsf{Y}]]=\sigma^2-c_1^t. \tag{H.14}$$

我们现在可以给出归纳证明。我们首先进行归纳步骤，因为基例 $t=0$ 的验证采用类似的论证，其中使用的是初始条件(4.38) 而非归纳假设。

归纳步骤。 现假设结论(H.1) 在直至 $t-1$的时刻均成立，我们验证其在时刻 t的有效性。

我们从 $\widetilde{a}_t$开始:

$$\sigma^2\widetilde{a}_t\overset{(4.29e)}{=}\mathbb{E}\left[\widetilde{\mathsf{P}}_t\mathsf{Z}\right]\overset{(4.29g)}{=}\mathbb{E}[g_t(\mathsf{P}_t;\mathsf{Y})\mathsf{Z}]\overset{(4.39e)}{=}\frac{1}{c_1^t}\mathbb{E}\left[g_{z1}(\mathsf{P}_t,\mathsf{Y};\tau_1^t)\mathsf{Z}\right]-\mathbb{E}[\mathsf{P}_t\mathsf{Z}]\overset{(4.33)\text{ and }(4.29h)}{=}\frac{1}{c_1^t}\mathbb{E}[\mathbb{E}[\mathsf{Z}\,|\,\mathsf{P}_t,\mathsf{Y}]\mathsf{Z}]-a_t\sigma^2$$
$$=\frac{1}{c_1^t}(\sigma^2-c_1^t)-\left(\tau_1^t-\frac{1}{\sigma^2}\right)\sigma^2=\sigma^2\left(\frac{1}{c_1^t}-\tau_1^t\right)\overset{(4.36d)}{=}\sigma^2\tau_2^t.$$

这意味着

$$\widetilde{a}_t=\tau_2^t, \tag{H.15}$$

如预期。在上面的推导中，第二行的第一个等号用到了观察结果 (H.14)和归纳假设 $a_t=\tau_1^t-\sigma^{-2}$。

同理，

$$\rho\widetilde{b}_t \overset{(4.29e)}{=} \mathbb{E}\Big[\widetilde{\mathsf{R}}_t\mathsf{B}_*\Big] \overset{(4.29g)}{=} \mathbb{E}[f_t(\mathsf{R}_t)\mathsf{B}_*] \overset{(4.39e)}{=} \frac{1}{v_1^t}\mathbb{E}\big[g_{x1}(\mathsf{R}_t;\gamma_1^t)\mathsf{B}_*\big] - \mathbb{E}[\mathsf{R}_t\mathsf{B}_*] \overset{(4.32)\text{ and }(4.29h)}{=} \frac{1}{v_1^t}\mathbb{E}[\mathbb{E}[\mathsf{B}_*\,|\,\mathsf{R}_t]\mathsf{B}_*] - b_t\rho,$$
$$= \frac{1}{v_1^t}(\rho - v_1^t) - \gamma_1^t\rho = \rho\bigg(\frac{1}{v_1^t} - \gamma_1^t\bigg) - 1 \overset{(4.36d)}{=} \rho\gamma_2^t - 1,$$

其中第二行的第一个等号用到了 (H.14) 和归纳假设 $b_t = \gamma_1^t$。这意味着

$$\widetilde{b}_t = \gamma_2^t - \frac{1}{\rho}. \tag{H.16}$$

接下来考虑 b_{t+1}:

$$b_{t+1} \overset{(4.29f)}{=} \mathbb{E}\Big[\widetilde{\Phi}_t(\Lambda_d)\Lambda_d\Big]\widetilde{a}_t \overset{(H.15)\text{ and }(4.39c)}{=} \frac{1}{v_2^t}\cdot\mathbb{E}\bigg[\frac{\tau_2^t\Lambda_d^2}{\gamma_2^t+\tau_2^t\Lambda_d^2}\bigg]$$
$$= \frac{1}{v_2^t}\bigg(1 - \gamma_2^t\cdot\mathbb{E}\bigg[\frac{1}{\gamma_2^t+\tau_2^t\Lambda_d^2}\bigg]\bigg) \overset{(4.36f)}{=} \frac{1}{v_2^t}\big(1-\gamma_2^t\cdot v_2^t\big) = \frac{1}{v_2^t} - \gamma_2^t \overset{(4.36h)}{=} \gamma_1^{t+1}, \tag{H.17}$$

如所需。接下来，转到 a_{t+1}:

$$a_{t+1} \overset{(4.29f)}{=} \frac{1}{\overline{\kappa}_2}\Big(\mathbb{E}\Big[\widetilde{\Psi}_t(\Lambda_n)\Lambda_n\Big]\widetilde{b}_t + \mathbb{E}\big[\Psi_t(\Lambda_n^2)\Lambda_n^2\big]\widetilde{a}_t\Big)$$
$$\overset{(H.2),\,(H.15),\,(H.16)\text{ and }(4.39d)}{=} \frac{1}{\overline{\kappa}_2}\bigg(\mathbb{E}\bigg[\frac{1}{c_2^t}\cdot\frac{\Lambda_n^2}{\gamma_2^t+\tau_2^t\Lambda_n^2}\bigg]\Big(\gamma_2^t-\frac{1}{\rho}\Big) + \mathbb{E}\bigg[\Big(\frac{1}{c_2^t}\cdot\frac{\Lambda_n^2}{\gamma_2^t+\tau_2^t\Lambda_n^2}-1\Big)\Lambda_n^2\bigg]\tau_2^t\bigg)$$
$$\overset{(4.36f)}{=} \frac{1}{\overline{\kappa}_2}\bigg\{\frac{1}{c_2^t}\cdot c_2^t\cdot\Big(\gamma_2^t-\frac{1}{\rho}\Big) + \frac{\tau_2^t}{c_2^t}\cdot\frac{1}{\tau_2^t}\bigg(\mathbb{E}\big[\Lambda_n^2\big] - \gamma_2^t\mathbb{E}\bigg[\frac{\Lambda_n^2}{\gamma_2^t+\tau_2^t\Lambda_n^2}\bigg]\bigg) - \tau_2^t\mathbb{E}\big[\Lambda_n^2\big]\bigg\}$$
$$\overset{(4.36f)}{=} \frac{1}{\overline{\kappa}_2}\bigg\{\gamma_2^t - \frac{1}{\rho} + \frac{1}{c_2^t}\big(\overline{\kappa}_2 - \gamma_2^t c_2^t\big) - \tau_2^t\overline{\kappa}_2\bigg\} = \frac{1}{\overline{\kappa}_2}\bigg\{\frac{1}{c_2^t}\overline{\kappa}_2 - \tau_2^t\overline{\kappa}_2 - \frac{1}{\rho}\bigg\} \overset{(3.9)}{=} \frac{1}{c_2^t} - \tau_2^t - \frac{1}{\sigma^2} \overset{(4.36h)}{=} \tau_1^{t+1} - \frac{1}{\sigma^2}. \tag{H.18}$$

在继续讨论 $\mathbb{E}\big[\mathsf{J}_{t+1}^2\big], \mathbb{E}\big[\mathsf{K}_{t+1}^2\big]$ 之前，利用归纳假设，我们推导出 $\mathbb{E}\big[\widetilde{\mathsf{P}}_t^2\big], \mathbb{E}\big[\widetilde{\mathsf{R}}_t^2\big]$ 关于 τ_2^t, γ_2^t 的表达式。根据条件期望的塔式性质，

$$\mathbb{E}\big[\widetilde{\mathsf{P}}_t^2\big] = \mathbb{E}\bigg[\Big(\frac{1}{c_1^t}\mathbb{E}[\mathsf{Z}\,|\,\mathsf{P}_t,\mathsf{Y}] - \mathsf{P}_t\Big)^2\bigg] = \frac{1}{(c_1^t)^2}\mathbb{E}[\mathsf{Z}\mathbb{E}[\mathsf{Z}\,|\,\mathsf{P}_t,\mathsf{Y}]] + \mathbb{E}\big[\mathsf{P}_t^2\big] - \frac{2}{c_1^t}\mathbb{E}[\mathsf{P}_t\mathsf{Z}]$$
$$= \frac{1}{(c_1^t)^2}(\sigma^2 - c_1^t) + \bigg[\Big(\tau_1^t - \frac{1}{\sigma^2}\Big)^2\sigma^2 + \Big(\tau_1^t - \frac{1}{\sigma^2}\Big)\bigg] - \frac{2}{c_1^t}\Big(\tau_1^t - \frac{1}{\sigma^2}\Big)\sigma^2$$
$$= \sigma^2\Big(\frac{1}{c_1^t} - \tau_1^t\Big)\Big(\frac{1}{c_1^t} - \tau_1^t + \frac{1}{\sigma^2}\Big) \overset{(4.36d)}{=} \sigma^2\tau_2^t\Big(\tau_2^t + \frac{1}{\sigma^2}\Big), \tag{H.19}$$

其中，第二行由归纳假设 $a_t = \mathbb{E}\big[\mathsf{J}_t^2\big] = \tau_1^t - \sigma^{-2}$ 得出。同理，

$$\mathbb{E}\big[\widetilde{\mathsf{R}}_t^2\big] = \mathbb{E}\bigg[\Big(\frac{1}{v_1^t}\mathbb{E}[\mathsf{B}_*\,|\,\mathsf{R}_t] - \mathsf{R}_t\Big)^2\bigg] = \frac{1}{(v_1^t)^2}\mathbb{E}[\mathsf{B}_*\mathbb{E}[\mathsf{B}_*\,|\,\mathsf{R}_t]] + \mathbb{E}\big[\mathsf{R}_t^2\big] - \frac{2}{v_1^t}\mathbb{E}[\mathsf{B}_*\mathsf{R}_t]$$
$$= \frac{1}{(v_1^t)^2}(\rho - v_1^t) + \big[(\gamma_1^t)^2\rho + \gamma_1^t\big] - \frac{2}{v_1^t}\cdot\gamma_1^t\rho = \frac{\rho}{v_1^t}\Big(\frac{1}{v_1^t} - \gamma_1^t\Big) - \Big(\frac{1}{v_1^t} - \gamma_1^t\Big) - \rho\gamma_1^t\Big(\frac{1}{v_1^t} - \gamma_1^t\Big)$$
$$= \rho\gamma_2^t\Big(\frac{1}{v_1^t} - \gamma_1^t - \frac{1}{\rho}\Big) \overset{(4.36d)}{=} \rho\gamma_2^t\Big(\gamma_2^t - \frac{1}{\rho}\Big), \tag{H.20}$$

其中，第二行由归纳假设 $b_t = \mathbb{E}[\mathsf{K}_t^2] = \gamma_1^t$ 得出。

利用表达式 (H.19) 和 (H.20)，现在验证 $\mathbb{E}[\mathsf{J}_{t+1}^2] = \tau_1^{t+1} - \sigma^{-2}$，$\mathbb{E}[\mathsf{K}_{t+1}^2] = \gamma_1^{t+1}$。首先考虑 $\mathbb{E}[\mathsf{J}_{t+1}^2]$:

$$
\begin{aligned}
\mathbb{E}[\mathsf{J}_{t+1}^2] &\overset{(4.29c)}{=} \mathbb{E}[\Psi_t(\Lambda_n^2)^2]\Big(\mathbb{E}\Big[\widetilde{\mathsf{P}}_t^2\Big] - \widetilde{a}_t^2\sigma^2\Big) + \mathbb{E}[\Psi_t(\Lambda_n^2)^2\Lambda_n^2]\widetilde{a}_t^2\rho + \mathbb{E}\Big[\widetilde{\Psi}_t(\Lambda_n)^2\Big]\mathbb{E}\Big[\widetilde{\mathsf{R}}_t^2\Big] - a_{t+1}^2\sigma^2 \\
&\quad + 2\mathbb{E}\Big[\Psi_t(\Lambda_n^2)\widetilde{\Psi}_t(\Lambda_n)\Lambda_n\Big]\widetilde{a}_t\widetilde{b}_t\rho, \\
&= \Big(\mathbb{E}\Big[\Lambda_n^2\widetilde{\Psi}_t(\Lambda_n)^2\Big] - 1\Big)\Big[\sigma^2\tau_2^t\Big(\tau_2^t + \frac{1}{\sigma^2}\Big) - (\tau_2^t)^2\sigma^2\Big] \\
&\quad + \Big(\mathbb{E}\Big[\Lambda_n^4\widetilde{\Psi}_t(\Lambda_n)^2\Big] + \mathbb{E}[\Lambda_n^2] - 2\mathbb{E}\Big[\Lambda_n^3\widetilde{\Psi}_t(\Lambda_n)\Big]\Big)(\tau_2^t)^2\rho \\
&\quad + \mathbb{E}\Big[\widetilde{\Psi}_t(\Lambda_n)^2\Big]\rho\gamma_2^t\Big(\gamma_2^t - \frac{1}{\rho}\Big) - \Big(\tau_1^{t+1} - \frac{1}{\sigma^2}\Big)^2\sigma^2 \\
&\quad + 2\Big(\mathbb{E}\Big[\Lambda_n^2\widetilde{\Psi}_t(\Lambda_n)^2\Big] - 1\Big)\tau_2^t\Big(\gamma_2^t - \frac{1}{\rho}\Big)\rho \\
&\overset{(4.36h)}{=} \mathbb{E}\Big[\Lambda_n^2\widetilde{\Psi}_t(\Lambda_n)^2\Big]\Big(\tau_2^t + 2\tau_2^t\Big(\gamma_2^t - \frac{1}{\rho}\Big)\rho\Big) \\
&\quad + \Big(\mathbb{E}\Big[\Lambda_n^4\widetilde{\Psi}_t(\Lambda_n)^2\Big] + \mathbb{E}[\Lambda_n^2] - 2\mathbb{E}\Big[\Lambda_n^3\widetilde{\Psi}_t(\Lambda_n)\Big]\Big)(\tau_2^t)^2\rho + \mathbb{E}\Big[\widetilde{\Psi}_t(\Lambda_n)^2\Big]\rho\gamma_2^t\Big(\gamma_2^t - \frac{1}{\rho}\Big) \\
&\quad - \Big(\tau_2^t + 2\tau_2^t\Big(\gamma_2^t - \frac{1}{\rho}\Big)\rho + \Big(\frac{1}{c_2^t} - \tau_2^t - \frac{1}{\sigma^2}\Big)^2\sigma^2\Big).
\end{aligned} \tag{H.21}
$$

~在上面的推导中，第二个等号用到了以下表达式：涉及 Ψ_t 的若干期望的 Ψ_t表达式 (H.19) 和 (H.20)，关于 $\mathbb{E}\big[\widetilde{\mathsf{P}}_t^2\big]$，$\mathbb{E}\big[\widetilde{\mathsf{R}}_t^2\big]$，以及刚刚在 (H.15)，(H.16) 和(H.18) 中证明的 $\widetilde{a}_t$，$\widetilde{b}_t$，a_{t+1}的表达式。

我们先考察涉及期望的项，不包括最后一行 (H.21)。

$$
\begin{aligned}
&\mathbb{E}\Big[\Lambda_n^2\widetilde{\Psi}_t(\Lambda_n)^2\Big]\Big(\tau_2^t + 2\tau_2^t\Big(\gamma_2^t - \frac{1}{\rho}\Big)\rho\Big) \\
&\quad + \Big(\mathbb{E}\Big[\Lambda_n^4\widetilde{\Psi}_t(\Lambda_n)^2\Big] + \overline{\kappa}_2 - 2\mathbb{E}\Big[\Lambda_n^3\widetilde{\Psi}_t(\Lambda_n)\Big]\Big)(\tau_2^t)^2\rho + \mathbb{E}\Big[\widetilde{\Psi}_t(\Lambda_n)^2\Big]\rho\gamma_2^t\Big(\gamma_2^t - \frac{1}{\rho}\Big) \\
&\overset{(H.6)}{=} \Big(\frac{1}{(c_2^t)^2}\mathbb{E}\Big[\frac{\Lambda_n^4}{(\gamma_2^t + \tau_2^t\Lambda_n^2)^2}\Big]\Big)\Big(\tau_2^t + 2\tau_2^t\Big(\gamma_2^t - \frac{1}{\rho}\Big)\rho\Big) \\
&\quad + \Big(\frac{\overline{\kappa}_2}{(\tau_2^t c_2^t)^2} - \frac{\gamma_2^t}{(\tau_2^t)^2 c_2^t} - \frac{\gamma_2^t}{\tau_2^t (c_2^t)^2}\mathbb{E}\Big[\frac{\Lambda_n^4}{(\gamma_2^t + \tau_2^t\Lambda_n^2)^2}\Big] + \overline{\kappa}_2 - 2\Big(\frac{\overline{\kappa}_2}{\tau_2^t c_2^t} - \frac{\gamma_2^t}{\tau_2^t}\Big)\Big)(\tau_2^t)^2\rho \qquad \text{(H.22)} \\
&\quad + \Big(\frac{1}{\gamma_2^t c_2^t} - \frac{\tau_2^t}{\gamma_2^t (c_2^t)^2}\mathbb{E}\Big[\frac{\Lambda_n^4}{(\gamma_2^t + \tau_2^t\Lambda_n^2)^2}\Big]\Big)\rho\gamma_2^t\Big(\gamma_2^t - \frac{1}{\rho}\Big). \qquad \text{(H.23)}
\end{aligned}
$$

初等代数表明，$\mathbb{E}\Big[\frac{\Lambda_n^4}{(\gamma_2^t + \tau_2^t\Lambda_n^2)^2}\Big]$ 前面的系数之和为零：

$$
\frac{1}{(c_2^t)^2}\Big(\tau_2^t + 2\tau_2^t\Big(\gamma_2^t - \frac{1}{\rho}\Big)\rho\Big) - \frac{\gamma_2^t}{\tau_2^t (c_2^t)^2}(\tau_2^t)^2\rho - \frac{\tau_2^t}{\gamma_2^t (c_2^t)^2}\rho\gamma_2^t\Big(\gamma_2^t - \frac{1}{\rho}\Big) = 0.
$$

因此，所有涉及该期望的项都相互抵消。此外，(H.21)至(H.23) 中不涉及 $\mathbb{E}\Big[\frac{\Lambda_n^4}{(\gamma_2^t + \tau_2^t\Lambda_n^2)^2}\Big]$ 的其余项相加得

$$
\Big(\frac{\overline{\kappa}_2}{(\tau_2^t c_2^t)^2} - \frac{\gamma_2^t}{(\tau_2^t)^2 c_2^t} + \overline{\kappa}_2 - 2\Big(\frac{\overline{\kappa}_2}{\tau_2^t c_2^t} - \frac{\gamma_2^t}{\tau_2^t}\Big)\Big)(\tau_2^t)^2\rho + \frac{1}{\gamma_2^t c_2^t}\cdot\rho\gamma_2^t\Big(\gamma_2^t - \frac{1}{\rho}\Big)
$$

$$
\begin{aligned}
&- \left(\tau_2^t + 2\tau_2^t\left(\gamma_2^t - \frac{1}{\rho}\right)\rho + \left(\frac{1}{c_2^t} - \tau_2^t - \frac{1}{\sigma^2}\right)^2\sigma^2\right) \\
&= \frac{1}{c_2^t} - \tau_2^t - \frac{1}{\sigma^2} \overset{(4.36h)}{=} \tau_1^{t+1} - \frac{1}{\sigma^2}.
\end{aligned}
$$

因此，我们证明了

$$
\mathbb{E}\left[\mathsf{J}_{t+1}^2\right] = \tau_1^{t+1} - \sigma^{-2}. \tag{H.24}
$$

最后，我们转向 $\mathbb{E}\left[\mathsf{K}_{t+1}^2\right]$:

$$
\begin{aligned}
\mathbb{E}\left[\mathsf{K}_{t+1}^2\right] &\overset{(4.29d)}{=} \mathbb{E}\left[\Phi_t(\Lambda_d^2)^2\right]\mathbb{E}\left[\widetilde{\mathsf{R}}_t^2\right] - b_{t+1}^2\rho + \mathbb{E}\left[\widetilde{\Phi}_t(\Lambda_d)^2\right]\left(\mathbb{E}\left[\widetilde{\mathsf{P}}_t^2\right] - \widetilde{a}_t^2\sigma^2\right) + \mathbb{E}\left[\widetilde{\Phi}_t(\Lambda_d)^2\Lambda_d^2\right]\widetilde{a}_t^2\rho \\
&\quad + 2\mathbb{E}\left[\Phi_t(\Lambda_d^2)\widetilde{\Phi}_t(\Lambda_d)\Lambda_d\right]\widetilde{a}_t\widetilde{b}_t\rho \\
&= \left(\frac{(\tau_2^t)^2}{(\gamma_2^t)^2}\mathbb{E}\left[\widetilde{\Phi}_t(\Lambda_d)^2\Lambda_d^2\right] - \frac{(\gamma_1^{t+1})^2}{(\gamma_2^t)^2}\right)\rho\gamma_2^t\left(\gamma_2^t - \frac{1}{\rho}\right) - (\gamma_1^{t+1})^2\rho \\
&\quad + \mathbb{E}\left[\widetilde{\Phi}_t(\Lambda_d)^2\right]\left(\sigma^2\tau_2^t\left(\tau_2^t + \frac{1}{\sigma^2}\right) - (\tau_2^t)^2\sigma^2\right) + \mathbb{E}\left[\widetilde{\Phi}_t(\Lambda_d)^2\Lambda_d^2\right](\tau_2^t)^2\rho \\
&\quad + 2\left(\frac{(\gamma_1^{t+1})^2}{\gamma_2^t\tau_2^t} - \frac{\tau_2^t}{\gamma_2^t}\mathbb{E}\left[\widetilde{\Phi}_t(\Lambda_d)^2\Lambda_d^2\right]\right)\tau_2^t\left(\gamma_2^t - \frac{1}{\rho}\right)\rho \\
&\overset{(4.36h)}{=} \mathbb{E}\left[\widetilde{\Phi}_t(\Lambda_d)^2\Lambda_d^2\right]\left(\frac{(\tau_2^t)^2}{(\gamma_2^t)^2}\rho\gamma_2^t\left(\gamma_2^t - \frac{1}{\rho}\right) + (\tau_2^t)^2\rho - 2\frac{\tau_2^t}{\gamma_2^t}\tau_2^t\left(\gamma_2^t - \frac{1}{\rho}\right)\rho\right) + \mathbb{E}\left[\widetilde{\Phi}_t(\Lambda_d)^2\right]\tau_2^t \qquad \text{(H.25)} \\
&\quad - \left(\frac{(1/v_2^t - \gamma_2^t)^2}{(\gamma_2^t)^2}\rho\gamma_2^t\left(\gamma_2^t - \frac{1}{\rho}\right) + (1/v_2^t - \gamma_2^t)^2\rho - 2\frac{(1/v_2^t - \gamma_2^t)^2}{\gamma_2^t\tau_2^t}\tau_2^t\left(\gamma_2^t - \frac{1}{\rho}\right)\rho\right), \qquad \text{(H.26)}
\end{aligned}
$$

其中，第二个等式使用了（H.19）和（H.20）中的 $\mathbb{E}\left[\widetilde{\mathsf{P}}_t^2\right]$、$\mathbb{E}\left[\widetilde{\mathsf{R}}_t^2\right]$ 的表达式，（H.15）至（H.17）中推导出的 $\widetilde{a}_t$、$\widetilde{b}_t$、b_{t+1} 的表达式，以及（H.8）中 $\mathbb{E}\left[\Phi_t(\Lambda_d^2)^2\right]$、$\mathbb{E}\left[\Phi_t(\Lambda_d^2)\widetilde{\Phi}_t(\Lambda_d)\Lambda_d\right]$ 的表达式。

我们使用（H.11）来展开第一行（H.25）中的各项：

$$
\begin{aligned}
&\mathbb{E}\left[\widetilde{\Phi}_t(\Lambda_d)^2\Lambda_d^2\right]\left(\frac{(\tau_2^t)^2}{(\gamma_2^t)^2}\rho\gamma_2^t\left(\gamma_2^t - \frac{1}{\rho}\right) + (\tau_2^t)^2\rho - 2\frac{\tau_2^t}{\gamma_2^t}\tau_2^t\left(\gamma_2^t - \frac{1}{\rho}\right)\rho\right) + \mathbb{E}\left[\widetilde{\Phi}_t(\Lambda_d)^2\right]\tau_2^t \\
&= \left(\frac{(\gamma_2^t)^2}{(\tau_2^t v_2^t)^2}\mathbb{E}\left[\frac{1}{(\gamma_2^t + \tau_2^t\Lambda_d^2)^2}\right] - \frac{2\gamma_2^t}{(\tau_2^t)^2 v_2^t} + \frac{1}{(\tau_2^t v_2^t)^2}\right)\left(\frac{(\tau_2^t)^2}{(\gamma_2^t)^2}\rho\gamma_2^t\left(\gamma_2^t - \frac{1}{\rho}\right) + (\tau_2^t)^2\rho - 2\frac{\tau_2^t}{\gamma_2^t}\tau_2^t\left(\gamma_2^t - \frac{1}{\rho}\right)\rho\right) \qquad \text{(H.27)} \\
&\quad + \left(\frac{1}{\tau_2^t v_2^t} - \frac{\gamma_2^t}{\tau_2^t(v_2^t)^2}\mathbb{E}\left[\frac{1}{(\gamma_2^t + \tau_2^t\Lambda_d^2)^2}\right]\right)\tau_2^t. \qquad \text{(H.28)}
\end{aligned}
$$

可以验证，所有涉及 $\mathbb{E}\left[\frac{1}{(\gamma_2^t+\tau_2^t\Lambda_d^2)^2}\right]$ 的项均相互抵消。此外，在（H.26）至（H.28）中不涉及 $\mathbb{E}\left[\frac{1}{(\gamma_2^t+\tau_2^t\Lambda_d^2)^2}\right]$ 的剩余各项合起来得到 $\frac{1}{v_2^t} - \gamma_2^t = \gamma_1^{t+1}$，其中等式由（4.36h）得出。因此，我们证明了

$$
\mathbb{E}\left[\mathsf{K}_{t+1}^2\right] = \gamma_1^{t+1} \tag{H.29}
$$

至此，归纳步骤完成。

综上，该结果的证明中，还需检查 (H.1) 在基例 $t = 0$ 中的有效性。

基例。 利用 定理 4.3 对 $t=0$ 以及初始条件 (4.38)，

$$\begin{bmatrix} r^0 & \beta_* \end{bmatrix} \xrightarrow{W_2} \begin{bmatrix} \mathsf{R}_0 & \mathsf{B}_* \end{bmatrix}, \qquad \begin{bmatrix} p^0 & z & \varepsilon \end{bmatrix} \xrightarrow{W_2} \begin{bmatrix} \mathsf{P}_0 & \mathsf{Z} & \mathsf{E} \end{bmatrix}. \tag{H.30}$$

现在来证明

$$\widetilde{a}_0 = \tau_2^0, \qquad \widetilde{b}_0 = \gamma_2^0 - \rho^{-1}, \tag{H.31}$$

我们重复导出(H.15) 和(H.16)的相同计算，但改用

$$\mathbb{E}[\mathsf{P}_0\mathsf{Z}] = \tau_1^0\sigma^2 - 1, \qquad \mathbb{E}[\mathsf{R}_0\mathsf{B}_*] = \gamma_1^0\rho \tag{H.32}$$

由 (H.30) 保证，而非归纳假设。将其用于 (H.17) 和 (H.18) 中的 $t=0$，即可得到

$$b_1 = \gamma_1^1, \qquad a_1 = \tau_1^1 - \sigma^{-2}. \tag{H.33}$$

此外，在 (H.19) 和 (H.20) 中的 $t=0$，用 (H.32) 替代归纳假设，可得

$$\mathbb{E}\left[\widetilde{\mathsf{P}}_0^2\right] = \sigma^2\tau_2^0(\tau_2^0 + \sigma^{-2}), \qquad \mathbb{E}\left[\widetilde{\mathsf{R}}_0^2\right] = \rho\gamma_2^0(\gamma_2^0 - \rho^{-1}). \tag{H.34}$$

在导出 (H.24) 和 (H.29)的计算中，结合上文给出的 (H.31)、(H.33) 和 (H.34) 以及 准备 步骤中的其他观察（这些观察不需要归纳假设），我们得到所需的 $\mathbb{E}\left[\mathsf{J}_1^2\right] = \tau_1^1 - \sigma^{-2}$，$\mathbb{E}\left[\mathsf{K}_1^2\right] = \gamma_1^1$。至此完成了对 $t=0$ 情形的验证(H.1)，从而完成整个证明。

□

I 定理4.6证明

在深入探讨　　证明之前，我们首先根据(4.42)中 Q_z 的定义以及(3.9) 中 σ^2 的定义可知

$$Q_z = \frac{\rho\mathbb{E}\left[\Lambda_d^2\right]}{\delta} = \rho\mathbb{E}\left[\Lambda_n^2\right] = \rho\overline{m}_2 = \rho\overline{\kappa}_2 = \sigma^2. \tag{I.1}$$

因此 $\widehat{Q}_z = \sigma^{-2}$。

为了证明这一结果，我们将论证，在(4.45)中的变量替换下，(4.44) 中变量的不动点方程与副本鞍点方程(4.41)一致。前者可通过从$(v_1^t,\ c_1^t,\ \gamma_2^t,\ \tau_2^t,\ v_2^t,\ c_2^t,\ \gamma_1^{t+1},\ \tau_1^{t+1})$在(4.36)中的递归中移除时间指标得到：

$$v_1 = \mathbb{E}[g'_{x1}(\mathsf{R};\gamma_1)], \qquad c_1 = \mathbb{E}[g'_{z1}(\mathsf{P},\mathsf{Y};\tau_1)], \tag{I.2a}$$

$$\gamma_1 + \gamma_2 = \frac{1}{v_1} = \frac{1}{v_2}, \qquad \tau_1 + \tau_2 = \frac{1}{c_1} = \frac{1}{c_2}, \tag{I.2b}$$

$$v_2 = \mathbb{E}\left[\frac{1}{\tau_2\Lambda_d^2 + \gamma_2}\right], \qquad c_2 = \mathbb{E}\left[\frac{\Lambda_n^2}{\tau_2\Lambda_n^2 + \gamma_2}\right], \tag{I.2c}$$

我们将逐个方程检验(I.2) 与(4.41) 之间的等价性。

由观察结果(I.1)可知，(4.41c) 与(I.2b)中的第一个方程、(4.41d) 与(I.2b)中的第二个方程、(4.41e) 与(I.2c)中的第一个方程，以及(4.41f)与(I.2c) 中的第二个方程是等价的，这一点显而易见。证明的其余部分将用于验证(4.41a) 与(I.2a)中的第一个方程，以及(4.41b) 与(I.2a)中的第二个方程之间的等价性。

首先考虑 (4.41a)。回顾 $\mathcal{Z}_0(b,\ a)$ 在 (4.40)中的定义，我们计算在 $(\sqrt{\widehat{q}_x}\xi,\ \widehat{q}_x)$处取值的对数配分函数的偏导数：

$$\partial_b \log \mathcal{Z}_0\left(\sqrt{\widehat{q}_x}\xi, \widehat{q}_x\right) = \frac{1}{\mathcal{Z}_0\left(\sqrt{\widehat{q}_x}\xi, \widehat{q}_x\right)} \int x P_{\mathsf{B}_*}(x) \exp\left(-\frac{1}{2}\widehat{q}_x x^2 + \sqrt{\widehat{q}_x}\xi x\right) \mathrm{d}x.$$

这使我们能够计算 (4.41a)右侧的期望：

$$
\begin{aligned}
&\mathbb{E}\Big[\mathcal{Z}_0\Big(\sqrt{\widehat{q}_x}\,\Xi,\widehat{q}_x\Big)\Big(\partial_b \log \mathcal{Z}_0\Big(\sqrt{\widehat{q}_x}\,\Xi,\widehat{q}_x\Big)\Big)^2\Big] \\
&= \mathbb{E}\left[\frac{1}{\mathcal{Z}_0\Big(\sqrt{\widehat{q}_x}\,\Xi,\widehat{q}_x\Big)}\left(\int x P_{\mathsf{B}_*}(x)\exp\Big(-\frac{1}{2}\widehat{q}_x x^2+\sqrt{\widehat{q}_x}\,\Xi x\Big)\,\mathrm{d}x\right)^2\right] \\
&= \int \frac{\exp(-\frac{1}{2}\xi^2)}{\sqrt{2\pi}}\frac{\Big(\int x P_{\mathsf{B}_*}(x)\exp\Big(-\frac{1}{2}\widehat{q}_x x^2+\sqrt{\widehat{q}_x}\,\xi x\Big)\,\mathrm{d}x\Big)^2}{\int P_{\mathsf{B}_*}(x)\exp\Big(-\frac{1}{2}\widehat{q}_x x^2+\sqrt{\widehat{q}_x}\,\xi x\Big)\,\mathrm{d}x}\,\mathrm{d}\xi \\
&= \frac{1}{\sqrt{\widehat{q}_x}}\int \frac{\left(\int x P_{\mathsf{B}_*}(x)\cdot\frac{1}{\sqrt{2\pi/\widehat{q}_x}}\exp\left(-\frac{1}{2}\widehat{q}_x\left(x-\frac{\xi}{\sqrt{\widehat{q}_x}}\right)^2\right)\mathrm{d}x\right)^2}{\int P_{\mathsf{B}_*}(x)\cdot\frac{1}{\sqrt{2\pi/\widehat{q}_x}}\exp\left(-\frac{1}{2}\widehat{q}_x\left(x-\frac{\xi}{\sqrt{\widehat{q}_x}}\right)^2\right)\mathrm{d}x}\,\mathrm{d}\xi \\
&= \mathbb{E}\Big[\mathbb{E}[\mathsf{B}_*\,|\,\mathsf{R}]^2\Big], \qquad \frac{\mathsf{R}}{\widehat{q}_x} = \mathsf{B}_* + \frac{1}{\sqrt{\widehat{q}_x}}\overline{\mathsf{K}},\ (\mathsf{B}_*,\overline{\mathsf{K}})\sim P_{\mathsf{B}_*}\otimes\mathcal{N}(0,1),
\end{aligned}
$$

其中，最后一步中我们作了变量替换 $\mathsf{R} = \sqrt{\widehat{q}_x}\,\Xi$。

另一方面，由 (H.12) 可知，(I.2a) 的第一个方程等价于

$$
v_1 = \rho - \mathbb{E}[\mathsf{B}_*\mathbb{E}[\mathsf{B}_*\,|\,\mathsf{R}]]. \tag{I.3}
$$

鉴于变量替换 $v_1 = \rho - q_x$ 以及条件期望的塔式性质，我们得出(I.3) 与 (4.41a)等价。

接下来，考虑 (4.41b)。回忆一下 $\mathcal{Z}_{\text{out}}(y;\xi,v)$ 的定义，我们计算对数配分函数的偏导数：

$$
\begin{aligned}
\partial_\xi \log \mathcal{Z}_{\text{out}}(y;\xi,v) &= \frac{1}{\mathcal{Z}_{\text{out}}(y;\xi,v)}\mathbb{E}\big[\partial_\xi Q\big(y\,\big|\,\sqrt{v}\Omega+\xi\big)\big] \\
&= \frac{1}{\mathcal{Z}_{\text{out}}(y;\xi,v)}\frac{1}{\sqrt{v}}\mathbb{E}\big[\partial_z Q\big(y\,\big|\,\sqrt{v}\Omega+\xi\big)\big] = \frac{1}{\mathcal{Z}_{\text{out}}(y;\xi,v)}\frac{1}{\sqrt{v}}\mathbb{E}\big[\Omega Q\big(y\,\big|\,\sqrt{v}\Omega+\xi\big)\big].
\end{aligned}
$$

使用简写 $\mathfrak{a} = \frac{1}{\sqrt{\widehat{Q}_z+\widehat{q}_z}}$ 和 $\mathfrak{b} = \sqrt{\widehat{Q}_z}\,\frac{\widehat{q}_z}{(\widehat{Q}_z+\widehat{q}_z)}$，我们可以将 (4.41b)右侧的期望写成

$$
\begin{aligned}
&\mathbb{E}\left[\int \mathcal{Z}_{\text{out}}\big(y;\mathfrak{b}\Xi,\mathfrak{a}^2\big)\big(\partial_\xi \log \mathcal{Z}_{\text{out}}\big(y;\mathfrak{b}\Xi,\mathfrak{a}^2\big)\big)^2\,\mathrm{d}y\right] \\
&= \frac{1}{\mathfrak{a}^2}\mathbb{E}\left[\int \frac{\mathbb{E}[\Omega Q(y\,|\,\mathfrak{a}\Omega+\mathfrak{b}\Xi)]^2}{\mathbb{E}[Q(y\,|\,\mathfrak{a}\Omega+\mathfrak{b}\Xi)]}\,\mathrm{d}y\right] \\
&= \frac{1}{\mathfrak{a}^2}\iint p_\Xi(\xi)\frac{\big(\int \omega p_\Omega(\omega)p_{\mathsf{Y}\,|\,\Omega,\Xi}(y\,|\,\omega,\xi)\big)^2\,\mathrm{d}\omega}{\int p_\Omega(\omega)p_{\mathsf{Y}\,|\,\Omega,\Xi}(y\,|\,\omega,\xi)\,\mathrm{d}\omega}\,\mathrm{d}y\,\mathrm{d}\xi \\
&= \frac{1}{\mathfrak{a}^2}\iint p_\Xi(\xi)p_{\mathsf{Y}\,|\,\Xi}(y\,|\,\xi)\left(\int \omega\frac{p_{\Omega,\mathsf{Y}\,|\,\Xi}(\omega,y\,|\,\xi)}{p_{\mathsf{Y}\,|\,\Xi}(y\,|\,\xi)}\,\mathrm{d}\omega\right)^2\mathrm{d}y\,\mathrm{d}\xi \\
&= \frac{1}{\mathfrak{a}^2}\iint p_{\Xi,\mathsf{Y}}(\xi,y)\left(\int \omega p_{\Omega\,|\,\mathsf{Y},\Xi}(\omega\,|\,y,\xi)\,\mathrm{d}\omega\right)^2\mathrm{d}y\,\mathrm{d}\xi \\
&= \frac{1}{\mathfrak{a}^2}\mathbb{E}\Big[\mathbb{E}[\Omega\,|\,\Xi,\mathsf{Y}]^2\Big], \qquad \text{where } (\Xi,\Omega)\sim\mathcal{N}(0,1)\otimes\mathcal{N}(0,1),\ \mathsf{Y}\sim Q(\cdot\,|\,\mathfrak{a}\Omega+\mathfrak{b}\Xi)
\end{aligned}
$$

$$= \frac{1}{\mathfrak{a}^2}\mathbb{E}\Big[\mathbb{E}[(\mathsf{M}-\mathfrak{b}\Xi)\,|\,\Xi,\mathsf{Y}]^2\Big], \qquad \text{where } \mathsf{M}=\mathfrak{a}\Omega+\mathfrak{b}\Xi$$
$$= \frac{1}{\mathfrak{a}^2}\Big(\mathbb{E}\Big[\mathbb{E}[\mathsf{M}\,|\,\Xi,\mathsf{Y}]^2\Big]+\mathfrak{b}^2-2\mathfrak{b}\mathbb{E}[\mathsf{M}\Xi]\Big)$$
$$= \frac{1}{\mathfrak{a}^2}\Big(\mathbb{E}\Big[\mathbb{E}[\mathsf{M}\,|\,\Xi,\mathsf{Y}]^2\Big]-\mathfrak{b}^2\Big),$$

倒数第二行是根据条件期望的塔式性质得到的。

为了将其与(I.2a)中的第二个方程联系起来，我们作如下观察。根据定理 4.4，在不动点处，P 具有分布

$$\mathsf{P}=\Big(\tau_1-\frac{1}{\sigma^2}\Big)\mathsf{Z}+\sqrt{\tau_1-\frac{1}{\sigma^2}}\overline{\mathsf{K}}, \qquad \text{where } (\mathsf{Z},\overline{\mathsf{K}})\sim\mathcal{N}(0,\sigma^2)\otimes\mathcal{N}(0,1).$$

特别地，$\mathsf{P}\sim\mathcal{N}(0,\,(\tau_1-\sigma^{-2})\tau_1\sigma^2)$。由此可以推导出反向信道

$$\mathsf{Z}=\frac{1}{\sqrt{\tau_1}}\Omega+\frac{1}{\tau_1}\mathsf{P}, \qquad \text{where } (\Omega,\mathsf{P})\sim\mathcal{N}(0,1)\otimes\mathcal{N}(0,(\tau_1-\sigma^{-2})\tau_1\sigma^2). \tag{I.4}$$

另一方面，回顾简写记号 $\mathfrak{a}$，$\mathfrak{b}$，利用观察结果 $\widehat{Q}_z=\sigma^{-2}$ 以及变量替换 $\widehat{q}_z=\tau_1-\sigma^{-2}$，我们有

$$\mathsf{M}=\frac{1}{\sqrt{\widehat{Q}_z+\widehat{q}_z}}\Omega+\sqrt{\frac{\widehat{q}_z}{\widehat{Q}_z(\widehat{Q}_z+\widehat{q}_z)}}\Xi=\frac{1}{\sqrt{\tau_1}}\Omega+\sqrt{\sigma^2-\frac{1}{\tau_1}}\Xi. \tag{I.5}$$

将(I.5)与(I.4)进行对比，我们发现

$$(\mathsf{P},\mathsf{Z})\overset{d}{=}\Big(\sqrt{(\tau_1-\sigma^2)\tau_1\sigma^2}\Xi,\mathsf{M}\Big).$$

因此，

$$\begin{aligned}\frac{1}{\mathfrak{a}^2}\Big(\mathbb{E}\Big[\mathbb{E}[\mathsf{M}\,|\,\Xi,\mathsf{Y}]^2\Big]-\mathfrak{b}^2\Big)&=\tau_1\Big(\mathbb{E}\Big[\mathbb{E}\Big[\mathsf{M}\,\Big|\,\sqrt{(\tau_1-\sigma^2)\tau_1\sigma^2}\Xi,\mathsf{Y}\Big]^2\Big]-\Big(\sigma^2-\frac{1}{\tau_1}\Big)\Big)\\&=\tau_1\Big(\mathbb{E}\Big[\mathbb{E}[\mathsf{Z}\,|\,\mathsf{P},\mathsf{Y}]^2\Big]-\Big(\sigma^2-\frac{1}{\tau_1}\Big)\Big).\end{aligned} \tag{I.6}$$

利用上述关系(I.6) 和变量替换 $q_z=\sigma^2-c_1$，$\widehat{q}_z=\tau_1-\sigma^{-2}$我们可以利用条件期望的塔式性质将(4.41b)写成

$$\sigma^2-c_1=\frac{1}{\tau_1}\Big\{\frac{\tau_1-\sigma^{-2}}{\sigma^{-2}}+\tau_1\Big(\mathbb{E}[\mathsf{Z}\mathbb{E}[\mathsf{Z}\,|\,\mathsf{P},\mathsf{Y}]]-\Big(\sigma^2-\frac{1}{\tau_1}\Big)\Big)\Big\}.$$

重新整理后，这恰好是

$$\sigma^2-c_1=\mathbb{E}[\mathsf{Z}\mathbb{E}[\mathsf{Z}\,|\,\mathsf{P},\mathsf{Y}]],$$

根据 (H.13)，这又等价于 (I.2a) 中的第二个方程。

Th是 comp letes t he proo f of the equiva lence b之间 (I.2) and (4.41) and th因此，定理 4.6。

J 示例和数值实验的详细内容

J.1 复制鞍点方程

对于固定的 $\gamma_1\geq 0$，$\tau_1\geq\sigma^{-2}$，令

$$(\mathsf{B}_*,\overline{\mathsf{J}})\sim P_{\mathsf{B}_*}\otimes\mathcal{N}(0,1), \qquad \mathsf{R}=\gamma_1\mathsf{B}_*+\sqrt{\gamma_1}\,\overline{\mathsf{J}}, \tag{J.1a}$$

$$(\mathsf{Z},\overline{\mathsf{K}})\sim\mathcal{N}(0,\sigma^2)\otimes\mathcal{N}(0,1),\qquad \mathsf{P}=(\tau_1-\sigma^{-2})\mathsf{Z}+\sqrt{\tau_1-\sigma^{-2}}\,\overline{\mathsf{K}},\qquad \mathsf{Y}\sim Q(\cdot\,|\,\mathsf{Z}).\tag{J.1b}$$

对于上述概率分布，定义函数

$$\mathcal{E}_x(\gamma_1)=\mathbb{E}[\mathsf{B}_*\mathbb{E}[\mathsf{B}_*\,|\,\mathsf{R}]],\qquad\qquad \mathcal{E}_z(\tau_1)=\mathbb{E}[\mathsf{Z}\mathbb{E}[\mathsf{Z}\,|\,\mathsf{P},\mathsf{Y}]].\tag{J.2}$$

回顾 Λ 在 (6.2) 中的构造，以及 Λ_d 在 (3.8)中的定义，定义函数

$$\mathcal{S}(\tau_2,\gamma_2)\coloneqq\mathbb{E}\big[(\tau_2\Lambda_d^2+\gamma_2)^{-1}\big]=\begin{cases}\delta\mathbb{E}\left[\left(\tau_2\widetilde{\Lambda}^2+\gamma_2\right)^{-1}\right]+(1-\delta)\gamma_2^{-1}, & \delta\le 1\\ \mathbb{E}\left[\left(\tau_2\delta\widetilde{\Lambda}^2+\gamma_2\right)^{-1}\right], & \delta>1\end{cases}.\tag{J.3}$$

命题 J.1。 方程 (v_1,v_2,γ_1) 的 (4.36) 不动点满足 $v_1=v_2=v$，并且 $(v,\gamma_1)\in[0,\rho]\times[0,\infty)$ 满足以下方程组

$$v=\rho-\mathcal{E}_x(\gamma_1),\qquad\qquad \gamma_1=\frac{\delta}{v}\big[1-\tau_1(v,\gamma_1)\big(\sigma^2-\mathcal{E}_z(\tau_1(v,\gamma_1))\big)\big],\tag{J.4}$$

其中 $\tau_1(v,\gamma_1)\ge\sigma^{-2}$ 是以下方程的唯一解

$$v=\mathcal{S}\left(\frac{\tau_1\gamma_1 v}{\delta-\gamma_1 v},\frac{1}{v}-\gamma_1\right).\tag{J.5}$$

此外，对于相位恢复这一特殊情形 $q(z,\varepsilon)=|z|$，若采用高斯分布先验 $P_{\mathsf{B}_*}=\mathcal{N}(0,\rho)$，$v\in[0,\rho]$ 可更明确地表示为方程 $v=\mathcal{F}(v)$的解，其中函数 $\mathcal{F}$ 定义为

$$\mathcal{F}(v)=\frac{\delta\rho v}{\rho-v}\big(1-\tau_1(v)\big(\sigma^2-\mathbb{E}\big[\mathsf{Y}^2\tanh(\mathsf{P}\mathsf{Y})^2\big]\big)\big),\tag{J.6}$$

其中

$$(\mathsf{P},\overline{\mathsf{K}})\sim\mathcal{N}(0,(\tau_1(v)-\sigma^{-2})\tau_1(v)\sigma^2)\otimes\mathcal{N}(0,1),\qquad \mathsf{Y}=\left|\frac{1}{\tau_1(v)}\mathsf{P}+\frac{1}{\sqrt{\tau_1(v)}}\overline{\mathsf{K}}\right|,\tag{J.7}$$

并且 $\tau_1(v)\ge\sigma^{-2}$ *(*对解 $\tau_1(v,\gamma_1)$ 到 (J.5)*)* 的记号稍作滥用）是以下方程的唯一解

$$v=\mathcal{S}\left(\frac{(\rho-v)\tau_1}{\rho\delta-(\rho-v)},\frac{1}{\rho}\right).$$

证明。 副本鞍点方程 (I.2) 如下：

$$v=\rho-\mathbb{E}[\mathsf{B}_*\mathbb{E}[\mathsf{B}_*\,|\,\mathsf{R}]],\qquad\qquad c=\sigma^2-\mathbb{E}[\mathsf{Z}\mathbb{E}[\mathsf{Z}\,|\,\mathsf{P},\mathsf{Y}]],\tag{J.8a}$$

$$\frac{1}{v}=\gamma_1+\gamma_2,\qquad\qquad \frac{1}{c}=\tau_1+\tau_2,\tag{J.8b}$$

$$v=\mathbb{E}\left[\frac{1}{\tau_2\Lambda_d^2+\gamma_2}\right],\qquad\qquad c=\mathbb{E}\left[\frac{\Lambda_n^2}{\tau_2\Lambda_n^2+\gamma_2}\right],\tag{J.8c}$$

其中我们使用如下记号 $v:=v_1=v_2$，$c:=c_1=c_2$，第一行中的随机变量的联合分布由(J.1)给出。 (J.8) 应在以下域内求解

$$(\gamma_1,\tau_1,\gamma_2,\tau_2,v,c)\in[0,\infty)\times[\sigma^{-2},\infty)\times[\rho^{-1},\infty)\times[0,\infty)\times[0,\rho]\times[0,\sigma^2].$$

注意到， c可以表示为其他变量的函数：

$$c=\mathbb{E}\left[\frac{\Lambda_n^2}{\tau_2\Lambda_n^2+\gamma_2}\right]=\frac{1}{\delta}\mathbb{E}\left[\frac{\Lambda_d^2}{\tau_2\Lambda_d^2+\gamma_2}\right]=\frac{1}{\delta\tau_2}(1-\gamma_2 v).\tag{J.9}$$

利用 $\gamma_2 = 1/v - \gamma_1$，$\tau_2 = 1/c - \tau_1$ 根据 (J.8b) 消去 γ_2，τ_2 代入 (J.9)并整理可得

$$c = \frac{1}{\tau_1}\Big(1 - \frac{\gamma_1 v}{\delta}\Big). \tag{J.10}$$

将(J.10) 右侧代入(J.8b) 的第二个方程中，关于 c，可得

$$\tau_2 = \frac{1}{c} - \tau_1 = \frac{\tau_1\gamma_1 v}{\delta - \gamma_1 v}. \tag{J.11}$$

类似地，将(J.8a) 的第二个方程代入并整理后，可得

$$\gamma_1 = \frac{\delta}{v}\big[1 - \tau_1(\sigma^2 - \mathcal{E}_z(\tau_1))\big], \tag{J.12}$$

其中 $\mathcal{E}_z(\tau_1)$ 定义见(J.2)。再结合 $\mathcal{E}_x(\gamma_1)$ （见(J.2) ）以及 $\mathcal{S}(\tau_2,\ \gamma_2)$ （见(J.3)），将(J.11)和(J.12)汇总，即可得到所需的方程组(J.4) 和(J.5)。

对于具有高斯先验 $P_{\mathsf{B}_*} = \mathcal{N}(0,\ \rho)$ 的相位恢复模型 $q(z,\ \varepsilon) = |z|$，有

$$\mathbb{E}[\mathsf{B}_* \,|\, \mathsf{R} = r] = \frac{\rho r}{\gamma_1\rho + 1}, \tag{J.13}$$

$$\mathbb{E}[\mathsf{Z} \,|\, \mathsf{P} = p, \mathsf{Y} = y] = \frac{\int \varphi_\sigma(z)\varphi_{\sqrt{\tau_1 - \sigma^{-2}}}(p - (\tau_1 - \sigma^{-2})z)\delta(y - |z|)z\,\mathrm{d}z}{\int \varphi_\sigma(z')\varphi_{\sqrt{\tau_1 - \sigma^{-2}}}(p - (\tau_1 - \sigma^{-2})z')\delta(y - |z'|)\,\mathrm{d}z'} = y\tanh(py), \tag{J.14}$$

因此，根据条件期望的塔式性质，

$$\mathcal{E}_x(\gamma_1) = \frac{\mathbb{E}[\mathsf{R}^2]}{(\gamma_1\rho + 1)^2} = \frac{\rho^2\gamma_1}{\gamma_1\rho + 1}, \qquad \mathcal{E}_z(\tau_1) = \mathbb{E}\big[\mathsf{Y}^2\tanh(\mathsf{PY})^2\big], \tag{J.15}$$

第二个期望是针对(J.7)中的联合分布求取的，(J.1b)在相位恢复模型中的特例与此等价。利用 $\mathcal{E}_x(\gamma_1)$ 在(J.15)中的表达式，我们可以根据 v，从(J.4) 的第一个方程求出 γ_1 关于 v 的表达式，得到

$$\gamma_1 = \frac{\rho - v}{\rho v}. \tag{J.16}$$

由此可得到更简洁的不动点方程 $v = \mathcal{F}(v)$ 其中 $\mathcal{F}$ 定义为 (J.6)。 □

J.2 合成数据

相位恢复。 在第 6.2 节的设定下，容易验证，函数 g 在 (4.16) 中由下式给出

$$\overline{g}(y) = y^2 - 1, \tag{J.17}$$

其中 γ 按照(4.2)定义。因此，相应的预处理函数（见 假设 (A5)）为$\mathcal{T}(y) = 1 - \frac{\gamma}{y^2 - 1 + \gamma}$，而来自 [61] 的预处理函数则特化为 $\mathcal{T}(y) = 1 - \frac{1}{y^2}$。弱恢复条件(4.18) 变为

$$\delta > \frac{\overline{m}_2'^2}{\overline{m}_4'}\frac{3}{2}.$$

Bayes-GVAMP 中使用的向量去噪器 g_{x1}、g_{z1} （见 (4.32) 和 (4.33)）具有显式表达式，这些表达式已在 (J.13)和(J.14)中导出。

泊松分布 reg式。 在此情形下，初等计算表明，g 在 (4.16) 中取值为 $\quad -$

$$\overline{g}(y) = \frac{2}{3}\left(\frac{1}{2} + y\right) - 1.$$

假设 (A5) 中的预处理函数假设 (A5)和 [61] 随后即可据此进行评估。临界条件(4.18) 变为

$$\delta > \frac{\overline{m}_2'^2}{\overline{m}_4'}\frac{7}{4}.$$

对于 Bayes-GVAMP，尽管去噪器 g_{x1} 仍由简单的高斯最小均方误差估计器 (J.13)给出，g_{z1} 但 (4.33) 中的去噪器及其二阶矩（(4.34)所需）更难计算：

$$g_{z1}(p, y; \tau_1) = \mathbb{E}[\mathsf{Z} \,|\, \mathsf{P} = p, \mathsf{Y} = y], \qquad \mathbb{E}\big[g_{z1}(\mathsf{P}, \mathsf{Y}; \tau_1)^2\big] = \mathbb{E}\Big[\mathbb{E}[\mathsf{Z} \,|\, \mathsf{P}, \mathsf{Y}]^2\Big], \tag{J.18}$$

其中

$$(\mathsf{Z}, \overline{\mathsf{K}}) \sim \mathcal{N}(0, \sigma^2) \otimes \mathcal{N}(0, 1), \qquad \mathsf{P} = (\tau_1 - \sigma^{-2})\mathsf{Z} + \sqrt{\tau_1 - \sigma^{-2}}\,\overline{\mathsf{K}}, \qquad \mathsf{Y} \sim \mathrm{Pois}(\mathsf{Z}^2).$$

在上述概率分布下，(J.18)可以等价地写成

$$g_{z1}(p, y; \tau_1) = \frac{\mathbb{E}\big[\mathsf{G}(p; \tau_1)^{2y+1}\big]}{\mathbb{E}[\mathsf{G}(p; \tau_1)^{2y}]}, \qquad \mathbb{E}\big[g_{z1}(\mathsf{P}, \mathsf{Y}; \tau_1)^2\big] = \mathbb{E}\left[\left(\frac{\mathbb{E}\big[\mathsf{G}(\mathsf{P}; \tau_1)^{2\mathsf{Y}+1}\big]}{\mathbb{E}[\mathsf{G}(\mathsf{P}; \tau_1)^{2\mathsf{Y}}]}\right)^2\right], \tag{J.19}$$

其中随机变量 $\mathsf{G}(p; \tau_1)$ 服从如下概率分布

$$\mathsf{G}(p; \tau_1) \sim \mathcal{N}\left(\frac{p}{\tau_1 + 2}, \frac{1}{\tau_1 + 2}\right).$$

评估 (J.19) 存在两个难点： g_{z1} 涉及非中心高斯分布的各阶矩，其数值在 $y \in \mathbb{Z}_{\geq 0}$取中等值时就可能迅速达到机器精度； g_{z1}的二阶矩涉及对 3 个变量 Z、K、Y的嵌套期望。为克服这一困难，我们利用以下观察。

命题 J.2。 固定 $\mu, a, b \in \mathbb{R}$ 和 $\sigma > 0$。令 $\mathsf{G} \sim \mathcal{N}(\mu, \sigma^2)$ 并对任意 $k \geq 0$记作，

$$R_{k+1}(\mu, \sigma^2) = \frac{\mathbb{E}\big[(a + b\mathsf{G})^{k+1}\big]}{\mathbb{E}[(a + b\mathsf{G})^k]}.$$

于是有 $R_1(\mu, \sigma^2) = a + b\mu$；对于 $k \in \mathbb{Z}_{\geq 1}$，

$$R_{k+1}(\mu, \sigma^2) = a + b\left(\mu + \sigma^2 k b \frac{1}{R_k(\mu, \sigma^2)}\right).$$

The p证明如下 这可直接由 Stein 引理应用于 R_{k+1}的分子得到；细节从略 s.

定理 J.2 使我们能够 g_{z1} 递归地（对 y）进行求值，而无需分别计算其分子和分母。对于 g_{z1}的二阶矩，我们进一步将 定理 J.2 与2维 Gauss– Hermite 求积公式 [86, 第 4.6 节1] 相结合，并采用以下近似方案（其中N取足够大， Y）：

$$\mathbb{E}\big[g_{z1}(\mathsf{P}, \mathsf{Y}; \tau_1)^2\big] \approx \sum_{z=1}^{N}\sum_{k=1}^{N} w_N(x_z) w_N(x_k) \sum_{y=0}^{Y} \frac{(\sigma x_z)^{2y} e^{-(\sigma x_z)^2}}{y!} R_{2y+1}\left(\frac{(\tau_1 - \sigma^{-2})\sigma x_z + \sqrt{\tau_1 - \sigma^{-2}}\, x_k}{\tau_1 + 2}, \frac{1}{\tau_1 + 2}\right)^2,$$

其中权重 $w_N(\cdot)$ 定义为

$$w_N(x) = \frac{2^{N-1}N!}{N^2 H_{N-1}(x)^2},$$

其中 H_N 表示次数为N 的物理学家厄米多项式，且$(x_i/\sqrt{2})_{1\le i\le N}$ 表示其根。在数值计算中，我们发现将求积的阶数设为 $N = 50$ ，并将Y 关于 y 的上界设为比泊松分布的均值大10 个标准差，既高效又准确。该近似方案是确定性的，避免了蒙特卡罗方法的随机误差；我们发现，即使样本量相当大，这种误差仍不可忽略。类似的近似方案结合定理 J.2 中的递归和求积，用于计算 GAMP 及其状态演化中的类似量。

J.3 真实数据

编码衍射图样。第 6.2 节中使用的实数编码衍射图样的构造方式与其复数对应版本类似，见 [17, 第2节。2]。对于 $\delta \in \mathbb{Z}_{\ge 1}$， $X \in \mathbb{R}^{d\delta\times d}$ 定义为

$$X = \begin{bmatrix} C_d^\top D_1 & \cdots & C_d^\top D_\delta \end{bmatrix}^\top,$$

其中， $C_d \in \mathbb{O}(d)$ 是离散余弦变换（DCT）矩阵，每个 $D_i \in \mathbb{R}^{d\times d}$ (其中 $1 \le i \le \delta$) 都是一个独立的对角矩阵，其对角元素独立同分布，服从某个随机变量 D 的概率分布。对于二元和三元编码衍射图样，law(D) 取为

$$\mathrm{law}(\mathsf{D}) = \mathrm{Bernoulli}(1/2) \qquad \text{and} \qquad \mathrm{law}(\mathsf{D}) = \frac{1}{2}\mathrm{Rademacher} + \frac{1}{2}\delta_0$$

分别。回顾 (3.8)，根据 C_d的正交性，我们有

$$\Lambda_d^2 \overset{d}{=} \sum_{i=1}^{\delta} \mathsf{D}_i^2,$$

其中，$(\mathsf{D}_i)_{1\le i\le\delta}$ 是 D 的独立同分布副本。不过请注意， X 在概率分布上既不是左正交不变的，也不是右正交不变的。

GTEx。 第 6.2 节中使用的设计矩阵 是通过依次应用以下标准预处理步骤获得的 [27] ， 处理的对象为 GTEx 数据集“皮肤－日晒部位（小腿）”和“肌肉－骨骼肌” [59]。

1. 删除全为零的行；
2. 依次纳入上一步中的每一行，只有当该行与所有已有行的重叠至多为 0.3 时，才将其纳入矩阵；
3. 对上一步收集的所有行进行标准化。

得到的矩阵完全是确定性的，尤其不满足假设（A2）。第6.2 节中的实验唯一涉及的随机性来自 $\beta_* \sim \mathcal{N}(0_d,\ I_d)$ 的生成。

K 辅助引理

定理 K.1 至 K.4 汇集了关于从[44, 附录 E 和 F.1] 以及 [57, 附录 G.1]中收集的经验测度的 Wasserstein 收敛的若干事实。

命题 K.1. 固定 $p \geq 1$、$k \geq 1$、$\ell \geq 0$。设 $E \in \mathbb{R}^{n\times \ell}$ 是一个满足 $E \xrightarrow{W_p} \mathsf{E}^\top$ 的确定性矩阵。若 $V \in \mathbb{R}^{n\times k}$ 的行独立同分布，且服从 $\mathsf{V} \in \mathbb{R}^k$ 的概率分布，并具有有限的、截至 p 阶的混合矩，则 $\begin{bmatrix} V & E \end{bmatrix} \xrightarrow{W_p} \begin{bmatrix} \mathsf{V}^\top & \mathsf{E}^\top \end{bmatrix}$ 其中V 与E独立。

命题 K.2. 固定 $p,\ q \geq 1$ 和 $k,\ \ell \geq 1$。假设 $V \in \mathbb{R}^{n\times k}$ 满足 $V \xrightarrow{W_{p+q}} \mathsf{V}^\top$。令 $g \in \mathrm{PG}^q_{k\to\ell}$(见 定义 3.2) 连续。那么 $g(V) \xrightarrow{W_p} g(\mathsf{V})^\top$，其中 g 逐行作用于 V。

命题 K.3. 固定 $p \geq 1, k \geq 1$。假设 $V \in \mathbb{R}^{n\times k}$ 满足 $V \xrightarrow{W_p} \mathsf{V}^\top$。令 $f \in \mathrm{PG}^p_{k\to 1}$ 除在概率为 0 的集合上外，关于V的概率分布处处连续。那么几乎必然，

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n f(V_{i,:}) = \mathbb{E}[f(\mathsf{V})].$$

命题 K.4. 固定 $p, k, \ell \geq 1$。假设 $V \in \mathbb{R}^{n\times k}$, $W \in \mathbb{R}^{n\times \ell}$ 和 $M_n, M \in \mathbb{R}^{k\times \ell}$ 几乎必然逐元素满足 $V \xrightarrow{W_p} \mathsf{V}^\top$, $W \xrightarrow{W_p} 0_{1\times\ell}$ 和 $M_n \to M$。那么 $VM_n + W \xrightarrow{W_p} \mathsf{V}^\top M$。

定理 K.5 是对 [44, 命题 F.2] 从 $m = 1$ 到 $m \geq 1$ 固定情形的推广。

命题 K.5. 固定 $p \geq 1$、k、$\ell \geq 0$ 和 $m \geq 1$。令 $O \sim \mathrm{Haar}(\mathbb{O}(n-\ell))$。令 $E \in \mathbb{R}^{n\times k}$、$V \in \mathbb{R}^{(n-\ell)\times m}$、$\Pi \in \mathbb{R}^{n\times(n-\ell)}$ 为满足以下条件的确定性矩阵：$E \xrightarrow{W_p} \mathsf{E}^\top$逐项满足 $n^{-1}V^\top V \to \Sigma$，且$\Pi$ 具有正交归一列。则 $\begin{bmatrix} \Pi OV & E \end{bmatrix} \xrightarrow{W_p} \begin{bmatrix} \mathsf{Z}^\top & \mathsf{E}^\top \end{bmatrix}$，其中 $\mathsf{Z} \sim \mathcal{N}(0_m, \Sigma)$ 与 E独立。

令 $j \geq 0$ 为固定值，d 为一个与 n共同增长的维数。固定常数 $j \geq 0$。令 $F \in \mathbb{R}^{d\times j}$ 为满足 $F \xrightarrow{W_p} \mathsf{F}^\top$的确定性矩阵。令 $U \in \mathbb{R}^{d\times m}$ 为 ΠOV的前 d 行（若 $d \leq n$），或者由 ΠOV与 $d-n$ 个服从 $\mathcal{N}(0_m, \Sigma)$ 分布的独立同分布行拼接而成（若 $d > n$）。则 $\begin{bmatrix} U & F \end{bmatrix} \xrightarrow{W_p} \begin{bmatrix} \mathsf{U}^\top & \mathsf{F}^\top \end{bmatrix}$，其中 $\mathsf{U} \sim \mathcal{N}(0_m, \Sigma)$ 与 F独立。

证明。记作 $n' = n-\ell$，并假设它足够大。注意，对 V 左乘 O 会使 V 的左奇异子空间均匀随机化，同时保留其奇异值。将 V 写成其奇异值分解的形式，我们有 $V = L\begin{bmatrix} \Lambda \\ 0_{(n'-m)\times m} \end{bmatrix} R^\top$。我们可以将 OV 的概率分布表示为

$$OV \overset{d}{=} O\begin{bmatrix} \Lambda \\ 0_{(n'-m)\times m} \end{bmatrix} R^\top \overset{d}{=} O\begin{bmatrix} R & 0_{m\times(n'-m)} \\ 0_{(n'-m)\times m} & I_{n'-m} \end{bmatrix}\begin{bmatrix} \Lambda \\ 0_{(n'-m)\times m} \end{bmatrix} R^\top = O'R\Lambda R^\top = O'(V^\top V)^{1/2},$$

其中 O' 包含 O 的前 m 列。我们使用部分 Haar 矩阵概率分布的众所周知的表示（例如参见 [99, 第 I-D 节]）：$O' \overset{d}{=} H(H^\top H)^{-1/2}$，其中 $H \in \mathbb{R}^{n'\times m}$ 由独立同分布的标准高斯条目组成。由于 Π 具有正交归一列，$\Pi^\top G \overset{d}{=} H$，其中 $G \in \mathbb{R}^{n\times m}$ 由独立同分布的标准高斯条目组成。结合以上两点，我们得到

$$\begin{aligned} \Pi OV &\overset{d}{=} \Pi\Pi^\top G(G^\top \Pi\Pi^\top G)^{-1/2}(V^\top V)^{1/2} \\ &= G(G^\top \Pi\Pi^\top G)^{-1/2}(V^\top V)^{1/2} - PG(G^\top \Pi\Pi^\top G)^{-1/2}(V^\top V)^{1/2}, \end{aligned} \tag{K.1}$$

其中 $P = I_n - \Pi\Pi^\top$。根据 K.1 定理，

$$\begin{bmatrix} G & E \end{bmatrix} \xrightarrow{W_p} \begin{bmatrix} \mathsf{G}^\top & \mathsf{E}^\top \end{bmatrix}, \tag{K.2}$$

其中 $\mathsf{G} \sim \mathcal{N}(0_m, I_m)$ 与 E 独立。由于 $\Pi^\top G$ 具有独立同分布的 $\mathcal{N}(0,1)$ 元素，我们有 $n^{-1}G^\top\Pi\Pi^\top G \to I_m$逐项，几乎必然。因此，回顾假设 $n^{-1}V^\top V \to \Sigma$，我们有

$$(G^\top \Pi\Pi^\top G)^{-1/2}(V^\top V)^{1/2} \to \Sigma^{1/2} \tag{K.3}$$

逐项，几乎必然。

接下来，我们证明

$$PG \xrightarrow{W_p} 0_m^\top. \tag{K.4}$$

PG 的概率分布可写为 $PG \overset{d}{=} \sum_{j=1}^{\ell} \pi_j g_j^\top$，其中 π_1，$\cdots$，$\pi_m \in \mathbb{R}^n$ 构成 P 的列空间的一组标准正交基，而 g_1，$\cdots$，g_m 是独立同分布的 $\mathcal{N}(0_m,\ I_m)$ 向量。将每个 π_j 写为 $\begin{bmatrix}\pi_{1,j} & \cdots & \pi_{n,j}\end{bmatrix}^\top$。对于任意固定的 $p \geq 1$,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left\|\sum_{j=1}^{\ell}\pi_{i,j}g_j^\top\right\|_2^p &\leq \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{\ell}\|\pi_{i,j}g_j^\top\|_2\right)^p \leq \frac{1}{n}\sum_{i=1}^{n}\left[\ell^{1-1/p}\left(\sum_{j=1}^{\ell}\|\pi_{i,j}g_j^\top\|_2^p\right)^{1/p}\right]^p \\ &= \frac{\ell^{p-1}}{n}\sum_{j=1}^{\ell}\sum_{i=1}^{n}|\pi_{i,j}|^p\|g_j\|_2^p = \ell^{p-1}\sum_{j=1}^{\ell}\|g_j\|_2^p\frac{\|\pi_j\|_p^p}{n},\end{aligned}$$

其中我们使用了 ℓ_2 范数的次可加性以及不等式 $\|x\|_p \leq \|x\|_r \leq d^{1/r-1/p}\|x\|_p$，该不等式对任意向量 $x \in \mathbb{R}^d$ 和 $0 \leq r \leq p$成立。若 $p \geq 2$，则 $n^{-1}\|\pi_j\|_p^p \leq n^{-1}\|\pi_j\|_2^p = n^{-1}$。若 $1 \leq p < 2$，则 $n^{-1}\|\pi_j\|_p^p \leq n^{-1}\big(n^{1/p-1/2}\|\pi_j\|_2\big)^p = n^{-p/2}$。两种情况下，均有 $n^{-1}\|\pi_j\|_p^p \to 0$。因此，几乎必然，

$$\frac{1}{n}\sum_{i=1}^{n}\|(PG)_{i,:}\|_2^p \to 0. \tag{K.5}$$

此外，对于任意有界利普希茨 $f\colon\ \mathbb{R}^m \to \mathbb{R}$，由 (K.5)，几乎必然有：

$$\left|\frac{1}{n}\sum_{i=1}^{n}\big[f((PG)_{i,:}) - f(0_m^\top)\big]\right| \leq \frac{1}{n}\sum_{i=1}^{n}C\big\|(PG)_{i,:} - 0_m^\top\big\|_2 \to 0. \tag{K.6}$$

综合 (K.5) 和 (K.6)，几乎必然有：

$$\frac{1}{n}\sum_{i=1}^{n}f((PG)_{i,:}) \to f(0_m^\top),$$

对于 $f(v) = \|v\|_2^p$ 以及任意有界利普希茨 f，因此 (K.4) 成立。结合 (K.1) 至 (K.4) 并应用 K.4 定理，得到关于 $\begin{bmatrix}\Pi OV & E\end{bmatrix}$的第一个收敛结果：

$$\begin{bmatrix}\Pi OV & E\end{bmatrix} \overset{d}{=} \begin{bmatrix}G & E\end{bmatrix}\begin{bmatrix}(G^\top\Pi\Pi^\top G)^{-1/2}(V^\top V)^{1/2} & 0_{m\times k}\\ 0_{k\times m} & I_k\end{bmatrix} - \begin{bmatrix}PG(G^\top\Pi\Pi^\top G)^{-1/2}(V^\top V)^{1/2} & 0_{n\times k}\end{bmatrix} \xrightarrow{W_p} \begin{bmatrix}\mathsf{Z}^\top & \mathsf{E}^\top\end{bmatrix},$$

其中 $\mathsf{Z} \sim \mathcal{N}(0_m,\ \Sigma)$ 与 E 相互独立。

接下来考虑 U。按如下方式构造 $\widetilde{G}$，W_1，$W_2 \in \mathbb{R}^{d\times m}$：

$$\widetilde{G} = \begin{cases}G_{1:d,:}, & d \leq n\\ \begin{bmatrix}G\\ G'\end{bmatrix}, & d > n\end{cases}, \qquad W_1 = \begin{cases}\big[PG(G^\top\Pi\Pi^\top G)^{-1/2}(V^\top V)^{1/2}\big]_{1:d,:}, & d \leq n\\ \begin{bmatrix}PG(G^\top\Pi\Pi^\top G)^{-1/2}(V^\top V)^{1/2}\\ 0_{(d-n)\times m}\end{bmatrix}, & d > n\end{cases},$$

$$W_2 = \begin{cases}0_{d\times m}, & d \leq n\\ \begin{bmatrix}0_{n\times m}\\ G'\big[(G^\top\Pi\Pi^\top G)^{-1/2}(V^\top V)^{1/2} - \Sigma\big]\end{bmatrix}, & d > n\end{cases},$$

其中，$G' \in \mathbb{R}^{(d-n)\times m}$（如果 $d > n$）包含元素服从 i.i.d. $\mathcal{N}(0,\ 1)$ 的独立同分布随机变量，并且与其他所有变量相互独立。根据构造，容易验证

$$U = \widetilde{G}(G^\top \Pi \Pi^\top G)^{-1/2}(V^\top V)^{1/2} - W_1 - W_2.$$

再次由 K.1 定理，

$$\begin{bmatrix} \widetilde{G} & F \end{bmatrix} \xrightarrow{W_p} \begin{bmatrix} \widetilde{\mathsf{G}}^\top & \mathsf{F}^\top \end{bmatrix}, \tag{K.7}$$

其中 $\widetilde{\mathsf{G}} \sim \mathcal{N}(0_m,\ I_m)$ 与 F 独立。与之前相同的论证表明 $W_1 \xrightarrow{W_p} 0_m^\top$。此外，$W_2 \xrightarrow{W_p} 0_m^\top$。为验证这一点，我们只需考虑 $d > n$ 的情形。几乎必然，

$$\frac{1}{d}\sum_{i=1}^{d-n}\left\| G'_{i,:}\left[(G^\top \Pi \Pi^\top G)^{-1/2}(V^\top V)^{1/2} - \Sigma\right]\right\|_2^p \le \left\|(G^\top \Pi \Pi^\top G)^{-1/2}(V^\top V)^{1/2} - \Sigma\right\|_2 \frac{1}{d}\sum_{i=1}^{d-n}\|G'_{i,:}\|_2^p \to 0.$$

利用这一点，可以验证 (3.3) 对任意有界利普希茨 f成立，从而证明 W_p-收敛W_2；论证与 (K.6) 类似。结合这些结果与 (K.7) 并应用 K.4 定理，即可得到 $\begin{bmatrix} U & F \end{bmatrix}$的收敛结果。

□

以下结果刻画了在满足某个线性系统条件下 Haar 矩阵的分布。

命题 K.6 ([88, 引理 4]). 固定 $k \ge 1$。令 $X,\ Y \in \mathbb{R}^{n\times k}$ 为秩为 k 的确定性矩阵，满足 $QY = X$，其中某个 $Q \in \mathbb{O}(n)$。若 $O \in \mathrm{Haar}(\mathbb{O}(n))$，则

$$O \,|\, \{OY = X\} \overset{d}{=} X(X^\top X)^{-1}Y^\top + \Pi_{X^\perp}\widetilde{O}\Pi_{Y^\perp}^\top = X(Y^\top Y)^{-1}Y^\top + \Pi_{X^\perp}\widetilde{O}\Pi_{Y^\perp}^\top,$$

其中 $\widetilde{O} \sim \mathrm{Haar}(\mathbb{O}(n-k))$ 独立于其余一切，且 $\Pi_{X^\perp}$、$\Pi_{Y^\perp} \in \mathbb{R}^{n\times(n-k)}$ 分别为张成 张成空间 $\left\{\{X_{:,i}\}_{i=1}^k\right\}^\perp$、张成空间$\left\{\{Y_{:,i}\}_{i=1}^k\right\}^\perp$ 的（任意）具有正交归一列的矩阵。