芝加哥大学

非线性随机矩阵及其在平方和层级中的应用

提交物理科学部全体教师审阅的学位论文，旨在申请哲学博士学位

计算机科学系

BY

Goutham Rajendran

伊利诺伊州芝加哥，
2022年8月

arXiv:2302.04462v1 [cs.CC] 9 Feb 2023

版权所有 © 2022 Goutham Rajendran
保留所有权利

谨以此文纪念我的父亲 *A. S. Rajendran* 博士（*1958–2017*）。他让我懂得教育的价值和追求知识的重要性，并塑造了今天的我。

目录

插图目录 vii

致谢 . viii

摘要 ACT xii

1 引言 性. 1
1.1 认证问题.2 1.2 平均情况分析.4 1.3 本文主题：随机矩阵.5 1.4 平方和层级.7 1.4.1 相关算法技术.9 1.5 针对平方和层级的下界.10 1.6 主要结果概要.11 1.6.1 通过矩阵埃夫龙-斯坦不等式实现非线性矩阵集中 . . . . . 11 1.6.2 平方和下界.12 1.7 未纳入本文的工作.14 1.7.1 稀疏独立集的 SoS 下界.14 1.7.2 因果推断.15 1.7.3 潜变量建模.15 1.7.4 因果表征学习.16 1.8 论文结构.17

2 非线性矩阵集中.18
2.1 引言.18 2.2 预备知识.31 2.3 Rademacher 随机变量的基本框架.33 2.4 应用.37 2.4.1 一个简单的张量网络.38 2.4.2 图矩阵.40 2.5 为什么将 ¨ıve 直接应用于一般乘积分布可能会失败 . 49 2.6 一般递归框架.51 2.7 [188] 的推广及引理 2.6.8 的证明.55 2.7.1 通过显式内核推广 [188].55 2.7.2 引理 2.6.8 的证明.61 2.7.3 界定 $\mathbf{\Delta}^{k}$, $_{a\,b}$，以及$\mathbf{V}_{k,\,a,\,b}$.65 2.7.4 界定 $\mathbf{\Delta}^{k}$, $_{a\,b}$，以及 $\mathbf{\Delta}^{k}$, $_{a\,b}$，.69 2.8 应用：稀疏图矩阵.75 2.8.1 简单图矩阵上的范数界.82

3 平方和层级3.1 平方和层级 ..... 88 3.1.1 多项式优化与凸松弛 ..... 88 3.1.2 平方和松弛88 ..... 90 3.2 假设检验 ..... 96 3.2.1 低次似然比 ..... 97 3.3 伪校准 ..... 99 3.3.1 证明 SoS 下界的策略 ..... 103 3.3.2 与低次区分器的联系 ..... 104 3.4 为什么选择平方和? 105

4 平方和下界的主要结果 107
4.1 Sherrington-Kirkpatrick 哈密顿量.107 4.1.1 我们的方法.110 4.1.2 相关工作. 114 4.2 稀疏主成分分析.115 4.3 张量主成分分析.124 4.4 植入的稍密子图.127 4.5 我们的方法.129 4.6 关于认证问题的平方和下界的相关工作 . 134 4.7 证明的组织结构. 134

5 谢尔 明顿-柯克帕特里克哈密顿量.136
5.1 技术预备知识.136 5.1.1 问题陈述.136 5.1.2 图矩阵.137 5.1.3 范数界.141 5.2 证明策略.142 5.3 伪校准.146 5.3.1 PAP 植入分布.147 5.3.2 伪校准技术.147 5.3.3 高斯情形下的伪校准.149 5.3.4 布尔情形下的伪校准.153 5.4 证明半正定性.156 5.4.1 非蜘蛛可忽略.158 5.4.2 消去单个蜘蛛.165 5.4.3 消除所有蜘蛛.177 5.4.4 完成证明.182 5.5 Sherrington-Kirkpatrick 下界.189 5.6 省略的技术细节.192 5.6.1 范数界.192 5.6.2 e (k) 的性质.194

6 分析工具与定性界.201 6.1 证明工具.201 6.1.1 形状与图矩阵.202 6.1.2 形状分解.210 6.1.3 系数矩阵.214 6.1.4 主要定理.217 6.2 植入的稍密子图的定性界.222 6.2.1 伪校准.222 6.2.2 定性工具界.224 6.3 张量主成分分析的定性界.225 6.3.1 伪校准.225 6.3.2 定性工具界.228 6.4 稀疏主成分分析的定性界.234 6.4.1 伪校准.234 6.4.2 定性工具界.237 6.4.3 定量界的直观解释.247

7 定量界.250
7.1 植入的稍密子图：完整验证.250 7.1.1 中间形状界.250 7.1.2 交项界.252 7.1.3 截断误差界.254 7.2 张量主成分分析：完整验证.257 7.2.1 中间形状界.258 7.2.2 交项界.260 7.2.3 截断误差界.263 7.3 稀疏主成分分析：完整验证.266 7.3.1 中间形状界.267 7.3.2 交项界.272 7.3.3 截断误差界.276

8 后续跟进及未来工作.281
8.1 非乘积分布的非线性集中性.281 8.2 平方和下界.281 8.2.1 稀疏独立集.282 8.2.2 植入仿射平面与最大割.284 8.2.3 唯一博弈.286 8.3 低度似然比假设.287 8.4 技术改进.287 8.5 对计算机科学的启示.288

参考文献 290

图目录

1.1 一个含 5 个顶点的图及两种可能的割.1 2.1 用于矩阵乘法和 [106] 中算法的张量网络. . . . . 20 2.2 图 τ 及其对应的展平张量网络 .22 2.3 左：对应邻接矩阵的形状；右：更复杂形状的示例.41 2.4 展示 τ_P 定义方式的示例。在此例中， P 约束蓝色边和红色边分别连接到 α 和 β。 U_{τ_P}、V_{τ_P}的顶点具有排序（此处未显示）.46 2.5 通过图示证明 $|S| \geq |S_\tau|$。绿色边可以出现在 τ中，橙色边不能，因此 S 阻断了从 U_τ 到 V_τ的所有路径.47 2.6 展示 τ_P 定义方式的示例。在此例中， P 约束蓝色边和红色边分别连接到 α_1 和 α_2 。此外， P 表示 γ_1、γ_2 中有些边处于激活状态（以实线边表示），另一些边未激活（以虚线边表示）。我们保留 τ_P中的实线边。 U_{τ_P}、V_{τ_P}的顶点也具有排序（此处未显示）.78 4.1 稀疏主成分分析的计算障碍图.118 5.1 傅里叶多项式 $h_3(d_u,i_1)h_1(d_u,i_2)h_2(d_u,w_1)h_1(d_u,j_1)h_1(d_u,j_2)$ 表示为图.138 5.2 两个平凡形状的示例.143 5.3 基本非蜘蛛形状 α的示意图.143 5.4 基本蜘蛛形状 α的示意图.144 5.5 形状 β_1 和 β_2的示意图.144 5.6 近似 $\beta_1 \times \beta_1 \approx \alpha$.145 5.7 构成 L_4的五种形状.169 5.8 图矩阵的一个令人惊讶的等式.174 6.1 形状示例.204 6.2 形状组合与分解的示意图.211 6.3 形状组合与分解的示意图.212 6.4 形状 $\sigma \circ \tau_1 \circ \sigma^T$、$\sigma \circ \tau_2 \circ \sigma^T$和 $\sigma \circ \sigma^T$。所有边的标签均为 1。. 248

T

致谢

在芝加哥大学度过的那些岁月，是我人生中最美好的时光之一，无论在学业、事业还是个人生活方面皆是如此。毫无疑问，这完全归功于我所处的那种滋养人心、鼓舞人心的环境，而非我所做的研究、取得的进展或工作的效率。寥寥数页不足以充分致谢那些促成这一切的人。不过，我还是想尽力一试。

首先，我要衷心感谢我的导师 Madhur Tulsiani 和 Aaron Potechin。这些年来，正是他们极大的耐心、慷慨、善良与指导支撑着我一路走来。他们对我的人生所产生的深远而积极的影响，再怎么强调也不为过。我找不到比他们更好的导师了。谨向他们致以由衷的谢意。

我还记得，2016年上 Madhur 的数学工具课程时，我深深为之着迷；后来，我主动找他，希望一起研究研究问题。从那以后，Madhur 一直慷慨地付出时间，我们也不断携手攻克许多难题。我从他身上学到了很多，不仅是与研究相关的，也包括其他方面。他对研究的品味极佳，学习他直觉而有效的研究方法，对我自身的成长大有裨益。每当看到他对计算机科学众多不同领域都了如指掌，我至今仍感到惊叹；此外，他的工作态度也令我由衷敬佩。除了研究之外，我也怀念我们围绕各种话题畅谈的时光。衷心感谢他给予我的智慧、持续不断的鼓励，以及在我尤其需要时始终不懈的支持。

Aaron 加入芝加哥大学时，我刚开始对平方和方法产生浓厚兴趣；如今，这已成为我的学位论文的主要主题之一。在接下来的几年里，我为这个领域作出了令自己非常自豪的贡献。这得益于他的指导，也得益于我借鉴了他的解题方法：我们从第一原理出发，不断攻克问题，直到问题迎刃而解，或无论如何都有了非凡的发现。我科研生涯中一些最难忘的时刻，都来自与 Aaron 的合作。我也感谢他认真审阅

了这篇学位论文，并提出了许多意见。

我由衷感谢芝加哥大学布斯商学院的 Bryon Aragam。我和同事 Bohdan 找他请教可以研究的问题时，他特意抽出时间向我们介绍自己的研究领域，还给了我们许多引人深思的问题和想法。最终，我们开展了非常出色的合作，产出了许多让我引以为傲的成果。令人惊喜的是，这些成果引出了我如今热衷并持续深入思考的主题。Bryon 思路清晰，沟通能力出色，而且有着非凡的才能，能够提炼任何问题或研究成果对计算机科学的影响。我很高兴有机会向他学习。

能与 Aravindan Vijayaraghavan 共事，我感到非常幸运。每当我们遇到瓶颈时，Aravindan 的乐观态度总能让我保持动力，因此我们始终能够以焕然一新的热情继续攻克问题。这段经历让我成长为一名更优秀的研究者。此外，他为人极其友善，也给了我许多睿智的建议，对此我深怀感激。

我要感谢 Janos Simon教授愿意担任我的委员会成员，并在我攻读博士学位期间一直给予我支持。他总是乐于倾听我所做的工作，并提出深思熟虑的见解。我很荣幸能向 Laci Babai教授学习。与他的每次交谈都令人振奋，他对研究和教学的无限热忱总能激励我、让我重新焕发活力。感谢 Borja Sotomayor 邀请我加入芝加哥大学社群的 ICPC编程竞赛活动，让我的编程技能始终保持娴熟。

博士学位学习的初期，我没有多少合作者，长时间困在各种概念或问题上，令人十分沮丧。然而，读博过半时，我意识到，与他人合作时，这种挫败感会大大减轻。从某种意义上说，攻读博士学位的主要收获之一，就是学会如何学习，以及如何让艰苦的挣扎不那么令人难受。对我而言，合作研究就是答案。此外，合作让我不仅享受最终目标，也享受研究过程

。我很幸运，在攻读博士学位期间与一些杰出的人共事，他们帮助我成长为一名研究者，也让我成为更好的个人。这些人包括 Bryon Aragam、Ainesh Bakshi、Xue Chen、Ming Gao、Mrinalkanti Ghosh、Fernando Granha Jeronimo、Chris Jones、Bohdan Kivva、Sidhanth Mohanty、Aaron Potechin、Pradeep Ravikumar、Madhur Tulsiani、Aravindan Vijayaraghavan、Jeff Xu，以及所有曾与我讨论数学的人。

我想特别提及其中两位——Chris 和 Bohdan。感谢开朗的 Chris，感谢我们一起花费数百小时研究问题、阅读论文、学习新主题和完成习题集。他也是我的亲密好友和知心伙伴，我们许多与研究无关的谈话和活动都令人难忘。我也很幸运能与聪慧的 Bohdan 共事，他惊人的直觉和清晰的思路至今仍激励着我。感谢他的善意与友谊。

最后，在学术方面，我想向我在芝加哥大学之前的教授和老师们致以诚挚的谢意。感谢 Sourav Chakraborty，他也是芝加哥大学博士学位毕业生，导师是 Laci Babai。Sourav 为我提供了职业建议，帮助我申请研究生院，并鼓励我申请芝加哥大学攻读博士学位（并最终入学）。感谢 Geevarghese Philip 将我引入研究领域；2015年夏季，他在宁静的马克斯·普朗克信息学研究所接待了我，我们在那里一起研究参数化算法。我感谢 Narayan Kumar、Madhavan Mukund、K. V. Subrahmanyam 和 B. Srivathsan 在我就读金奈数学研究所期间，于大学生涯的各个阶段给予我帮助，也感谢他们让我在那里度过的时光如此难忘。感谢 Sadagopan Rajesh 先生带我接触奥林匹克数学，这或许改变了我的整个人生。感谢我的好友 Amir Goharshady 教我 C++ ，并带我接触竞赛编程；直到今天我仍乐在其中，而这也促使我的研究方向从数学转向计算机科学。

感谢芝加哥大学和 TTIC 为我提供计算集群，以便开展部分

研究，也感谢这里出色的工作人员和设施，让我在此的生活十分愉快。

研究生院固然主要是一个学术场所，但我的生活远比预想中有趣得多。我会怀念那些踢足球（soccer）的日子——踢的次数多得简直不合常理，尤其是在 Stagg和South球场，有时也会去 Jackman球场、Henry Crown球场、Midway Plaisance、Stony Island球场、Fire Pitch，以及芝加哥各处随机找到的球场。也许最好的收获，就是足球让我结识了生命中一些最美好的人。除了踢足球，我在这座城市也遇到了许多了不起的人，这在一定程度上要感谢芝加哥大学国际学生之家在校园里组织的社交活动。我很幸运，拥有一个温暖体贴的社交圈，帮助我成长为更好的人。以下是一些与我亲近的人，有些甚至比家人还亲；他们按姓氏的字典序排列（如果不小心漏掉了谁的名字，我先在此致歉）：Sudarshan Babu、Alan Chang、Alexis Comte、Leonardo Coregliano、Adam Dziedzic、Animesh Fatehpuria、Nicolas Fierro、Mrinalkanti Ghosh、Mina Gian、Amir Goharshady、Neng Huang、Jafar Jafarov、Fernando Jeronimo、Reza Jokar、Chris Jones、Marie Kim、Akash Kumar、Hyunku Kwon、Kunal Marwaha、Matt McPartlon、Omar Montasser、Tushant Mittal、Sidhanth Mohanty、Sayan Mukherjee、Nathan Mull、Kshitij Patel、Wojciech Nadara、Rachit Nimavat、Jair Pinedo、Aravind Reddy、Suhail Rehman、Harry Ren、Amar Risbud、Diego Rojas、Aritra Sen、Hy Truong Son、Harsha Srinivas、Zihan Tan、Akilesh Tangella、Shubham Toshniwal、Shubhendu Trivedi、Xiaoyan Wang、Jeff Xu、Xifan Yu、Ivan Zelich、Meiqing Zhang和Wei Zou。

最后，我要感谢家人，他们给了我快乐的童年，培养了我无忧无虑、乐观开朗的性格，并在人生的每个阶段给予我坚定的支持。为此，我感谢我的兄弟 Aravind Raj、姐姐 R. Ishwarya、母亲 R. Anandhi 和父亲 A. S. Rajendran 博士。我的大部分个性、特征、热情、思想以及人生观，都深受已故父亲的影响。我谨以这篇学位论文，献给他，并缅怀他慈爱的身影。

摘要

我们在非线性随机矩阵理论中开发了新的工具，并将其用于研究平方和（SoS）层级在平均情形问题上的表现。

SoS 层级是一种强大的优化技术，在组合优化、稳健统计学和机器学习中的各种问题上取得了巨大成功。它是一系列凸松弛，使我们能够灵活地权衡运行时间与近似保证。近期研究表明，它对于从高维含噪数据中恢复结构极为有用。它也仍然是我们反驳臭名昭著的唯一博弈猜想的最佳方法。

在这项工作中，我们分析了 SoS 层级在源自统计学、理论计算机科学和统计物理学的基本问题上的表现。具体而言，我们针对 Sherrington-Kirkpatrick 哈密顿量、植入的稍密子图、张量主成分分析和稀疏主成分分析问题，证明了亚指数时间平方和下界。这些平方和下界涉及对大型随机矩阵的分析，而我们的主要贡献也正在于此。这些结果有力地支持了低次似然比假设的正确性，并增进了我们对该假设的理解。这一重要猜想预测了有界时间算法在假设检验中的能力。

我们还开发了通用工具，用于分析由独立随机变量构成的随机矩阵的行为。为此，我们以 Efron-Stein 不等式的矩阵版本为基础，并对其加以推广。具体而言，我们关于矩阵集中性的通用定理涵盖了文献中已有的多项结果。我们预计，这些随机矩阵理论的思想还将有其他重要应用。

第1章

引言

算法设计、数学优化和计算复杂性是计算机科学中紧密相连的领域，在发展初期，它们大多各自平行发展。近几十年来，这些领域的研究呈爆发式增长，并常常相互借鉴思想，如今它们之间已不再有明显的界限。事实上，如今可以将这些计算机科学领域理解为都在试图实现同一个目标——哪些问题容易解决，哪些问题难以解决?

早期研究人员主要关注搜索问题。给定一个输入，目标是搜索所需的隐藏结构。通常，这可以等价地重新表述为：在各种约束下优化适当目标函数的问题。

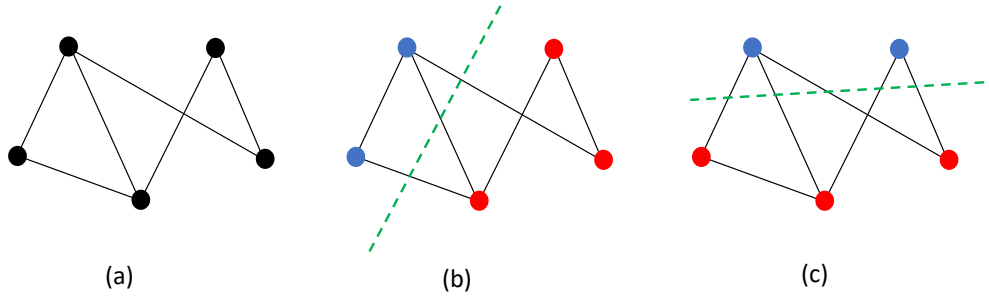


图1.1：一个包含5个顶点和两种可能割的示例图。

例如，考虑最大割 问题，其输入是一个图，目标是将顶点集合分拆为两个子集，使端点位于不同部分的边数最大化。以图 1.1 中的图（a）为例，（b）和（c）展示了两种可能的分拆，其中蓝色顶点构成一部分，红色顶点构成另一部分。那么，（b）中的分拆割出 3 条边，而（c）中的分拆割出 5 条边，即与绿色线相交的边。通过简单的奇偶性论证，很容易看出我们不可能得到超过 5 条边。

在搜索问题的表述中，我们希望搜索算法输出一个能割出尽可能多边的分拆。在优化问题的表述中，我们希望优化算法正确输出最优值。

计算问题的另一种表述是判定问题。给定一个输入，目标是判断是否存在某种隐藏结构，或者目标值是否满足某些性质；同时，算法只能返回布尔输出——例如“真”或“假”，或者“是”或“否”。在上述最大割的例子中，从判定问题的角度来看，可以问：给定图中的最大割是否至少包含边总数的 0.6（比如说）？

这些类型的问题都密切相关，在许多情况下，本质上都归结为寻找算法。出于实用考虑，我们要求算法具备效率、准确性等各种性质。这推动了可计算性、复杂性理论和优化领域丰富理论的发展。在这篇学位论文中，我们也将考察相关问题类型的视角，即认证问题和假设检验。正如我们将看到的，这些其他表述与前者以及彼此之间都有关联，但目前尚不清楚这些联系究竟有多深；试图理解这一点，是理论计算机科学中的一项重要课题。尽管如此，所有这些表述背后的目标，都是寻找高效算法，以从数据中检测并提取结构；或者论证除非我们愿意在效率或准确性等其他方面作出妥协，否则不存在此类算法。

1.1 认证问题

与搜索问题或判定问题不同，认证问题要求在给定输入后，求出一个以概率 1 成立的目标值界限，并提供该输出界限的证书。算法的质量通常根据该界限接近真实最优值的程度来衡量。

在运行　　以最大割为例，给定一个图，任务可能是输出一个

始终不小于最大割大小的值。一种简单的算法是直接返回图中边的总数。确实，这是一种有效的认证算法，但我们可以进一步探讨是否有更好的方法。

这是一种截然不同的算法设计方法。设想我们要最大化某个目标函数，因此希望得到最优值的上界。那么，设计认证算法可以理解为从上方 着手解决问题，而不是从下方着手；后者才是更常见的算法设计思路。

线性规划松弛的概念已经提供了这样的认证算法。对于一个可以表述为整数规划的问题（许多问题都可以），一种自然的认证解求法是将搜索空间从整数变量扩展到实变量，并根据需要添加其他适当的约束。这称为松弛该规划。这样可以尝试用更快的算法计算解，但代价是只能得到近似解。更重要的是，无论输入是什么，返回解得到的目标值都是最优解的确定上界。这正是认证算法所需要的。衡量返回结果的质量，往往取决于所采用的松弛类型以及问题特有的结构。

在许多情形下，可以通过考察问题的松弛来获得近似算法：先求得一个非整数解，再将其舍入为有效解。对于最大割问题，Goemans 和 Williamson 在其开创性工作中采用了半定规划松弛；这种松弛比线性规划松弛更强。 [89]

在本学位论文中，我们将聚焦于这类认证算法中的一个特定类别，即平方和（SoS）层级，有时也称为 Lasserre 层级。SoS 层级是一系列针对给定规划的凸松弛。由于它是一种

松弛，因此可用于认证。凭借其在最大割、约束满足等各种基础优化问题上的巨大成功，SoS 层级已成为一种强大的优化技术。相关结果进一步表明，在一大类半定规划松弛中，SoS 层级是最优的松弛 [150], ；并且在假设著名的唯一博弈猜想成立的情况下，它是每个约束满足问题的最佳近似算法 [195]。本学位论文的一个主要目标是理解这一强大技术的局限。我们尤其关注所谓的平均情形设定，下面将对此作出定义。

1.2 平均情况分析

本书的一个重要主题是研究问题的随机实例，这称为平均情况分析。与传统的最坏情形算法设计不同，后者旨在设计一种能在最坏情况下的输入上表现良好的算法；近年来，针对输入从某个分布中随机抽取的问题，研究取得了令人振奋的进展。例如，在最大割问题中，我们可以假设输入来自 Erd˝os-R´enyi 随机图族：图中的顶点数量预先选定，每条边都以 0.5 的概率独立地存在。

在平均情形算法设计中，我们希望设计出能够以高概率在平均情形输入上表现良好的算法，而不是要求它们对所有输入都表现良好。这一点很重要，因为研究某个问题的最坏情况复杂度，未必能揭示该问题本身的困难性。这是因为算法的最坏情况输入可能极为人为且刻意构造。换句话说，在现实场景中，我们遇到的各种优化或搜索问题，其输入不太可能是这样的实例。实践中也能看到这一点。例如，Klee 和 Minty [135], 证明，线性规划的单纯形法[56] 在最坏情况下会慢到呈指数级，但

在实际应用中却表现极佳。许多研究都试图解释这一现象，例如[33, 217, 34, 218], Spielman 和 Teng 的研究尤为突出，他们因此获得了 2008年 的 G¨odel 奖。

人们投入了巨大努力，试图理解各种问题的平均情形复杂度。旨在设计平均情形算法的研究，有助于更深入地理解问题的核心，也能推动最坏情形算法的设计。例如，著名的最稠密 k子图问题 [30]便是如此。本文将聚焦于平均情况分析。

在我们的研究中，反复出现的基础数学对象是大型随机矩阵。我们常常希望了解它们的行为。

1.3 本文的主题：随机矩阵 s

随机矩阵在计算机科学中十分常见，尤其是在优化和统计学领域。算法分析常常需要研究某些随机矩阵的行为，而这些矩阵可以由输入构造出来。即使在计算机科学之外，随机矩阵理论本身也是一门基础学科，自 20 世纪初以来便一直受到研究，其应用也延伸至数学和物理学的诸多分支。简要综述请参见 [81]。

过去几十年来，人们投入了大量精力发展随机矩阵理论，参见 Tropp [227] 的著作。例如，矩阵 Bernstein 不等式研究矩阵的随机加权和的行为；Wigner 半圆律研究从高斯正交系综中抽取的随机矩阵的特征值分布。另一方面，要理解非线性随机矩阵的行为，可用的工具则较少。在这类矩阵中，每个矩阵项都是输入的非线性函数，例如低次多项式。

在我们的设定中，在尝试分析 SoS 层级以应对各种问题时，这种情况经常发生。 s

无论是尝试通过 SoS 设计算法，还是研究 SoS 算法的局限性（例如，
[22, 107, 211, 167, 119]），都是如此。因此，我们从理解非线性随机矩阵的行为这一重要课题入手。在本论文的第一部分，我们特别关注集中行为。我们强调，这本身就是一个重要的研究方向。

为了界定随机矩阵相对于其均值的波动（以两者之差的谱范数或 Schatten t-范数衡量），一种简单而强大的技术已得到广泛应用（包括上述许多文献），这就是所谓的迹方法。在该方法中，将（中心化后的）随机矩阵提升到较大的幂次，并对所得矩阵的期望迹进行界定。虽然这种方法能够得到令人满意的结果，但它往往需要巧妙的观察和极其复杂的组合学技巧。

另一种方法如下。考虑一个由多个独立输入变量构成函数的随机矩阵。我们可以通过研究重新采样一个均匀随机选取的输入矩阵项时，该矩阵的偏离程度，来分析其行为。界定这些局部波动，就能界定随机矩阵的整体波动。这一技术导出了 Efron-Stein 不等式。最初，这些不等式是为标量随机变量（可视为 1×1 矩阵）而提出的。在这一特殊情形下，它们极为强大，因为已有研究表明，它们能够推出许多标准的集中不等式。近期，文献 [188] 提出了 Efron-Stein 不等式的矩阵版本。在这项工作中，我们以此为基础，建立一个证明大型随机矩阵集中性的一般框架。

在本论文的第二部分，对平方和算法进行分析时，出现的根本困难是分析一个大型非线性随机矩阵的行为。特别是，我们希望论证：对于输入的选择，该随机矩阵以高概率是半正定的。为此，我们给出了该矩阵的近似 Cholesky分解，而证明大量依赖于我们

在上文中得到的集中结果。

总而言之，本工作的主旨是研究非线性随机矩阵：我们既建立了一个分析集中性的通用框架，也将其用于研究基本问题上的算法。

1.4 平方和层级

对于具有多项式不等式约束的规划形式的优化问题，已有许多研究提出通用方法来松弛该规划，以便高效地获得良好解。一些较为主流的方法包括 Lov´asz-Schrijver 层级 [155] 和 Sherali-Adams 层级 [214]。粗略地说，这些算法层级将规划提升到一个更大的变量集合，并通过各种约束将这些变量联系起来；随后对更大的规划进行松弛并求解，最后将解投影回原始变量空间。它们由一个称为度的整数参数化；度越大，松弛越紧，但运行时间也越长。

平方和（SoS）层级是一种类似的优化技术，利用了半定规划的幂。对于多项式优化问题，平方和层级最初由 Shor [216], Nesterov [173], Parillo [185], Lasserre[148] 和 Grigoriev [91, 92], 独立研究，提供了一列由称为平方和层级度的整数参数化的凸松弛。随着层级的度 d 增加，我们会得到逐渐更强的凸松弛，并且可以在 $n^{O(d)}$ 时间内求解。这使平方和层级几乎成为算法设计中的一种黑箱工具。多项研究表明，无论在最坏情况还是平均情况设定下，它都是解决各种问题的强大算法。

以最大割问题为例。具有开创性的 Goemans-Williamson 算法 [88] 通过半定规划松弛，为该问题实现了 ≈ 0.878 的近似比。事实证明，这个算法正是 2 度

SoS 层级。人们猜想这个近似比是最优的，且越来越多的证据表明事实确实如此。这凸显了 SoS 层级为何如此强大。

事实上，利用 SoS 层级为组合优化问题（例如 [89, 10, 94, 196]）以及源自统计学和机器学习的问题（例如 [22, 23, 108, 193, 140]）设计高效算法，已经取得了巨大成功。实际上，对于最稀疏割 [10], 、最大割 [89], 、张量主成分分析 [108] 以及所有 Max-k-约束满足问题 [195]等许多基本问题，平方和都实现了最先进的近似保证。正如前文所述，对于一大类问题，平方和松弛已被证明是所有半定规划松弛中最高效的 [150]。

“平方和”这一术语源自证明复杂性中的一种对偶视角。除了作为一种算法技术，平方和也可以等价地视为对多项式优化问题最优值界限的证明或证书。这项研究可以追溯到希尔伯特第十七问题，并由此推动了对一种称为正定性定理的证明复杂性结果的研究。该定理给出了能够证明多项式系统无解的条件，参见例如 [221, 194, 205]。Lasserre [148] 和 Parillo [185, 186] 最初发现了其算法意义，由此将平方和解释为一种优化技术，正如我们在本文中所研究的那样。这种对偶性可以完全形式化，并催生了所谓的“从证明到算法”框架，该框架已取得巨大成功，尤其是近年来在稳健统计学领域，参见例如 [140, 120, 101, 18]。有句格言是：如果我们能为某个搜索问题的可辨识性结果找到一个“简单”的证明，就可以将其自动化，从而得到一种算法。我们在此不对此进行详细探讨，读者可参阅专著 [80]。

接下来，我们将讨论平方和下界。不过在此之前，我们先介绍一些近期在学界逐渐受到关注的相关技术。

1.4.1 相关算法技术

除了搜索、判定和认证之外，研究人员还考虑了其他相关类型的问题。设输入是从两个已知分布之一中采样得到的，我们希望判断它是从哪个分布中采样得到的。这类问题通常称为假设检验。例如，一个分布可以是 Erd˝os-R´enyi 随机图的分布，另一个则可以是其中植入了大割的 Erd˝os-R´enyi 随机图的分布。显然，这类问题与随机图上的最大割问题属于不同类型。研究这些相关表述不仅本身很有趣，也能为搜索或认证变体提供不同视角和有趣的见解。另一类称为恢复问题的问题，是指当输入从后一种分布中采样得到时，恢复其中植入的结构。

对于迄今考虑过的所有类型的问题，除了平方和之外，还考虑过其他几种算法框架，其中有些还得到了深入研究。例子包括

- Lov´asz-Schrijver 和 Sherali–Adams 层级——如前所述，这些层级将一个规划提升到包含更多变量的空间中，然后放松任何整数性约束。随后将得到的解投影回原始变量，再将其舍入为整数解。平方和层级涵盖了这些层级；换言之，平方和层级至少与这些层级一样强大[80]。

- 低次多项式——在假设检验中，可以使用低次多项式尝试区分两个分布。更确切地说，如果存在一个低次多项式，使其在两个分布下的期望值表现不同，且方差不太大，就可以用它来区分这两个分布。这与平方和层级有关，我们稍后会更详细地探讨这一点。

- 统计查询算法——在假设检验中，统计查询模型（SQ）是 [123] 提出的一种广受欢迎的受限算法类别。在该模型中，对于某个底层分布，我们可以通过查询函数的期望值（允许一定误差）来间接访问它。获得这个预言机的访问权限后，我们希望进行假设检验。SQ 算法涵盖了统计学和机器学习中的广泛算法技术，包括谱方法、矩方法和张量方法（例如参见 [76, 77]）。SQ 算法也被用于研究信息与计算之间的权衡，而且更广泛地说，也在其他背景下得到研究 [75]。此外，大量研究致力于理解 SQ 算法的局限性（例如 [76, 78, 61]）。近期研究 [39] 表明，在温和条件下，低次多项式与统计查询算法的能力等价。

- 近似消息传递以及其他统计物理学技术，例如置信传播，参见综述 [240]。

- 局部算法，参见例如 [70, 71, 111]。

- 有界规模的电路计算模型，参见例如 [207, 208]。

1.5 平方和层级的下界

平方和层级在各种问题上的巨大成功，使得研究其局限性成为一个重要的研究方向，而这正是本学位论文致力于探讨的内容。具体而言，我们将聚焦于平均情形问题；正如我们将看到的，大多数技术难点都归结为对非线性随机矩阵的分析。为此，我们将开发各种技术。

研究针对平方和层级的下界有许多重要原因。平方和层级足够通用，能够涵盖广泛的一类算法推理 [80]。具体而言，平方和涵盖 Lov´asz-Schrijver 和 Sherali–Adams 层级，

在一些温和的限制条件下，也涵盖统计查询算法和基于低次多项式的算法。因此，平方和下界向算法设计者揭示了问题的内在困难性，并表明若想突破算法障碍，就需要寻找平方和层级无法涵盖的算法。其次，在平均情形问题设定中，诸如 P $\neq$NP 等标准复杂性理论假设尚未被证明能够揭示高效算法的局限。相反，针对平方和等强大技术的下界，已成为计算困难性的有力证据 [104, 110]。因此，理解平方和层级在这些问题上的能力，是理解这些问题可近似性的重要一步。另请参阅综述 [25, 166] ，了解更多相关内容。

与我们讨论过的其他一些算法类别相比，关于平方和下界的研究相对较少，这可以归因于证明此类下界在技术上的巨大难度。例如，文献 [92, 210, 137] 研究了随机约束满足问题的平方和下界。一系列研究 [72, 164, 59,20, 180] 探讨了随机图上最大团的平方和下界。其他一些平方和下界研究（不包括本论文中的研究）见文献 [156, 139, 165, 143, 138]。

1.6 主要结果总结

在本研究的第一部分，我们研究非线性随机矩阵的集中行为。在第二部分，我们研究针对若干基本问题的平方和层级下界。

1.6.1 通过矩阵 *Efron-Stein* 实现非线性矩阵集中

我们首先给出一个关于随机矩阵集中性的通用定理，其中矩阵的元素是独立随机变量的多项式。著名的矩阵伯恩斯坦不等式回答了仅含线性多项式时的这一问题。然而，理解非线性多项式的

情形同样重要，但这也带来了重大挑战。在文献中的各种应用中，这类随机矩阵通常采用所谓的迹方法进行处理。尽管这种方法能够得到所需结果，有时效果极佳，但应用该方法通常极其复杂。在本文中，我们提出了一种通过矩阵版埃夫隆–斯坦不等式证明矩阵集中的替代方法。我们提出了一个通用的矩阵集中不等式，其证明依赖于强大的可交换对方法。我们展示了该不等式的一些应用，并预计它在我们所探讨范围之外也将具有重要的应用价值。

1.6.2 平方和下界

我们针对计算机科学中的各种基本问题，证明了 SoS 层级的强次指数时间下界。我们所有的应用都从所谓的伪校准启发式出发，将问题归结为分析一个大型随机矩阵（称为矩矩阵）的行为。我们的概念性和技术性创新都体现在这一步。以下是我们给出的结果。

Sherrington-Kirkpatrick 哈密顿量

统计物理中的一个重要问题——Sherrington-Kirkpatrick 问题，旨在对从高斯正交系综中采样得到的随机矩阵的二次型进行优化，其中优化对象为布尔向量。长期以来，人们已经知道，真实的最优值经缩放后会集中于某个特定常数。近期，有人针对这一优化问题提出了一种高效算法。另一方面，人们普遍认为，除了简单的谱算法之外，认证很难实现。我们通过给出针对该问题的平方和下界，为这一观点提供了有力证据。要完成这项工作，我们需要理解矩矩阵的零空间，并在应用矩阵集中工具之前将其化为零。从概念上说，这项工作让我们深入了解了平方和算法在其他

最大割和高斯混合学习等基本问题上的表现。

稀疏主成分分析

稀疏主成分分析是主成分分析（PCA）的一种变体，而主成分分析是统计学和机器学习中的一项基础方法。我们研究尖峰 Wishart 模型，它是这一问题最自然的版本，但已证明很难用平方和进行分析。先前工作预测了恢复稀疏分量时存在的计算障碍，该障碍体现为维度、稀疏度和样本数量之间的权衡。我们通过证明针对亚指数时间平方和方法的下界，确认了这一障碍；这些下界与已知算法相匹配。本研究将随机矩矩阵拆分为不同的矩阵，并运用创新的组合收费论证，研究这些矩阵之间的相互作用。从概念上说，本研究确认了这一问题的计算障碍图；该图早已得到预测，并长期以来被认为是正确的。

植入的稍密子图

在给定图中寻找稠密子图是一项重要问题，多年来无论从算法角度还是算法困难性角度，都受到了广泛研究。对于该问题在某些参数范围内的随机实例，其困难性已形成猜想，通常称为 PDS 猜想。该问题也被用作典型困难问题，通过归约到各种其他问题来研究它们的计算障碍。此外，这些困难实例也被用作密码方案的基础。因此，针对该问题证明平方和下界，将有助于进一步验证这一猜想。在这项工作中，我们针对某些参数范围证明了此类亚指数时间下界；人们普遍认为，在这些范围内求解该问题需要亚指数时间。

张量主成分分析是这样一个问题的平均情形版本：在球面上优化齐次多项式。由于该问题与众多领域相关，因此它是优化领域中的一项基础且重要的问题。在这项工作中，我们证明了与该问题已知算法相匹配的平方和下界，从而确定了该问题中平方和方法的计算障碍。这项工作还为一般齐次多项式优化的可近似性-不可近似性阈值提供了见解，并表明随机实例可能并非该问题最困难的实例。

1.7 未纳入的研究

本学位论文包含了我在攻读博士学位期间所开展研究的主体内容，但遗憾的是，还有其他研究方向未能纳入其中。这些研究包括以下工作。

1.7.1 稀疏独立集的平方和下界

在我们的工作中 [119], ，我们证明了稀疏 Erd˝os-R´enyi 随机图上最大独立集问题的 SoS 下界，与 Lov´asz theta 函数在低阶项范围内相匹配。为此，我们在本学位论文所开发工具的基础上，进一步发展了多种新技术。具体而言，这项工作首次探索了这一重要研究方向：理解 SoS 在稀疏随机图上的局限性。值得指出的是，在这项工作中，我们第2章提出的非线性矩阵集中工具非常有用。我们将在第8章详细阐述这一结果，因为它建立在本学位论文中将要发展的许多工作之上。

1.7.2 因果推断

因果推断研究如何从观测数据中发现并理解因果关系，在医学、《遗传学》、经济学、流行病、人工智能等领域有着广泛的应用。在我们的工作 [200], 中，我们专注于从数据中学习一类称为贝叶斯网络（BN）的因果模型。这是一个经典且基础的问题，因为贝叶斯网络结构紧凑、模块化，并提供直观的因果解释，因此在各个领域都非常有用。我们提出并研究了一种解决这一问题的实用新算法。该算法效率高，可证明不同于广泛使用的贪心等价搜索算法；而且，由于它是一种通用的基于评分的学习算法，因此适用范围广。此外，在一些受近期研究启发并对其加以推广的统计假设下，即使对于非参数模型，我们的算法也能在不对线性、可加性、独立噪声或忠实性作出任何假设的情况下，可证明地恢复真实的贝叶斯网络。它还揭示了与其他机器学习领域之间有趣的潜在联系，例如聚类、前向-后向贪心方法和核方法。

1.7.3 潜变量建模

在我们的工作中 [132], ，我们研究了一个相对少受关注但十分重要的问题：对观测数据进行潜变量建模。承接上一节，我们现在考虑观测变量背后未被观测到（有时甚至无法测量！）的潜在原因或混杂因素。我们聚焦于概率混合模型这一设定，它自然地出现在机器学习、经济学、金融、生物学等领域。在对模型作出一些自然假设的前提下，我们开发了一种算法，以观测数据为输入，揭示隐变量及其背后的因果关系。以往与这一问题相关的工作通常聚焦于线性模型等特殊设定。与之不同的是，我们针对高度非线性的混合模型设定提出了一种算法，该算法基于现有算法运行

，用于混合模型阶数估计（这比密度估计更容易）。

1.7.4 因果表征学习

因果表征学习是机器学习领域一个令人振奋的新分支，它以原始的非结构化数据作为输入，旨在学习生成这些数据的底层生成模型。在此基础上，它还旨在学习所学潜变量之间的因果关系，因此得名因果表征学习。具体而言，该领域汇集了两个长期分别发展的领域中的思想，即因果推断和潜变量建模，也就是上文介绍的两个主题。在我们的工作中， [133], 我们在这一方向上证明了一个有趣且出人意料的结果。我们表明，具有高斯混合先验的广泛一类生成模型是可辨识的（这意味着可以从原始数据中恢复该模型）。具体而言，我们的模型具有通用近似能力，并已被广泛应用于许多有关深度表征学习的实践工作中（但尚未获得理论验证） [62, 116, 235]。

在深度学习领域，人们投入了大量精力来识别潜在特征以及生成观测数据的机制。与手工设计数据的低级特征不同，这一过程主要通过学习低级表征的算法实现自动化。因此学到的模型对各种下游任务都很有用，例如抽样、预测、分类、聚类、干预等。其中一个重要方法是变分自编码器 [131, 204]。过去十年间，人们对变分自编码器进行了多种改进，并将其应用于广泛领域。训练过程一项备受期待的特性是稳定性，即反复训练是否会得到相同的潜变量生成模型。这可以用可识别性这一数学概念来刻画；它是一项至关重要的基本条件，能够保证存在唯一的参数和生成机制来生成数据。撇开计算可行性不谈，可识别性是实现稳定且可重复训练的必要条件。除了训练的稳定性之外

，这也为机器学习中的其他重要考量铺平了道路，例如，人们日益需要学习稳健、可解释、易于说明且公平的数据表示。

在我们的工作中 [134], ，我们证明，对于常用的采用高斯混合先验的变分自编码器，只要扭曲机制是仿射的（特别是，采用 ReLU 激活的深度神经网络满足这一性质），即可保证可识别性；重要的是，这一结论无需假设存在辅助信息。这显著改进了近期大量研究（由 [124] 开创）的成果，这些研究证明，在存在辅助变量或侧信息的情况下，可识别性成立。此外，先前的几项工作通过实证观察发现，高斯混合先验通常能使变分自编码器的训练稳定且可重复，从而暗示了可识别性。我们的工作为这些观察提供了理论依据。

1.8 论文的组织结构

在第2章中，我们推导非线性矩阵集中结果，并展示这些结果在文献中出现的各种非线性随机矩阵上的应用。接着，我们在第3章介绍平方和层级，并介绍用于证明 SoS 下界的伪校准技术，同时展示它与低次算法之间的联系。在第4章中，我们正式陈述本文要证明的主要 SoS 下界，并将其置于已知先前工作的背景下。在第5章中，我们证明 Sherrington-Kirkpatrick 问题的 SoS 下界。接下来的两章，即第6章和第7章，我们将证明植入的稍密子图、张量主成分分析和稀疏主成分分析的 SoS 下界。最后，我们在第8章讨论后续工作和未来可能的研究方向。

第2章

非线性矩阵集中

本章将介绍我们通过 Efron-Stein 不等式研究非线性矩阵集中的技术。本章内容改编自 [201]。虽然我们发展了可用于研究非线性集中的通用技术，且本章内容完全自成一体，但我们对图矩阵的应用将作为很好的热身，帮助读者过渡到本学位论文后续章节中平方和下界的技术部分。

2.1 引言

在优化、统计学和谱算法中，我们常常希望了解各种随机矩阵的集中性。为此，我们可以借助强大的矩阵偏差不等式理论 [227]。例如，矩阵伯恩斯坦不等式适用于如下形式的随机矩阵：

$$\mathbf{M} \;=\; x_1 \cdot \mathbf{C_1} + \cdots + x_n \cdot \mathbf{C_n}$$

其中 x_1、.....、x_n 是相互独立的标量随机变量，而 $\mathbf{C_1}$、.....、$\mathbf{C_n}$ 是固定矩阵。当随机矩阵（例如）$\mathbf{M}$ 是独立随机变量的线性函数时，有大量此类不等式可供使用。然而，近期的若干研究要求我们理解由标量随机变量构成的非线性函数，尤其是低次多项式函数形式的随机矩阵。这正是我们研究的重点。

作为一个激励性示例，考虑如下获得的随机矩阵 $\mathbf{M} \in \mathbb{R}^{[n]^2 \times [n]^2}$

$$\mathbf{M} \;=\; \mathbf{A_1} \otimes \mathbf{A_1} + \cdots + \mathbf{A_m} \otimes \mathbf{A_m}\,,$$

其中，$\mathbf{A_1}$、……、$\mathbf{A_m} \in \mathbb{R}^{[n]\times[n]}$ 是相互独立的随机矩阵，其元素为独立同分布，并在 $\{-1,1\}$ 中均匀分布。容易看出，矩阵 $\mathbf{M}$ 的元素是描述 $\mathbf{A_1}$、……、$\mathbf{A_m}$ 元素的独立随机变量的二次多项式函数。Hopkins 等人 [107, 100], 分析了此类矩阵的集中性，并利用它为主成分分析（PCA）的一种变体设计了谱算法。该矩阵是我们在本文中研究的更一般情形的一个特例。

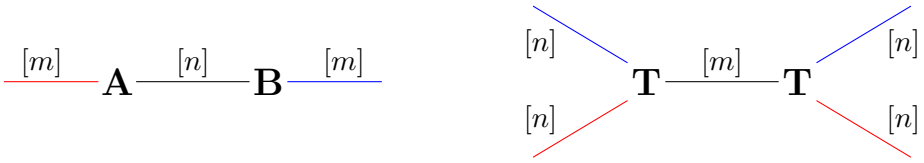
矩阵值多项式函数。 在上面的示例中，矩阵的矩阵项是相互独立的（拉ademacher）随机变量的低次多项式。在本文中，我们考虑一个一般情形：取一个由独立同分布随机变量组成的 n-元组 $Z = (Z_1、……、Z_n)$[1] ，其分布于 Ω。我们考虑由矩阵值函数 $\mathbf{F}(Z)$ 给出的随机矩阵，该函数取值于 $\mathbb{R}^{\mathcal{I}\times\mathcal{J}}$，其中索引集为任意的 $\mathcal{I}$、$\mathcal{J}$；每个矩阵项 $\mathbf{F}[I,J](Z)$ 都是 Z_1、……、Z_n 的多项式。我们建立了一个用于分析此类矩阵集中性的一般框架。当 Z_1、……、Z_n 是在 $\{-1,1\}$ 中均匀分布的相互独立的拉ademacher变量时，我们的矩阵集中结果更易于表述，但对一般情形同样适用。

此类非线性随机矩阵的特殊情形已应用于谱算法和下界中的若干问题。下面我们简要讨论几个示例。

1. **张量网络**。Moitra 和 Wein [167], 将上述随机矩阵视为“展平张量网络”的一种特殊情形；他们还研究了通过稍大的张量网络得到的谱算法。张量网络是一个图，其中的节点对应张量（例如下图）。两个节点之间的一条边对应于某个维度上的共享指标，而每个节点的度等于相应张量的阶（维度的数量）。这类网络展示了如何将不同阶的张量相乘，

1. 我们的框架也适用于变量不一定同分布的情况，只要它们彼此独立。

从而得到更大的张量。例如，下图中的第一个网络展示了两个矩阵$\mathbf{A}$ $\cdot\mathbf{B}$ $\mathbf{A}$ $\in \mathbb{R}^{m\times n}$和$\mathbf{B}$ $\in \mathbb{R}^{n\times m}$进行简单乘法所对应的网络，其中红色和蓝色的边分别表示行索引和列索引。同样，下图中的第二个网络展示了 Hopkins 等人 [106], 所采用方法对应的网络，其中$\mathbf{T}$ $\in \mathbb{R}^{n\times n\times m}$ 是一个具有独立同分布元素的随机张量，其元素取自 $\{-1, 1\}$。虽然后一个网络得到的是一个四阶张量，但他们通过“展平”将其转换为 $\mathbb{R}^{n^2\times n^2}$ 中的矩阵，其中行由红色边上的索引表示，列由蓝色边上的索引表示。图中还标出了各条边对应的索引集（尽管图示中通常会省略这些信息）。Moitra 和 Wein [167] 在其连续多参考对齐问题算法中分析了一个更大的张量网络，其图由 10 个节点组成。



Fig图 2.1：用于矩阵乘法的张量网络以及 [106] 中的算法

2. 图矩阵。 非线性集中性的另一种情形源于对所谓“图矩阵”的分析 [163, 1]。图矩阵在针对平均情形问题的下界中发挥着重要作用，这些问题所针对的是基于强大的平方和（SoS）半定规划层级、运行于多项式时间甚至亚指数时间内的算法 [164, 59, 103, 197, 20, 165, 86, 191, 119]。

令$\mathbf{X}$ 为 $\{\pm 1\}$中随机图的$\mathcal{G}_n$邻接矩阵，$1/2$ 即，当 $i \neq j$ 时$\mathbf{X}[i, j]$ 服从均匀分布，当 $i = j$时取 0。图矩阵是与一个称为“形状”的小型图模式出现次数相对应的随机矩阵。形状 τ是一个小型固定图，其中包含两个有序子集 U_τ，V_τ它们由顶点构成。为简单起见，令

τ为一个固定大小的形状，其中顶点集 $V(\tau)$被分拆为两个有序集合 $V(\tau) = U_\tau \sqcup V_\tau$。对于这样的形状 τ，相应的图矩阵$\mathbf{M}_\tau$的行和列分别由 $[n]^{|U_\tau|}$ 和 $[n]^{|V_\tau|}$ 进行索引，并且我们将行列索引 I 和 J 视为定义了一个映射 $\varphi:\ U_\tau \sqcup V_\tau \to [n]$。相应的矩阵项为

$$\mathbf{M}_\tau[I,J] \ = \ \mathbf{M}_\tau[\varphi(U_\tau),\varphi(V_\tau)] \ = \ \begin{cases} \prod_{(u,v)\in E(\tau)} \mathbf{X}[\varphi(u),\varphi[v]] & \text{if } \varphi \text{ is injective} \\ 0 & \text{otherwise} \end{cases}$$

对于一般图矩阵（在第 2.4.2 节中正式定义），U_τ，V_τ是图 τ 顶点集的任意有序子集，我们对所有可行的单射映射 φ [2] 求和。例如，考虑图 2.2 所示的情形，其中 τ 是一个由三个顶点 $\{u_1, v_1, v_2\}$ 组成的三角形，且 $U_\tau = (u_1)$ 和 $V_\tau = (v_1,\ v_2)$。那么，对应的矩阵为

$$\mathbf{M}_\tau[i_1,(i_2,i_3)] \ = \ \mathbf{X}[i_1,i_2]\cdot\mathbf{X}[i_2,i_3]\cdot\mathbf{X}[i_3,i_1]\,,$$

其中 $\mathbf{X}$ 会自动保证单射性。

图矩阵与张量网络密切相关（忽略对 φ 的单射性约束）。例如，上述矩阵可视为下方展平的张量网络，其中张量 $\mathbf{I}$ 表示一个“三阶对角”张量：当所有指标相等时，其元素为 1，否则为 0。

分析集中性 回顾一下，我们的目标是分析多项式随机矩阵的集中性。为了说明我们的方法，首先考虑获取.....的问题。

2. 在后续章节中，出于技术原因，我们改用另一种定义：对不同的傅里叶特征求和，而不是对不同的单射映射求和

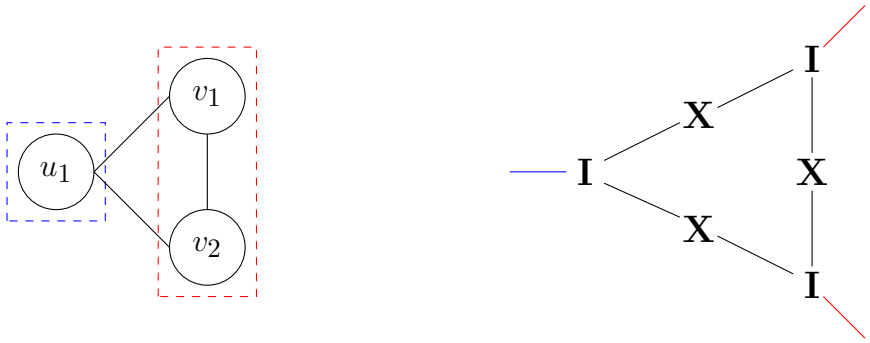


图 2.2：图 τ 及其对应的展平张量网络

关于均值为零的标量 多项式 $f(Z)$ 的集中界。为了获得这样的界，根据马尔可夫不等式，只需计算矩估计

$$\mathbb{P}\left[|f(Z)| \geq \lambda\right] \;=\; \mathbb{P}\left[(f(Z))^{2t} \geq \lambda^{2t}\right] \;\leq\; \lambda^{-2t} \cdot \mathbb{E}\left[(f(Z))^{2t}\right]$$

虽然在某些情形下，可以通过直接展开计算 $\mathbb{E}[(f(Z))^{2t}]$ ，但这通常需要细致分析各项的结构，而其度会随着 t 增长，因此间接方法可能更方便。一种这样的方法基于超收缩不等式。特别是对于拉ademacher变量，超收缩不等式 [174] 给出：对于度为 d_p 的多项式 f ，有

$$\mathbb{E}\left[(f(Z))^{2t}\right] \;\leq\; (2t-1)^{d_p \cdot t} \cdot \left(\mathbb{E}\left[(f(Z))^{2}\right]\right)^{t}.$$

因此，对于（标量）多项式函数，超收缩不等式利用 $(f(Z))_2$ 给出矩估计，这很方便，因为 $(f(Z))^2$ 是一个度为固定 的多项式，因此更容易理解。事实上，通常可以方便地利用 f 的傅里叶系数对其进行分析。

上述论证的矩阵类比涉及 Schatten-$2t$ 范数 $\|.\|_{2t}$，其定义如下：对于一个具有非零奇异值 $\sigma_1, \ldots, \sigma_r$ 的矩阵 $\mathbf{M}$， $\|\mathbf{M}\|_{2t}^{2t} := \sum_{j\in[r]} \sigma_j^{2t}$。

对于函数 $\mathbf{F}$ ，若有 $\mathbb{E}[\mathbf{F}(Z)] = 0$，则利用 Schatten 范数可得以下界。

$$\mathbb{P}\left[\sigma_1(\mathbf{F}) \geq \lambda\right] \;\leq\; \lambda^{-2t} \cdot \mathbb{E}\,\|\mathbf{F}\|_{2t}^{2t} \;=\; \lambda^{-2t} \cdot \mathbb{E}\,\mathrm{tr}\left[(\mathbf{F}(Z)\mathbf{F}(Z)^{\intercal})^t\right]$$

张量网络（涉及高斯变量）[167] 和图矩阵 [1, 119] 的已知范数界依赖于对上述迹的直接展开。他们将展开中的项视为由 $2t$ 个网络/形状副本构成，因此会考察由 $2t$ 个网络/形状副本构成的图，这些副本的顶点集可能彼此重叠。为了分析这类图，他们都依赖复杂的组合学方法。

关于拉ademacher变量 [26]的矩阵值函数，也已知一些超收缩不等式 [26]。不过，这些不等式的形式涉及 Schatten-p 范数 $p \in [1, 2]$ ，而且（据我们所知）尚不能推出矩阵集中。为了解决这个问题，我们考虑另一种基于 Efron-Stein 不等式的间接方法。在标量情形下，Efron-Stein 不等式给出的标量界略弱于上述界。有趣的是，事实证明，这种方法确实可以推广到矩阵情形。

Efron-Stein 不等式。 Efron-Stein 不等式将独立随机变量的函数的全局方差，界定为通过一次改变一个变量而得到的局部方差估计。对于 $i \in [n]$ 和元组 $Z = (Z_1, \ldots\ldots, Z_n)$，令 $Z^{(i)}$ 表示元组 $\widetilde{}(Z_1, \ldots\ldots, Z_{i-1}, Z_i, Z_{i+1}, \ldots\ldots, Z_n)$，其中 Z_i 是 Z_i的一个独立副本。对于标量函数$f(Z)$，Efron-Stein 不等式指出

$$\mathrm{Var}\left[f(Z)\right] \;=\; \mathbb{E}\left[(f(Z) - \mathbb{E}\,f)^2\right] \;\leq\; \frac{1}{2} \cdot \sum_{i \in [n]} \mathbb{E}\left[\left(f(Z) - f\left(Z^{(i)}\right)\right)^2\right] \;=\; \mathbb{E}\left[V(Z)\right],$$

其中 $V(Z) := \sum_{i \in [n]} \mathbb{E}\left[\left(f(Z) - f\left(Z^{(i)}\right)\right)^2 | Z\right]$。对于拉ademacher变量， $\mathbb{E}[V(Z)]$ 等于布尔傅里叶分析中的总影响，而且事实上，也可以通过傅里叶分析得出上述不等式。事实上，当 f 是一个度为 d_p的多项式时，

两边相差一个因子 d_p。

Boucheron 等人提出了埃夫隆-斯坦不等式的一个矩版本 [35]，他们给出了以 $V(Z)$ 为依据的界（事实上，是以更精细的量 $V_+(Z)$ 和$V_-(Z)$ 为依据），这些量可作为方差的替代指标。其结果表明，对于函数 f，

$$\mathbb{E}\left[(f(Z)-\mathbb{E}\,f)^{2t}\right] \;\le\; (C_0\cdot t)^t\cdot\mathbb{E}\left[(V(Z))^t\right].$$

Paulin、Mackey 和 Tropp [188], 通过可交换对方法，得到了上述不等式一个优美的矩阵推广（下文的定理 2.1.1）（另见 [114]，其中给出了另一种证明）。他们的不等式适用于取值为厄米矩阵的函数$\mathbf{H}$。不过，我们也可以将其用于非厄米函数$\mathbf{F}$，只需将其应用于厄米扩张$\mathbf{H} = \begin{bmatrix} 0 & \mathbf{F} \\ \mathbf{F} & 0 \end{bmatrix}$ 即可。

T

定理 2.1.1 ([188])。 设 $\mathbf{H}(Z)$ 是一个以独立随机变量 $Z=(Z_1,\dots,Z_n)$ 为自变量、取值为厄米矩阵的函数，且满足 $\mathbb{E}\,\|\mathbf{H}\|<\infty$。那么，对于每个自然数 $t\ge 1$,

$$\mathbb{E}\,\mathrm{tr}\left[(\mathbf{H}-\mathbb{E}\,\mathbf{H})^{2t}\right] \;\le\; (4t-2)^t\cdot\mathbb{E}\,\mathrm{tr}\left[\mathbf{V}^t\right],$$

其中，$\mathbf{V}(Z)$是定义为如下的方差代理量

$$\mathbf{V}(Z):=\frac{1}{2}\cdot\sum_{i=1}^{n}\mathbb{E}\left[\left(\mathbf{H}(Z)-\mathbf{H}\left(Z^{(i)}\right)\right)^2\mid Z\right].$$

拉ademacher变量的一个简单界。 方差代理的形式表明，可以对拉ademacher变量的多项式函数（例如，度为 d_p的函数）采用递归方法。特别是，我们再次考虑标量情形，即 Boucheron 等人提出的 Efron-Stein 不等式

等。 [35], 其中，方差代理量可以写成

$$V(Z) \;=\; \frac{1}{2}\cdot\sum_{i\in[n]}\mathbb{E}\left[\left(f(Z)-f\left(Z^{(i)}\right)\right)^2 \mid Z\right] \;=\; \frac{1}{2}\cdot\sum_{i\in[n]}\mathbb{E}\left[(Z_i-\widetilde{Z}_i)^2\cdot\left(\frac{\partial f(Z)}{\partial Z_i}\right)^2 \mid Z\right]$$
$$= \sum_{i\in[n]}\left(\frac{\partial f(Z)}{\partial Z_i}\right)^2 \;=\; \|\mathbf{f}_1(Z)\|_2^2\,,$$

其中，f1(Z) 是一个向量值函数，定义为 $\mathbf{f}_1[i](Z) \;=\; {}^{\partial f}(Z)\,{}_{\partial Z_i}$ 。因此，要估计$\mathbb{E}\,(f(Z))^{2t}$，我们只需估计 $\mathbb{E}\,\|$f1(Z)$\|_2^{2t}$，此时 f1(Z) 是一个向量值函数。关键观察是，f1(Z) 的矩阵项的度至多为 d_p-1。这表明，我们可以递归地应用该不等式，直到得到常数多项式，而常数多项式是我们完全理解的。对于矩阵值函数 F(Z)，我们也可以利用定理 2.1.1进行类似的计算。这会得到两个偏导数矩阵 F0,1 和 F1,0，其中额外的索引 i 要么加到行索引中，要么加到列索引中。迭代这一过程，可得以下结果；我们用偏导算子 $\nabla_\alpha(f) \;=\; \left(\prod_i{}_{\alpha_i=1}\frac{\partial}{\partial Z_i}\right)(f)$ 来表述该结果，其中$\alpha\in\{0,1\}^n$ （按矩阵的各矩阵项扩展）。

—

定理 2.1.2 (拉德马赫递归**)**。 设 $\mathbf{F}:\ \{-1,1\}^n \to \mathbb{R}^{\mathcal{I}\times\mathcal{J}}$ 是一个次数至多为 d_p 的矩阵值多项式函数。那么，对于每个自然数 $t\geq 1$,

$$\mathbb{E}\,\|\mathbf{F}-\mathbb{E}\,\mathbf{F}\|_{2t}^{2t} \;\leq\; \sum_{1\leq a+b\leq d_p}(16td_p)^{(a+b)\cdot t}\cdot\left\|\mathbb{E}\,\mathbf{F}_{a,b}\right\|_{2t}^{2t}\,,$$

其中 $\mathbf{F}_{a,\,b}$ 是一个由集合 $\mathcal{I}\times\{0,1\}^n$ 和 $\mathcal{J}\times\{0,1\}^n$ 索引的偏导数矩阵，其中

$$\mathbf{F}_{a,b}[(\cdot,\alpha),(\cdot,\beta)] = \begin{cases}\nabla_{\alpha+\beta}(\mathbf{F}) & \text{if } |\alpha|=a, |\beta|=b, \alpha\cdot\beta=0\\ 0 & \text{otherwise}\end{cases}$$

与标量情形的超收缩界类似，上述界由少量（$O\left(d_p^2\right)$ ）矩阵给出，这些矩阵源自固定度（不随 O 增长）的多项式

，但重要的是，它们是确定性矩阵。由于它们是确定性的，分析起来要容易得多。当我们将此定理应用于情形$\mathbf{F} = \mathbf{M}_\tau$（形状 τ 的图矩阵）时，得到的界以组合对象表示，这些对象称为形状 τ 的“顶点分隔集”。这恢复了 Ahn 等人 [1] 给出的界，而且或许令作者惊讶的是，与 Ahn 等人 [1]提出的巧妙观察相比，这为顶点分隔集等组合结构提供了一种直接的替代推导。我们在第2.4节介绍这一点以及 Rademacher 框架的其他应用。

将框架扩展到一般乘积分布。我们工作的一个重要贡献是展示了如何将上述框架扩展到任意乘积分布（具有有界矩）。一个推动性示例是所谓“稀疏图矩阵”的范数界。在稀疏图矩阵中，变量 Z_i 可以被视为一个 $\mathcal{G}_{n',\ p}$ 图的（归一化）边，也就是说，$Z_i = -\sqrt{\frac{1-p}{p}}$以概率 p取值，并以概率 $1-p$取值。这些变量是 p-偏置傅里叶分析中的标准变量，[175] 并且被选取为满足 $\mathbb{E}\, Z_i = 0$ 和 $\mathbb{E}\, Z_i^2 = 1$。在分析 $\mathcal{G}_{n,\ p}$ 图上 $p = o(1)$ 的平均情形问题时，稀疏图矩阵自然出现，而不是分析$\mathcal{G}_{n,\ 1/2}$ 图。

直到最近，人们对稀疏图矩阵的范数界还知之甚少。困难部分源于这样一个事实：当 $p = o(1)$ 时，稀疏图矩阵的范数界必须正确依赖于 p，而不仅仅依赖于 n，这一点非常重要。Jones 等人最近通过迹幂方法得到了这类范数界，[119], 该方法涉及精细的组合计数论证。另一方面，我们使用自己的框架也能得到类似的范数界，而且过程更加机械化。我们还可以直接将框架应用于更一般的次高斯随机变量情形，此时我们的界将取决于这些分布的次高斯范数。

为了扩展我们的 对于一般乘积分布的框架，我们可以从拉德马赫情形中汲取灵感

并尝试递归地应用埃夫隆-斯坦不等式。不幸的是，这个想法行不通。再次考虑标量情形，就能看出问题所在。假设 $Z_1, \dots, Z_n$ 是独立同分布的，且对所有 $i \in [n]$， $\mathbb{E}\, Z_i = 0$ 和 $\mathbb{E}\, Z_i^2 = 1$ 成立。为简单起见，还假设 $f(Z)$是一个度为 d_p的多线性多项式。像之前一样分析方差代理量，可得

$$V(Z) \;=\; \frac{1}{2} \cdot \sum_{i \in [n]} \mathbb{E}\left[(Z_i - \widetilde{Z}_i)^2 \cdot \left(\frac{\partial f(Z)}{\partial Z_i} \right)^2 \mid Z \right] \;=\; \frac{1}{2} \sum_{i \in [n]} \mathbb{E}\left[(Z_i - \widetilde{Z}_i)^2 | Z \right] \cdot \left(\frac{\partial f(Z)}{\partial Z_i} \right)^2 .$$

在拉德马赫情形中，我们有 $\mathbb{E}[(Z_i - \widetilde{Z}_i)^2 | Z] = 2$。这使我们得到与偏导数对应的多项式，而重要的是，这些多项式的度严格更低。然而，对于一般的乘积分布，我们得到的则是 $\mathbb{E}[(Z_i - \widetilde{Z}_i)^2 | Z] = 1 + Z_i^2$。这又带回了一个项 $\left(Z_i \cdot \frac{\partial f}{\partial Z_i} \right)^2$，其中方形内的多项式的度可能仍等于 d_p。这意味着在递归的下一步中，我们可能又要考虑关于 Z_i 的导数，并且可能再次得到相同的多项式 f。因此，递归陷入停滞！矩阵也会出现类似问题，详见第2.5节。为了解决这一问题，我们推广了 [188] 的工作。

通过显式内核进行推广 [188] 。 为了解决上述问题，我们修改了 [188] 的证明，而我们的证明技巧本身也可能具有独立的研究价值。

我们首先回顾矩阵 Efron-Stein 不等式（定理 2.1.1）在[188]中的证明方法。他们的基本策略是利用可交换对 [219, 220, 44, 45], 的理论，尤其是核斯坦对。核斯坦对是具有“核”的随机矩阵可交换对，其中的“核”是一个二元函数，能够“再现”该对中的矩阵。更具体地说，考虑一对随机变量（Z， Z'），这意味着（Z'， Z）具有相同的分布。对于这对可交换对，如果二元矩阵值函数$\mathbf{K}$（z， z'）满足以下条件，则称其为矩阵值函数$\mathbf{F}$ 的核:

- 反对称性：$\mathbf{K}(z', z) = -\mathbf{K}(z, z')$ 对所有输入 (z, z') 均成立。

- 再生性质： $\mathbb{E}[\mathbf{K}(Z, Z') \mid Z] = \mathbf{F}(Z)$。

若存在这样的核$\mathbf{K}$ ，则称随机变量对$\mathbf{F}(Z)$，$\mathbf{F}(Z')$为核斯坦对。

在 [220, 44], Paulin、Mackey 和 Tropp [188] 的思想基础上，他们首先通过将其表示为耦合马尔可夫链的极限，证明核的存在性。通过研究该核耦合的演化过程，他们证明了核的解析性质。随后，他们利用这个核，运用强大的可交换对方法来计算随机矩阵的矩，进而推出集中性。

对于厄米随机矩阵$\mathbf{X}$，他们引入了两个矩阵——条件方差$\mathbf{V_X}$ ，它衡量对 Z的一个坐标重新抽样时$\mathbf{X}$ 的平方波动；以及核条件方差 $\mathbf{V_K}$ ，它衡量对 Z的一个坐标重新抽样时核的平方波动。有了这两个矩阵，他们便可对任意参数 $s > 0$，用$\mathbf{V_X} + s^{-1}\mathbf{V_K}$ 的 Schatten$2t$范数来界定$\mathbf{X}$ 的 Schatten t范数。最后，他们适当选择 s ，使这两个量近似相等，此时问题便简化为方差代理量$\mathbf{V}$，从而证明定理 2.1.1。

在我们的设定中，这样的 s 选择不可行，因为无论如何选择 s，条件方差项 $s\mathbf{V_X}$ 都会支配 $\mathbf{X}^2$ ，或者核条件方差项$s^{-1}\mathbf{V_K}$ 会支配 $\mathbf{X}^2$。这会使主不等式定理 2.1.1 变得平凡。

为了解决这个问题，我们将利用现有矩阵的结构，即 $\mathbf{F} = \mathbf{DGD}$，其中 $\mathbf{D}$ 是一个对角矩阵，它编码了所有已经进行过微分的变量，而 $\mathbf{G}$ 是由其余变量构成的多项式矩阵。由于 $\mathbf{D}$ 是一个度数较低的简单对角矩阵，$\mathbf{F}$ 所表现出的大多数偏差实际上很可能也会出现在$\mathbf{G}$中。为了捕捉这一直觉，我们只考虑内层矩阵 $\mathbf{G}$的核，而不是整个$\mathbf{F}$ 的核。我们称之为内核。

这有助于我们避免问题的根本原因，即对已经遇到过的变量求导（这些变量对应于$\mathbf{D}$中的条目）。因此，递归不会

停滞!

然而，一般而言，这并不可实现，因为$\mathbf{D}$ 与$\mathbf{G}$ 的核可能会以出乎意料的方式相互作用。为研究这种相互作用，我们构造了显式的多项式核（定理 2.7.3）（相较于 [188]，后者证明了核对所有函数的存在性）。

我们研究这一显式内核如何与$\mathbf{D}$ 相互作用（见引理 2.7.6），并利用它得到 [188] 所提出不等式的一个推广（之所以称为推广，是因为令$\mathbf{D} = \mathbf{I}$ 即可恢复他们的结果），该推广陈述于引理 2.7.10 中。

一个微妙的问题是，$\mathbf{X}$ 的条件方差仍可能由于对角矩阵$\mathbf{D}$而产生额外偏差（这些矩阵仍然涉及随机变量）。我们使用詹森算子迹不等式（适用于不交换的平均） [96]来控制这些额外偏差（该不等式陈述于引理 2.2.4 中）。综合这些思想，我们可以得到 Efron-Stein 不等式的一个版本，其中方差代理量仅对应于内核的条件方差。在多项式函数的情形下，该不等式推广了 [188] 的工作。

利用上述修正后的 Efron-Stein 不等式，我们无法保证中间步骤中的矩阵$\mathbf{F}$ 度数较低；但另一方面，内层矩阵$\mathbf{G}$ 的度数会在每一步降低。因此，我们可以递归地应用该不等式，得到最终的界。最终的界以简化矩阵$\mathbf{DGD}$ 的范数界形式给出，其中$\mathbf{G}$ 是确定性矩阵，$\mathbf{D}$ 是仍为 Z 的函数的对角矩阵。尽管这些矩阵是随机的，但利用简单的标量集中工具便可轻松分析。

主定理在第 2.6 节中陈述，特别是定理 2.6.6；证明则见第 2.7 节。虽然我们的证明建立在 [188]，的工作之上，但这里的论证是自成一体的。

应用。我们的框架适用于文献中得到的许多非线性集中结果 [22, 85, 107, 163, 1, 106, 211, 100, 105, 167, 119]。我们将在以下几项

应用中展示这一点，这些应用见第2.4节和第 2.8 节。我们预计，未来类似的应用也会受益于我们的框架，因为它将任务机械地简化为分析简单得多的矩阵。

在第 2.4.2 节中，我们推导了稠密图矩阵的范数界。在早期的工作中，稠密图矩阵已被广泛用于半定规划层级的分析，尤其是平方和（SoS）层级
[164, 59, 103, 197, 20, 165, 86, 191]。关于更多应用以及对图矩阵的详细讨论，请参见 [1]。

在第 2.8 节中，我们推导了稀疏图矩阵的范数界。直到最近，人们对稀疏图矩阵的理解一直相对有限； [119] 通过迹幂方法获得了此类矩阵的范数界。他们利用这些界证明了稀疏图上最大独立集问题的 SoS 下界。

潜在的扩展 在这项工作中，我们假设输入服从乘积分布。换句话说，变量 Z_1、.....、Z_n 彼此独立。一个自然的扩展是考虑它们不独立的情形。当输入为均匀 d-正则图，或输入从具有全局约束的分布中采样时，这一扩展对许多问题都有重要应用，等等。在此类情形下，输入变量并不独立，但仍有可能利用类似的思想来分析集中性。

更具体地说，要研究非独立情形下的集中性，可以结合我们的框架，采用 Huang 和 Tropp [114] 近期关于由 Poincar´e 不等式推导矩阵集中的研究。为此，我们只需构造一个收敛到目标分布的马尔可夫过程。

本章结构 我们先在第2.2节介绍预备知识。在第2.3节，我们陈述并证明拉德马赫递归。在第2.4节，我们展示这一框架的一些应用。在第2.5节，我们解释为什么类似的想法在一般情形下可能不足以奏效。随后，我们在第 2.6 节提出一般框架，并在

第 2.7 节证明该框架。最后，我们在第 2.8 节将一般框架应用于稀疏图矩阵。

2.2 预备知识

记号 我们使用粗体字母（如 $\mathbf{I}$、$\mathbf{M}$、$\mathbf{X}$.....）表示矩阵。矩阵 $\mathbf{X} \in \mathbb{R}^{\mathcal{I}\times\mathcal{J}}$的矩阵项记为 $\mathbf{X}[I,J]$，其中 $I \in \mathcal{I}$，$J \in \mathcal{J}$。令 $\mathbb{H}^n$ 表示$n \times n$ 实对称矩阵的集合。矩阵 $\mathbf{X} \in \mathbb{H}^n$ 的迹等于 $\sum_{i\in[n]} \mathbf{X}[i,i]$，记作 $\mathbf{tr}\mathbf{X}$。

多重指标记法

对于任意一对向量 α、$\beta \in \mathbb{N}^n$ 和标量 $c \in \mathbb{N}$，我们逐项定义 $\alpha+\beta$、$\alpha\cdot\beta$、$c\alpha$。我们还定义序 $\alpha \le \beta$ 和 $\alpha \trianglelefteq \beta$：如果对于每个 i都有 $\alpha_i \le \beta_i$，则称 $\alpha \le \beta$；如果对于每个 i，α_i 只能是 0 或 β_i，则称$\alpha \trianglelefteq \beta$。我们用 $|\alpha|_0$ 表示α 的非零项数量，用 $|\alpha|_1$表示 α各矩阵项之和。对于布尔向量 $\gamma \in \{0,1\}^n$，我们定义 $1-\gamma$为将其所有比特翻转后的向量。

导数

对于变量 Z_1、...、Z_n和 $\alpha \in \mathbb{N}^n$，定义单项式 $Z^\alpha := \prod_{i=1}^n Z_i^{\alpha_i}$。这构成了多项式的标准基。

对于 $\alpha \in \mathbb{N}^n$，我们定义线性算子 ∇_α，其作用于多项式的方式是先定义它对元素 Z^β 的作用如下，然后再将其线性扩展到所有多项式。

$$\nabla_\alpha(Z^\beta) = \begin{cases} Z^{\beta-\alpha} & \text{if } \alpha \trianglelefteq \beta \\ 0 & \text{o.w.} \end{cases}$$

通俗地说，对于一个写成标准基多项式线性组合的多项式 f

多项式 Z^β, $\nabla_\alpha(f)$ 会筛选出恰好包含幂 $Z_i^{\alpha_i}$ 的项，其中 i 满足 $\alpha_i \neq 0$，然后截断这些幂。换句话说，它是 f 中 Z^α 的系数。特别要注意，对于任何满足 $\alpha_i \neq 0$ 的 i， $\nabla_\alpha(f)$ 都不依赖于 Z_i。

假设 f 是多线性的；在拉德马赫情形下，当我们处理 $Z_i \in \{-1, 1\}$时，可以这样假设。对于非零指标为 i_1、.....、$i_k \in [n]$的 $\alpha \in \{0,1\}^n$，有 $\nabla_\alpha(f) = \frac{\partial}{\partial Z_{i_1}} \cdots \frac{\partial}{\partial Z_{i_k}} f$。因此，这个线性算子推广了偏导算子。但请注意，一般而言， ∇ 并不只是标准的偏导算子。

矩阵分析

作用于多项式的线性算子，也可以自然地定义为通过作用于矩阵的每个矩阵项来作用于矩阵。

我们将 $\mathbf{I}_m$定义为 $m \times m$单位矩阵。若上下文明确，则省略下标。对于矩阵 $\mathbf{F}$、$\mathbf{G}$，定义 $\mathbf{F} \oplus \mathbf{G}$为矩阵 $\begin{bmatrix} 0 & \\ & \mathbf{F}\,\mathbf{G}0 \end{bmatrix}$。对于矩阵 $\mathbf{F}$，将其厄米扩张定义为$\mathbf{F} \oplus \mathbf{F}^T$。用 $\preceq$表示勒文纳序，也就是说，对于$\mathbf{A}$、$\mathbf{B} \in \mathbb{H}^n$，$\mathbf{A} \preceq \mathbf{B}$当且仅当$\mathbf{B} - \mathbf{A}$是半正定矩阵。

定义 2.2 .1. 对于矩阵 $\mathbf{F}$ 和整数 $t \geq 0$，定义 *Schatten* $2t$-范数为

$$\|\mathbf{F}\|_{2t}^{2t} = \mathrm{tr}[(\mathbf{F}\mathbf{F}^T)^t]$$

事实 2.2.2. 对于实对称矩阵 $\mathbf{X}1$，...，$\mathbf{X}n$，有

$$(\mathbf{X}_1 + \ldots + \mathbf{X}_n)^2 \preceq n(\mathbf{X}_1^2 + \ldots + \mathbf{X}_n^2)$$

事实 2.2.3. 对于半正定矩阵$\mathbf{X}$、$\mathbf{X}1$.....$\mathbf{X}n$，满足$\mathbf{X} \preceq \mathbf{X}_1 + \ldots + \mathbf{X}n$

且对于任意整数 $t \geq 1$,

$$\mathrm{tr}[\mathbf{X}^t] \leq n^{t-1}(\mathrm{tr}[\mathbf{X}_1^t] + \ldots + \mathrm{tr}[\mathbf{X}_n^t])$$

证明。 根据 Hölder 不等式， $n^{t-1}(\mathrm{tr}[\mathbf{X}_1^t] + \ldots + \mathrm{tr}[\mathbf{X}_n^t]) \geq (\|\mathbf{X}_1\|_t + \ldots + \|\mathbf{X}_n\|_t)^t$。根据 Schatten 范数的三角不等式，这至少为 $\|\mathbf{X}_1 + \ldots + \mathbf{X}_n\|_t^t$。最后，由于$\mathbf{X}_1 + \ldots + \mathbf{X}_n \succeq \mathbf{X} \succeq 0$，我们可以利用迹函数的单调性（见 [190, 命题1]），其中我们在 $x \in [0, \infty$ 上使用递增函数 $f(x) = x^t$。由此得证。

■

引理 2.2.4 (詹森算子迹不等式)。 *[96, 推论 2.5]* 设 f 是定义在区间 I 上的凸连续函数，并假设 $0 \in I$ 且 $f(0) \leq 0$。那么，对于所有整数 m、$n \geq 1$，对每个由实对称 $m \times m$ 矩阵组成的元组 $\mathbf{B}1$、.....、$\mathbf{B}n$ （其谱包含于 I），以及每个由满足 $\sum_{i=1}^n \mathbf{A}_i^T \mathbf{A}_i \preceq \mathbf{I}$ 的 $m \times m$ 矩阵组成的元组 $\mathbf{A}1$、.....、$\mathbf{A}n$，我们有

$$\mathrm{tr}[f(\sum_{i=1}^n \mathbf{A}_i^T \mathbf{B}_i \mathbf{A}_i)] \leq \mathrm{tr}[\sum_{i=1}^n \mathbf{A}_i^T f(\mathbf{B}_i) \mathbf{A}_i]$$

2.3 拉ademacher随机变量的基本框架

令 $Z = (Z_1, \ldots\ldots, Z_n)$ 从 $\{-1, 1\}^n$ 中均匀抽样。我们将考虑矩阵值函数 $\mathbf{F}$: $\{-1, 1\}^n \to \mathbb{R}^{\mathcal{I} \times \mathcal{J}}$，其行和列分别由任意集合 $\mathcal{I}$， $\mathcal{J}$ 索引，使得对于所有 $I \in \mathcal{I}$, $J \in \mathcal{J}$,

$$\mathbf{F}[I, J] \ = \ f_{I,J}(Z)$$

其中 $f_{I,\ J}$是关于 Z_1、.....、 Z_n的多项式。由于 $Z_i \in \{-1, 1\}$，不失一般性，我们可以假设 $f_{I,\ J}$ 是多线性的。令 d_p 为 $\mathbf{F}$ 中任意一个 $f_{I,\ J}$ 的最高次数。在本节中，我们将给出一个一般框架，借助它可以得到任意整数 $t \geq 1$下 $\mathbb{E}\,\|\mathbf{F} - \mathbb{E}\,\mathbf{F}\|_{2t}^{2t}$的界。

定理 2.1.2 (拉德马赫递归)。 设 $\mathbf{F}: \{-1,1\}^n \to \mathbb{R}^{\mathcal{I}\times\mathcal{J}}$ 是一个次数至多为 d_p的矩阵值多项式函数。那么，对每个自然数 $t \geq 1$,

$$\mathbb{E}\,\|\mathbf{F}-\mathbb{E}\,\mathbf{F}\|_{2t}^{2t} \;\leq \sum_{1\leq a+b\leq d_p} (16td_p)^{(a+b)\cdot t}\cdot \left\|\mathbb{E}\,\mathbf{F}_{a,b}\right\|_{2t}^{2t},$$

其中 $\mathbf{F}_{a,b}$ 是一个偏导数矩阵，其行和列分别由集合 $\mathcal{I}\times\{0,1\}^n$ 和 $\mathcal{J}\times\{0,1\}^n$索引，并且

$$\mathbf{F}_{a,b}[(\cdot,\alpha),(\cdot,\beta)] = \begin{cases} \nabla_{\alpha+\beta}(\mathbf{F}) & \text{if } |\alpha|=a, |\beta|=b, \alpha\cdot\beta=0 \\ 0 & \text{otherwise} \end{cases}$$

注记 2.3.1。 注意，虽然上文为方便起见，将矩阵 $\mathbf{F}_{a,b}$ 表述为其行和列分别由 $\mathcal{I}\times\{0,1\}^n$和 $\mathcal{J}\times\{0,1\}^n$ 索引，但我们只需考虑具有 $|\mathcal{I}|\cdot\binom{n}{a}$ 行和 $|\mathcal{J}|\cdot\binom{n}{b}$ 列的子矩阵，因为所有其他元素都将为零（当 $|\alpha|\neq a$ 或 $|\beta|\neq b$时）。

注记 2.3.2。 为了从矩估计中获得高概率范数界，我们可以令$t = \mathrm{polylog}(n)$，并应用马尔可夫不等式。由于我们不试图优化对数因子的依赖关系，因此也不试图优化主定理中 t 的指数。

为证明这一点，我们将证明引理 2.3.3，然后递归地应用它。

对于每个 $i \leq n$，定义随机向量

$$Z^{(i)} := (Z_1,\ldots,Z_{i-1},\widetilde{Z}_i,Z_{i+1},\ldots,Z_n)$$

~其中 Z_i 是 Z_i 的一个独立副本，也就是说，它是从 $\{-1,1\}$ 独立地重新采样得到的。

设 $\mathbf{X} := \mathbf{F}-\mathbb{E}\mathbf{F}$。当输入为 Z 时，我们将这些矩阵记为 F、X 等；当输入为 $Z^{(i)}$ 时，则将相应的矩阵记为 $\mathbf{F}^{(i)}$、$\mathbf{X}^{(i)}$ 等。也就是说，对于 $I\in\mathcal{I}, J\in\mathcal{J}$，有 $\mathbf{F}^{(i)}[I,J] = f_{I,J}(Z^{(i)})$。定义 $\mathbf{X}_{a,b} = \mathbf{F}_{a,b} - \mathbb{E}\,\mathbf{F}_{a,b}$。

引理 2.3.3。 对于整数 $a, b \geq 0$，有

$$\mathbb{E}\left\|\mathbf{X}_{a,b}\right\|_{2t}^{2t} \leq (16td_p)^t(\mathbb{E}\left\|\mathbf{X}_{a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\mathbf{X}_{a+1,b}\right\|_{2t}^{2t} + \left\|\mathbb{E}\,\mathbf{F}_{a,b+1}\right\|_{2t}^{2t} + \left\|\mathbb{E}\,\mathbf{F}_{a+1,b}\right\|_{2t}^{2t})$$

利用这个引理，我们可以完成主定理的证明。

定理 *2.1.2* 的证明。 注意到 $\mathbf{X}$ 是$\mathbf{X}_{0,0}$ 的一个主子矩阵，其余矩阵项均为 0，因此我们可以反复应用引理 2.3.3，直到$\mathbf{X}_{a,b} = 0$，当 $a + b > d_p$ 时即可达到这一情形。

■

在本节余下部分，我们将证明引理 2.3.3。我们先介绍一个基本事实。令$\mathbf{e}_i \in \{0,1\}^n$ 为只有一个非零项的向量（$\mathbf{e}_i$）$_i = 1$。

命题 2.3.4。 对于多线性多项式 $f(Z) = f(Z_1, \ldots, Z_n)$，有

$$f(Z) - f(Z^{(i)}) \;=\; (Z_i - \widetilde{Z}_i) \cdot \nabla_{\mathbf{e}_i} f(Z)$$

引理 *2.3.3* 的证明。 考虑厄米扩张 $\mathbf{F}_{a,b} = \mathbf{F}_{a,b} \oplus \mathbf{F}_{a,b}^T$. 定义 $\mathbf{X}_{a,b} = \mathbf{F}_{a,b} - \mathbb{E}\mathbf{F}_{a,b} = \mathbf{X}_{\overline{a,b}} \oplus \mathbf{X}_{a,b}^T$. 根据定理 2.1.1 应用于$\mathbf{X}_{a,b}$,

$$\mathbb{E}\,\mathrm{tr}\left[\overline{\mathbf{X}}_{a,b}^{2t}\right] \leq (2(2t-1))^t\,\mathbb{E}\,\mathrm{tr}\left[\mathbf{V}_{a,b}^t\right]$$

其中 $\mathbf{V}_{a,\ b}$ 是方差代理量

$$\mathbf{V}_{a,b} = \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}[(\overline{\mathbf{X}}_{a,b} - \overline{\mathbf{X}}_{a,b}^{(i)})^2 | Z]$$

首先，通过简单的计算,

$$\mathbb{E}\,\mathrm{tr}\left[\overline{\mathbf{X}}_{a,b}^{2t}\right] = \mathbb{E}\,\mathrm{tr}\left[(\mathbf{X}_{a,b}\mathbf{X}_{a,b}^{\intercal})^t\right] + \mathbb{E}\,\mathrm{tr}\left[(\mathbf{X}_{a,b}^{\intercal}\mathbf{X}_{a,b})^t\right] = 2\,\mathbb{E}\left\|\mathbf{X}_{a,b}\right\|_{2t}^{2t}$$

and

$$\begin{aligned}\mathbf{V}_{a,b} &= \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}[(\overline{\mathbf{X}}_{a,b}-\overline{\mathbf{X}}_{a,b}^{(i)})^2|Z] \\ &= \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}\left[\begin{bmatrix}(\mathbf{X}_{a,b}-\mathbf{X}_{a,b}^{(i)})(\mathbf{X}_{a,b}-\mathbf{X}_{a,b}^{(i)})^\intercal & 0 \\ 0 & (\mathbf{X}_{a,b}-\mathbf{X}_{a,b}^{(i)})^\intercal(\mathbf{X}_{a,b}-\mathbf{X}_{a,b}^{(i)})\end{bmatrix}|Z\right] \\ &= \frac{1}{2}\begin{bmatrix}\sum_{i=1}^{n}\mathbb{E}[(\mathbf{F}_{a,b}-\mathbf{F}_{a,b}^{(i)})(\mathbf{F}_{a,b}-\mathbf{F}_{a,b}^{(i)})^\intercal|Z] & 0 \\ 0 & \sum_{i=1}^{n}\mathbb{E}[(\mathbf{F}_{a,b}-\mathbf{F}_{a,b}^{(i)})^\intercal(\mathbf{F}_{a,b}-\mathbf{F}_{a,b}^{(i)})|Z]\end{bmatrix}\end{aligned}$$

我们将使用以下结论，并在稍后证明它。

命题 2.3.5. 我们有以下关系。

$$\sum_{i=1}^{n}\mathbb{E}[(\mathbf{F}_{a,b}-\mathbf{F}_{a,b}^{(i)})(\mathbf{F}_{a,b}-\mathbf{F}_{a,b}^{(i)})^\intercal|Z] = 2(b+1)\mathbf{F}_{a,b+1}\mathbf{F}_{a,b+1}^\intercal$$

$$\sum_{i=1}^{n}\mathbb{E}[(\mathbf{F}_{a,b}-\mathbf{F}_{a,b}^{(i)})^\intercal(\mathbf{F}_{a,b}-\mathbf{F}_{a,b}^{(i)})|Z] = 2(a+1)\mathbf{F}_{a+1,b}^\intercal\mathbf{F}_{a+1,b}$$

由此得出 $\mathbb{E}\mathrm{tr}\left[\mathbf{V}_{a,b}^t\right] = (b+1)^t\,\mathbb{E}\left\|\mathbf{F}_{a,b+1}\right\|_{2t}^{2t} + (a+1)^t\,\mathbb{E}\left\|\mathbf{F}_{a+1,b}\right\|_{2t}^{2t}$。因此，我们得到

$$\begin{aligned}2\,\mathbb{E}\,\|\mathbf{X}_{a,\ b}\|_{2t}^{2t} &= \mathbb{E}\,\mathrm{tr}\left[\underline{\mathbf{X}_{a}}^{2t}{}_{,}\right. \\ &\left._{b}\right] \\ &\le (2(2t-1))^t\,\mathbb{E}\,\mathrm{tr}\left[\mathbf{V}_{a,b}^t\right] \\ &\le (2(2t-1))^t((b+1)^t\,\mathbb{E}\,\|\mathbf{F}_{a,b+1}\|_{2t}^{2t} + (a+1)^t\,\mathbb{E}\,\|\mathbf{F}_{a+1,b}\|_{2t}^{2t}) \\ &\le (2(2t-1))^t((b+1)^t\,\mathbb{E}\,\|\mathbf{X}_{a,b+1}+\mathbb{E}\,\mathbf{F}_{a,b+1}\|_{2t}^{2t} + (a+1)^t\,\mathbb{E}\,\|\mathbf{X}_{a+1,b}+\mathbb{E}\,\mathbf{F}_{a+1,b}\|_{2t}^{2t}) \\ &\le (16t)^t((b+1)^t(\mathbb{E}\,\|\mathbf{X}_{a,b+1}\|_{2t}^{2t} + \|\mathbb{E}\,\mathbf{F}_{a,b+1}\|_{2t}^{2t}) + (a+1)^t(\mathbb{E}\,\|\mathbf{X}_{a+1,b}\|_{2t}^{2t} + \|\mathbb{E}\,\mathbf{F}_{a+1,b}\|_{2t}^{2t}) \\ &\le (16td_p)^t(\mathbb{E}\,\|\mathbf{X}_{a,b+1}\|_{2t}^{2t} + \|\mathbb{E}\,\mathbf{F}_{a,b+1}\|_{2t}^{2t} + \mathbb{E}\,\|\mathbf{X}_{a+1,b}\|_{2t}^{2t} + \|\mathbb{E}\,\mathbf{F}_{a+1,b}\|_{2t}^{2t})\end{aligned}$$

■

接下来证明该命题。

命题 *2.3.5* 的证明。我们将证明第一个等式。第二个等式的证明类似。对于$I \in \mathcal{I}$，$J \in \mathcal{J}$，α，$\beta \in \{0,1\}^n$，我们有

$$(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)})[(I,\alpha),(J,\beta)] = \begin{cases} \nabla_{\alpha+\beta}(f_{I,J}(Z) - f_{I,J}(Z^{(i)})) & \text{if } |\alpha|_0 = a, |\beta|_0 = b, \alpha \cdot \beta = 0 \\ 0 & \text{o.w.} \end{cases}$$

根据命题 2.3.4，第一个表达式可化简为 $(Z_i - Z_i')\nabla_{\mathbf{e}_i}\nabla_{\alpha+\beta} f_{I,J}(Z)$。定义矩阵 $\mathbf{F}_{a,b,i}$ 使其与 $\mathbf{F}_{a,b}$ 具有相同的行和列，且其唯一的非零项为

$$\mathbf{F}_{a,b,i}[(I,\alpha),(J,\beta+\mathbf{e}_i)] = \nabla_{\mathbf{e}_i}\nabla_{\alpha+\beta} f_{I,J}(Z) \text{ if } |\alpha|_0 = a, |\beta|_0 = b, \beta \cdot \mathbf{e}_i = 0, \alpha \cdot (\beta + \mathbf{e}_i) = 0$$

那么，很容易看出 $\sum_{i=1}^n \mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^\intercal = (b+1)\mathbf{F}_{a,b+1}\mathbf{F}_{a,b+1}^\intercal$ 以及 $(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)})(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)})^\intercal = (Z - \widetilde{Z}_i)^2\mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^\intercal$ 后一个等式意味着

$$\mathbb{E}[(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)})(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)})^\intercal | Z] = \mathbb{E}[(Z_i - \widetilde{Z}_i)^2 \mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^\intercal | Z] = 2\mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^\intercal$$

因此，

$$\sum_{i=1}^n \mathbb{E}[(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)})(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)})^\intercal | Z] = 2\sum_{i=1}^n \mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^\intercal = 2(b+1)\mathbf{F}_{a,b+1}\mathbf{F}_{a,b+1}^\intercal$$

■

2.4 应用

为阐释我们的框架，我们将其应用于文献中已有研究的非线性随机矩阵，从而获得集中界。第一个例子是一个简单的张量网络，它出现在一种变体的主成分分析谱算法分析中

（PCA） [107, 100]。第二个应用是为稠密图矩阵 [163, 1]获得范数界。在第二个应用中，范数界由一种称为形状的最小顶点分隔集的组合结构决定。我们将看到这一概念如何在我们的框架下自然出现，而先前推导出此类界的工作采用了迹幂方法，并需要非平凡的组合学洞见。

2.4.1 一个简单的张量网络

我们考虑 [107, 100] 中的以下结果。

引理 2.4.1 ([100], 定理 6.7.1)**.** 设 $c \in \{1, 2\}$，并令 $d \geq 1$ 为一个整数。设$\mathbf{A}_1, \ldots,$ $\mathbf{A}_{n^c}$ 为从 $\{-1, 1\}^{n^d \times n^d}$ 中均匀采样得到的独立同分布随机矩阵。那么，以概率 $1 - O(n^{-100})$,

$$\left\| \sum_{k \leq n^c} \mathbf{A}_k \otimes \mathbf{A}_k - \mathbb{E} \sum_{k \leq n^c} \mathbf{A}_k \otimes \mathbf{A}_k \right\| \leq C\sqrt{d} n^{(2d+c)/2} (\log n)^{1/2}$$

其中 $C > 0$ 是一个绝对常数。

利用我们的框架，我们将证明一个略微放宽的不等式版本，其中$\sqrt{d}(\log\ n)^{1/2}$ 被 $\log n$ 替代。我们指出，我们并未尝试优化主导项 $n^{(2d+c)/2}$ 前面的这些额外因子，因此，更仔细的分析或许能够得到稍好一些的界。

放宽界的证明。 令 i, j-th 项为 $\mathbf{A}_k$ be $a_{k,i,j}$. 令 $\mathbf{F} = \sum_{i \leq n^c} \mathbf{A}_k \otimes \mathbf{A}_k - \mathbb{E} \sum_{i \leq n^c} \mathbf{A}_k \otimes \mathbf{A}_k$ 为变量 $a_{k,i,j}$ 上的随机矩阵，其中 $k \leq n^c, i, j \leq n^d$. 因此$\mathbb{E}\mathbf{F} = 0$，我们要寻找 $\|\mathbf{F}\|$ 的界。矩阵项为

$$\mathbf{F}[(i_1, i_2), (j_1, j_2)] = \begin{cases} \sum_{k \leq n^c} a_{k,i_1,j_1} a_{k,i_2,j_2} & \text{if } (i_1, j_1) \neq (i_2, j_2) \\ 0 & \text{if } (i_1, j_1) = (i_2, j_2) \end{cases}$$

非零项是 2 度齐次多项式。利用定理 2.1.2,

$$\mathbb{E}\,\|\mathbf{F}\|_{2t}^{2t} \leq (32t)^{2t}(\left\|\mathbb{E}\,\mathbf{F}_{2,0}\right\|_{2t}^{2t} + \left\|\mathbb{E}\,\mathbf{F}_{1,1}\right\|_{2t}^{2t} + \left\|\mathbb{E}\,\mathbf{F}_{0,2}\right\|_{2t}^{2t})$$

我们将逐一考察这些项。在接下来的论证中，我们只考虑满足 $(i_1, j_1) \neq (i_2, j_2)$ 的指标 i_1, i_2, j_1, j_2。

1. $\mathbb{E}\,\mathbf{F}_{2,0}$ 在行 $((i_1, i_2)\ \{(k, i_1, j_1), (k, i_2, j_2)\})$ 和列 (j_1, j_2) 上有非零项，且这些项均为 1。置换行和列不会改变 Schatten 范数。因此，我们可以按 k、i_1、i_2 对行分组，并在每组中对行和列里的 j_1、j_2 排序。由此得到一个矩阵，其中有 n^{2d+c} 个维度为 $n^{2d} \times n^{2d}$ 的单位矩阵，上下堆叠在一起。根据定义，很容易计算出该矩阵的 Schatten-$2t$ 范数为 $\left\|\mathbb{E}\mathbf{F}_{2,0}\right\|_{2t}^{2t} = n^{c+4d}n^{t(2d+c)}$。

2. $\mathbb{E}\,\mathbf{F}_{1,1}$ 的非零项位于以下行和列中：行 $((i_1, i_2)\ \{(k, i_1, j_1)\})$ 与列 $((j_1, j_2),\{(k, i_2, j_2)\})$；或行 $((i_1, i_2)\ \{(k, i_2, j_2)\})$ 与列 $((j_1, j_2)\ \{(k, i_1, j_1)\})$，且所有这些项均为 1。因此，我们可以将对应于这 2 个项集合的 $\mathbb{E}\mathbf{F}_{1,1} = \mathbf{A} + \mathbf{B}$ 写出。与前一种情形中的论证相同，我们可以得到 $\|\mathbf{A}\|_{2t}^{2t} = n^{c+4d}n^{t(2d+c)}$，其中我们将行按 k, i_2, j_1 分组；以及 $\|\mathbf{B}\|_{2t}^{2t} = n^{c+4d}n^{t(2d+c)}$，其中我们将行按 k, i_1, j_2分组。因此，$\left\|\mathbb{E}\,\mathbf{F}_{1,1}\right\|_{2t}^{2t} \leq 2^{2t}(\|\mathbf{A}\|_{2t}^{2t} + \|\mathbf{B}\|_{2t}^{2t}) = 2^{2t+1}n^{c+4d}n^{t(2d+c)}$。

3. 情形 $\mathbb{E}\,\mathbf{F}_{0,2}$ 与 $\mathbb{E}\mathbf{F}_{2,0}$相同。

将它们结合起来，$\mathbb{E}\,\|\mathbf{F}\|_{2t}^{2t} \leq (C't)^{2t}n^{c+4d}n^{t(2d+c)}$ 其中 $C' > 0$为绝对常数。现在，我们应用马尔可夫不等式，得到

$$\begin{aligned} Pr[\|\mathbf{F} - \mathbb{E}\,\mathbf{F}\| \geq \theta] \ &\leq\ Pr[\|\mathbf{F} - \mathbb{E}\,\mathbf{F}\|_{2t}^{2t} \geq \theta^{2t}] \ \leq\ \theta^{-2t}\,\mathbb{E}\,\|\mathbf{F} - \mathbb{E}\,\mathbf{F}\|_{2t}^{2t} \\ &\leq\ \theta^{-2t}(C't)^{2t}n^{c+4d}n^{t(2d+c)} \end{aligned}$$

现在令 $\theta = \varepsilon^{-1/(2t)}(C't)n^{(c+4d)/t}n^{(2d+c)/2}$，使该表达式至多为 ε。代入 $\varepsilon = n^{-100}$ 并令 $t =$取对数 n，可得 $\|\mathbf{F} - \mathbb{E}\mathbf{F}\| \leq Cn^{(2d+c)/2}$ 取对数 n 以概率 $1 - n^{-100}$成立，其中 $C > 0$ 为绝对常数。

■

2.4.2 图矩阵

本节首先定义图矩阵，然后展示如何利用我们的框架得到稠密图矩阵的范数界，即 $G \sim \mathcal{G}_{n,\ 1/2}$的情形。处理稀疏图矩阵，即 $G \sim \mathcal{G}_{n,\ p}$ 且 $p = o(1)$ 的情形，可能不适用于我们的基本框架，具体原因将在第2.5节中说明。相反，我们在第 2.6 节提出的一般框架能够很好地处理这种情形，并在第 2.8 节中得到稀疏图矩阵的范数界。

定义

定义 $\mathcal{G}_{n,\ p}$ 为顶点集 $[n]$ 上具有 n 个顶点的 Erd˝os-R´enyi 随机图，其中每条边都以概率 p 独立地出现。用变量 $G_{i,\ j} \in \Omega = \{-\sqrt{\frac{1-p}{p}}, \sqrt{\frac{p}{1-p}}\}$ 对该图进行编码，其中 $-\sqrt{\frac{1-p}{p}}$表示边 $\{i,j\}$ 存在，$\sqrt{\frac{p}{1-p}}$ 表示其不存在，对所有 $1 \leq i,\ j \leq n$均如此。

因此，每个 $G_{i,\ j}$ 对于 $i < j$，都是从 Ω 中抽样得到的，其中 $G_{i,\ j}$ 以概率 p 取值 $-\sqrt{\frac{1-p}{p}}$，否则取值 $\sqrt{\frac{p}{1-p}}$。此处，Ω 已经过归一化，使得$\mathbb{E}_{x\sim\Omega}[x] = 0$，$\mathbb{E}_{x\sim\Omega}[x^2] = 1$。这与 p-偏置傅里叶分析中的标准做法一致。

当 $p = 1/2$ 时，我们处于稠密图矩阵的情形。此时，$\mathcal{G}_{n,\ 1/2}$ 可以看作是从 Ω 中独立且均匀地抽样得到的 $G_{i,\ j},\ i < j = \{-1, 1\}$。

对于边集 $E \subseteq \binom{[n]}{2}$，定义 $G_E := \prod_{e\in E} G_e$。当 $p = 1/2$ 时，G_E 对应于图上函数的傅里叶基。

定义 $\mathcal{I}$ 为 $[n]$的子元组集合，包括空元组。图矩阵的行和列以 $\mathcal{I}$为索引。每个图矩阵都可以简洁地表示为一个附带额外信息的图，这种表示称为形状。

定义 2.4.2 （形状）。 形状是一个元组 $\tau = (V(\tau)\ E(\tau)\ U_\tau, V_\tau)$，其中$(V(\tau)\ E(\tau))$是一个图，且 U_τ、V_τ是顶点的有序子集。

定义 2.4.3 (实现)。 给定一个形状 τ， τ 的一个实现是一个单射

$\varphi : V(\tau) \to [n].$

定义 2.4.4 (图矩阵)。 设 τ 为一个形状。对于 τ，图矩阵$\mathbf{M}_\tau$： $\{\pm 1\}^{\binom{n}{2}} \to \mathbb{R}^{\mathcal{I}\times\mathcal{I}}$定义为一个矩阵值函数，其 I, J 项定义如下。

$$\mathbf{M}_\tau[I,J] := \sum_{\substack{Realization\ \varphi \\ \varphi(U_\tau)=I, \varphi(V_\tau)=J}} G_{\varphi(E(\tau))} = \sum_{\substack{Realization\ \varphi \\ \varphi(U_\tau)=I, \varphi(V_\tau)=J}} \prod_{(u,v)\in E(\tau)} G_{\varphi(u),\varphi(v)}$$

换言之，我们对所有将 U_τ, V_τ分别映射到 I, J的 τ 的实现求和，并且对于每一个这样的实现，都有一个与该实现产生的傅里叶特征相对应的项。

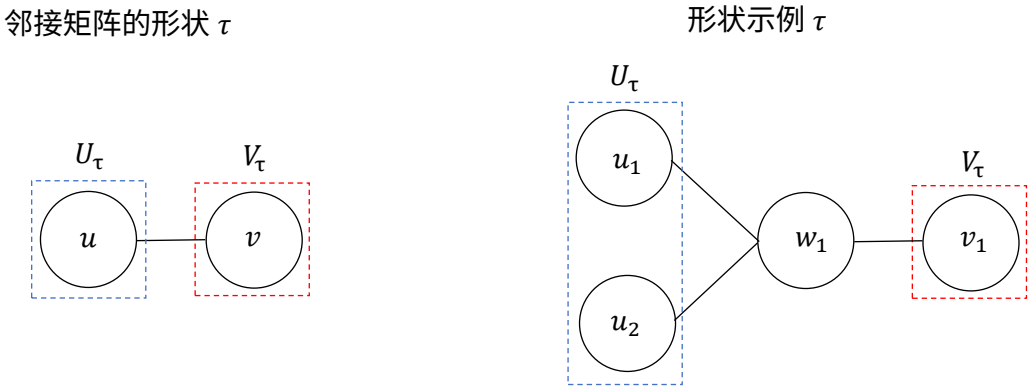


图 2.3：左：邻接矩阵对应的形状；右：更复杂形状的示例

以下示例展示了一些简单图矩阵。

例 2.4.5 （邻接矩阵）。 设 τ 为图 *2.3* 左侧的形状，它有两个顶点 V（τ） $= \{u, v\}$和一条边 E（τ） $= \{\{u, v\}\}$。 U_τ、V_τ分别是（u）、（v）*41*

其中我们用元组表示排序。那么，$\mathbf{M}_\tau$的非零项为$\mathbf{M}_\tau[\,(i)\,、\,(j)\,]\,(G) = G_{i,\,j}$ 对所有 $i \neq j$均成立。如果将 $G \in \{\pm 1\}^{\binom{n}{2}}$ 视为一个图，那么$\mathbf{M}_\tau$ 的主子矩阵是 ± 1 的邻接矩阵，其对角线上为零，其他项均为0。

例 *2.4.6*。在图 *2.3* 中，考虑右侧的形状 τ。我们有 $U_\tau = (u_1, u_2)$ $V_\tau = (v_1)$ $V(\tau) = \{u_1, u_2, v_1, w_1\}$和 $E(\tau) = \{\{u_1, w_1\}, \{u_2, w_1\}, \{w_1, v_1\}\}$。$M_\tau$是一个矩阵，其行和列由 $[n]$ 的子元组索引。它的非零项位于行 I 和列 J 中，分别对应 $|I| = |U_\tau| = 2$ 和 $|J| = |V_\tau| = 1$。更具体地说，对于所有互不相同的 a_1、a_2、b_1，对应于行 (a_1, a_2) 和列 (b_1) 的矩阵项为 $\sum_{c_1 \in [n] \setminus \{a_1, a_2, b_1\}} G_{a_1,c_1} G_{a_2,c_1} G_{c_1,b_1}$。这里，每一项都是通过实现 φ 得到的，该实现将 u_1, u_2, w_1, v_1 分别映射到 a_1, a_2, c_1, b_1。简而言之，

$$\mathbf{M}_\tau = \begin{array}{c} \\ row\ (a_1, a_2) \rightarrow \end{array} \begin{array}{c} column\ (b_1) \\ \downarrow \\ \left(\begin{array}{c} \vdots \\ \cdots\cdots \sum_{c_1 \in [n] \setminus \{a_1, a_2, b_1\}} G_{a_1,c_1} G_{a_2,c_1} G_{c_1,b_1} \cdots\cdots\cdots \\ \vdots \end{array} \right) \end{array}$$

直观地说，图矩阵是傅里叶基的对称化，其中的对称性通过对形状为 τ 的“自由”顶点 $V(\tau) \setminus U_\tau \setminus V_\tau$ 的所有实现求和来体现。有关图矩阵的更多示例，以及它们为何可以成为一种实用工具，请参见 [1]。

稠密图矩阵的范数界

在本节中，我们研究所谓“稠密图矩阵”的集中性，这一术语指的是设置 $p = 1/2$ 中的图矩阵 M_τ。由于随机图的边

从 $\mathcal{G}_{n,\ 1/2}$ 中采样的图可以被视为相互独立的 Rademacher 随机变量，因此我们可以在此设置中应用我们的框架。

特别地，我们将得到关于 $\mathbb{E}\,\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\|_{2t}^{2t}$ 的界。$G_{i,j} \in \{-1,1\}$ 对应于第 2.3 节中的 Z_is；对于固定的形状 τ，$\mathbf{M}_\tau$ 将是我们感兴趣并要分析的矩阵 $\mathbf{F}$ 。对于 $I, J \in \mathcal{I}$，只有当至少存在一个 τ 的实现将 U_τ, V_τ 分别映射到 I, J 时，$\mathbf{M}_\tau[I,J]$ 才是非零多项式。特别地，必须有 $|I| = |U_\tau|$ 和 $|J| = |V_\tau|$。在这种情况下，$\mathbf{M}_\tau[I,J]$ 是一个次数为 $|E(\tau)|$ 的齐次多项式。

根据定理 2.1.2，我们有

$$\mathbb{E}\,\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\|_{2t}^{2t} \;\le\; \sum_{\substack{a+b\ge 1\\ a,b\ge 0}} (16t|E(\tau)|)^{(a+b)t}\, \left\|\mathbb{E}\,\mathbf{M}_{\tau,a,b}\right\|_{2t}^{2t}$$

其中，对于整数 $a, b \ge 0$，$\mathbf{M}_{\tau,a,b}$ 定义为一个矩阵，其行和列均由 $\mathcal{I} \times \{0,1\}^{\binom{n}{2}}$ 编号，使得对于所有 $I, J \in \mathcal{I}$，我们有

$$\mathbf{M}_{\tau,a,b}[(I,\alpha),(J,\beta)] \;=\; \begin{cases} \nabla_{\alpha+\beta}\mathbf{M}_\tau[I,J] & \text{if } |\alpha|_0 = a, |\beta|_0 = b, \alpha\cdot\beta = 0 \\ 0 & \text{o.w.} \end{cases}$$

对于任意次数为 d 的多线性齐次多项式 f，由于 $\mathbb{E}[G_{i,j}] = 0$ 对所有 i,j 都成立，因此当 $|\alpha|_0 < d$ 时，我们有 $\nabla_\alpha f = 0$。所以，对于所有 $a+b < |E(G)|$，$\mathbb{E}\mathbf{M}_{\tau,a,b} = 0$。此外，当 $a+b \ne |E(G)|$ 时，$\mathbb{E}\mathbf{M}_{\tau,a,b} = 0$；否则，$\mathbb{E}\,\mathbf{M}_{\tau,a,b} = \mathbf{M}_{\tau,a,b}$。因此，我们可以进一步将上述表达式简化为

$$\mathbb{E}\,\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\|_{2t}^{2t} \;\le\; \sum_{\substack{a+b=|E(\tau)|\\ a,b\ge 0}} (16t|E(\tau)|)^{|E(\tau)|t}\, \left\|\mathbf{M}_{\tau,a,b}\right\|_{2t}^{2t}$$

接下来需要分析 $\left\|\mathbf{M}_{\tau,\ a,\ b}\right\|_{2t}^{2t}$ 对于 $a+b = |E(G)|$。我们将看到，分析这些矩阵要简单得多，因为它们是确定性矩阵，而且简单的计算

使用 Frobenius 范数界会很有效。为了陈述最终的界，我们需要定义形状的顶点分隔集这一概念。

备注 2.4.7. 正如我们将看到的，在分析这些确定性矩阵的 *Frobenius* 范数时，最小顶点分隔集的概念会自然出现。在以往的迹方法计算中（例如 *[163], [1]*），这需要巧妙的组合观察。

定义 2.4.8 (顶点分隔集**)**。 对于形状 τ，定义顶点分隔集为顶点 $S \subseteq V$ 的一个子集(τ)，使得在 $\tau \setminus S$ 中不存在从 U_τ 到 V_τ 的路径； $\tau \setminus S$ 是删除 S 中所有顶点（包括与它们相连的所有边）后得到的形状。

对于形状 τ，用 S_τ 表示大小最小的顶点分隔集。此外，令 I_τ 表示 V（τ）$\setminus U_\tau \setminus V_\tau$ 中的孤立顶点集（度为 0 的顶点）；这些顶点的存在实质上会使矩阵缩放一个标量因子。

定理 2.4.9. 对于形状 τ 和任意整数 $t \geq 1$,

$$\mathbb{E}\,\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\|_{2t}^{2t} \leq \left(C^{t|E(\tau)|} n^{|V(\tau)|} t^{t|E(\tau)|} |E(\tau)|^{2t|E(\tau)|}\right) n^{t(|V(\tau)|-|S_\tau|+|I_\tau|)}$$

其中 $C > 0$ 为绝对常数。

忽略低阶项后，同样的结果此前已在 [163, 1]中证明。为理解这一界，假设 τ 的顶点数为常数。令 $t \approx$polylog(n)，则可得

$$\|\mathbf{M}_\tau\| = \widetilde{\mathrm{O}}\left(\sqrt{n}^{|V(\tau)|-|S_\tau|+|I_\tau|}\right)$$

~以高概率成立，其中 O 隐去了对数因子。这是通过对 $\mathbb{E}\,\|\mathbf{M}_\tau\|_{2t}^{2t}$ 的界应用马尔可夫不等式得到的。如果 τ至少有一条边，那么 $\mathbb{E}\,\mathbf{M}_\tau = 0$，定理 2.4.9 给出了这样的界。如果 τ没有边，那么很容易得到这样的界；为完整起见，我们将其纳入引理 2.4.10。推论 2.4.11

精确表述了上述高概率界。因此， n 的这个幂基本上控制着范数界，并且在应用中被大量使用（例如 [20, 86, 191]）。**定理 2.4.9 的证明。** 我们首先论证可以假设 $I_\tau = \emptyset$。理由如下。τ 中每个度为 0 的不同顶点，本质上至多将矩阵缩放一个 n 的因子。而在不等式的右侧， I_τ 中的每个顶点相应地贡献一个 n^{2t} 的因子，分别来自 $n^{t|V_{(\tau)}|}$ 和 $n^{t|I_\tau|}$；其他变化只会使不等式变弱。

现在，固定 $a, b \geq 0$，使得 $a + b = |E(\tau)|$，并考虑 $\mathbf{M}_{\tau,a,b}$。对于 $I, J \in \mathcal{I}, \alpha, \beta \in \{0,1\}^{\binom{n}{2}}$，使得 $|\alpha|_0 = a, |\beta|_0 = b, \alpha \cdot \beta = 0$，根据定义，

$$\begin{aligned}\mathbf{M}_{\tau,a,b}[(I,\alpha),(J,\beta)] &= \nabla_{\alpha+\beta}\left(\sum_{\varphi:\varphi(U_\tau)=I,\varphi(V_\tau)=J}\prod_{u,v\in E(\tau)} G_{\varphi(u),\varphi(v)}\right)\\ &= |\{\varphi \mid \varphi(U_\tau) = I, \varphi(V_\tau) = J, \varphi(E(\tau)) = \mathrm{Supp}(\alpha+\beta)\}|\end{aligned}$$

其中 Supp(.) 表示支撑集。现在，我们将通过把这些确定性矩阵重新解释为对应于不同形状的图矩阵，来得到它们的范数界。

令 $P = (E_1, E_2)$ 表示 $E(\tau) = E_1 \sqcup E_2$ 的分拆，它分为两个有序集合 E_1, E_2，其中 $\sqcup$ 表示不交并。于是，我们可以写出 $\mathbf{M}_{\tau,a,b} = \sum_{P\in\mathcal{P}} \mathbf{M}_{\tau,a,b,P}$，其中

$$\mathbf{M}_{\tau,a,b,P}[(I,\alpha),(J,\beta)] = |\{\varphi \mid \varphi(U_\tau) = I, \varphi(V_\tau) = J, \varphi(E_1) = \mathrm{Supp}(\alpha), \varphi(E_2) = \mathrm{Supp}(\beta)\}|$$

设有序分拆的集合 P 为 $\mathcal{P}$。于是， $|\mathcal{P}| \leq (4|E(\tau)|)^{|E_{(\tau)}|}$，因此，根据事实 2.2.3，

$$\left\|\mathbf{M}_{\tau,a,b}\right\|_{2t}^{2t} \leq (4|E(\tau)|)^{t|E(\tau)|}\sum_{P\in\mathcal{P}}\left\|\mathbf{M}_{\tau,a,b,P}\right\|_{2t}^{2t}$$

每个$\mathbf{M}_{\tau,\ a,\ b,\ P}$都可以解释为一种不同形状 τ_P 的图矩阵，其顶点集相同且没有边。令 $V(\tau_P) = V(\tau)$ $E(\tau_P) = \emptyset$，并设 $U_{\tau_P} = U_\tau \cup V(E_1)$ $V(\tau_P) =$

$V_\tau \cup V(E_2)$，采用规范排序。于是，$\mathbf{M}_{\tau,\ a,\ b}$与$\mathbf{M}_{\tau_P}$仅相差行和列的重命名。图示参见图 2.4。

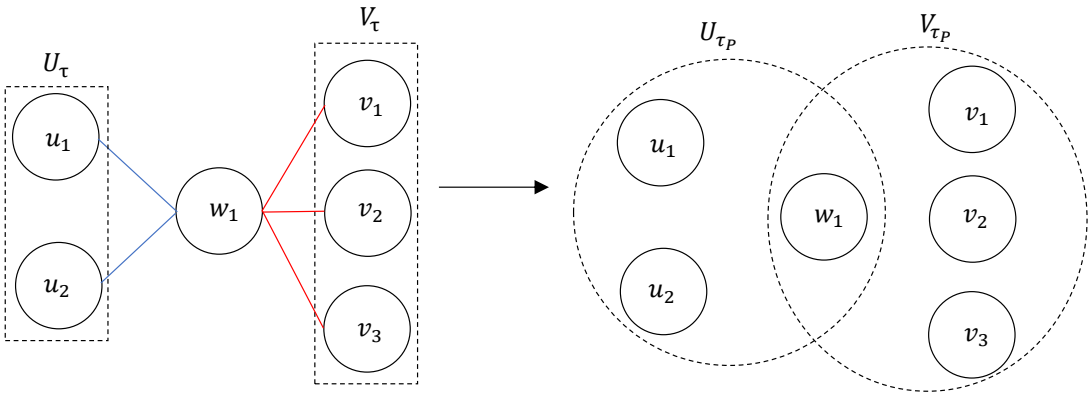


图 2.4：示例说明了如何定义 τ_P。在此例中，P约束蓝色边和红色边分别连接到 α和 β。U_{τ_P}，V_{τ_P}对顶点进行了排序（此处未显示）。

该图矩阵具有分块对角结构，其索引由公共顶点集合 $S = U_{\tau_P} \cap V_{\tau_P}$ 的实现给出。事实上，对于 $K \in [n]^S$，令 $\mathbf{M}_{\tau_P,\ K}$为$\mathbf{M}_{\tau_P}$中满足 $\varphi(S) = K$的块。那么，对于 $K \neq K'$，有$\mathbf{M}_{\tau_P,\ K}\mathbf{M}_{\tau_P,\ K'} = \mathbf{M}_{\tau_P,\ K}\mathbf{M}_{\tau_P,\ K'}^\dagger = 0$，因此，

$$\begin{aligned}
\mathbb{E}\left\|\mathbf{M}_{\tau,a,b}\right\|_{2t}^{2t} &\leq (4|E(\tau)|)^{t|E(\tau)|} \sum_{P\in\mathcal{P}} \left\|\mathbf{M}_{\tau_P}\right\|_{2t}^{2t} \\
&= (4|E(\tau)|)^{t|E(\tau)|} \sum_{P\in\mathcal{P}} \sum_{T\in[n]^S} \left\|\mathbf{M}_{\tau_P,T}\right\|_{2t}^{2t} \\
&\leq (4|E(\tau)|)^{t|E(\tau)|} \sum_{P\in\mathcal{P}} \sum_{T\in[n]^S} \left(\left\|\mathbf{M}_{\tau_P,T}\right\|_2^2\right)^t
\end{aligned}$$

其中，我们用适当的 Frobenius 范数幂对 Schatten 范数进行了界定。

对于任意固定的 $K \in [n]^S$，$\mathbf{M}_{\tau_P,\ K}$的各项取值属于 $\{0,1\}$，且非零项的数量至多为 $n^{|V(\tau)|-|S|}$，因为 S中顶点的实现已固定，而其他顶点各自至多有 n种选择。因此，$\left\|\mathbf{M}_{\tau_P,\ K}\right\|_2^2 \leq n^{|V(\tau)|-|S|}$。

最后，我们对 $|S|$ 进行界定，以估计在 P 的所有可能取值中，这一项最大能有多大。我们论证 S 阻断了从 U_τ 到 V_τ 的所有路径。为此，考虑一条从 U_τ 到

V_τ 的任意路径，它必定包含一条边 $(u,\ v) \in E(\tau)$，使得 $u \in U_{\tau_P},\ v \in V_{\tau_P}$。我们必有 $(u,\ v) \in E_1$，此时 $u,\ v \in U_{\tau_P}$ 和 $v \in S$；或者 $(u,\ v) \in E_2$，此时 $u,\ v \in V_{\tau_P}$ 和 $u \in S$。无论哪种情况，S 都必须包含 u 或 v。由此可知，S 必须是 τ 的顶点分隔集，从而得到 $|S| \geq |S_\tau|$。图示证明见图 2.5。

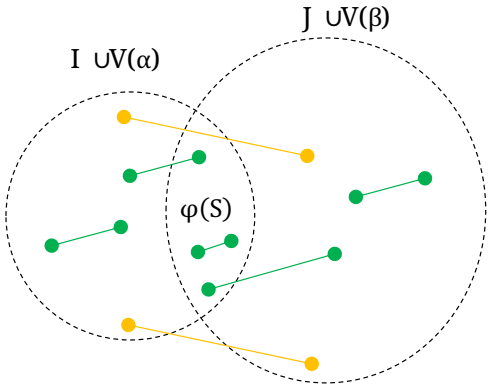


图 2.5：$|S| \geq |S_\tau|$ 的图示证明。绿色边可以出现在 τ 中，橙色边则不能，因此 S 阻断了从 U_τ 到 V_τ 的所有路径。

我们还有平凡的上界 $|S| \leq |V(\tau)|$。最终可得

$$\begin{aligned}\left\|\mathbf{M}_{\tau,a,b}\right\|_{2t}^{2t} &\leq (4|E(\tau)|)^{t|E(\tau)|} \sum_{P\in\mathcal{P}} \sum_{T\in[n]^S} n^{t(|V(\tau)|-|S_\tau|)} \\ &\leq (4|E(\tau)|)^{t|E(\tau)|}(4|E(\tau)|)^{|E(\tau)|}n^{|V(\tau)|}n^{t(|V(\tau)|-|S_\tau|)}\end{aligned}$$

结合之前的讨论，我们得到

$$\begin{aligned}\mathbb{E}\left\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\right\|_{2t}^{2t} &\leq \sum_{a+b=|E(\tau)|} (16t|E(\tau)|)^{|E(\tau)|t} \left\|\mathbf{M}_{\tau,a,b}\right\|_{2t}^{2t} \\ &\leq \sum_{a+b=|E(\tau)|} (16t|E(\tau)|)^{|E(\tau)|t}(4|E(\tau)|)^{t|E(\tau)|}(4|E(\tau)|)^{|E(\tau)|}n^{|V(\tau)|}n^{t(|V(\tau)|-|S_\tau|)} \\ &\leq \left(C^{t|E(\tau)|}n^{|V(\tau)|}t^{t|E(\tau)|}|E(\tau)|^{2t|E(\tau)|}\right)n^{t(|V(\tau)|-|S_\tau|)}\end{aligned}$$

其中 $C > 0$ 是一个绝对常数。 ■

在证明中，a 如上所述，我们对没有边的形状 τ_P 的分析普遍适用 l

于任何没有边的形状 τ 。为完整起见，我们在以下引理中明确陈述这一点。

引理 2.4.10. 对于不含边的形状 τ 以及任意整数 $t \geq 1$,

$$\mathbb{E}\,\|\mathbf{M}_\tau\|_{2t}^{2t} \leq n^{|U_\tau \cap V_\tau|} n^{t(V(\tau)-|U_\tau \cap V_\tau|+|I_\tau|)}$$

注意，这与定理 2.4.9 的形式相同，因为对于不含边的形状 τ，最小顶点分隔集 S_τ 恰好就是 $U_\tau \cap V_\tau$。

下面的推论利用马尔可夫不等式，得到图矩阵范数的高概率范数界。

推论 2.4.11. 对于形状 τ，对任意常数 $\varepsilon > 0$，以概率 $1-\varepsilon$,

$$\|\mathbf{M}_\tau\| \leq (C|E(\tau)|\log(n^{|V(\tau)|}/\varepsilon))^{|E(\tau)|} \cdot \sqrt{n}^{|V(\tau)|-|S_\tau|+|I_\tau|}$$

其中 $C > 0$ 是一个绝对常数。

证明。 如果 $E(\tau) = \emptyset$，我们使用引理 2.4.10。否则， $\mathbb{E}\,\mathbf{M}_\tau = 0$，并使用定理 2.4.9。由马尔可夫不等式可得,

$$\begin{aligned} Pr[\|\mathbf{M}_\tau\| \geq \theta] &\leq Pr[\|\mathbf{M}_\tau\|_{2t}^{2t} \geq \theta^{2t}] \\ &\leq \theta^{-2t}\,\mathbb{E}\,\|\mathbf{M}_\tau\|_{2t}^{2t} \\ &\leq \theta^{-2t}\left((C')^{t|E(\tau)|} n^{|V(\tau)|} t^{t|E(\tau)|} |E(\tau)|^{2t|E(\tau)|}\right) n^{t(|V(\tau)|-|S_\tau|+|I_\tau|)} \end{aligned}$$

其中 $C' > 0$ 为绝对常数。现在我们设定

$$\theta = \left(\varepsilon^{-1/(2t)} (C'')^{|E(\tau)|} n^{|V(\tau)|/(2t)} t^{|E(\tau)|/2} |E(\tau)|^{|E(\tau)|}\right) \sqrt{n}^{|V(\tau)|-|S_\tau|+|I_\tau|}$$

对于绝对常数 $C'' > 0$，要使该表达式至多为 ε。令 $t = \frac{1}{2}\log(n^{|V(\tau)|}/\varepsilon)$ –

以完成证明。 ■

2.5 为什么对 [188] 的朴素应用可能无法适用于一般乘积分布

在本节中，我们详细阐述处理不一定是拉ademacher 的随机变量时所产生的困难。在这种情形下，请注意，我们也不能假设多项式项是多线性的。

回顾一下问题设定：给定一个随机矩阵 $\mathbf{F}$ ，其元素是随机变量 $Z_1, \ldots, Z_n$ 的低次多项式，这些变量是从任意分布中独立抽样得到的。我们希望通过控制 $\mathbb{E}\,\|\mathbf{F} - \mathbb{E}\,\mathbf{F}\|_{2t}^{2t}$，得到关于$\mathbf{F}$ 偏离其均值程度的集中界。

基于第2.3节中的思路，我们可以尝试使用矩阵埃夫龙-斯坦不等式（定理 2.1.1），并希望得到类似的递归框架。现在我们来讨论这样做会发生什么。假设 $\mathbb{E}[Z_i] = 0$, $\mathbb{E}[Z_i^2] = 1$。我们可以采用与定理 2.1.2 的证明类似的方法。因此，我们将$\mathbf{X}$ 视为$\mathbf{X}_{0,0}$ 的一个主子矩阵，并按照引理 2.3.3 的步骤进行。主要变化会出现在命题 2.3.5 中。具体来说，等式$\mathbb{E}[(Z_i - \widetilde{Z}_i)^2|Z] = 2$ 不再成立。取而代之的是，我们将得到 $\mathbb{E}[(Z_i - \widetilde{Z}_i)^2|Z] = 1 + Z_i^2$。因此，我们得到表达式

$$\sum_{i=1}^{n}(1+Z_i^2)\mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^{\intercal} = \sum_{i=1}^{n}\mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^{\intercal} + \sum_{i=1}^{n}Z_i^2\mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^{\intercal}$$

第一项可以像在基本框架中一样处理。遗憾的是，第二项将成为一个困难的来源。为了解决这一困难，我们可以尝试再次对一个适当构造的矩阵应用矩阵 Efron-Stein 不等式。为此，我们可以将第二项理解为先对变量 Z_i 求导，然后放回该变量后得到的。相比之下，在处理 Rademacher 随机变量时，我们并不需要将其放回。但在我们这样做之后，当对这些额外的矩阵进行递归时，新的第二项将包含左侧作为一个子项，从而得到一个平凡的不等式，并使递归停滞。

在处理 Rademacher 随机变量时，我们并不需要将其放回。但在我们这样做之后，当对这些额外的矩阵进行递归时，新的第二项将包含左侧作为一个子项，从而得到一个平凡的不等式，并使递归停滞。

为了更清楚地说明这一点，考虑最简单的情形 $a = b = 0$。那么，正如我们之前所看到的，第一项 $\sum_{i=1}^{n} \mathrm{F}_{a,b,i}\mathrm{F}_{a,b,i}$ 将等于 $\mathrm{F}_{0,1}\mathrm{F}_{0,1}$。类似地，为了计算第二项 $\sum_{i=1}^{n} Z_i^2 \mathrm{F}_{a,b,i}\mathrm{F}_{a,b,i}$，我们定义矩阵 H，使其与 $\mathrm{F}_{0,1}$ 相同，只是现在每个矩阵项都乘上了 Z_i，其中 i 是该列中求导所针对的变量。也就是说，$\mathrm{H}[I,(J,\mathrm{e}_i)] = Z_i\mathrm{F}_{0,1}[I,(J,\mathrm{e}_i)]$。注意，在 H 的定义中，对 Z_i 求导后又将它放了回来。于是，第二项将是 HH，我们可以希望再次递归地对矩阵 H 使用 Efron-Stein。

我们可以这样做，并在作出上述适当修改后，仿照引理 2.3.3 的证明继续推导。但由于 $\beta_i = 1$ 已经如此，对 Z_i 求导并将它放回后，会得到同一个矩阵 $\mathbf{H}$！因此，我们最终得到如下形式的不等式

$$\mathbb{E}\,\|\mathbf{H}\|_{2t}^{2t} \le \mathrm{O}(t)^t(\mathbb{E}\,\|\mathbf{H}\|_{2t}^{2t} + \text{ other nonnegative terms})$$

事实上，这是一个同义反复，对我们并无帮助。

粗略估计一下，假设对于我们的分布，存在一个参数 L 使得 $1 + Z_i^2 \le L$，那么我们就能得到一个类似的框架，但在递归的每一步都会产生 $\sqrt{L}$ 的损失。遗憾的是，这个界并不紧。例如，如果我们对 $G \sim \mathcal{G}_{n,p}$ 的中心化归一化邻接矩阵进行这项计算，就会得到一个范数界 $\widetilde{\mathrm{O}}(\sqrt{n(1-p)}\sqrt{p})$，其中 $\widetilde{\mathrm{O}}$ 隐藏了对数因子。。当 p 为常数，甚至为逆多对数级时，这个界是紧的。但对于某个常数 $0 < \theta < 1$，当 $p = n^{-\theta}$ 时，这个界并不紧，因为在这种情形下，真实的范数界已知为 $\mathrm{O}(\sqrt{n})$（参见 [82, 229] 的早期研究；更紧的界请参见 [27] 及其中的参考文献）。

如果我们深入探究事情的细节，这个例子说明，对于某些类型的矩阵，矩阵 Efron-Stein 不等式定理 2.1.1 会变成同义反复，

给出 $\mathbf{V} = O(1)\mathbf{X}\mathbf{X}^\intercal +$ 其他半正定矩阵。

但在我们的一般框架中，上述坏矩阵会在我们对已经求过导的变量再次求导时出现。换言之，当前的方差代理量$\mathbf{V}$ 定义并未考虑我们是否已经对某个变量 Z_i求过导。因此，对于一般递归，我们深入研究 [188] 的证明，并利用在我们的框架中得到的中间矩阵的结构性质对其进行修改。

2.6 一般递归框架

现在假设 $Z_1, \ldots, Z_n$ 是从具有有限阶矩的分布 Ω 中抽样得到的独立同分布随机变量。为简便起见，我们假设它们同分布，但只要它们~相互独立，我们的方法也很容易推广到不同分布的情形。对于每个 $i \le n$，定义 Z_i 为 Z_i 的独立副本，并定义向量 $Z^{(i)} := (Z_1, \ldots, Z_{i-1}, \widetilde{Z}_i, Z_{i+1}, \ldots, Z_n)$。定义 Z' 为如下随机向量：从 $[n]$ 中均匀随机抽取 i，然后令 $Z' = Z^{(i)}$。

设$\mathbf{F} \in \mathbb{R}[Z]^{\mathcal{I}\times\mathcal{J}}$ 是一个矩阵，其行和列分别由任意集合 $\mathcal{I}, \mathcal{J}$索引，使得对于所有 $I \in \mathcal{I}, J \in \mathcal{J}$，$\mathbf{F}[I, J]$ 都是关于 $Z_1, \ldots, Z_n$的多项式。令 d_p为$\mathbf{F}[I, J]$ 所有元素中的最高次数 I, J并令 d为 Z_i在$\mathbf{F}[I, J]$ 所有元素中的最高次数，以及 $i \le n$。

与拉德马赫情形类似，令$\mathbf{X} := \mathbf{F} - \mathbb{E}\mathbf{F}$。当输入为 Z 时，我们将矩阵记作$\mathbf{F}$,$\mathbf{X}$等；当输入为 $Z^{(i)}$ 时，将相应的矩阵记作$\mathbf{F}^{(i)}$,$\mathbf{X}^{(i)}$等。在本节中，我们将给出一个一般框架，借助它可以得到任意整数 $t \ge 1$ 下 $\mathbb{E}\,\|\mathbf{F} - \mathbb{E}\mathbf{F}\|_{2t}^{2t}$ 的界。

我们先介绍一些预备知识，以便陈述主定理。

定义 2.6.1 (空间 $\mathcal{S}$)。令 $\mathcal{S}$为 $Z_1, \ldots, Z_n$ 中度不超过 d_p 的零均值多项式空间。

对于 $\alpha \neq 0$，我们还定义中心化单项式

$$\chi_\alpha(Z) = \prod_{\alpha_i>0} (Z_i^{\alpha_i} - \mathbb{E}[Z_i^{\alpha_i}])$$

根据定义，对所有 $\alpha \neq 0$，$|\alpha|_1 \leq d_p$，$\chi_\alpha \in \mathcal{S}$。以下命题显而易见。

命题 2.6.2。集合 $\{\chi_\alpha(Z)|1 \leq |\alpha|_1 \leq d_p\}$ 构成 $\mathcal{S}$ 的一组基。

对于一般框架，我们在这个基下进行处理，因为正如我们将在第 2.7 节看到的，“内核矩阵”在这个基下便于表述。∇ 算子也能很好地作用于我们的多项式 χ_β。事实上，注意到若 $0 \overset{\alpha \lhd \beta}{\neq} \alpha$，则 $\nabla_\alpha(\chi_\beta) = \begin{cases} \chi_{\beta-\alpha} \\ \end{cases}$；否则为 0。

对于 $\mathcal{S}$ 中的多项式 $f(Z)$，记 $\widehat{f}(\alpha)$ 为 $\chi_\alpha(Z)$ 在f的展开中的系数，也就是说，

$$f(Z) = \sum_{0\neq\alpha\in\mathbb{N}^n} \widehat{f}(\alpha)\chi_\alpha(Z)$$

我们可以自然地将这个记号推广到均值为零的矩阵。因此，我们可以写作 $\widehat{\mathbf{X}} = \sum_{\alpha\neq 0}\mathbf{X}(\alpha)\chi_\alpha(Z)$，其中 $\mathbf{X}(\alpha)$ 是确定性矩阵。为了应用我们的递归框架，我们根据 $|\alpha|_0$ 将这个和分组成若干项。对于 $k \geq 1$，定义 $\widehat{\mathbf{X}}_k = \sum_{|\alpha|_0=k}\mathbf{X}(\alpha)\chi_\alpha(Z)$。于是，

$$\mathbf{X} = \sum_{k\geq 1} \mathbf{X}_k$$

注意，当 $k > d_p$ 时，$\mathbf{X}_k = 0$。

定义 2.6.3（索引集 $\mathcal{K}$）。我们将 $\mathcal{K} \subseteq \mathbb{N}^n \times \{0,1\}^n$ 定义为满足以下条件的有序对集合：（α、γ），其中 $|\alpha|_1 \leq d_p$、$\alpha \in \mathbb{N}^n$，且 $\gamma \leq \alpha$、$\gamma \in \{0,1\}^n$。

定义非零项为$\mathbf{D}_1 \in \mathbb{R}[Z]^{\mathcal{I}\times\mathcal{K}} \times \mathbb{R}[Z]^{\mathcal{I}\times\mathcal{K}}$ 的对角矩阵

$$\mathbf{D}_1[(I,\alpha,\gamma),(I,\alpha,\gamma)] = \sqrt{\mathbb{E}[Z^{2\alpha\cdot(1-\gamma)}]}Z^{\alpha\cdot\gamma}$$

同理，定义非零项为$\mathbf{D}_2 \in \mathbb{R}[Z]^{\mathcal{J}\times\mathcal{K}} \times \mathbb{R}[Z]^{\mathcal{J}\times\mathcal{K}}$的对角矩阵

$$\mathbf{D}_2[(J,\alpha,\gamma),(J,\alpha,\gamma)] = \sqrt{\mathbb{E}[Z^{2\alpha\cdot(1-\gamma)}]}Z^{\alpha\cdot\gamma}$$

定义 *2.6.4*（矩阵 $G_{k,a,b}$、$F_{k,a,b}$）。对于满足 $k \geq 1, a, b \geq 0$ 的整数 k, a, b，定义矩阵 $G_{k,a,b}$，其行和列分别由 $\mathcal{I}\times\mathcal{K}$ 和 $\mathcal{J}\times\mathcal{K}$ 索引，并满足对所有 $(I, \alpha_1, \gamma_1) \in \mathcal{I}\times\mathcal{K}$、$(J, \alpha_2, \gamma_2) \in \mathcal{J}\times\mathcal{K}$，

$$\mathbf{G}_{k,a,b}[(I,\alpha_1,\gamma_1),(J,\alpha_2,\gamma_2)] = \begin{cases} \nabla_{\alpha_1+\alpha_2}\mathbf{X}_k[I,J] & if\ |\alpha_1|_0 = a, |\alpha_2|_0 = b, \alpha_1\cdot\alpha_2 = 0 \\ 0 & o.w. \end{cases}$$

此外，定义 $\mathbf{F}_{k,a,b} := \mathbf{D}_1\mathbf{G}_{k,a,b}\mathbf{D}_2$。

注意，当 $k > d_p$ 时，$\mathbf{F}_{k,a,b} = 0$。

命题 2.6.5。对于整数 k, a, b，满足 $k \geq 1, a, b \geq 0$，设 $a + b < k$。则矩阵 $G\ f$ 的每个非零项$\mathbf{G}_{k,a,b}$都具有以下性质：$f(\alpha)$ 仅当 $|\alpha|_0 = k-a-b$证明。$\mathbf{X}_k$ 的非零项仅包含恰好 k 个变量，而 $\nabla_{\alpha_1+\alpha_2}$ 要么使该项归零，要么恰好截去 $|\alpha_1+\alpha_2|_0 = |\alpha_1|_0+|\alpha_2|_0 = a+b$ 个变量。

■

这也立即意味着，当 $a + b < k$ 时，$\mathbb{E}[\mathbf{G}_{k,\ a,\ b}] = 0$。最后，当$k = a + b$时，我们有$\mathbf{G}_{k,\ a,\ b}$是一个独立于 Z_i的确定性矩阵。这些矩阵构成了主定理中出现的$\mathbf{F}_{a+b,\ a,\ b}$。

现在我们可以陈述主定理了。

定理 2.6.6 (一般递归)。设随机变量元组 Z 和函数$\mathbf{F}$ 如上所述。那么，对于所有整数 $t \geq 1$,

$$\mathbb{E}\left\|\mathbf{F} - \mathbb{E}\,\mathbf{F}\right\|_{2t}^{2t} \leq \sum_{a,b\geq 0, a+b\geq 1} (Ct^2 dd_p^4)^{(a+b)t}\,\mathbb{E}\left\|\mathbf{F}_{a+b,a,b}\right\|_{2t}^{2t}$$

其中 $C > 0$为绝对常数。

注意，$\mathbf{F}_{a+b,a,b} = \mathbf{D}_1\mathbf{G}_{a+b,a,b}\mathbf{D}_2$ ，其中$\mathbf{D}_1$,$\mathbf{D}_2$ 是对角矩阵，$\mathbf{G}_{a+b,a,b}$是一个与 Z 相互独立的确定性矩阵。要分析此类矩阵的 Schatten 范数的期望，我们可以采用简单得多的技巧。例如，我们可以利用 Frobenius 范数的适当幂得到一个简单的界，并应用标准的标量集中工具。我们将在第 2.8 节看到一个例子。

备注 2.6.7. 我们没有尝试优化定理 *2.6.6* 中期望前面的因子，我们推测这些因子还可以改进。

我们通过反复应用以下技术引理来证明主定理，其证明推迟到下一节。

Lemma 2.6.8. 对于所有整数 $t \geq 1$、整数 $k \geq 1$、 a、 $b \geq 0$，使得 $a + b < k$ ，

$$\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} \leq (Ct^2 dd_p^2)^t (\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})$$

其中 $C > 0$ 是一个绝对常数。

利用这个引理，我们可以完成主定理的证明。

定理 *2.6.6* 的证明。 利用事实 2.2.3，我们有 $\mathbb{E}\|\mathbf{X}\|_{2t}^{2t} \leq d_p^{2t}\sum_{k=1}^{d_p}\mathbb{E}\|\mathbf{X}_k\|_{2t}^{2t}$。注意，对于任意 $k \geq 1$，矩阵 $\mathbf{X}k$ 是$\mathbf{F}_{k,0,0}$ 的一个主子矩阵，其余所有元素均为 0，因此 $\mathbb{E}\|\mathbf{X}_k\|_{2t}^{2t} = \mathbb{E}\left\|\mathbf{F}_{k,0,0}\right\|_{2t}^{2t} = \frac{1}{2}\mathbb{E}\left\|\mathbf{F}_{k,0,0}\right\|_{2t}^{2t}$。因此，

$$\mathbb{E}\|\mathbf{X}\|_{2t}^{2t} \leq \frac{1}{2}d_p^{2t}\sum_{k=1}^{d_p}\mathbb{E}\left\|\overline{\mathbf{F}}_{k,0,0}\right\|_{2t}^{2t}$$

我们现在应用 对所有项反复应用引理 2.6.8，直到 $k = a + b$，最终得到 g

$$\mathbb{E}\|\mathbf{X}\|_{2t}^{2t} \leq \frac{1}{2}d_p^{2t}(Ct^2 dd_p^2)^{(a+b)t}\sum_{a,b\geq 0, a+b\geq 1}\mathbb{E}\left\|\overline{\mathbf{F}}_{a+b,a,b}\right\|_{2t}^{2t}$$

观察到 $\mathbb{E}\left\|\mathbf{F}_{a+b,\ a,\ -b}\right\|_{2t}^{2t} = 2\,\mathbb{E}\left\|\mathbf{F}_{a+b,\ a,\ b}\right\|_{2t}^{2t}$，证明得证。 ■

2.7 对 [188] 的推广及引理 2.6.8 的证明

本节将采用第 2.1 节中介绍的高层次策略来证明引理 2.6.8。这需要推广 [188], 中的结果，而其中的证明技巧本身也可能具有独立的研究价值。

2.7.1 通过显式内核推广 *[188]*

在我们的设定中，注意到 $(Z,\ Z')$ 与 $(Z',\ Z)$ 具有相同的分布。这就是所谓的变量可交换对，对我们的分析极为有用。特别地，$Z,\ Z'$ 具有相同的分布，并且对于每个可积函数 f，都有 $\mathbb{E}\, f(Z,\ Z') = \mathbb{E}\, f(Z',\ Z)$。

定义 2.7.1 (拉普拉斯算子 $\mathcal{L}$)。 定义空间 $\mathcal{S}$ 上的算子 $\mathcal{L}$ 为

$$\mathcal{L}(f)(Z) = \mathbb{E}[f(Z) - f(Z')|Z]$$

对所有多项式 $f \in \mathcal{S}$。

注意，该算子定义良好，因为对于任意 $f \in \mathcal{S}$, $\mathbb{E}[L(f)] = \mathbb{E}[\mathbb{E}[f(Z) - f(Z')|Z]] = \mathbb{E}[f(Z) - f(Z')] = 0$，因此，$L(f) \in \mathcal{S}$。

引理 2.7.2。 对于所有 $\alpha \in \mathbb{N}^n$, χ_α 都是 $\mathcal{L}$ 的特征向量，其特征值为 $\frac{|\alpha|_0}{n}$。

证明。 回想一下，Z' 是通过均匀随机选择 $i \in [n]$，然后设置$Z' = Z^{(i)}$得到的。因此，

$$\begin{aligned}\mathcal{L}(\chi_\alpha)(Z) &= \mathbb{E}[\chi_\alpha(Z) - \chi_\alpha(Z')|Z] \\ &= \frac{1}{n}\sum_{i\le n}\mathbb{E}[\chi_\alpha(Z) - \chi_\alpha(Z^{(i)})|Z]\end{aligned}$$

当 $\alpha_i = 0$时， $\chi_\alpha(Z) - \chi_\alpha(Z_{(i)}) = 0$。否则， $\mathbb{E}[\chi_\alpha(Z) - \chi_\alpha(Z_{(i)})|Z] = \chi_\alpha(Z)$。因此，上述表达式可简化为 $\frac{|\alpha|_0}{n}\chi_\alpha(Z)$。 ■

定理 2.7.3 (显式核)。对于任意均值中心化多项式 $f \in \mathcal{S}$，存在一个关于 $2n$ 个变量 z_1、...、z_n、z'_1、...、z'_n 的多项式 K_f，统称为(z, z')，具有以下性质

1. $K_f(z', z) = -K_f(z, z')$
2. $\mathbb{E}[K_f(Z, Z')|Z] = f(Z)$ 其中(Z, Z') 是我们上文考虑的可交换对。

证明。利用命题 2.6.2 和引理 2.7.2，在多项式基 χ_α 下，算子 $\mathcal{L}$ 是一个对角矩阵，其对角元素非零，因此， $\mathcal{L}^{-1}$ 存在，且可显式表示为

$$\mathcal{L}^{-1}(f)(Z) = \sum_\alpha \frac{n}{|\alpha|_0}\widehat{f}(\alpha)\chi_\alpha(Z)$$

然后我们取 $K_f(z, z') = \mathcal{L}^{-1}(f)(z) - \mathcal{L}^{-1}(f)(z')$。第一个条件显然成立；对于第二个条件，我们有

$$\mathbb{E}[K_f(Z, Z')|Z] = \mathbb{E}[\mathcal{L}^{-1}(f)(Z) - \mathcal{L}^{-1}(f)(Z')|Z] = \mathcal{L}(\mathcal{L}^{-1}(f)) = f$$

■

如定理 2.7.3 的证明所示， $\mathcal{L}$ 有一个定义良好的逆 $\mathcal{L}^{-1}$。现在我们定义矩阵 $\mathbf{K}_{k,\ a,\ b}$ 并称之为内核。

定义 2.7.4 （内核矩阵 $\mathbf{K}_{k,\ a,\ b}$）。对于满足$a + b < k$的整数 $k \geq 1$、a、$b \geq 0$，定义矩阵 $\mathbf{K}_{k,\ a,\ b} \in \mathbb{R}[Z]^{\mathcal{I}\times\mathcal{K}} \times \mathbb{R}[Z]^{\mathcal{J}\times\mathcal{K}}$，将变量 $(z、z') = (z_1、\ldots、z_n、z'_1、\ldots、z'_n)$ 作为输入，定义如下：

$$\mathbf{K}_{k,a,b}(z, z') = \mathcal{L}^{-1}(\mathbf{G}_{k,a,b})(z) - \mathcal{L}^{-1}(\mathbf{G}_{k,a,b})(z')$$

在本节其余部分中，除非另有明确说明，固定满足 $a+b<k$ 的整数 $k\geq 1$、a、$b\geq 0$。那么，内核 $\mathbf{K}_{k,\ a,\ b}$ 是良定义的。

引理 2.7.5。$\mathbf{K}_{k,\ a,\ b}(Z,\ Z')\ =\ \frac{n}{k-a-b}(\mathbf{G}_{k,\ a,\ b}(Z)\ -\mathbf{G}_{k,\ a,\ b}(Z'))$

证明。

$$\begin{aligned}\mathbf{K}_{k,a,b}(Z,Z') &= \mathcal{L}^{-1}(\mathbf{G}_{k,a,b})(Z)-\mathcal{L}^{-1}(\mathbf{G}_{k,a,b})(Z')\\ &= \sum_{|\alpha|_0=k-a-b}\widehat{\mathbf{G}_{k,a,b}}(\alpha)(\mathcal{L}^{-1}(\chi_\alpha)(Z)-\mathcal{L}^{-1}(\chi_\alpha)(Z'))\\ &= \frac{n}{k-a-b}\sum_{|\alpha|_0=k-a-b}\widehat{\mathbf{G}_{k,a,b}}(\alpha)(\chi_\alpha(Z)-\chi_\alpha(Z'))\\ &= \frac{n}{k-a-b}(\mathbf{G}_{k,a,b}(Z)-\mathbf{G}_{k,a,b}(Z'))\end{aligned}$$

■

下面的引理阐述了内核的重要性质，包括它如何与$\mathbf{D}_1$ 和$\mathbf{D}_2$相互作用。

引理 2.7.6。$\mathbf{K}_{k,\ a,\ b}$ 满足以下性质

1. $\mathbf{K}_{k,a,b}(z',z)=-\mathbf{K}_{k,a,b}(z,z')$
2. $\mathbb{E}[\mathbf{K}_{k,a,b}(Z,Z')|Z]=\mathbf{G}_{k,a,b}(Z)$
3. $(\mathbf{D}_1(Z)-\mathbf{D}_1(Z'))\mathbf{K}_{k,a,b}(Z,Z')=\mathbf{K}_{k,a,b}(Z,Z')(\mathbf{D}_2(Z)-\mathbf{D}_2(Z'))=0.$

证明。 第一个等式根据定义显然成立。对于第二个等式，注意到$\mathbb{E}[\mathbf{G}_{k,a,b}]=0$ 和 $\mathbf{K}_{k,a,b}$ 的定义是将$\mathbf{G}_{k,a,b}$ 的每个矩阵项 f 替换为定理 2.7.3 中所示的核多项式 K_f。现在，我们证明第三个等式。

考虑矩阵 $(\mathbf{D}_1(Z) - \mathbf{D}_1(Z'))\mathbf{K}_{k,a,b}(Z, Z')$，其 $[(I, \alpha_1, \gamma_1), (J, \alpha_2, \gamma_2)]$ 矩阵项为

$$\frac{n}{k-a-b}\sqrt{\mathbb{E}[Z^{2\alpha_1\cdot(1-\gamma_1)}]}(Z^{\alpha_1\cdot\gamma_1} - (Z')^{\alpha_1\cdot\gamma_1})(\nabla_{\alpha_1+\alpha_2}\mathbf{X}_k[I,J](Z) - \nabla_{\alpha_1+\alpha_2}\mathbf{X}_k[I,J](Z'))$$

这里用到了引理 2.7.5。我们将证明这一项恒等于 0。必有 $Z' = Z(i)$，其中某个 $i \le n$。如果 $(\alpha_1 \cdot \gamma_1)_i = 0$，那么 $Z^{\alpha_1\cdot\gamma_1} = (Z')^{\alpha_1\cdot\gamma_1}$，上述项便为 0。否则，$(\alpha_1 + \alpha_2)_i \neq 0$，因此对于任意多项式 f，$\nabla_{\alpha_1+\alpha_2}$ 只含有与 Z_i 无关的项；此时 $\nabla_{\alpha_1+\alpha_2}\mathbf{X}_k[I,J](Z) = \nabla_{\alpha_1+\alpha_2}\mathbf{X}_k[I,J](Z')$。在这种情形下，上述项同样为 0。另一个等式的证明类似。

■

我们之所以称$\mathbf{K}_{k,a,b}$ 为内核，是因为如上所见，它在分解$\mathbf{F} = \mathbf{DGD}$中充当内层矩阵 $\mathbf{G}$ 的核。

由于我们需要使用厄米扩张，因此定义

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \end{bmatrix}$$

我们将在推导过程中广泛使用以下基本事实。

事实 2.7.7。对于任意矩阵 $\mathbf{A} \in \mathbb{R}[Z]^{\mathcal{I}\times\mathcal{K}} \times \mathbb{R}[Z]^{\mathcal{J}\times\mathcal{K}}$，$\mathbf{D}\overline{\mathbf{A}\mathbf{D}} = \overline{\mathbf{D}_1\mathbf{A}\mathbf{D}_2}$。

证明。我们有

$$\begin{aligned}\mathbf{D}\overline{\mathbf{A}}\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \end{bmatrix}\begin{bmatrix} 0 & \mathbf{A} \\ \mathbf{A}^\intercal & 0 \end{bmatrix}\begin{bmatrix} \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \end{bmatrix} &= \begin{bmatrix} 0 & \mathbf{D}_1\mathbf{A} \\ \mathbf{D}_2\mathbf{A}^\intercal & 0 \end{bmatrix}\begin{bmatrix} \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \end{bmatrix} \\ &= \begin{bmatrix} 0 & \mathbf{D}_1\mathbf{A}\mathbf{D}_2 \\ \mathbf{D}_2\mathbf{A}^\intercal\mathbf{D}_1 & 0 \end{bmatrix} \\ &= \overline{\mathbf{D}_1\mathbf{A}\mathbf{D}_2}\end{aligned}$$

■

我们从 [188]中的一个结果的推广版本开始。

引理 2.7.8。 设 $\mathbf{K} = \mathbf{K}_{k,\overline{a,b}}$. 对于定义在变量 Z 上的任意与$\mathbf{K}$维度相同的对称矩阵值函数 $\mathbf{R}$ ，若满足 $\mathbb{E}\left\|\mathbf{K}(Z, Z')\mathbf{R}(Z)\right\| < \infty$，则有

$$\mathbb{E}[\overline{\mathbf{F}}_{k,a,b}(Z)\mathbf{R}(Z)] = \frac{1}{2}\mathbb{E}[\mathbf{D}(Z)\mathbf{K}(Z, Z')\mathbf{D}(Z)(\mathbf{R}(Z) - \mathbf{R}(Z'))]$$

证明。 由引理 2.7.6 可得

$$\begin{aligned}\mathbb{E}[\overline{\mathbf{F}}_{k,a,b}(Z)\mathbf{R}(Z)] &= \mathbb{E}[\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z)\mathbf{R}(Z)] \\ &= \mathbb{E}[\mathbf{D}(Z)\,\mathbb{E}[\mathbf{K}(Z, Z')|Z]\mathbf{D}(Z)\mathbf{R}(Z)] \\ &= \mathbb{E}[\mathbf{D}(Z)\mathbf{K}(Z, Z')\mathbf{D}(Z)\mathbf{R}(Z)]\end{aligned}$$

其中，第一个等式由引理 2.7.6 的条件 2 得出，第二个等式则由期望的穿越性质得出。继续推导,

$$\begin{aligned}\mathbb{E}[\overline{\mathbf{F}}_{k,a,b}(Z)\mathbf{R}(Z)] &= \mathbb{E}[\mathbf{D}(Z)\mathbf{K}(Z, Z')\mathbf{D}(Z)\mathbf{R}(Z)] \\ &= \mathbb{E}[\mathbf{D}(Z')\mathbf{K}(Z', Z)\mathbf{D}(Z')\mathbf{R}(Z')] \\ &= -\mathbb{E}[\mathbf{D}(Z')\mathbf{K}(Z, Z')\mathbf{D}(Z')\mathbf{R}(Z')] \\ &= -\mathbb{E}[\mathbf{D}(Z)\mathbf{K}(Z, Z')\mathbf{D}(Z')\mathbf{R}(Z')] \\ &= -\mathbb{E}[\mathbf{D}(Z)\mathbf{K}(Z, Z')\mathbf{D}(Z)\mathbf{R}(Z')]\end{aligned}$$

这里，第二个等式由以下事实得出：$(Z,\ Z')$ 与 $(Z',\ Z)$ 具有相同的分布，因此我们可以交换它们。第三、第四和第五个等式分别由引理 2.7.6 的条件 1、3、3 得出。将这两个显示式相加，即得所需结果。

■

定义 2.7.9 （矩阵 $\mathbf{U}_{k,\ a,\ b}$，$\mathbf{V}_{k,\ a,\ b}$）。 我们定义以下矩阵

$$\mathbf{U}_{k,a,b} = \mathbb{E}[(\overline{\mathbf{F}}_{k,a,b}(Z) - \overline{\mathbf{F}}_{k,a,b}(Z'))^2|Z]$$

$$\mathbf{V}_{k,a,b} = \mathbb{E}[(\mathbf{D}(Z)\overline{\mathbf{K}}_{k,a,b}(Z,Z')\mathbf{D}(Z))^2|Z]$$

$\mathbf{U}_{k,a,b}$ 的定义与 [188], 中的定义基本相同，在那里它被称为条件方差。在我们的设定中，$\mathbf{V}_{k,a,b}$ 的定义略有不同。这使我们能够利用$\mathbf{F}_{k,a,b}$ 所呈现的特定乘积结构，以及引理 2.7.6 中内核的特殊性质。 —

现在我们将证明一个与 [188] 中所示引理类似的引理。

引理 2.7.10。 对于任意 $s > 0$ 以及任意整数 $t \geq 1$,

$$\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} \leq \left(\frac{2t-1}{4}\right)^t \mathbb{E}\left\|s\mathbf{U}_{k,a,b} + s^{-1}\mathbf{V}_{k,a,b}\right\|_t^t$$

为证明这一点，我们需要以下不等式。

引理 2.7.11 （多项式均值迹不等式， [188]）。 对于所有矩阵 $\mathbf{A},\mathbf{B},\mathbf{C} \in \mathbb{H}^d$、所有整数 $q \geq 1$ 以及所有 $s > 0$,

$$\operatorname{tr}[\mathbf{C}(\mathbf{A}^q - \mathbf{B}^q)]| \leq \frac{q}{4}\operatorname{tr}[(s(\mathbf{A}-\mathbf{B})^2 + s^{-1}\mathbf{C}^2)(\mathbf{A}^{q-1} + \mathbf{B}^{q-1})]$$

引理 *2.7.10* 的证明。 我们先令 $\mathbf{R}(Z) = \mathbf{F}_{k\ a\ b,}^{2t-1}(Z)$，然后应用引理 2.7.8。 —

$$\begin{aligned}\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} &= \mathbb{E}\operatorname{tr}[\overline{\mathbf{F}}_{k,a,b}\cdot\overline{\mathbf{F}}_{k,a,b}^{2t-1}] \\ &= \frac{1}{2}\mathbb{E}[\mathbf{D}(Z)\overline{\mathbf{K}}_{k,a,b}(Z,Z')\mathbf{D}(Z)(\overline{\mathbf{F}}_{k,a,b}^{2t-1}(Z) - \overline{\mathbf{F}}_{k,a,b}^{2t-1}(Z'))]\end{aligned}$$

应用引理 2.7.11,

$$
\begin{aligned}
&\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} \\
&\leq (\frac{2t-1}{8})\,\mathbb{E}\,\mathrm{tr}[(s(\overline{\mathbf{F}}_{k,a,b}(Z)-\overline{\mathbf{F}}_{k,a,b}(Z'))^2+s^{-1}(\mathbf{D}(Z)\overline{\mathbf{K}}_{k,a,b}(Z,Z')\mathbf{D}(Z))^2)(\overline{\mathbf{F}}_{k,a,b}^{2t-2}(Z)+\overline{\mathbf{F}}_{k,a,b}^{2t-2}(Z'))] \\
&=(\frac{2t-1}{4})\,\mathbb{E}\,\mathrm{tr}[(s(\overline{\mathbf{F}}_{k,a,b}(Z)-\overline{\mathbf{F}}_{k,a,b}(Z'))^2+s^{-1}(\mathbf{D}(Z)\overline{\mathbf{K}}_{k,a,b}(Z,Z')\mathbf{D}(Z))^2)\overline{\mathbf{F}}_{k,a,b}^{2t-2}(Z)]
\end{aligned}
$$

其中，最后一行用到了这样一个事实：(Z,Z') 与 (Z',Z) 具有相同的分布，并应用了引理 2.7.6 的条件 3。根据 $\mathbf{U}_{k,a,b}$ 和 $\mathbf{V}_{k,a,b}$ 的定义，我们得到

$$
\begin{aligned}
\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} &\leq \frac{2t-1}{4}\,\mathbb{E}\,\mathrm{tr}[(s\mathbf{U}_{k,a,b}+s^{-1}\mathbf{V}_{k,a,b})\overline{\mathbf{F}}_{k,a,b}^{2t-2}] \\
&\leq \frac{2t-1}{4}\left(\mathbb{E}\left\|s\mathbf{U}_{k,a,b}+s^{-1}\mathbf{V}_{k,a,b}\right\|_t^t\right)^{1/t}(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t})^{(t-1)/t}
\end{aligned}
$$

其中，迹部分使用了 H¨older 不等式，期望部分使用了 H¨older 不等式。整理后即可得到所需结果。 ■

2.7.2 引理 *2.6.8* 的证明

引理 2.7.10 表明，要给 $\mathbb{E}\left\|\mathbf{F}_{k,a,b}\right\|_{2t}^{2t}$ 设定上界，只需给 $\mathbb{E}\left\|\mathbf{U}_{k,a,b}\right\|_t^t$ 和 $\mathbb{E}\left\|\mathbf{V}_{k,a,b}\right\|_t^t$ 设定上界。事实上，这正是我们的策略。为了给 $\mathbb{E}\left\|\mathbf{U}_{k,a,b}\right\|_t^t$ 设定上界，我们将通过下文定义的矩阵来为它设定上界。

定义 2.7.12 (矩阵 $\boldsymbol{\Delta}_{1}^{k,\ a\ b\ ,}, \boldsymbol{\Delta}_{2}^{k,\ a\ b\ ,}, \boldsymbol{\Delta}_{3}^{k,\ a\ b\ ,})$。定义这些矩阵

$$\boldsymbol{\Delta}_1^{k,a,b}=\mathbb{E}[((\mathbf{D}(Z)-\mathbf{D}(Z'))\overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z))^2|Z]$$

$$\boldsymbol{\Delta}_2^{k,a,b}=\mathbb{E}[(\mathbf{D}(Z)(\overline{\mathbf{G}}_{k,a,b}(Z)-\overline{\mathbf{G}}_{k,a,b}(Z'))\mathbf{D}(Z))^2|Z]$$

$$\boldsymbol{\Delta}_3^{k,a,b}=\mathbb{E}[(\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)(\mathbf{D}(Z)-\mathbf{D}(Z')))^2|Z]$$

引理 2.7.13。$\mathbf{U}_{k,\ a,\ b} \preceq 3\ (\boldsymbol{\Delta}_{1}^{k,\ \ a\ b}+\boldsymbol{\Delta}_{2}^{k,\ \ a\ b}+\boldsymbol{\Delta}_{3}^{k,\ \ a\ b,})$。

为证明这个引理，我们将使用下面的引理。

引理 2.7.14. 我们有如下关系

$$(\mathbf{D}(Z)-\mathbf{D}(Z'))(\overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z)-\overline{\mathbf{G}}_{k,a,b}(Z')\mathbf{D}(Z'))=0$$

$$(\overline{\mathbf{G}}_{k,a,b}(Z)-\overline{\mathbf{G}}_{k,a,b}(Z'))(\mathbf{D}(Z)-\mathbf{D}(Z'))=0$$

证明概要。 证明与引理 2.7.6 中第三个等式的证明类似。当 Z' 取为 $Z_{(i)}$ 时，对于某个 $i\leq n$，若 $\mathbf{D}(Z)$ $-\mathbf{D}(Z')$ 的某个对角元非零，则 $\mathbf{G}_{k,a,b}(Z)\mathbf{D}(Z)$ $-\mathbf{G}_{k,a,b}(Z')\mathbf{D}(Z')$ 中对应的行将为 0。第二个等式的证明类似。 □

■

引理 *2.7.13* 的证明。 我们有

$$\begin{aligned}
&(\overline{\mathbf{F}}_{k,a,b}(Z)-\overline{\mathbf{F}}_{k,a,b}(Z'))^2\\
&=(\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z)-\mathbf{D}(Z')\overline{\mathbf{G}}_{k,a,b}(Z')\mathbf{D}(Z'))^2\\
&=\Big(\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)(\mathbf{D}(Z)-\mathbf{D}(Z'))+\mathbf{D}(Z)(\overline{\mathbf{G}}_{k,a,b}(Z)-\overline{\mathbf{G}}_{k,a,b}(Z'))\mathbf{D}(Z')\\
&\qquad+(\mathbf{D}(Z)-\mathbf{D}(Z'))\overline{\mathbf{G}}_{k,a,b}(Z')\mathbf{D}(Z')\Big)^2\\
&=\Big(\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)(\mathbf{D}(Z)-\mathbf{D}(Z'))+\mathbf{D}(Z)(\overline{\mathbf{G}}_{k,a,b}(Z)-\overline{\mathbf{G}}_{k,a,b}(Z'))\mathbf{D}(Z)\\
&\qquad+(\mathbf{D}(Z)-\mathbf{D}(Z'))\overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z)\Big)^2
\end{aligned}$$

其中最后一个等式由引理 2.7.14 得出。对 Z取条件期望并应用事实 2.2.2，我们立即得到 $\mathbf{U}_{k,a,b}\preceq 3(\mathbf{\Delta}^{k}_{,\ 1\ a\ b}+\mathbf{\Delta}^{k,,}_{\ 2\ a\ b}+\mathbf{\Delta}^{k,,}_{\ 3\ a\ b,})$。 ■

在后续各节中，我们将证明关于迄今所考虑的矩阵的以下技术性界。

引理 2.7.15. 对于所有整数 $t \geq 1$,

$$\mathbb{E}\left\|\boldsymbol{\Delta}_2^{k,a,b}\right\|_t^t \leq \frac{(2d_p)^t}{n^t}(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})$$

引理 2.7.16. 对于所有整数 $t \geq 1$,

$$\mathbb{E}\left\|\mathbf{V}_{k,a,b}\right\|_t^t \leq (2d_p)^t n^t(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})$$

引理 2.7.17. 对于所有整数 $t \geq 1$,

$$\mathbb{E}\left\|\boldsymbol{\Delta}_1^{k,a,b}\right\|_t^t \leq \frac{(8dd_p)^t}{n^t}\,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}$$

引理 2.7.18. 对于所有整数 $t \geq 1$,

$$\mathbb{E}\left\|\boldsymbol{\Delta}_3^{k,a,b}\right\|_t^t \leq \frac{(4d_p)^t}{n^t}\,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}$$

假设上述引理成立，我们便可以完成引理 2.6.8 的证明。为方便起见，我们重新陈述该引理。

Lemma 2.6.8. 对于所有整数 $t \geq 1$, 整数 $k \geq 1, a, b \geq 0$ 使得 $a + b < k$,

$$\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} \leq (Ct^2 dd_p^2)^t(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})$$

对于某个绝对常数 $C > 0$。

引理 *2.6.8* 的证明。利用引理 2.7.10 和引理 2.7.13，我们得到，对于任意 $s>0$,

$$\begin{aligned}\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} &\leq (\frac{2t-1}{4})^t\,\mathbb{E}\left\|s\mathbf{U}_{k,a,b}+s^{-1}\mathbf{V}_{k,a,b}\right\|_t^t \\ &\leq t^t(s^t\,\mathbb{E}\left\|\mathbf{U}_{k,a,b}\right\|_t^t+s^{-t}\,\mathbb{E}\left\|\mathbf{V}_{k,a,b}\right\|_t^t) \\ &\leq (9st)^t(\mathbb{E}\left\|\mathbf{\Delta}_1^{k,a,b}\right\|_t^t+\mathbb{E}\left\|\mathbf{\Delta}_2^{k,a,b}\right\|_t^t+\mathbb{E}\left\|\mathbf{\Delta}_3^{k,a,b}\right\|_t^t)+t^ts^{-t}\,\mathbb{E}\left\|\mathbf{V}_{k,a,b}\right\|_t^t\end{aligned}$$

令 $\rho=s/n$。由于该不等式对任意选择的 $s>0$ 都成立，因此对任意选择的 $\rho>0$ 也成立。现在，利用引理 2.7.17、引理 2.7.18,

$$\begin{aligned}(9st)^t(\mathbb{E}\left\|\mathbf{\Delta}_1^{k,a,b}\right\|_t^t+\mathbb{E}\left\|\mathbf{\Delta}_3^{k,a,b}\right\|_t^t) &\leq (9st)^t\left(\frac{(8dd_p)^t}{n^t}+\frac{(4d_p)^t}{n^t}\right)\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} \\ &=\rho^t(C_1tdd_p)^t\,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}\end{aligned}$$

其中 $C_1>0$ 是一个绝对常数。利用引理 2.7.15、引理 2.7.16,

$$\begin{aligned}(9st)^t\,\mathbb{E}\left\|\mathbf{\Delta}_2^{k,a,b}\right\|_t^t+t^ts^{-t}\,\mathbb{E}\left\|\mathbf{V}_{k,a,b}\right\|_t^t &\leq \left((9st)^t\frac{(2d_p)^t}{n^t}+t^ts^{-t}(2d_p)^tn^t\right)(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t}+\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t}) \\ &\leq(\rho^tC_2^t+\rho^{-t}C_3^t)(td_p)^t(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t}+\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})\end{aligned}$$

其中 C_2、， $C_3>0$ 是绝对常数。因此,

$$\begin{aligned}\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} &\leq \rho^t(C_1tdd_p)^t\,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} \\ &\quad+(\rho^tC_2^t+\rho^{-t}C_3^t)(td_p)^t(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t}+\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})\end{aligned}$$

我们选择 $\rho>0$ 使得 $\rho^t(C_1tdd_p)^t=\frac{1}{2}$，从而得到

$$\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}\leq\frac{1}{2}\,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}+\frac{1}{2}(Ct^2dd_p^2)^t(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t}+\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})$$

其中 $C>0$ 为绝对常数。重新排列即可得到所需结果。 ■

2.7.3 界定 $\boldsymbol{\Delta}_{2}^{k,a\,b}$ 和 $\mathbf{V}_{k,a,b}$

下面的引理将 $\mathbf{V}_{k,a,b}$ 与 $\boldsymbol{\Delta}_{2}^{k,a\,b}$ 联系起来，二者相差一个因子 n^2，这对我们来说已经足够。接下来，我们可以专注于界定 $\boldsymbol{\Delta}_{2}^{k,a\,b}$。

引理 2.7.19. $\mathbf{V}_{k,a,b} \preceq n^2 \boldsymbol{\Delta}_{2}^{k,a\,b}$，

证明。 根据引理 2.7.5,

$$\begin{aligned}
\mathbf{V}_{k,a,b} &= \mathbb{E}[(\mathbf{D}(Z)\overline{\mathbf{K}}_{k,a,b}(Z,Z')\mathbf{D}(Z))^2|Z] \\
&= \mathbb{E}[(\mathbf{D}(Z)\left(\frac{n}{k-a-b}(\overline{\mathbf{G}}_{k,a,b}(Z)-\overline{\mathbf{G}}_{k,a,b}(Z'))\right)\mathbf{D}(Z))^2|Z] \\
&\preceq n^2\,\mathbb{E}[(\mathbf{D}(Z)(\overline{\mathbf{G}}_{k,a,b}(Z)-\overline{\mathbf{G}}_{k,a,b}(Z'))\mathbf{D}(Z))^2|Z] \\
&= n^2\boldsymbol{\Delta}_2^{k,a,b}
\end{aligned}$$

■

对于 $1 \le i \le n$ 和 $1 \le l \le d$，令 $\mathbf{e}_{i,l} \in \mathbb{N}^n$ 表示向量 α，其满足 $\alpha_i = l$ 和 $\alpha_j = 0$，其中 $j \neq i$。我们注意到以下简单命题。

命题 2.7.20. 对于任意多项式 f，若 Z_i 的度至多为 d，

$$f(Z) - f(Z^{(i)}) = \sum_{1\le l\le d}(Z_i^l - \widetilde{Z}_i^{\,l})\nabla_{\mathbf{e}_{i,l}}(f)$$

现在我们重新陈述并证明引理 2.7.15。

引理 2.7.15. 对于所有整数 $t \ge 1$,

$$\mathbb{E}\left\|\boldsymbol{\Delta}_2^{k,a,b}\right\|_t^t \le \frac{(2d_p)^t}{n^t}(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})$$

证明。 考虑

$$\begin{aligned}\mathbf{\Delta}_2^{k,a,b} &= \mathbb{E}[(\mathbf{D}(Z)(\overline{\mathbf{G}}_{k,a,b}(Z) - \overline{\mathbf{G}}_{k,a,b}(Z'))\mathbf{D}(Z))^2|Z] \\ &= \mathbb{E}\left[\begin{bmatrix}\mathbf{M}\mathbf{M}^\intercal & 0 \\ 0 & \mathbf{M}^\intercal\mathbf{M}\end{bmatrix}|Z\right] \\ &= \begin{bmatrix}\mathbb{E}[\mathbf{M}\mathbf{M}^\intercal|Z] & 0 \\ 0 & \mathbb{E}[\mathbf{M}^\intercal\mathbf{M}|Z]\end{bmatrix}\end{aligned}$$

wh其中 $\mathbf{M} = \mathbf{D}_1(Z)(\mathbf{G}_{k,a,b}(Z) - \mathbf{G}_{k,a,b}(Z'))\mathbf{D}_2(Z)$. 利用命题 2.7.20 ,

$$\begin{aligned}\mathbb{E}[\mathbf{M}\mathbf{M}^T|Z] &= \mathbb{E}[\mathbf{D}_1(Z)(\mathbf{G}_{k,a,b}(Z) - \mathbf{G}_{k,a,b}(Z'))\mathbf{D}_2(Z)\cdot\mathbf{D}_2(Z)(\mathbf{G}_{k,a,b}(Z) - \mathbf{G}_{k,a,b}(Z'))^\intercal\mathbf{D}_1(Z)|Z] \\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\mathbf{D}_1(Z)(\mathbf{G}_{k,a,b}(Z) - \mathbf{G}_{k,a,b}(Z^{(i)}))\mathbf{D}_2(Z)\cdot\mathbf{D}_2(Z)(\mathbf{G}_{k,a,b}(Z) - \mathbf{G}_{k,a,b}(Z^{(i)}))^\intercal\mathbf{D}_1(Z)|Z] \\ &= \frac{1}{n}\sum_{i=1}^n\sum_{l=1}^d \mathbb{E}[(Z_i^l - \widetilde{Z}_i^l)^2|Z]\cdot\mathbf{D}_1(Z)(\nabla_{\mathbf{e}_{i,l}}\mathbf{G}_{k,a,b})(Z)\mathbf{D}_2(Z)\cdot\mathbf{D}_2(Z)(\nabla_{\mathbf{e}_{i,l}}\mathbf{G}_{k,a,b})(Z)^\intercal\mathbf{D}_1(Z)\end{aligned}$$

定义 $\mathbf{N}_{i,l}(Z) := \mathbf{D}_1(Z)(\nabla_{\mathbf{e}_{i,l}}\mathbf{G}_{k,a,b})(Z)\mathbf{D}_2(Z)$。于是,

$$\begin{aligned}\mathbb{E}[\mathbf{M}\mathbf{M}^T|Z] &= \frac{1}{n}\sum_{i=1}^n\sum_{l=1}^d \mathbb{E}[(Z_i^l - \widetilde{Z}_i^l)^2|Z]\cdot\mathbf{N}_{i,l}(Z)\mathbf{N}_{i,l}(Z)^\intercal \\ &\preceq \frac{2}{n}\sum_{i=1}^n\sum_{l=1}^d (Z_i^{2l} + \mathbb{E}[Z_i^{2l}])\cdot\mathbf{N}_{i,l}(Z)\mathbf{N}_{i,l}(Z)^\intercal\end{aligned}$$

类似地,

$$\mathbb{E}[\mathbf{M}^\intercal\mathbf{M}|Z] \preceq \frac{2}{n}\sum_{i=1}^n\sum_{l=1}^d (Z_i^{2l} + \mathbb{E}[Z_i^{2l}])\cdot\mathbf{N}_{i,l}(Z)^\intercal\mathbf{N}_{i,l}(Z)$$

断言 2.7.21. 我们有以下关系

$$\sum_{i=1}^n\sum_{l=1}^d (Z_i^{2l} + \mathbb{E}[Z_i^{2l}])\cdot\mathbf{N}_{i,l}(Z)\mathbf{N}_{i,l}(Z)^\intercal = (b+1)\mathbf{F}_{k,a,b+1}\mathbf{F}_{k,a,b+1}^\intercal$$

$$\sum_{i=1}^{n}\sum_{l=1}^{d}(Z_i^{2l}+\mathbb{E}[Z_i^{2l}])\cdot\mathbf{N}_{i,l}(Z)^{\intercal}\mathbf{N}_{i,l}(Z)=(a+1)\mathbf{F}_{k,a+1,b}^{\intercal}\mathbf{F}_{k,a+1,b}$$

利用这一断言，我们得到

$$\mathbb{E}[\mathbf{M}\mathbf{M}^T|Z]\preceq\frac{2(b+1)}{n}\mathbf{F}_{k,a,b+1}\mathbf{F}_{k,a,b+1}^{\intercal}\preceq\frac{2d_p}{n}\mathbf{F}_{k,a,b+1}\mathbf{F}_{k,a,b+1}^{\intercal}$$

$$\mathbb{E}[\mathbf{M}^{\intercal}\mathbf{M}|Z]\preceq\frac{2(a+1)}{n}\mathbf{F}_{k,a+1,b}^{\intercal}\mathbf{F}_{k,a+1,b}\preceq\frac{2d_p}{n}\mathbf{F}_{k,a+1,b}^{\intercal}\mathbf{F}_{k,a+1,b}$$

因此，

$$\begin{aligned}\mathbb{E}\left\|\mathbf{\Delta}_2^{k,a,b}\right\|_t^t&=\mathbb{E}\left\|\mathbb{E}[\mathbf{M}\mathbf{M}^{\intercal}|Z]\right\|_t^t+\mathbb{E}\left\|\mathbb{E}[\mathbf{M}^{\intercal}\mathbf{M}|Z]\right\|_t^t\\&\le\frac{(2d_p)^t}{n^t}(\mathbb{E}\left\|\mathbf{F}_{k,a,b+1}\right\|_{2t}^{2t}+\mathbb{E}\left\|\mathbf{F}_{k,a+1,b}\right\|_{2t}^{2t})\\&\le\frac{(2d_p)^t}{n^t}(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t}+\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})\end{aligned}$$

■

剩下的工作是证明该断言。

断言 2.7.21 的证明。我们将证明第一个关系，第二个关系与此类似。固定 $i\le n$、$l\le d$，考虑 $\sum_{i=1}^{n}\sum_{l=1}^{d}(Z_i^{2l}+\mathbb{E}[Z_i^{2l}])\mathbf{N}_{i,l}(Z)\mathbf{N}_{i,l}(Z)$ 的任意非零项 $[(I_1,\alpha_1,\gamma_1)、(I_2,\alpha_2,\gamma_2)]$，其中 $I_1、I_2\in\mathcal{I}$、$(\alpha_1,\gamma_1)、(\alpha_2,\gamma_2)\in\mathcal{K}$。必有 $|\alpha_1|_0=|\alpha_2|_0=a$，在这种情形下，该矩阵项等于

$$\begin{aligned}\sum_{\substack{(J,\alpha_3,\gamma_3)\in\mathcal{J}\times\mathcal{K}\\|\alpha_3|=b\\\alpha_1\alpha_3=\alpha_2\alpha_3=0}}&(Z_i^{2l}+\mathbb{E}[Z_i^{2l}])\cdot(\sqrt{\mathbb{E}[Z^{2\alpha_1\cdot(1-\gamma_1)+2\alpha_3\cdot(1-\gamma_3)}]}Z^{\alpha_1\cdot\gamma_1+\alpha_3\cdot\gamma_3}\nabla_{\mathbf{e}_{i,l}}\nabla_{\alpha_1+\alpha_3}\mathbf{X}_k[I_1,J])\\&\cdot(\sqrt{\mathbb{E}[Z^{2\alpha_2\cdot(1-\gamma_2)+2\alpha_3\cdot(1-\gamma_3)}]}Z^{\alpha_2\cdot\gamma_2+\alpha_3\cdot\gamma_3}\nabla_{\mathbf{e}_{i,l}}\nabla_{\alpha_2+\alpha_3}\mathbf{X}_k[I_2,J])\end{aligned}$$

注意，求和中的项仅在$\mathbf{e}_{i,\,l}\cdot(\alpha_1+\alpha_3)=\mathbf{e}_{i,\,l}\cdot(\alpha_2+$

$\alpha_3) = 0$。因此，该求和可以写成

$$\sum_{\substack{(J,\alpha_3,\gamma_3)\in\mathcal{J}\times\mathcal{K}\\ |\alpha_3|=b+1\\ \mathbf{e}_{i,l}\preceq\alpha_3,\alpha_1\alpha_3=\alpha_2\alpha_3=0}} (\sqrt{\mathbb{E}[Z^{2\alpha_1\cdot(1-\gamma_1)+2\alpha_3\cdot(1-\gamma_3)}]}Z^{\alpha_1\cdot\gamma_1+\alpha_3\cdot\gamma_3}\nabla_{\alpha_1+\alpha_3}\mathbf{X}_k[I_1,J]) \cdot (\sqrt{\mathbb{E}[Z^{2\alpha_2\cdot(1-\gamma_2)+2\alpha_3\cdot(1-\gamma_3)}]}Z^{\alpha_2\cdot\gamma_2+\alpha_3\cdot\gamma_3}\nabla_{\alpha_2+\alpha_3}\mathbf{X}_k[I_2,J])$$

将这个矩阵项对所有 $i\le n, l\le d$ 求和后，可简化为

$$(b+1)\cdot\sum_{\substack{(J,\alpha_3,\gamma_3)\in\mathcal{J}\times\mathcal{K}\\ |\alpha_3|=b+1\\ \alpha_1\alpha_3=\alpha_2\alpha_3=0}} (\sqrt{\mathbb{E}[Z^{2\alpha_1\cdot(1-\gamma_1)+2\alpha_3\cdot(1-\gamma_3)}]}Z^{\alpha_1\cdot\gamma_1+\alpha_3\cdot\gamma_3}\nabla_{\alpha_1+\alpha_3}\mathbf{X}_k[I_1,J]) \cdot (\sqrt{\mathbb{E}[Z^{2\alpha_2\cdot(1-\gamma_2)+2\alpha_3\cdot(1-\gamma_3)}]}Z^{\alpha_2\cdot\gamma_2+\alpha_3\cdot\gamma_3}\nabla_{\alpha_2+\alpha_3}\mathbf{X}_k[I_2,J])$$

$(b+1)$ 这个因子是因为指标 i 可以从 α_3 中所有活跃指标里选取。但这恰好是 (I_1,α_1,γ_1), (I_2,α_2,γ_2)] 恒等矩阵 $(b+1)\mathbf{F}_{k,a,b+1}\mathbf{F}_{k,a,b+1}$ 的矩阵项，从而证明了该结论。

T ■

我们重述并证明引理 2.7.16。

引理 2.7.16。对于所有整数 $t\ge 1$,

$$\mathbb{E}\left\|\mathbf{V}_{k,a,b}\right\|_t^t \le (2d_p)^t n^t(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})$$

证明。利用引理 2.7.19 和引理 2.7.15，我们得到

$$\begin{aligned}\mathbb{E}\left\|\mathbf{V}_{k,a,b}\right\|_t^t &\le n^{2t}\,\mathbb{E}\left\|\mathbf{\Delta}_2^{k,a,b}\right\|_t^t\\ &\le (2d_p)^t n^t(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})\end{aligned}$$

■

2.7.4 界定 $\boldsymbol{\Delta}^{k,a,b}_1$ 和 $\boldsymbol{\Delta}^{k,a,b}_3$

将 $\sqcup$ 定义为集合的不交并。对于 $1 \le i \le n$ 和 $1 \le l \le d$，定义对角矩阵 $\boldsymbol{\Pi}_{i,l}$，$\boldsymbol{\Pi}'_{i,l}$，$\boldsymbol{\Pi}_i$，$\boldsymbol{\Pi}'_i \in \mathbb{R}^{(\mathcal{I}\times\mathcal{K})\sqcup(\mathcal{J}\times\mathcal{K})\times(\mathcal{I}\times\mathcal{K})\sqcup(\mathcal{J}\times\mathcal{K})}$（维度与 $\mathbf{D}$ 相同）为

$$\boldsymbol{\Pi}_{i,l}[(I,\alpha,\beta),(I,\alpha,\beta)] = \begin{cases} 1 & \text{if } (\alpha\cdot\gamma)_i \neq 0 \text{ and } \alpha_i = l \\ 0 & \text{o.w.} \end{cases} \qquad \boldsymbol{\Pi}_{i}[(I,\alpha,\beta),(I,\alpha,\beta)] = \begin{cases} 1 & \text{if } (\alpha\cdot\gamma)_i \neq 0 \\ 0 & \text{o.w.} \end{cases}$$

$$\boldsymbol{\Pi}'_{i,l}[(I,\alpha,\beta),(I,\alpha,\beta)] = \begin{cases} 1 & \text{if } \alpha_i \neq 0 \text{ and } \alpha_i = l \\ 0 & \text{o.w.} \end{cases} \qquad \boldsymbol{\Pi}'_{i}[(I,\alpha,\beta),(I,\alpha,\beta)] = \begin{cases} 1 & \text{if } \alpha_i \neq 0 \\ 0 & \text{o.w.} \end{cases}$$

对于所有 $I \in \mathcal{I}\sqcup\mathcal{J}$。注意，对于所有 $i \le n$、$\boldsymbol{\Pi}_i = \sum_{l=1}^{d} \boldsymbol{\Pi}_{i,l}$。

此外，对于所有 $1 \le i \le n$，我们将置换矩阵 $\boldsymbol{\Sigma}_i \in \mathbb{R}^{(\mathcal{I}\times\mathcal{K})\sqcup(\mathcal{J}\times\mathcal{K})\times(\mathcal{I}\times\mathcal{K})\sqcup(\mathcal{J}\times\mathcal{K})}$ 定义如下。考虑 $\mathcal{I}\times\mathcal{K}$ 上的置换 σ_1，该置换对所有满足 $\alpha_i \neq 0$ 的 $(I,\ \alpha,\ \gamma)$ 交换 $(I,\ \alpha,\ \gamma)$ 和 $(I,\ \alpha,\ \gamma+\mathsf{e}_i)$。这里，$\mathsf{e}_i \in \{0,1\}^n$ 恰好有一个非零项，位于第 i 个位置，而 $\gamma+\mathsf{e}_i$ 是 $\mathbb{F}_2$ 上通常的加法。σ_1 保持其他位置不变。令 $\boldsymbol{\Sigma}^{(1)}_i$ 为 σ 的置换矩阵。类似地，令 $\boldsymbol{\Sigma}^{(2)}_i$ 为 $\mathcal{J}\times\mathcal{K}$ 上置换 σ_2 的置换矩阵，该置换对所有满足 $\alpha_i \neq 0$ 的 $(J,\ \alpha,\ \gamma)$ 交换 $(J,\ \alpha,\ \gamma)$ 和 $(J,\ \alpha,\ \gamma+\mathsf{e}_i)$，并保持所有其他位置不变。然后，我们定义 $\boldsymbol{\Sigma}_i = \begin{bmatrix} \boldsymbol{\Sigma}^{(1)}_i & 0 \\ 0 & \boldsymbol{\Sigma}^{(2)}_i \end{bmatrix}$。以下事实很容易验证。

事实 2.7.22. $\boldsymbol{\Pi}'_{i,l}\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i\boldsymbol{\Pi}'_{i,l}$ 和 $\boldsymbol{\Pi}'_i\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i\boldsymbol{\Pi}'_i$。

现在我们可以证明引理 2.7.17，为方便起见，现重述如下 .

引理 2.7.17. 对于所有整数 $t \ge 1$，

$$\mathbb{E}\left\|\boldsymbol{\Delta}^{k,a,b}_1\right\|^t_t \le \frac{(8dd_p)^t}{n^t}\,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|^{2t}_{2t}$$

证明。 首先,

$$\begin{aligned}\boldsymbol{\Delta}_1^{k,a,b} &= \mathbb{E}[((\mathbf{D}(Z)-\mathbf{D}(Z'))\overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z))^2|Z] \\ &= \mathbb{E}[(\mathbf{D}(Z)-\mathbf{D}(Z'))\overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z)\cdot\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)(\mathbf{D}(Z)-\mathbf{D}(Z'))|Z] \\ &= \mathbb{E}[(\mathbf{D}(Z)-\mathbf{D}(Z'))\mathbf{M}(Z)(\mathbf{D}(Z)-\mathbf{D}(Z'))|Z]\end{aligned}$$

其中我们定义 $\mathbf{M}(Z) = \mathbf{G}_{\overline{k},a,b}(Z)\mathbf{D}(Z)\cdot\mathbf{D}(Z)\mathbf{G}_{k,a,\overline{b}}(Z)$。回忆一下， $Z' = Z(i)$ 对某个从 $[n]$ 中均匀随机选取的 i 成立。注意到，对所有 i，$\mathbf{D}(Z) - \mathbf{D}(Z(i)) = \boldsymbol{\Pi}_i(\mathbf{D}(Z) - \mathbf{D}(Z(i)))$，我们得到

$$\begin{aligned}\boldsymbol{\Delta}_1^{k,a,b} &= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[(\mathbf{D}(Z)-\mathbf{D}(Z^{(i)}))\mathbf{M}(Z)(\mathbf{D}(Z)-\mathbf{D}(Z^{(i)}))]|Z] \\ &= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i(\mathbf{D}(Z)-\mathbf{D}(Z^{(i)}))\mathbf{M}(Z)(\mathbf{D}(Z)-\mathbf{D}(Z^{(i)}))\boldsymbol{\Pi}_i]|Z] \\ &\preceq 2\left(\mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i\mathbf{D}(Z)\mathbf{M}(Z)\mathbf{D}(Z)\boldsymbol{\Pi}_i]|Z] + \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i\mathbf{D}(Z^{(i)})\mathbf{M}(Z)\mathbf{D}(Z^{(i)})\boldsymbol{\Pi}_i]|Z]\right) \\ &\preceq 2\left(\mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i\overline{\mathbf{F}}_{k,a,b}^2\boldsymbol{\Pi}_i] + \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i\mathbf{D}(Z^{(i)})\mathbf{M}(Z)\mathbf{D}(Z^{(i)})\boldsymbol{\Pi}_i]|Z]\right) \\ &\preceq 2(\boldsymbol{\Delta}_{10}+\boldsymbol{\Delta}_{11})\end{aligned}$$

其中我们定义

$$\boldsymbol{\Delta}_{10} = \mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i\overline{\mathbf{F}}_{k,a,b}^2\boldsymbol{\Pi}_i], \qquad \boldsymbol{\Delta}_{11} = \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i\mathbf{D}(Z^{(i)})\mathbf{M}(Z)\mathbf{D}(Z^{(i)})\boldsymbol{\Pi}_i]|Z]$$

在区间 $[0,\infty)$ 上对凸连续函数 $f(x) = x^t$ 应用引理 2.2.4，$\mathbf{B}_i = \mathbf{F}_{k,}^2\ a,\ b,\ \mathbf{A}_i = \frac{1}{\sqrt{d_p}}\boldsymbol{\Pi}_i$ 其中我们注意到 $\sum_{i=1}^n \mathbf{A}_i\mathbf{A}_i^T = \frac{1}{d_p}\sum_{i=1}^n \boldsymbol{\Pi}_i^2 \preceq$ **恒等矩阵**，得到 —

$$\begin{aligned}
\mathbb{E}\,\|\boldsymbol{\Delta}_{10}\|_t^t = \mathbb{E}\,\mathrm{tr}[\boldsymbol{\Delta}_{10}^t] = \mathbb{E}\,\mathrm{tr}[\Big(\mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i\overline{\mathbf{F}}_{k,a,b}^2\boldsymbol{\Pi}_i]\Big)^t] &= \frac{1}{n^t}\,\mathbb{E}\,\mathrm{tr}[\Big(\sum_{i=1}^n \boldsymbol{\Pi}_i\overline{\mathbf{F}}_{k,a,b}^2\boldsymbol{\Pi}_i\Big)^t] \\
&\le \frac{d_p^{t-1}}{n^t}\,\mathbb{E}\,\mathrm{tr}[\Big(\sum_{i=1}^n \boldsymbol{\Pi}_i\overline{\mathbf{F}}_{k,a,b}^{2t}\boldsymbol{\Pi}_i\Big)] \\
&\le \frac{d_p^{t-1}}{n^t}\,\mathbb{E}\,\mathrm{tr}[\Big(\sum_{i=1}^n \boldsymbol{\Pi}_i^2\Big)\overline{\mathbf{F}}_{k,a,b}^{2t}] \\
&\le \frac{d_p^{t}}{n^t}\,\mathbb{E}\,\mathrm{tr}[\overline{\mathbf{F}}_{k,a,b}^{2t}] \\
&= \frac{d_p^{t}}{n^t}\,\mathbb{E}\,\big\|\overline{\mathbf{F}}_{k,a,b}\big\|_{2t}^{2t}
\end{aligned}$$

现在，考虑

$$\begin{aligned}
\boldsymbol{\Delta}_{11} &= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i\mathbf{D}(Z^{(i)})\mathbf{M}(Z)\mathbf{D}(Z^{(i)})\boldsymbol{\Pi}_i]|Z] \\
&= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[(\sum_{l=1}^d \boldsymbol{\Pi}_{i,l})\mathbf{D}(Z^{(i)})\mathbf{M}(Z)\mathbf{D}(Z^{(i)})(\sum_{l=1}^d \boldsymbol{\Pi}_{i,l})]|Z] \\
&\preceq d\cdot\mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\sum_{l=1}^d \boldsymbol{\Pi}_{i,l}\mathbf{D}(Z^{(i)})\mathbf{M}(Z)\mathbf{D}(Z^{(i)})\boldsymbol{\Pi}_{i,l}]|Z] \\
&= d\cdot\mathop{\mathbb{E}}_{i\in[n]}[\sum_{l=1}^d \frac{\mathbb{E}[Z_i^{2l}]}{Z_i^{2l}}\boldsymbol{\Pi}_{i,l}\mathbf{D}(Z)\mathbf{M}(Z)\mathbf{D}(Z)\boldsymbol{\Pi}_{i,l}] \\
&= \frac{d}{n}\sum_{i=1}^n\sum_{l=1}^d \frac{\mathbb{E}[Z_i^{2l}]}{Z_i^{2l}}\boldsymbol{\Pi}_{i,l}\mathbf{D}(Z)\mathbf{M}(Z)\mathbf{D}(Z)\boldsymbol{\Pi}_{i,l} \\
&= \frac{d}{n}\sum_{i=1}^n\sum_{l=1}^d \boldsymbol{\Pi}_{i,l}\boldsymbol{\Sigma}_i\mathbf{D}(Z)\mathbf{M}(Z)\mathbf{D}(Z)\boldsymbol{\Sigma}_i^\intercal\boldsymbol{\Pi}_{i,l} \\
&= \frac{d}{n}\sum_{i=1}^n\sum_{l=1}^d \boldsymbol{\Pi}_{i,l}\boldsymbol{\Sigma}_i\overline{\mathbf{F}}_{k,a,b}^2\boldsymbol{\Sigma}_i^\intercal\boldsymbol{\Pi}_{i,l}
\end{aligned}$$

现在，我们对 dd_p 项应用引理 2.2.4：$\mathbf{B}_{i,\ l} = \mathbf{F}_{k,\ a,\ b}^2$ 和 $\mathbf{A}_{i,\ l} = \frac{1}{\sqrt{d_p}}\boldsymbol{\Pi}_{i,\ l}\boldsymbol{\Sigma}_i$，其中

我们观察到

$$\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{A}_{i,l}\mathbf{A}_{i,l}^{T}=\frac{1}{d_p}\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Pi}_{i,l}\mathbf{\Sigma}_i\mathbf{\Sigma}_i^{\intercal}\mathbf{\Pi}_{i,l}^{\intercal}=\frac{1}{d_p}\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Pi}_{i,l}^{2}\preceq\mathbf{I}$$

得到

$$\begin{aligned}\mathbb{E}\left\|\mathbf{\Delta}_{11}\right\|_t^t=\mathbb{E}\,\mathrm{tr}[\mathbf{\Delta}_{11}^t]&\leq\frac{d^t}{n^t}\,\mathbb{E}\,\mathrm{tr}[(\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Pi}_{i,l}\mathbf{\Sigma}_i\overline{\mathbf{F}}_{k,a,b}^{2}\mathbf{\Sigma}_i^{\intercal}\mathbf{\Pi}_{i,l})^t]\\&\leq\frac{(dd_p)^t}{n^t}\,\mathbb{E}\,\mathrm{tr}[\left(\frac{1}{d_p}\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Pi}_{i,l}\mathbf{\Sigma}_i\overline{\mathbf{F}}_{k,a,b}^{2t}\mathbf{\Sigma}_i^{\intercal}\mathbf{\Pi}_{i,l}\right)]\\&=\frac{(dd_p)^t}{n^t}\,\mathbb{E}\,\mathrm{tr}[\left(\frac{1}{d_p}\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Sigma}_i^{\intercal}\mathbf{\Pi}_{i,l}\mathbf{\Pi}_{i,l}\mathbf{\Sigma}_i\overline{\mathbf{F}}_{k,a,b}^{2t}\right)]\end{aligned}$$

为简化这一过程，我们使用事实 2.7.22 得到

$$\begin{aligned}\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Sigma}_i^{\intercal}(\mathbf{\Pi}_{i,l})^2\mathbf{\Sigma}_i\preceq\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Sigma}_i^{\intercal}(\mathbf{\Pi}_{i,l}')^2\mathbf{\Sigma}_i=\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Pi}_{i,l}'\mathbf{\Sigma}_i^{\intercal}\mathbf{\Sigma}_i\mathbf{\Pi}_{i,l}'=\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Pi}_{i,l}'\mathbf{\Pi}_{i,l}'\\\preceq d_p\mathbf{I}\end{aligned}$$

因此，

$$\mathbb{E}\left\|\mathbf{\Delta}_{11}\right\|_t^t\leq\frac{(dd_p)^t}{n^t}\,\mathbb{E}\,\mathrm{tr}[\overline{\mathbf{F}}_{k,a,b}^{2t}]=\frac{(dd_p)^t}{n^t}\,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}$$

将它们合并，并使用事实 2.2.3，

$$\begin{aligned}\mathbb{E}\left\|\mathbf{\Delta}_1^{k,a,b}\right\|_t^t&\leq4^t(\mathbb{E}\left\|\mathbf{\Delta}_{10}\right\|_t^t+\mathbb{E}\left\|\mathbf{\Delta}_{11}\right\|_t^t)\\&\leq\frac{(8dd_p)^t}{n^t}\,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}\end{aligned}$$

■

现在我们重新陈述并证明引理 2.7.18。

引理 2.7.18. 对于所有整数 $t \geq 1$,

$$\mathbb{E}\left\|\boldsymbol{\Delta}_3^{k,a,b}\right\|_t^t \leq \frac{(4d_p)^t}{n^t}\,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}$$

证明。 回顾一下，对于从 $[n]$ 中均匀采样的 $Z' = Z_{(i)}$，有:

$$\begin{aligned}
\boldsymbol{\Delta}_3^{k,a,b} &= \mathbb{E}[(\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)(\mathbf{D}(Z)-\mathbf{D}(Z')))^2|Z] \\
&= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[(\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)(\mathbf{D}(Z)-\mathbf{D}(Z^{(i)})))^2]|Z] \\
&= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[(\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)\boldsymbol{\Pi}_i(\mathbf{D}(Z)-\mathbf{D}(Z^{(i)})))^2]|Z]
\end{aligned}$$

这里用到了如下事实：对于所有 i，$\mathbf{D}(Z) - \mathbf{D}(Z_{(i)}) = \boldsymbol{\Pi}_i(\mathbf{D}(Z) - \mathbf{D}(Z_{(i)}))$。定义 $\mathbf{M}(Z) = \mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}$ to get

$$\begin{aligned}
\boldsymbol{\Delta}_3^{k,a,b} &= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i(\mathbf{D}(Z)-\mathbf{D}(Z^{(i)}))^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal]|Z] \\
&\preceq 2(\mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z)^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal]|Z] + \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z^{(i)})^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal]|Z]) \\
&= 2(\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z)^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal] + \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z^{(i)})^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal]|Z]) \\
&= 2(\boldsymbol{\Delta}_{30} + \boldsymbol{\Delta}_{31})
\end{aligned}$$

其中我们定义

$$\boldsymbol{\Delta}_{30} = \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z)^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal], \qquad \boldsymbol{\Delta}_{31} = \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z^{(i)})^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal]|Z]$$

我们有

$$\begin{aligned}\boldsymbol{\Delta}_{30} = \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z)^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal] &= \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\mathbf{D}(Z)\boldsymbol{\Pi}_i\boldsymbol{\Pi}_i\mathbf{D}(Z)\mathbf{M}(Z)^\intercal] \\ &= \mathbf{M}(Z)\mathbf{D}(Z)(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Pi}_i^2)\mathbf{D}(Z)\mathbf{M}(Z)^\intercal \\ &\preceq \frac{d_p}{n}\mathbf{M}(Z)\mathbf{D}(Z)\mathbf{D}(Z)\mathbf{M}(Z)^\intercal \\ &= \frac{d_p}{n}\overline{\mathbf{F}}_{k,a,b}^2\end{aligned}$$

对于另一项，根据事实 2.7.22,

$$\begin{aligned}\boldsymbol{\Delta}_{31} &= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z^{(i)})^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal]|Z] \\ &= \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\boldsymbol{\Sigma}_i\mathbf{D}(Z)^2\boldsymbol{\Sigma}_i\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal] \\ &\preceq \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i'\boldsymbol{\Sigma}_i\mathbf{D}(Z)^2\boldsymbol{\Sigma}_i\boldsymbol{\Pi}_i'\mathbf{M}(Z)^\intercal] \\ &= \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Sigma}_i\boldsymbol{\Pi}_i'\mathbf{D}(Z)^2\boldsymbol{\Pi}_i'\boldsymbol{\Sigma}_i\mathbf{M}(Z)^\intercal] \\ &= \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}\boldsymbol{\Sigma}_i\boldsymbol{\Pi}_i'\mathbf{D}(Z)^2\boldsymbol{\Pi}_i'\boldsymbol{\Sigma}_i\overline{\mathbf{G}}_{k,a,b}\mathbf{D}(Z)]\end{aligned}$$

注意到 $\mathbf{G}_{k,\ a,\ \overline{b}}\boldsymbol{\Sigma}_i = \mathbf{G}_{k,\ \overline{a},\ b}$，因为 $\mathbf{G}$ 的元素只取决于 α 而不取决于 γ，所以对 γ 进行置换不会对矩阵产生任何影响。因此,

$$
\begin{aligned}
\mathbf{\Delta}_{31} &\preceq \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}\mathbf{\Pi}_i'\mathbf{D}(Z)^2\mathbf{\Pi}_i'\overline{\mathbf{G}}_{k,a,b}\mathbf{D}(Z)] \\
&\preceq \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}\mathbf{D}(Z)\mathbf{\Pi}_i'\mathbf{\Pi}_i'\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}\mathbf{D}(Z)] \\
&= \mathop{\mathbb{E}}_{i\in[n]}\overline{\mathbf{F}}_{k,a,b}\mathbf{\Pi}_i'\mathbf{\Pi}_i'\overline{\mathbf{F}}_{k,a,b} \\
&= \frac{1}{n}\sum_{i=1}^{n}\overline{\mathbf{F}}_{k,a,b}\mathbf{\Pi}_i'\mathbf{\Pi}_i'\overline{\mathbf{F}}_{k,a,b} \\
&\preceq \frac{d_p}{n}\overline{\mathbf{F}}_{k,a,b}^2
\end{aligned}
$$

其中我们用到了以下事实： $\sum_{i=1}^{n}\mathbf{\Pi}_i'\mathbf{\Pi}_i' \preceq d_p\mathbf{I}$。将它们合在一起，

$$
\mathbb{E}\left\|\mathbf{\Delta}_3^{k,a,b}\right\|_t^t \le 2^t(\mathbb{E}\left\|\mathbf{\Delta}_{30}\right\|_t^t + \mathbb{E}\left\|\mathbf{\Delta}_{31}\right\|_t^t) \le 2^t\cdot 2\frac{d_p^t}{n^t}\,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} \le \frac{(4d_p)^t}{n^t}\,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}
$$

■

2.8 应用：稀疏图矩阵

现在我们考虑稀疏图矩阵，即 $G\sim\mathcal{G}_{n,\ p}$ $p\le\frac{1}{2}$ 的情形。与稠密图矩阵的主要区别在于边因子的贡献。按照第2.5节中的说明，朴素地用每条边贡献的绝对值来界定时，形状中的每条边都会贡献一个 $\sqrt{\frac{1-p}{p}}$ 因子。但在许多情况下，这些界并不紧。事实上，即使在邻接矩阵这一基本情形中，这些界也不紧。在本节中，我们利用一般递归得到更紧的界。正如我们将看到的，改进后的界仅包含顶点分隔集内边对应的边因子。

设$\mathbf{M}_\tau$ 为与形状 τ 相对应的图矩阵，其中我们使用 p-偏置傅里叶特征 $G_{i,\ j}$。在本节中，我们将推导 $\mathbb{E}\left\|\mathbf{M}_\tau-\mathbb{E}\,\mathbf{M}_\tau\right\|_{2t}^{2t}$的界，并利用它得到

对 $\|\mathbf{M}_\tau\|$的高概率界。由于其中许多细节与第 2.4.2 节以及定理 2.4.9 的证明相似，我们将略过部分细节。建议读者先阅读该节。

其中的$G_{i,\ j}$对应于第 2.6 节中的 Z_is，而 $\mathbf{F}$ 对应于 $\mathbf{M}_\tau$。令 $\mathcal{I}$ 表示 $[n]$的子元组集合。$\mathbf{M}_\tau$ 的每个非零矩阵项都是次数为$|E(\tau)|$的齐次多项式。如果 $E(\tau) = \emptyset$，那么，$\mathbf{M}_\tau - \mathbb{E}\mathbf{M}_\tau = 0$，因此我们可以专注于 τ至少含有一条边的情形。此外，由于 $V(\tau)$ 中的度为 0 的顶点只会将矩阵缩放至多 n 倍，我们可以单独处理这些顶点；在主要分析中，我们假设 τ中不存在此类顶点。

我们将使用定理 2.6.6，但在我们的应用中，矩阵和定理陈述都可以大幅简化。沿用第 2.6 节的记号。由于我们处理的是多线性多项式，在 $\mathcal{K}$的定义中，我们可以将注意力限制在 $\alpha \in \{0,1\}^{\binom{n}{2}}$ 上，因为对于任何其他 $\alpha \in \mathbb{N}^n$，$\mathbf{G}_{a+b,\ a,\ b}$ 的相应行或列将为 0，因此$\mathbf{F}_{a+b,\ a,\ b}$也将为 0。所以，我们相应地重新定义 $\mathcal{K}$，使其只包含这些 $(\alpha,\ \gamma)$，从而$\mathcal{K} \subseteq \{0,1\}^n \times \{0,1\}^n$。

接下来，对角矩阵 $\mathbf{D}_1$,$\mathbf{D}_2$ 都将等于对角矩阵 $\mathbf{D} \in \mathbb{R}[Z]^{\mathcal{I}\times\mathcal{K}} \times \mathbb{R}[Z]^{\mathcal{I}\times\mathcal{K}}$，其非零项为

$$\mathbf{D}[(I,\alpha,\gamma),(I,\alpha,\gamma)] = \sqrt{\mathbb{E}[\prod_{i,j} G_{ij}^{2\alpha_{ij}(1-\gamma)_{ij}}]} \prod_{i,j} G_i^{\alpha_{ij}\gamma_{ij}} = \prod_{i,j} G_i^{\alpha_{ij}\gamma_{ij}}$$

这里我们用到了这样一个事实：对于任意 i,j, $\mathbb{E}[G_{ij}^2] = 1$。

对于满足 $a+b = |E(\tau)|$ 的整数 $a,\ b \geq 0$，定义矩阵 $\mathbf{M}_{\tau,\ a,\ b}$为矩阵$\mathbf{G}_{a+b,\ a,\ b}$。我们采用这一记号，以便与第 2.4.2 节保持一致。也就是说，$\mathbf{M}_{\tau,\ a,\ b}$的行和列由满足 $\mathcal{I}\times\mathcal{K}$ 的索引标记，对于所有 $(I,\ \alpha_1,\ \gamma_1)$，$(J,\ \alpha_2,\ \gamma_2) \in \mathcal{I}\times\mathcal{K}$,

$$\mathbf{M}_{\tau,a,b}[(I,\alpha_1,\gamma_1),(J,\alpha_2,\gamma_2)] = \begin{cases} \nabla_{\alpha_1+\alpha_2}\mathbf{M}_\tau[I,J] & \text{if } |\alpha_1|_0 = a, |\alpha_2|_0 = b, \alpha_1\cdot\alpha_2 = 0 \\ 0 & \text{o.w.} \end{cases}$$

这与第 2.4.2 节中定义的 $\mathbf{M}_{\tau,\ a,\ b}$ 矩阵几乎完全相同，区别在于现在行列索引中包含 γ。因此，对于满足 $|\alpha_1|_0 = a$，$|\alpha_2|_0 = b$，$\alpha_1 \cdot \alpha_2 = 0$ 的 I，$J \in \mathcal{I}$，$(\alpha_1,\ \gamma_1)$，$(\alpha_2,\ \gamma_2) \in \mathcal{K}$，第 $(I,\ \alpha_1,\ \gamma_1)$ 行、$(J,\ \alpha_2,\ \gamma_2)$ 列的矩阵项为 φ 的实现数目，其中 τ 满足

- U_τ，V_τ 在 φ 下分别映射到 I，J，并且
- 在 φ 下，将 τ 的边映射为 α_1 和 α_2 中的边，并将其视为一个集合.

根据定理 2.6.6，对于整数 $t \geq 1$，

$$\begin{aligned}\mathbb{E}\left\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\right\|_{2t}^{2t} &\leq \sum_{a,b\geq 0, a+b\geq 1} (Ct^2 d d_p^4)^{(a+b)t}\,\mathbb{E}\left\|\mathbf{F}_{a+b,a,b}\right\|_{2t}^{2t} \\ &= \sum_{a,b\geq 0, a+b=|E(\tau)|} (Ct^2|E(\tau)|^4)^{t|E(\tau)|}\,\mathbb{E}\left\|\mathbf{D}\mathbf{M}_{\tau,a,b}\mathbf{D}\right\|_{2t}^{2t}\end{aligned}$$

其中 $C > 0$ 是一个绝对常数。

现在，我们想要分析 $\mathbb{E}\left\|\mathbf{D}\mathbf{M}_{\tau,\ a,\ b}\mathbf{D}\right\|_{2t}^{2t}$。与定理 2.4.9 的证明一样，令 P 指明 $E(\tau)$ 中哪些边分别映射到 α_1，α_2，以及映射的顺序。此外，我们现在在 P 中存储额外信息，以指明 γ_1，γ_2（相对于 α_1，α_2）中的哪些元素被设为 1。将这类信息的集合 P 记为 $\mathcal{P}$，那么 $|\mathcal{P}| \leq (4|E(\tau)|)^{t|E(\tau)|}2^{|E(\tau)|}$。因此，

$$\mathbb{E}\left\|\mathbf{D}\mathbf{M}_{\tau,a,b}\mathbf{D}\right\|_{2t}^{2t} \leq (8|E(\tau)|)^{t|E(\tau)|}\sum_{P\in\mathcal{P}}\mathbb{E}\left\|\mathbf{D}\mathbf{M}_{\tau,a,b,P}\mathbf{D}\right\|_{2t}^{2t}$$

其中我们定义$\mathbf{M}_{\tau,\ a,\ b,\ P}$，其定义方式与$\mathbf{M}_{\tau,\ a,\ b}$ 类似，但还需满足额外条件：φ，α_1，α_2，γ_1，γ_2必须符合 P。

此时，与定理2.4.9的证明不同，请注意，这里的矩阵$\mathbf{M}_{\tau,a,b,P}$的行和列由 $\mathcal{I} \times \mathcal{K}$索引。我们将再次定义形状 τ_P，它等于矩阵$\mathbf{D}\mathbf{M}_{\tau,a,b,P}\mathbf{D}$的非零块，允许对行和

列重新命名。 $V(\tau_P)$ U_{τ_P}, V_{τ_P}的定义方式与第 2.4.2 节相同；但为纳入$\mathbf{D}$ 对这些项的作用，我们只保留在 γ_1 或 γ_2中活跃的边，具体由 P规定。示意见图2.6。

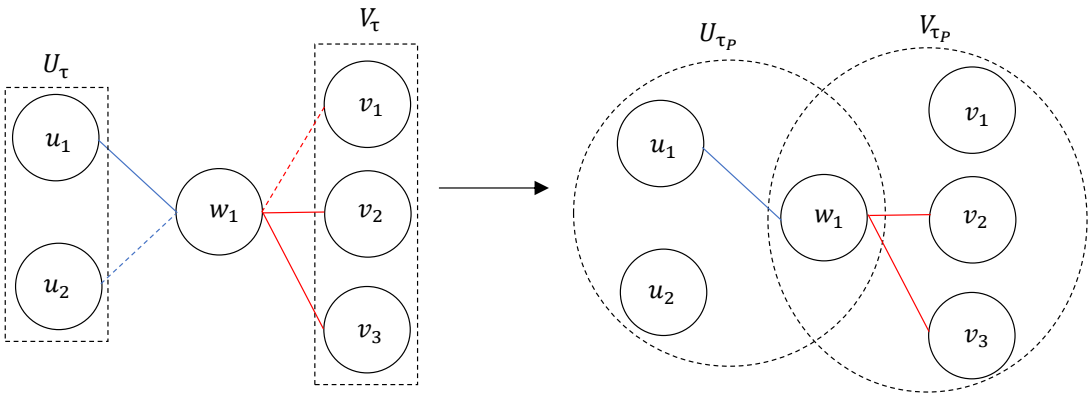


图 2.6：示例说明了如何定义 τ_P。在此示例中， P 约束蓝色边和红色边分别指向 α_1 和 α_2 。此外， P 表示在 γ_1 中， γ_2 有些边处于活动状态（以实线边表示），有些边则未处于活动状态（以虚线边表示），在 γ_1 中， γ_2也是如此。我们保留 τ_P 中的实线边。 U_{τ_P} 中， V_{τ_P}的顶点也有一个排序（此处未显示）。

然后，通过对 $\mathbf{DM}_\tau$、a、b、$P\mathbf{D}$ 的行和列进行类似的重新命名，并去掉 γ，我们得到 $\mathbf{M}_{\tau_P}$。因此，我们得到如下界

$$\mathbb{E}\left\|\mathbf{DM}_{\tau,a,b}\mathbf{D}\right\|_{2t}^{2t} \leq (8|E(\tau)|)^{t|E(\tau)|} \sum_{P\in\mathcal{P}} \mathbb{E}\left\|\mathbf{M}_{\tau_P}\right\|_{2t}^{2t}$$

我们希望分析矩阵 $\mathbf{M}_{\tau_P}$ 的范数界。注意， τ_P 是具有以下性质的形状

- 在 $V(\tau_P) \setminus U_{\tau_P} \setminus V_{\tau_P}$中没有顶点
- 每条边要么完全包含在 U_{τ_P} 中，要么完全包含在 V_{τ_P} 中

将这样的形状称为简单。

在下面的引理中，我们将证明简单形状的范数界；其证明推迟到下一节。回想一下，在引理 2.4.10 中，我们分析了没有边的简单形状的范数界（因为在这种情形下，图分布并不重要）。

对简单形状的分析非常相似，但这次我们使用标量集中性工具来界定 Frobenius 范数。

对于顶点集合 S，用 $E(S)$ 表示两个端点都在 S 中的边的集合。

引理 2.8.1。对于所有偶整数 $t \geq 2$，如果 τ 是一个简单形状，

$$\mathbb{E}\,\|\mathbf{M}_\tau\|_{2t}^{2t} \leq \left(n^{|V(\tau)|}(Ct)^{t|E(\tau)|}|V(\tau)|^{t|V(\tau)|}\right) \max_{U_\tau \cap V_\tau \subseteq S \subseteq V(\tau)} \left(\frac{1-p}{p}\right)^{t|E(S)|} n^{t(|V(\tau)|-|S|)}$$

对于某个绝对常数 $C > 0$。

对于简单形状，与相应稠密图矩阵的范数界相比，主要区别在于，S 内的每条边都会贡献一个 $\sqrt{\frac{1-p}{p}}$ 因子。在处理稀疏图矩阵时，边的贡献是不可避免的，但我们已经发现，无需考虑形状中的所有边，只需考虑其中的一个子集。

利用这个引理，我们可以得到一般图矩阵的范数界。我们回顾顶点分隔集的定义。

定义 2.4.8（顶点分隔集）。对于形状 τ，定义顶点分隔集为顶点 $S \subseteq V(\tau)$ 的一个子集，使得在 $\tau \setminus S$ 中不存在从 U_τ 到 V_τ 的路径；$\tau \setminus S$ 是删除 S 中所有顶点（包括与它们相 *incident* 的所有边）后得到的形状。

令 I_τ 为 $V(\tau) \setminus U_\tau \setminus V_\tau$ 中所有孤立顶点（度为 0 的顶点）的集合，因此它们本质上只是用一个标量因子缩放矩阵。现在我们陈述本节的主定理。

定理 2.8.2。对于所有偶整数 $t \geq 2$，以及任意形状 τ，

$$\begin{aligned}&\mathbb{E}\,\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\|_{2t}^{2t} \\ &\leq \left(n^{|V(\tau)|}|V(\tau)|^{t|V(\tau)|}(Ct^3|E(\tau)|^5)^{t|E(\tau)|}\right) \max_{\text{vertex separator } S} \left(\frac{1-p}{p}\right)^{t|E(S)|} n^{t(|V(\tau)|-|S|+|I_\tau|)}\end{aligned}$$

其中最大值取遍所有顶点分隔集 S。

为了解释这一界，如果我们假设 τ 中有常数个顶点，那么通过选择 $t \approx \mathrm{polylog}(n)$，可得

$$\|\mathbf{M}_\tau\| = \widetilde{\mathrm{O}}\left(\max_{\text{vertex separator } S} \left(\sqrt{\frac{1-p}{p}}\right)^{|E(S)|} \sqrt{n}^{|V(\tau)-|S|+|I_\tau|} \right)$$

~以高概率成立，其中 O 隐含了对数因子。如果 τ 至少有一条边，则该结果可由定理 2.8.2 推出；但即使 τ 没有边，该结果同样适用，此时我们可以直接使用简单得多的引理 2.4.10。上述刻画的精确形式见推论 2.8.3。

定理 2.8.2 给出了稀疏图矩阵情形下范数界关于 p, n 的正确依赖关系。通过迹幂方法, [119]也得到了相同的界（除低阶项外）；他们利用这些界证明了稀疏图上最大独立集问题的半定规划下界。

定理 2.8.2 的证明。 如果 $E(\tau) = \emptyset$，那么 $\mathbf{M}_\tau = \mathbb{E}\,\mathbf{M}_\tau$，结论即得。因此，假设$E(\tau) \neq \emptyset$。由于 I_τ 中的顶点只会将矩阵缩放至多 n倍，我们可以单独处理它们，而我们的界也会得到由这些顶点产生的适当的 n 幂。因此，可以假设 $I_\tau = \emptyset$。延续我们之前的讨论，对于某个绝对常数$C_1 > 0$,

$$\begin{aligned}
\mathbb{E}\,\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\|_{2t}^{2t} &\leq \sum_{a,b\geq 0, a+b=|E(\tau)|} (C_1 t^2 |E(\tau)|^4)^{t|E(\tau)|}\, \mathbb{E}\left\|\mathbf{D}\mathbf{M}_{\tau,a,b}\mathbf{D}\right\|_{2t}^{2t} \\
&\leq \sum_{a,b\geq 0, a+b=|E(\tau)|} (C_1 t^2 |E(\tau)|^4)^{t|E(\tau)|} (8|E(\tau)|)^{t|E(\tau)|} \sum_{\psi\in\Gamma_{a,b}} \mathbb{E}\left\|\mathbf{M}_\psi\right\|_{2t}^{2t}
\end{aligned}$$

其中 $\Gamma_{a,\,b}$ 是根据上述讨论，我们为 $\mathbf{D}\mathbf{M}_{\tau,\,a,\,b}\mathbf{D}$得到的简单形状集合。

利用引理 2.8.1，对于某个绝对常数 $C_2 > 0$，我们有

$$\mathbb{E}\,\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\|_{2t}^{2t} \le \left(n^{|V(\tau)|}|V(\tau)|^{t|V(\tau)|}(C_2 t^3|E(\tau)|^5)^{t|E(\tau)|}\right) \cdot \sum_{a,b\ge 0, a+b=|E(\tau)|} \sum_{\psi\in\Gamma_{a,b}} \max_{U_\psi\cap V_\psi\subseteq S\subseteq V(\psi)} \left(\frac{1-p}{p}\right)^{t|E(S)|} n^{t(|V(\psi)|-|S|)}$$

对于任意 a，b，考虑任意可得到的简单形状 $\psi \in \Gamma_{a,\,b}$。正如定理 2.4.9 的证明中所观察到的（特别参见图 2.5），$U_\psi \cap V_\psi$ 必定是τ的顶点分隔集。因此，任意 $S \supseteq U_\psi \cap V_\psi$都必定是 τ的顶点分隔集。容易看出，当 S遍历所有满足 $U_\psi \cap V_\psi \subseteq S \subseteq V(\psi)$ 的集合时，它也遍历了 τ的所有顶点分隔集。

此外，不同 ψ 的数量至多为 $4^{|E}(\tau)|$，因为每条边要么连到 U_ψ，要么连到V_ψ，并且对于每种这样的选择，它要么在 γ 中处于激活状态，要么不处于激活状态。因此，

$$\begin{aligned}
&\mathbb{E}\,\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\|_{2t}^{2t} \\
&\le \left(n^{|V(\tau)|}|V(\tau)|^{t|V(\tau)|}(C_2 t^3|E(\tau)|^5)^{t|E(\tau)|}\right) 4^{|E(\tau)|} \max_{\text{vertex separator } S} \left(\frac{1-p}{p}\right)^{t|E(S)|} n^{t(|V(\tau)|-|S|)} \\
&\le \left(n^{|V(\tau)|}|V(\tau)|^{t|V(\tau)|}(Ct^3|E(\tau)|^5)^{t|E(\tau)|}\right) \max_{\text{vertex separator } S} \left(\frac{1-p}{p}\right)^{t|E(S)|} n^{t(|V(\tau)|-|S|)}
\end{aligned}$$

其中 $C > 0$为某个绝对常数。 ■

以下推论通过马尔可夫不等式，得出图矩阵范数的高概率范数界。我们假设该图至少有一条边；否则它是确定性的，其范数界已在引理 2.4.10 和推论 2.4.11 中分析过。在那里我们观察到，如果随机矩阵是确定性的，那么稀疏图矩阵与稠密图矩阵之间的区别并不重要。

推论 2.8.3。对于至少有一条边的形状 τ，对于任意常数 $\varepsilon > 0$，以概率 $1 - \varepsilon$，

$$\|\mathbf{M}_\tau\| \le \left(|V(\tau)|^{|V(\tau)|/2}(C|E(\tau)|^5 \log^3(n^{|V(\tau)|}/\varepsilon))^{|E(\tau)|/2}\right) \cdot \max_{\textit{vertex separator } S} \left(\sqrt{\frac{1-p}{p}}\right)^{|E(S)|} \sqrt{n}^{|V(\tau)-|S|+|I_\tau|}$$

其中 $C > 0$ 为绝对常数。

证明。 由于 $|E(\tau)| \geq 1, \mathbb{E}\,\mathbf{M}_\tau = 0$。应用马尔可夫不等式,

$$
\begin{aligned}
&Pr[\|\mathbf{M}_\tau\| \geq \theta] \\
&\quad \leq Pr[\|\mathbf{M}_\tau\|_{2t}^{2t} \geq \theta^{2t}] \\
&\quad \leq \theta^{-2t}\,\mathbb{E}\,\|\mathbf{M}_\tau\|_{2t}^{2t} \\
&\quad \leq \theta^{-2t}\left(n^{|V(\tau)|}|V(\tau)|^{t|V(\tau)|}(C't^3|E(\tau)|^5)^{t|E(\tau)|}\right) \max_{\text{vertex separator } S}\left(\frac{1-p}{p}\right)^{t|E(S)|} n^{t(|V(\tau)|-|S|+|I_\tau|)}
\end{aligned}
$$

其中 $C' > 0$ 是一个绝对常数。现在我们设

$$
\begin{aligned}
\theta = &\left(\varepsilon^{-1/(2t)}(C'')^{|E(\tau)|}n^{|V(\tau)|/(2t)}|V(\tau)|^{|V(\tau)|/2}t^{3|E(\tau)|/2}|E(\tau)|^{5|E(\tau)|/2}\right) \\
&\cdot \max_{\text{vertex separator } S}\left(\sqrt{\frac{1-p}{p}}\right)^{|E(S)|}\sqrt{n}^{|V(\tau)-|S|+|I_\tau|}
\end{aligned}
$$

其中 $C'' > 0$ 是一个绝对常数，使该表达式至多为 ε。令 $t = \frac{1}{2}\log(n^{|V(\tau)|}/\varepsilon)$，即可完成证明。

■

2.8.1 简单图矩阵的范数界

在本节中，我们将证明引理 2.8.1。首先，我们回顾来自 [212] 的以下标量集中结果。

Schudy-Sviridenko 矩界

本节中的定义和主要界均来自 [212]。

定义 2.8.4. 若对于任意整数 $i \geq 1$，随机变量 Z 都是具有实参数 $L > 0$ 的中心矩有界随机变量，

$$
\mathbb{E}[|Z - \mathbb{E}[Z]|^i] \leq i \cdot L \cdot \mathbb{E}[|Z - \mathbb{E}[Z]|^{i-1}]
$$

命题 2.8.5. p 偏置的 *Bernoulli* 随机变量 Z 是具有实参数 $L = \sqrt{\frac{1-p}{p}}$ 的中心矩有界随机变量。

证明。 我们有 $\mathbb{E}[Z] = 0$，并且对于 $p \leq \frac{1}{2}$、$|Z| \leq \sqrt{\frac{1-p}{p}}$，因此，

$$\begin{aligned}\mathbb{E}[|Z - \mathbb{E}[Z]|^i] &= p\sqrt{\frac{p}{1-p}}^i + (1-p)\sqrt{\frac{1-p}{p}}^i \\ &\leq \sqrt{\frac{1-p}{p}}\left(p\sqrt{\frac{p}{1-p}}^{i-1} + (1-p)\sqrt{\frac{1-p}{p}}^{i-1}\right) \\ &= \sqrt{\frac{1-p}{p}}\,\mathbb{E}[|Z - \mathbb{E}[Z]|^{i-1}]\end{aligned}$$

因此，我们可以取 $L = \sqrt{\frac{1-p}{p}}$ ■

对于变量 x_1、……、x_n 上给定的多线性多项式 $f(x)$，我们可以自然地将其与顶点为 $[n]$ 、带权超边为 $E(H)$ 的超图 H 关联起来，其中每个$h \in E(H)$对应于 $f(x)$ 的一个不同项。每条超边 h 都是顶点 $V(h)$ 的一个子集，并具有实值权重 w_h ，该权重就是该单项式在 f 中的系数。因此，

$$f(x) = \sum_{h \in E(H)} w_h \prod_{v \in V(h)} x_v$$

A假设 f 的度为 d_p，那么 H 的每条超边至多有 d_p 个顶点 .

现在，对于给定的一组独立随机变量 Y_1、……、Y_n、一个具有相关超图 H 和权重 w 的多线性多项式 f，以及一个整数 $r \geq 0$，定义

$$\mu_r(f, Y) = \max_{S \subseteq [n], |S| = r} \left(\sum_{h \in E(H), S \subseteq V(h)} |w_h| \prod_{v \in V(h) \setminus S} \mathbb{E}[|Y_v|] \right)$$

引理 2.8.6 ([212], 引理 5.1)。 给定 n 个相互独立、中心矩有界且具有相同参数 $L > 0$ 的随机变量 $Y_1, \ldots, Y_n$，以及一个度为 d_p 的多线性多项式

$f(x)$。令 $t \geq 2$ 为偶整数，则

$$\mathbb{E}[|f(Y) - \mathbb{E}[f(Y)]|^t] \leq \max\left\{\left(\sqrt{tR_4^{d_p}\mathrm{Var}[f(Y)]}\right)^t, \max_{r\in[d_p]}(t^r R_4^{d_p} L^r \mu_r(f,Y))^t\right\}$$

其中 $R_4 \geq 1$ 是某个绝对常数。

在我们的设定中，我们还可以像 [212], 中所示的那样，用 μ_r 来界定方差，这将简化我们的计算。

引理 2.8.7 ([212],引理 1.5)。 对于与引理 *2.8.6* 相同的设定，

$$\mathrm{Var}[f(Y)] \leq 2d_p 4^{d_p} \max_{r\in[d_p]}(\mu_0(f,Y)\mu_r(f,Y)4^r L^r)$$

引理 2.8.1 的证明

我们现在准备证明引理 2.8.1，为方便起见，我们先重述该引理。

引理 2.8.1. 对于所有偶整数 $t \geq 2$，若 τ 是简单形状，

$$\mathbb{E}\,\|\mathbf{M}_\tau\|_{2t}^{2t} \leq \left(n^{|V(\tau)|}(Ct)^{t|E(\tau)|}|V(\tau)|^{t|V(\tau)|}\right)\max_{U_\tau\cap V_\tau\subseteq S\subseteq V(\tau)}\left(\frac{1-p}{p}\right)^{t|E(S)|} n^{t(|V(\tau)|-|S|)}$$

其中 $C > 0$ 为绝对常数。

我们将用与引理 2.4.10 相同的方法证明这一点：将每个对角块的 Schatten 范数界定为其 Frobenius 范数的适当幂。在这种情形下，为了界定 Frobenius 范数的期望幂，我们使用上一节中的标量集中不等式。

引理 *2.8.1* 的证明。 首先，我们注意到 $\mathbf{M}_\tau$ 具有一种分块对角结构，其索引由公共顶点集合 $S_0 = U_{\tau_P} \cap V_{\tau_P}$ 的实现给出。对于 $T \in [n]^{S_0}$，令 $\mathbf{M}_{\tau,T}$ 为

$\mathbf{M}_\tau$中对应于 $\varphi(S_0) = T$ 的块。于是，$\mathbf{M}_{\tau,T}\mathbf{M}_{\tau,T'}$ 与 $\mathbf{M}_{\tau,T}\mathbf{M}_{\tau,T'}$ 对于 $T \neq T'$ 也如此，

$$\mathbb{E}\left\|\mathbf{M}_\tau\right\|_{2t}^{2t} = \sum_{T\in[n]^{S_0}} \mathbb{E}\left\|\mathbf{M}_{\tau,T}\right\|_{2t}^{2t} \leq \sum_{T\in[n]^{S_0}} \mathbb{E}(\left\|\mathbf{M}_{\tau,T}\right\|_2^2)^t$$

其中，我们用 Frobenius 范数的幂对 Schatten 范数进行了界定。

固定 $T \in [n]^{S_0}$，并考虑 $\mathbb{E}\left\|\mathbf{M}_{\tau,T}\right\|_2^2$。令 $\mathcal{R}$ 为满足 $\varphi(S_0) = T$ 的 τ 的实现 φ 的集合。于是，对于 $\varphi \in \mathcal{R}$ 和 $e \in E(S_0)$，$\varphi(e)$ 的值是固定的。利用这一点，

$$\begin{aligned}
\left\|\mathbf{M}_{\tau,T}\right\|_2^2 &= \sum_{\varphi\in\mathcal{R}} \prod_{e\in E(\tau)} G_{\varphi(e)}^2 \\
&= \prod_{e\in E(S_0)} G_{\varphi(e)}^2 \sum_{\varphi\in\mathcal{R}} \prod_{e\in E(\tau)\setminus E(S_0)} G_{\varphi(e)}^2 \\
&\leq L^{|E(S_0)|} \sum_{\varphi\in\mathcal{R}} \prod_{e\in E(\tau)\setminus E(S_0)} G_{\varphi(e)}^2
\end{aligned}$$

其中，$L = \frac{1-p}{p}$ 是 $p \leq \frac{1}{2}$ 上 G_{ij}^2 的一个上界。定义量

$$A = \max_{S_0\subseteq S\subseteq V(\tau)} L^{|E(S)|} n^{|V(\tau)|-|S|}$$

命题 2.8.8。$\mathbb{E}\left(\left\|\mathbf{M}_{\tau,T}\right\|_2\right)^t \leq (Ct)^{t|E(\tau)|}|V(\tau)|^{t|V(\tau)|}A^t$，其中 $C > 0$ 为绝对常数。

利用这一命题，我们有

$$\begin{aligned}
\mathbb{E}\left\|\mathbf{M}_\tau\right\|_{2t}^{2t} &\leq \sum_{T\in[n]^{S_0}} \mathbb{E}(\left\|\mathbf{M}_{\tau,T}\right\|_2)^t \\
&\leq n^{|S_0|}(Ct)^{t|E(\tau)|}|V(\tau)|^{t|V(\tau)|}A^t \\
&= n^{|V(\tau)|}(Ct)^{t|E(\tau)|}|V(\tau)|^{t|V(\tau)|} \max_{U_\tau\cap V_\tau\subseteq S\subseteq V(\tau)} \left(\frac{1-p}{p}\right)^{t|E(S)|} n^{t(|V(\tau)|-|S|)}
\end{aligned}$$

如所需。 ∎

剩下的就是证明这个断言。

命题 *2.8.8* 的证明。对于 $1 \le i,\ j \le n$，定义变量 $Y_{ij} = G_{ij}^2$，使其满足 $\mathbb{E}[|Y_{ij}|] = 1$。令 $f(Y)$ 为多项式 $L^{|E_{(S_0)}|}\sum_{\varphi\in\mathcal{R}}\prod_{e\in E(\tau)\setminus E(S_0)} Y_{\varphi(e)}$。只需证明 $\mathbb{E}[f(Y)^t] \le (Ct)^{t|E_1|}A^t$。

我们首先证明，对于足够大的常数 $C' > 0$，$\mathbb{E}[(f(Y)-\mathbb{E}[f(Y)])^t] \le (C't)^{t|E_{(\tau)}|}|V(\tau)|^{t|V_{(\tau)}|}A^t$。

f 是一个次数为 $|E(\tau) \setminus E(S_0)|$的齐次多重线性多项式。若有 $E(\tau) \setminus E(S_0) = \emptyset$，则 f是一个常数，因此不等式显然成立，因为 $f(Y) = \mathbb{E}[f(Y)]$。现在，假设 $E(\tau) \setminus E(S_0) \ne \emptyset$。我们应用引理 2.8.6。设 f对应的超图为H，权重为 w。那么，

$$\mathbb{E}[|f(Y) - \mathbb{E}[f(Y)]|^t] \le \max\left\{\left(\sqrt{tR_4^{|E(\tau)\setminus E(S_0)|}\mathrm{Var}[f(Y)]}\right)^t, \max_{r\in[|E(\tau)\setminus E(S_0)|]}(t^r R_4^{|E(\tau)\setminus E(S_0)|}L^r\mu_r(f,Y))^t\right\}$$

对于所有 $r \ge 0$，我们将证明 $L^r\mu_r(f,\ Y) \le |V(\tau)|^{|V_{(\tau)}|}A$。根据定义，

$$\mu_r(f,Y) = \max_{F\subseteq\binom{[n]}{2},|F|=r}\sum_{h\in E(H),F\subseteq V(h)}|w_h|$$

考虑任意一组边标签 $F \subseteq \binom{[n]}{2}$，$|F| = r$。那么，$\sum_{h\in E(H)\ F\subseteq V(h)}|w_h|$至多为$L^{|E(S_0)|}c$，其中 c是满足 $\varphi(E(\tau))$ 包含 F的实现 $\varphi \in \mathcal{R}$的数量。假设F除了 $\varphi(S_0) = T$之外，还包含 v个新标签。那么 $c \le |V(\tau)|^v n^{|V_{(\tau)}|-|S_0|-v}$，因为我们可以先选择并标记获得这些 v 标签的顶点集合，然后自由地标记其余顶点，而每个顶点至多有 n 种选择。

注意到 $L^{|E_{(S_0)}|}L^r n^{|V_{(\tau)}|-|S_0|-v} \le A$，因为在 S 的定义中，我们可以将 S 设为 S 与这些 v 个顶点的任意有效选取的并集。综合以上各点，我们得到

$$\begin{aligned}L^r\mu_r(f,Y) &\le L^r\max_{F\subseteq\binom{[n]}{2},|F|=r}\sum_{h\in E(H),F\subseteq V(h)}|w_h|\\ &\le |V(\tau)|^{|V(\tau)|}A\end{aligned}$$

这意味着

$$\max_{r\in[|E(\tau)\setminus E(S_0)|]}(t^r R_4^{|E(\tau)\setminus E(S_0)|}L^r\mu_r(f,Y))^t \le |V(\tau)|^{t|V(\tau)|}(R_4 t)^{t|E(\tau)|}A^t$$

并使用引理 2.8.7,

$$\begin{aligned}\mathrm{Var}[f(Y)] &\le 2|E(\tau)|4^{|E(\tau)|}\max_{r\in[|E(\tau)\setminus E(S_0)|]}(\mu_0(f,Y)\mu_r(f,Y)4^r L^r)\\ &\le 2|E(\tau)|16^{|E(\tau)|}|V(\tau)|^{2|V(\tau)|}A^2\end{aligned}$$

将它们合并，得到

$$\begin{aligned}\mathbb{E}[(f(Y)-\mathbb{E}[f(Y)])^t] &\le \max\left\{\left(\sqrt{2tR_4^{|E(\tau)|}|E(\tau)|16^{|E(\tau)|}|V(\tau)|^{2|V(\tau)|}A^2}\right)^t, |V(\tau)|^{t|V(\tau)|}(R_4 t)^{t|E(\tau)|}A^t\right\}\\ &\le (C't)^{t|E(\tau)|}|V(\tau)|^{t|V(\tau)|}A^t\end{aligned}$$

其中 $C'>0$ 为绝对常数。

最后，$\mathbb{E}[f(Y)] \le L^{|E(S_0)|}|\mathcal{R}| \le L^{|E(S_0)|}n^{|V(\tau)\setminus S_0|} \le A$由此可得

$$\begin{aligned}\mathbb{E}[f(Y)^t] &\le 2^t(\mathbb{E}[(f(Y)-\mathbb{E}[f(Y)])^t]+\mathbb{E}[f(Y)]^t)\\ &\le 2^t((C't)^{t|E(\tau)|}|V(\tau)|^{t|V(\tau)|}A^t + A^t)\\ &\le (Ct)^{t|E(\tau)|}|V(\tau)|^{t|V(\tau)|}A^t\end{aligned}$$

对于某个绝对常数 $C>0$。 ■

第3章

平方和层级

本章将正式介绍平方和（SoS）层级。随后，我们稍作岔开，定义用于假设检验的低次区分器及相关概念，为讨论 SoS 下界做好铺垫。接着，我们回到 SoS，介绍一种称为伪校准的启发式方法，它将是我们推导 SoS 下界时使用的基本工具。最后，我们将展示伪校准与低次区分器之间的形式化联系，并简要介绍 SoS 近期取得的成功。

3.1 平方和层级

我们首先定义多项式优化问题的凸松弛。随后，SoS 层级将成为一类特殊的凸松弛。有关更详细的介绍，请参见例如 [149, 25, 80]。

3.1.1 多项式优化与凸松弛

在多项式优化中，给定取实值的多元多项式 p，g_1，……，g_m，它们定义在 n 个变量 x_1，……，x_n 上，统称为 x，我们的任务是：

$$\text{maximize } p(x) \text{ such that } g_1(x) = 0, \ldots, g_m(x) = 0$$

一般来说，我们也可以允许不等式约束，例如 $g_i(x) \geq 0$。为便于技术处理，在我们的设定中，我们只考虑等式约束；不过，只要稍作修改，许多理论也可以推广到存在不等式约束的情形。另一种方法是将每个不等式 $g_i(x) \geq 0$ 替换为 $g_i(x) = y^2$，其中 y 是我们可以引入的一个新变量。

在这种表述下，许多优化问题都可以表述为多项式优化问题。

例 3.1.1（最大割）。 给定一个图 $G = (V, E)$，我们希望将顶点集合分拆为两个子集，使得端点分别位于不同子集中的边数达到最大。为将其表述为多项式优化问题，设该图有 n 个顶点，并令 $x_1, \ldots, x_n$ 为变量，每个顶点对应一个变量。我们希望满足 $x_i \in \{-1, 1\}$，其中所有满足 $x_i = -1$ 的顶点 i 构成一个子集，其余顶点构成另一个子集。我们可以通过多项式约束 $x_i^2 = 1$ 来实现这一集合包含约束。对于任意一条边 $(i, j) \in E$，当且仅当 $x_i x_j = -1$ 时，该边被割开。因此，被割开的边总数为 $\sum_{(i,j)\in E} \frac{1}{2}(1 - x_i x_j)$。因此，多项式表述为

$$\max_{x\in\mathbb{R}^n} \sum_{(i,j)\in E} \frac{1}{2}(1 - x_i x_j) \text{ such that}$$

$$x_i^2 = 1 \text{ for all } i \leq n$$

示例 3.1.2（最大团）。 给定一个图 $G = (V, E)$，我们希望找到一个由构成团的顶点组成的最大子集。再次令 $x_1, \ldots, x_n$ 为变量，每个顶点对应一个变量。这一次，我们希望施加 $x_i \in \{0, 1\}$，为此可以轻松地使用多项式约束 $x_i^2 = x_i$，其意图是让所有满足 $x_i = 1$ 的顶点 i 构成一个团。为施加这个团约束，我们可以针对所有非边 $(i, j) \notin E$ 添加多项式约束 $x_i x_j = 0$。最后，为了使子集大小最大化，我们只需最大化 $\sum_{i\leq n} x_i$。因此，该多项式优化问题为

$$\max_{x\in\mathbb{R}^n} \sum_{i\leq n} x_i \text{ such that}$$

$$x_i x_j = 0 \text{ for all } (i, j) \notin E$$

$$x_i^2 = x_i \text{ for all } i \leq n$$

这些问题也可以有其他等价的表述。一般而言，许多优化问题都可以用这种方式表述，因此，通用多项式优化涵盖了计算机科学中出现的一大类基本问题。

由于精确求解最大割或最大团是 NP难的， [121], 精确求解这些多项式优化问题同样是 NP难的。因此，我们转而考虑凸松弛。

多项式优化问题的凸松弛会扩大解向量 x 的搜索空间，将其扩展到一个更大的空间，使我们能够在其中高效地进行优化。我们将介绍一种实现方法。我们找出一个包含空间 $\mathcal{S} = \{g_1(x) = 0, \ldots, g_m(x) = 0\}$ 的凸空间 $\mathcal{C}$ ，二者之间相差一个映射；也就是说，对每个 $x \in \mathcal{S}$，都存在一个对应的 $y \in \mathcal{C}$，使得 y 是 x的代表。我们还找出一个凸函数 $\tilde{p}(y)$，使得若 y是 x的代表，则 $\tilde{p}(y) = p(x)$。然后，我们只需在 $\mathcal{C}$上优化 $\tilde{p}(y)$。已有大量研究探讨如何在凸体上高效优化凸函数；在合理的假设下，这是可行的（例如，参见 [181]）。显然，根据上述性质，我们得到的解至少不小于最优解（在最大化情形下），同时还具有可高效计算的优点。我们希望为问题设计能够产生良好近似的凸松弛。SoS 层级就是这样一族凸松弛。

3.1.2 平方和松弛

SoS 层级有时也称为 Lasserre 层级，最早由 [185, 148, 216] 独立地进行研究，[173, 91, 92]也在其他背景下对其进行了研究。它是用于多项式优化的一族凸松弛，其参数由一个称为度的整数确定。随着度的增加，我们会得到逐渐更紧的松弛，但优化所需的时间也会更长。

现在，我们通过所谓的伪期望

算子视角，正式介绍平方和层级。

定义 3.1.3 (伪期望值)。 给定关于 n 个变量 $x_1, \ldots, x_n$ 的多元多项式约束 $g_1 = 0, \ldots, g_m = 0$，度为 d 的伪期望值是一个线性映射 $\widetilde{\mathbb{E}}$，其定义域为由 $x_1, \ldots, x_n$ 构成的次数至多为 d 的多项式，值域为 $\mathbb{R}$，并满足以下条件：

1. $\widetilde{\mathbb{E}}[1] = 1$,
2. $\widetilde{\mathbb{E}}[f \cdot g_i] = 0$ 对于每个 $i \in [m]$ 和满足 度$(f \cdot g_i) \leq d$ 的多项式 f。
3. $\widetilde{\mathbb{E}}[f^2] \geq 0$ 对于每个满足 度$(f^2) \leq d$ 的多项式 f。

任何满足上述性质的线性映射 $\widetilde{\mathbb{E}}$ 都称为满足约束 $g_1 = 0, \ldots, g_m = 0$ 的度为 d 的伪期望算子。

定义 3.1.4 (度 d 平方和)。 该多项式优化问题的度为 d 的平方和松弛

$$maximize\ p(x)\ such\ that\ g_1(x) = 0, \ldots, g_m(x) = 0$$

是一个程序，它在所有满足约束 $g_1 = 0, \ldots, g_m = 0$ 的度为 d 的伪期望算子 $\widetilde{\mathbb{E}}$ 中，最大化 $\widetilde{\mathbb{E}}[p(x)]$。

伪期望值背后的直觉是，伪期望值所满足的条件，恰好是解的某个分布上的实际期望值会满足的条件，因此，对伪期望值进行优化就得到了该问题的一个松弛。

主要的观察是，平方和松弛可以高效求解！这是因为伪期望值所满足的条件可以用半定规划来刻画。具体而言，定义 3.1.3 中的第 3 项可以用一个称为矩矩阵的矩阵来重新表述。

~ ~**定义 3.1.5 (** $\mathbb{E}$的矩矩阵。 给定一个度为 d 的伪期望算子 $\mathbb{E}$，定义其对应的矩矩阵 Λ 为一个矩阵，其行和列由单项式 p 和 q 索引，使得对应于行 p 和列 q 的矩阵项为

$$\Lambda[p,q] := \widetilde{\mathbb{E}}\,[pq]\,.$$

容易验证，定义 3.1.3 中的第 3 项等价于 $\Lambda \succeq 0$。因此，可以通过半定规划求解 d 度平方和松弛，例如参见 [228]。一般来说，对于 d 度平方和，我们可以在 $n^{O(d)}$ 时间内[1]求解。因此，常数次数平方和可以在多项式时间内求解。

分析用于最大团的 2 度平方和

为了说明这一技术的用法，我们来分析 Erd˝os-R´enyi 随机图$G_{n,\ 1/2}$上的最大团问题的二次平方和松弛。我们使用例 3.1.2 中的程序。

令 A 为从 $G_{n,1/2}$ 中抽样得到的图 G 的邻接矩阵，令 J 为元素全为 1 的矩阵。那么，根据随机矩阵理论，对于 G 的选择，以高概率有 $\lambda_{max}(A - J/2) = O(\sqrt{n})$，其中 $\lambda_{max}(.)$ 表示最大奇异值。现在，假设一组顶点 S 构成一个团，并令 $\mathbf{1}_S$ 表示集合 S 的指示向量，那么

$$\begin{aligned}\frac{k(k-1)}{2} &= \langle \mathbf{1}_S, (A - J/2)\mathbf{1}_S \rangle \\ &\leq \|\mathbf{1}_S\|^2 \cdot \lambda_{max}(A - J/2) \\ &\leq k \cdot O(\sqrt{n})\end{aligned}$$

1.由于比特复杂度方面的问题，这种说法并不完全准确 [176]，但对于大多数值得关注的问题，这种情况并不会发生 [198]

这表明 $k \leq O(\sqrt{n})$。

这一简单论证的关键在于，这是一个平方和能够捕捉的低次证明，更具体地说，是二次证明。也就是说，如果我们求解二次平方和松弛，那么我们将能够证明 $\mathbb{E}[\sum x_i] = O(\sqrt{n})$ 以高概率成立。

要形式化地说明这一点，我们从以下不等式开始： $O(\sqrt{n})I - (A - J/2) \succeq 0$以高概率成立。这意味着

$$x^\intercal (O(\sqrt{n})I - (A - J/2))x = \sum p_i(x)^2$$

它是次数至多为 1 的多项式的平方和。简单计算可得

$$x^\intercal (A - J/2)x = \frac{1}{2}(\sum_{i=1}^{n} x_i)^2 - \sum_{i,j} x_i x_j \mathbf{1}_{(i,j)\notin E(G)}$$

对于我们的程序变量 x，有 $x_i^2 = x_i$ 和 $x_i x_j \mathbf{1}_{(i,j)\notin E(G)} = 0$。因此，

$$\sum p_i(x)^2 = O(\sqrt{n})(\sum_{i=1}^{n} x_i) - \frac{1}{2}(\sum_{i=1}^{n} x_i)^2$$

对两边应用 $\widetilde{\mathbb{E}}$。最后，我们利用以下事实：对于多项式 $p(x)$，有 $\widetilde{\mathbb{E}}[p(x)^2] \geq \widetilde{\mathbb{E}}[p(x)]^2$，，这成立是因为它可重新排列为 $\widetilde{\mathbb{E}}[(p(x) - \widetilde{\mathbb{E}}[p(x)])^2] \geq 0$；这又成立是因为左侧是一个平方多项式的伪期望，根据定义它是非负的。这个简单的事实本质上说明伪方差非负。利用 $\mathbb{E}$ 的线性，最后得到

$$
\begin{aligned}
O(\sqrt{n})\widetilde{\mathbb{E}}[\sum_{i=1}^{n} x_i] - \frac{1}{2}(\widetilde{\mathbb{E}}[\sum_{i=1}^{n} x_i])^2 &\geq O(\sqrt{n})\widetilde{\mathbb{E}}[\sum_{i=1}^{n} x_i] - \frac{1}{2}\widetilde{\mathbb{E}}[(\sum_{i=1}^{n} x_i)^2] \\
&= \widetilde{\mathbb{E}}[O(\sqrt{n})(\sum_{i=1}^{n} x_i) - \frac{1}{2}(\sum_{i=1}^{n} x_i)^2] \\
&= \widetilde{\mathbb{E}}[\sum p_i(x)^2] \\
&= \sum \widetilde{\mathbb{E}}[p_i(x)^2] \\
&\geq 0
\end{aligned}
$$

因此，$\widetilde{\mathbb{E}}[\sum_{i=1}^{n} x_i] = O(\sqrt{n})$，这正是我们想要证明的。

这表明，对于 Erdős-Rényi 随机图，二次平方和松弛以高概率给出最大团大小的上界 $O(\sqrt{n})$。相比之下，真正的最大团大小为$(2 + o(1))\log\ n$ [162]。尽管付出了巨大努力，多项式时间算法只有在 $k = \Omega(\sqrt{n})$时才能检测出植入的 k-团。因此，对于这个问题，平方和已经达到了已知的最佳保证，误差仅相差常数因子。文献 [20] 证明，高阶平方和（最高到 $O(\log n)$度）未必能带来多大改进；这是一种我们将在本文中研究的平方和下界。

平方和的另一种视角

在多项式优化问题中，若要在约束$g_1(x) = 0, \ldots, g_m(x) = 0$下最大化 $p(x)$，如果不存在满足 $g_1 = 0, \ldots, g_m = 0$ 且使得 $\mathbb{E}[p] > c$成立的度为 d 的伪期望算子 $\widetilde{\mathbb{E}}$，那么我们就称度为 d 的平方和能够证明 $\mathbb{E}[p(x)] \leq c$。

度 d SoS 证明 $p(x) \leq c$ 给定 $g_1(x) = 0, \ldots, g_m(x) = 0$ 是一个表达式 of

形式

$$-1 = \sum_{i \le m} g_i(x) q_i(x) + \sum_{i \le a} s_i(x)^2 + (p(x) - c) \sum_{i \le b} t_i(x)^2$$

其中 q_1、……、q_m、s_1、……、s_a、t_1、……、t_b 是 x 中的多项式，使得上述表达式右侧的每一项的度至多为 d。事实上，这种表达式的存在自动意味着：只要 $g_1(x) = 0$、……、$g_m(x) = 0$，就有 $p(x) \le c$。

~当度为 d 的平方和证明证实 $\mathbb{E}[p(x)] \le c$ 时，根据对偶性，这意味着存在一个度为 d 的平方和证明，证明在给定 $g_1(x) = 0$……、$g_m(x) = 0$ 的情况下，$p(x) \le c$。Krivine 和 Stengle 的 Positivstellensatz [142, 221] 指出，对于任意 c，要么存在 x，使得 $p(x) > c$、$g_1(x) = 0$、……、$g_m(x) = 0$；要么存在一个平方和证明，证明在给定 $g_1(x) = 0$、……、$g_m(x) = 0$ 的情况下，$p(x) \le c$。

对于固定的 d，d 次 SoS 确实可以理解为寻找最优值 c，使得存在一个关于 $\mathbb{E}[p(x)] \le c$ 的 d 次 SoS 证明。这也直观地解释了为什么更高阶的 SoS 会带来更紧的松弛。对于大多数源自组合优化问题的程序，n 次 SoS 通常能找到最优界，其中 n 是变量的数量。因此，例如，n 次 SoS 能精确输出图的最大团的大小。为了实现高效算法，我们通常希望使用常数次数的 SoS。因此，对于平方和下界，次数越高，下界越强。在这项工作中，我们所有的下界都是针对 n^{ε} 次 SoS 的，这对应于亚指数时间！

我们在这里研究的视角是对偶视角，也就是寻找简单证明，这符合我们的目的。还有一种原始视角，可以直接将 SoS 视为该程序的半定规划松弛。这有时对算法设计很有用。

与上文展示的最大团应用类似，已有正式研究表明，在最坏情况和平均情况设置下，SoS 层级都能为许多基础问题取得当前最先进的近似保证。这些问题包括约束满足问题、[195]，最大割 [89]，最稀疏割、[10]，张量主成分分析 [107]，等。

因此，研究 SoS 的局限性自然要从研究 SoS 下界入手。

在讨论 SoS 下界之前，我们先更详细地介绍适用于我们目的的假设检验问题框架。

3.2 假设检验

设 Ω 为一个样本空间。设 ν, μ 是 Ω^n 上的概率分布。假设检验问题是指在给定一个样本的情况下，区分 ν, μ 的问题。形式上，输入 $x \sim \Omega^n$ 从以下二者之一中抽样得到

- H_0: $x \sim \mu$
- H_1: $x \sim \nu$。

我们的目标是以高概率确定它来自哪个分布。这就是一般意义上的假设检验问题；传统上，H_0 称为原假设，H_1 称为备择假设。我们滥用记号，将 H_0, H_1 也分别用来表示概率分布 μ, ν。

例如，H_0 可以是 Erd˝os-R´enyi 随机图的分布，而 H_1 可以是含有大型植入团的 Erd˝os-R´enyi 随机图的分布。给定该图，我们希望确定它来自这两种分布中的哪一种，换句话说，就是判断它是否含有一个大型团。

假设检验 f 是一个函数 $f: \Omega^n \to \{0,1\}$。给定输入 x，如果 $f(x) = 0$，那么我们报告说 x 来自零假设分布 H_0；否则，我们报告说 x 来自备择分布 H_1。

成功的假设检验是指满足以下条件的检验 f：当 b 从 $\{0,1\}$ 中均匀随机选取，且 x 从 H_b中采样时，我们有 $\mathbb{E}_b \Pr_{x \sim H_b}[f(x) \neq b] \leq o(1)$。也就是说，检验 f的成功概率为 $1 - o(1)$。这里，为简单起见，我们不区分类 1 错误和类型 2 错误。

实际上，要使检验有用，它应当能够高效地计算。在不考虑计算效率时，著名的奈曼-皮尔逊引理精确刻画了最佳假设检验。要定义这一检验，我们需要以下标准定义。

定义 3.2.1（似然比）。 对于给定的假设检验问题，定义输入 x 的似然比为 $LR(x) = \Pr^{H_1}(x)\ \Pr^{H_0}(x)$。

引理 3.2.2（奈曼-皮尔逊引理）。 对于给定的假设检验问题，使 $\mathbb{E}_b \Pr_{x\sim H_b}[f(x) \neq b]$ 最小的检验f是似然比检验

$$f(x) = \begin{cases} 1 & if\ LR(x) > 1 \\ 0 & o.w. \end{cases}$$

在本文中，我们将重点关注可高效计算的检验 f。

3.2.1 低次似然比

考虑一个给定的假设检验问题。我们关注一类特殊的、可高效计算的假设检验，这类检验涉及低次多元多项式，称为低次区分器。本节将对此作简要介绍，并请读者参阅 [110, 144] 以了解更详细的论述。

本节中，为了使多项式定义良好，假设 $\Omega \subseteq \mathbb{R}$。此外，假设 H_0具有有限阶矩。我们将考虑由多元多项式$f : \mathbb{R}^n \to \mathbb{R}$导出的区分器。若 f 的度至多为 D，则称该区分器的度为 D。由于多项式的输出不一定是布尔值，我们需要另行定义该区分器的成功。我们采用 [110]中的以下定义。

定义 3.2.3 (度 D 区分器**)**。 对于一个假设检验问题，多变量多项式 f 是一个成功的度 D 区分器，当且仅当

- *(*低度*)* f 是一个次数至多为 D 的多元多项式。

- *(归一化)* $\mathbb{E}_{x\sim H_0}[f(x)] = 0, \mathbb{E}_{x\sim H_0}[f(x)^2] = 1$

- （可区分性）$\lim_{n\to\infty} \mathbb{E}_{x\sim H_1}[f(x)] \to \infty$。

归一化可确保多项式按适当的尺度缩放。注意，归一化是相对于零假设分布进行的。非正式地说，如果归一化后的 f 在备择分布上随着极限推进而取得无界值，它就是一个成功的区分器。事实上，在应用中，可以通过对多项式的值进行适当的阈值化来构造假设检验。

对度的限制体现了我们希望施加于区分算法的计算限制。研究这类低次区分器在假设检验问题中的能力，是一个活跃研究领域。例如，我们可以提出这样的问题：对于输入规模为 $n^{O(1)}$的假设检验问题，如果度为 $O(\log n)$ 的区分器都无法奏效，那么该问题对所有多项式时间算法来说都很难吗?

首先自然会问：对于给定的假设检验问题，最佳的度为 D 的区分器是什么？先前工作已经回答了这个问题：它就是将似然比 $LR(x) = \Pr^{H_1(x)} \Pr^{H_0(x)}$ 投影到度为 D 的多项式上。

为使其精确，对于 $f, g\colon \mathbb{R}^n \to \mathbb{R}$，定义内积 $\langle f, g\rangle = \mathbb{E}_{x\sim H_0} f(x)g(x)$。然后，我们可以通过此内积规范地定义函数 f 到次数不超过 D 的多项式的投影 $f^{\le D}$ 。取一组次数至多为 D 的多元多项式的正交归一基 $\chi_0 = 1, \chi_1, \dots, \chi_t$，其中 $\chi_0 = 1$ 为常函数。于是，$f^{\le D}(x) = \sum_{i\le t}\langle f, \chi_t\rangle\chi_t(x)$。

以下引理隐含于先前工作中（例如 [109, 102]）。为完整起见，我们给出证明。

引理 3.2.4. 对于一个假设检验问题，使 $\mathbb{E}_{x\sim H_1} f(x)$ 最大化的最优度数 D 检验 f 是归一化的低次似然比 $\frac{LR^{\le D}-1}{\|LR^{\le D}-1\|}$。此外，其值为 $\mathbb{E}_{x\sim H_1}[f(x)] = \left\|LR^{\le D} - 1\right\|$。

证明。设 f 是一个归一化的度数为 D 的多项式，且满足 $f = \sum_{i=0}^{t} c_t \chi_t$。于是，$c_0 = \mathbb{E}[f] = 0$且 $\sum c_i^2 = \mathbb{E}[f^2] = 1$。因此，

$$\mathop{\mathbb{E}}_{x\sim H_1} f(x) = \sum_{1\le i\le t} c_i \mathop{\mathbb{E}}_{x\sim H_1} \chi_i \le \sqrt{(\sum_{1\le i\le t} c_i^2)(\sum_{1\le i\le t} (\mathop{\mathbb{E}}_{x\sim H_1} \chi_i)^2)} = \sqrt{\sum_{1\le i\le t} (\mathop{\mathbb{E}}_{x\sim H_1} \chi_i)^2}$$

另一方面，多项式 $g = \frac{LR^{\le D}-1}{\|LR^{\le D}-1\|}$ 可达到等式。事实上，我们有 $\mathbb{E}_{x\sim H_0}[g] = 0$ ，因为 $\mathbb{E}_{x\sim H_0}[LR^{\le D}(x)] = \mathbb{E}_{x\sim H_0}[LR(x)] = 1$ ；显然，由于我们按范数进行了缩放，有 $\mathbb{E}_{x\sim H_0}[g(x)^2] = 1$ 。最后，

$$\mathop{\mathbb{E}}_{x\sim H_1} g(x) = \frac{1}{\left\|LR^{\le D} - 1\right\|} \sum_{1\le i\le t} \langle LR(x), \chi_i\rangle^2$$

通过注意到 $\langle LR(x)\ \chi_i\rangle = \mathbb{E}_{x\sim H_0}[LR(x)\chi_i(x)] = \mathbb{E}_{x\sim H_1}[\chi_i(x)]$，我们完成证明。计算该值是直接的。 ■

低次似然比假设 [104, 110, 146] 假设：如果 H_0，H_1 是足够良好 的分布，那么当且仅当存在一个成功的度 D 区分器时，才存在一个运行时间为 $n^{O(D)}$ 的成功假设检验。特别地，根据上述讨论，如果 $\left\|LR^{\le D} - 1\right\| = O(1)$，那么我们预计不存在运行时间为 $n^{O(D)}$ 的成功假设检验。

这项工作的主要贡献之一，是通过展示强 SoS 下界，为这一猜想在许多基本问题中成立提供有力证据。为了更形式化地理解这种联系，我们将介绍伪校准，并将其与低次区分器联系起来。

3.3 伪校准

考虑一个我们试图证明其平方和下界的优化问题。为了在随机实例上得到平方和整数间隙，我们需要为该问题的随机输入实例构造有效的伪期望

值。当然，这些伪期望值会依赖于输入。

伪校准是一种由 [20] 引入的启发式方法：通过考虑一个由目标值较大的问题实例支撑的植入分布，并利用该植入分布作为构造伪期望值的指导，几乎机械地构造出这样的候选伪期望值。需要注意的是，出于历史原因，我们使用“随机分布”这一术语，而不使用“零假设分布”；使用“植入分布”这一术语，而不使用“备择分布”。

遗憾的是，伪校准并不能保证这些候选伪期望值及其对应的矩矩阵具有可行性，而这必须针对不同问题分别进行验证。除半正定性条件外，这种可行性验证相对容易。本文的主要贡献正是在于此：我们分析了所构造的随机矩矩阵的行为。

事实上，对于我们的应用，伪校准用于从随机与植入问题中获得候选伪期望算子 $\mathbb{E}$ 以及相应的矩矩阵 Λ 。这将作为我们所有应用的起点。伪校准为许多问题给出了下界，例如文献 [92, 210, 137, 49, 165],中所研究的问题，这使其成为一种颇具吸引力但尚未得到充分理解的技术。

这里，我们不打算详细阐述伪校准的动机和原理。相反，我们将简要介绍这一启发式方法及其背后的直觉，并展示一个使用示例。详细论述见 [20]。

令 ν 表示随机分布， μ 表示植入分布。令 v 表示输入， x 表示我们平方和松弛中的变量。主要思想是，对于从 ν 中采样得到的输入 v ，以及任意一个次数至多为平方和次数的多项式 $f(x)$，伪-~校准提出，对于任意低次测试 $g(v)$， $\mathbb{E}[f]$ 的相关性应当与之匹配

在植入分布和随机分布中。也就是说，

$$\mathop{\mathbb{E}}_{v\sim\nu}[\widetilde{\mathbb{E}}[f(x)]g(v)] = \mathop{\mathbb{E}}_{(x,v)\sim\mu}[f(x)g(v)]$$

这里，记号 $(x,v) \sim \mu$ 表示在植入分布 μ 中，输入是 v，而 x 表示该实例中的植入结构。例如，在植入团问题中，x 是该团的指示向量。如果存在多个，则任选一个。

令 $\mathcal{F}$ 表示输入 v 的多项式傅里叶基。通过从 $\mathcal{F}$ 中选择不同的基函数作为 g 的选项，使其度至多为某个截断参数 D，我们就能得到 $\mathbb{E}[f(x)]$ 关于 v 的所有低阶傅里叶系数。此外，将高阶系数设为 0，这样候选伪期望算子就可以写成

$$\widetilde{\mathbb{E}}f(x) = \sum_{\substack{g\in\mathcal{F}\\ deg(g)\le n^{\varepsilon}}} \mathop{\mathbb{E}}_{v\sim\nu}[\widetilde{\mathbb{E}}[f(x)]g(v)]g(v) = \sum_{\substack{g\in\mathcal{F}\\ deg(g)\le n^{\varepsilon}}} \mathop{\mathbb{E}}_{(x,v)\sim\mu}[[f(x)]g(v)]g(v)$$

系数 $\mathbb{E}_{(x,\ v)\sim\mu}[[f(x)]g(v)]$ 在许多情形下都可以显式计算，因此可得到显式的伪期望算子 $\mathbb{E}$。

对伪校准的一种直观理解如下。通常，植入分布会被选为在保留植入结构的同时具有最大熵的分布。这符合这样一种理念：随机实例对 SoS 来说很难处理，例如植入团问题中的均匀伯努利分布，或张量主成分分析中的高斯分布。通过条件化，使低阶矩与这样的植入分布相匹配，伪校准可以被理解为只观察低阶傅里叶特征，从而在随机分布和植入分布之间进行某种插值。这种直觉已被证明是有效的，因为伪校准已成功用于构造多种稠密及稀疏问题的 SoS 下界。

伪校准的一个优点是，这种构造会自动满足伪期望应具有的一些良好性质。根据 υ 构造，它是线性的。对于植入分布中满足以下形式的所有多项式等式： $f(x) = 0$ ，以下关系成立： $\mathbb{E}[f(x)] = 0$。对于植入分布中满足以下形式的其他多项式等式： $f(x, \upsilon) = 0$，等式 $\mathbb{E}[f(x, \upsilon)] = 0$ 近似成立。在大多数情况下，可以对 $\mathbb{E}$ 作轻微调整，使其精确满足这些等式。

条件 $\mathbb{E}[1] = 1$ 并非自动满足，但在大多数应用中，我们通常要求 $\mathbb{E}[1] = 1 \pm o(1)$。事实上，所有已知成功应用伪校准的情形都是如此。满足这一条件后，我们只需将最终的伪期望算子设为 $\mathbb{E}'$ 定义为 $\mathbb{E}'[f(x)] = \mathbb{E}[f(x)]/\mathbb{E}[1]$。

我们指出，条件 $\mathbb{E}[1] = 1 \pm o(1)$ 在预测可近似性与不可近似性之间的正确阈值方面相当成功[104, 110, 146]。在我们将伪校准与低次区分器联系起来时，这一点至关重要。

示例：植入团 作为热身，我们回顾植入团的伪校准计算。这里，随机分布 ν 为 $G(n, \frac{1}{2})$。

植入分布 μ 如下。给定整数 k，先从 $G(n, \frac{1}{2})$ 中采样 G' ，然后从顶点中选取一个随机子集 S ，其中每个顶点都以概率 $\frac{k}{n}$ 独立地被选中。对于 S 中所有不同顶点的点对 i, j，如果边 (i, j) 尚不存在，就将其加入图中。令 G 为所得的图。

输入由 $G \in \{-1, 1\}^{\binom{[n]}{2}}$ 给出，其中 $G_{i,j}$ 在边 (i, j) 存在时为 1，否则为 -1。令 $x_1, \ldots, x_n$ 为平方和程序中的布尔变量，使得 x_i 表示 i 是否属于该团。

给定一组顶点 $V \subseteq [n]$，定义 $x_V = \prod_{v \in V} x_v$。给定一组可能的边 $E \subseteq \binom{[n]}{2}$，定义 $\chi_E = (-1)^{|E \setminus E(G)|} = \prod_{(i,j) \in E} G_{i,j}$。

伪校准表示，对于所有小的 V 和 E,

$$\mathop{\mathbb{E}}_{G\sim\nu}\left[\tilde{E}[x_V]\chi_E\right] = \mathop{\mathbb{E}}_{\mu}[x_V\chi_E]$$

根据标准的傅里叶分析，这意味着如果我们取

$$c_E = \mathop{\mathbb{E}}_{\mu}[x_V\chi_E] = \left(\frac{k}{n}\right)^{|V\cup V(E)|}$$

其中 $V(E)$ 是 E 中各边端点的集合，那么对于所有较小的 V,

$$\widetilde{\mathbb{E}}[x_V] = \sum_{E:E \text{ is small}} c_E\chi_E = \sum_{E:E \text{ is small}} \left(\frac{k}{n}\right)^{|V\cup V(E)|}\chi_E$$

~由于 $\mathbb{E}[x_V]$ 的取值已知，根据多重线性，我们可以自然地将其扩展~，从而得到任意次数至多为平方和次数的多项式 f 的取值 $\mathbb{E}[f(x)]$ 。

在这里，我们只为较小的 E 设置傅里叶系数，并将其他较大的傅里叶系数设为 0。截断参数的选择通常取决于具体问题，但有一些基本要求 [104]。现在我们概述证明平方和下界的一般策略。我们在所有结果中都采用这一策略。

3.3.1 证明平方和下界的策略

在这项工作中， 证明 SoS 下界的一般策略可概括如下 .

- 给定一个随机分布，找出一个合适的植入分布
- 针对这两个分布进行伪校准，并得到一个候选伪期望算子
- 证明矩矩阵满足约束

这种方法中技术上最具挑战性的部分通常是证明矩矩阵是半正定的。我们的许多贡献都集中在这一步，我们在此分析由此得到的随机矩矩阵的行为。现在，我们将伪校准与低次区分器联系起来。

3.3.2 与低次区分器的联系

我们现在可以将伪校准与低次测试联系起来。回顾一下，在伪校准中，我们将高阶傅里叶系数设为 0。这称为截断。具体而言，我们进行截断，使得到的伪期望关于输入的度至多为 D in$\widetilde{}$ $\widetilde{}$。根据构造，$\mathbb{E}[\mathbb{E}[1]] = 1$, 我们想要了解 $\mathbb{E}[1]\widetilde{}$ 偏离 1 的程度。以下引理指出，$\mathbb{E}[1]$ 的方差表现得如同最优度为 D 的区分器的平方值。

$\widetilde{}$**引理 3.3.1.** 伪校准伪期望 $\mathbb{E}$ 截断至度 D，满足

$$\mathrm{var}(\widetilde{\mathbb{E}}[1]) = \left\| LR^{\leq D} - 1 \right\|^2$$

$\widetilde{}$ $\widetilde{}$证明. 伪校准将所有 $i \leq t$ 的 $\mathbb{E}_{x\sim H_0}[\mathbb{E}[1]\chi_i] = \mathbb{E}_{x\sim H_1}[\chi_i]$ 设为 $\mathbb{E}[1] =$。因此，$1 + \sum_{1\leq i\leq t} \mathbb{E}_{x\sim H_1}[\chi_i]\chi_i$，从而 var($\widetilde{\mathbb{E}}$[1]) $= \sum_{1\leq i\leq t}(\mathbb{E}_{x\sim H_1}\chi_i)^2 = \left\| LR^{\leq D} - 1 \right\|^2$。 ∎

我们的 SoS 下界证明中的一个关键步骤，是在伪-$\widetilde{}$校准后验证 $\mathbb{E}[1]$表现良好。特别是，对于强 SoS 下界，我们预期$\sim\mathbb{E}[1] = 1 + o(1)$。虽然这并非形式上必需，但在我们的应用中通常如此，而且我们预计，通过这种方法获得强 SoS 下界必须满足这一条件。

但如果确实如此，而且我们证明了平方和下界，请注意，这已经是支持低次似然比假设的有力证据。具体而言，根据引理 3.2.4 和引理 3.3.1，最优的度 D 区分器无法区分这

两个分布 μ, ν。我们的下界表明，强大的 SoS 层级同样无法区分这两个分布；这是迈向一般性假设的重要一步。

在该领域，一个重要的开放问题是证明：对于足够良好的分布$\sim\mu$, ν，进行伪校准后，$\mathbb{E}[1] = 1 + o(1)$ 蕴含强 SoS 下界的存在。

3.4 为什么选择平方和?

我们简要回顾平方和在过去十年中的成果，尤其是在鲁棒机器学习领域的成果。鲁棒机器学习是机器学习的一个分支，其底层数据集含有噪声，而噪声可能是随机的，也可能具有对抗性。近年来，鲁棒机器学习受到了广泛关注，这是因为它在机器学习及其他下游应用中具有多种用途，其中包括自动驾驶等安全关键型应用。例如，计算机视觉 [223, 90, 237, 97, 213, 236, 79, 147]和语音识别 [113, 232, 202, 203, 152, 3, 172, 177]领域已经积累了大量实践成果。在这一重要领域，平方和最近针对长期存在的开放问题，带来了突破性算法 [15, 154, 101,136, 80, 140, 16, 17, 211]。主要成果包括

- 鲁棒学习高维高斯混合分布。这是一个极其重要的问题，长期以来受到广泛关注，相关研究不断推进，最终 culminated in [15, 154]。

- 存在离群值时，回归 [136], 矩估计 [140], 聚类 [16]和子空间恢复 [17] 等基本问题的高效算法。

此外，人们认为平方和算法是许多统计问题的最优鲁棒算法。另一方面，平方和算法促成了快速算法的设计，可用于张量分解等问题 [106, 211]。

广义而言，由于平方和能够涵盖各种各样的算法技术，它已成为算法与优化领域的一项基础工具。过去如此，如今它仍然是组合优化的一种极其灵活的工具 [89, 10, 94, 196]；但正如我们上文所见，它也被广泛应用于统计学和机器学习（除上述参考文献外，另请参见 [21, 23, 108, 193]）。这为本文后续内容奠定了基础：我们将分析平方和在统计学和统计物理领域中若干值得关注的问题上的应用。

第4章

平方和下界的主要结果

本章正式陈述本文证明的平方和下界主要结果，并将其置于先前工作的背景下。本章内容改编自 [87, 191], ，相关结果最初发表于该处。不过，本章与这些工作不同之处在于，我们着重介绍了这些工作的近期进展，提及了最近发现的与其他问题之间的联系，并依次呈现证明技巧；这在教学上有所帮助，因为这些证明的核心原理并非完全不同。

4.1 Sherrington-Kirkpatrick哈密顿量

我们首先定义高斯正交系综 GOE(n)，这是一种用于$n \times n$ 矩阵的随机矩阵模型。

定义 4.1.1. 高斯正交系综，记为$\mathrm{GOE}(n)$，是 $\frac{1}{\sqrt{2}}(A + A^\intercal)$ 的分布，其中 A 是一个由独立同分布标准高斯条目构成的随机 $n \times n$ 矩阵。

等价地，我们可以将 GOE(n) 定义为对称矩阵上的概率分布 W，满足 $W_{ii} \sim \mathcal{N}(0, 2)$（对于 $i \leq n$）以及 $i \neq j$， $W_{ij} = W_{ji} \sim \mathcal{N}(0, 1)$（相互独立）。

我们考虑主要的优化任务

$$\mathrm{OPT}(W) := \max_{x \in \{\pm 1\}^n} x^\intercal W x, \tag{4.1}$$

其中， W 是 $\mathbb{R}^{n \times n}$ 中的一个随机对称矩阵。这是一项在计算机科学和统计物理中都会出现的重要任务。

在计算机科学中，W 的一种自然选择是将其设为图的拉普拉斯矩阵 [99,节 4]。那么，该问题就等价于最大割问题；众所周知，该问题在最坏情况下是 NP难问题 [122]。只需注意到，$x \in \{\pm 1\}^n$ 可以被视为对 $[n] = \{1, 2, \ldots, n\}$进行二分划分的编码，即可立即得出这一等价关系。

尤其是一个有趣的特殊情形：考虑从 Erd˝os-R´enyi 图 $G(n, \frac{d}{n})$（平均度为 d）中抽样得到的稀疏随机图，或均匀随机选取的 d-正则图，其中 $d \geq 3$ 是一个固定整数。在这种情形下，已知最大割的真实规模渐近为 $n(\frac{d}{4} + f(d)\sqrt{d})$。此外，文献 [58]（最初在文献 [239]中提出猜想）证明了 $\lim_{d\to\infty} f(d) = \frac{1}{2}P^* \approx 0.382$，其中

$$P^* := \frac{1}{2} \lim_{n\to\infty} \mathbb{E}_{W\sim \mathrm{GOE}(n)}[\frac{1}{n^{3/2}} \mathrm{OPT}(W)] \approx 0.7632$$

称为帕里西常数。这已经充分说明了研究 $W \sim$GOE(n) 时的式 (4.1) 是很有意义的。有趣的是，这个问题还有另一个绝妙的动机。

在统计物理中，当 $W \sim$ GOE(n) 时，我们的目标（忽略缩放因子）就是著名的 Sherrington-Kirkpatrick模型的哈密顿量。在这里，x 可以看作是自旋玻璃模型中自旋取值的编码。$-W_{i,j}$ 刻画自旋 x_i 与 x_j 之间的相互作用（其中$-W_{i,j} \geq 0$ 为铁磁作用，而 $-W_{i,j} < 0$ 为反铁磁作用）。因此，最优值（相差一个符号）对应于系统的最小能量，即基态。文献 [183, 184, 54] 通过非严格的方法预测了 $P^* \approx 0.7632$。这一结果最终在文献 [224, 179, 93]中得到了严格证明。

在这项工作中，当 $W \sim$ GOE(n) 时，我们将聚焦于这个平均情形优化问题。首先自然要问的是，是否存在一种多项式时间算法，给定$W \sim$ GOE(n)，能够计算出一个 x，使其达到接近 OPT(W) 的值。Montanari [168] 近期的一项突破性工作表明，对于任意 $\varepsilon > 0$，存在一种多项式时间算法，给定 W 后输出 x，并且以高概率达到 $(2P^* - \varepsilon)$ $n^{3/2}$

（假设一个广受认可的猜想成立）。

现在我们转向认证：对于任意输入 W，是否存在一种高效算法，能够证明 OPT(W) 的一个上界?

一种简单的算法是谱算法：我们只需输出 W 的最大特征值（经过适当缩放），作为上界。注意，GOE(n) 是一种特殊的威格纳矩阵系综，因此满足半圆律；在此情形下，这说明 W 的最大特征值以概率 $1-o_n(1)$ 为 $(2+o_n(1))\cdot\sqrt{n}$。因此，一个平凡的谱界表明，OPT(W) $\leq (2+o_n(1))\cdot n^{3/2}$，且这一结论以概率 $1-o_n(1)$ 成立。

现在，我们可以问：认证方面是否有可能超越这种谱算法? 具体来说，我们可以考察平方和作为认证算法的表现如何。通过 W 的谱范数得到的自然上界 $(2+o_n(1))\cdot n^{3/2}$，也是度-2 平方和松弛的值 [170]。Mohanty–Raghavendra– Xu [165] 和 Kunisky–Bandeira [145] 最近开展的两项独立研究表明，度-4 平方和的表现不会好多少；而 [19] 给出的启发式论证则表明，即使是度为 (n/log n) 的平方和，也无法证明比平凡谱界更强的结论。因此我们提出以下问题：

更高次数的平方和能否为 *Sherrington–Kirkpatrick* 问题证明更好的上界，并有望更接近真实界 $2\cdot P^*\cdot n^{3/2}$?

在这项工作中，我们对上述问题给出了否定的回答：我们证明，即使次数高达 n^δ，平方和也无法优于基本的谱算法。

定理 4.1.2. 存在一个常数 $\delta>0$，使得以高概率，对于 $W\sim$ 高斯正交系综(n)，存在一个度为 n^δ 的平方和解，其 *Sherrington–Kirkpatrick* 问题的取值至少为 $(2-o_n(1))\cdot n^{3/2}$。

Kunisky [143] 独立且同时进行的研究也证明了上述定理针对6度平方和的一个特殊情形，但采用了不同的技术。

我们将在第5章介绍该定理的证明。上述定理及其证明最初发表于 [87], ，本文内容即改编自该文。现在，我们介绍该定理证明背后的高层次思路。

4.1.1 我们的方法

为证明定理 4.1.2，我们首先引入一个新的平均情形问题，称为植入仿射平面（PAP），并直接证明其平方和下界。然后，我们利用 PAP 下界证明 Sherrington–Kirkpatrick 问题的下界。PAP 问题可以非正式地描述如下（正式定义见定义 5.1.1）。

定义 4.1.3（PAP 的非正式表述）。 给定 m 个随机向量 $d_1, \ldots, d_m$ 属于 $\mathbb{R}^n$，我们能否证明不存在向量 $v \in \mathbb{R}^n$，使得对所有 $u \in [m]$, $\langle v, d_u \rangle^2 = 1$ 都成立？换言之，我们能否证明这些 m 个随机向量并非全都包含在两个与原点距离相等的平行超平面中？

当我们将 v 限制为 $\{\pm \frac{1}{\sqrt{n}}\}^n$中的一个布尔向量时，这个问题可以编码为以下多项式系统的可行性问题

$$\begin{aligned} \exists v \in \mathbb{R}^n \text{ s.t.} \qquad & \forall i \in [n],\ v_i^2 = \frac{1}{n}, \\ & \forall u \in [m],\ \langle v, d_u \rangle^2 = 1. \end{aligned}$$

因此，它是平方和的理想候选。然而，我们证明，对于输入，以高概率平方和无法反驳一个随机实例。对 v 施加布尔限制实际上使下界结果更强，因为平方和甚至无法反驳 $\mathbb{R}^n$中的更小向量子集。在本文中，我们将考虑两种不同的随机分布，即d_1, ……, d_m 是从多元正态分布中独立抽取的样本，以及它们是布尔超立方体上的均匀分布中的独立样本。

定理 4.1.4。 对于高斯情形和布尔情形，存在一个常数 $c > 0$，使得对于所有 $\varepsilon > 0$ 和 $\delta \le c\varepsilon$，当 $m \le n^{3/2-\varepsilon}$ 时，以高概率存在一个关于植入仿射平面的可行度为-n^{δ} 平方和解。

事实证明，上文介绍的植入仿射平面问题与以下“随机子空间中的布尔向量”问题密切相关。我们将其称为植入布尔向量问题； [165] 在研究平方和方法计算 Sherrington–Kirkpatrick 哈密顿量的性能时提出了这个问题。

植入布尔向量问题旨在证明随机子空间远未包含布尔向量。具体而言，我们希望证明以下量的一个上界：

$$\mathrm{OPT}(V) := \frac{1}{n} \max_{b \in \{\pm 1\}^n} b^{\intercal} \Pi_V b,$$

其中， V是 $\mathbb{R}^n$中的一个均匀随机 p维子空间[1] ，而 Π_V是投影到V上的投影算子。简而言之，植入仿射平面问题与此问题的联系在于，PAP 向量 v表示向量 b 在V的一组基所张成空间中的线性组合系数。

[165] 的一个论证表明，当 $p \ll n$时，以高概率有 OPT(V) $\approx \frac{2}{\pi}$；他们还表明，假设 $p \ge n^{0.99}$，以高概率存在一个值为 $1 - o_n(1)$的 4 度平方和解。他们提出的问题是，是否存在能够证明更紧上界的多项式时间算法；我们则在更大的范围内排除了基于平方和的算法，涉及更高的平方和次数以及随机子空间的维度 p 。

定理 4.1.5. 存在一个常数 $c > 0$ ，使得对于所有 $\varepsilon > 0$ 和 $\delta \le c\varepsilon$，当$p \ge n^{2/3+\varepsilon}$时，以高概率，对 V，存在一个值为 1 的 *Planted Boolean Vector* 的 n^{δ} 度平方和解。

我们的技术贡献主要在于植入

1. V可以由一个基指定，该基由 p 个来自 $\mathcal{N}(0, I)$ 的独立同分布样本组成。

仿射平面问题的平方和下界，即定理 4.1.4。随后我们证明，高斯情形下的植入仿射平面问题与植入布尔向量问题等价。从 Sherrington-Kirkpatrick 到植入布尔向量问题的归约由 Mohanty– Raghavendra–Xu [165] 提出。

作为证明 PAP 下界的起点，我们采用伪校准来构造一个~良好的候选平方和解 $\mathbb{E}$。遗憾的是，算子 $\mathbb{E}$ 并未精确满足 PAP 约束“$\langle v, d_u\rangle^2 = 1$”，而只是在极小误差范围内满足这些约束。在原始工作中，~我们采用了一种有趣且相当通用的方法，将 $\mathbb{E}$ 舍入为一个邻近的伪期望~算子 $\mathbb{E}'$，使其精确满足这些约束。为简洁起见，本论文省略了这一部分，详情请见原始工作 [87]。

对于度数 D，候选平方和解可以看作一个（伪）矩矩阵$\mathcal{M}$，其行和列以子集 I, $J \subset [n]$ 为索引，大小受 $D/2$ 限制，且其元素为

$$\mathcal{M}[I, J] := \widetilde{\mathbb{E}}[v^I v^J].$$

矩阵 $\mathcal{M}$ 是输入 d_1、.. . 、 d_m的随机函数，而分析中最具挑战性的部分是证明 $\mathcal{M}$ 以高概率为半正定（PSD）。

与 [20], 类似，我们将 $\mathcal{M}$ 分解为图矩阵的线性组合，即 $\mathcal{M} = \sum_\alpha \lambda_\alpha \cdot M_\alpha$，其中 M_α 是与形状 α 相关联的图矩阵。简言之，每个图矩阵汇总了 $\mathcal{M}$ 各元素的傅里叶展开中形状为 α 的所有项——形状 α 非正式地说是一个带标签边的图，其大小有 poly(D) 的上界。 $\mathcal{M}$ 的图矩阵分解在半正定分析中特别有用，因为单个图矩阵 M_α 的算子范数（以高概率）由图 α 的简单组合性质决定。与 [20] 相比，一个技术上的不同之处是，我们的图矩阵有两种顶点类型: 和 ；这些图矩阵属于 Ahn 等人在 [1] 中提出的一般框架。

□ ○

要证明矩阵 $\mathcal{M}$ 是半正定的，我们需要研究分解中非零系数对应的图矩阵。矩阵 $\mathcal{M}$ 可以拆分为多个块，每个对角块在分解中都包含一个（经缩放的）单位矩阵。从图矩阵视角来看，这意味着分解中出现了某些“平凡”形状，并带有适当的系数。如果我们能够将所有其他出现的图矩阵的范数相对于这些平凡形状进行界定，并证明它们的范数总和相较于这些经缩放的单位矩阵块之和可以忽略不计，那么我们就能顺利解决问题。

遗憾的是，这种方法行不通。由于满足 PAP 约束“$\langle v, d_u \rangle^2 = 1$”，矩阵 $\mathcal{M}$ 的核并非平凡，因此根本无法证明分解 $\mathcal{M}$ 中所有非平凡形状的贡献都具有较小的范数。事实上，分解 $\mathcal{M}$ 中出现的某些形状 α 具有$\|\lambda_\alpha \cdot M_\alpha\|$ 较大的性质。结果表明，所有这类形状都具有一种简单的图论子结构，因此我们将这些形状称为蜘蛛。

为了解决零空间问题，我们将范围限制在 Null($\mathcal{M}$)$^\perp$，它是 $\mathcal{M}$的零空间的补空间。我们证明，蜘蛛中存在的子结构意味着该蜘蛛在 Null($\mathcal{M}$)$^\perp$中接近零矩阵。因此，对于蜘蛛 α，我们几乎可以任意加减 M_α，同时保持 $\mathcal{M}$ 在 Null($\mathcal{M}$)$^\perp$上的作用。我们的策略是通过对每个蜘蛛 α减去 $\lambda_\alpha \cdot M_\alpha$ 来“消灭”蜘蛛。然而，由于 M_α 只是近似属于 Null($\mathcal{M}$)$^\perp$，这一策略可能会引入新的图矩阵项，尤其可能引入新的蜘蛛。为此，我们递归地消灭这些蜘蛛，同时仔细分析所有图矩阵的系数如何变化。消灭所有蜘蛛后，得到的矩阵矩阵变为

$$\sum_{0 \le k \le D/2} \frac{1}{n^k} \cdot I_k + \sum_{\gamma\text{: non-spiders}} \lambda'_\gamma \cdot M_\gamma,$$

对于某些新的系数 λ'_γ。这里，I_k 是一个在第 k 个块中含有恒等矩阵、其余元素为 0 的矩阵

采用一种新颖的收费论证，我们最终证明后一个项相较于前一个项可以忽略不计，从而确立了 $\mathcal{M} \succeq 0$。

4.1.2 相关工作

关于 Sherrington-Kirkpatrick 哈密顿量问题的 4 度平方和下界，由 Mohanty–Raghavendra–Xu [165] 和 Kunisky–Bandeira [145]独立证明。Kunisky [143] 同期独立开展的工作则得到了 6 度平方和下界。在本文中，我们针对某个常数 $\delta > 0$证明了一个更强的 n^{δ} 度平方和下界。我们的结果是通过将 Sherrington-Kirkpatrick 问题归约为“随机子空间中的布尔向量”问题而得到的；后者等价于我们提出的、基于正态分布的全新植入仿射平面问题。将 Sherrington-Kirkpatrick 问题归约为“随机子空间中的布尔向量”问题的方法由 Mohanty–Raghavendra–Xu [165]提出。Mohanty–Raghavendra–Xu [165] 和 Kunisky–Bandeira [145] 的结果建立在 Montanari 和 Sen [170]证明的 2 度平方和下界之上。

针对 $p \geq n^{0.99}$ “随机子空间中的布尔向量”问题的四度平方和下界由 Mohanty–Raghavendra–Xu 在 [165] 首次提出该问题时证明。我们将对 p 的依赖改进为 $p \geq n^{2/3+\varepsilon}$，适用于任意 $\varepsilon > 0$，并针对某个绝对常数 $c > 0$得到更强的度为-$n^{c\varepsilon}$ 平方和下界。

有趣的是，近期的工作 [238]针对 $m \geq n+1$ 的植入仿射平面搜索变体，提出了一种多项式时间算法，并达到了统计最优性。具体而言，他们超越了先前已知的多项式时间算法，包括基于平方和的方法，而这些算法都要求 $m \gg n^2$ [161]。这种新算法是一种基于格的方法，利用了该问题中特有的代数结构。因此，他们的算法无法抵御微小扰动，也就是说，他们要求这些点恰好位于平面上。另一方面，诸如 [161] 的工作所提出的谱算法则能抵御噪声。正因为必须没有噪声，这种基于格的算法与以下情形有类似之处：

高斯消元法在没有噪声时可以超越平方和下界。具体而言，这意味着这种基于格的算法并未推翻我们的认证下界，也没有推翻第3.2.1节中描述的低度似然比假设。

4.2 稀疏主成分分析

主成分分析（PCA） [118] 是一种广泛使用的数据处理和降维方法，在机器学习、统计学、工程学、生物学等领域有着众多应用。给定一个数据集，PCA 会将数据投影到由主成分张成的低维空间中。其核心思想是，PCA 能够去除噪声等低阶信息，同时保留数据中大量供下游任务使用的内在信息。

然而，尽管 PCA 具有许多最优性优势，它也存在缺点。首先，由于主成分是所有原始变量的线性组合，因此很难对其进行解释 [158]。其次，众所周知，PCA 在高维环境下无法得到良好的估计量 [14, 187, 117]。

为了解决这些问题，人们常采用一种称为稀疏主成分分析的 PCA 变体。稀疏主成分分析在寻找数据的主成分时，额外加入稀疏度约束。具体而言，考虑给定的数据 v_1，v_2、.....、$v_m \in \mathbb{R}^d$。在稀疏主成分分析中，我们希望在额外约束主成分的稀疏度至多为 k的条件下，找到数据的首要主成分。也就是说，我们要找到一个向量 $v \in \mathbb{R}^d$，使 $\sum_{i=1}^m \langle v, v_i \rangle^2$ 最大化，并满足 $\|v\|_0 \leq k$。

稀疏主成分分析已广泛应用于众多领域，包括医学、计算生物学、经济学、图像与信号处理、金融，当然也包括机器学习和统计学（例如 [230, 171, 159, 225, 51, 2]）。此外，稀疏主成分分析的一项重要优势是其各个成分更容易解释。一个典型例子是从文档中提取主题 [57, 182]。此外，可解释性对于机器学习中的算法公平性也有重要益处。

大量研究致力于探讨稀疏主成分分析及其变体。多项研究提出并考察了相关算法，例如 [4, 157, 141, 60, 231,28, 156, 61, 104, 37, 64, 50, 66]。例如，在标准主成分分析基础上进行阈值化等简单的主成分分析变体 [117, 50]，在某些参数设置下效果良好。这自然引出了一个问题：更复杂的算法能否在这些设置或其他参数设置下取得更好的效果?

另一方面，也有从不可近似性角度开展的研究（例如 [28, 104, 37, 141, 64, 231], 我们将在陈述主要结果后更详细地概述）。具体而言，由于证明无条件下界十分困难，许多这类不可近似性结果都依赖于各种其他猜想。尽管已有这些先前工作，要准确理解高效算法在这一问题上的能力极限，仍是一个活跃的研究领域。考虑到稀疏主成分分析的重要性，以及它对于众多应用的基础性作用，这一点并不奇怪。

因此，我们自然会提出这个问题（先前的研究[156, 104, 110]也提出过这一问题，并将其列为一个开放问题）：

平方和算法能否在稀疏主成分分析问题上胜过已知算法?

在这项工作中，我们证明，即使允许使用亚指数时间，平方和算法也无法胜过已知的谱算法！因此，这表明，目前使用的阈值化算法或其他谱算法，在某种意义上是解决这一问题的最优算法。

为了证明我们的结果，我们将考虑稀疏主成分分析的随机实例，并证明它们对平方和而言天然地难以处理。具体来说，我们重点研究稀疏主成分分析的 Wishart 随机模型。与此前研究过的其他随机模型（例如 Wigner 随机模型）相比，这一模型是更自然的建模假设。

请特别注意，由于我们证明的是不可能性结果，我们的模型假设只会强化这些结果。换言之，如果平方和算法对这个受限版本的稀疏主成分分析不起作用，那么它对更一般的模型也不会起作用，例如具有

一般协方差或多重尖峰的模型。现在我们来描述该模型。

稀疏主成分分析的 Wishart 模型，也称为尖峰协方差模型，最初由 [117] 提出。在该模型中，我们观测到来自分布 $\mathcal{N}(0, I_d + \lambda uu^T)$ 的 m 个向量 $v_1, \ldots, v_m \in \mathbb{R}^d$，其中 u 是一个 k-稀疏单位向量，即 $\|u\|_0 \leq k$，我们的目标是恢复主成分 u。这里，向量的稀疏度是非零项的数量，而 λ 被称为信噪比。

随着信噪比 λ 降低，由于 u 在数据中留下的特征变得越来越微弱，恢复 u 会更加困难，甚至可能变得不可能。然而，如果样本数量 m 增加，这种情况或许可以得到缓解。因此，这里存在 m, n 与 k 之间的权衡。较早提出的算法已经能够在不同情形下恢复 u。例如，如果样本数量非常大，即 $m \gg \max(\frac{d}{\lambda}, \frac{d}{\lambda^2})$，那么标准主成分分析将会奏效。如果情况并非如此，通过假设稀疏度相对于样本数量而言并不太大，即 $m \gg \frac{k^2}{\lambda^2}$，我们可能仍然能够恢复 u。为此，我们使用一种称为对角线阈值化的标准主成分分析变体。在不同情形下也得到了类似结果，但有些情形仍然难以通过算法攻克。

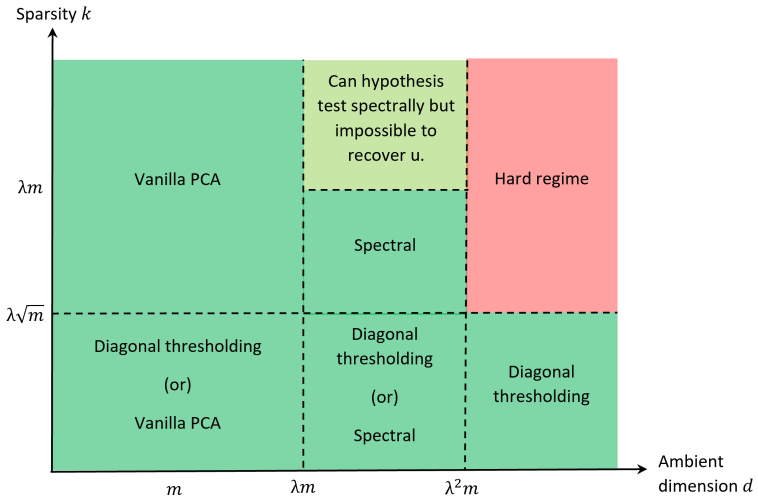
我们的结果完善了这一图景：在迄今仍难以被高效算法攻克的参数范围内，强大的平方和算法同样无法恢复主成分。下面给出定理的非正式表述，正式表述留待定理 4.2.2。

定理 4.2.1。 对于稀疏主成分分析的 *Wishart* 模型，当样本数量 $m \ll \min(\frac{d}{\lambda^2}, \frac{k^2}{\lambda^2})$ 时，亚指数时间 *SoS* 算法无法恢复主成分。

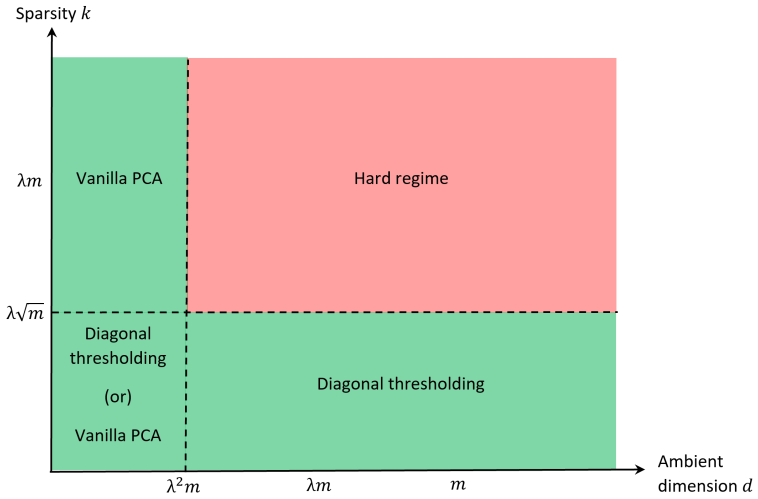
特别是，该定理解决了 [156]和 [104, 110]提出的一个开放问题。

在几乎所有其他情形下，都存在恢复主成分 u。我们很快会概述这类算法，并在图 4.1 中简洁地加以总结。之所以说几乎所有其他情形，是因为有一种有趣的情形，即 $\frac{d}{\lambda^2} \leq m \leq \min(d,k)\,\lambda$，在图 4.1 中以浅绿色标出，其特点是

从信息论角度来看，我们无法恢复u，但可以对稀疏主成分分析进行假设检验。也就是说，在这种情形下，我们可以区分纯随机的无尖峰样本与尖峰样本。然而，即使使用指数时间暴力算法，我们也无法恢复主成分。



(a) 信噪比 $\lambda \geq 1$



(b) 信噪比 $\lambda < 1$

图 4.1：稀疏主成分分析的计算障碍图

现在，我们更正式地陈述一下我们的结果。首先，我们假设 u 的各项从 $\{-\frac{1}{\sqrt{k}}, 0, \frac{1}{\sqrt{k}}\}$ 中选取，使得稀疏度为 k（因此范数为 1）。

请特别注意，这一假设只会强化我们的结果：如果平方和连这个特定的 u 情形下的问题都无法解决，那么对于 u 任意取值的一般问题，它也不可能做得更好。

设给定数据集中的向量为 v_1、.....、v_m。令它们构成矩阵 $S \in \mathbb{R}^{m\times d}$ 的各行。令 $\Sigma = \frac{1}{m}\sum_{i=1}^{m} v_i v_i^T$ 为样本协方差矩阵。那么，标准的主成分分析目标是最大化 $x^T\Sigma x$ 并求出 $x = \sqrt{k}u$。因此，稀疏主成分分析问题可以改写为

$$\text{maximize } m \cdot x^T\Sigma x = \sum_{i=1}^{m}\langle x, v_i\rangle^2 \text{ such that } x_i^3 = x_i \text{ for all } i \le d \text{ and } \sum_{i=1}^{d} x_i^2 = k$$

其中，程序变量为 x_1、.....、x_d。约束 $x_i^3 = x_i$ 确保 x 的各元素属于 $\{-1, 0, 1\}$；结合这些约束，最后一个条件 $\sum_{i=1}^{d} x_i^2 = k$ 确保稀疏度为 k（但我们指出，出于技术原因，在下界中我们只能在误差不超过 $o(1)$ 的范围内满足这一条件）。然后，向量 u 可以通过设定 $u = \frac{1}{\sqrt{k}}x$ 求得。

现在，我们将考虑由平方和算法得到的稀疏主成分分析的一系列凸松弛。具体而言，我们将考虑 d^{ε} 的 SoS 次数，其中 $\varepsilon > 0$ 是一个较小的常数。请注意，这对应于关于输入规模具有亚指数运行时间的平方和算法$_{d^{O(1)}}$。

我们的主要结果表明，对于低于某一阈值的 m，当向量v_1、.....、v_m从未加尖峰的标准高斯分布 $\mathcal{N}(0, I_d)$ 中抽样时，亚指数时间平方和算法得到的最优值将接近 $m + m\lambda$。当向量 v_1、.....、v_m 确实从尖峰高斯分布$\mathcal{N}(0, I_d + \lambda uu^T)$ 中抽样且 $x = \sqrt{k}u$ 时，这也是最优值。因此，平方和无法区分 $\mathcal{N}(0, I_d)$ 与$\mathcal{N}(0, I_d + \lambda uu^T)$，因而无法解决稀疏主成分分析问题。形式化地，

定理 4.2.2。对于所有充分小的常数 $\varepsilon > 0$，假设 $m \le \frac{d^{1-\varepsilon}}{\lambda^2}$、$m \le \frac{k^{2-\varepsilon}}{\lambda^2}$，

并且对于某些 $A > 0$、 $d^A \leq k \leq d^{1-A\varepsilon}$、 $\frac{\sqrt{\lambda}}{\sqrt{k}} \leq d^{-A\varepsilon}$，则对于绝对常数 $C > 0$，以高概率（关于一个随机的 $m \times d$ 输入矩阵 S，其具有高斯条目），稀疏主成分分析的次指数时间平方和算法，其度为 $d^{C\varepsilon}$ ，最优值至少为 $m + m\lambda - o$（1）。

换句话说，亚指数时间的平方和程序无法证明，对于具有高斯条目的随机数据集，不存在具有 k 个非零项且满足 $m \cdot u^T \Sigma u \approx m + m\lambda$ 的单位向量 u。

这里有几点需要说明。

1. 请注意，当输入向量 $v_1, \ldots, v_m$ 确实从尖峰模型 $\mathcal{N}(0, I_d + \lambda uu^T)$ 中采样，且 $x = \sqrt{k}u$时， $m + m\lambda$ 大约等于平方和程序的值。因此，亚指数时间的平方和程序无法区分完全随机分布与尖峰分布，因而无法解决稀疏主成分分析问题。

2. 常数 A 可以看作是 ≈ 0 ；引入它是出于技术原因，以确保我们的界具有充分衰减。具体而言， k, λ的大多数取值都满足该定理的条件。

3. 出于技术原因，在我们的下界中，约束 $\sum_{i=1}^{k} x_i^2 = k$ 仅在误差为 $o(1)$ 的范围内成立。我们将精确满足这一约束的问题留待未来研究。

非正式地说，我们的主要结果表明，当 $m \ll \min\left(\frac{d}{\lambda^2}, \frac{k^2}{\lambda^2}\right)$ 时，亚指数时间的平方和方法无法恢复主成分 u。这就是定理 4.2.1 的内容。

为了证明我们的结果，我们采用第 3.3.1 节中的策略。对于稀疏主成分分析的 Wishart 模型，我们使用以下分布。

- 随机分布 ν： $v_1, \ldots, v_m$ 从 $\mathcal{N}(0, I_d)$ 中采样，并令 S 为行向量分别为 $v_1, \ldots, v_m$ 的恒等矩阵。

• 植入分布 μ：从 $\{-\frac{1}{\sqrt{k}}, 0, \frac{1}{\sqrt{k}}\}^d$ 中采样 u，其中各个值的取值概率分别为 $\frac{k}{2d}$，$1-\frac{k}{d}$，$\frac{k}{2d}$。然后按如下方式采样 $v_1, \ldots, v_m$。对于每个$i \in [m]$，以概率 Δ 从 $\mathcal{N}(0, I_d + \lambda uu^T)$ 中采样 v_i，以概率$1-\Delta$ 从 $\mathcal{N}(0, I_d)$ 中采样 v_i。最后，令 S为行向量分别为$v_1, \ldots, v_m$ 的 $m \times d$矩阵。

在第 6.4 节中，我们计算通过伪校准得到的平方和解。我们证明以下定理。

定理 4.2.3. 存在一个常数 $C > 0$，使得对于所有充分小的常数$\varepsilon > 0$，若 $m \leq \frac{d^{1-\varepsilon}}{\lambda^2}$、$m \leq \frac{k^{2-\varepsilon}}{\lambda^2}$，并且存在一个常数 A，使得 $0 < A < \frac{1}{4}$、$d^{4A} \leq k \leq d^{1-A\varepsilon}$ 和 $\frac{\sqrt{\lambda}}{\sqrt{k}} \leq d^{-A\varepsilon}$，那么以高概率，通过伪校准得到的、针对度为 $d^{C\varepsilon}$ 的平方和解是可行的。

由于我们使用的是平均情况分布，这个平方和下界也是认证的下界。我们的证明概述见第 4.5 节。由此定理可推得定理 4.2.2。

算法方面的先前工作 鉴于稀疏主成分分析的重要性广为人知，理论和实践中都已有大量工作致力于为稀疏主成分分析设计算法，
[4, 157, 141, 60, 231, 28, 156, 61, 104, 37, 64, 50, 66] 仅举几例。

现在，我们将本研究的结果置于已知稀疏主成分分析算法的背景下，并解释为何它在可近似性与不可近似性之间实现了紧密的权衡。结合本研究与先前工作，我们对稀疏主成分分析在多对数因子范围内哪些参数范围容易求解、哪些被猜想为困难，已有了完整的认识。在图 4.1a 和图 4.1b 中，我们将不同的参数范围划分为以下几类。

• 对角线阈值化：在此情形下，对角线阈值化 [117, 4] 可以恢复稀疏向量。协方差阈值化 [141, 60]和平方和算法 [67] 也可以

用于此情形。这些替代算法的优点在于，协方差阈值化对对数因子的依赖性更佳，而平方和算法在存在对抗性误差时仍可发挥作用。

- 标准 PCA：标准 PCA（即普通 PCA）可以恢复该向量，也就是说，我们不需要利用该向量是稀疏向量这一事实（例如参见 [29, 67]）。

- 谱算法：一种高效的谱算法可以恢复稀疏向量（例如参见 [67] ）.

- 可以检验但无法恢复：一种简单的谱算法可以解决稀疏主成分分析的假设检验问题，但从信息论角度而言，不可能恢复稀疏向量 [67, 附录 E]。

- 困难：在这一范围内，猜想算法难以恢复稀疏主成分。我们将在下文对此进行更详细的讨论。

在图 4.1a 和图 4.1b 中，对应于对角线阈值化、标准 PCA 和谱算法的范围以深绿色表示，而对应于谱算法*和困难情形的范围则分别以浅绿色和红色表示。

关于困难性的先前工作 先前工作探讨了统计查询下界 [39],基本半定规划下界 [141], 基于猜想的困难问题的归约 [29, 28, 37, 84,231], 基于低次猜想的下界 [64, 67], 基于统计物理的下界[64, 13], 等。我们注意到， [67],曾预测过与我们类似的阈值行为，但重要的是，他们假设了一个称为低次似然猜想的猜想。同样，许多其他下界也依赖于各种猜想。作为背景，低次似然猜想是比 P $\neq$ NP 更强的假设。相比之下，我们的结果是无条件的，不依赖任何猜想。

与这些其他下界相比，针对平方和算法的下界研究此前仅有两项 [141, 29, 156] ，且仅涉及 2 度和 4 阶平方和。

具体而言， [141, 29] 研究了 2 度平方和下界，尽管他们并未以这种方式表述。此外， [156] 得出了 4 次平方和下界，但这些下界损失很大，也就是说，它们仅适用于困难区域的一个真子集 $m \ll \frac{k^2}{\lambda^2}$ 和 $m \ll \frac{d}{\lambda^2}$。此外，这些先前工作所用的思路无法推广到更高阶。其他平方和下界的缺失可归因于证明此类下界的困难。在本文中，我们大幅强化了这些已知结果，并证明了度数为 d^ε 的平方和算法几乎紧致的下界，这些算法对应于亚指数运行时间 $d^{d^{O_{(\varepsilon)}}}$。我们指出，平方和算法的度数越高，能力越强，因此我们的结果立即蕴含了这些先前结果；即便是在 4 阶平方和这一特殊情形下，我们也改进了已知的宽松下界。总之，定理 4.2.2 涵盖了所有这些早期已知结果，并大幅改进了此前已知的平方和下界，为该参数范围内稀疏主成分分析的困难性提供了有力证据。

文献 [104] 也给出了稀疏主成分分析的 SoS 下界，但它与我们的工作有三个重要区别。首先，他们研究的是与稀疏主成分分析相关、但在性质上有所不同的 Wigner 模型。他们的方法不适用于稀疏主成分分析的 Wishart 模型，而后者在实践中更为自然。我们克服了这一局限，研究 Wishart 模型。需要强调的是，他们的方法不足以处理如此一般的情形，而要克服这一点绝非单纯的技术细节。另一方面，我们的方法可以轻松地重现他们的结果。其次，虽然他们概述了其下界证明的大致思路，却没有给出证明。相比之下，我们的证明完全明确。最后，他们假设输入分布的条目取值于 $\{\pm 1\}$，也就是说，他们研究的是 PCA 的 ± 1 变体。相比之下，我们研究的是更贴近现实的情形，即分布为$\mathcal{N}(0, 1)$。同样，我们的方法也可以轻松地重现他们的结果。

4.3 张量主成分分析

我们还运用这些方法，在相关的张量主成分分析（Tensor PCA）问题上取得了有力的结果。张量主成分分析最初由 [206], 提出，是将 PCA 推广到高阶张量的一种方法。形式上，给定一个形如$\lambda u^{\otimes k} + B$ 的 k 阶张量，其中 $u \in \mathbb{R}^n$ 是一个单位向量，且 $B \in \mathbb{R}^{[n]^k}$ 的条目相互独立并服从高斯分布，我们希望恢复主成分 u。这里， λ 称为信噪比。

张量主成分分析是一种极为有用的统计与计算技术，可用于利用数据的高阶矩。它最初被设想用于潜变量建模，确实也已在这一领域得到多种应用（例如 [7, 132, 6]）。在这里，我们先计算一个包含输入数据统计信息的张量，再对其进行分解，以恢复潜变量。由于这种技术用途广泛，它在机器学习领域备受关注，并应用于主题建模、视频处理、协同过滤、社区检测等（例如参见 [112, 8, 206, 7, 6, 68, 153] 及其中的参考文献。）

与稀疏主成分分析类似，张量主成分分析也引起了学界的广泛关注，研究者既关注相关算法（例如 [11, 32, 108, 106, 206, 242, 233, 130, 5]），也关注其可近似性与困难性（例如 [169, 146, 38, 104], 在陈述我们的主要结果后，我们将提供更详细的概述）。值得注意的是，许多困难性结果都是有条件的，也就是说，它们依赖于各种猜想，有时甚至比 P $\neq$ NP 更强。此外，统计学界也对此广泛关注，例如 [115, 189, 151, 46, 47],，原因在于它与随机矩阵理论和统计物理之间存在引人入胜的联系。

在这项工作中，我们研究了亚指数时间平方和算法在张量主成分分析中的表现。我们的主要结果非正式地陈述如下，并在定理 4.3.2 中给出了正式表述。

定理 4.3.1。 对于张量主成分分析，亚指数时间平方和算法无法恢复

当信噪比为 $\lambda \ll n^{\frac{k}{4}}$ 时的主成分。

特别地，这解决了相关工作 [108, 32, 104, 110]提出的一个未解问题。

让我们将这一定理形式化。回顾一下，我们给定了一个形如$A = \lambda u^{\otimes k} + B$ 的 k 阶 A 张量，其中 $u \in \mathbb{R}^n$ 是一个单位向量， $B \in \mathbb{R}^{[n]^k}$ 具有相互独立的高斯条目；我们希望恢复主成分 u。张量主成分分析可以改写为以下程序

$$\text{maximize } \langle A, x^{\otimes k} \rangle = \langle A, \underbrace{x \otimes \ldots \otimes x}_{k \text{ times}} \rangle \text{ such that } \sum_{i=1}^{n} x_i^2 = 1$$

其中程序变量为 x_1、.....、 x_n。随后，主成分 u 将只是返回的解 x。与稀疏主成分分析一样，我们指出，出于技术原因，在下界中我们只会使单位向量条件满足 $o(1)$误差，而精确满足该条件则留待未来工作。对于这个问题，我们将再次考虑亚指数时间平方和算法，特别是 n^{ε} 度平方和算法。这是亚指数时间的，因为输入规模为 $n^{O(1)}$。

接下来我们证明，如果信噪比 λ 低于某个阈值，那么针对未加尖峰输入 $A \sim \mathcal{N}(0, I_{[n]^k})$ 的亚指数时间 SoS 算法的最优值将接近λ；而当 $A = \lambda u^{\otimes k} + B$， $B \sim \mathcal{N}(0, I_{[n]^k})$ 且 $x = u$ 时，这也是加尖峰情形下的最优值。换言之，SoS 无法区分未加尖峰分布与加尖峰分布，因此也无法恢复主成分 u。

定理 4.3.2。设 $k \geq 2$ 为一个整数。对于所有足够小的 $\varepsilon > 0$，若 $\lambda \leq n^{\frac{k}{4} - \varepsilon}$，则对于某个绝对常数 $C > 0$，以高概率，在随机张量 $A \sim \mathcal{N}(0, I_{[n]^k})$ 上，张量主成分分析的度为 $n^{C\varepsilon}$ 的亚指数时间 *SoS* 算法的最优值至少为$\lambda - o(1)$。

因此，亚指数时间的平方和方法无法证明，对于采样得到的随机张量 A

自 $\mathcal{N}(0, I_{[n]^k})$，不存在单位向量 u 使得 $\langle A, u \otimes \ldots \otimes u\underbrace{k \text{ ti} \quad \text{mes}}\rangle \approx \lambda$。

我们再次指出，当张量 A 确实从尖峰模型$A = \lambda u^{\otimes k} + B$中采样得到时，平方和程序的最优值在 $x = u$时约为 λ。因此，这表明亚指数时间平方和算法无法解决张量主成分分析。

非正式地说，该定理表明，当信噪比 $\lambda \ll n^{\frac{k}{4}}$时，平方和算法无法解决张量主成分分析，正如定理 4.3.1 中所述。

为展示我们关于张量主成分分析的结果，我们采用第 3.3.1 节中的策略，在该策略中使用以下分布。设 $k \geq 2$ 为一个整数。

- 随机分布 ν：从 $\mathcal{N}$中采样 A（0，$I_{[n]^k}$）。
- 植入分布 μ：设 λ，$\Delta > 0$。从 $\{-\frac{1}{\sqrt{\Delta n}}, 0, \frac{1}{\sqrt{\Delta n}}\}^n$ 中采样 u，其中各取值的概率分别为 $\frac{\Delta}{2}$，$1-\Delta$，$\frac{\Delta}{2}$。然后从$\mathcal{N}(0, I_{[n]^k})$中采样 B。令 $A = B + \lambda u^{\otimes k}$。

在第 6.3 节中，我们应用伪校准，并证明以下定理。

定理 4.3.3。令 $k \geq 2$ 为一个整数。存在常数 C、$C_\Delta > 0$，使得对于所有充分小的常数 $\varepsilon > 0$，如果 $\lambda \leq n^{\frac{k}{4}-\varepsilon}$且 $\Delta = n^{-C_\Delta \varepsilon}$，那么以高概率，由伪校准得到的、针对 $n^{C\varepsilon}$ 度平方和的平方和解是可行的。

该定理也可以自然地解释为张量主成分分析认证问题的平方和下界。我们将在第 4.5 节中给出证明概要。由此定理，定理 4.3.2 可作为推论得到。

先前工作 关于张量主成分分析的算法已在文献 [11, 32, 108,106, 206, 242, 233, 130, 5] 中得到研究。文献 [32] 证明，度为 q 的平方和算法可以为张量主成分分析问题给出 $2^{O(k)}(n \cdot \text{polylog}(n^{k/4}))$ 的上界。当 $q = n^\varepsilon$ 时，这给出的上界为 $n^{\frac{k}{4}-O(\varepsilon)}$。因此，我们的结果是紧的，也让我们得以深入了解张量主成分分析的计算阈值。

$$q^{k/4-1/2}$$

张量主成分分析的下界已通过多种形式得到研究，包括统计查询下界 [39, 69], 、从猜想的困难问题出发的归约 [241, 38], 、基于低次猜想的下界 [104, 110, 146], 、基于景观行为的证据 [12, 169], 等等。与许多依赖各种猜想的相关工作相比，我们指出，我们的下界是无条件的，不依赖任何猜想。

在 [104], 中，与稀疏主成分分析类似，他们针对张量主成分分析的另一种变体给出了一个类似的定理。然而，他们没有给出证明，而我们给出了明确的证明。具体而言，他们对张量主成分分析的 ± 1 变体只陈述了结果而未给出证明，而我们研究的是更现实的情形，其中分布为 $\mathcal{N}(0, 1)$。我们指出，他们的技术无法得到我们的结果；但另一方面，我们的技术可以得到他们的结果。

4.4 植入的略微更稠密子图

在植入稠密子图问题中，给定一个随机图 G ，其中植入了一个大小为 k 的稠密子图，我们需要找出这个植入的稠密子图。这是 k-团问题的一个自然推广 [121] ，多年来一直受到大量研究（例如 [74, 73, 126, 30, 31, 36, 160]）。在本文中，我们研究植入稠密子图问题的以下认证变体。

给定从 *Erd˝os-R´enyi* 模型 $G(n, 12)$ 中采样得到的随机图 G ，证明其最稠密子图在 k 个顶点上的边密度上界。

我们使用第 3.3.1 节中的策略，为这个问题证明一个高次数的平方和下界。具体而言，我们使用以下分布。

- 随机分布 ν：从 $G(n, 12)$ 中采样 G –
- 植入分布 μ：令 k 为一个整数，并令 $p > \frac{1}{2}$。从$G(n, 12)$中采样一个图 G' 。从顶点中随机选取一个子集 S，其中每个顶点被独立选中的概率为 $\frac{k}{n}$。对于 S 中的所有顶点对 $\bar{i}, j$，重新随机化

–

边 (i, j)，其中 (i, j) 属于该图的概率现在为 p。令 G 为所得图。

在第6.2节中，我们通过伪校准计算所得到的候选平方和解。我们的主定理如下，随后将在第 4.5 节给出证明概要。

定理 4.4.1。设 $C_p > 0$。存在一个常数 $C > 0$，使得对于所有充分小的常数 $\varepsilon > 0$，如果 $k \leq n^{\frac{1}{2}-\varepsilon}$ 且 $p = \frac{1}{2} + \frac{n^{-C_p\varepsilon}}{2}$，那么以高概率，伪校准针对度为 $n^{C\varepsilon}$ 的平方和所给出的候选解是可行的。

相关工作 对于随机分布和植入分布的许多不同参数范围（例如，在 $G_{n,p}$ 中植入 $G_{k,q}$，其中 $p < q$ 为常数），并且当 $k = o(\sqrt{n})$ 时，植入稠密子图问题的较易判别版本的困难性，此前已在文献中被提出为正式猜想（通常称为 PDS 猜想）（例如参见 [95, 48, 40, 41]）。这也促成了对其他问题的许多归约 [37]，尽管尚不清楚这些归约能否在平方和框架中实现，同时不损失参数依赖性。

在我们的情形中，我们考虑轻度植入的更稠密子图版本：对于 $k \leq n^{\frac{1}{2}-\varepsilon}$，我们植入一个密度为 $\frac{1}{2} + n^{-O(1)(\varepsilon)}$ 的子图，即 $p = \frac{1}{2}$，$q = \frac{1}{2} + n^{-O(1)(\varepsilon)}$。人们普遍认为，这需要亚指数时间。我们的工作通过针对强大的 SoS 层级给出无条件下界，为此提供了有力证据；即使我们考虑 $n^{O(\varepsilon)}$ 层，这也对应于 $n^{n^{O(\varepsilon)}}$ 的运行时间！我们预计，这将使该问题成为归约的自然起点，用于证明各种问题具有亚指数时间困难性。

在 SoS 文献中，[20] 表明，对于 $k \leq n^{\frac{1}{2}-\varepsilon}$ 和某个常数 $\varepsilon > 0$，度为 $o(\log n)$ 的平方和无法区分从 $G(n, 1/2)$ 中采样的完全随机图与含有一个植入的 k-团的随机图。这意味着，如果 $k \leq n^{\frac{1}{2}-\varepsilon}$，度为 $o(\log n)$ 的 SoS 无法证明最稠密 k-子图的边密度优于 1。

在定理 4.4.1 中，我们证明，对于 $k \leq n^{\frac{1}{2}-\varepsilon}$和某个常数 $\varepsilon > 0$，度为 $n^{\Omega(\varepsilon)}$的 SoS 无法证明边密度优于 $\frac{1}{2} + n^{O(\varepsilon)}$。在我们的设定中，SoS 的次数为 $n^{\Omega(\varepsilon)}$，远高于此前已知结果所用的度数 $o(\log n)$。据我们所知，这是首个证明该问题存在如此高次数下界的结果。

我们注意到，当取 $k = n^{\frac{1}{2}-\varepsilon}$ 时，最稠密的 k-子图的真实边密度为 $\frac{1}{2} + \sqrt{\log(n/k)}$ $\sqrt{k}$ $+ \frac{1}{\sqrt{k}}$ $\approx \frac{1}{2} + n^{1/4-\varepsilon/2}$，正如 [83, 推论 2] 所示；然而，根据定理 4.4.1，SoS 最优值却可以达到 $\frac{1}{2} + \frac{1}{n^{\varepsilon}}$。这凸显了最优值之间的显著差异。

4.5 我们的方法

在本节中，我们简要介绍如何证明定理 4.2.3、定理 4.3.3 和定理 4.4.1。我们自然地从伪校准开始，除了正定性之外，所有约束都可以通过构造轻松证明其成立。为了证明正定性，进而得到下界，我们将基本应用一个称为技术工具的一般元定理。该技术工具使我们能够为某些类型的“带噪”问题证明 SoS 下界。

在本文中，我们陈述并使用这一技术工具；其证明可见于最初提出该工具的原始著作 [191]。为了证明半正定性，这一技术工具从矩矩阵 Λ 构造出某些系数矩阵，并给出关于这些系数矩阵的条件；这些条件足以保证 Λ 以高概率为半正定。该技术工具涉及的一些思想，是对用于证明植入团平方和下界 [20]的技术的推广。本节将非正式地概述这一技术工具，并说明其中一些条件的由来。

形状和图矩阵 我们首先介绍形状和图矩阵，它们最初由 [20, 163]提出，后来在 [1]中得到推广。它们将非常便于

我们的分析。

形状 α 是包含有关顶点的额外信息的图。对于每个形状 α，都有一个矩阵值函数 M_α （即其元素取决于输入的矩阵），我们称之为图矩阵。图矩阵类似于傅里叶基，但适用于呈现某种对称性的矩阵值函数。在我们的设定中， Λ将是这样的矩阵值函数，因此我们可以将 Λ 分解为图矩阵 $\Lambda = \sum_{\text{形状 } \alpha} \lambda_\alpha M_\alpha$的线性组合。

形状和图矩阵具有若干使其非常便于处理的性质。首先， $\|M_\alpha\|$ 可以以高概率根据形状 α的简单组合性质进行界定。其次，如果两个形状 α 和 β 以某种方式相匹配，我们可以将它们组合成一个更大的形状 $\alpha \circ \beta$。我们将这一操作称为形状组合。第三，每个形状 α 都可以规范分解为三个形状，即 α的左部、中部和右部，我们将其称为 σ、 τ和 σ'^T。对于这一规范分解，我们有$\alpha = \sigma \circ \tau \circ \sigma'^T$和 $M_\alpha \approx M_\sigma M_\tau M_{\sigma'^T}$。这一分解对于我们的分析至关重要。

平方和下界的一般框架现在我们简要概述这套机制的策略。

1. 将矩矩阵 Λ 分解为图矩阵 M_α的线性组合 $\Lambda = \sum_{\text{形状 } \alpha} \lambda_\alpha M_\alpha$ 。2. 对于每个形状 α，将 α 分解为左部分 σ、中间部分 τ和右部分σ'^T。3. 根据系数 λ_α 以及形状 α 分解为左部分、中间部分和右部分的方式，构造系数矩阵 H_{Id_U} 和 H_τ。4. 根据系数矩阵 H_{Id_U} 和 H_τ，得到 Λ的近似半正定分解。

5. 证明误差项（我们称之为交项）可以由 Λ的近似半正定分解加以界定。

这与 [20] 的工作大体相似，他们证明了植入团问题的 SoS 下界。

这些技术工具表明，只需将分析归纳为系数矩阵上的三个条件，这项分析便能成功。使用这些技术工具证明 SoS 下界的大致蓝图如下。

1. 构造一个候选矩矩阵 Λ。

2. 将矩矩阵 Λ 分解为图矩阵 M_α 的线性组合 $\Lambda = \sum_{\text{形状 }\alpha} \lambda_\alpha M_\alpha$ （类似于傅里叶分解），并找出相应的系数矩阵。

3. 验证系数矩阵满足所需的条件。

条件背后直觉的概述

现在，我们将说明并概述在技术工具中提出的条件。

给出近似半正定因式分解 如上所述，我们将矩矩阵 Λ 分解为图矩阵 M_α的线性组合 $\Lambda = \sum_{\text{形状 }\alpha} \lambda_\alpha M_\alpha$。然后，我们将每个矩阵 α 分解为左部、中部和右部 σ、 τ 和 σ'^T。现在我们有

$$\Lambda = \sum_{\alpha=\sigma\circ\tau\circ\sigma'^T} \lambda_{\sigma\circ\tau\circ\sigma'^T} M_{\sigma\circ\tau\circ\sigma'^T}$$

我们首先考虑这些项 $\sum_{\sigma、\sigma'} \lambda_{\sigma\circ\sigma'^T} M_{\sigma\circ\sigma'^T} \approx \sum_{\sigma、\sigma'} \lambda_{\sigma\circ\sigma'^T} M_\sigma M_{\sigma'^T}$ 其中 τ 对应一个单位矩阵，可以忽略不计。

如果存在实数 v_σ，使得对所有左形状 σ 都有 $\lambda_{\sigma\circ\sigma'^T} = v_\sigma v_{\sigma'}$，那么我们将有

$$\sum_{\sigma,\sigma'} \lambda_{\sigma\circ\sigma'^T} M_\sigma M_{\sigma'^T} = \sum_{\sigma,\sigma'} v_\sigma v_{\sigma'} M_\sigma M_{\sigma'^T} = (\sum_\sigma v_\sigma M_\sigma)(\sum_\sigma v_\sigma M_\sigma)^T \succeq 0$$

这表明这些项的贡献是半正定的。事实上，植入团分析中确实如此。不过，一般情况下未必成立。为了解决这个问题，我们注意到，v_σ 的存在可以放宽如下：设 H为一个矩阵，其行和列由左形状 σ 索引，并满足 $H(\sigma, \sigma') = \lambda_{\sigma\circ\sigma'^T}$。经过缩放后，$H$ 将成为我们的系数矩阵之一。如果 H 是半正定的，那么这些项的贡献也将是半正定的。事实上，这将是主定理中的半正定质量条件。

处理具有非平凡中间部分的项遗憾的是，我们也有一些项$\lambda_{\sigma\circ\tau\circ\sigma'^T} M_{\sigma\circ\tau\circ\sigma'^T}$，其中 τ是非平凡的。他们的策略是将这些项归入其他项。为简单起见，我们将说明如何处理其中一项。以下不等式是一个出发点。对于左形状 σ、中间形状 τ、右形状 σ'^T以及实数 a, b,

$$(aM_\sigma - bM_{\sigma'} M_{\tau^T})(aM_\sigma - bM_{\sigma'} M_{\tau^T})^T \succeq 0$$

重新排列后可得

$$\begin{aligned} ab(M_\sigma M_\tau M_{\sigma'^T} + (M_\sigma M_\tau M_{\sigma'^T})^T) &\preceq a^2 M_\sigma M_{\sigma^T} + b^2 M_{\sigma'} M_{\tau^T} M_\tau M_{\sigma'^T} \\ &\preceq a^2 M_\sigma M_{\sigma^T} + b^2 \|M_\tau\|^2 M_{\sigma'} M_{\sigma'^T} \end{aligned}$$

如果 $\lambda^2_{\sigma\circ\tau\circ\sigma'^T} \|M_\tau\|^2 \leq \lambda_{\sigma\circ\sigma^T} \lambda_{\sigma'\circ\sigma'^T}$，那么我们可以选择 a, b使得 $a^2 \leq \lambda_{\sigma\circ\sigma^T}$,

$b^2 \|M_\tau\|^2 \leq \lambda_{\sigma' \circ \sigma'^T}$且 $ab = \lambda_{\sigma \circ \tau \circ \sigma'^T}$。这将近似意味着

$$\lambda_{\sigma \circ \tau \circ \sigma'^T}(M_{\sigma \circ \tau \circ \sigma'^T} + M^T_{\sigma \circ \tau \circ \sigma'^T}) \preceq \lambda_{\sigma \circ \sigma^T} M_{\sigma \circ \sigma^T} + \lambda_{\sigma' \circ \sigma'^T} M_{\sigma' \circ \sigma'^T}$$

这将为我们提供一种方法，用平凡中间部分的项来抵消具有非平凡中间部分的项。

虽然我们可以尝试逐项应用这个不等式，但它还不足以推出主要的技术工具结果。相反，他们将这个不等式推广，使其适用于固定 τ 的整个形状集合 σ，σ'。这将引出中间形状界条件。

处理交项上述计算中有一个重要的技术细节。每当我们将 α 分解为左部、中间部分和右部 σ、τ 和 σ'^T 时，$M_\sigma M_\tau M_{\sigma'^T}$ 仅近似等于 $M_\alpha = M_{\sigma \circ \tau \circ \sigma'^T}$。分析中必须仔细处理所有其他误差项。我们称这些项为交项。

我们利用这样一个事实：这些交项本身也是图矩阵。因此，我们递归地将它们分解为 $\sigma_2 \circ \tau_2 \circ \sigma_2'^T$，并应用前述思路。为系统地实施这一方法，这套技术工具采用了若干概念，例如交模式和广义交权衡引理。妥善处理交项是他们工作中技术性最强的部分之一。这项分析将引出交项界条件。

应用技术工具为了将这些技术工具应用于我们关注的问题，我们需要验证系数应满足的谱条件，然后便可使用主定理。植入的稍密子图这一应用很直接，也很适合作为热身，帮助我们理解这些技术工具。在张量主成分分析和稀疏主成分分析的应用中，非零系数对应的图矩阵形状具有良好的结构性质，这些性质对我们的分析至关重要。我们利用这种结构，并

采用新颖的收费论证来验证这些技术工具的条件。本文将完成这项工作。

4.6 认证问题的平方和下界相关工作

[137]证明，对于谓词的解服从平衡两两独立分布的随机约束满足问题（约束满足问题），以高概率，需要 $\Omega(n)$ 度的平方和方法，才能证明这些约束满足问题无解。尽管他们没有这样表述，但 [137] 使用的伪期望值也可以通过伪校准 [199, 42]推导出来。证明矩矩阵为半正定的分析方法则截然不同。能否将这些分析统一起来，是一个有趣的问题。

[165]证明，在适当条件下，可以将 2 度平方和解提升为 4 次平方和解，并利用这一点，为平均情况d-正则最大割问题和 Sherrington-Kirkpatrick问题得到了 4次平方和下界。他们的构造受到伪校准的启发，分析也采用了图矩阵。

[143] 最近提出了一种将二阶 SoS 下界提升到更高层级的技术，并将其用于构造 Sherrington-Kirkpatrick 问题的 6 次下界。有趣的是，他们的构造并非通过伪校准实现。

4.7 证明的组织

我们在第5章中证明 Sherrington-Kirkpatrick 下界，即定理 4.1.2。对于植入的稍密子图、张量主成分分析和稀疏主成分分析的证明，即定理 4.4.1、定理 4.3.3 和定理 4.2.3，分别分布在第6章和第7章中。后面这些证明分为定性版本和定量版本。定性定理陈述

阐明了我们所证明的不等式的本质，并用于展示我们所需界的主要形式，同时避免陷入细节。另一方面，定量定理在定性对应定理的基础上，陈述所需的精确界。在第6章中，我们介绍技术工具，并在第6.2节、第 6.3 节和第 6.4 节中，分别对植入的稍密子图、张量主成分分析和稀疏主成分分析定性地验证这些技术工具的条件。虽然这些节仅验证了定性条件，但其中的结果是精确的，并将在第7章中再次使用；在第7章中，我们将在第7.1节、第 7.2 节和第 7.3 节中完整验证这些技术工具的条件。

第5章

Sherrington-Kirkpatrick 哈密顿量

本章将正式证明 Sherrington-Kirkpatrick 哈密顿量认证问题的平方和下界，特别是定理 4.1.2。本章内容改编自 [87], ，该成果最初发表于此处。本章与原作的主要区别在于，我们省略了关于精确满足约束的技术部分。

5.1 技术预备知识

本节先记录正式的问题陈述，然后定义并讨论平方和下界中的一个主要对象：图矩阵。

对于向量或变量 $v \in \mathbb{R}^n$ 和 $I \subseteq [n]$，我们使用记号 $v^I := \prod_{i \in I} v_i$。当某个陈述以高概率（w.h.p.）成立时，意指其以概率 $1 - o_n(1)$ 成立。特别是，不要求 n 很小。

5.1.1 问题陈述

我们引入分布 $\mathcal{D}$ 上的植入仿射平面问题。

定义 5.1.1 (植入仿射平面（PAP）问题**)**。给定 d_1、.. .、$d_m \sim \mathcal{D}$，其中每个 d_u 都是 $\mathbb{R}^n$ 中的一个向量，判断是否存在 $v \in \{\pm \frac{1}{\sqrt{n}}\}^n$，使得

$$\langle v, d_u \rangle^2 = 1,$$

对于每个 $u \in [m]$ 都成立。

我们的结果适用于高斯情形 $\mathcal{D} = \mathcal{N}(0, I)$，也适用于布尔情形，其中 $\mathcal{D}$是从 $\{\pm 1\}^n$ 中均匀采样的；不过，我们在第8.2.2节中猜想，类似的平方和

界在 $\mathcal{D}$ 满足更一般条件时也成立。

请注意，在这两种设定中，解向量 v 被限制为布尔向量（也就是说，其各个分量只能是 $\frac{1}{\sqrt{n}}$ 或 $\frac{-1}{\sqrt{n}}$），而针对这一受限版本问题的平方和下界，比 v 可以是来自 $\mathbb{R}^n$的任意向量时的下界更强。

正如我们在第4章中所见，Sherrington–Kirkpatrick（SK）问题源自统计物理学[215]中的自旋玻璃模型。

定义 5.1.2（Sherrington-Kirkpatrick 问题）。给定 W ~高斯正交系综(n)，计算

$$\mathrm{OPT}(W) := \max_{x\in\{\pm 1\}^n} x^\intercal W x.$$

植入布尔向量问题由 Mohanty–Raghavendra–Xu[165], 提出，当时被称为“随机子空间中的布尔向量”。

定义 5.1.3（植入布尔向量问题）。给定一个均匀随机的p维子空间 V，它是 $\mathbb{R}^n$ 的子空间，输入形式为投影算子 Π_V，投影到 V上，计算

$$\mathrm{OPT}(V) := \frac{1}{n}\max_{b\in\{\pm 1\}^n} b^\intercal \Pi_V b.$$

5.1.2 图矩阵

为了研究 $\mathcal{M}$，我们使用图矩阵框架对其进行分解。图矩阵最初是在植入团问题的背景下发展起来的，是一种随机矩阵，其元素是底层随机对象的对称函数——在我们的情形中，该随机对象是向量集合 d_1、……、d_m。我们采用 [1] 中给出的一般表述和结果。对我们的目的而言，以下定义已经足够。

我们研究的图有两种顶点类型：圆形 和正方形 。令 $\mathcal{C}_m$为由 m 个圆形组成的集合，圆形的标签从 1 到 m，记作 $_1$，$_2$，……，$_m$；再令 $\mathcal{S}_n$ 为由 n 个正方形组成的集合，正方形的标签从 1 到 n，记作 $_1$，$_2$，……，$_n$。我们将研究

二分图，其中圆形和正方形之间由边连接，边上带有正整数标签。当不存在重边时（即图为简单图），这类图与向量 d_u上的傅里叶特征一一对应。连接 u 和 i 、标签为 l 的边表示 $h_l(d_{u,\ i})$，其中 $\{h_k\}$ 是傅里叶基（例如厄米特多项式）。

○ □

$$\text{带标签边的简单图} \quad \Longleftrightarrow \quad \prod_{\substack{u\in\mathcal{C}_m,\\ i\in\mathcal{S}_n}} h_{l(u,i)}(d_{u,\ i})$$

图 5.1 展示了一个以带标签边的图表示的傅里叶多项式示例。未标记的边默认标记为 1。

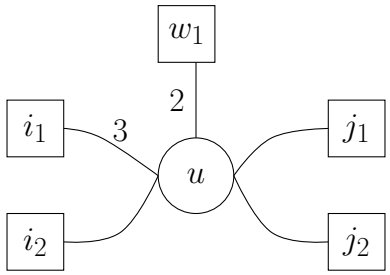


图 5.1：傅里叶多项式 $h_3(d_{u,\ i_1})h_1(d_{u,\ i_2})h_2(d_{u,\ w_1})h_1(d_{u,\ j_1})h_1(d_{u,\ j_2})$ 的图表示。

定义顶点 v 的度，记为 $\deg(v)$，为与 v 相邻的标签之和；而 $|E|$ 为所有标签之和。为了直观理解，通常只需处理简单图；在这种情况下，这些量可视为隐式多重图中的边重数。

定义 5.1。 **4** （正规）. 如果一个边标记图没有重边，我们就称它为正规边标记图 .

这些定义允许存在“不正规”的边标记多重图，这类图可简化图矩阵的乘法（第 5.4.2 节）。

定义 5.1.5 （矩阵指标）. 矩阵指标是由 $\mathcal{C}_m \cup \mathcal{S}_n$ 中的元素构成的集合 A 。

我们令 $A(□\,i)$ 或 $A(○\,u)$ 为 0 或 1，以表示该顶点是否属于 A。

定义 5.1.6 (带状图**)**。带状图是一个无向边标记图 R ，定义为$R = (V(R)\ E(R)\ A_R,\ B_R)$，其中 $V(R) \subseteq \mathcal{C}_m \cup \mathcal{S}_n$和 A_R、B_R是两个矩阵指标（可能不互不相交），对应两个特定顶点集 A_R、$B_R \subseteq V(R)$。此外， $E(R)$中的所有边都连接方形和圆形。

我们分别将 A_R 和 B_R 视为 R的“左侧”和“右侧”。我们还定义“中间顶点”集合 C_R: $= V(R) \setminus (A_R \cup B_R)$。如果 $e \notin E(R)$，那么我们将其标签定义为 $l(e) = 0$。我们还滥用记号，用 $l(i, u)$ 表示 $l(\{\boxed{i}, \textcircled{u}\})$。

□ ○

与上图类似，每个带状图对应一个傅里叶多项式。这个傅里叶多项式位于矩阵 M_R 的一个矩阵项中。在下面的定义中， $h_k(x)$ 是对应于各自情形的傅里叶基。在高斯情形中，它们是（未归一化的）厄米特多项式；在布尔情形中，它们就是奇偶函数，表示为

$$h_0(x) = 1, \qquad h_1(x) = x, \qquad h_k(x) = 0 \ \ (k \geq 2)$$

定义 5.1.7 (带状图的矩阵**)**。 矩阵 M_R 的行和列由 $\mathcal{C}_m \cup \mathcal{S}_n$的子集索引，其唯一的非零项定义为

$$M_R[I, J] = \begin{cases} \prod\limits_{\substack{e \in E(R), \\ e = \{\boxed{i}, \textcircled{u}\}}} h_{l(e)}(d_{u,i}) & I = A_R, J = B_R \\ \\ 0 & Otherwise \end{cases}$$

接下来，我们将描述带状图的形状。它本质上是指忘记所有顶点标签、仅保留图结构和指定的顶点集合后得到的带状图。

定义 5.1.8 (指标形状**)**。 指标形状是由形式变量组成的集合 U 。此外，每个变量都被标记为“圆形”或“方形”。

对于 $_i$ 或 $_u$ 是否属于 U，我们分别令 $U(\ _i)$ 和 $U(\ _u)$ 为 0 或 1。

定义 5.1.9 (形状**)。** 形状是一个无向边标记图 α，由 $\alpha = (V(\alpha)\ E(\alpha)\ U_\alpha, V_\alpha)$ 给出，其中 $V(\alpha)$ 是一个形式变量集，每个变量都标记为“圆形”或“方形”。U_α 和 V_α 是指标形状（可能含有相同的变量），使得 $U_\alpha, V_\alpha \subseteq V(\alpha)$。边集 $E(\alpha)$ 中只能包含连接圆形变量与方形变量的边。

我们还用 $W_\alpha := V(\alpha) \setminus (U_\alpha \cup V_\alpha)$ 表示该形状的“中间顶点”。

备注 5.1.10。 我们将滥用记号，用 $_i, _j, _u, _v, \ldots$ 同时表示带状图的顶点和形状的顶点。如果它们是带状图顶点，则这些顶点是 $\mathcal{C}_m \cup \mathcal{S}_n$ 的元素；如果它们是形状顶点，则对应于具有相应类型的形式变量。

定义 5.1.11 (平凡形状**)。** 若形状 α 满足 $U_\alpha = V_\alpha$、$W_\alpha = \emptyset$ 和 $E(\alpha)$，$= \emptyset$，则称其为平凡形状。

定义 5.1.12 (形状的转置**)。** 对于形状 $\alpha = (V(\alpha)\ E(\alpha)\ U_\alpha, V_\alpha)$，其转置定义为形状 $\alpha^\intercal = (V(\alpha)\ E(\alpha)\ V_\alpha, U_\alpha)$。

对于形状 α 和单射 $\sigma\colon V(\alpha) \to \mathcal{C}_m \cup \mathcal{S}_n$，我们以自然的方式将实现 $\sigma(\alpha)$ 定义为一个带状图：使用映射 σ 为所有变量添加标签。我们还要求 σ 保持类型，即它将方形变量映射为 $\mathcal{S}_n$，将圆形变量映射为 $\mathcal{C}_m$。由此得到的带状图称为形状为 α 的带状图；注意，这将所有带状图按其形状[1][2]进行分拆。

最后，给定一个形状 α，图矩阵 M_α 由所有形状为 α 的带状图的傅里叶特征组成。

1. 按形状的等式关系取分拆：若两个形状的变量之间存在保持类型的双射，且该双射将一个形状变为另一个形状，则这两个形状相等。下文对形状集合进行运算时，默认每个不同的形状只计一次。2. 请注意，根据我们的定义，一个形状的两个实现可能对应同一个带状图。

定义 5.1.13 (图矩阵**)。** 给定一个形状 $\alpha = (V(\alpha)\ E(\alpha)\ U_\alpha, V_\alpha)$，图矩阵 M_α 为

$$M_\alpha = \sum_{R \text{ is a ribbon of shape } \alpha} M_R$$

PAP 的矩矩阵将被定义为使用图矩阵 M_α，其左侧和右侧都只有方形顶点，没有圆形顶点。然而，在分析过程中，我们会对左侧或右侧带有圆形顶点的图矩阵进行分解和相乘。

5.1.3 范数界

与仅含单一类型顶点的图矩阵的范数界类似（见第 2 章），在我们的设定中，图矩阵的谱范数由图相对简单的组合性质决定，精确到对数因子。对于子集 $S \subseteq \mathcal{C}_m \cup \mathcal{S}_n$，我们定义权重 $w(S) := (\#\text{ 圆形顶点 in } S) \cdot \log n(m) + (\#\text{ 方形顶点 in } S)$。注意$n^{w(S)} = m^{\#\text{ 圆形顶点 in } S} \cdot n^{\#\text{ 方形顶点 in } S}$。

定义 5.1.14 (最小顶点分隔集**)。** 对于形状 α，集合 $S_{\min}$ 是一个最小顶点分隔集，当且仅当所有从 U_α 到 V_α 的路径都经过 $S_{\min}$，并且在所有此类分隔集中，$w(S_{\min})$ 取最小值。

令 W_{iso} 表示 W_α中的孤立顶点集。那么，对于所有形状 α，以下范数界以高概率成立（形式化陈述见第 5.6.1 节）：

$$\|M_\alpha\| \leq \widetilde{\mathrm{O}}\left(n^{\frac{w(V(\alpha)) - w(S_{\min}) + w(W_{iso})}{2}}\right)$$

事实上，我们的证明对输入 $d_1, \ldots, d_m$ 唯一要求的概率性质，是上述范数界对分析中出现的所有形状都成立。因此，以下我们假设引理 5.6.3（高斯情形）和引理 5.6.1（布尔情形）中的范数界成立。

5.2 证明策略

现在我们更详细地说明植入仿射平面下界的思路。为了证明定理 4.1.4，固定一个常数 $\varepsilon > 0$ 和一个满足 $n \le m \le n^{3/2-\varepsilon}$ 的随机实例 $d_1, \dots, d_m$。我们将构造一个伪期望算子，并证明它以高概率在度 $D = 2 \cdot n^{\delta}$ 范围内是半正定的。

我们首先进行伪校准，以获得一个伪期望算子 $\mathbb{E}$。算子 $\mathbb{E}$将精确满足“布尔性”约束“$v_i^2 = \frac{1}{n}$”，但由于伪校准中的截断误差，它可能无法精确满足约束“$\langle v, d_u \rangle^2 = 1$”。令截断参数 n^{τ} 大于平方和解的度 D，即 $\delta \ll \tau$，则截断误差足够小，我们可以将 $\mathbb{E}$ 舍入为一个相近的 $\mathbb{E}'$，使其精确满足这些约束。具体而言，我们将 $\widetilde{\mathbb{E}} \in \mathbb{R}^{\binom{[n]}{\le D}}$ 视为一个向量，并将这些约束表示为一个矩阵 Q，使得 $\mathbb{E}$ 满足这些约束当且仅当它属于 Q的零空间。然后，$\mathbb{E}'$ 便是将 $\mathbb{E}$ 投影到 Null(Q) 上的结果。最终，我们构造出一个精确满足这些约束的矩矩阵 $\mathcal{M}_{fix} = \mathcal{M} + \mathcal{E}$，且$\|\mathcal{E}\|$ 极小。为简洁起见，本文省略这一技术细节，详情请见 [87] 。

进行伪校准后，在这两种情形下，我们基本上都会得到图矩阵分解

$$\mathcal{M} = \sum_{\text{形状 } \alpha} \lambda_\alpha M_\alpha = \sum_{\substack{\text{形状 } \alpha: \\ \deg(\square_i)+U(\square_i)+V(\square_i) \text{ 偶}, \\ \deg(\textcircled{u}) \text{ 偶数}}} \frac{1}{n^{\frac{|U_\alpha|+|V_\alpha|}{2}}} \cdot \left(\prod_{\textcircled{u} \in V(\alpha)} h_{\text{度}(\textcircled{u})}(1) \right) \cdot \frac{M_\alpha}{n^{|E(\alpha)|/2}}$$

He在两种情形下，re $h_k(1)$ 都是取值于 1 的第 k 个厄米特多项式。

在对 $\mathcal{M}$ 的这种分解中，平凡形状将是主导项，我们将利用它们来界定其他项。回想一下，如果 $U_\alpha = V_\alpha$、$W_\alpha = \emptyset$ 以及 $E(\alpha) = \emptyset$，则形状 $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$ 是平凡的。这些形状在

$\mathcal{M}$ 主对角线的不同块上贡献缩放单位矩阵，其中平凡形状 α 贡献系数为 $n^{-|U_\alpha|}$ 的单位矩阵。图 5.2 展示了两个平凡形状。

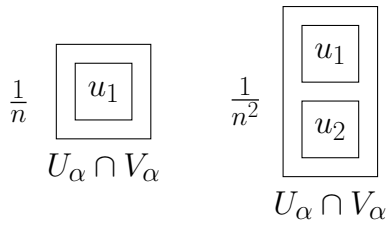


图 5.2：两个平凡形状的示例。

令 $\mathcal{M}_{\text{平凡形状}}$ 表示上述 $\mathcal{M}$ 分解中的平凡形状对角矩阵。为了证明 $\mathcal{M} \succeq 0$，我们尝试采用一个简单策略：证明所有其他项的范数都可以由这个对角矩阵 $\mathcal{M}_{\text{平凡形状}}$“承担”。对于若干种形状，这一策略确实行得通。举例来说，我们来看图 5.3 中描绘的其中一种形状 α。

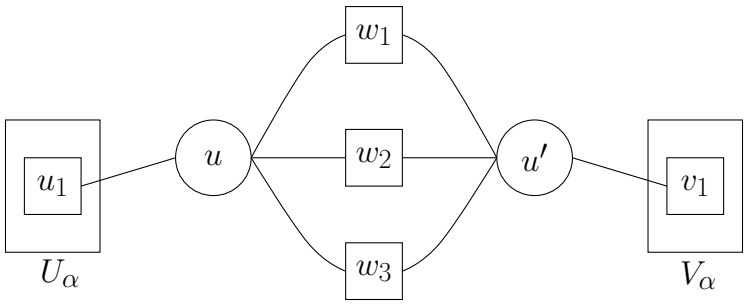


图 5.3：基本非蜘蛛形状 α 的示意图。

这个图矩阵有 $|\lambda_\alpha| = \Theta(\frac{1}{n^5})$。根据图矩阵范数界，以高概率，这个图矩阵的范数为 $\tilde{O}(n^2 m)$：有四个方形顶点和两个不属于最小顶点分隔集的圆形顶点。因此，对于这种形状 α，以高概率，$|\lambda_\alpha| \|M_\alpha\|$ 为 $\tilde{O}\left(\frac{m}{n^3}\right)$，从而 $\lambda_\alpha M_\alpha \preceq \frac{1}{n} Id$（它是恒等矩阵的倍数，出现在相应的 $\mathcal{M}_{\text{平凡形状}}$ 块中）。

遗憾的是，分解中出现的一些形状 α 太大，无法计入 $\mathcal{M}_{\text{triv}}$。这些形状具有某种子结构（实际上，就是用于投影伪期望算子的矩阵 Q 中出现的相同结构！）

其范数无法用前面的论证处理，我们将其称为蜘蛛。下图展示了这样一种蜘蛛形状（也说明了这个术语的由来）：

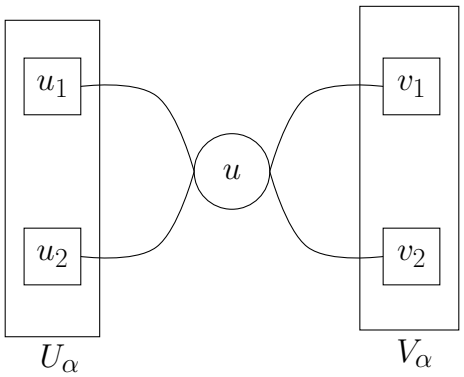


图5.4：基本蜘蛛形状的图示 α。

该图的范数 $\|\lambda_\alpha M_\alpha\|$ 为 $\widetilde{\Omega}(\frac{1}{n^2})$，通过范数界可以轻松估计得出（系数为 $\lambda_\alpha = \frac{-2}{n^4}$，最小顶点分隔集为 u，且不存在孤立顶点）。这个值太大，无法与 $\frac{1}{n^2}Id$进行界定；$\frac{1}{n^2}Id$ 是该蜘蛛的块上 M_{triv}的系数。

为了避开这个蜘蛛及其他蜘蛛，我们将范围限制在向量 $x \perp \mathrm{Null}(M)$内，并注意到该蜘蛛 α 满足 $x\ M_\alpha \approx 0_\intercal$更准确地说，考虑以下论证。考虑图5.5中的两个形状：β_1 和 β_2 （注意 β_2中的边上标有标签2）。

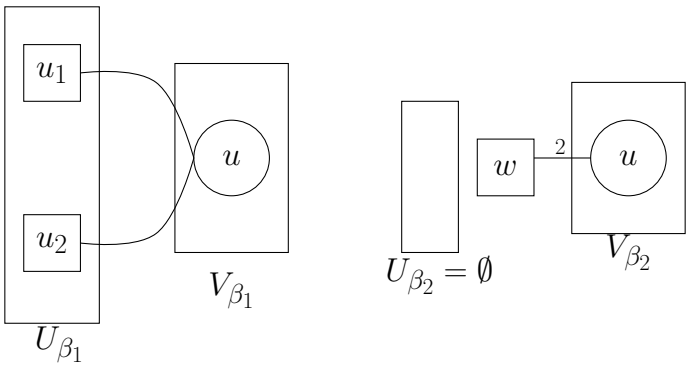


图5.5：形状 β_1 和 β_2的图示。

我们断言，矩阵 $2M_{\beta_1} + \frac{1}{n}M_{\beta_2}$ 的每一列都属于 $\mathcal{M}$ 的零空间。共有 m 个非零列，它们由对 V 的赋值进行索引，而 V 可以是单个圆形

$\bigcirc\ \bigcirc$，$_2\bigcirc$.....，$_m$。非零行是 $\emptyset$ 中的 β_2 和 $\{\boxed{i},\boxed{j}\}$，其中 $i \neq j$ 属于 β_1。固定 $I \subseteq [n]$ 后，乘积矩阵 $\mathcal{M}(2M_{\beta_1} + \frac{1}{n}M_{\beta_2})$ 的矩阵项 $(I,\ _u)$ 为

$$
\begin{aligned}
&2\sum_{i<j}\widetilde{\mathbb{E}}[v^I v_i v_j]\cdot d_{ui}d_{uj} + \frac{1}{n}\widetilde{\mathbb{E}}[v^I]\cdot\sum_i (d_{ui}^2-1)\\
&= 2\sum_{i<j}\widetilde{\mathbb{E}}[v^I v_i v_j]\cdot d_{ui}d_{uj} + \widetilde{\mathbb{E}}[v^I v_i^2]\cdot\sum_i d_{ui}^2 - \widetilde{\mathbb{E}}[v^I] \qquad (\widetilde{\mathbb{E}}\text{ satisfies }\text{“}v_i^2 = \frac{1}{n}\text{”})\\
&= \sum_{i,j}\widetilde{\mathbb{E}}[v^I v_i v_j]d_{ui}d_{uj} - \widetilde{\mathbb{E}}[v^I]\\
&= \widetilde{\mathbb{E}}[v^I(\langle v, d_u\rangle^2 - 1)]\\
&= 0 \qquad\qquad (\widetilde{\mathbb{E}}\text{ satisfies }\text{“}\langle v, d_u\rangle^2 = 1\text{”})
\end{aligned}
$$

换句话说，约束“$\langle v, d_u\rangle^2 = 1$”会产生一个形状 $2\beta_1 + \frac{1}{n}\beta_2$，它位于矩矩阵的零空间中。另一方面，我们可以近似地沿蜘蛛 α 的中心顶点进行分解，此时形状 β_1 会出现在左侧。因此

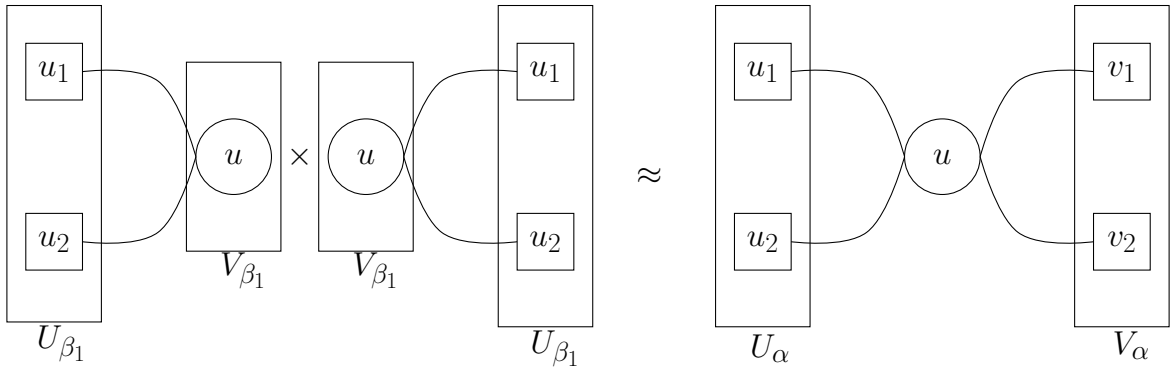


图 5.6：近似 $\beta_1 \times \beta_1 \approx \alpha$。

$M_\alpha \approx M_{\beta_1}M_{\beta_1}^\intercal \approx (M_{\beta_1} + \frac{1}{2n}M_{\beta_2})M_{\beta_1}^\intercal$。矩阵 $M_{\beta_1} + \frac{1}{2n}M_{\beta_2}$ 的列都属于 $\mathcal{M}$ 的零空间，因此对于 $x \perp \mathrm{Null}(\mathcal{M})$，我们有 $x\ M_\alpha \approx 0$。更正式地说，我们能够找到系数 c_β，使得矩阵的所有列

$$A = M_\alpha + \sum_\beta c_\beta M_\beta$$

都属于 Null($\mathcal{M}$)。接着我们观察到以下事实：

事实 5.2.1。 如果 $x \perp$ 零空间($\mathcal{M}$) 并且 $\mathcal{M}A = 0$, 那么 $x\,(AB+\mathcal{M})x = x\,(B\ A\ +\mathcal{M})x = x^{\intercal}\mathcal{M}x$。

利用这个事实，我们可以任意将 A 的倍数加到 $\mathcal{M}$ 上，而不改变 $\mathcal{M}$ 在零空间($\mathcal{M}$)$^{\perp}$ 上的作用。一个恰当的选择是减去 $\lambda_\alpha A$ ，这样就能从 $\mathcal{M}$ 中“消去”蜘蛛。对所有蜘蛛都这样做，我们得到一个矩阵，它在零空间($\mathcal{M}$)$^{\perp}$ 上的作用等价，并且由于它不含蜘蛛，其最小特征值较大，这表明 $\mathcal{M}$ 是半正定的。这里有两个问题：首先，系数 c_β 可能会影响非蜘蛛项的系数；其次，进一步的交项 M_β 本身可能是蜘蛛（尽管它们的方形顶点总是少于 α）。因此，我们必须递归地消去这些蜘蛛，直到 $\mathcal{M}$ 的分解中不再有蜘蛛。所得矩阵在非蜘蛛项上会有一些新的系数

$$\mathcal{M}' = \sum_{\text{non-spiders } \beta} \lambda'_\beta M_\beta.$$

我们必须界定系数 λ'_β 的累积量。为此，我们考察每个蜘蛛所创建的蜘蛛网 及非蜘蛛，并利用c_β 和 λ_α 的界来论证各项贡献不会失控，这通过一种利用这些图结构的有趣分摊方案实现。

5.3 伪校准

正如我们在第3章中所看到的，要将伪校准技术应用于平均情形可行性问题——在我们的例子中即 PAP 问题——需要设计一个由可行实例构成支持集的植入分布。这在第5.3.1节中完成。在第5.3.2节中，我们回顾应用伪校准的具体细节。随后，我们分别在高斯情形（第5.3.3节）和布尔情形（第5.3.4节）下进行伪校准。

5.3.1 PAP 植入分布

我们在高斯和布尔情形下，正式定义植入仿射平面问题的随机分布和植入分布。伪校准机制需要这两个（分布）族，以定义候选伪期望$\widetilde{\mathbb{E}}$算子 $\mathbb{E}$。对于高斯情形，我们有以下分布。

定义 5.3.1 (高斯 PAP 分布)。 高斯 *PAP* 分布如下。

1. （随机分布） m 独立同分布向量 $d_u \sim \mathcal{N}(0, I)$。
2. （植入分布）从 $\left\{\pm \frac{1}{\sqrt{n}}\right\}^n$ 中均匀抽取一个向量 v，以及符号$b_u \in_R \{\pm 1\}$，并从 $\mathcal{N}(0, I)$ 中抽取 m 个向量 d_u，条件为 $\langle d_u, v\rangle = b_u$。对于布尔情形，我们有以下分布。

定义 5.3.2 (布尔 PAP 分布)。 布尔 *PAP* 分布如下

1. （随机分布） m 独立同分布向量 $d_u \in_R \{-1, +1\}^n$。
2. （植入分布）从 $\left\{\pm \frac{1}{\sqrt{n}}\right\}^n$ 中均匀采样一个向量 v，以及符号$b_u \in_R \{\pm 1\}$，并从满足 $\langle d_u, v\rangle = b_u$ 条件的 $\{\pm 1\}^n$ 中抽取 m 个向量 d_u。

5.3.2 伪校准技术

我们将用简写 $\mathbb{E}_{\mathrm{ra}}$ 和 $\mathbb{E}_{\mathrm{pl}}$ 分别表示随机分布和植入分布下的期望。伪校准提供了一种构造候选伪期望算子 $\mathbb{E}$的方法。伪校准的基本思想是，对于数据 $t = t(d) = t(d_1, \ldots, d_m)$ 的每个低次测试， $\mathbb{E}_{\mathrm{ra}}\mathbb{E}f(v)$ 都应与$\mathbb{E}_{\mathrm{pl}}f(v)$ 相匹配，

$$\mathbb{E}_{\mathrm{ra}} t(d)\widetilde{\mathbb{E}}f(v) = \mathbb{E}_{\mathrm{pl}} t(d)f(v).$$

进行伪校准时，可以自由选择用于表示多项式 $f(v)$ 的“外部”基，也可以自由选择应与植入分布相一致的低次测试“内部”基。虽然我们尝试使用其他基来简化分析，但最终还是选择了标准基：每种情形下，内部基均采用傅里叶基（高斯情形下使用厄米特函数，布尔情形下使用奇偶函数），外部基则采用坐标基 v^I 。

当内部基在随机分布下是正交归一的（例如傅里叶基），伪校准条件会给出 $\mathbb{E}f(v)$ 在正交归一基下的系数公式（不过它只给出低次函数 $t(d)$ 的系数）。具体而言，令内部基以 $\alpha \in \mathcal{F}$ 为索引，则伪校准条件作为 d 的函数强制满足

$$\widetilde{\mathbb{E}} f(v) = \sum_{\substack{\alpha \in \mathcal{F}: \\ |\alpha| \leq n^\tau}} \left(\mathbb{E}_{\mathrm{pl}} t_\alpha(d) f(v) \right) t_\alpha(d).$$

这里我们用“$|\alpha| \leq n^\tau$”来描述低次测试集合。伪校准条件不会规定函数 $t_\alpha(d)$ 在 $|\alpha| > n^\tau$ 时的任何系数，一种经济的选择是将这些系数设为零。

进行伪校准时，我们的伪期望算子保证是线性的，因为上面的表达式关于 f 是线性的。它保证满足所有形如“$f(v) = 0$”的约束。它将近似满足形如“$f(v, d) = 0$”的约束，不过只能做到截断误差范围内。

事实5.3.3 (证明见 [87])。 如果 $p(v)$ 是一个在植入$\widetilde{\ }$分布上恒为零的多项式，那么 $\mathbb{E}[p]$ 就是零函数。如果 $p(v, d)$ 是一个在植入分布上恒为零的多项式，那么 $\mathbb{E}[p]$ 仅有那些大小介于 n^τ ±次数$d(p)$之间的傅里叶系数非零。

截断会在约束中引入微小误差，我们能够在 [87], 中处理这种误差

为简洁起见，本文省略了相关内容。

对于伪校准，我们仅保留大小至多为 n^τ 的傅里叶系数。参数之间的关系为 $\delta \le c\tau \le c'\varepsilon$，其中 $c' < c < 1$ 是绝对常数。我们假设这些参数足够小，使得所有证明都能顺利进行。

$\widetilde{}$伪校准默认也不满足条件 $\mathbb{E}[1] = 1$。不过，只需将算子除以 $\mathbb{E}[1]$ 即可轻松修正这一点$\widetilde{}$。正如注 5.4.9 中将指出的那样，$\widetilde{}$在未归一化的伪校准中，以高概率有 $\mathbb{E}[1] = 1 + o_n(1)$，因此引入的误差不会影响任何引理的结论。

5.3.3 高斯情形下的伪校准

我们先计算高斯情形下的伪校准。这里自然选择的傅里叶基是厄米特多项式。令 $\alpha \in \mathbb{N}^m$ 表示厄米特多项式的一个指标。定义 $\alpha! := \prod_{u,i} \alpha_{u,i}!$ 和 $|\alpha| := \sum_{u,i} \alpha_{u,i}$，以及 $|\alpha_u| := \sum_i \alpha_{u,i}$。我们将$h_\alpha(d_1, \dots, d_m)$ 记作未归一化的厄米特多项式，因此 $h_\alpha/\sqrt{\alpha!}$ 构成向量 $d_1, \dots, d_m$ 各分量所对应多项式空间的一组标准正交基，其内积为 $\langle p, q\rangle = \underline{\mathbb{E}}_{d_1,\dots,d_m \sim \mathcal{N}(0,I)} [p \cdot q]$。

我们可以将 α 视为一个自然数 $m \times n$ 矩阵，并据此定义$\alpha \in (\mathbb{N}^m)^n$。

$^\intercal$

引理 5.3.4。对于任意 $I \subseteq [n]$，伪校准值为

$$\widetilde{\mathbb{E}} v^I = \sum_{\substack{\alpha : |\alpha| \le n^\tau, \\ |\alpha_u| \text{ even}, \\ |(\alpha^\intercal)_i| \equiv I_i \pmod 2}} \left(\prod_{u=1}^{m} h_{|\alpha_u|}(1) \right) \cdot \frac{1}{n^{|I|/2 + |\alpha|/2}} \cdot \frac{h_\alpha(d_1, \dots, d_m)}{\alpha!}.$$

换言之，非零傅里叶系数是那些行和为偶数，且列和与 I 的奇偶性相符的系数。

证明。 截断伪校准值定义为

$$\widetilde{\mathbb{E}}v^I = \sum_{\alpha:|\alpha|\le n^\tau} \frac{h_\alpha(d_1,\dots,d_m)}{\alpha!} \cdot \mathbb{E}_{\mathrm{pl}}[h_\alpha(d_1,\dots,d_m)\cdot v^I]$$

因此，我们开始计算植入矩。对于这项计算，以下引理至关重要。这里，我们用生成函数给出该引理的简短证明。另一种组合证明见 [87]。

引理 5.3.5. 令 $\alpha \in \mathbb{N}^n$。当 v 固定且 b 固定（不一定是 ± 1）并且$d \sim N(0, I)$ 在给定 $\langle v, d\rangle = b\,\|v\|$ 的条件下，

$$\mathbb{E}_d[h_\alpha(d)] = \frac{v^\alpha}{\|v\|^{|\alpha|}} \cdot h_{|\alpha|}(b).$$

证明。 只需在 $\|v\| = 1$ 时证明该断言，因为左侧与$\|v\|$无关。将 $d = bv + (I - vv)x$ 表示为其中 $x \sim N(0, I)$ 是一个标准正态变量。现在我们要

$$\mathbb{E}_{x\sim N(0,I)} h_\alpha\left(bv + (I - vv^\intercal)x\right).$$

厄米多项式生成函数为

$$\sum_{\alpha\in\mathbb{N}^n} \mathbb{E}_{x\sim N(0,I)} h_\alpha\left(bv + (I - vv^\intercal)x\right)\frac{t^\alpha}{\alpha!} = \mathbb{E}_x \exp\left(\langle bv + (I - vv^\intercal)x, t\rangle - \frac{\|t\|_2^2}{2}\right)$$

$$= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot \exp\left(\langle bv + (I - vv^\intercal)x, t\rangle - \frac{\|t\|_2^2}{2} - \frac{\|x\|_2^2}{2}\right) dx.$$

配方后，

$$
\begin{aligned}
&= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot \exp\left(\langle bv, t\rangle - \frac{\langle v, t\rangle^2}{2} - \frac{1}{2} \cdot \|x - (t - \langle v, t\rangle v)\|_2^2 \right) \, dx \\
&= \exp\left(\langle bv, t\rangle - \frac{\langle v, t\rangle^2}{2} \right) \\
&= \exp\left(b\langle v, t\rangle - \frac{1}{2} \cdot \langle v, t\rangle^2 \right).
\end{aligned}
$$

如何按 t 展开这个表达式的泰勒级数？$\exp(by - \frac{y^2}{2})$ 的泰勒展开为 $\sum_{i=0}^{\infty} h_i(b) \frac{y^i}{i!}$。也就是说，$\exp(by - \frac{y^2}{2})$ 关于 y 的 i 阶导数在 0 处的取值为 $h_i(b)$。利用 $y = \langle v, t\rangle$ 的链式法则，我们的表达式关于 t 的 α 阶导数在 0 处的取值为 $v^\alpha \cdot h_{|\alpha|}(b)$。当 $\|v\| = 1$ 时，这正是我们所需的表达式；再结合前面关于 $\|v\|$ 中齐次性的说明，证明即告完成。

■

现在我们可以完成计算了。要计算 $\mathbb{E}_{\mathrm{pl}}[h_\alpha(d_1, \ldots, d_m) \cdot v^I]$，对 v 和 b_u 求边缘分布，并将条件独立的 b_u 与 d_u 分解。

$$
\begin{aligned}
\mathbb{E}_{\mathrm{pl}}[h_\alpha(d_1, \ldots, d_m) v^I] &= \mathbb{E}_{v, b_u} v^I \prod_{u=1}^{m} \mathbb{E}_d \left[h_{\alpha_u}(d_u) \mid v, b_u \right] \\
&= \mathbb{E}_{v, b_u} v^I \cdot \prod_{u=1}^{m} \frac{v^{\alpha_u}}{\|v\|^{|\alpha_u|}} \cdot h_{|\alpha_u|}(b_u) \qquad \text{(Lemma 5.3.5)} \\
&= \left(\mathbb{E}_v \frac{v^{I + \sum_{u=1}^m \alpha_u}}{\|v\|^{\sum_{u=1}^m |\alpha_u|}} \right) \cdot \left(\prod_{u=1}^{m} \mathbb{E}_{b_u} h_{|\alpha_u|}(b_u) \right)
\end{aligned}
$$

如果度为奇数，那么厄米特多项式在对 b_u 求期望时，其期望值为零；否则，b_u 被提升为偶数幂，可以替换为 1。这要求对所有 u，$|\alpha_u|$ 都是偶数。范数 $\|v\|$ 恒为 1，因此可以省略。如果每个 $|(\alpha)_i|$ 的奇偶性都与 I_i 匹配，分子将为 $\frac{1}{n^{|I|/2 + |\alpha|/2}}$；否则为 0。伪校准计算至此完成。

T

■

现在我们可以用图矩阵来表示 $\mathcal{M}$。

定义 5.3.6。 令 $\mathcal{L}$ 为满足以下性质的所有正规形状 α 的集合

- U_α 和 V_α 仅包含方形顶点和 $|U_\alpha|, |V_\alpha| \leq n^\delta$
- W_α 没有度为 0 的顶点
- $\deg(i) + U_\alpha(i) + V_\alpha(i)$ 对所有 $i \in V(\alpha)$ 均为偶数
- $\deg(u)$ 为偶数，并且 $\deg(u) \geq 4$ 对所有 $u \in V(\alpha)$
- $|E(\alpha)| \leq n^\tau$

注 5.3.7。 请注意， $\mathcal{L}$ 中的形状在 $U_\alpha \cap V_\alpha$中可以有孤立顶点。

注 5.3.8. $\mathcal{L}$ 囊括了将 $\mathcal{M}$ 写成图矩阵的形式时系数非零的所有形状。约束 $\deg(u) \geq 4$ 的出现，是因为伪校准告诉我们 $\deg(u)$ 为偶数， u 无法被孤立出来，并且 $h_2(1) = 0$。



对于形状 α，我们定义

$$\alpha! := \prod_{e \in E(\alpha)} l(e)!$$

请注意，这等于该形状对应的厄米特多项式指标的阶乘。

定义 5.3.9. 对于任意形状 α，若 $\alpha \in \mathcal{L}$，则定义

$$\lambda_\alpha := \left(\prod_{u \in V(\alpha)} h_{\deg(u)}(1) \right) \cdot \frac{1}{n^{(|U_\alpha| + |V_\alpha| + |E(\alpha)|)/2}} \cdot \frac{1}{\alpha!}$$

否则，定义 $\lambda_\alpha := 0$。

推论 5.3.10. 除脚注外[3]， $\mathcal{M} = \sum \lambda_\alpha M_\alpha$。 形状 α

3. 严格来说，图矩阵 M_α 的行和列由 $\mathcal{C}_m \cup \mathcal{S}_n$的所有子集作为索引。取自 $\binom{\mathcal{S}_n}{\leq D/2}$ 的行和列构成的子矩阵，等于 $\widetilde{\mathbb{E}}$的矩矩阵。

5.3.4 布尔情形下的伪校准

现在我们给出布尔情形下的伪校准。接下来，我们需要介绍布尔立方体切片上向量的记号。

定义 5.3.11 (切片**)**。 设 $v \in \{\pm 1\}^n$ 和 $\theta \in \mathbb{Z}$。切片 $\mathcal{S}_v(\theta)$ 定义为

$$\mathcal{S}_v(\theta) \coloneqq \{d \in \{\pm 1\}^n \mid \langle v, d\rangle = \theta\}.$$

我们用 $\mathcal{S}_v(\pm\theta)$ 表示 $\mathcal{S}_v(\theta) \cup \mathcal{S}_v(-\theta)$，用 $\mathcal{S}(\theta)$ 表示 $\mathcal{S}_v(\theta)$，其中 v 是全 *1* 向量。

注 5.3.12。 采用切片的记号，布尔情形下的植入分布也可以等价地描述为

1. 均匀采样 $v \in \{\frac{\pm 1}{\sqrt{n}}\}^n$，然后

2. 从 $\mathcal{S}_{\sqrt{n}\cdot v}(\pm\sqrt{n})$ 中独立且均匀地采样 d_1、.....、d_m。 – –

并非对每个 n 都存在植入分布，但这无关紧要，因为我们仍可通过相同的公式定义伪期望。

我们还需要计算单项式在切片 $\mathcal{S}(\sqrt{n})$ 上的期望，因为它们会出现在伪校准傅里叶系数的描述中。

定义 5.3.13。 $e(k)$**:** $= \mathbb{E}_{x \in_R \mathcal{S}(\sqrt{n})} [x_1 \cdots x_k]$.

现在我们来计算 $\widetilde{\mathbb{E}} v^\beta$ 的傅里叶系数，其中 $\beta \in \mathbb{F}_2^n$。当$d_1, \ldots, d_m \in_{\mathrm{R}} \{\pm 1\}^n$时，傅里叶基是奇偶函数的集合。因此，特征可以表示为$\alpha \in (\mathbb{F}_2^n)^m$，其中 α由 m 个向量组成 $\alpha_1, \ldots, \alpha_m \in \mathbb{F}_2^n$。更确切地说，与 α 相关联的特征 χ_α 定义为

$$\chi_\alpha(d_1, \ldots, d_m) := \prod_{u=1}^{m} d_u^{\alpha_u}$$

我们用 $|\alpha|$ 表示 α 中非零项的数量，并类似地定义 $|\alpha_u|$。将α 视为一个元素属于 $\mathbb{F}_2$ 的 $m \times n$ 矩阵，我们还定义 $\alpha \in (\mathbb{F}_2^n)^m$。 $^\intercal$

引理 5.3.14。我们有

$$\widetilde{\mathbb{E}} v^\beta = \frac{1}{n^{|\beta|/2}} \sum_{\substack{\alpha\colon |\alpha| \le n^\tau, \\ |\alpha_u| \text{ even}, \\ |\alpha_i^\intercal| \equiv \beta_i \pmod 2}} \prod_{u=1}^m e(|\alpha_u|) \cdot \chi_{\alpha_u}(d_u).$$

非零系数的集合与高斯情形中的结构类似： α 的各行必须包含偶数个元素，而第 i 列的奇偶性必须与β_i匹配。

证明。给定 $\alpha \in (\mathbb{F}_2^n)^m$，其中 $|\alpha| \le n^\tau$，伪校准方程按构造强制满足

$$\mathbb{E}_{d_1,\dots,d_m \in \{\pm 1\}^n} (\widetilde{\mathbb{E}} v^\beta)(d_1,\dots,d_m) \cdot \chi_\alpha(d_1,\dots,d_m) = \mathbb{E}_{\mathrm{pl}} v^\beta \cdot \chi_\alpha(d_1,\dots,d_m).$$

计算上式右侧可得

$$\begin{aligned}
\mathbb{E}_{v \in \{\pm 1\}^n} \mathbb{E}_{d_1,\dots,d_m \in_{\mathrm{R}} \mathcal{S}_v(\pm\sqrt{n})} \left[v^\beta \prod_{u=1}^m \chi_{\alpha_u}(d_u) \right] &= \mathbb{E}_{v \in \{\pm 1\}^n} \mathbb{E}_{d_1,\dots,d_m \in_{\mathrm{R}} \mathcal{S}(\pm\sqrt{n})} \left[v^\beta \prod_{u=1}^m \chi_{\alpha_u}(v) \chi_{\alpha_u}(d_u) \right] \\
&= \mathbb{E}_{v \in \{\pm 1\}^n} \chi_{\alpha_1+\cdots+\alpha_m+\beta}(v) \mathbb{E}_{d_1,\dots,d_m \in \mathcal{S}(\pm\sqrt{n})} \left[\prod_{i=1}^m \chi_{\alpha_i}(d_i) \right] \\
&= \mathbf{1}_{[\alpha_1+\cdots+\alpha_m=\beta]} \cdot \prod_{i=1}^m \mathbb{E}_{d_i \in \mathcal{S}(\pm\sqrt{n})} [\chi_{\alpha_i}(d_i)] \\
&= \mathbf{1}_{[\alpha_1+\cdots+\alpha_m=\beta]} \cdot \prod_{i=1}^m \mathbf{1}_{[|\alpha_i| \equiv 0 \pmod 2]} \cdot \prod_{i=1}^m e(|\alpha_i|).
\end{aligned}$$

由于我们有每个特征的傅里叶系数的一般表达式，应用傅里叶反演即可完成证明。

■

现在我们可以用图矩阵来表示矩矩阵。

定义 5.3.15。 令 $\mathcal{L}_{bool}$ 为 $\mathcal{L}$ 中定义 *5.3.6* 所述形状的集合，其中所有边标签均为 *1*。

注记 5.3.16。 $\mathcal{L}_{bool}$ 涵盖了将$\mathcal{M}$ 表示为图矩阵的线性组合时所有系数非零的形状。与注记 *5.3.8* 类似，由于 $e(2) = 0$ （见断言 *5.6.5*），对于 $\mathcal{L}_{bool}$ 中的形状，我们有相同的条件 $deg(\ u) \geq 4$。

○

定义 5.3.17。 对于所有形状 α，若 $\alpha \in \mathcal{L}_{bool}$，则定义

$$\lambda_\alpha := \frac{1}{n^{(|U_\alpha|+|V_\alpha|)/2}} \prod_{ⓤ \in V(\alpha)} e(\deg(ⓤ))$$

否则，令 $\lambda_\alpha := 0$。

Corollary 5.3.18. $\mathcal{M} = \sum_{shapes\ \alpha} \lambda_\alpha M_\alpha$

统一分析

事实证明，布尔情形的分析大体上可由高斯情形的分析推得。首先，布尔伪校准本质上等同于去除了所有包含至少一条标签为 $k \geq 2$ 的边的形状后的高斯伪校准。图矩阵上的系数实际上会略有不同，但二者都满足一个足以满足我们需求的上界（精确陈述见命题 5.4.13）。

为统一分析中的记号，我们方便地将布尔情形下图的边函数设为

$$h_k(x) = \begin{cases} 1 & \text{if } k = 0 \\ x & \text{if } k = 1 \\ 0 & \text{if } k \geq 2 \end{cases}$$

这种对 $h_k(x)$ 的选择保留了这样一个事实： $\{h_0(x) = 1,\ h_1(x) = x\}$ 在布尔情形下是正交多项式基，同时将标签较大的图置零。

在分析过程中，我们可能会将两个图矩阵相乘，从而产生带有不合法的平行边的图矩阵（所谓的“交项”）。对于一对固定的顶点 u, i， u 和 i 之间标签为 $l_1, \ldots, l_s$ 的平行边，对应于正交多项式 $\prod_{j=1}^{s} h_{l_j}(d_{u,i}) =: q(d_{u,i})$ 的乘积。我们会将这个乘积重新表示为正交多项式族中的多项式线性组合，即对于某些系数 $\lambda_i \in \mathbb{R}$， $q(d_{u,i}) = \sum_{i=0}^{\deg(q)} \lambda_i \cdot h_i(d_{u,i})$。在布尔情形下，多项式 $q(d_{u,i})$ 将是 $h_0(d_{u,i}) = 1$ 或 $h_1(d_{u,i}) = d_{u,i}$。然而，在高斯情形下，对应的厄米特多项式 h_i 可能有多达 $\deg(q)$ 个非零且可能更大的系数 λ_i。对于以这种方式产生的图，我们将始终通过应用三角不等式和范数界来界定其对 $\mathcal{M}$ 的贡献。由于我们使用高斯情形下较大的系数 λ_i 来证明界，因此使用布尔情形下的 0/1 系数时，同样的界依然适用。

在分析在两种情形之间有所不同时，我们会分别讨论。

5.4 证明半正定性

观察构成 $\mathcal{M}$ 的各种形状，其中具有 k 个方形顶点的平凡形状，会在 $\mathcal{M}$ 的度为 $2k$ 的子矩阵上贡献一个单位矩阵。我们的最终目标是用这些单位矩阵来界定所有形状。

定义 5.4.1 (块)。 对于 $k, l \in \{0, 1, \ldots, D/2\}$， $\mathcal{M}$ 的 (k, l) 块是一个子矩阵，其行来自 $\binom{[n]}{k}$，列来自 $\binom{[n]}{l}$。注意，当 $\mathcal{M}$ 表示为图矩阵之和时，这恰好将 $\mathcal{M}$ 限制为满足 $|U_\alpha| = k$ 和 $|V_\alpha| = l$ 的形状 α。

我们定义参数 $\eta := 1/\sqrt{n}$。平凡形状位于 $\mathcal{M}$ 的对角块中，并在 (k, k) 块的对角线上贡献一个因子 $\frac{1}{n^k} = \eta^{2k}$。原则上，我们

可以通过考虑 v 缩放后的矩，而不是v 本身的矩，使 η 尽可能小[4] 。出乎意料的是，事实证明这种缩放有助于我们证明半正定性（更多详情见[87] ）。事实证明，在分析中，将 v 伪校准为单位向量（等价地，使用 $\eta = 1/\sqrt{n}$）就足够了。

–

为了实现用恒等项界定 $\mathcal{M}$ 的目标，我们将界定 $\mathcal{M}$ 每个块上的矩阵范数，并应用以下引理来得出半正定性。

引理 5.4.2. 设对称矩阵 $\mathcal{A} \in \mathbb{R}^{\binom{[n]}{\leq D} \times \binom{[n]}{\leq D}}$ 满足：对某个参数$\eta \in (0,1)$,

1. 对每个 $k \in \{0, 1, \ldots, D\}$，$(k, k)$ 块的最小奇异值至少为 $\eta^{2k}(1 - \frac{1}{D+1})$

——

2. 对每个满足 $k \neq l$ 的 $k, l \in \{0, 1, \ldots, D\}$，$(k, l)$ 块的范数至多为 $\frac{\eta^{k+l}}{D+1}$。那么$\mathcal{A} \succeq 0$。

证明。 我们需要证明，对所有向量 x, x $\mathcal{A}x \geq 0$ 都成立。给定一个向量 x，令 $x_0, \ldots, x_D$ 为其在块 $0, \ldots, D$ 中的分量。注意到

$$x^\intercal \mathcal{A} x \geq \sum_{k \in [0,D]} \eta^{2k} \left(1 - \frac{1}{D+1}\right) \|x_k\|^2 - \sum_{k \neq l \in [0,D]} \frac{\eta^{k+l}}{D+1} \|x_k\| \|x_l\|$$

$$= \left(\|x_0\|, \eta \|x_1\|, \ldots, \eta^D \|x_D\|\right) \begin{pmatrix} 1 - \frac{1}{D+1} & -\frac{1}{D+1} & \cdots & -\frac{1}{D+1} \\ -\frac{1}{D+1} & 1 - \frac{1}{D+1} & \cdots & -\frac{1}{D+1} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{D+1} & -\frac{1}{D+1} & \cdots & 1 - \frac{1}{D+1} \end{pmatrix} \begin{pmatrix} \|x_0\| \\ \eta \|x_1\| \\ \vdots \\ \eta^D \|x_D\| \end{pmatrix} \geq 0.$$

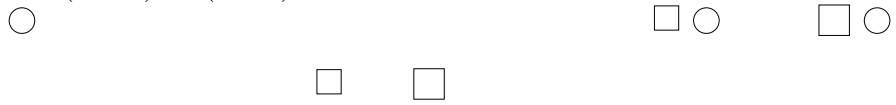
■

我们首先定义蜘蛛，它们是一些特殊形状 α，在分解 $\mathcal{M}$ 时需要单独处理。通俗地说，这些形状包含特殊的子结构，使得它们的

4. 尽管对于极小的 η，伪校准截断误差可能会变得不可忽略。

范数界相对于单位矩阵而言不会变得可忽略不计。随后我们证明，非蜘蛛形状的范数有界。

定义 5.4.3（左蜘蛛）。左蜘蛛是一种真形状 $\alpha = (V(\alpha)\ E(\alpha)\ U_\alpha, V_\alpha)$，其性质是存在两个不同的度为 1 的方形顶点 $i, j \in U_\alpha$，以及一个圆形顶点 $u \in V(\alpha)$，使得 (α) 包含边 (i, u) 和 (j, u)（这些必然是分别与 i 和 j 相连的唯一边）。



顶点 i 和 j 称为 α 的端点。由于度的奇偶性，端点必须位于 $U_\alpha \setminus (U_\alpha \cap V_\alpha)$ 中。

定义 5.4.4（右蜘蛛）。如果形状 $\alpha = (V(\alpha)\ E(\alpha)\ U_\alpha, V_\alpha)$ 的恒等矩阵 $\alpha = (V(\alpha)\ E(\alpha)\ V_\alpha, U_\alpha)$ 是左蜘蛛，则称其为右蜘蛛。α 的端点也称为 α 的端点。

定义 5.4.5（蜘蛛）。形状 α 是蜘蛛，当且仅当它是左蜘蛛或右蜘蛛。**注 5.4.6**。一个蜘蛛可以有许多对端点。对于每一种可能的蜘蛛形状，我们都选出一对端点，以便在下文中可以讨论蜘蛛的“那一对”端点。

5.4.1 非蜘蛛可以忽略不计

现在我们将证明，非蜘蛛的范数很小。需要指出的是，非蜘蛛的这个范数界在很大程度上依赖于假设 $m \leq n^{3/2-\varepsilon}$。

引理 5.4.7。若 $\alpha \in \mathcal{L}$ 既不是平凡形状，也不是蜘蛛，则

$$\frac{1}{n^{|E(\alpha)|/2}} n^{\frac{w(V(\alpha))-w(S_{\min})}{2}} \leq \frac{1}{n^{\Omega(\varepsilon|E(\alpha)|)}}$$

其中 S_{min} 是 α的最小顶点分隔集。

证明。 证明的思路如下。每个不属于最小顶点分隔集的方形顶点都会为范数界贡献 $\sqrt{n}$，而每个不属于最小顶点分隔集的圆形顶点都会贡献 $\sqrt{m}$。为此，我们将尝试从每条边中取出因子 $\frac{1}{\sqrt{n}}$，并将其分配给该边的两个端点，使得每个不属于最小顶点分隔集的方形顶点所分配到的因子不大于 $\frac{1}{\sqrt{n}}$，而每个不属于最小顶点分隔集的圆形顶点所分配到的因子不大于 $\frac{1}{\sqrt{m}}$。

注释 5.4.8。 实际上，我们将使用一组 S方形顶点，而不是最小顶点分隔集，使得 $w(S) \leq w(S\min)$。详情请见下文的实际分配方案。

为了说明我们所采用的分配方案，我们先介绍两种不太奏效的尝试。为简单起见，在前两种尝试中，我们假设 $U_\alpha \cap V_\alpha = \emptyset$作为 $U_\alpha \cap V_\alpha$ 中的顶点，基本上可以忽略不计。

尝试 1：取每条边，给其方形端点分配一个 $\frac{1}{\sqrt[4]{n}}$ 的因子，并给其圆形端点分配一个 $\frac{1}{\sqrt[8]{m}}$ 的因子。

采用这种分配方案，由于每个圆形顶点的度数至少为 4，因此分配给每个圆形顶点的因子不超过 $\frac{1}{\sqrt{m}}$。由于 W_α 中的每个方形顶点的度数至少为 2，因此 W_α 中的每个方形顶点分配到的因子不超过 $\frac{1}{\sqrt{n}}$。然而， $U_\alpha \cup V_\alpha$ 中的方形顶点的度数可能只有 1，此时它们分配到的因子为 $\frac{1}{\sqrt[4]{n}}$，这还不够小。

为解决这一问题，我们可以让所有与方形顶点相接的边将其全部的 $\frac{1}{\sqrt{n}}$ 因子赋予该方形顶点$U_\alpha \cup V_\alpha$。

注 5.4.9。 对于分析 $\mathbb{E}[1]$，第一次尝试作为 $U_\alpha = V_\alpha = \emptyset$ 是可行的。因此，只要 $m \leq n^{2-\varepsilon}$， $\widetilde{\mathbb{E}}[1] = 1 \pm o_n(1)$ 就以高概率成立。

尝试 2：对于每条连接 $U_\alpha \cup V_\alpha$ 中的方形顶点与圆形顶点的边，我们给方形顶点分配因子 $\frac{1}{\sqrt{n}}$，而不给圆形顶点分配任何因子。对于所有其他边，我们给其方形端点分配因子 $\frac{1}{\sqrt[4]{n}}$，给其圆形端点分配因子 $\frac{1}{\sqrt[6]{m}}$（这是可行的，因为 $m \leq n^{\frac{3}{2}-\varepsilon}$）。

采用这种分配方案，每个方形顶点都被赋予一个 $\frac{1}{\sqrt{n}}$ 因子。由于 α 不是蜘蛛，没有圆形顶点同时与 U_α 或 V_α 中的两个顶点相邻。因此，任何未同时与 U_α 中的一个方形顶点和 V_α 中的一个方形顶点相邻的圆形顶点，都必须与 W_α 中至少 3 个方形顶点相邻，因此被赋予的因子不大于 $\frac{1}{\sqrt{m}}$。然而，可能存在同时与 U_α 中的一个方形顶点和 V_α 中的一个方形顶点相邻的圆形顶点。这些圆形顶点可以被赋予 $\frac{1}{\sqrt[3]{m}}$ 因子，但这个因子还不够小。

为了解决这个问题，注意到每当有一个圆形顶点同时与 U_α 中的一个方形顶点和 V_α 中的一个方形顶点相邻时，就会形成一条从 U_α 到 V_α、长度为 2 的路径。任何顶点分隔集都必须包含这条路径上的一个顶点，因此我们可以将这两个方形顶点中的一个放入 S，并且不给它赋予 $\frac{1}{\sqrt{n}}$ 因子。

实际分配方案：根据这些观察，我们采用以下分配方案。这里我们不再假设 $U_\alpha \cap V_\alpha$ 为空。

1. 按照以下方式选择一组方形顶点 $S \subseteq U_\alpha \cup V_\alpha$。从 $S = U_\alpha \cap V_\alpha$ 开始。每当有一个圆形顶点同时与 $U_\alpha \setminus V_\alpha$ 中的一个方形顶点和 $V_\alpha \setminus U_\alpha$ 中的一个方形顶点相邻时，就将这两个方形顶点中的一个放入 S（任选其一）。注意，$w(S) \leq w(S$最小值$)$

2. 对于每条与 S 中某个方形顶点相接的边，将因子 $\frac{1}{\sqrt[3]{m}}$ 分配给其圆形端点，而不给这个方形顶点分配任何因子。

3. 对于 $(U_\alpha \cup V_\alpha) \setminus S$中与方形顶点相连的每条边，向方形顶点赋予因子 $\frac{1}{\sqrt{n}}$，圆形顶点不赋予任何因子。

4. 对于所有其他边，向其方形端点赋予因子 $\frac{1}{\sqrt[4]{n}}$，向其圆形端点赋予因子 $\frac{1}{\sqrt[6]{m}}$。

现在，每个不在 S 中的方形顶点都被分配了一个 $\frac{1}{\sqrt{n}}$ 的因子；由于 α 不是蜘蛛，所有圆形顶点都被分配了一个不大于 $\frac{1}{\sqrt{m}}$ 的因子。

现在我们将这一论证形式化。

分别设 $\mathcal{C}_\alpha$ 和 $\mathcal{S}_\alpha$ 为 α 中的圆形顶点集合和方形顶点集合。我们有 $n^{w(V(\alpha))-w(S\text{最小})_2} \leq n^{0.5|\mathcal{S}_\alpha \setminus S_{min}| + (0.75-\frac{\varepsilon}{2})|\mathcal{C}_\alpha \setminus S_{min}|}$。因此，只需证明

$$|E(\alpha)| - |\mathcal{S}_\alpha \setminus S_{min}| - (1.5-\varepsilon)|\mathcal{C}_\alpha \setminus S_{min}| \geq \Omega(\varepsilon|E(\alpha)|)$$

设 $Q = U_\alpha \cap V_\alpha$，$P = (U_\alpha \cup V_\alpha) \setminus Q$，并设 P' 为 P 中度为 1 且不在 S_{min} 中的顶点集合。设 E_1 为与 P' 相 incident 的边集合，并令 $E_2 = E(\alpha) \setminus E_1$。

对于每个顶点 i（分别为 u），设 E_2 中与其 incident 的边数为 $\deg'(i)$（分别为 $\deg'(u)$）。由于 α 是二分的，我们有 $|E_2| = \sum_{i \in \mathcal{S}_\alpha} \deg'(i) = \sum_{u \in \mathcal{C}_\alpha} \deg'(u)$。于是得到

$$|E(\alpha)| = |E_1| + |E_2| = |P'| + \frac{1}{2}\Big(\sum_{i \in \mathcal{S}_\alpha} \deg'(i) + \sum_{u \in \mathcal{C}_\alpha} \deg'(u)\Big)$$

我们还有 $|S_\alpha \setminus S_{min}| \leq |P'| + |\mathcal{S}_\alpha \cap W_\alpha| + |\mathcal{S}_\alpha \cap (P \setminus P')| \leq |P'| + \frac{1}{2}\sum_{i \in \mathcal{S}_\alpha} \text{度}'(i)$，因为每个 $P' \cup Q$ 之外的方形顶点的度数至少为 2，且不与 E_1 中的任何边关联。因此，只需证明

$$\frac{1}{2}\sum_{u \in \mathcal{C}_\alpha} \deg'(u) - (1.5-\varepsilon)|\mathcal{C}_\alpha \setminus S_{min}| \geq \Omega(\varepsilon|E(\alpha)|)$$

现在注意到，在 E_1 中，每个 $u \in \mathcal{C}_\alpha$ 至多与两条边关联。这是因为，若它在 E_1 中与至少 3 条边相邻，那么要么 u 与 U_α 中至少两个度为 1 的顶点相邻，要么 u 与 V_α 中至少两个度为 1 的顶点相邻。然而，这不可能发生，因为 α 不是蜘蛛。这意味着 $\deg'(u) \geq \deg(u) - 2$。

此外请注意，若 $u \in \mathcal{C}_\alpha \setminus S_{min}$，则有 $\deg'(u) \geq \deg(u) - 1$。这是因为，根据前面的论证，$\deg'(u) = \deg(u) - 2$ 只有在存在 $i \in U_\alpha$，$j \in V_\alpha$ 使得 (i, u)，$(j, u) \in E_1$ 时才会发生。但请注意，根据 P' 的定义，我们有 i，$j \notin S_{min}$；并且根据假设，$u \notin S_{min}$。这意味着存在一条从 U_α 到 V_α 的路径，且不经过 S_{min}，这就产生了矛盾。

最后，我们将 ε 设得足够小，使得下列不等式成立；这两个不等式都可由以下事实推出：对于所有 $u \in \mathcal{C}_\alpha$，$\deg(u) \geq 4$。

1. 对任意 $u \in \mathcal{C}_\alpha \cap S_{min}$，我们有 $\frac{\deg(u)-2}{2} \geq \frac{\varepsilon}{10})\deg(u)$。2. 对任意 $u \in \mathcal{C}_\alpha \setminus S_{min}$，我们有 $\frac{\deg(u)-1}{2} - 1.5 + \varepsilon \geq \frac{\varepsilon}{10})\deg(u)$。

利用这一点，我们得到

$$
\begin{aligned}
&\frac{1}{2}\sum_{u \in \mathcal{C}_\alpha} \deg'(u) - (1.5-\varepsilon)|\mathcal{C}_\alpha \setminus S_{min}| \\
&\qquad \geq \sum_{u \in \mathcal{C}_\alpha \cap S_{min}} \frac{\deg(u)-2}{2} + \sum_{u \in \mathcal{C}_\alpha \setminus S_{min}} \frac{\deg(u)-1}{2} - (1.5-\varepsilon)|\mathcal{C}_\alpha \setminus S_{min}| \\
&\qquad \geq \sum_{u \in \mathcal{C}_\alpha \cap S_{min}} \frac{\varepsilon}{10}\deg(u) + \sum_{u \in \mathcal{C}_\alpha \setminus S_{min}} \left(\frac{\deg(u)-1}{2} - 1.5 + \varepsilon\right) \\
&\qquad \geq \sum_{u \in \mathcal{C}_\alpha \cap S_{min}} \frac{\varepsilon}{10}\deg(u) + \sum_{u \in \mathcal{C}_\alpha \setminus S_{min}} \frac{\varepsilon}{10}\deg(u) \\
&\qquad = \sum_{u \in \mathcal{C}_\alpha} \frac{\varepsilon}{10}\deg(u) = \Omega(\varepsilon|E(\alpha)|)
\end{aligned}
$$

■

由于 $\mathcal{L}_{bool} \subseteq \mathcal{L}$，上述结果也适用于 $\mathcal{L}_{bool}$ 中非平凡的非蜘蛛形状。

C推论 5.4.10. 如果 $\alpha \in \mathcal{L}_{bool}$ 不是平凡形状，也不是蜘蛛，那么

$$\frac{1}{n^{|E(\alpha)|/2}} n^{\frac{w(V(\alpha))-w(S_{\min})}{2}} \leq \frac{1}{n^{\Omega(\varepsilon|E(\alpha)|)}}$$

推论 5.4.11. 如果 $\alpha \in \mathcal{L}$ 不是平凡形状，也不是蜘蛛，那么以高概率

$$\frac{1}{n^{|E(\alpha)|/2}} \|M_\alpha\| \leq \frac{1}{n^{\Omega(\varepsilon|E(\alpha)|)}}$$

证明。 利用引理 5.6.3 中的范数界，我们得到

$$\|M_\alpha\| \leq 2 \cdot (|V(\alpha)| \cdot (1 + |E(\alpha)|) \cdot \log(n))^{C \cdot (|V_{rel}(\alpha)| + |E(\alpha)|)} \cdot n^q \frac{w(V(\alpha)) - w(S_{\min}) + w(W_{iso})}{2}$$

我们有 $W_{iso} = \emptyset$。注意，由于 $V_{rel}(\alpha)$ 中没有零度顶点，因此有 $|V_{rel}(\alpha)| \leq 2|E(\alpha)|$；又因为我们还有 $|V(\alpha)| \cdot (1 + |E(\alpha)|) \cdot \log n \leq n^{O(\tau)}$，所以因子 $2 \cdot (|V(\alpha)| \cdot (1 + |E(\alpha)|) \cdot \log(n))^{C \cdot (|V_{rel}(\alpha)| + |E(\alpha)|)}$ 可以吸收到 $n^{\varepsilon |E(\alpha)|}$ 中。结论由引理 5.4.7 得出。

■

这说明非平凡非蜘蛛形状的范数为 $o_n(1)$（暂时忽略额外因子 η）。现在我们将展示如何利用这个范数界来控制 $\mathcal{M}$ 中一个块内所有非蜘蛛的总范数，即推论 5.4.14。我们首先需要几个命题，这些命题在之后消去蜘蛛后也会派上用场。

命题 5.4.12. 顶点数至多为 L 且边数恰为 k 的正规形状数量至多为 $L^{8(k+1)}$。

证明。 以下过程涵盖了所有形状（尽管许多形状会被构造多次）：

- 选择四个集合 $U \cap V, U \setminus (U \cap V)$ $V \setminus (U \cap V)$ W 中方形变量和圆形变量的数量。这会产生一个因子 L^8。
- 在 每两个顶点之间放置一条边。这会产生一个因子 L^{2k} .

■

命题 5.4.13. $|\lambda_\alpha| \leq \eta^{|U_\alpha| + |V_\alpha|} \cdot |E_{(\alpha}|^{3 \cdot |E\ \alpha\ |}) () () 其中$，按照约定，我们假设 $0^0 = 1$。

证明。**(高斯情形)** 回想一下，系数 λ_α 要么为零，要么由以下公式定义

$$\lambda_\alpha = \eta^{|U_\alpha|+|V_\alpha|} \cdot \left(\prod_{ⓤ \in V(\alpha)} h_{\deg(ⓤ)}(1) \right) \cdot \frac{1}{n^{|E(\alpha)|/2}} \cdot \frac{1}{\alpha!}$$

数列 $h_k(1)$ 满足递推关系 $h_0(1) = h_1(1) = 1$，$h_{k+1}(1) = h_k(1) - k h_{k-1}(1)$。我们可以通过归纳法证明 $|h_k(1)| \le k^k$，因此，

$$\prod_{ⓤ \in V(\alpha)} \left| h_{\deg(ⓤ)}(1) \right| \le \prod_{ⓤ \in V(\alpha)} (\deg(ⓤ))^{\deg(ⓤ)} \le |E(\alpha)|^{|E(\alpha)|}.$$

（布尔情形）在布尔情形下，系数 λ_α 定义为

$$\lambda_\alpha = \eta^{|U_\alpha|+|V_\alpha|} \cdot \left(\prod_{ⓤ \in V(\alpha)} e(\deg(ⓤ)) \right)$$

利用推论 5.6.16，我们有 $|e(k)| \le k^{3k} \cdot n^{-k/2}$。因此，

$$|\lambda_\alpha| = \eta^{|U_\alpha|+|V_\alpha|} \cdot \prod_{ⓤ \in V(\alpha)} |e(\deg(ⓤ))| \le \eta^{|U_\alpha|+|V_\alpha|} \cdot \frac{|E(\alpha)|^{3|E(\alpha)|}}{n^{|E(\alpha)|/2}}.$$

■

推论 5.4.14. 对于 $k, l \in \{0, 1, \ldots, D/2\}$，令 $\mathcal{B}_{k,l} \subseteq \mathcal{L}$ 表示在 (k, l) 块上的非平凡非蜘蛛 $\alpha \in \mathcal{L}$ 集合，即 $|U_\alpha| = k, |V_\alpha| = l$。$\mathcal{B}_{k,l}$ 中非蜘蛛的总范数满足

$$\sum_{\alpha \in \mathcal{B}_{k,l}} |\lambda_\alpha| \, \|M_\alpha\| = \eta^{k+l} \cdot \frac{1}{n^{\Omega(\varepsilon)}}$$

证明。

$$
\begin{aligned}
\sum_{\alpha\in\mathcal{B}_{k,l}} |\lambda_\alpha| \, \|M_\alpha\| &\le \sum_{\alpha\in\mathcal{B}_{k,l}} \eta^{k+l} \cdot \frac{|E(\alpha)|^{3|E(\alpha)|}}{n^{|E(\alpha)|/2}} \, \|M_\alpha\| && \text{(Proposition 5.4.13)} \\
&\le \eta^{k+l} \cdot \sum_{\alpha\in\mathcal{B}_{k,l}} \left(\frac{|E(\alpha)|^3}{n^{\Omega(\varepsilon)}} \right)^{|E(\alpha)|} && \text{(Corollary 5.4.11)} \\
&\le \eta^{k+l} \cdot \sum_{\alpha\in\mathcal{B}_{k,l}} \left(\frac{n^{3\tau}}{n^{\Omega(\varepsilon)}} \right)^{|E(\alpha)|} && (\alpha \in \mathcal{L}) \\
&\le \eta^{k+l} \cdot \sum_{\alpha\in\mathcal{B}_{k,l}} \frac{1}{n^{\Omega(\varepsilon|E(\alpha)|)}} \\
&\le \eta^{k+l} \cdot \sum_{i=1}^{\infty} \frac{n^{O(\tau i)}}{n^{\Omega(\varepsilon i)}} \\
&= \eta^{k+l} \cdot \frac{1}{n^{\Omega(\varepsilon)}}
\end{aligned}
$$

其中最后一个不等式用到了命题 5.4.12 以及事实 $|E(\alpha)| \ge 1$ 对于 $\alpha \in \mathcal{B}_{k,\,l}$。 ∎

5.4.2 消去单个蜘蛛

我们在“证明策略”一节中看到，形状 $2\beta_1 + \frac{1}{n}\beta_2$ 属于一个矩矩阵的零空间，该矩阵满足约束“$\langle v, d_u\rangle^2 = 1$”。形状 β_1 正是蜘蛛中会出现的那类子结构！因此，很自然地希望：如果 α 是左蜘蛛，那么 $\mathcal{M}_{fix} M_\alpha = 0$。但这并不完全成立，因为 $\langle v, d_u\rangle^2$ “缺少”了一些项：在 α的实现中，端点必须与 α中的其他正方形不同，这使得乘积 $\mathcal{M}_{fix} M_\alpha$中无法出现所有顶点对 $i,\ j$对应的项。我们可以加入更小的“交项”（称为 α的塌缩），使端点可以取遍所有顶点对 $i,\ j$。加入这些项后，我们将构造一个满足 $\mathcal{M}_{fix} L = 0$的矩阵 L。

我们首先定义通过合并两个

顶点将一种形状塌缩为另一种形状的含义。这里，我们只定义合并两个方形顶点的情况，因为在我们对交项的分析中，只会发生这类合并。

定义 5.4.15 (非正规坍缩**)**。 设 α 为一个形状，且 $V(\alpha)$ 中有两个不同的方形顶点 i 和 j。我们将 i、j 的非正规坍缩定义为：

- 从 $V(\alpha)$ 中移除 i、j，并将它们替换为一个新的顶点 k。
- 若边 $\{i, u\}$ 和 $\{j, u\}$ 存在，则分别将其替换为 $\{k, u\}$，并保留相同的标签（注意，可能存在重边，因此新形状可能不是正规的）。
- 令 $U(k) = U(i) + U(j)$ (模 2) 且 $V(k) = V(i) \uplus V(j)$ (模 2)。

非正规坍缩具有平行边，但我们可以将它们转换回正规形状之和。具体做法是，对于每组平行边，在傅里叶基中展开傅里叶特征的乘积。例如，两条标签为 1 的平行边应展开为

$$h_1(z)^2 = (z^2 - 1) + 1 = h_2(z) + h_0(z)$$

定义 5.4.16 (形状坍缩**)**。 设 α 是一个具有两个不同方形顶点 i、j 的形状。若 β 出现在 i、j 的非正规坍缩的展开中，我们就称 β 是 i、j 的（正规）坍缩。

注释 5.4.17。 如果 $l_1, \ldots, l_k$ 是一组平行边的标签，那么乘积 $h_{l_1}(z) \cdots h_{l_k}(z)$ 的奇偶性取决于 $l_1 + \cdots + l_k$ 的奇偶性。因此，非零傅里叶系数将是奇偶性相匹配的项。因此，在布尔情形和高斯情形中，给定非正规坍缩的正规坍缩形状，可通过将每组平行边替换为一条边 e 得到，使得 $l(e) \leq l_1 + \ldots + l_k$ 且 $l(e) \equiv l_1 + \cdots + l_k$ (模 2)。

注 5.4.18。 考察定义并结合上一条注释，我们得到如下结论。

1. 形状塌缩不会改变圆形顶点的数量，但方形顶点的数量会减少1。

2. $\alpha \in \mathcal{L}$ 具有这样的性质：顶点的度为奇数，当且仅当它们位于$(U_\alpha \cup V_\alpha) \setminus (U_\alpha \cap V_\alpha)$中。当 α 塌缩时，这一性质保持不变。

现在，我们定义所需的形状 L_k，它们位于 $\mathcal{M}_{fix}$的零空间中。

定义 5.4.19。 对于 $k \geq 2$，定义形状 ℓ_k，它位于 $\{\boxed{1}, \ldots, \boxed{k}, \textcircled{1}\}$ 上并包含两条边 $\{\{\boxed{1}, \textcircled{1}\}, \{\boxed{2}, \textcircled{1}\}\}$。 ℓ_k 的左侧由 $U_{\ell_k} = \{\boxed{1}, \ldots, \boxed{k}\}$组成。右侧由$V_{\ell_k} = \{\boxed{3}, \ldots, \boxed{k}, \textcircled{1}\}$组成。

定义 5.4.20。 将 ℓ_k 的“补全”版本 L_k 定义为以下形状对应的 $c_\beta M_\beta$ 之和，其中 β 为系数：

- $(L_{k,\ 1})$： ℓ_k，系数为 *2*。
- $(L_{k,\ 2})$：如果 $k \geq 3$，则在 ℓ_k 中将 $\boxed{1}$ 和 $\boxed{3}$ 塌缩，系数为 $\frac{2}{n}$。
- $(L_{k,\ 3})$：如果 $k \geq 4$，则在 ℓ_k 中将 $\boxed{1}$ 和 $\boxed{3}$ 塌缩，并将 $\boxed{2}$ 和 $\boxed{4}$ 塌缩，系数为 $\frac{2}{n^2}$。
- $(L_{k,\ 4})$：将 $\boxed{1}$ 和 $\boxed{2}$ 塌缩，以一条带有标签 *2* 的边替代这些边，系数为 $\frac{1}{n}$。
- $(L_{k,\ 5})$：如果 $k \geq 3$，则将 $\boxed{1}$、$\boxed{2}$ 和 $\boxed{3}$ 塌缩，以一条带有标签 *2* 的边替代这些边，系数为 $\frac{1}{n}$。有关带/形状的图示，请参见下方的图 5.7。

引理 5.4.21. $\mathcal{M}_{fix} L_k = 0$

证明。这些形状的构造方式保证：如果我们将顶点1和3, . . . , k固定为 $u \in \mathcal{C}_m$ 和 $S \in \binom{\mathcal{S}_n}{k-2}$，从而确定一个部分实现，那么正方形1 和2 仍可实现为任意$j_1, j_2 \in [n]$。也就是说，恰好成立以下等式，

$$\begin{aligned}(\mathcal{M}_{fix} L_k)_I &= \sum_{\substack{u \in \mathcal{C}_m, \\ S \in \binom{\mathcal{S}_n}{k-2}}} \left(\sum_{\substack{j_1, j_2 \in [n]: \\ j_1 \neq j_2}} \widetilde{\mathbb{E}}[v^I v^S v_{j_1} v_{j_2}] d_{uj_1} d_{uj_2} + \sum_{j_1 \in [n]} \widetilde{\mathbb{E}}[v^I v^S v_{j_1}^2](d_{uj_1}^2 - 1) \right) \\ &= \sum_{\substack{u \in \mathcal{C}_m, \\ S \in \binom{\mathcal{S}_n}{k-2}}} \widetilde{\mathbb{E}}[v^I v^S (\langle v, d_u \rangle^2 - 1)] \\ &= 0\end{aligned}$$

为了说明系数是如何产生的，我们分析 R 由哪些带组成，并观察它们如何对输出作出贡献。带状图/形状的图示见下方的图 5.7。设该带部分实现为 u 和 $S = \{j_3, \ldots, j_k\}$。令 $(M_{fix} L_k)_{I(u,S)}$ 表示 $(M_{fix} L_k)_I$ 中具有此部分实现的项。用这种记号，我们要证明

$$(\mathcal{M}_{fix} L_k)_{I(u,S)} = \sum_{\substack{j_1, j_2 \in [n]: \\ j_1 \neq j_2}} \widetilde{\mathbb{E}}[v^I v^S v_{j_1} v_{j_2}] d_{uj_1} d_{uj_2} + \sum_{j_1 \in [n]} \widetilde{\mathbb{E}}[v^I v^S v_{j_1}^2](d_{uj_1}^2 - 1).$$

1. 如果我们取一个带 R，其中有 $A_R = \{j_1, \ldots, j_k\}$、$B_R = \{j_3, \ldots, j_k\} \cup \{u\}$ 和 $E(R) = \{\{j_1, u\}, \{j_2, u\}\}$，其中 $j_1 \neq j_2$ 且 $j_1, j_2 \notin S$，那么

$$(\mathcal{M}_{fix} M_R)_{I(u,S)} = \widetilde{\mathbb{E}}[v^I v^S v_{j_1} v_{j_2}] d_{uj_1} d_{uj_2}.$$

这条带状图必须“覆盖”两个有序对（j_1，j_2）和（j_2，j_1），因此我们希望每条这样的带状图 R 在 L_k 中的系数为 2。

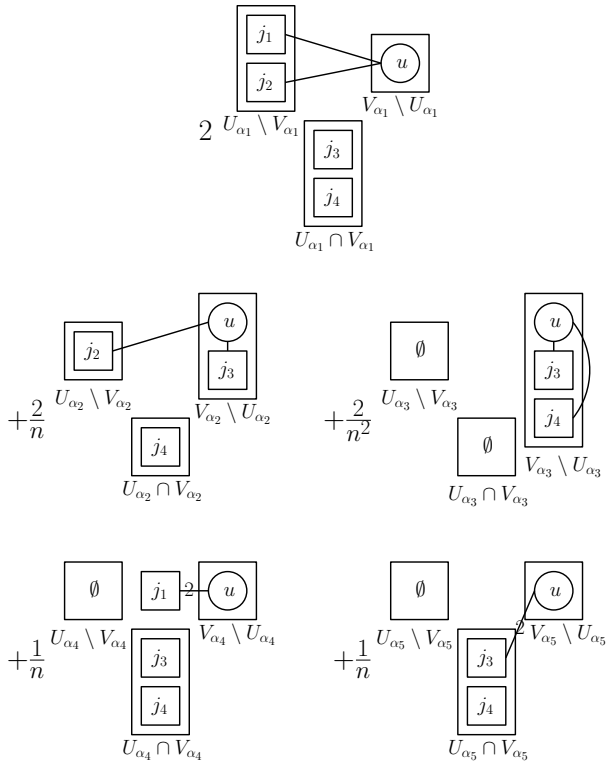


图 5.7：构成 L_4 的五种形状。

2. 如果我们取一条带状图 R，其中 $A_R = \{\boxed{j_1}, \ldots, \boxed{j_k}\} \setminus \{\boxed{j_1}, \boxed{j_3}\}$、$B_R = \{\boxed{j_3}, \ldots, \boxed{j_k}\} \cup \{\textcircled{u}\}$ 和 $E(R) = \{\{\boxed{j_3}, \textcircled{u}\}, \{\boxed{j_2}, \textcircled{u}\}\}$，且 $j_1 = j_3 \in S$，那么

$$(\mathcal{M}_{fix} M_R)_{I(u,S)} = \widetilde{\mathbb{E}}[v^I v^{S \setminus \{j_3\}} v_{j_2}] d_{uj_3} d_{uj_2} = n \widetilde{\mathbb{E}}[v^I v^S v_{j_1} v_{j_2}] d_{uj_1} d_{uj_2}.$$

在 L_k 中取出 $\frac{2}{n}$ 的系数，对于这种与 S 重叠的情形，涵盖了两对 (j_1, j_2) 和 (j_2, j_1)。

3. 如果我们取一个带状图 R，其中有 $A_R = \{\boxed{j_1}, \ldots, \boxed{j_k}\} \setminus \{\boxed{j_1}, \boxed{j_2}, \boxed{j_3}, \boxed{j_4}\}$、$B_R = \{\boxed{j_3}, \ldots, \boxed{j_k}\} \cup \{\textcircled{u}\}$ 和 $E(R) = \{\{\boxed{j_3}, \textcircled{u}\}, \{\boxed{j_4}, \textcircled{u}\}\}$，且 $j_1 = j_3 \in S$ 和 $j_2 = j_4 \in S$，那么

$$(\mathcal{M}_{fix} M_R)_{I(u,S)} = \widetilde{\mathbb{E}}[v^I v^{S \setminus \{j_3, j_4\}}] d_{uj_3} d_{uj_4} = n^2 \widetilde{\mathbb{E}}[v^I v^S v_{j_1} v_{j_2}] d_{uj_1} d_{uj_2}.$$

在 L_k 中取 $\frac{2}{n^2}$ 的系数，涵盖了此情形下的两对 (j_1, j_2) 和 (j_2, j_1)

与 S 重叠。

4. 如果我们取一条带状图 R，其中有 $A_R = \{\boxed{j_1}, \ldots, \boxed{j_k}\} \setminus \{\boxed{j_1}, \boxed{j_2}\}$、$B_R = \{\boxed{j_3}, \ldots, \boxed{j_k}\} \cup \{\textcircled{u}\}$ 和 $E(R) = \{\{\boxed{j_1}, \textcircled{u}\}_2\}$，其中 $j_1 = j_2 \notin S$，那么

$$(\mathcal{M}_{fix} M_R)_{I(u,S)} = \widetilde{\mathbb{E}}[v^I v^S](d_{uj_1}^2 - 1) = n\widetilde{\mathbb{E}}[v^I v^S v_{j_1}^2](d_{uj_1}^2 - 1).$$

取 $\frac{1}{n}$ 在 L_k 中的系数,– 涵盖了这些项。

5. 如果我们取一条带状图 R，其中有 $A_R = \{\boxed{j_1}, \ldots, \boxed{j_k}\} \setminus \{\boxed{j_1}, \boxed{j_2}\}$、$B_R = \{\boxed{j_3}, \ldots, \boxed{j_k}\} \cup \{\textcircled{u}\}$ 和 $E(R) = \{\{\boxed{j_3}, \textcircled{u}\}_2\}$，其中 $j_1 = j_2 = j_3 \in S$，那么

$$(\mathcal{M}_{fix} M_R)_{I(u,S)} = \widetilde{\mathbb{E}}[v^I v^S](d_{uj_3}^2 - 1) = n\widetilde{\mathbb{E}}[v^I v^S v_{j_1}^2](d_{uj_1}^2 - 1).$$

取 $\frac{1}{n}$ 在 L_k 中的系数,– 涵盖了这些项。

■

图矩阵的一个关键事实是，图矩阵相乘近似等于一个新的图矩阵 $M_\alpha \cdot M_\beta \approx M_\gamma$，其中 γ 是将 V_α 与 U_β 粘合后的结果（并且如果 V_α，U_β 的每种类型顶点数量不同，乘积就为零）。近似中的误差项是 α 和 β 中变量之间的交项（塌缩）。

定义 5.4.22. 如果形状 α 和 β 的方形顶点数和圆形顶点数分别相同，我们就称它们可复合。如果形状 γ 是形状 α 和 β 的粘合，也就是说，γ 的图是 α 和 β 的图的不交并，然后在某个保持类型的双射下识别 V_α 和 U_β，并且 $U_\gamma = U_\alpha$ 和 $V_\gamma = V_\beta$。

命题 5.4.23。 设 α, β 是可组合的形状。假设 $V(\alpha) \setminus V_\alpha$ 的顶点全都是方形。令 $\{\gamma_i\}$ 为 α 与 β 的不同粘合，并令 $\mathcal{I}$ 为非正规塌缩的集合

在 $V(\alpha) \setminus V_\alpha$ 中塌缩任意数量的正方形（可能为零）得到的，其中 $V(\beta) \setminus U_\beta$ 中的正方形各不相同，并且$\widetilde{}$中的任意粘合 γ_i。那么存在系数 c_γ，使得对于 $\gamma \in \mathcal{I}$

$$M_\alpha \cdot M_\beta = \sum_{\gamma \in \widetilde{\mathcal{I}}} c_\gamma M_\gamma.$$

此外，系数满足 $\left|c_\gamma\right| \leq 2^{|V(\alpha)\setminus V_\alpha|} |V(\gamma)|^{|V(\alpha)\setminus U_\alpha|}$。

证明。乘积 $M_\alpha \cdot M_\beta$ 是一个矩阵，它是输入 $(d_1, \ldots, d_m)$ 的对称函数。其所在空间由所有可能形状 γ（不局限于 $\mathcal{I}$）对应的 M_γ 张成，因此，如果允许所有形状 γ，便存在系数 c_γ。我们需要证明 $\sim M_\alpha \cdot M_\beta$确实属于形状 $\mathcal{I}$ 张成的空间，只需说明 $M_\alpha \cdot M_\beta$中的所有带状图都具有 $\mathcal{I}$中的形状。扩展定义可得，

$$M_\alpha \cdot M_\beta = \left(\sum_{R \text{ is a ribbon of shape } \alpha} M_R \right) \left(\sum_{S \text{ is a ribbon of shape } \beta} M_S \right) = \sum_{\substack{R \text{ is a ribbon of shape } \alpha, \\ S \text{ is a ribbon of shape } \beta}} M_R M_S.$$

要使 $M_R M_S$ 非零，我们要求 $B_R = A_S$ 作为集合相等； R 可以在 B_R 内任意分配标签，从而产生 α 和 β的不同粘合。固定 R 和 S，并令 γ 为在此 R 和 S下 α 与 β 对应的粘合。

矩阵 $M_R M_S$ 只有一个非零项；我们断言，它是带状图 T 的一个傅里叶特征，而 T 是由 γ塌缩而成的。 R 中位于 B_R 之外的标签可能与 S 中位于 A_S之外的标签重叠，而 T 的形状自然是将 γ 中具有相同标签的顶点塌缩后的结果。

要界定所出现的系数 c_γ，只需界定带状图 M_T上的系数；该系数受贡献带状图数量 R, S的界定，而如果形状为 α的带状图 R和形状为 β的带状图 S满足 $M_R M_S = M_T$，我们就说它们对 T有贡献。根据 T，我们可以完全恢复集合 A_R和 B_S。 V的标签（R）$\setminus A_R$必须来自 T的标签；可以用至多 $|V（\gamma）|^{|V（\alpha）\setminus U_\alpha|}$种方式选取它们。这也决定了 $B_R = A_S$。剩下的只是确定 S 的图结构。由于不正规的收缩不会丢失任何边，

知道 R 的标签后，我们就能准确知道 T 中哪些边必须来自 R 和 S。顶点 $V(T) \setminus V(R)$ 必须来自 S， B_R也同样如此；从 $V(R) \setminus B_R$ 中选取一个子集纳入 $2^{|V}$，共有 $(\alpha)^{\setminus V_\alpha|}$ 种方式。

■

设 α 是一个左蜘蛛，其端点为 i、j，它们与圆形 u 相邻。回想一下，我们的目标是论证 $\mathcal{M}M_\alpha \approx 0$。为此，我们可以尝试将 M_α 沿着顶点分隔集 $S = U_\alpha \cup \{u\} \setminus \{i, j\}$ 进行因式分解，该分隔集将 α 分隔为

$$M_\alpha \approx L_{|U_\alpha|} \cdot M_{\mathrm{body}(\alpha)}$$

其中我们定义

定义 5.4.24。设 α 是一个左蜘蛛，其端点为 i、j。将 $\mathrm{body}(\alpha)$ 定义为这样一种形状：其图为 α，删除 i 和 j，并且 $U_{\mathrm{body}(\alpha)} = U_\alpha \cup \{u\} \setminus \{i, j\}$、$V_{\mathrm{body}(\alpha)} = V_\alpha$。右蜘蛛形的定义类似。

由引理 5.4.21 可知，近似式的右侧属于 $\mathcal{M}$ 的零空间。现在我们将这一近似因式分解形式化。

~~**定义 5.4.25**。设 α 是一个蜘蛛，其端点为 i、j。将 $\mathcal{I}_\alpha$ 定义为通过执行以下至少一个步骤可由 α 得到的形状集合：

- 用方形顶点对 i 在 α 中进行非正规坍缩
- 用方形顶点对 j 在 α 中进行非正规坍缩

令 $\mathcal{I}_\alpha$ 为通过相同过程但使用正规坍缩所能得到的真形状集合。

在上述定义中，我们允许 i，j 与两个不同的正方形塌缩，或彼此塌缩，或二者都与同一个第三顶点塌缩。出于技术原因，我们需要将 $\mathcal{I}_\alpha$ 细化为两个形状集合，并对其中一个集合的系数采用更紧的界。

定义 5.4.26. 令 $\mathcal{I}(1)\alpha$ 为通过执行以下至少一个步骤可从 α 得到的形状集合:

- 塌缩 i，其中 $\mathrm{body}(\alpha) \setminus U_\alpha$中有一个方形顶点
- 塌缩 j，其中 $\mathrm{body}(\alpha) \setminus U_\alpha$ 中有一个方形顶点（与 i 若发生塌缩时的情况不同）

令 $\mathcal{I}(2)\alpha := \mathcal{I}_\alpha \setminus \mathcal{I}(1)\alpha$，并类似地定义非正规的版本 $\widetilde{\mathcal{I}}(1)\alpha, \widetilde{\mathcal{I}}(2)\alpha$。

引理 5.4.27. 设 α 是一个左蜘蛛，其端点为 i, j。存在系数 c_β，使得对于$\sim\beta \in \mathcal{I}_\alpha$

$$L_{|U_\alpha|} \cdot M_{\mathrm{body}(\alpha)} = 2M_\alpha + \sum_{\beta \in \widetilde{\mathcal{I}}_\alpha} c_\beta M_\beta,$$

$$|c_\beta| \le \begin{cases} 40\,|V(\alpha)|^3 & \beta \in \widetilde{\mathcal{I}}_\alpha^{(1)} \\ \frac{40|V(\alpha)|^3}{n} & \beta \in \widetilde{\mathcal{I}}_\alpha^{(2)} \end{cases}.$$

证明。 首先，我们可以验证 M_α 的系数为 2。只有 L_k 的 ℓ_k 项具有最多的正方形数，并且它在 L_k中带有一个因子 2。

$\widetilde{\ }\mathcal{I}_\alpha$ 中的形状按定义就是出现在该图$\widetilde{\ }$矩阵乘积中的交项，而且 $\mathcal{I}_\alpha$ 中的形状按定义就是 ℓ_k 项对应的交项。利用命题 5.4.23，对于 $L_{|U_\alpha|}$ 中的五种形状，每种形状贡献的系数都不超过 $4\,|V(\alpha)|^3$。ℓ_k 上的系数为 2，因此 $\widetilde{\mathcal{I}}(1)\alpha$的系数至多为 $8\,|V(\alpha)|^3$。$L_{|U_\alpha|}$ 中另外四种形状的最大系数为 $\frac{2}{n}$，因此它们对 $\widetilde{\mathcal{I}}(2)\alpha$ 上的系数的总贡献至多为 $\frac{32|V(\alpha)|^3}{n}$。

$-$

$\blacksquare$

现在，我们想将非正规形状转变为来自 $\mathcal{I}_\alpha$的正规形状。遗憾的是，展开一个非正规形状时，并不能简单地逐条边分别展开（尽管对非正规的带状图来说确实如此）。带状图对称性还会带来额外的困难。要了解这一困难，请看下方图 5.8 中的例子。

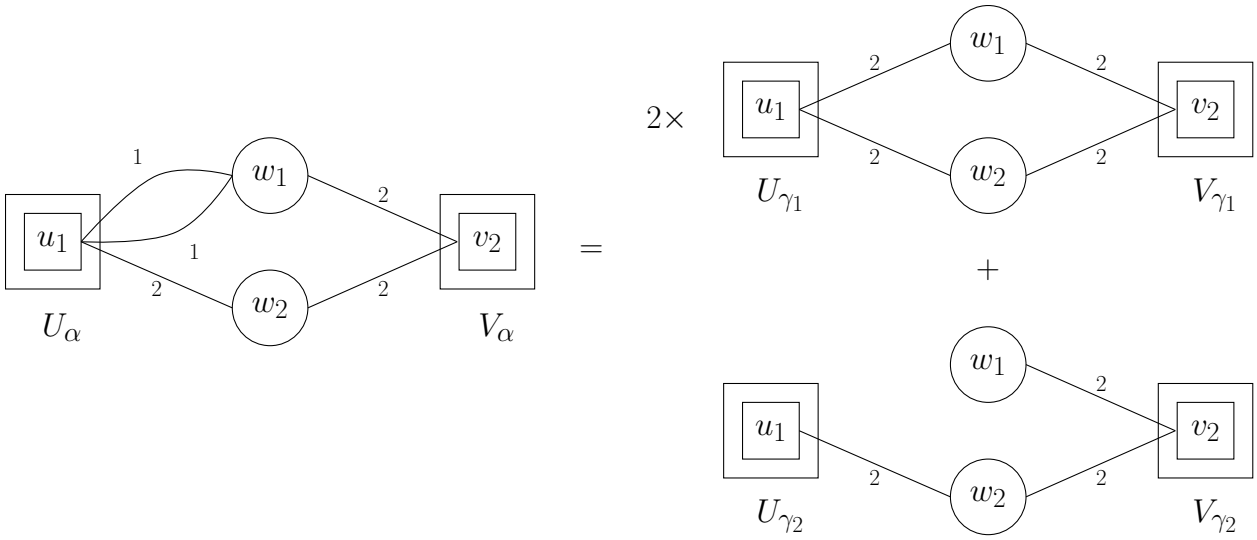


图 5.8：图矩阵的一个令人惊讶的等式。

人们会认为右侧两个形状的系数都应为 1，因为 $h_1(z)^2 = h_2(z) + h_0(z)$。然而，在左侧形状中，两个圆可以区分，因此对所有带状图求和时，会包含一个带有 $w_1 = i$, $w_2 = j$ 的带状图，以及另一个带有 $w_1 = j$, $w_2 = i$ 的带状图。在右上方的形状中，两个圆无法区分，因此圆被赋予 $\{i, j\}$ 的图/带状图会被计算两次。在右下方的形状中，两个圆可以区分，所以所有带状图都只计算一次。为了界定新的系数，我们使用形状自同构这一概念。

定义 5.4.28. 形状 α 的自同构是一个函数 $\varphi\colon V(\alpha) \to V(\alpha)$，它保持集合 U_α, V_α 不变，并且是底层边标记图的一个自同构。令自同构群(α) 表示 α的自同构群。

命题 5.4.29. 设 α 是一个非正规形状，令 $\mathcal{P}$ 为通过扩展 α可以得到的真形状集合。那么存在系数 $|c_\gamma| \le C_{Fourier} \cdot C_{Aut}$使得

$$M_\alpha = \sum_{\gamma \in \mathcal{P}} c_\gamma M_\gamma$$

其中， $C_{Fourier}$ 是展开中傅里叶系数绝对值的上界，并且

C_{Aut} = 最大值$_{\gamma\in\mathcal{P}}$ |自同构群(
γ |) 自同构群( )。
| α |

证明。 给出特定带状图的图矩阵实现数量恰好等于自同构数量，因此

$$M_\alpha = \frac{1}{|\mathrm{Aut}(\alpha)|} \sum_{\text{realizations } \sigma} M_{\sigma(\alpha)}$$

将每个非正规的带状图 $M_{\sigma(\alpha)}$ 展开为系数至多为 $C_{Fourier}$的正规带状图。由于 α 与任意 γ 的实现相同，这恰好对所有 γ 及 γ 的所有实现求和。 γ 的每个实现上的傅里叶系数都相同；记其为 c'_γ，其中 $\left|c'_\gamma\right| \le C_{Fourier}$。继续进行，

$$\begin{aligned} &= \frac{1}{|\mathrm{Aut}(\alpha)|} \sum_{\gamma\in\mathcal{P}} c'_\gamma \sum_{\text{realizations } \sigma} M_{\sigma(\gamma)} \\ &= \sum_{\gamma\in\mathcal{P}} c'_\gamma \frac{|\mathrm{Aut}(\gamma)|}{|\mathrm{Aut}(\alpha)|} M_\gamma \end{aligned}$$

■

命题 5.4.30。 设 $l_1 \le \cdots \le l_k \in \mathbb{N}$ 且设 $L = l_1 + \cdots + l_k$。假设 $L \ge 1$。在 h_{l_1}的傅里叶展开$(z) \cdots h_{l_k}(z)$中，最大系数的绝对值被 $(2L)^{L-l_k}$所界定。

证明。 在布尔情形中，系数为 1。在高斯情形中，乘积中 $h_p(z)$ 的“线性化系数”由正交性给出，为

$$\frac{\mathbb{E}_{z\sim\mathcal{N}(0,1)}[h_{l_1}(z)\cdots h_{l_k}(z)\cdot h_p(z)]}{\mathbb{E}_{z\sim\mathcal{N}(0,1)}[h_p^2(z)]} = \frac{\mathbb{E}_{z\sim\mathcal{N}(0,1)}[h_{l_1}(z)\cdots h_{l_k}(z)\cdot h_p(z)]}{p!}$$

例如， [209, 示例 G（续）] 中的一个公式表明， $\mathbb{E}[h_{l_1}\cdots h_{l_k}\cdot h_p]$ 等于“分块完美匹配”的数量：将 $l_1+\cdots+l_k+p$ 个元素划分为

大小为 l_i 或 p 的块，并且同一块中的任意两个元素都不匹配的完美匹配。分块完美匹配的数量上界为：

- 从块 $l_1, \ldots, l_{k-1}$ 到 $[L]$ 的偏函数至多有 $(L+1)^{L-l_k}$ 种选法。
- 如果这构成一个有效的部分匹配，且还剩下 p 个未匹配元素，就用 p 阶块中的元素以 $p!$ 种方式将它们匹配。

因此，系数不超过 $(L+1)^{L-l_k} \le (2L)^{L-l_k}$。 ∎

命题 *5.4.31*。 *</style}*对于形状 α，令 $\alpha \pm e$ 表示添加或删除边 e 后得到的形状。那么

$$\frac{|\mathrm{Aut}(\alpha \pm e)|}{|\mathrm{Aut}(\alpha)|} \;\le\; |V(\alpha)|^2 .$$

证明。 我们证明这两个群有一个相同的大子群。将 $\mathrm{Aut}(\alpha \pm e)$ 和 $\mathrm{Aut}(\alpha)$ 视为对集合 $\binom{V}{2}^{\alpha}$ 的群作用。令 G^e 表示边 e 的稳定子，注意到 $\mathrm{Aut}(\alpha \pm e)^e = \mathrm{Aut}(\alpha)^e$。根据轨道-稳定子引理，指数 $|G : G^e|$ 等于 e 的轨道大小，该大小至少为 1，至多为 $|V(\alpha)|^2$。因此，

$$\frac{|\mathrm{Aut}(\alpha \pm e)|}{|\mathrm{Aut}(\alpha)|} = \frac{|\mathrm{Aut}(\alpha \pm e) : \mathrm{Aut}(\alpha \pm e)^e|}{|\mathrm{Aut}(\alpha) : \mathrm{Aut}(\alpha)^e|} \le |V(\alpha)|^2 . \qquad ∎$$

引理 5.4.32. 如果 α 是一个左蜘蛛，则对于每个 $\beta \in \mathcal{I}_\alpha$，都存在系数 c_β，使得

$$L_{|U_\alpha|} \cdot M_{\mathrm{body}(\alpha)} = 2M_\alpha + \sum_{\beta \in \mathcal{I}_\alpha} c_\beta M_\beta,$$

$$|c_\beta| \le \begin{cases} 160\,|V(\alpha)|^7\,|E(\alpha)|^2 & \beta \in \mathcal{I}_\alpha^{(1)} \\ \frac{160|V(\alpha)|^7|E(\alpha)|^2}{n} & \beta \in \mathcal{I}_\alpha^{(2)} \end{cases}.$$

~证明。 我们将引理 5.4.27 中的每个 $M_\beta, \beta \in \mathcal{I}_\alpha$ 都表示为正规形状。我们利用以下关于 $C_{Fourier}$ 和 C_{Aut} 的界应用命题 5.4.29。 β 中唯一的不正规之处来自合并（至多）两个端点，而这两个端点各自仅关联一条

边。因此，任意平行边的标签集合要么是 $\{1, k\}$，要么是 $\{1, 1, k\}$，对于某个 $k \leq |E(\alpha)|$。根据命题 5.4.30，我们有 $C_{Fourier} \leq 4\,|E(\alpha)|^2$。$\beta$中至多有两条额外的平行边，因此根据命题 5.4.31，我们有 $C_{Aut} \leq |V(\alpha)|^4$。所以，系数至多增加 $C_{Fourier} \cdot C_{Aut} \leq 4\,|E(\alpha)|^2\,|V(\alpha)|^4$。

■

推论 5.4.33。 如果 α 是一个右蜘蛛形，则存在系数 c_β，其界与引理 *5.4.32* 中给出的相同，使得

$$M_{\mathrm{body}(\alpha)} \cdot L^{\mathsf{T}}_{|U_\alpha|} = 2M_\alpha + \sum_{\beta \in \mathcal{I}_\alpha} c_\beta M_\beta.$$

推论 5.4.34。 如果 $x \perp$ 零空间$(\mathcal{M}_{fix})$ 且 α 是一个蜘蛛，则对于某个 c_β，其界与引理 *5.4.32* 中给出的相同，

$$x^\top (M_\alpha - \sum_{\beta \in \mathcal{I}_\alpha} c_\beta M_\beta) x = 0$$

证明。 对于左蜘蛛，由于

$$\mathcal{M}_{fix}(2M_\alpha + \sum_{\beta \in \mathcal{I}_\alpha} c_\beta M_\beta) = \mathcal{M}_{fix} \cdot L_{|U_\alpha|} \cdot M_{\alpha'} = 0$$

我们现在可以使用事实 5.2.1。对于右蜘蛛形，证明类似。 ■

5.4.3 消去所有蜘蛛

策略是从矩矩阵 $\mathcal{M}$ 开始，反复应用推论 5.4.34，直到分解中不再有蜘蛛。对于每只蜘蛛，通过推论 5.4.34 将其消去后，只会留下交项。其中一些交项本身可能是更小的蜘蛛，此时我们会一再应用该推论，直到只剩下非蜘蛛。这个过程中难点在于如何界定每个非蜘蛛上累积的总系数

为了描述这一过程，我们定义蜘蛛 α的网，它是一个有向无环图，用来描述蜘蛛消除过程。为便于区分，我们将该图的顶点称为“节点”。

定义 5.4.35（网 α）。蜘蛛 α 的网 W（α）是一个有根有向无环图，其节点为形状，根节点为 α。每个蜘蛛节点 γ 都有边连向节点 β，每个形状 $\beta \in \mathcal{I}_\gamma$ 对应一个节点。非蜘蛛节点是该有向无环图的叶节点／汇点。

备注 5.4.36。有向无环图结构的出现，是因为对于任意蜘蛛 γ， $\mathcal{I}_\gamma$ 中的每个形状的方形顶点数都严格少于 γ 。因此，网 W（α）| 的高度至多为 $|V$（α）。

W（α）的每个节点 γ 也有一个关联值 v_γ，该值通过以下过程定义：

- 首先，设置 $v_\alpha = 1$，并将所有其他 γ 设为 $v_\gamma = 0$。
- 从根节点开始，按拓扑序处理；每个蜘蛛节点 γ 都会将 $v_\gamma c_\beta$ 加到每个子形状 $\beta \in \mathcal{I}_\gamma$ 对应的 v_β 上，其中 c_β 是推论 5.4.34 中的系数。

命题 5.4.37。若 $x \perp$ 零空间($\mathcal{M}_{fix}$)，则

$$x^\intercal (M_\alpha - \sum_{leaves\ \gamma\ of\ W(\alpha)} v_\gamma M_\gamma) x = 0.$$

证明。从方程 $x\ M_\alpha x = x\ v_\alpha M_\alpha x$ 开始。在每一步中，我们选取拓扑顺序最靠前的蜘蛛 γ；在此情形下，这指的是 $W(\alpha)$ 中最靠近根节点的蜘蛛。若该蜘蛛出现在方程右侧，我们便利用推论 5.4.34，将 $v_\gamma M_\gamma$ 替换为 $\sum_{\beta \in \text{子节点}(\gamma)} v_\gamma c_\beta M_\beta$。根据 v_γ 的定义，这一过程最终会得到方程

$$x^\intercal M_\alpha x = x^\intercal (\sum_{\text{leaves } \gamma \text{ of } W(\alpha)} v_\gamma M_\gamma) x$$

■

命题 5.4.38。 对于 $W(\alpha)$ 中的任意节点 β， $|\text{父节点}(\beta)| \leq 4|V(\alpha)|^3 \cdot |E(\alpha)|^2$，其中 $parents(\beta)$ 是 $W(\alpha)$ 中满足 $\beta \in \mathcal{I}_\gamma$ 的节点 γ 所组成的集合。

证明。 以下过程涵盖所有可能将其端点塌缩以形成 β 的父左蜘蛛 γ。从 $\gamma = \beta$ 开始，

- 选择一个圆形顶点 $u \in V(\gamma)$ 作为端点的邻点。
- 选择一个方形顶点 $i \in V(\gamma)$ 作为第一个端点塌缩后形成的顶点。通过在 U_γ 中添加一个新的方形，并用一条标签为 1 的边将其连接到 u，将它“反塌缩”。翻转 U_γ 的值（i）。将 $\{i, u\}$ 的标签修改为不超过 $|E(\alpha)|$ 的任意数。
- 选择一个方形顶点 $j \notin V(\gamma)$ 作为第二个端点。也可以按照上述相同方式，在 γ 中添加一个新的方形，将其反塌缩。

该过程最多可以通过 $|V(\alpha)|^3 |E(\alpha)| (|E(\alpha)| + 1) \leq 2|V(\alpha)|^3 |E(\alpha)|^2$ 种方式执行。乘以 2 是为了将右蜘蛛也考虑在内。

■

我们将每条父子边（γ, β）标记为“类型 1”边（若满足 $\beta \in \mathcal{I}_{(1)}\gamma$），或“类型 2”边（若满足 $\beta \in \mathcal{I}_{(2)}\gamma$）。

命题 5.4.39。 设 p 是 $W(\alpha)$ 中的一条路径，其中有 $\#_1(p)$ 条类型 *1* 边和 $\#_2(p)$ 条类型 *2* 边。那么，$\#_1(p) \leq |E(\alpha)| + 2\#_2(p)$。

证明。 对于一个形状 γ，令 S_γ 为 γ 中方形顶点的集合。那么， $S_\gamma \cap W_\gamma$ 将是 γ 中属于正方形的中间顶点集合。我们断言，量 $\left|\mathcal{S}_\gamma \cap W_\gamma\right| + \left|U_\gamma \setminus (U_\gamma \cap V_\gamma)\right| + \left|V_\gamma \setminus (U_\gamma \cap V_\gamma)\right|$ 在塌缩过程中会减少。

固定一对相邻的形状（γ, β），它们构成一条类型 1 边。根据 $\mathcal{I}$（1）γ 的定义，每个端点要么（1）不与任何对象塌缩，要么（2）与 W_γ 的一个顶点塌缩，要么（3）与 $V_\gamma \setminus U_\gamma$ 的一个顶点塌缩（如果 γ 是左蜘蛛；若是右蜘蛛形，则为 $U_\gamma \setminus V_\gamma$）。此外，情形（2）或（3）必须至少发生在一个端点上，而且它们不会彼此塌缩。

如果出现情形（2），那么$|\mathcal{S}_\beta \cap W_\beta| < |\mathcal{S}_\gamma \cap W_\gamma|$而 $|U_\beta \setminus (U_\beta \cap V_\beta)| = |U_\gamma \setminus (U_\gamma \cap V_\gamma)|$且 $|V_\beta \setminus (U_\beta \cap V_\beta)| = |V_\gamma \setminus (U_\gamma \cap V_\gamma)|$。另一方面，如果出现情形（3），那么 $W_\beta = W_\gamma$而 $|U_\beta \setminus (U_\beta \cap V_\beta)| < |U_\gamma \setminus (U_\gamma \cap V_\gamma)|$且 $|V_\beta \setminus (U_\beta \cap V_\beta)| < |V_\gamma \setminus (U_\gamma \cap V_\gamma)|$。在所有情形下，$|\mathcal{S}_\beta \cap W_\beta| + |U_\beta \setminus (U_\beta \cap V_\beta)| + |V_\beta \setminus (U_\beta \cap V_\beta)| < |\mathcal{S}_\gamma \cap W_\gamma| + |U_\gamma \setminus (U_\gamma \cap V_\gamma)| + |V_\gamma \setminus (U_\gamma \cap V_\gamma)|$，如所需。

现在我们对 α 的这个表达式进行界定。根据 $\mathcal{L}$ 的定义，即定义 5.3.6，在伪校准中出现的蜘蛛，其 W_α, $U_\alpha \setminus (U_\alpha \cap V_\alpha)$ 和 $V_\alpha \setminus (U_\alpha \cap V_\alpha)$ 中的方形顶点的度至少为 1，且只能与圆形顶点相连。因此，它们的数量至多为 $|E(\alpha)|$。所以，最初有 $|\mathcal{S}_\alpha \cap W_\alpha| + |U_\alpha \setminus (U_\alpha \cap V_\alpha)| + |V_\alpha \setminus (U_\alpha \cap V_\alpha)| \leq |E(\alpha)|$。

最后，p 中的每条类型 2 边至多使量 $|\mathcal{S}_\gamma \cap W_\gamma| + |U_\gamma \setminus (U_\gamma \cap V_\gamma)| + |V_\gamma \setminus (U_\gamma \cap V_\gamma)|$ 增加 2。因此，我们得到所需的不等式 $\#_1(p) \leq |E(\alpha)| + 2\#_2(p)$。

■

推论 5.4.40. $\#_2(p) \geq \frac{|p|}{3} - \frac{|E(\alpha)|}{3}$。

证明。 代入 $|p| = \#_1(p) + \#_2(p)$ 并重新整理。 ■

最后，我们可以对每个非蜘蛛项的累积量进行界定，该项只依赖于蜘蛛 α 的参数。

引理 5.4.41. 存在绝对常数 C_1、C_2，使得对于 $W(\alpha)$ 的所有叶节点 γ，

$$|v_\gamma| \leq (C_1 \cdot |V(\alpha)| \cdot |E(\alpha)|)^{C_2|E(\alpha)|}.$$

证明。 为了界定 $|v_\gamma|$，我们将对从 α 出发并以 γ 结束的 $W(\alpha)$ 中所有路径 $p = (\beta_0 = \alpha, \ldots, \beta_r = \gamma)$ 的贡献求和。这条路径对 v_γ 的贡献为系数 c_β 的乘积。

注记 5.4.42. 这里重要的是，类型 *2* 边的系数界更强 $|c_\beta| \leq C \cdot (|V(\alpha)|\,|E(\alpha)|)^{O(1)/n} \ll 1$。

在开始证明之前，我们先引入一些方便的记号，并回顾若干事实。对于连续的形状 β_{i-1},β_i （即， β_i 是 β_{i-1} 的子形状），我们用 c_{β_i} 表示将推论 5.4.34 应用于 β_{i-1} 所得的系数。根据命题 5.4.38，$W(\alpha)$ 的入度可界为 $B_1\cdot(|V(\alpha)|\,|E(\alpha)|)^{B_2}$，其中常数为 B_1，B_2。因此，以 γ 结尾的长度为 r 的路径数至多为 $(B_1\,|V(\alpha)|\,|E(\alpha)|)^{B_2 r}$。利用推论 5.4.34，取 B_1,B_2 足够大，使得对于类型 1 边，c_{β_i} 至多为 $B_1\cdot(|V(\alpha)|\,|E(\alpha)|)^{B_2}$（对于类型 2 边，则至多为 $B_1\cdot(|V(\alpha)|\,|E(\alpha)|)^{B_2}/n$）。

$$
\begin{aligned}
|v_\gamma| &\le \sum_{r=0}^{\infty} \sum_{\substack{p=(\beta_0=\alpha,\dots,\beta_r=\gamma)\\ \text{path from } \alpha \text{ to } \gamma \text{ in } W(\alpha)}} \prod_{i=1}^{r} |c_{\beta_i}| \\
&\le \sum_{r=0}^{\infty} \sum_{\substack{p=(\beta_0=\alpha,\dots,\beta_r=\gamma)\\ \text{path from } \alpha \text{ to } \gamma \text{ in } W(\alpha)}} \left(B_1\cdot(|V(\alpha)|\,|E(\alpha)|)^{B_2}\right)^{\#_1(p)} \left(B_1\cdot(|V(\alpha)|\,|E(\alpha)|)^{B_2}/n\right)^{\#_2(p)} \\
&\le \sum_{r=0}^{\infty} \sum_{\substack{p=(\beta_0=\alpha,\dots,\beta_r=\gamma)\\ \text{path from } \alpha \text{ to } \gamma \text{ in } W(\alpha)}} \left(B_1\cdot(|V(\alpha)|\,|E(\alpha)|)^{B_2}\right)^{|E(\alpha)|+2\#_2(p)} \left(B_1\cdot(|V(\alpha)|\,|E(\alpha)|)^{B_2}/n\right)^{\#_2(p)} \\
&= \sum_{r=0}^{\infty} \sum_{\substack{p=(\beta_0=\alpha,\dots,\beta_r=\gamma)\\ \text{path from } \alpha \text{ to } \gamma \text{ in } W(\alpha)}} \left(B_1\cdot(|V(\alpha)|\,|E(\alpha)|)^{B_2}\right)^{|E(\alpha)|} \left(B_1'\cdot(|V(\alpha)|\,|E(\alpha)|)^{B_2'}/n\right)^{\#_2(p)}
\end{aligned}
$$

对于某些常数 B_1',B_2'，第一个不等式由推论 5.4.34 得出，第二个不等式由命题 5.4.39 得出。我们将上述求和拆成两部分，$r\le 3|E(\alpha)|$ 和 $r>3|E(\alpha)|$。对于 $r\le 3\,|E(\alpha)|$，将 #2(p) 项上界为 1，并将路径数上界为 $(B_1\,|V(\alpha)|\,|E(\alpha)|)^{B_2 r}$，可得界 $(B_1''\,|V(\alpha)|\,|E(\alpha)|)^{B_2''|E(\alpha)|}$，其中 B_1'',B_2'' 是某些常数。对于较大的 r，我们利用推论 5.4.40 得到 #2(p) $\ge r/9=|E(\alpha)|/3$ 的下界。应用相同的路径数上界，较大 r 对应项的总贡献利用分母中 n 的幂可界为 1（假设 δ,τ 足够小）。

■

我们将消除所有蜘蛛的结果定义为一个新矩阵 $\mathcal{M}^+$。

定义 5.4.43。 将矩阵 $\mathcal{M}^+$ 定义为消除所有蜘蛛的结果，

$$\mathcal{M}^+ := \mathcal{M} - \sum_{spiders\ \alpha} \lambda_\alpha \left(M_\alpha - \sum_{leaves\ \gamma\ of\ W(\alpha)} v_\gamma M_\gamma \right)$$

5.4.4 完成证明

证明的最后一步是说明，蜘蛛消除过程完成后，在 $\mathcal{M}^+$ 中新产生的非蜘蛛项也具有较小的范数。为此，我们希望证明一个类似于推论 5.4.11 的结论。在该证明中，我们利用了 $\mathcal{L}$ 中非蜘蛛项的特殊结构性质，证明了伪校准中的非蜘蛛项可以忽略。但现在， $\mathcal{M}^+$ 中的非蜘蛛项未必具有 $\mathcal{L}$ 的性质——例如，可能存在度为 2 的圆形顶点或孤立顶点。为处理可能更大的范数，我们将利用这样一点：这些新非蜘蛛项的系数 β 带有它们所在网中蜘蛛项 α 的系数 λ_α。由于 α 的顶点/边比 β更多， λ_α 中 $\frac{1}{n}$ 的幂大于$\eta^{|U_\beta|+|V_\beta|} \cdot \frac{1}{n^{|E(\beta)|/2}}$的“预期伪校准”系数。我们通过仔细的收费论证证明，这些额外的 $\frac{1}{n}$ 因子足以抵消孤立顶点或较小的顶点分隔集带来的影响。

引理 5.4.44. 如果 β 是一个非平凡非蜘蛛，并且对于某个蜘蛛 $\alpha \in \mathcal{L}$， $\beta \in W(\alpha)$，那么

$$\eta^{|U_\alpha|+|V_\alpha|} \cdot \frac{1}{n^{|E(\alpha)|/2}} \cdot n^{\frac{w(V(\beta))-w(S_{\min})+w(W_{iso})}{2}} \leq \eta^{|U_\beta|+|V_\beta|} \cdot \frac{1}{n^{\Omega(\varepsilon|E(\alpha)|)}}$$

其中， S_{min} 和 W_{iso} 分别是 β 的最小顶点分隔集，以及 $V(\beta) \setminus (U_\beta \cup V_\beta)$ 的孤立顶点集。

证明。 我们先说明证明思路。假设我们尝试使用与证明引理 5.4.7 相同的分配方案。由于两个原因，这行不通。首先，根据注 5.4.18， β 中的圆形顶点仍然具有偶数度，但现在它们的度可能为 0 或 2。要使之前的分配方案奏效，我们需要它们

的度数至少为 4，这样才能产生必要的边衰减，以处理范数界。其次，方形顶点现在可能具有度 0，因此无法从边获得任何衰减。

第一个问题相对容易处理。由于 β 是通过塌缩 α 得到的，β 中度为 0 或 2 的圆形顶点，起初在 α 中的度数必定至少为 4。因此，我们可以固定从 α 到 β 的一个特定收缩序列，然后为了分析，假设被移除的边仍然存在。在这种情况下，引理 5.4.7 中相同的收费论证依然成立。通过考察这条收缩链中的非正规塌缩序列，可以将这一点形式化。

为处理第二个问题，我们来更仔细地分析度为 0 的方形顶点是如何出现的。固定从 α 到 β 的一个收缩序列，并考虑其中一个特定步骤：γ 塌缩到 γ'，并产生了一个度为 0 的方形顶点。设在 γ 中塌缩的两个方形顶点为 i, j，并设在 γ' 中产生的度为 0 的方形顶点为 k。根据注 5.4.18，由于 k 的度为 0，它必定不在 $(U_{\gamma'} \cup V_{\gamma'}) \setminus (U_{\gamma'} \cap V_{\gamma'})$ 中，因此，$U_{\gamma'}(k) = V_{\gamma'}(k) = 0$ 或 $U_{\gamma'}(k) = V_{\gamma'}(k) = 1$。不过在后一种情形下，该顶点不会影响范数界，因为它位于 $U_{\gamma'} \cap V_{\gamma'}$ 中，所以可以放心地忽略。请注意，它不一定会始终留在这个集合中，因为后续塌缩可能会将该顶点塌缩掉；但这并不成问题，因为如果发生这种塌缩，我们可以为此收费。

所以，假设我们有 $U_{\gamma'}(k) = V_{\gamma'}(k) = 0$。但根据收缩定义，$i$ 或 j 中至少有一个曾出现在 $U_\gamma \setminus (U_\gamma \cap V_\gamma)$ 或 $V_\gamma \setminus (U_\gamma \cap V_\gamma)$ 中。同样根据收缩定义，我们有 $U_{\gamma'}(k) = U_\gamma(i) + U_\gamma(j)$(模 2)，且 $V_{\gamma'}(k) = V_\gamma(i) + V_\gamma(j)$(模 2)。综合这些结果，我们立即得到，这种情况发生的唯一可能是 $i, j \in U_\gamma \setminus (U_\gamma \cap V_\gamma)$，或者 $i, j \in V_\gamma \setminus (U_\gamma \cap V_\gamma)$。

当这种塌缩发生时，注意到 $|U_\gamma| + |V_\gamma| \geq |U_{\gamma'}| + |V_{\gamma'}| + 2$。这正是归一化因子 $\eta = \frac{1}{\sqrt{n}}$ 开始发挥衰减作用的地方。这个不等式意味着，与“预期的伪校准”系数 β 相比，我们可以多获得一个衰减因子 $\eta^2 = \frac{1}{n}$。我们将用这个因子来抵扣新的

0 度方形顶点。

现在，我们将这些想法形式化。

令 $Q = U_\beta \cap V_\beta$，$P = (U_\beta \cup V_\beta) \setminus Q$，并令 P' 为 β 中不属于 S_{min} 的 1 度方形顶点集合。令 s_0 为 $V(\beta) \setminus Q$ 中的 0 度方形顶点数。$P' \cup Q \cup S_{min}$ 之外的所有方形顶点的度数至少为 2，设其数量为 $s_{\geq 2}$。

根据奇偶性约束、注 5.4.18，以及 $U_\beta \cup V_\beta$ 中没有圆形顶点这一事实，β 中的所有圆形顶点的度都是偶数。令 c_0 为 β 中的 0 度圆形顶点数。令 c_2，$c_{\geq 4}$ 分别为 $V(\beta) \setminus S_{min}$ 中的 2 度圆形顶点数和度数至少为 4 的圆形顶点数。那么，我们有

$$n^{\frac{w(V(\beta))-w(S_{\min})+w(W_{iso})}{2}} \leq n^{\frac{|P'|+s_{\geq 2}+(1.5-\varepsilon)(c_2+c_{\geq 4})}{2}} \cdot n^{s_0+(1.5-\varepsilon)c_0}$$

利用 $\eta = \frac{1}{\sqrt{n}}$，只需证明

$$\begin{aligned} &|E(\alpha)| + (|U_\alpha| + |V_\alpha| - |U_\beta| - |V_\beta|) \\ &\qquad \geq |P'| + s_{\geq 2} + (1.5-\varepsilon)(c_2 + c_{\geq 4}) + 2s_0 + 2(1.5-\varepsilon)c_0 + \Omega(\varepsilon\,|E(\alpha)|) \end{aligned}$$

将 α 塌缩为 β 的方式可能有很多种，任选一种。我们首先对 0 度方形顶点进行收费论证。

引理 5.4.45. $|U_\alpha| + |V_\alpha| - |U_\beta| - |V_\beta| \geq 2s_0$

证明。在塌缩过程中，每一步中，中间形状 γ 中度为 1 的顶点 $i \in U_\gamma \setminus (U_\gamma \cap V_\gamma)$ 或 $i \in V_\gamma \setminus (U_\gamma \cap V_\gamma)$ 会与另一个方形顶点 k 塌缩。我们有，只有当 i 与 $k \in U_\gamma$（分别为 $k \in V_\gamma$）塌缩时，$|U_\gamma| + |V_\gamma|$ 才会减少。无论哪种情况，该量恰好减少 2，我们将其分配给这个新合并的顶点。在 $V(\beta)$ 中，每个度为 0 的方形顶点 $\in Q$ 必定是由一次塌缩产生的，因此至少有 2 的加性量分配给它。这就证明了 $|U_\alpha| + |V_\alpha| - |U_\beta| - |V_\beta| \geq 2s_0$。

■

现在我们将证明一个结构引理。

引理 5.4.46. 在 $V(\beta) \setminus S_{min}$ 中度至少为 2 的任意顶点 u，至多与 P' 中的 1 个顶点相邻。

证明. 注意到，u 不可能与 P' 中的 3 个顶点相邻；否则，其中至少 2 个会在 $U_\beta \setminus Q$ 或 $V_\beta \setminus Q$ 中，这意味着 β 会是一个蜘蛛，矛盾。若 u 与 P' 中的 2 个顶点相邻，则其中一个分别在 $U_\beta \setminus Q$ 中，另一个在 $V_\beta \setminus Q$ 中。由于这两个顶点都不在 S_{min} 中，因此 u 位于 S_{min} 中，因为不存在一条从 U_β 到 V_β 且不经过 S_{min} 的路径。这是矛盾。因此，u 至多与 P' 中的 1 个顶点相邻。

■

这个引理立即推出 $|P'| \leq c_2 + c_{\geq 4}$。

为了计入 α 中不属于 β的边，我们令 β 为将α 不当收缩到 β后得到的结果；注意，$|E(\alpha)| = \left|E(\widetilde{\beta})\right|$。我们将正规收缩时消失的边称为“幻影”边。令 $\deg_{\widetilde{\beta}}(i)$（分别为 $\deg_{\widetilde{\beta}}(u)$）表示 i（分别为 u）在 β中的顶点度。注意，$V(\beta)$中的任意圆形顶点 u都有 $deg_{\widetilde{\beta}}(u) \geq 4$。

引理 5.4.47. $|E(\alpha)| \geq |P'| + s_{\geq 2} + (1.5 - \varepsilon)(c_2 + c_{\geq 4}) + 2(1.5 - \varepsilon)c_0 + \Omega(\varepsilon |E(\alpha)|)$

证明。我们将使用以下分摊方案。与 P' 相邻的 β 的每条边都向相邻的方形顶点分摊 1，该顶点属于 P'。β 的每条其他边都向相邻的方形顶点分摊$\frac{1}{2} - \frac{\varepsilon}{10}$，并向相邻的圆形顶点分摊$1 - \frac{\varepsilon}{10}$。每条虚边都向相邻的圆形顶点 u分摊。因此，总计有 $\frac{\varepsilon}{10}(|E(\alpha)| - |P'|)$尚未分摊。

P' 中的所有方形顶点都已分配到值 1。并且注意到，β 中所有度数至少为 2 的方形顶点，都从 β的关联边中至少分配到 1，总值为 $s_{\geq 2}$。因此，方形顶点获得的分配总值至少为 $|P'| + s_{\geq 2}$。

考虑 $V(\beta)$ 中的任意一个度数为 0 的圆顶点 u。它必定至少关联 4 条虚边，因此必须至少分配到值 $4(1-\frac{\varepsilon}{10}) > 2(1.5-\varepsilon)$。因此，$V(\beta)$ 中的度数为 0 的圆顶点获得的分配总值至少为 $2(1.5-\varepsilon)c_0$。

假设 u 在 $V(\beta)$ 中的度为 2。那么，它至少与 2 条虚边关联。根据引理 5.4.46，它至多还与 P' 中的一个顶点相邻，因此，它必定被分配了至少 $2(1-\frac{\varepsilon}{10}) + (deg_{\widetilde{\beta}}(u) - 3)(\frac{1}{2}-\frac{\varepsilon}{10})$ 的值。这至少为 $1.5-\varepsilon+\frac{\varepsilon}{10}$。

假设 u 在 $V(\beta)$ 中的度至少为 4。根据引理 5.4.46，它至多与 P' 中的一个顶点相邻。因此，它必定被分配了至少 $(deg_{\widetilde{\beta}}(u)-1)(\frac{1}{2}-\frac{\varepsilon}{10})$ 的值。利用 $deg_{\widetilde{\beta}}(u) \geq 4$，这至少为 $1.5-\varepsilon+\frac{\varepsilon}{10}$。

这意味着

$$|E(\alpha)| \geq |P'| + s_{\geq 2} + 2(1.5-\varepsilon)c_0 + (1.5-\varepsilon+\frac{\varepsilon}{10})(c_2+c_{\geq 4}) + \frac{\varepsilon}{10}(|E(\alpha)|-|P'|)$$

利用 $|P'| \leq c_2 + c_{\geq 4}$ 即可完成证明。 ■

将引理 5.4.45 和引理 5.4.47 相加，得到所需结果。 ■

推论 5.4.48. 若 β 是一个非平凡非蜘蛛，且对于某个蜘蛛 $\alpha \in \mathcal{L}$，有 $\beta \in W(\alpha)$，则

$$\eta^{|U_\alpha|+|V_\alpha|} \cdot \frac{1}{n^{|E(\alpha)|/2}} \left\|M_\beta\right\| \leq \eta^{|U_\beta|+|V_\beta|} \cdot \frac{1}{n^{\Omega(\varepsilon|E(\alpha)|)}}$$

证明. 由引理 5.6.3 可得

$$\left\|M_\beta\right\| \leq 2 \cdot (|V(\beta)| \cdot (1+|E(\beta)|) \cdot \log(n))^{C\cdot(|V_{rel}(\beta)|+|E(\beta)|)} \cdot n^{\frac{w(V(\beta))-w(S_{\min})+w(W_{iso})}{2}}$$

我们有 $|V(\beta)|\cdot(1+|E(\beta)|)\cdot\log(n) \leq n^{O(\tau)}$。此外，$|V_{rel}(\beta)| \leq 2(|E(\alpha)|+|E(\beta)|)$，因为 $V_{rel}(\beta)$ 中所有的零度顶点在收缩链中本应有 $V_{rel}(\alpha)$ 的顶点塌缩到其中，而 $V_{rel}(\alpha)$ 中没有零度顶点。最后，由于 $|E(\alpha)| \geq |E(\beta)|$，因子 $2\cdot(|V(\beta)|\cdot(1+|E(\beta)|)\cdot\log(n))^{C\cdot(|V_{rel}(\beta)|+|E(\beta)|)}$ 可以吸收到 $n^{\frac{1}{\varepsilon}|E\Omega(\varepsilon|E(\alpha)|)}$ 中。

结论由引理 5.4.44 得出。 ■

命题 5.4.49。 如果 β 是平凡形状，则 $\lambda_\beta^+ = \lambda_\beta$。

证明。 对于任何 α，平凡形状都不可能出现在 $W(\alpha)$ 中，因为蜘蛛的每次塌缩都会保留其圆形顶点。 ■

引理 5.4.50。 对于 $k, l \in \{0, 1, \ldots, D/2\}$，令 $\mathcal{B}_{k,l}$ 表示块 (k, l) 上非平凡非蜘蛛的集合。那么

$$\sum_{\beta \in \mathcal{B}_{k,l}} \left|\lambda_\beta^+\right| \left\|M_\beta\right\| \leq \eta^{k+l} \cdot \frac{1}{n^{\Omega(\varepsilon)}}$$

证明。

$$\sum_{\beta \in \mathcal{B}_{k,l}} \left\|\lambda_\beta^+ M_\beta\right\| \leq \sum_{\beta \in \mathcal{B}_{k,l}} \left|\lambda_\beta\right| \left\|M_\beta\right\| + \sum_{\beta \in \mathcal{B}_{k,l}} \sum_{\substack{\text{spiders } \alpha: \\ \beta \in W(\alpha)}} \left|v_\beta\right| \left|\lambda_\alpha\right| \left\|M_\beta\right\|$$

为了界定第一项，我们此前在推论 5.4.14 中验证过，伪校准中出现的非平凡非蜘蛛的总范数（即这一项）为 $\eta^{k+l} o_n(1)$。对于第二项，根据引理 5.4.41，我们有一个关于一只蜘蛛在一个非蜘蛛上的累积量 v_γ 的上界，因此它至多为

$$\leq \sum_{\beta \in \mathcal{B}_{k,l}} \sum_{\substack{\text{spiders } \alpha: \\ \beta \in W(\alpha)}} \left(C_1 \left|V(\alpha)\right| \cdot \left|E(\alpha)\right|\right)^{C_2 |E(\alpha)|} \cdot \left|\lambda_\alpha\right| \left\|M_\beta\right\|.$$

利用系数的界 $|\lambda_\alpha|$，命题 5.4.13，

$$\leq \sum_{\beta \in \mathcal{B}_{k,l}} \sum_{\substack{\text{spiders } \alpha: \\ \beta \in W(\alpha)}} \left(C_1 \left|V(\alpha)\right| \cdot \left|E(\alpha)\right|\right)^{C_2 |E(\alpha)|} \cdot \eta^{|U_\alpha| + |V_\alpha|} \cdot \frac{\left|E(\alpha)\right|^{3|E(\alpha)|}}{n^{|E(\alpha)|/2}} \cdot \left\|M_\beta\right\|$$

应用塌缩的非蜘蛛的范数界，推论 5.4.48,

$$\leq \eta^{k+l} \cdot \sum_{\beta \in \mathcal{B}_{k,l}} \sum_{\substack{\text{spiders } \alpha: \\ \beta \in W(\alpha)}} \left(\frac{C_1 \, |V(\alpha)| \cdot |E(\alpha)|}{n^{\Omega(\varepsilon)}} \right)^{C_2' |E(\alpha)|}$$

$$\leq \eta^{k+l} \cdot \sum_{\beta \in \mathcal{B}_{k,l}} \sum_{\substack{\text{spiders } \alpha: \\ \beta \in W(\alpha)}} \left(\frac{C_1 n^{\tau} \cdot n^{\tau}}{n^{\Omega(\varepsilon)}} \right)^{C_2' |E(\alpha)|} .$$

将所有蜘蛛的和用所有形状的和来界定。根据命题 5.4.12，含有 i 条边的形状数量为 $n^{O(\tau(i+1))}$。按边数求和，注意到 $|E(\alpha)| \geq \max(|E(\beta)|, 2)$，因为蜘蛛总是至少有 2 条边。

$$\leq \eta^{k+l} \sum_{\beta \in \mathcal{B}_{k,l}} \sum_{i=\max(|E(\beta)|,2)}^{\infty} n^{O(\tau(i+1))} \cdot \left(\frac{C_1 n^{\tau} \cdot n^{\tau}}{n^{\Omega(\varepsilon)}} \right)^{C_2' i}$$

$$\leq \eta^{k+l} \sum_{\beta \in \mathcal{B}_{k,l}} \frac{1}{n^{\Omega(\varepsilon \max(|E(\beta)|,2))}}$$

$$\leq \eta^{k+l} \sum_{i=0}^{\infty} \frac{n^{O(\delta(i+1))}}{n^{\Omega(\varepsilon \max(i,2))}}$$

$$= \eta^{k+l} \cdot \frac{1}{n^{\Omega(\varepsilon)}} \qquad \blacksquare$$

推论 5.4.51. 对于 $k \in \{0, \ldots, D/2\}$, $\mathcal{M}^+$ 的 (k,k) 块的最小奇异值至少为 $\eta^{2k}(1 - n^{-\Omega(\varepsilon)})$；对于 k、$l \in \{0, \ldots, D/2\}$、$l \neq k$，(k,l) 非对角块的范数至多为 $\eta^{k+l} \cdot n^{-\Omega(\varepsilon)}$。

证明。 根据命题 5.4.49，单位矩阵以系数 η^{2k} 出现在 (k,k) 块上。根据构造， $\mathcal{M}^+$ 不含蜘蛛形状。根据引理 5.4.50，(k,l) 块上非蜘蛛形状的总范数至多为 $\eta^{k+l} \cdot n^{-\Omega(\varepsilon)}$。

$\blacksquare$

定理 5.4.52。以高概率 $\mathcal{M}_{fix} \succeq 0$。

证明。 对于任意 $x \in$零空间$(\mathcal{M}_{fix})$，我们当然有 $x^{\intercal} \mathcal{M}_{fix} x = 0$。对于任意 $x \perp$零空间$(\mathcal{M}_{fix})$

且 $\|x\|_2 = 1$,

$$\begin{aligned} x^\intercal \mathcal{M}_{fix} x &= x^\intercal (\mathcal{M} + \mathcal{E}) x \\ &= x^\intercal \mathcal{M}^+ x + x^\intercal \left(\sum_{\text{spiders } \alpha} \lambda_\alpha \left(\mathcal{M}_\alpha - \sum_{\text{leaves } \gamma \text{ of } W(\alpha)} v_\gamma M_\gamma \right) \right) x + x^\intercal \mathcal{E} x \\ &= x^\intercal (\mathcal{M}^+ + \mathcal{E}) x \end{aligned}$$

其中最后一个等式由命题 5.4.37 得出。由于 $\mathcal{E}$ 的范数界远小于 $\eta^D = n^{-n^\delta}$ （见 [87]），推论 5.4.51 中对 $\mathcal{M}^+$每个块的范数界也适用于 $\mathcal{M}^+ + \mathcal{E}$的各块。因此，我们利用引理 5.4.2 得出 $\mathcal{M}^+ + \mathcal{E} \succeq 0$，且上述表达式非负。

■

5.5 Sherrington-Kirkpatrick 下界

在此，我们证明定理 4.1.5 和定理 4.1.2。

回顾一下，在植入布尔向量问题中，我们希望优化

$$\mathrm{OPT}(V) := \frac{1}{n} \max_{b \in \{\pm 1\}^n} b^\intercal \Pi_V b,$$

其中， V 是 $\mathbb{R}^n$的一个均匀随机 p维子空间。

定理 4.1.5。 存在一个常数 $c > 0$，使得对所有 $\varepsilon > 0$ 和 $\delta \le c\varepsilon$，当 $p \ge n^{2/3+\varepsilon}$时，关于 V，以高概率存在一个度为 n^δ 的植入布尔向量问题的平方和解，其值为 1。

~证明。 我们希望构造一个关于布尔变量 $b_1, \ldots, b_n$ 的平方和解 $\mathbb{E}$，使得~ $\mathbb{E}[b_\intercal \Pi_V b] = n$。 我们不直接从 $\mathbb{R}^n$中均匀随机抽取一个 p 维子空间 V，而是先从 $\mathcal{N}(0, I)$ 中抽取 $d_1, \ldots, d_n$ 个独立同分布的 p 维高斯向量，然后构造一个n×p 矩阵 A，其行是 $d_1, \ldots, d_n$，最后令 V 为其列向量张成的空间

A。由于 A 的列是各向同性、独立同分布的随机高斯向量，因此 V 是 $\mathbb{R}^n$中的一个均匀随机 p 维子空间[5] 。

我们将把 V作为植入布尔向量问题的输入，而向量d_1、.....、 d_n 将用于构造植入仿射平面问题的伪期望算子[6]。由于 $n \leq p^{3/2-\Omega(\varepsilon)}$，根据定理 4.1.4，对于所有 $\delta \leq c\varepsilon$ ，当$c > 0$为常数时，以高概率存在一个度为-n^{δ} 的伪期望算子 $\widetilde{\mathbb{E}}'$ ，其作用于形式变量$v = (v_1$、.....、$v_p)$，使得对于每个 $u \in [n]$都有 $\widetilde{\mathbb{E}}'[\langle v, d_u\rangle^2] = 1$。

$\widetilde{\mathbb{E}}$通过 $\mathbb{E}[b_u]$ 定义 $\mathbb{E} := \mathbb{E}'[\langle v, d_u\rangle]$ 对所有 $u \in [n]$ 成立，并利用多重线性将其扩展到$\{b_u\}$ 上的所有多项式。由于 $\widetilde{\mathbb{E}}'[\langle v, d_u\rangle^2] = 1$，该定义是良定义的。注意， $\widetilde{\mathbb{E}}$ 是一个有效的、与 $\mathbb{E}'$具有相同度的$\widetilde{\mathbb{E}}$伪期望算子。最后，注意到

$$\frac{1}{n}\widetilde{\mathbb{E}}[b^{\intercal}\Pi_V b] = \frac{1}{n}\widetilde{\mathbb{E}}'[v^{\intercal}A^{\intercal}\Pi_V Av] = \frac{1}{n}\widetilde{\mathbb{E}}'[v^{\intercal}A^{\intercal}Av] = 1.$$

■

现在，我们使用 [165] 提出的归约和证明来给出 Sherrington-Kirkpatrick 问题的下界。为完整起见，我们在此将其纳入。回顾一下，SK 问题是计算

$$\mathrm{OPT}(W) := \max_{x\in\{\pm 1\}^n} x^{\intercal}Wx,$$

其中， W 采样自高斯正交系综(n)。

定理 4.1.2。 存在一个常数 $\delta > 0$ ，使得以高概率，对于 $W \sim$ 高斯正交系综(n)，*Sherrington–Kirkpatrick* 问题存在一个度为 n^{δ} 的平方和解，其值至少为 $(2 - o_n(1)) \cdot n^{3/2}$。

我们将使用以下关于高斯正交系综(n)随机矩阵理论的标准结果。

5. 除了一个零测集事件之外。

6. 注意，在 Planted BooleanVector 问题中，向量 d_u 并不是“给定的”，尽管 $\widetilde{\ }$ 的构造无论如何都不要求具有任何意义上的算法性。

事实 5.5.1. 设 $\lambda_1 \geq \ldots \geq \lambda_n$ 为 $W \sim$ 高斯正交系综(n) 的特征值，其对应的归一化特征向量为 $w_1, \ldots, w_n$。那么，

1. 对于每个 $p \in [n]$，w_1、...、w_p 的张成空间是 $\mathbb{R}^n$ 中一个均匀随机的 p 维子空间（参见例如 *[178,* 第 *2]* 节）。

2. W以高概率，$\lambda_{n^{0.67}} \geq (2 - o(1))\sqrt{n}$（*Wigner* 半圆律的推论 *[234]*）

定理 4.1.2 的证明： 设 $p = n^{0.67}$ 和 $W \sim$ 为 GOE(n)。令 $\lambda_1 \geq \ldots \geq \lambda_n$ 为 W 的特征值，其对应的正交归一特征向量组为 w_1，...，w_n。根据事实 5.5.1，我们有 $\lambda_p \geq (2 - o(1))\sqrt{n}$，并且 w_1，...，w_p 张成了 $\mathbb{R}^n$ 的一个均匀随机 p 维子空间 V。

我们将 V 视为布尔植入向量问题的输入。根据定理 4.1.5，对于某个常数 $\delta > 0$，以高概率存在一个度为 n^δ 的伪期望算子 $\widetilde{\mathbb{E}}$，使得 $\widetilde{\mathbb{E}}[x_i^2] = 1$ 且 $\widetilde{\mathbb{E}}[\sum_{i=1}^p \langle x, w_i\rangle^2] = \widetilde{\mathbb{E}}[x^\intercal \Pi_V x] = n$。现在，

$$\begin{aligned}
\widetilde{\mathbb{E}}[x^\intercal W x] &= \widetilde{\mathbb{E}}[\sum_{i=1}^n \lambda_i \langle x, w_i\rangle^2] \\
&\geq \lambda_p \widetilde{\mathbb{E}}[x^\intercal \Pi_V x] - |\lambda_n| \widetilde{\mathbb{E}}[\sum_{i=p+1}^n \langle x, w_i\rangle^2] \\
&\geq (2 - o(1)) n^{3/2} - |\lambda_n| \widetilde{\mathbb{E}}[\langle x, x\rangle - \sum_{i=1}^p \langle x, w_i\rangle^2] \\
&= (2 - o(1)) n^{3/2}.
\end{aligned}$$

■

备注 5.5.2. 使用与上述相同的证明，即使我们只能证明某些 $m = \omega(n)$下植入仿射平面的平方和下界，也能得到定理 *4.1.2*。因此，将 m 提升至 $n^{3/2-\varepsilon}$（即定理 *4.1.4*）只能带来有限的改进。

5.6 省略的技术细节

5.6.1 范数界

我们使用的精确范数界是通过在 [1], 应用迹幂方法得到的；不过从定性的角度看，第2章中的界也适用。论文 [1] 采用了略有不同的矩阵指标定义。他们将矩阵索引片段 定义为一个由 $\mathcal{C}_m$ 或 $\mathcal{S}_n$ 中的不同元素组成的元组，并附带一个表示重数的固定整数。矩阵指标则是矩阵索引片段构成的集合。对于给定的方形顶点集合，将正方形按递增顺序排列成一个元组，并赋予重数 1，我们的图矩阵 M_α 便是那些矩阵的一个子矩阵。因此，相同的范数界仍然适用。

布尔范数界：

引理 5.6.1. 设 $V_{rel}(\alpha) := V(\alpha) \setminus (U_\alpha \cap V_\alpha)$。存在一个普适常数 C，使得对所有适当形状 α，以下范数界以高概率成立：

$$\|M_\alpha\| \leq 2 \cdot (|V(\alpha)| \cdot \log(n))^{C \cdot |V_{rel}(\alpha)|} \cdot n^{\frac{w(V(\alpha)) - w(S_{\min}) + w(W_{iso})}{2}}$$

证明。根据 [1], 的推论 8.13，对于固定形状 α，其概率至少为 $1 - \varepsilon$，

$$\|M_\alpha\| \leq 2\,|V(\alpha)|^{|V_{rel}(\alpha)|} \cdot \left(6e \left\lceil \frac{\log\left(\frac{n^{w(S_{\min})}}{\varepsilon}\right)}{6\,|V_{rel}(\alpha)|} \right\rceil \right)^{|V_{rel}(\alpha)|} \cdot n^{\frac{w(V(\alpha)) - w(S_{\min}) + w(W_{iso})}{2}}$$

令 N_k 为 k 个顶点上的不同形状（圆或正方形）的数量，我们将推论应用于 $\varepsilon = 1/(mnN_{|V(\alpha)|})$。通过并界，所有大小为 k 的形状的失败概率至多为 $1/mn$；又由于一个形状中的顶点数至多为 $m + n \leq 2m$，因此我们得到一个以高概率对所有形状均成立的界。接下来只需化简这个精确的界。

命题 5.6.2. $N_k \leq 8^k 2^{k^2}$

证明。 以下过程可以构造出 k 个顶点上的所有形状：从 k 个形式变量开始，将每个变量指定为圆形或方形，决定每个变量是否属于 U_α 和/或 V_α，然后在 k^2 对变量之间放置任意数量的边。

■

我们还对 $n^{w(S_{\text{最小值}})} \leq (mn)^{|V(\alpha)|}$ 进行界定。

$$\begin{aligned}
\|M_\alpha\| &\leq 2\,|V(\alpha)|^{|V_{rel}(\alpha)|} \cdot \left(6e \left\lceil \frac{\log\left(n^{w(S_{\min})} \cdot mnN_{|V(\alpha)|}\right)}{6\,|V_{rel}(\alpha)|} \right\rceil\right)^{|V_{rel}(\alpha)|} \cdot n^{\frac{w(V(\alpha))-w(S_{\min})+w(W_{iso})}{2}} \\
&\leq 2\,|V(\alpha)|^{|V_{rel}(\alpha)|} \cdot \left(12e \log\left(n^{w(S_{\min})} \cdot mnN_{|V(\alpha)|}\right)\right)^{|V_{rel}(\alpha)|} \cdot n^{\frac{w(V(\alpha))-w(S_{\min})+w(W_{iso})}{2}} \\
&\leq 2\,|V(\alpha)|^{|V_{rel}(\alpha)|} \cdot \left(12e \log\left((mn)^{|V(\alpha)|} \cdot mn \cdot 8^{|V(\alpha)|} 2^{|V(\alpha)|^2}\right)\right)^{|V_{rel}(\alpha)|} \cdot n^{\frac{w(V(\alpha))-w(S_{\min})+w(W_{iso})}{2}} \\
&\leq 2\,|V(\alpha)|^{|V_{rel}(\alpha)|} \cdot \left(100e\,|V(\alpha)|^2 \log(mn)\right)^{|V_{rel}(\alpha)|} \cdot n^{\frac{w(V(\alpha))-w(S_{\min})+w(W_{iso})}{2}} \\
&\leq 2 \cdot (|V(\alpha)| \cdot \log(mn))^{3\cdot|V_{rel}(\alpha)|} \cdot n^{\frac{w(V(\alpha))-w(S_{\min})+w(W_{iso})}{2}}
\end{aligned}$$

注意，现在我们假设 $m \leq n^2$。 ■

对于 Hermite 形状，我们有如下范数界。对于 Hermite 形状 α，将总大小 定义为 $|U_\alpha| + |V_\alpha| + |W_\alpha| + |E(\alpha)|$。

引理 5.6.3。 令 $V_{rel}(\alpha) := V(\alpha) \setminus (U_\alpha \cap V_\alpha)$ 为集合。存在一个普适常数 C，使得对于所有总大小至多为 n 的真形状 α ，以下范数界均以高概率成立：

$$\|M_\alpha\| \leq 2 \cdot (|V(\alpha)| \cdot (1 + |E(\alpha)|) \cdot \log(n))^{C\cdot(|V_{rel}(\alpha)|+|E(\alpha)|)} \cdot n^{\frac{w(V(\alpha))-w(S_{\min})+w(W_{iso})}{2}}$$

证明从 [1, 推论 8.15] 开始进行了相同的计算。请注意，在我们的记号中， $l(\alpha) = |E(\alpha)|$。还有一个区别： [1] 使用归一化 Hermite 多项式，而我们使用未归一化 Hermite 多项式；这会产生额外的项 $\prod_{e\in E(\alpha)} l(e)! \leq (1 + |E(\alpha)|)^{|E(\alpha)|}$。我们必须用下面的命题替换命题 5.6.2：

命题 5.6.4。 总大小为 k 的 *Hermite* 形状数量至多为 $k2^k(k+1)^{2k+k^2}$。

证明。 这样的形状至多有 k 个不同的变量顶点。每个变量顶点要么是圆形，要么是方形。每个变量在 U_α 中的重数介于 0 和（至多） k之间，在 V_α 中的重数也介于 0 和 k之间。k^2 个可能的顶点对，其边重数在 $E(\alpha)$ 中介于 0 和 k之间。

■

5.6.2 $e(k)$的性质

本节将确立分析中所用 $e(k)$ 的一些性质。回顾一下，$e(k) = \mathbb{E}_{x\in\mathcal{S}(\sqrt{n})}[x_1 \dots x_k]$，其中 $\mathcal{S}(\sqrt{n}) := \{x \in \{\pm 1\}^n \mid \sum_{i=1}^n x_i = \sqrt{n}\}$。

断言 5.6.5. $e(2) = 0$。

证明。 固定 $y \in \mathcal{S}(\sqrt{n})$。注意，$(\sum_{i=1}^n y_i)^2 = n$，这意味着 $\sum_{i<j} y_i y_j = 0$。利用这一事实，我们得到

$$\mathbb{E}_{x\in\mathcal{S}(\sqrt{n})}[x_1 x_2] = \mathbb{E}_{\sigma\in S_n} y_{\sigma(1)} y_{\sigma(2)} = 0,$$

证明至此完成。 ■

定义 5.6.6. 我们称由非负整数组成的元组 $\lambda = (\lambda_1, \dots, \lambda_k)$ 为 k 的一个分拆，条件是 $\sum_{i=1}^k \lambda_i = k$ 且 $\lambda_1 \geq \cdots \geq \lambda_k$。我们用记号 $\lambda \vdash k$ 表示 k的一个分拆。我们称 λ_i 为 λ的一行／部分。

接下来，我们将讨论可以用整数分拆作索引的多项式。因此，我们现在为分拆及其相关对象约定一些记号。

定义 5.6.7. 分拆 $\lambda = (\lambda_1, \dots, \lambda_k)$ 的转置记作 λ^t，定义为 $\lambda_i^t = \left|\{j \in [k] \mid \lambda_j \geq i\}\right|$。

Re注记 5.6.8. 对于一个分拆 $\lambda \vdash k$，λ_1^t 是 λ的行数/部分数。

定义 5.6.9. 分拆的自同构群 $\mathrm{Aut}(\lambda) \leq S_{\lambda_1^t}$ 是由行 $\lambda_i = \lambda_j$的对换 (i,j) 生成的群。

注记 5.6.10。 设 $\lambda \vdash k$和 $p_1(\lambda) \ldots, p_k(\lambda)$ 满足 $p_i(\lambda) = \left|\{j \in [\lambda_1^t] \mid \lambda_j = i\}\right|$。那么 自同构群$(\lambda) \simeq S_{p_1} \times \cdots \times S_{p_k}$。

引理 5.6.11。我们有

$$\sum_{\lambda \vdash k} \frac{\lambda!}{\lambda_1! \cdots \lambda_k!} \cdot \frac{(n)_{\lambda_1^t}}{|\mathrm{Aut}(\lambda)|} \cdot \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \ldots x_k^{\lambda_k}\right] = n^{k/2}.$$

证明。 对于 $x \in \mathcal{S}(\sqrt{n})$，我们有 $(\sum_{i=1}^n x_i)^k = n^{k/2}$。然后将前面的方程中的 $(\sum_{i=1}^n x_i)^k$ 展开，并对两边取关于 $\mathcal{S}(\sqrt{n})$ 的期望，即可得到该引理的结论（适当合并各项后）。

■

断言 5.6.12。令 $\lambda \vdash k$。我们有

$$(n)_{\lambda_1^t} \cdot \left|\mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \ldots x_k^{\lambda_k}\right]\right| \leq 3^{k^3} \cdot n^{k/2}.$$

证明。 我们对 k 进行归纳。对于 $k = 1$，我们有 $n \cdot \left|\mathbb{E}_{\overline{x} \in \mathcal{S}(\sqrt{n})} [x_1]\right| = \sqrt{n} \leq 3 \cdot n^{1/2}$。现在，假设 $k \geq 2$。我们分三种情形讨论：

1. 情形 $\lambda_1 \geq 3$: 设 λ'为从 λ_1中移除两个方格后由 λ 得到的分拆。注意， $\lambda_1^t = (\lambda')_1^t \leq k-2$ 和$\mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \cdots x_{k-2}^{\lambda_{k-2}}\right] = \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \cdots x_{k-2}^{\lambda_{k-2}}\right]$。由归纳假设，我们有 $(n)_{(\lambda')_1^t} \cdot \left|\mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda'_1} \ldots x_{k-2}^{\lambda'_{k-2}}\right]\right| \leq 3^{(k-2)^2} \cdot n^{(k-2)/2}$。

2. 情形 $\lambda_1 = 2$: 设 λ' 为从 λ 中移除 λ_1后得到的分拆。注意到

$\lambda_1^t = (\lambda')_1^t + 1 \le k-2$。根据归纳假设，我们有

$$\begin{aligned}(n)_{\lambda_1^t} \cdot \left|\mathbb{E}_{x\in\mathcal{S}(\sqrt{n})}\left[x_1^{\lambda_1}\dots x_{k-2}^{\lambda_{k-2}}\right]\right| &\le n\cdot(n)_{(\lambda')_1^t}\cdot\left|\mathbb{E}_{x\in\mathcal{S}(\sqrt{n})}\left[x_1^{\lambda'_1}\dots x_{k-2}^{\lambda'_{k-2}}\right]\right| \\ &\le 3^{(k-2)^3}\cdot n^{k/2}.\end{aligned}$$

3. 情形 $\lambda_1 = 1$: 为了给 $(n)_k \cdot \mathbb{E}_{x\in\mathcal{S}(\sqrt{n})}\left[x_1^{\lambda_1}\dots x_k^{\lambda_k}\right]$ 取一个上界，我们使用引理 5.6.11 以及前面的两种情形。设 $p(k)$ 为分拆函数，即 $p(k) = |\{\lambda \vdash k\}|$。由此可得

$$\begin{aligned}(n)_k \cdot \left|\mathbb{E}_{x\in\mathcal{S}(\sqrt{n})}\left[x_1^{\lambda_1}\dots x_k^{\lambda_k}\right]\right| &\le n^{k/2} + \sum_{\lambda\vdash k:\ \lambda_1\ge 2}\frac{\lambda!}{\lambda_1!\cdots\lambda_k!}\cdot\frac{(n)_{\lambda_1^t}}{|\mathrm{Aut}(\lambda)|}\cdot\left|\mathbb{E}_{x\in\mathcal{S}(\sqrt{n})}\left[x_1^{\lambda_1}\dots x_k^{\lambda_k}\right]\right| \\ &\le n^{k/2} + k!\sum_{\lambda\vdash k:\ \lambda_1\ge 2}(n)_{\lambda_1^t}\cdot\left|\mathbb{E}_{x\in\mathcal{S}(\sqrt{n})}\left[x_1^{\lambda_1}\dots x_k^{\lambda_k}\right]\right| \\ &\le n^{k/2} + k!\sum_{\lambda\vdash k:\ \lambda_1\ge 3}(n)_{\lambda_1^t}\cdot\left|\mathbb{E}_{x\in\mathcal{S}(\sqrt{n})}\left[x_1^{\lambda_1}\dots x_k^{\lambda_k}\right]\right| + \\ &\qquad k!\sum_{\lambda\vdash k:\ \lambda_1=2}(n)_{\lambda_1^t}\cdot\left|\mathbb{E}_{x\in\mathcal{S}(\sqrt{n})}\left[x_1^{\lambda_1}\dots x_k^{\lambda_k}\right]\right| \\ &\le 3^{(k-2)^3}\cdot k!\cdot(1+p(k)+k)\cdot n^{k/2} \le 3^{k^3}\cdot n^{k/2},\end{aligned}$$

如所求。

■

断言 5.6.13。 假设 $k < \sqrt{n}/2$。我们有

$$\left|\mathbb{E}_{x\in\mathcal{S}(\sqrt{n})}\left[x_1\dots x_k\right]\right| \le 2\cdot 3^{k^3}\cdot n^{-k/2}.$$

证明。 由断言 5.6.12 以及关于 k 的界可得。 ■

注记 5.6.14。 在断言 *5.6.13* 中，因子 3^{k^3} 损失过大，无法得到关于 $k = n^{\varepsilon}$ 的有意义的界，其中 $\varepsilon > 0$ 是一个常数。

通过对断言 5.6.12 中的思路加以完善，我们在下文证明一个更强的引理，它将给出关于 $e(k)$ 的更紧界，足以满足我们的应用需求。

引理 5.6.15. 存在一个普适常数 $C \geq 1$，使得

$$\sum_{\lambda \vdash k} \frac{\lambda!}{\lambda_1! \cdots \lambda_k!} \cdot \frac{(n)_{\lambda_1^t}}{|\mathrm{Aut}(\lambda)|} \cdot \left| \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right] \right| \leq k^{C \cdot k} \cdot n^{k/2}. \tag{5.1}$$

特别地，对于 $n \geq 6$，等式 (5.1) 在 $C = 2$ 下成立。

证明. 我们对 k 进行归纳。对于 $k = 1$，我们有 $n \cdot \left| \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} x_1 \right| \leq \sqrt{n}$，如所需。利用断言 5.6.5 中的 $e(2) = 0$ 以及公式（5.1）的情形 $k = 1$，可得引理 5.6.15 对于 $k = 2$ 也成立。现在考虑 $k \geq 3$。令 $\Lambda_1 = \{\lambda \vdash k \mid \lambda_1 = 1\}$、$\Lambda_2 = \{\lambda \vdash k \mid \lambda_1 = 2\}$ 和 $\Lambda_{\geq 3} = \{\lambda \vdash k \mid \lambda_1 = 3\}$。注意到 $\Lambda_1 \sqcup \Lambda_2 \sqcup \Lambda_{\geq 3} = \{\lambda \vdash k\}$ 和 $|\Lambda_1| = 1$。

为方便起见，定义 a_λ 为公式（5.1）左端与 $\lambda \vdash k$ 相关的项，即，

$$a_\lambda = \frac{\lambda!}{\lambda_1! \cdots \lambda_k!} \cdot \frac{(n)_{\lambda_1^t}}{|\mathrm{Aut}(\lambda)|} \cdot \left| \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right] \right|.$$

首先，我们对公式（5.1）左端中与来自 $\Lambda_{\geq 3}$ 的分拆相关的项的贡献进行界定。令 λ' 为从 λ 中移除 λ_1 的两个方格后得到的分拆。注意，$\lambda_1^t = (\lambda')_1^t \leq k - 2$ 且 $\mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_{k-2}^{\lambda_{k-2}} \right] = \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_{k-2}^{\lambda_{k-2}} \right]$。因此，

$$\begin{aligned} a_\lambda &= \frac{\lambda!}{\lambda_1! \cdots \lambda_k!} \cdot \frac{(n)_{\lambda_1^t}}{|\mathrm{Aut}(\lambda)|} \cdot \left| \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right] \right| \\ &= \frac{k(k-1)}{\lambda_1(\lambda_1 - 1)} \cdot \frac{\left|\mathrm{Aut}(\lambda')\right|}{|\mathrm{Aut}(\lambda)|} \frac{\lambda'!}{\lambda_1'! \cdots \lambda_k'!} \cdot \frac{(n)_{(\lambda')_1^t}}{|\mathrm{Aut}(\lambda')|} \cdot \left| \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1'} \dots x_{k-2}^{\lambda_{k-2}'} \right] \right| \\ &= k^2 \cdot \frac{\left|\mathrm{Aut}(\lambda')\right|}{|\mathrm{Aut}(\lambda)|} \cdot a_{\lambda'} \leq k^3 \cdot a_{\lambda'}, \end{aligned}$$

因为 $\left|\mathrm{Aut}(\lambda')\right| / |\mathrm{Aut}(\lambda)| \leq k - 2 \leq k$。对于每个 $\lambda' \vdash k - 2$，我们都可以通过向 λ' 的单行添加两个块，以 $k - 2 \leq k$ 种方式构造一个分拆 $\lambda \vdash k$。因此，我们有

$$\sum_{\lambda \in \Lambda_{\geq 3}} a_\lambda \leq k \cdot \sum_{\lambda' \vdash k-2} k^3 \cdot a_{\lambda'} \leq k^4 \cdot k^{C \cdot (k-2)} \cdot n^{(k-2)/2}, \tag{5.2}$$

其中最后一个等式由归纳假设得出。

现在，我们对公式（5.1）左端中与来自 Λ_2 的分拆 λ 相关联的各项 a_λ 的贡献进行估计。令 $i \geq 1$ 为 λ 中大小为二的部分的数量，并令 λ' 为从 λ 中移除这些大小为二的部分 i 后得到的分拆。注意，$\lambda'^t_1 = \lambda^t_1 + i \leq k-1$。我们有

$$\begin{aligned} a_\lambda &= \frac{\lambda!}{\lambda_1! \cdots \lambda_k!} \cdot \frac{(n)_{\lambda^t_1}}{|\mathrm{Aut}(\lambda)|} \cdot \left| \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right] \right| \\ &\leq n^i \cdot \frac{(k)_i}{2^i} \cdot \frac{\left|\mathrm{Aut}(\lambda')\right|}{|\mathrm{Aut}(\lambda)|} \cdot \frac{\lambda'!}{\lambda'_1! \cdots \lambda'_k!} \cdot \frac{(n)_{(\lambda')^t_1}}{|\mathrm{Aut}(\lambda')|} \cdot \left| \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1 \dots x_{k-2i} \right] \right| \\ &= n^i \cdot \frac{(k)_i}{2^i} \cdot \frac{1}{i!} \cdot \frac{\lambda'!}{\lambda'_1! \cdots \lambda'_k!} \cdot \frac{(n)_{(\lambda')^t_1}}{|\mathrm{Aut}(\lambda')|} \cdot \left| \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1 \dots x_{k-2i} \right] \right|, \end{aligned}$$

其中，在最后一个等式中，我们使用了 $\left|\mathrm{Aut}(\lambda')\right| / |\mathrm{Aut}(\lambda)| = 1/(i!)$。由于 $\lambda \in \Lambda_2$ 由其大小为二的部分的数量唯一确定，应用归纳假设可得

$$\begin{aligned} \sum_{\lambda \in \Lambda_2} a_\lambda &\leq \sum_{i=1}^{\lfloor k/2 \rfloor} n^i \cdot \frac{(k)_i}{2^i} \cdot \frac{1}{i!} \cdot \left(\frac{\lambda'!}{\lambda'_1! \cdots \lambda'_k!} \cdot \frac{(n)_{(\lambda')^t_1}}{|\mathrm{Aut}(\lambda')|} \cdot \left| \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1 \dots x_{k-2i} \right] \right| \right) \\ &\leq \sum_{i=1}^{\lfloor k/2 \rfloor} n^i \cdot \frac{(k)_i}{2^i} \cdot \frac{1}{i!} \cdot k^{C \cdot (k-2i)} \cdot n^{(k-2i)/2} \\ &\leq k^{C \cdot (k-1)} \cdot n^{k/2} \cdot \sum_{i=0}^{\infty} k^{-C \cdot i} \leq \frac{3}{2} \cdot k^{C \cdot (k-1)} \cdot n^{k/2}, \end{aligned}$$

其中，在最后一个不等式中，我们使用了 $k \geq 3$ 和 $C \geq 1$。

最后，我们考虑情形 $\lambda_1 = 1$。为了给 a_λ 建立界，我们使用引理 5.6.11 和前面的两种情形。由此可得

$$a_\lambda \le n^{k/2} + \sum_{\mu\in\Lambda_2} a_\mu + \sum_{\mu\in\Lambda_{\ge 3}} a_\mu \le n^{k/2} + k^4 \cdot k^{C\cdot(k-2)} \cdot n^{(k-2)/2} + \frac{3}{2}\cdot k^{C\cdot(k-1)} \cdot n^{k/2}$$
$$= k^{C\cdot k} \cdot n^{k/2}\left(\frac{1}{k^{C\cdot k}} + \frac{k^4}{n\cdot k^{2\cdot C}} + \frac{3}{2\cdot k^C}\right).$$

我们可以将公式（5.1）的左端估计为

$$\sum_{\mu\in\Lambda_1} a_\mu + \sum_{\mu\in\Lambda_2} a_\mu + \sum_{\mu\in\Lambda_{\ge 3}} a_\mu \le k^{C\cdot k}\cdot n^{k/2}\left(\frac{1}{k^{C\cdot k}} + \frac{2\cdot k^4}{n\cdot k^{2\cdot C}} + \frac{3}{k^C}\right)$$
$$\le k^{C\cdot k}\cdot n^{k/2},$$

其中 $C > 0$ 是一个足够大的常数。特别地，对于 $n \ge 6$，常数 C 可以取为 2。

■

推论 5.6.16. 我们有

$$\left|\mathbb{E}_{x\in\mathcal{S}(\sqrt{n})}\left[x_1\dots x_k\right]\right| \le k^{3\cdot k}\cdot n^{-k/2}.$$

证明。假设 $k \le \sqrt{n}$。注意，引理 5.6.15 表明，对于满足 λ_1 的 $\lambda \vdash k$，存在一个常数 $C > 0$，使得

$$\frac{\lambda!}{\lambda_1!\cdots\lambda_k!}\cdot\frac{(n)_{\lambda_1^t}}{|\mathrm{Aut}(\lambda)|}\cdot\left|\mathbb{E}_{x\in\mathcal{S}(\sqrt{n})}\left[x_1^{\lambda_1}\dots x_k^{\lambda_k}\right]\right| = (n)_k\cdot\left|\mathbb{E}_{x\in\mathcal{S}(\sqrt{n})}\left[x_1\dots x_k\right]\right|$$
$$\le k^{C\cdot k}\cdot n^{k/2}.$$

化简并利用假设 $k \le \sqrt{n}$，我们得到

$$\left|\mathbb{E}_{x\in\mathcal{S}(\sqrt{n})}\left[x_1\dots x_k\right]\right| \le \frac{k^{C\cdot k}\cdot n^{-k/2}}{\prod_{i=1}^{k-1}\left(1-\frac{i}{n}\right)} \le 2\cdot k^{C\cdot k}\cdot n^{-k/2}.$$

此外，对于 $n \geq 6$，引理 5.6.15 允许我们选取 $C = 2$。由于 $\left|\mathbb{E}_{x \in \mathcal{S}(\sqrt{n})}[x_1]\right| = 1/\sqrt{n}$，因此对 k 的所有取值，较简单的界都适用

$$\left|\mathbb{E}_{x \in \mathcal{S}(\sqrt{n})}[x_1 \dots x_k]\right| \leq k^{3 \cdot k} \cdot n^{-k/2},$$

现在可以去掉假设 $n \geq 6$，因为对于 $k \geq 2$，我们有 $(k^3/\sqrt{n})^k \geq 1$，其中 1 是平凡界。同样，我们最初关于 $k \leq \sqrt{n}$ 的假设也可以去掉，因为在 $k > \sqrt{n}$ 的范围内，该界也变成了平凡界。

■

第6章

分析工具与定性界

本章首先陈述我们用于证明结果的主要技术工具。该技术工具是一条元定理，它表明，在满足若干线性代数条件的情况下，一个大型随机矩阵以高概率是半正定（PSD）的。这与上一章中 Sherrington-Kirkpatrick 哈密顿量的平方和下界里的半正定性论证思路相似，不过这里涉及的内容要复杂得多。

这些技术工具最初发表于 [191, 192], ，完整证明可在其中找到。在第6.1节，我们陈述这些技术工具。与那项工作相比，我们大幅简化了陈述和应用这些技术工具所需的定义。由于我们不提供证明，而只是应用定理，因此可以精简这些定义。例如，我们不定义带状图，不正式定义技术矩阵 $M_\tau^{orth}(H)$，并且采用简化的设定，使用拉ademacher或高斯变量，而非具有有限阶矩的任意分布。此外，我们将这些技术工具的条件解释为问题本身的条件，而不是像 [191]中那样，将其视为关于矩矩阵的计算性线性代数条件。在这一框架下，便能清楚地看出它与第3章和第8章所述的低度似然比假设之间可能存在的联系。

在第6.1节陈述这些技术工具之后，我们给出将其应用于所关注问题的定性界。本节内容也改编自 [191]。主要区别在于，我们通过修正若干错字并澄清一些技术论证，改进了表述。

6.1 证明工具

本节将介绍我们用于证明平方和下界的技术工具。正如我们在第5章中已经看到的，本文证明平方和下界的一般思路是

将矩矩阵分解为图矩阵，即输入各项的矩阵值函数，然后通过给出一个近似半正定分解来证明其半正定性。这些技术工具采用了类似的方法，并利用傅里叶系数的某些衰减性质以及与第2章中得到的类似范数界，给出一个近似半正定分解。

考虑一个假设检验问题 $\mathcal{P}$。我们假设采用第3.3节中的设定。因此， $\mathcal{P}$ 是一个区分两个分布的问题——随机分布（原假设）和植入分布（备择假设）。正如我们之前所见，我们可以使用伪校准技术来得到一个候选矩矩阵 Λ，使得 Λ 有可能作为随机分布上相关优化任务的平方和下界。这些技术工具给出了一般条件，可确保以高概率满足可行性。具体而言，这些技术工具是一组关于Λ 的线性代数条件，它们能以高概率推出 Λ 的正定性；而正如我们之前所见，其他所需的可行性条件则可通过伪校准轻松得到。

在这项工作中，我们稍微偏离这一观点（最初发表于 [191]），转而将这些条件视为问题 $\mathcal{P}$ 本身的性质。因此，这些技术工具可以理解为一项可行性主张：在满足问题 $\mathcal{P}$的某些条件时，经过伪校准的伪期望算子是可行的。为了用这种语言表述这些技术工具，我们需要先给出接下来的一些定义。

6.1.1 形状与图矩阵

考虑输入分布为拉ademacher $G_{n,\ 1/2}$ 图且输入项为 $\chi_e \in \{-1, 1\}$ 的情形。对于 $T \subseteq \binom{[n]}{2}$，令 $\chi_T = \prod_{e \in T} \chi_e$ 为标准傅里叶基。在这种情形下，形状已在第2章中定义。这里，出于技术原因，我们稍微修改定义，使行和列由 $[n]$ 的子元组而非 $[n]$的子集索引。第2章中发展的技术仍可用于

界定此类图矩阵的范数。

定义 6.1.1（拉ademacher $G_{n,\ 1/2}$ 输入的形状）。形状 α =（V（α） E（α） U_α, V_α）是一个以顶点 V（α）和边 E（α）构成的图，并带有两组指定的顶点元组 $U_\alpha, V_\alpha \subseteq V$（$\alpha$）。注意， U_α, V_α 是有序子集（元组）。

正如我们先前所见，可以为每个形状定义相应的矩阵，这些矩阵称为图矩阵。回顾一下，实现是从 $V(\alpha)$ 到 $[n]$ 的单射映射。与第2章相比，这里的主要区别在于，在图矩阵的定义中，我们对对应于不同特征的实现 φ 求和，而不是对所有实现 φ 求和。

为准确表述这一概念，我们使用以下定义。若两个实现（从 $V(\alpha)$ 到 $[n]$ 的单射映射）φ，φ' 满足 $\varphi(U_\alpha) = \varphi'(U_\alpha)$ $\varphi(V_\alpha) = \varphi'(V_\alpha)$ 作为元组相等，且 $\varphi(E(\alpha)) = \varphi'(E(\alpha))$ 作为集合相等，则称它们等价。将 α 的非等价实现集合记作 $\mathrm{Real}(\alpha)$。

定义 6.1.2（Rademacher $G_{n,\ 1/2}$ 输入情形下的图矩阵）。对于一个形状α，图矩阵 M_α 是一个矩阵值函数，其行和列分别由大小为 $|U_\alpha|$、$|V_\alpha|$ 的 $[n]$ 子元组索引，定义如下：它将输入图$G \in \{\pm 1\}^{\binom{n}{2}}$（其对应的傅里叶特征为 χ_E）映射为一个矩阵，其中第 A、B 项为

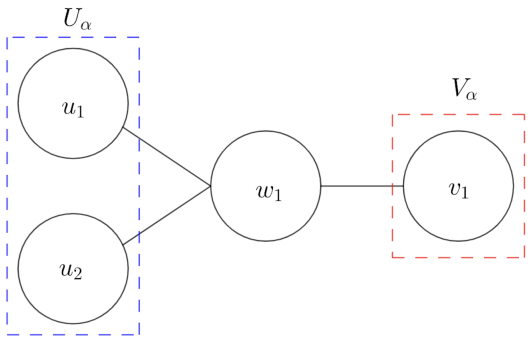
$$M_\alpha(A,B) = \sum_{\substack{\varphi(U_\alpha)=A, \varphi(V_\alpha)=B \\ \varphi \in \mathrm{Real}(\alpha)}} \chi_{E(\alpha)}$$

定义 6.1.3（形状转置）。对于形状 α =（V（α） E（α） U_α， V_α），将其转置 α^T 定义为 α^T =（V（α） E（α） V_α， U_α）。注意， $M_{\alpha^T} = M_\alpha^T$是矩阵转置。

例 6.1.4。在图 *6.1* 中，考虑图示的形状 α。我们有 $U_\alpha = (u_1, u_2)$ $V_\alpha = (v_1)$ $V(\alpha)$ $= \{u_1, u_2, v_1, w_1\}$ 以及 $E(\alpha) = \{\{u_1, w_1\}, \{u_2, w_1\}, \{w_1, v_1\}\}$。$M_\alpha$ 是一个矩阵，其行和列分别由长度为 $|U_\alpha| = 2$ 和 $|V_\alpha| = 1$ 的元组索引。

非零项的行和列分别由(a_1, a_2) 和 b_1 索引，对所有互不相同的 a_1, a_2, b_1均如此；相应的矩阵项为 $M_\alpha((a_1, a_2),(b_1)) = \sum_{c_1 \in [n] \setminus \{a_1,a_2,b_1\}} \chi_{a_1,c_1} \chi_{a_2,c_1} \chi_{c_1,b_1}$。这里，单射 φ 将顶点 u_1, u_2, w_1, v_1分别映射到 a_1, a_2, c_1, b_1，并且我们对所有此类映射求和（因为它们两两不等价）。简而言之，

$$M_\alpha = \begin{array}{c} \\ row\ (a_1, a_2) \rightarrow \end{array} \begin{array}{c} column\ (b_1) \\ \downarrow \\ \begin{pmatrix} \vdots \\ \cdots\cdots \sum_{c_1 \in [n] \setminus \{a_1,a_2,b_1\}} \chi_{a_1,c_1} \chi_{a_2,c_1} \chi_{c_1,b_1} \cdots\cdots\cdots \\ \vdots \end{pmatrix} \end{array}$$



Shape α

图 6.1：形状示例

一些简单的矩阵，例如图的邻接矩阵和单位矩阵，也属于图矩阵，如下所示

- 设 α 为一个有两个顶点 u， v 其中 $U_\alpha = (u)$ $V_\alpha = (v)$，且恰有一条边 $\{u, v\}$。那么， M_α 的行和列由 $[n]$ （更具体地说，是长度为 1 的元组）作为索引，其中 i， j 项在 $i \neq j$ 时为 G_{ij}，否则为 0。因此， M_α 就是图 G的 ± 1邻接矩阵。

- 设 α 为恰有 1 个顶点 u、没有边且 $U_\alpha = V_\alpha = (u)$ 的形状。那么，M_α 是大小为 $n \times n$的单位矩阵。

关于图矩阵的更多示例，以及它们为何是实用的工具，请参见 [1]。现在我们定义一些术语，用来描述图矩阵的行和列。

矩阵指标和指标形状 在上述拉ademacher $G_{n,1/2}$ 输入的设定中，矩阵指标 A 是一个指标元组 $(a_1, \ldots, a_{|A|})$，其中 $a_i \in [n]$。当 SoS 变量为 $y_1, \ldots, y_n$时，我们将单项式 $\prod_{i \le |A|} y_{a_i}$ 与该矩阵指标 A 关联起来。根据这一定义，图矩阵的行和列都是矩阵指标。

定义指标形状 $U = (u_1, \ldots, u_{|U|})$ 为形式变量 u_i组成的元组，换言之，即未指定的指标。若 $|U| = t$，则称任意长度为 t 的矩阵指标 A 具有形状U。若两个指标形状 U, V等价，则记作 $U \equiv V$，其中 $|U| = |V|$。最后，定义 U的权重为 $w(U) = |U|$，其自同构群为 $Aut(U) = S_{|U|}$（即 U中元素的置换）。后一项定义用于描述系数。

形状定义 我们称形状 α 为正规形状，如果除 $U_\alpha \cup V_\alpha$ 外没有孤立顶点（即没有 0 度顶点）。如果 U_α 和 V_α 作为集合相等，且它们构成 α中的全部顶点，并且 α中没有边，则称形状 α 为平凡形状。

路径是 $V(\alpha)$ 中顶点组成的数列，其中每一对相邻顶点在 $V(\alpha)$ 中构成一条边。 α 的顶点分隔集是顶点集合 S，使得从U_α 到 V_α 的每条路径都经过 S。正如我们在第2章中所看到的，图矩阵M_α 的范数界取决于 α的最小顶点分隔集的大小。将顶点分隔集 S 的权重定义为 $|S|$。

上述定义足以应用于植入式稍稠密子图问题。但在处理张量主成分分析和稀疏主成分分析时，我们需要推广形状和图矩阵的概念。这些推广后的形状和图矩阵已在 [1] 中得到研究。现在，我们来介绍所需的推广。

张量主成分分析的定义

在张量主成分分析的应用中，输入是一个张量 $A \in \mathbb{R}^{[n]^k}$。为将其纳入考虑，我们相应地修改形状和指标形状的定义。现在，输入项取自分布 $\mathcal{N}(0, 1)$，而非 $\{-1, 1\}$。因此，我们将使用厄米特多项式基。设标准的未归一化厄米特多项式记为 $h_0(x) = 1$，$h_1(x) = x$，$h_2(x) = x^2 - 1$，.....。接着，我们使用基$h_a(A) := \prod_{e\in[n]^k} h_e(A_e)$ over $a \in \mathbb{N}^{[n]^k}$。相应地，我们将修改表示形状的图，使其具有元数为 k 的带标签超边。因此，带有标签 t 的超边 e 将对应于厄米特多项式 $h_t(A_e)$。

定义 6.1.5（超边）。 在带状图和形状的底层图中，我们将使用元数为 k 的带标签超边，而不是标准边。超边的标签 e（记作 l_e）是 $\mathbb{N}$ 的一个元素，它对应于在该矩阵项上求值的厄米特多项式。

请注意，我们的超边是有序的，因为张量 A 不一定是对称的。对于变量 x_1、.....、x_n，矩矩阵的行和列现在将对应于形如 $\prod_{i\leq n} x_i^{p_i}$ 的单项式，其中 $p_i \geq 0$。为了描述这一点，我们采用指标形状片段和指标形状的概念。直观地说，我们按幂将上述单项式乘积拆分成若干组，每一组都构成一个指标形状片段。

定义 6.1.6（指标形状片段）。 指标形状片段 $U_i = ((U_{i,1}, \ldots, U_{i,t})\ p_i)$ 是一个指标元组 $(U_{i,1}, \ldots, U_{i,t})$ 以及一个幂 $p_i \in \mathbb{N}$。令 $V(U_i)$ 表示该指标形状片段的顶点集合 $\{U_{i,1}, \ldots, U_{i,t}\}$。上下文明确时，我们用 U_i 代替 $V(U_i)$。

如果我们将 $U_{i、1}$、.....、$U_{i、t}$ 实现为指标 a_1、.....、$a_t \in [n]$，那么这个指标形状片段的实现对应于单项式 $\prod_{j\leq t} x_{a_j}^{p_i}$。

定义 6.1.7（指标形状）。指标形状 U 是一组幂各不相同的指标形状片段 U_i 。令 V（U）为顶点 $\cup_i V$（U_i）的集合。当上下文清楚时，我们用 U 代替 V（U）。

注意到，指标形状的每个实现都对应矩矩阵的一行或一列。指标形状的等价性类似地定义：对于两个指标形状 U，V，若对于所有幂 p，U 和 V 中幂为 p 的指标形状片段长度相同，我们就记作 $U \equiv V$。我们还将 U 的自同构群定义为 $Aut(U) = \prod_{U_i \in U} Aut(U_i)$，其中指标形状片段 U_i 的自同构群为 $Aut(U_i) = S_{|U_i|}$。在形状的定义中，现在应以指标形状取代指定顶点集。

定义 6.1.8（形状）。形状是元组 $\alpha =$（V（α） E（α） U_α, V_α），其中（V（α） E（α））是一个超边元数为 k 的图，且 U_α, V_α 是指标形状，使得 $U_\alpha, V_\alpha \subseteq V$（$\alpha$）。

形状 α 若在 $U_\alpha \cup V_\alpha$之外没有孤立顶点、没有多重边，且所有边都带有非零标签，则称为正规形状。为了定义顶点分隔集的概念，我们相应地将路径的概念修改为适用于超边，而非边。形式上，路径是一个顶点数列 u_1，.....，u_t，使得对于所有 $i \leq t-1$，u_i, u_{i+1} 都属于同一条超边。顶点分隔集的概念按照上述超边和路径的概念作相同定义。最后，平凡形状 τ 的定义也类似，唯一的变化是我们现在要求 $U_\tau \equiv V_\tau$，而不是说它们作为集合相等。

稀疏主成分分析的定义

给定作为输入的 m 向量 v_1，.....，$v_m \in \mathbb{R}^d$。与张量主成分分析类似，我们将使用多项式的厄米特基，因为其条目采样自分布$\mathcal{N}(0, 1)$。具体而言，如果将未归一化 Hermite 多项式记为 $h_0(x) = 1$，$h_1(x) = x$，$h_2(x) = x^2 - 1$，.....，那么，我们将在 $a \in \mathbb{N}^{m \times n}$ 上使用基 $h_a(v) := \prod_{i \in [m],\ j \in [n]} h_{a_{i,\ j}}(v_{i,\ j})$。为表示这一基，我们将修改表示形状的图，使其

成为二分图，具有两种顶点类型，并且带标签的边连接不同类型的顶点。因此，一条边 (i, j)，其标签为 t，连接类型 1 顶点 i 与类型 2 顶点 j，将对应于 $h_t(v_{i,\ j})$。

形式上，我们有两种顶点：一种顶点对应于 m 输入向量，我们称之为类型 1 顶点；另一种顶点对应于空间的环境维数，我们称之为类型 2 顶点。对于具有这类顶点的形状，边将连接不同类型的顶点，从而形成二分图。类型 1 顶点 i 与类型 2 顶点 j 之间的一条边对应于输入矩阵项 $v_{i,\ j}$。每条边都带有一个标签 $\mathbb{N}$，该标签对应于在该矩阵项上求值的厄米特多项式。

我们的平方和程序中将有变量 x_1，.....，x_n，因此我们将像在张量主成分分析中那样使用指标形状片段和指标形状，因为现在矩矩阵的行和列将对应于形如 $\prod_{i\leq n} x_i^{p_i}$ 的单项式，其中 $p_i \geq 0$。但是，由于我们有两种顶点类型，因此需要稍微修改指标形状片段和指标形状的概念。

定义 6.1.9 (指标形状片段**)**。 指标形状片段 $U_i = ((U_{i,1}, \ldots, U_{i,t})\ t_i, p_i)$ 是一个沿着类型 $t_i \in \{1, 2\}$ 且具有幂 $p_i \in \mathbb{N}$的指标元组 $(U_{i,1}, \ldots, U_{i,t})$。 $V(U_i)$ 是该指标形状片段的顶点集合 $\{U_{i,1}, \ldots, U_{i,t}\}$。在上下文明确时，我们用$U_i$ 代替 $V(U_i)$。

对于类型为 $t_i = 2$ 的指标形状片段 $((U_{i,1}, \ldots, U_{i,t})\ t_i, p_i)$，如果我们将 $U_{i_1}, \ldots, U_{i_t}$实现为指标 $a_1, \ldots, a_t \in [n]$，那么这个指标形状片段对应于单项式$\prod_{j\leq n} x_{a_j}^{p_i}$。

定义 6.1.10 (指标形状**)**。 指标形状 U 是由指标形状片段 U_i 组成的集合，这些片段的类型或幂互不相同。 $V(U)$ 是顶点集合 $\cup_i V(U_i)$。在上下文明确时，我们用 U 代替 $V(U)$。

每个指标形状的实现都对应矩矩阵的一行或一列。对于我们的矩矩阵，唯一的非零行对应于仅含类型 2 指标形状片段的指标形状

，因为唯一的平方和变量是 $x_1 \ldots,\ x_n$，但为了进行分析，我们需要使用允许包含两种类型指标形状片段的索引形状的广义概念。

与之前的定义类似，对于两个指标形状 U、V，我们记作 $U \equiv V$，如果对于所有类型 t 和所有幂 p，U 和 V 中类型为 t 且幂为 p 的指标形状片段长度相同。由于我们使用的是标准图，原有的路径和顶点分隔集概念仍然适用，但现在我们将使用最小权顶点分隔集，而不是最小顶点分隔集，其权重定义如下。

定义 6.1.11 (指标形状的权重**)**。 假设有一个指标形状 $U = \{U_1, U_2\} \in \mathcal{I}_{mid}$，其中 $U_1 = ((U_{1,1}, \ldots, U_{1,|U_1|}), 1, 1)$ 是类型为1的指标形状片段，且 $U_2 = ((U_{2,1}, \ldots, U_{2,|U_2|}), 2, 1)$ 是类型为2的指标形状片段。于是，定义该指标形状的权重为 $w(U) = \sqrt{m}^{|U_1|}\sqrt{n}^{|U_2|}$。

— –

顶点分隔集也适用该定义。我们还将 U 的自同构群定义为 $Aut(U) = \prod_{U_i \in U} Aut(U_i)$，其中指标形状片段 U_i 的自同构群为 $Aut(U_i) = S_{|U_i|}$。现在给出形状的修订定义。

定义 6.1.12 （形状）。 形状是元组 $\alpha =$ （V（α） E（α） U_α, V_α），其中 （V（α） E（α））是一个具有两种顶点类型的图，只有不同类型的顶点之间才有带标签的边，并且 U_α, V_α 是指标形状，使得 $U_\alpha, V_\alpha \subseteq V$（$\alpha$）。

其余定义也与此类似。形状 α 是正规的，当且仅当除 $U_\alpha \cup V_\alpha$之外没有孤立顶点、没有多重边，且所有边的标签均非零。在平凡形状 τ的定义中，与张量主成分分析中的做法一样，我们要求 $U_\tau \equiv V_\tau$，而不是说它们作为集合相等。

6.1.2 形状分解

与第 5 章 Sherrington-Kirkpatrick 下界中的下界策略相比，这套技术工具的主要策略是将形状 α 分解为另外三种形状 σ，τ，σ'^T，使得 $M_\alpha \approx M_\sigma M_\tau M_{\sigma'^T}$。接着，思路是论证矩矩阵的图矩阵系数也以类似方式分解，最终得到一个半正定分解，说明该矩矩阵是半正定的。

我们首先需要定义形状的复合。若 $U_\beta \equiv V_\alpha$，则称形状 α 和 β 可以复合。在这种情况下，将它们的复合定义为形状 $\alpha \circ \beta$，它是通过拼接 α，β 并将 U_β，V_α 粘合在一起而得到的。形式化地说，$\alpha \circ \beta$ 满足 $U_{\alpha\circ\beta} = U_\alpha$、$V_{\alpha\circ\beta} = V_\beta$；令 $U_\beta = V_\alpha$ 后，取 $V(\alpha \circ \beta) = V(\alpha) \cup V(\beta)$，最后取 $E(\alpha \circ \beta) = E(\alpha) \cup E(\beta)$。

请注意，这样做可能会导致拼接后的形状变得不正规，因为边可能重复。需要指出的是，形状组合不一定满足交换律，但满足结合律。

例 6.1.13. 图*6.2*展示了一个只有一种顶点类型时形状组合的例子。观察形状 $\sigma \circ \sigma'^T$ 和 $\sigma \circ \tau \circ \sigma'^T$ 如何由形状 σ、τ 和 σ'^T 得到。

例6.1.14. 图*6.3* 展示了一个顶点有两种类型时的形状组合示例。我们用方形和圆形来表示这两种顶点。请注意，形状 $\sigma \circ \sigma'^T$ 和 $\sigma \circ \tau \circ \sigma'^T$ 是如何由形状 σ、τ 和 σ'^T 得到的。

此前，我们定义了最小顶点分隔集（以及相应的最小权顶点分隔集）的概念。下文中，我们将二者统称为最小权重顶点分隔集。定义最左侧（分别为最右侧）最小权重顶点分隔集 S（分别为 T）为满足以下条件的最小权重顶点分隔集：对于

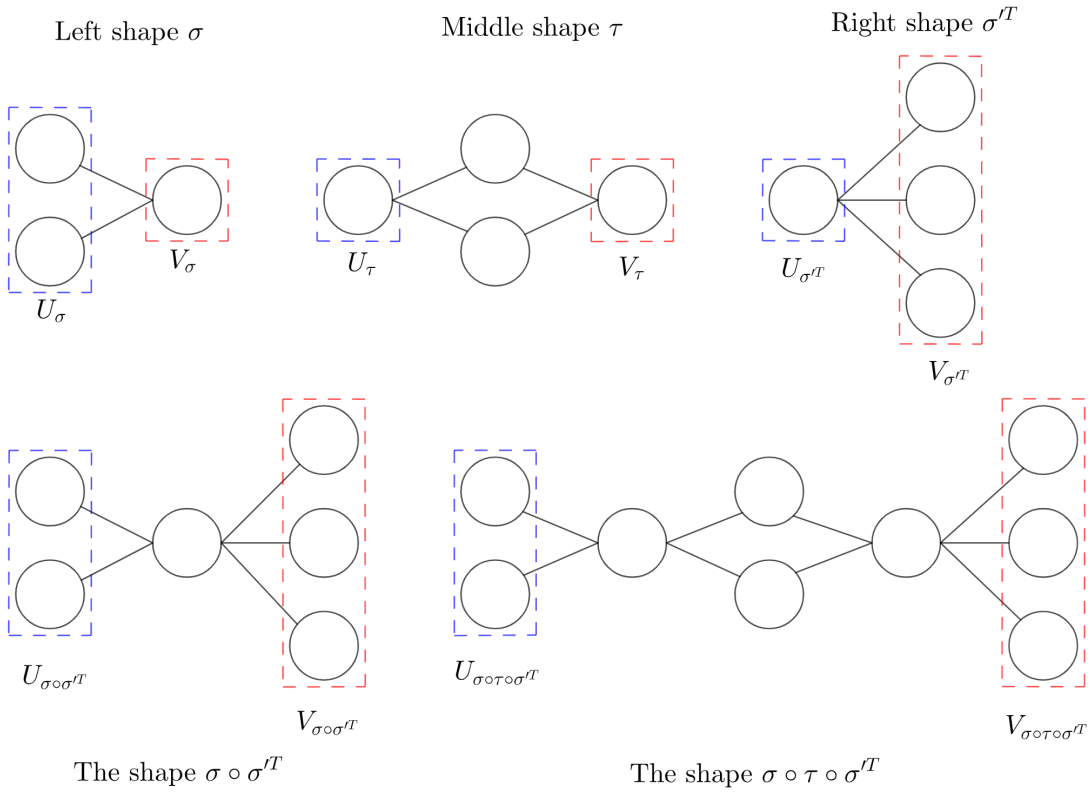


图 6.2：形状组合与分解示意图。

任意其他最小权重顶点分隔集 S'（分别为 T'），S 将 U_α 与 S' 分隔开（分别将T' 与 V_α分隔开）。文献 [20, 191]，证明了这一定义是良定的。

有了这些定义，我们现在可以定义如何将形状 α 分解为左部、中部和右部 σ，τ，σ'^T。

定义 6.1.15（形状分解）。设 α 为一个形状。设 S 和 T 分别为 α 的最左侧和最右侧最小权重顶点分隔集，并设有 S 和 T的一些序 O_S、O_T。

- 我们将 α 的左部分 σ 定义为如下形状：取 α 中所有无需经过 S 即可从 U_α 到达的顶点所构成的导出子图（但包括 S 的顶点），并移除 S 内的所有边和超边；然后我们取

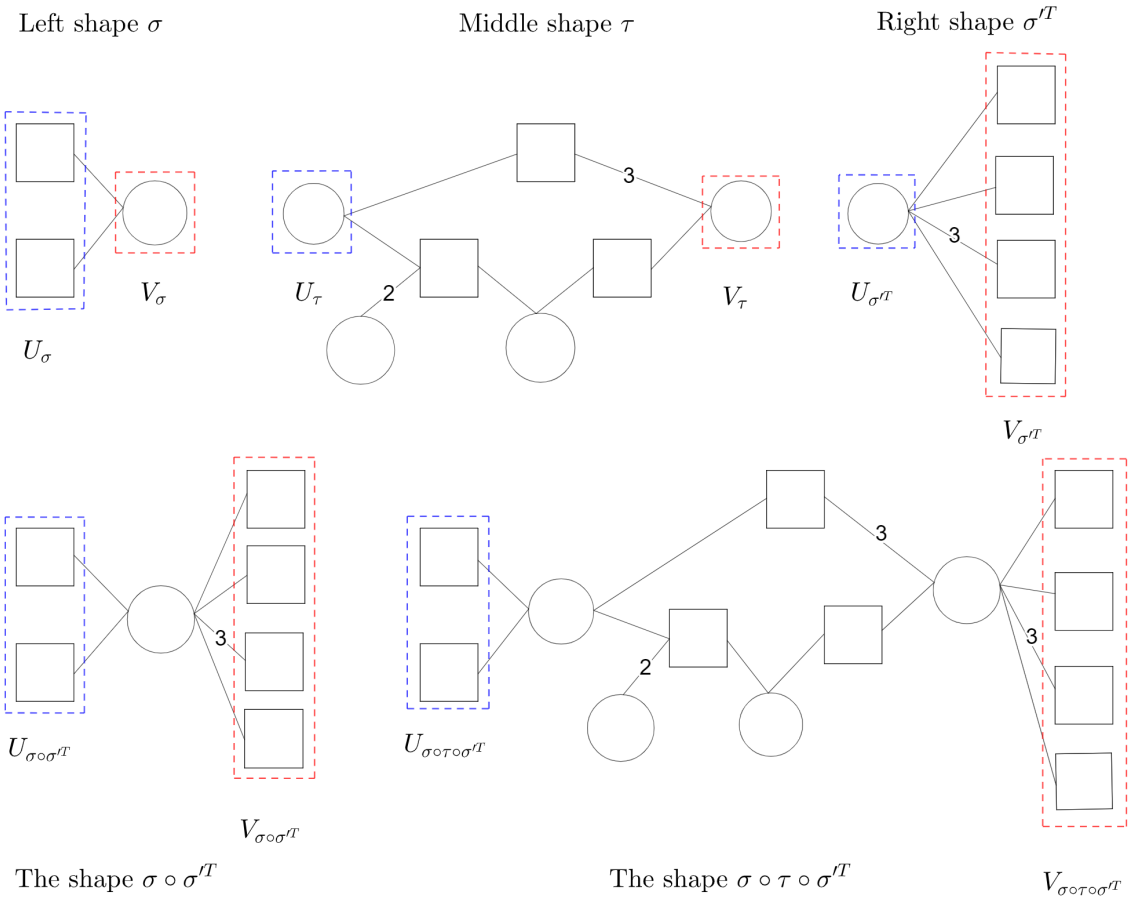


图 6.3：形状组合与分解示意图。

$$U_\sigma = U_\alpha \ and \ V_\sigma = (S, O_S).$$

• 我们将 α 的右部 σ'^T 定义为如下形状：取 α 中所有无需经过 T 即可从 V_α 到达的顶点所构成的导出子图（但包括 T 的顶点），并移除 T 内的所有边和超边；然后我们取$V_{\sigma'^T} = V_\alpha$和 $U_{\sigma'^T} = （T，\ O_T）$。

• 最后，我们将 τ 的中间部分 α 定义为这样一种形状：它由 α 中所有无法在不接触 S 和 T 的情况下分别从 U_α 和 V_α 到达的顶点所诱导的子图构成（但包含 S 和 T 的顶点）；此外，还包含完全位于 S 内的边或超边，以及完全位于 T 内的边或超边，并取 $U_\tau = (S, O_S)$和 $V_\tau = (T, O_T)$。

根据定义，显然 $\alpha = \sigma \circ \tau \circ \sigma'^{T}$。

例6.1.16. 图*6.2*展示了一个分解示例，其中只有一种顶点类型。

1. 如果从形状 α（记作 $\sigma \circ \sigma'^{T}$）开始，可以看到存在唯一的最小顶点分隔集，它由度为 5 的中间顶点构成，也就是既不属于 $U_{\sigma\circ\sigma'^{T}}$ 也不属于 $V_{\sigma\circ\sigma'^{T}}$ 的那个顶点。于是，α 被分解为左部分 σ、一个平凡中间部分 τ（图中未显示），其中包含 $V(\tau) = \{u\}$, $U_\tau = V_\tau = (u)$ $E(\tau) = \emptyset$，以及右部 σ'^{T}。

2. 如果从形状 α（记作 $\sigma \circ \tau \circ \sigma'^{T}$）开始，那么最左侧的最小顶点分隔集是度为 4 的顶点，而最右侧的最小顶点分隔集是度为 5 的顶点。于是，α 被分解为左部分 σ、中间部分 τ 和右部 σ'^{T}，它们都显示在图中。

例6.1.17。 图*6.3*展示了一个分解示例，其中包含两种顶点类型。我们用方形和圆形来图示表示这两种顶点。在此例中，我们假设，包含单个圆形顶点的集合的权重低于包含两个方形顶点的集合。

1. 如果从形状 $\sigma \circ \sigma'^{T}$ 开始，那么它可以唯一分解为左形状 σ 和右形状 σ'^{T} 的组合。在这种情形下，中间形状（图中未显示）是平凡的。

2. 如果从形状 $\sigma \circ \tau \circ \sigma'^{T}$ 开始，那么它可以唯一分解为左形状 σ、中间形状 τ 和右形状 σ'^{T} 的组合，这些形状在图中均有显示。

此时，左形状、中间形状和右形状的定义是很自然的。我们称形状 σ 为左形状，如果 σ 是真形状，V_σ 是 σ的最左且最右的最小权重分隔集，$V(\sigma)$中的每个顶点 $\setminus V_\sigma$都可以从 U_σ到达而不接触

V_σ，并且 σ 在 V_σ内没有完全包含其中的超边。类似地，我们称形状 τ 为适当中间形状，如果 τ 是真形状， U_τ 是 τ的最左最小权重分隔集，并且 V_τ是 τ的最右最小权重分隔集。我们不定义不适当中间形状；它们在机制证明中是必需的，但在此处并不需要。最后，我们称形状 σ^T 为右形状，如果它是左形状的转置。

6.1.3 系数矩阵

我们已经具备了关于形状和图矩阵的所有必要定义。为了应用这些技术工具，我们将给定的矩矩阵 Λ 分解为 $\Lambda = \sum \lambda_\alpha M_\alpha$ ，其中求和遍历所有形状 α。随后，我们将系数 λ_α 汇集成矩阵，称为系数矩阵，接下来将对其进行定义。技术工具的条件将以这些系数矩阵的形式给出。

我们先介绍一些关于不同指标形状集合的记号。给定一个矩矩阵 Λ，定义 $\mathcal{I}(\Lambda)$ 为矩阵形状 U 的集合，其中某个行或列索引 Λ的形状为 U。定义 $w_{max} = \max\{w(U) : U \in \mathcal{I}(\Lambda)\}$为 $\mathcal{I}(\Lambda)$ 中指标形状可能具有的最大权重。最后，定义 $\mathcal{I}_{mid}$为 $\mathcal{I}_{mid} = \{U : w(U) \leq w_{max},\ \forall U_i \in U,\ p_i = 1\}$。注意，在拉ademacher $G_{n,\ 1/2}$ 输入的情形下，我们有 $\mathcal{I}_{mid} = \{U : |U| \leq w_{max}\}$。

在伪校准中，我们只保留满足我们所选某些截断参数的形状。形式上，截断参数的满足条件定义如下。

定义 6.1.18 (Rademacher $G_{n,\ 1/2}$ 输入的设定下的截断参数)。对于整数 $D_{sos}, D_V \geq 0$，若形状 α 满足截断参数 D_{sos}, D_V，则称其满足以下条件：

- 与 U_α 和 V_α 对应的单项式的度至多为 $\frac{D_{sos}}{2}$ ——
- 左侧部分 t σ、 α 的中间部分 τ 和右部 σ' 满足边界 ds

$|V(\sigma)|,\ |V(\tau)|,\ |V(\sigma')| \le D_V$。

定义 6.1.19（张量主成分分析和稀疏主成分分析的截断参数）。对于整数D_{sos}、D_V、$D_E \ge 0$，称形状 α 满足截断参数 D_{sos}、D_V、D_E，如果

- 单项式 U_α 和 V_α 的度至多为 $\frac{D_{sos}}{2}$ ——
- α 的左部分 σ、中间部分 τ 和右部 σ'^T 满足界$|V(\sigma)|$、$|V(\tau)|$、$|V(\sigma'^T)| \le D_V$
- 对于每个 $e \in E(\alpha)$，$l_e \le D_E$。

我们还需要定义分析 Λ时可能出现的形状集合。给定一个矩矩阵 Λ，定义 $\mathcal{L} = \{\sigma$：σ是一个左形状 $U_\sigma \in \mathcal{I}(\Lambda)$ $V_\sigma \in \mathcal{I}_{mid}$，$|V(\sigma)| \le D_V$，$\forall e \in E(\sigma)$ $l_e \le D_E\}$。此外，给定 $V \in \mathcal{I}_{mid}$，定义 $\mathcal{L}_V = \{\sigma \in \mathcal{L}$： $V_\sigma \equiv V\}$。最后，给定 $U \in \mathcal{I}_{mid}$，定义 $\mathcal{M}_U = \{\tau$： τ是一个非平凡的正规中间形状 $U_\tau \equiv V_\tau \equiv U$，$|V(\tau)| \le D_V$，$\forall e \in E(\tau)$ $l_e \le D_E\}$。

现在我们可以定义系数矩阵了。给定一个矩矩阵 Λ，系数矩阵是一个矩阵 H，其行和列由左形状 σ，$\sigma' \in \mathcal{L}$进行索引。 H 称为 SoS 对称的，如果 $H(\sigma,\ \sigma')$ 在对称群的作用下保持不变；也就是说，如果我们对 U_σ 的顶点和 $U_{\sigma'}$ 的顶点（其中仅在同一类型内进行置换）应用相同的置换，那么矩阵项不会改变。

对于形状 τ，如果对于任意 $V_\sigma \ne U_\tau$ 或 $V_\tau \ne U_{\sigma'^T}$，系数矩阵 H 都满足 $H(\sigma,\ \sigma') = 0$，我们就称系数矩阵 H 为一个 τ- 系数矩阵。给定一个指标形状 U，我们定义 Id_U 为具有 $U_{Id_U} = V_{Id_U} = U$、不含其他顶点且不含边的形状。

如前所述，矩矩阵 Λ 的系数 λ_α 汇集起来构成系数矩阵，用以表述技术工具条件。给定一个形状 τ 和一个 τ-系数矩阵 H，我们考虑如下定义的矩阵值函数 $M_\tau^{fact}(H)$:

$$M_\tau^{fact}(H) = \sum_{\sigma \in \mathcal{L}_{U_\tau}, \sigma' \in \mathcal{L}_{V_\tau}} H(\sigma, \sigma') M_\sigma M_\tau M_{\sigma'}^T$$

这个定义的动机如下。首先注意到， Λ 本质上是一个形如 $\sum_{\tau}\sum_{\sigma,\ \sigma'} H(\sigma,\ \sigma')$ $M_{\sigma\circ\tau\circ\sigma'^{T}}$ 的表达式。对于每个 τ，其中的内层表达式看起来有点像 $M_{\tau}^{fact}(H)$。不过，两者在技术上有所不同。回想一下，如果展开图矩阵 M_{σ}，M_{τ} 和 $M_{\sigma'}^{T}$ 的定义，那么 $M_{\tau}^{fact}(H)$ 会分别对来自集合 Real(σ)，Real(τ)，Real(σ'^{T}) 的非等价实现求和。先验而言，无法保证对于从这些集合中选取的每组实现 φ_1，φ_2，φ_3，相应的标签子集 $\varphi_1(V(\sigma))$ $\varphi_2(V(\tau))$ $\varphi_3(V(\sigma'^{T}))\subseteq[n]$ 彼此不相交。不过，如果我们强制它们彼此不相交，就会得到一个与我们所需矩阵密切相关的矩阵。[191] 将通过强制实现彼此不相交而得到的这个矩阵称为 $M_{\tau}^{orth}(H)$。

正如他们所指出的，由于涉及自同构群的额外项，$M_{\tau}^{orth}(H)=\sum_{\sigma,\ \sigma'} H(\sigma,\ \sigma')$ $M_{\sigma\circ\tau\circ\sigma'^{T}}$并不成立。尽管如此，由于这一强制要求实现之间不重叠的条件，Λ 可以很容易地用 M^{orth}表示。事实上，正如他们通过仔细计数所证明的，$\Lambda=\sum_{U\in\mathcal{I}_{mid}} M_{Id_U}^{orth}($ $H_{Id_U})+\sum_{U\in\mathcal{I}_{mid}}\sum_{\tau\in\mathcal{M}_U} M_{\tau}^{orth}(H_{\tau})$，其中 H_{Id_U} 和 H_{τ}（将在下文正式定义）是由λ_{α}组装而成的简单系数矩阵

尽管我们希望使用 M^{orth}，但这套技术工具实际使用的是 M^{fact}，因为借助 M^{fact} 证明半正定性更容易，这是由于其乘积结构。该技术工具中的证明策略是证明从 M^{fact}到 M^{orth} （因此也包括 Λ）时产生的误差项以高概率可以忽略不计，从而完成半正定性证明。

给定矩阵值函数 Λ，我们构造以下系数矩阵。

定义 6.1.20。 给定矩阵值函数 $\Lambda=\sum_{\alpha:\ \alpha\text{ 是正规函数}}\lambda_{\alpha}M_{\alpha}$,

1. 对每个指标形状 $U\in\mathcal{I}_{mid}$以及每个 σ，$\sigma'\in\mathcal{L}_U$，设 $H_{Id_U}(\sigma,\sigma')=\frac{1}{|Aut(U)|}\lambda_{\sigma\circ\sigma'^{T}}$
2. 对每个 $U\in\mathcal{I}_{mid}$，$\tau\in\mathcal{M}_U$以及 σ，$\sigma'\in\mathcal{L}_U$，设 $H_{\tau}(\sigma,\sigma')=\frac{1}{|Aut\ U_{\tau}\ ()|\cdot|Aut(V_{\tau})|}\lambda_{\sigma\circ\tau\circ\sigma'^{T}}$

我们还需要给出最后一个定义，即系数矩阵 $H^{-\gamma,\ \gamma}$。为了处理近似半正定分解中的误差项，技术工具必须进一步分解左

形状 σ 为 $\sigma = \sigma_2 \circ \gamma$，其中 σ_2 和 γ 本身都是左形状。为了描述这一操作，需要以下定义。

给定矩矩阵 Λ，定义 $\Gamma = \{\gamma : \gamma$ 是一个非平凡左形状，满足 U_γ，$V_\gamma \in \mathcal{I}_{mid}$，$|V(\gamma)| \leq D_V$，$\forall e \in E(\gamma)\ l_e \leq D_E\}$。此外，给定 U，$V \in \mathcal{I}_{mid}$，使得 $w(U) > w(V)$，定义 $\Gamma_{U,\ V} = \{\gamma \in \Gamma : U_\gamma \equiv U,\ V_\gamma \equiv V\}$。最后，给定 $U \in \mathcal{I}_{mid}$，定义 $\Gamma_{U,\ *} = \{\gamma \in \Gamma : U_\gamma \equiv U\}$。

最后，我们在给定截断参数 D_V 的情况下定义系数矩阵 $H^{-\gamma,\ \gamma}$。给定一个形状 τ，满足 $U_\tau \equiv V_\tau$，左形状 $\gamma \in \Gamma_{*,\ U_\tau}$（因此，$\gamma \in \Gamma_{*,\ V_\tau}$），以及一个 τ-系数矩阵 H，定义 $H^{-\gamma,\ \gamma}$ 为一个 $(\gamma \circ \tau \circ \gamma^T)$-系数矩阵，其元素为

- $H^{-\gamma,\ \gamma}(\sigma,\ \sigma') = H(\sigma \circ \gamma,\ \sigma' \circ \gamma)$ 当且仅当 $|V(\sigma \circ \gamma)| \leq D_V$ 且 $|V(\sigma' \circ \gamma)| \leq D_V$。
- $H^{-\gamma,\ \gamma}(\sigma,\ \sigma') = 0$ 当 $|V(\sigma \circ \gamma)| > D_V$ 或 $|V(\sigma' \circ \gamma)| > D_V$ 时。

6.1.4 主要定理

对于问题 $\mathcal{P}$，令 $\Lambda_{\mathcal{P}}$ 为通过伪校准获得的矩矩阵。接下来，我们陈述技术工具为以高概率证明正定性所要求的条件。我们将使用以下系数矩阵之间的距离概念，它有助于界定截断误差。

定义 6.1.21。给定函数 $B_{norm}(\alpha)$，定义两个 τ 系数矩阵 H_τ 和 H'_τ 之间的距离 $d_\tau(H_\tau, H'_\tau)$ 为

$$d_\tau(H_\tau, H'_\tau) = \sum_{\sigma \in \mathcal{L}_{U_\tau}, \sigma' \in \mathcal{L}_{V_\tau}} |H'_\tau(\sigma, \sigma') - H_\tau(\sigma, \sigma')| B_{norm}(\sigma) B_{norm}(\tau) B_{norm}(\sigma')$$

我们还定义 Id_{Sym}，它是单位矩阵的 SoS 对称类比。对于矩阵指标 A，用 p_A 表示它对应的形式单项式（关于 SoS 程序变量）。定义 Id_{Sym} 为一个矩阵，使得 Id_{Sym} 的行和列

由矩阵指标 A，B 索引，其指标形状属于 $\mathcal{I}(\Lambda)$，且若 $p_A = p_B$，则为 $Id_{Sym}(A,\ B) = 1$；若 $p_A \neq p_B$，则为 $Id_{Sym}(A,\ B) = 0$。

为了陈述我们的参数，我们再介绍一些关于形状的记号。定义$\mathcal{M}'$为所有形状 α 的集合，使得 $|V(\alpha)| \leq 3D_V$、$\forall e \in E(\alpha)\ l_e \leq D_E$，并且所有边$e \in E(\alpha)$的重数都至多为 $3D_V$。注意，对于 Rademacher $G_{n,\ 1/2}$ 输入的情形，后两个条件并非必需；但在高斯输入的情形下则是必需的。在 Rademacher$G_{n,\ 1/2}$ 输入的情形，对于形状 α，定义 S_α为 α 的最左侧最小顶点分隔集，并定义 I_α 为 $V(\alpha) \setminus (U_\alpha \cup V_\alpha)$中孤立的顶点集合。在高斯 $\mathcal{N}(0,\ 1)$输入的情形，对于形状 $\alpha \in \mathcal{M}'$，如果删除所有重数至少为 2 的边，则定义$S_{\alpha,\ min}$ 为 α 的最左侧最小顶点分隔集。此外，定义 I_α 为$V(\alpha) \setminus (U_\alpha \cup V_\alpha)$中的顶点集合，使得与该顶点相 incident 的所有边的重数都至少为 2。

拉ademacher $G_{n,\ 1/2}$ 输入情形下的参数选择

我们首先说明本文将使用的一些参数，然后说明主定理陈述所需的主要条件；主定理将在下文给出。

设 $\varepsilon > 0$ 和 D_V，D_E 为截断参数。定义

- $q = 3\left\lceil D_V \ln n + \frac{\ln(\frac{1}{\varepsilon})}{3} + D_V \ln 5 + 3D_V^2 \ln 2 \right\rceil$
- $B_{vertex} = 6D_V \sqrt[4]{2eq}$
- $B_{norm}(\alpha) = B_{vertex}^{|V(\alpha)\setminus U_\alpha| + |V(\alpha)\setminus V_\alpha|} n^{\frac{w(V(\alpha)) + w(I_\alpha) - w(S_\alpha)}{2}}$
- $B(\gamma) = B_{vertex}^{|V(\gamma)\setminus U_\gamma| + |V(\gamma)\setminus V_\gamma|} n^{\frac{w(V(\gamma)\setminus U_\gamma)}{2}}$
- $N(\gamma) = (3D_V)^{2|V(\gamma)\setminus V_\gamma| + |V(\gamma)\setminus U_\gamma|}$
- $c(\alpha) = 100(3D_V)^{|U_\alpha\setminus V_\alpha| + |V_\alpha\setminus U_\alpha| + 2|E(\alpha)|} 2^{|V(\alpha)\setminus(U_\alpha \cup V_\alpha)|}$

在我们的应用中，如前所述，我们证明了度为 n^{ε} 的 SoS 的 SoS 下界，其中输入规模为 $n^{O(1)}$。在这种情况下，我们取 D_V， D_E 为 $n^{O(\varepsilon)}$ 的量级。因此，为简单起见，我们可以将这些参数理解为

$$q = n^{O(\varepsilon)}, B_{vertex} = n^{O(\varepsilon)}, B_{norm}(\alpha) = n^{O(\varepsilon)|V(\alpha)|} n^{\frac{w(V(\alpha))+w(I_\alpha)-w(S_\alpha)}{2}}$$

$$B(\gamma) = n^{O(\varepsilon)|V(\gamma)|} n^{\frac{w(V(\gamma)\setminus U_\gamma)}{2}}, N(\gamma) = n^{O(\varepsilon)|V(\gamma)|}, c(\alpha) = n^{O(\varepsilon)|V(\alpha)|}$$

高斯输入超图情形下的参数选择 现在，我们陈述在高斯输入超图的情形下，对更一般的技术工具陈述所需的参数。在此情形下，顶点类型至多有 t_{max} 种，并令 k 为超边的最大元数。在张量主成分分析的情形下，我们取 $t_{max} = 1$；在稀疏主成分分析的情形下，我们取 $k = t_{max} = 2$。对于所有 $\varepsilon > 0$ 和截断参数 D_V， D_E，定义

1. $q = \left\lceil 3D_V \ln n + \ln(\frac{1}{\varepsilon}) + (3D_V)^k \ln(D_E + 1) + 3D_V \ln 5 \right\rceil$ 2. $B_{vertex} = 6qD_V$ 3. $B_{edge}(e) = \left(400 D_V^2 D_E^2 q\right)^{l_e}$ 4. $B_{norm}(\alpha) = 2eB^{|V_{vertex}(\alpha)\setminus U_\alpha|+|V(\alpha)\setminus V_\alpha|} \left(\prod_{e\in E} B_{edge}(e)\right) n^{w(V(\alpha))+w(I_\alpha)-w(S_{\alpha,min})}$ 2 5. $B(\gamma) = B^{|V_{vertex}(\gamma)\setminus U_\gamma|+|V(\gamma)\setminus V_\gamma|} \left(\prod_{e\in E} B_{edge}(e)\right) n^{w(V(\gamma)\setminus U_\gamma)}$ 2 6. $N(\gamma) = (3D_V)^{2|V(\gamma)\setminus V_\gamma|+|V(\gamma)\setminus U_\gamma|}$ 7. $c(\alpha) = 100(3t_{max}D_V)^{|U_\alpha\setminus V_\alpha|+|V_\alpha\setminus U_\alpha|+k|E(\alpha)|}(2t_{max})^{|V(\alpha)\setminus(U_\alpha\cup V_\alpha)|}$

在我们的应用中，我们可以再次以更加简单的方式理解上述参数。更具体地说， k 是一个常数，而我们使用平方和次数 n^{ε}。于是，我们可以认为

形状 α 或 γ 中的每个顶点或边本质上都贡献了一个 n^{ε}的因子。因此，我们可以理解为

$$q = n^{O(\varepsilon)}, B_{vertex} = n^{O(\varepsilon)}, B_{edge} = n^{O(\varepsilon)|E(\alpha)|}$$

$$B_{norm}(\alpha) = n^{O(\varepsilon)(|V(\alpha)|+|E(\alpha)|)} n^{\frac{w(V(\alpha))+w(I_\alpha)-w(S_{\alpha,min})}{2}}$$

$$B(\gamma) = n^{O(\varepsilon)(|V(\gamma)|+|E(\gamma)|)} n^{\frac{w(V(\gamma)\setminus U_\gamma)}{2}}$$

$$N(\gamma) = n^{O(\varepsilon)|V(\gamma)|}, c(\alpha) = n^{O(\varepsilon)(|V(\alpha)|+|E(\alpha)|)}$$

机制陈述

如上所述，考虑适合该问题的参数选择。现在，我们可以用相应构造的伪校准矩阵 $\Lambda_{\mathcal{P}}$ 和系数矩阵 H_{Id_U}， H_τ 来陈述关于问题 $\mathcal{P}$ 的条件。

定义 6.1.22（半正定质量）。我们称 $\mathcal{P}$ 满足（半正定质量），如果对于所有 $U \in \mathcal{I}_{mid}$、$H_{Id_U} \succeq 0$,

定义 6.1.23（中间形状界）。我们称 $\mathcal{P}$ 满足（中间形状界），如果对于所有 $U \in \mathcal{I}_{mid}$ 和 $\tau \in \mathcal{M}_U$,

$$\begin{bmatrix} \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} & B_{norm}(\tau) H_\tau \\ B_{norm}(\tau) H_\tau^T & \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} \end{bmatrix} \succeq 0$$

定义 6.1.24（交项界）。对于某些 *SoS* 对称系数矩阵$\{H'_\gamma : \gamma \in \Gamma\}$， $\mathcal{P}$ 满足（交项界），即对于所有 U， $V \in \mathcal{I}_{mid}$ 其中 $w(U) > w(V)$，且对于所有 $\gamma \in \Gamma_{U,\ V}$,

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H^{-\gamma,\gamma}_{Id_V} \preceq H'_\gamma$$

定义 6.1.25 （截断误差界）。 对于某些 *SoS* 对称系数矩阵$\{H'_\gamma : \gamma \in \Gamma\}$， $\mathcal{P}$ 满足 （截断误差界）的条件是以下条件成立：当 $\|M_\alpha\| \leq B_{norm}(\alpha)$ 对所有 $\alpha \in \mathcal{M}'$ 成立时，

$$\sum_{U \in \mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) \succeq 6 \left(\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)|c(\gamma)} \right) Id_{sym}$$

最后，我们可以陈述主定理。

定理 6.1.26. 对于所有 $\varepsilon > 0$，若取上述定义的参数，且有 *SoS* 对称系数矩阵 $\{H'_\gamma : \gamma \in \Gamma\}$ 使得 $\mathcal{P}$ 满足 (半正定质量) , (中间形状界) , (交项界) 和 (截断误差界)，则其概率至少为 $1 - \varepsilon$, $\Lambda_{\mathcal{P}} \succeq 0$。

对于我们感兴趣的问题 $\mathcal{P}$，我们进行伪校准，将其分解为图矩阵，给出关于 $\mathcal{P}$ 的所需条件，并调用相关技术工具来证明我们的下界。

为我们的应用选择 H'_γ

在我们的应用中，我们按如下方式选择 H'_γ。

1. $H'_\gamma(\sigma, \sigma') = H_{Id_U}(\sigma, \sigma')$，当且仅当 $|V(\sigma \circ \gamma)| \leq D_V$且 $|V(\sigma' \circ \gamma)| \leq D_V$。 2. $H'_\gamma(\sigma, \sigma') = 0$，当 $|V(\sigma \circ \gamma)| > D_V$或 $|V(\sigma' \circ \gamma)| > D_V$ 时。

那么，我们需要界定的截断误差为

$$d_{Id_{U_\gamma}}(H_{Id_{U_\gamma}}, H'_\gamma) = \sum_{\substack{\sigma,\sigma' \in \mathcal{L}_{U_\gamma} : V(\sigma) \leq D_V, V(\sigma') \leq D_V, \\ |V(\sigma \circ \gamma)| > D_V \text{ or } |V(\sigma' \circ \gamma)| > D_V}} B_{norm}(\sigma) B_{norm}(\sigma') H_{Id_{U_\gamma}}(\sigma, \sigma')$$

6.2 植入的稍稠密子图的定性界

6.2.1 伪校准

我们将针对以下一对随机分布和植入分布进行伪校准，分别记为 ν 和 μ。

- 随机分布 ν：从 $G(n, 12)$ 中采样 G –
- 植入分布 μ：设 k 为一个整数，并设 $p > \frac{1}{2}$。从 $G(n, 12)$ 中采样一个图 G'。从顶点中随机选取一个子集 S，其中每个顶点都以概率 $\frac{k}{n}$ 独立地被选中。对于 S 中顶点的所有配对 i, j，重新随机化边 (i, j)；此时 (i, j) 属于该图的概率为 p。令 G 为所得的图。 –

我们假设输入给定为 $G_{i, j}$，其中 $i, j \in \binom{[n]}{2}$，且 $G_{i, j}$ 在边 (i, j) 存在于图中时为 1，否则为 -1。我们使用傅里叶基 χ_E，其定义为 $\chi_E(G) := \prod_{(i, j)\in E} G_{i, j}$。对于子集 $I \subseteq [n]$，定义 $x_I := \prod_{i\in I} x_I$。

引理 6.2.1. 设 $I \subseteq [n], E \subseteq \binom{[n]}{2}$. 那么，

$$\mathop{\mathbb{E}}_{\mu}[x_I\chi_E(G)] = \left(\frac{k}{n}\right)^{|I\cup V(E)|} (2p-1)^{|E|}$$

证明. 从 μ 中抽取 (G, S) 时，我们根据 $I \cup V(E) \subseteq S$ 是否成立进行条件化。

$$\begin{aligned}\mathop{\mathbb{E}}_{(G,S)\sim\mu}[x_I\chi_E(G)] = Pr_{(G,S)\sim\mu}[I \cup V(E) \subseteq S] \mathop{\mathbb{E}}_{(G,S)\sim\mu}[x_I\chi_E(G)|I \cup V(E) \subseteq S] \\ + Pr_{(G,S)\sim\mu}[I \cup V(E) \not\subseteq S] \mathop{\mathbb{E}}_{(G,S)\sim\mu}[x_I\chi_E(G)|I \cup V(E) \not\subseteq S]\end{aligned}$$

我们声称第二项为 0。具体而言，$\mathbb{E}_{(G, S)\sim\mu}[x_I\chi_E(G)|I \cup V(E) \not\subseteq S] = 0$ 因为当 $I \cup V(E) \not\subseteq S$ 时，要么 S 不包含 I 中的顶点，要么边 $(i, j) \in E$ 位于

在 S之外。如果 S 不包含 I中的顶点，那么 $x_I = 0$ ，因此该量为 0。而如果边 $(i, j) \in E$ 位于 S之外，由于这条边以概率 12 被抽取，通过取期望，该量 $\mathbb{E}_{(G, S)\sim\mu}[x_I\chi_E(G)|I \cup V(E) \not\subseteq S]$ 为 0。

最后，注意到 $Pr_{(G,S)\sim\mu}[I \cup V(E) \subseteq S] = \binom{k}{n}^{|I\cup V(E)|}$ 以及

$$\mathop{\mathbb{E}}_{(G,S)\sim\mu}[x_I\chi_E(G)|I \cup V(E) \subseteq S] = \mathop{\mathbb{E}}_{(G,S)\sim\mu}[\chi_E(G)|V(E) \subseteq S] = (2p-1)^{|E|}$$

最后一个等式成立，是因为对于每条边 $e \in E$，由于 e 以概率 p独立出现， χ_e 的期望值为 $1 \cdot p + (-1) \cdot (1-p) = 2p-1$。 ■

将 SoS 的次数定义为 $D_{sos} = n^{C_{sos}\varepsilon}$ ，其中 $C_{sos} > 0$ 是我们稍后选择的某个常数。并将截断参数定义为 $D_V = n^{C_V\varepsilon}$ ，其中 $C_V > 0$是某个常数。

注 6.2.2 (参数选择)。 我们首先将 ε 设为一个充分小的常数。基于这一选择，我们将 C_V 设为一个充分小的常数，以满足证明中用到的所有《不等式》。基于这些选择，我们可以将 C_{sos} 设得充分小，以满足所用的《不等式》。

现在我们将描述矩矩阵 Λ的分解。

定义6.2.3。 若形状 α 满足以下性质：

- α 是正规的，
- α 满足截断参数 D_{sos}, D_V。

则定义

$$\lambda_\alpha = \left(\frac{k}{n}\right)^{|V(\alpha)|}(2p-1)^{|E(\alpha)|}$$

推论6.2.4。 $\Lambda = \sum \lambda_\alpha M_\alpha$。

6.2.2 定性工具界

本节将证明半正定质量条件，以及中间形状界和交项界的定性版本。

引理 6.2.5（半正定质量）。对于所有 $U \in \mathcal{I}_{mid}$， $H_{Id_U} \succeq 0$

虽然直接证明这一点很容易，但我们希望引入恰当的记号，使本引理以及接下来要给出的定性界都能立即得出。因此，我们先陈述这些定性条件，再将它们一并证明。现在，我们定义以下量，用以刻画顶点对傅里叶系数的贡献，其中顶点属于 τ， γ。

定义 6.2.6。对于 $U \in \mathcal{I}_{mid}$和 $\tau \in \mathcal{M}_U$，定义 $S(\tau) = \binom{k}{n}^{|V(\tau)|-|U_\tau|}(2p-1)^{|E(\tau)|}$。对于所有 $U, V \in \mathcal{I}_{mid}$，其中 $w(U) > w(V)$ 且 $\gamma \in \Gamma_{U,V}$，定义 $\bar{S}(\gamma) = \binom{k}{n}^{|V(\gamma)|-\frac{|U_\gamma|+|V_\gamma|}{2}}(2p-1)^{|E(\gamma)|}$。

现在可以陈述我们的定性界，我们稍后将证明它们。

引理 6.2.7（定性中间形状界）。对于所有 $U \in \mathcal{I}_{mid}$ 和 $\tau \in \mathcal{M}_U$，

$$\begin{bmatrix} \frac{S(\tau)}{|Aut(U)|} H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)}{|Aut(U)|} H_{Id_U} \end{bmatrix} \succeq 0$$

在下面的定性交集项界中，我们使用第 6.1.4 节中H'_γ 的典范定义。

引理 6.2.8（定性交集项界）。对于所有 U、$V \in \mathcal{I}_{mid}$，其中 $w(U) > w(V)$，以及所有 $\gamma \in \Gamma_{U,V}$，

$$\frac{|Aut(V)|}{|Aut(U)|} \cdot \frac{1}{\bar{S}(\gamma)^2} H^{-\gamma,\gamma}_{Id_V} = H'_\gamma$$

为了证明这些界，我们定义以下量，用来捕捉 σ 内各顶点对傅里叶系数的贡献。

定义 6.2.9。 对于形状 $\sigma \in \mathcal{L}$，定义 $T(\sigma) = \binom{k}{n}^{|V(\sigma)|-\frac{|V_\sigma|}{2}}(2p-1)^{|E(\sigma)|}$。对于 $U \in \mathcal{I}_{mid}$，定义 v_U 为以 $\sigma \in \mathcal{L}$ 为索引的向量，使得 $v_U(\sigma) = T(\sigma)$（若满足 $\sigma \in \mathcal{L}_U$），否则为 0。

以下命题可直接由定义6.2.3得出。

命题 6.2.10。对于所有 $U \in \mathcal{I}_{mid}$， $\rho \in \mathcal{P}_U$， $H_{Id_U} = \frac{1}{|Aut(U)|} v_U v_U^T$。

命题 6.2.11。对于任意 $U \in \mathcal{I}_{mid}$和 $\tau \in \mathcal{M}_U$， $H_\tau = \frac{1}{|Aut(U)|^2} S(\tau) v_U v_U^T$。

第一个命题意味着，对于所有 $U \in \mathcal{I}_{mid}$, $H_{Id_U} \succeq 0$，这就是引理 6.2.5 中的半正定质量条件。引理 6.2.7 和引理 6.2.8 也很容易得出。

引理 *6.2.7* 的证明。

$$\begin{bmatrix} \frac{S(\tau)}{|Aut(U)|} H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)}{|Aut(U)|} H_{Id_U} \end{bmatrix} = \begin{bmatrix} \frac{S(\tau)}{|Aut(U)|} v_U v_U^T & \frac{S(\tau)}{|Aut(U)|^2} v_U v_U^T \\ \frac{S(\tau)}{|Aut(U)|^2} v_U v_U^T & \frac{S(\tau)}{|Aut(U)|} v_U v_U^T \end{bmatrix} \succeq 0$$

■

引理 6.2.8 的证明。固定满足 $|V(\sigma \circ \gamma)|, |V(\sigma' \circ \gamma)| \le D_V$ 的 $\sigma, \sigma' \in \mathcal{L}_U$。注意，$|V(\sigma)| - \frac{|V_\sigma|}{2} + |V(\sigma')| - \frac{|V_{\sigma'}|}{2} + 2(|V(\gamma)| - \frac{|U_\gamma|+|V_\gamma|}{2}) = |V(\sigma \circ \gamma \circ \gamma^T \circ \sigma'^T)|$。利用定义 6.2.3，我们可以很容易地验证 $\lambda_{\sigma\circ\gamma\circ\gamma^T\circ\sigma'^T} = T(\sigma)T(\sigma')S(\gamma)^2$。因此，我们有 $H'^{-\gamma,\gamma}_{Id_V}(\sigma,\sigma') = \frac{|Aut(U)|}{|Aut(V)|} S(\gamma)^2 H_{Id_U}(\sigma,\sigma')$。由于只要 $|V(\sigma \circ \gamma)|, |V(\sigma' \circ \gamma)| \le D_V$，就有 $H'_\gamma(\sigma,\sigma') = H_{Id_U}(\sigma,\sigma')$，证毕。

■

6.3 张量主成分分析的定性界

6.3.1 伪校准

定义 6.3.1 (松弛参数**)**。 将松弛参数定义为 $\Delta = n^{-C_\Delta \varepsilon}$ ，其中 $C_\Delta > 0$ 为一个常数。

我们将针对以下一对随机分布和植入分布进行伪校准，分别记为 ν 和 μ。

- 随机分布 ν：从 $\mathcal{N}(0, I_{[n]^k})$ 中采样 A。
- 植入分布 μ：令 λ，$\Delta > 0$。从 $\{-\frac{1}{\sqrt{\Delta n}}, 0, \frac{1}{\sqrt{\Delta n}}\}^n$ 中采样 u，其取值分别以概率 $\frac{\Delta}{2}$，$1-\Delta$，$\frac{\Delta}{2}$ 取得。然后从 $\mathcal{N}(0, I_{[n]^k})$ 中采样 B。设 $A = B + \lambda u^{\otimes k}$。

令厄米特多项式为 $h_0(x) = 1$，$h_1(x) = x$，$h_2(x) = x^2 - 1$，.....。对于 $a \in \mathbb{N}^{[n]^k}$ 和变量 A_e（其中 $e \in [n]^k$），定义 $h_a(A) := \prod_{e\in[n]^k} h_e(A_e)$。我们将使用这个厄米特基。

引理 6.3.2。令 $I \in \mathbb{N}^n, a \in \mathbb{N}^{[n]^k}$。对于 $i \in [n]$，令 $d_i = \sum_{i\in e\in[n]^k} a_e$。令 c 为使得 $I_i + d_i$ 非零的 i 的数量。那么，如果 $I_i + d_i$ 全为偶数，我们有

$$\mathop{\mathbb{E}}_{\mu}[u^I h_a(A)] = \Delta^c \left(\frac{1}{\sqrt{\Delta n}}\right)^{|I|} \prod_{e\in[n]^k} \left(\frac{\lambda}{(\Delta n)^{\frac{k}{2}}}\right)^{a_e}$$

否则，$\mathbb{E}_\mu[u^I h_a(v)] = 0$。

证明。当 $A \sim \mu$ 时，对于所有 $e \in [n]^k$，我们有 $A_e = B_e + \lambda \prod_{i\le k} u_{e_i}$。其中 $B_e \sim \mathcal{N}(0, 1)$。下面分析所需的期望何时非零。我们可以先以 u 为条件，并利用以下事实：对于固定的 t，$\mathbb{E}_{g\sim\mathcal{N}(0,\,1)}[h_k(g+t)] = t^k$，从而得到

$$\mathop{\mathbb{E}}_{(u_i,w_e)\sim\mu}[u^I h_a(A)] = \mathop{\mathbb{E}}_{(u_i)\sim\mu}[u^I \prod_{e\in[n]^k} (\lambda \prod_{i\le k} u_{e_i})^{a_e}] = \mathop{\mathbb{E}}_{(u_i)\sim\mu}[\prod_{i\in[n]} u_i^{I_i+d_i}] \prod_{e\in[n]^k} \lambda^{a_e}$$

注意，只有当所有 $I_i + d_i$ 均为偶数时，这才非零，此时

$$\mathop{\mathbb{E}}_{(u_i)\sim\mu}[\prod_{i\in[n]} u_i^{I_i+d_i}] = \Delta^c \left(\frac{1}{\sqrt{\Delta n}}\right)^{\sum_{i\le n} I_i+d_i} = \Delta^c \left(\frac{1}{\sqrt{\Delta n}}\right)^{|I|} \prod_{e\in[n]^k} \left(\frac{1}{(\Delta n)^{\frac{k}{2}}}\right)^{a_e}$$

这里用到了 $\sum_{e\in[n]^k} a_e = k\sum_{i\in[n]} d_i$ 这一事实。证明至此完成。 ■

将 SoS 次数定义为 $D_{sos} = n^{C_{sos}\varepsilon}$，其中 $C_{sos} > 0$ 是我们稍后选取的某个常数。并将截断参数定义为 $D_V = n^{C_V\varepsilon}$，$D_E = n^{C_E\varepsilon}$，其中 C_V，$C_E > 0$ 是一些常数。

注 6.3.3 (参数选择)。 我们首先将 ε 设为一个充分小的常数。根据 ε 的选择，我们将常数 $C_\Delta > 0$ 设得足够小，以确保植入分布定义良好。基于这些选择，我们仿照注 *6.2.2*，依次选择 C_V、C_E、C_{sos}。

图矩阵的底层图具有以下结构：其中有 n 个同类型的顶点，边是元数为 k 的有序超边。对于张量主成分分析，我们将使用以下记号。

- 对于指标形状 U 和顶点 i，将 $deg^U(i)$ 定义如下：若 $i \in V(U)$，则它是满足 $i \in V(A)$ 的唯一指标形状片段 $A \in U$ 的幂。否则，它为 0。

- 对于指标形状 U，定义 $deg(U) = \sum_{i\in V(U)} deg^U(i)$。这也是与 U 对应的单项式的度。

- 对于形状 α 和顶点 $i \in \alpha$，令 $deg^\alpha(i) = \sum_{i\in e\in E(\alpha)} l_e$。

- 对于任意形状 α，令 $deg(\alpha) = deg(U_\alpha) + deg(V_\alpha)$。

现在我们来描述矩矩阵 Λ的分解。

定义 *6.3.4. </style}*如果形状 α 满足以下性质：

- $deg^\alpha(i) + deg^{U_\alpha}(i) + deg^{V_\alpha}(i)$ 对于所有 $i \in V(\alpha)$均为偶数，

- α 是正规的，

- α 满足截断参数 D_{sos}、D_V、D_E。

然后定义

$$\lambda_\alpha = \Delta^{|V(\alpha)|}\left(\frac{1}{\sqrt{\Delta n}}\right)^{deg(\alpha)} \prod_{e\in E(\alpha)}\left(\frac{\lambda}{(\Delta n)^{\frac{k}{2}}}\right)^{l_e}$$

否则，定义 $\lambda_\alpha = 0$。

推论 6.3.5。 $\Lambda = \sum \lambda_\alpha M_\alpha$。

6.3.2 定性技术工具界

与植入的稍密子图一节中的做法一样，我们通过先陈述半正定质量条件以及定性中间形状界和交集项界，再引入适当的记号，以统一的方式证明它们。

引理 6.3.6 (半正定质量**)**。对所有 $U \in \mathcal{I}_{mid}$， $H_{Id_U} \succeq 0$

我们定义以下量，以刻画顶点 τ, γ对傅里叶系数的贡献。

定义 6.3.7. 对于 $U \in \mathcal{I}_{mid}$和 $\tau \in \mathcal{M}_U$，若对所有顶点 $i \in V(\tau) \setminus U_\tau \setminus V_\tau$， $deg^\tau(i)$ 均为偶数，则定义

$$S(\tau) = \Delta^{|V(\tau)|-|U_\tau|}\prod_{e\in E(\tau)}\left(\frac{\lambda}{(\Delta n)^{\frac{k}{2}}}\right)^{l_e}$$

否则，定义 $S(\tau) = 0$。对于所有满足 $w(U) > w(V)$且 $\gamma \in \Gamma_{U,V}$ 的 $U, V \in \mathcal{I}_{mid}$，若对 $V(\gamma) \setminus U_\gamma \setminus V_\gamma$中的所有顶点 i，$deg^\gamma(i)$均为偶数，则定义

$$S(\gamma) = \Delta^{|V(\gamma)|-\frac{|U_\gamma|+|V_\gamma|}{2}}\prod_{e\in E(\gamma)}\left(\frac{\lambda}{(\Delta n)^{\frac{k}{2}}}\right)^{l_e}$$

否则，定义 $S(\gamma) = 0$。

现在，我们用这些量来表述定性界。

引理 6.3.8 (定性中间形状界**)**。 对于所有 $U \in \mathcal{I}_{mid}$ 和 $\tau \in \mathcal{M}_U$，

$$\begin{bmatrix} \frac{S(\tau)}{|Aut(U)|} H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)}{|Aut(U)|} H_{Id_U} \end{bmatrix} \succeq 0$$

我们再次使用第 6.1.4 节中的 H'_γ 的典范定义。

引理 6.3.9 (定性交集项界**)**。 对于所有 $U, V \in \mathcal{I}_{mid}$ 满足 $w(U) > w(V)$ 以及所有 $\gamma \in \Gamma_{U,V}$，

$$\frac{|Aut(V)|}{|Aut(U)|} \cdot \frac{1}{S(\gamma)^2} H^{-\gamma,\gamma}_{Id_V} \preceq H'_\gamma$$

半正定质量条件的证明

我们引入一些记号，以便于证明定性界，同时也有助于揭示系数矩阵的结构。根据定义 6.3.4，复合形状 σ，σ' 时，要使 $\lambda_{\sigma\circ\sigma'}$ 非零，请注意，$\lambda_{\sigma\circ\sigma'}$ 中的所有顶点 i 都应满足 $deg^{\sigma\circ\sigma'}(i) + deg^{U_{\sigma\circ\sigma'}}(i) + deg^{V_{\sigma\circ\sigma'}}(i)$ 为偶数。为方便地部分捕捉这一概念，我们将引入奇偶向量的概念。

定义 6.3.10。 将奇偶向量 ρ 定义为其矩阵项属于 $\{0,1\}$的向量。对于$U \in \mathcal{I}_{mid}$，定义 $\mathcal{P}_U$ 为奇偶向量 ρ 的集合，其中其坐标由 U索引。

定义 6.3.11。 对于左形状 σ，定义 $\rho_\sigma \in \mathcal{P}_{V_\sigma}$，称为 σ的奇偶向量，它是满足对于每个顶点 $i \in V_\sigma$，ρ_σ 的第 i 个矩阵项等于$deg^{U_\sigma}(i) + deg^\sigma(i)$ 的奇偶性（即 $(\rho_\sigma)_i \equiv deg^{U_\sigma}(i) + deg^\sigma(i)$（模 2）的奇偶向量。对于 $U \in \mathcal{I}_{mid}$和 $\rho \in \mathcal{P}_U$，令 $\mathcal{L}_{U,\rho}$ 为满足 $\rho_\sigma = \rho$的所有左形状 $\sigma \in \mathcal{L}_U$ 的集合，也就是具有奇偶向量 ρ的所有左形状的集合。

对于形状 τ， 对于 τ系数矩阵 H_τ和奇偶向量 $\rho \in \mathcal{P}_{U_\tau}$， $\rho' \in \mathcal{P}_{V_\tau}$，定义 e

当 $\sigma \in \mathcal{L}_{U_\tau}$，$\rho$，$\sigma' \in \mathcal{L}_{V_\tau}$，$\rho'$时，矩阵 τ-系数矩阵 H_τ，ρ，ρ'为 H_τ，ρ，$\rho'(\sigma, \sigma') = H_\tau(\sigma, \sigma')$，否则为 0。以下命题显然成立。

命题 6.3.12。对于任意形状 τ和 τ-系数矩阵 H_τ，我们有等式$H_\tau = \sum_{\rho \in \mathcal{P}_{U_\tau}, \rho' \in \mathcal{P}_{V_\tau}} H_{\tau, \rho, \rho'}$

命题 6.3.13。对于任意 $U \in \mathcal{I}_{mid}$，$H_{Id_U} = \sum_{\rho \in \mathcal{P}_U} H_{Id_U, \rho, \rho}$

证明。对于任意 σ，$\sigma' \in \mathcal{L}_U$，根据定义 6.3.4，注意到要使 $H_{Id_U}(\sigma, \sigma')$ 非零，必须有 $\rho_\sigma = \rho_{\sigma'}$。

■

我们定义以下量，以捕捉 σ 内顶点对傅里叶系数的贡献。

定义 6.3.14。对于形状 $\sigma \in \mathcal{L}$，如果 $deg^\sigma(i) + deg^{U_\sigma}(i)$ 对于所有顶点 $i \in V(\sigma) \setminus V_\sigma$ 都是偶数，则定义

$$T(\sigma) = \Delta^{|V(\sigma)| - \frac{|V_\sigma|}{2}} \left(\frac{1}{\sqrt{\Delta n}} \right)^{deg(U_\sigma)} \prod_{e \in E(\sigma)} \left(\frac{\lambda}{(\Delta n)^{\frac{k}{2}}} \right)^{l_e}$$

否则，定义 $T(\sigma) = 0$。对于 $U \in \mathcal{I}_{mid}$和 $\rho \in \mathcal{P}_U$，将 v_ρ 定义为按 $\sigma \in \mathcal{L}$ 索引的向量，使得当 $\sigma \in \mathcal{L}_U$时，$v_\rho(\sigma)$ 为 $T(\sigma)$，否则为 0。

利用这一记号，可以很容易地证明半正定质量条件。

半正定质量条件引理 *6.3.6* 的证明。对于所有 $U \in \mathcal{I}_{mid}, \rho \in \mathcal{P}_U$，定义 6.3.4 意味着 $H_{Id_U, \rho, \rho} = \frac{1}{|Aut(U)|} v_\rho v_\rho^T$。因此，

$$H_{Id_U} = \sum_{\rho \in \mathcal{P}_U} H_{Id_U, \rho, \rho} = \frac{1}{|Aut(U)|} \sum_{\rho \in \mathcal{P}_U} v_\rho v_\rho^T \succeq 0$$

■

定性中间形状界

下一个命题说明了这样一个事实：复合形状 σ, τ, σ'^T 时，要使 $\lambda_{\sigma\circ\tau\circ\sigma'^T}$ 非零，合并顶点的度的奇偶性必须相应地相加。

命题 6.3.15。对于所有 $U \in \mathcal{I}_{mid}$ 和 $\tau \in \mathcal{M}_U$，存在两组奇偶向量 P_τ、$Q_\tau \subseteq \mathcal{P}_U$ 以及一个双射 $\pi:\ P_\tau \to Q_\tau$，使得 $H_\tau = \sum_{\rho\in P_\tau} H_{\tau,\ \rho,\ \pi(\rho)}$。

证明。根据定义 6.3.4，为使 $H_\tau(\sigma,\ \sigma')$ 非零，在 $\sigma\circ\tau\circ\sigma'$ 中，对于所有 $i \in U_\tau \cup V_\tau$，$deg^{U_\sigma}(i) + deg^{U_{\sigma'}}(i) + deg^{\sigma\circ\tau\circ\sigma'^T}(i)$ 都必须是偶数。换言之，对于任意 $\rho \in \mathcal{P}_U$，至多存在一个 $\rho' \in \mathcal{P}_U$，使得若取 $\sigma \in \mathcal{L}_{U,\ \rho,\ \sigma'} \in \mathcal{L}_U$，且 $H_\tau(\sigma,\ \sigma')$ 非零，那么 σ' 的奇偶性为 ρ'。此外，注意到 ρ' 确定了 ρ。接着，我们将 P_τ 取为满足 ρ' 存在的 ρ 的集合，将 Q_τ 取为 ρ' 的集合，并在此情形下定义 $\pi(\rho) = \rho'$。

■

直接验证定义 6.3.4 的条件可得以下命题。

命题 6.3.16. 对于任意 $U \in \mathcal{I}_{mid}$和 $\tau \in \mathcal{M}_U$，假设我们取 $\rho \in P_\tau$。设 π 为命题 *6.3.15* 中的双射，使得 $\pi(\rho) \in Q_\tau$。那么，$H_{\tau,\rho,\pi(\rho)} = \frac{1}{|Aut(U)|^2} S(\tau) v_\rho v_{\pi(\rho)}^T$。

现在我们可以证明定性中间形状界。

定性中间形状界引理 6.3.8 的证明。令 P_τ, Q_τ, π 来自命题 6.3.15。对于 $\rho, \rho' \in \mathcal{P}_U$，令 $W_{\rho,\rho'} = v_\rho (v_{\rho'})^T$。于是，$H_{Id_U} = \sum_{\rho\in\mathcal{P}_U} H_{Id_U,\rho,\rho} = \frac{1}{|Aut(U)|} \sum_{\rho\in\mathcal{P}_U} W_{\rho,\rho}$ 以及 $H_\tau = \sum_{\rho\in P_\tau} H_{\tau,\rho,\pi(\rho)} = \frac{1}{|Aut(U)|^2} S(\tau) \sum_{\rho\in P_\tau} W_{\rho,\pi(\rho)}$。我们有

$$\begin{bmatrix} \frac{S(\tau)}{|Aut(U)|} H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)}{|Aut(U)|} H_{Id_U} \end{bmatrix} = \frac{S(\tau)}{|Aut(U)|^2} \begin{bmatrix} \sum_{\rho\in\mathcal{P}_U} W_{\rho,\rho} & \sum_{\rho\in P_\tau} W_{\rho,\pi(\rho)} \\ \sum_{\rho\in P_\tau} W_{\rho,\pi(\rho)}^T & \sum_{\rho\in\mathcal{P}_U} W_{\rho,\rho} \end{bmatrix}$$

由于 $S(\tau)|Aut(U)|^2 \geq 0$，只需证明 $\begin{bmatrix} \sum_{\rho\in\mathcal{P}_U} W_{\rho} & \sum_{\rho\in P_\tau} W_{\rho,\pi(\rho} \\ \sum_{\rho\in P_\tau} W^T_{\rho\,\pi} & \rho \end{bmatrix} \succeq 0^{),(\rho)} \sum_{\rho\in\mathcal{P}_U} W_{\rho,\circ}$

考虑

$$\begin{bmatrix} \sum_{\rho\in\mathcal{P}_U} W_{\rho,\rho} & \sum_{\rho\in P_\tau} W_{\rho,\pi(\rho)} \\ \sum_{\rho\in P_\tau} W^T_{\rho,\pi(\rho)} & \sum_{\rho\in\mathcal{P}_U} W_{\rho,\rho} \end{bmatrix} = \begin{bmatrix} \sum_{\rho\in\mathcal{P}_U\setminus P_\tau} W_{\rho,\rho} & 0 \\ 0 & \sum_{\rho\in\mathcal{P}_U\setminus Q_\tau} W_{\rho,\rho} \end{bmatrix} + \begin{bmatrix} \sum_{\rho\in P_\tau} W_{\rho,\rho} & \sum_{\rho\in P_\tau} W_{\rho,\pi(\rho)} \\ \sum_{\rho\in P_\tau} W^T_{\rho,\pi(\rho)} & \sum_{\rho\in P_\tau} W_{\pi(\rho),\pi(\rho)} \end{bmatrix}$$

我们有 $\sum_{\rho\in\mathcal{P}_U\setminus P_\tau} W_{\rho,\ \rho} = \sum_{\rho\in\mathcal{P}_U\setminus P_\tau} v_\rho v_\rho^T \succeq 0_\circ^\rho$ 类似地，$\begin{bmatrix} 0 & \sum_{\rho\in\mathcal{P}_U\setminus Q_\tau} W_{\rho,\ \rho} \succeq 0 \\ \sum_{\rho\in\mathcal{P}_U\setminus P_\tau} W_{\rho,\ 0} & \sum_{\rho\in\mathcal{P}_U\setminus Q_\tau} W_{\rho,\ \rho} \end{bmatrix}$ 是半正定的。因此，上述表达式中的第一项

对于第二项,

$$\begin{aligned} \begin{bmatrix} \sum_{\rho\in P_\tau} W_{\rho,\rho} & \sum_{\rho\in P_\tau} W_{\rho,\pi(\rho)} \\ \sum_{\rho\in P_\tau} W^T_{\rho,\pi(\rho)} & \sum_{\rho\in P_\tau} W_{\pi(\rho),\pi(\rho)} \end{bmatrix} &= \sum_{\rho\in P_\tau} \begin{bmatrix} W_{\rho,\rho} & W_{\rho,\pi(\rho)} \\ W^T_{\rho,\pi(\rho)} & W_{\pi(\rho),\pi(\rho)} \end{bmatrix} \\ &= \sum_{\rho\in P_\tau} \begin{bmatrix} v_\rho v_\rho^T & v_\rho (v_{\pi(\rho)})^T \\ v_{\pi(\rho)} (v_\rho)^T & v_{\pi(\rho)} (v_{\pi(\rho)})^T \end{bmatrix} \\ &= \sum_{\rho\in P_\tau} \begin{bmatrix} v_\rho \\ v_{\pi(\rho)} \end{bmatrix} \begin{bmatrix} v_\rho & v_{\pi(\rho)} \end{bmatrix} \\ &\succeq 0 \end{aligned}$$

■

定性交集项界

与命题 6.3.15 类似，下一个命题揭示了这样一个事实：当我们复合形状 σ，γ，γ^T，σ'^T 时，若要使 $\lambda_{\sigma\circ\gamma\circ\gamma'^T\circ\sigma'^T}$ 非零，合并顶点的度的奇偶性就必须相应地相加。

我们使用以下记号。对于所有 $U,V\in\mathcal{I}_{mid}$，其中 $w(U)>w(V)$，对于 $\gamma\in\Gamma_{U,V}$ 以及奇偶向量 $\rho,\rho'\in\mathcal{P}_U$，定义 $\gamma\circ\gamma^T$-系数矩阵 $H^{-\gamma,\overset{\gamma}{Id_V},\rho,\rho'}$ 为 $H^{-\gamma,\overset{\gamma}{Id_V},\rho,\rho'}(\sigma,\sigma')=H^{-\gamma,\overset{\gamma}{Id_V}}(\sigma,\sigma')$，若 $\sigma\in\mathcal{L}_{U,\rho},\sigma'\in\mathcal{L}_{U,\rho'}$ 成立；否则为 0。

命题 6.3.17。对于所有 $U,V\in\mathcal{I}_{mid}$，其中 $w(U)>w(V)$，对于所有 $\gamma\in\Gamma_{U,V}$，存在一个奇偶向量集合 $P_\gamma\subseteq\mathcal{P}_U$，使得 $H^{-\gamma,\overset{\gamma}{Id_V}}=\sum_{\rho\in P_\gamma}H^{-\gamma,\overset{\gamma}{Id_V},\rho,,\rho}$。

证明。任取 $\rho\in\mathcal{P}_U$。对于 $\sigma\in\mathcal{L}_{U,\rho}$，$\sigma'\in\mathcal{L}_U$，由于 $H^{-\gamma,\overset{\gamma}{Id_V}}(\sigma,\sigma')=\frac{\lambda_{\sigma\circ\gamma\circ\gamma^T\circ\sigma'^T}()}{|Aut\ V|}$，$H^{-\gamma,\overset{\gamma}{Id_V}}(\sigma,\sigma')$ 非零当且仅当 $\lambda_{\sigma\circ\gamma\circ\gamma^T\circ\sigma'^T}$ 非零。根据定义 6.3.4，该量非零的必要条件（但非充分条件）是 σ' 的奇偶向量也必须是 ρ。另请注意，存在一个奇偶向量集合 P_γ ρ，使得 $H^{-\gamma,\overset{\gamma}{Id_V},\rho,,\rho}$ 非零，且它们的和恰为 $H^{-\gamma,\overset{\gamma}{Id_V}}$。

■

对于所有 U，$V\in\mathcal{I}_{mid}$，其中 $w(U)>w(V)$，对于所有 $\gamma\in\Gamma_{U,V}$ 以及奇偶向量 $\rho\in\mathcal{P}_U$，定义矩阵 $H'_{\gamma,\rho,\rho}$ 如下：若 σ，$\sigma'\in\mathcal{L}_{U,\rho}$，则 $H'_{\gamma,\rho,\rho}(\sigma,\sigma')=H'_\gamma(\sigma,\sigma')$，否则为 0。以下命题可由定义直接得出。

命题 6.3.18。对于所有 U、$V\in\mathcal{I}_{mid}$，其中 $w(U)>w(V)$，对于 $\gamma\in\Gamma_{U,V}$、$H'_\gamma=\sum_{\rho\in P_\gamma}H'_{\gamma,\rho,\rho}$。

命题 6.3.19。对于所有 U、$V\in\mathcal{I}_{mid}$，其中 $w(U)>w(V)$，对于所有 $\gamma\in\Gamma_{U,V}$ 和 $\rho\in P_\gamma$，

$$H^{-\gamma,\gamma}_{Id_V,\rho,\rho}=\frac{|Aut(U)|}{|Aut(V)|}S(\gamma)^2H'_{\gamma,\rho,\rho}$$

证明。固定 σ，$\sigma'\in\mathcal{L}_{U,\rho}$，使得 $|V(\sigma\circ\gamma)|$，$|V(\sigma'\circ\gamma)|\le D_V$。注意 $|V(\sigma)|-\frac{|V_\sigma|}{2}+|V(\sigma')|-\frac{|V_{\sigma'}|}{2}+2(|V(\gamma)|-\frac{|U_\gamma|+|V_\gamma|}{2})=|V(\sigma\circ\gamma\circ\gamma^T\circ\sigma'^T)|$。根据定义 6.3.4，

我们可以很容易地验证 $\lambda_{\sigma\circ\gamma\circ\gamma^T\circ\sigma'^T} = T(\sigma)T(\sigma')S(\gamma)^2$。因此，$H^{-\gamma,\gamma}_{Id_V,\rho,\rho}(\sigma,\sigma') = \frac{|Aut\ U|}{|Aut\ V|}S(\gamma)^2 H_{Id_U,\rho,\rho}(\sigma,\sigma')$。由于只要有 $|V(\sigma\circ\gamma)|, |V(\sigma'\circ\gamma)| \le D_V$，就有 $H'_{\gamma,\rho,\rho}(\sigma,\sigma') = H_{Id_U,\rho,\rho}(\sigma,\sigma')$，故证明完成。

■

由此，我们可以证明定性交集项界。

定性交集项界引理 *6.3.9* 的证明。我们有

$$\frac{|Aut(V)|}{|Aut(U)|}\cdot\frac{1}{S(\gamma)^2}H^{-\gamma,\gamma}_{Id_V} = \sum_{\rho\in P_\gamma}\frac{|Aut(V)|}{|Aut(U)|}\cdot\frac{1}{S(\gamma)^2}H^{-\gamma,\gamma}_{Id_V,\rho,\rho} = \sum_{\rho\in P_\gamma}H'_{\gamma,\rho,\rho} \preceq \sum_{\rho\in\mathcal{P}_U}H'_{\gamma,\rho,\rho}$$
$$= H'_\gamma$$

其中我们用到了这样一个事实：对于所有 $\rho\in\mathcal{P}_U$，我们有 $H'_{\gamma,\ \rho,\ \rho}\succeq 0$。

■

6.4 稀疏主成分分析的定性界

6.4.1 伪校准

定义 6.4.1（松弛参数）。定义松弛参数为 $\Delta = d^{-C_\Delta\varepsilon}$，其中常数为 $C_\Delta > 0$。

我们将针对以下一对随机分布和植入分布进行伪校准，分别记为 ν 和 μ。

- 随机分布 ν：$v_1,\dots,v_m$ 从 $\mathcal{N}(0,I_d)$ 中采样，并取 S 为行向量为 $v_1,\dots,v_m$ 的恒等矩阵。

- 植入分布 μ：从 $\{-\frac{1}{\sqrt{k}},0,\frac{1}{\sqrt{k}}\}^d$ 中抽取 u，其中各值分别以概率 $\frac{k}{2d}$、$1-\frac{k}{d}$、$\frac{k}{2d}$ 取到。然后按如下方式抽取 v_1、……、v_m。对于每个 $i\in[m]$，以概率 Δ 从 $\mathcal{N}$ 中抽取 $v_i(0, I_d+\lambda uu^T)$，以概率 $1-\Delta$ 从 $\mathcal{N}$ 中抽取 $v_i(0, I_d)$。最后，令 S 为行 v_1、……、v_m 组成的 $m\times d$ 矩阵。

我们将再次使用厄米特多项式基。对于 $a \in \mathbb{N}^{m\times d}$ 和变量v_i、j（其中 $i \in [m]$、$j \in [n]$），定义 $h_a(v) := \prod_{i\in[m],\, j\in[n]} h_{a_i,\, j}(v_i,\, j)$。对于非负整数 t，定义 $t!! = (2t)!$ $t!\ 2^t = 1 \times 3 \times \ldots \times t$，如果 t 为奇数则取该值，否则取 0。

引理 6.4.2。设 $I \in \mathbb{N}^d, a \in \mathbb{N}^{m\times d}$。对于 $i \in [m]$，令 $e_i = \sum_{j\in[d]} a_{ij}$；对于 $j \in [d]$，令$f_j = I_j + \sum_{i\in[m]} a_{ij}$。设 c_1（分别为 c_2）表示满足 $e_i > 0$（分别为$f_j > 0$）的 i（分别为 j）的数量。那么，如果 e_i, f_j 均为偶数，则有

$$\mathop{\mathbb{E}}_{\mu}[u^I h_a(v)] = \left(\frac{1}{\sqrt{k}}\right)^{|I|} \left(\frac{k}{d}\right)^{c_2} \Delta^{c_1} \prod_{i\in[m]} (e_i - 1)!! \prod_{i,j} \frac{\sqrt{\lambda}^{a_{ij}}}{\sqrt{k}^{a_{ij}}}$$

否则，$\mathbb{E}_\mu[u^I h_a(v)] = 0$。

证明。$v_1, \ldots\ldots, v_m \sim \mu$ 可以写成 $v_i = g_i + \sqrt{\lambda} b_i l_i u$，其中 $g_i \sim \mathcal{N}(0, I_d)$ $l_i \sim \mathcal{N}(0, 1)$ $b_i \in \{0, 1\}$，其中 $b_i = 1$ 以概率 Δ取值。我们来分析所求期望何时非零。首先，我们可以在给定 b_i，l_i，u 的条件下，利用以下事实：对于固定的 t，$\mathbb{E}_{g\sim\mathcal{N}(0,\, 1)}[h_k(g + t)] = t^k$，从而得到

$$\begin{aligned}\mathop{\mathbb{E}}_{(u,l_i,b_i,g_i)\sim\mu}[u^I h_a(v)] &= \mathop{\mathbb{E}}_{(u,l_i,b_i)\sim\mu}[u^I \prod_{i,j} (\sqrt{\lambda} b_i l_i u_j)^{a_{ij}}] \\ &= \mathop{\mathbb{E}}_{(u,l_i,b_i)\sim\mu}[\prod_{i\in[m]} (b_i l_i)^{e_i} \prod_{j\in[d]} u_j^{f_j}] \prod_{i,j} \sqrt{\lambda}^{a_{ij}}\end{aligned}$$

要使其非零，满足 $e_i > 0$ 的 c_1 指标集合 i 不应被重新采样，否则 $b_i = 0$；其中每个指标都以概率 Δ 独立地发生。满足 $f_j > 0$ 的 c_2指标集合 j 应当使 u_j 非零；其中每个指标都以概率 $\frac{k}{d}$ 独立地发生。由于 l_i, u_j 在 ν 中的期望为零，因此我们需要 e_i, f_j 为偶数。于是期望变为

$$\Delta^{c_1} \left(\frac{k}{d}\right)^{c_2} \mathop{\mathbb{E}}_{(u,l_i)\sim\mu}[\prod_{i\in[m]} l_i^{e_i} \prod_{j\in[d]} u_j^{f_j}] \prod_{i,j} \sqrt{\lambda}^{a_{ij}} = \left(\frac{1}{\sqrt{k}}\right)^{|I|} \left(\frac{k}{d}\right)^{c_2} \Delta^{c_1} \prod_{i\in[m]} (e_i - 1)!! \prod_{i,j} \frac{\sqrt{\lambda}^{a_{ij}}}{\sqrt{k}^{a_{ij}}}$$

最后一个等式成立，是因为对于每个满足 u_j 非零的 j，都有 $u_j^t = (\frac{1}{\sqrt{k}})^t$

并且 $\mathbb{E}_{g\sim\mathcal{N}(0,\ 1)}[g^t] = (t-1)!!$，如果 t 是偶数。 ■

将 SoS 次数定义为 $D_{sos} = d^{C_{sos}\varepsilon}$，其中 $C_{sos} > 0$ 是我们稍后选定的某个常数。将截断参数定义为 $D_V = d^{C_V\varepsilon}$，$D_E = d^{C_E\varepsilon}$，其中 C_V 和 $C_E > 0$ 是某些常数。关于参数选择，尽管我们研究的是一个不同的问题，备注 6.3.3 仍可直接应用。

图矩阵的底层图具有以下结构：存在两种类型的顶点——d 类型 1 顶点对应空间的维度，m 类型 2 顶点对应不同的输入向量。形状对应于二分图，边连接不同类型的顶点。对于稀疏主成分分析，我们将使用以下记号。

- 对于形状 α 和类型 $t \in \{1,2\}$，令 $V_t(\alpha)$ 表示 $V(\alpha)$ 中属于类型 t 的顶点。令 $|\alpha|_t = |V_t(\alpha)|$。

- 对于指标形状 U 和顶点 i，定义 $deg^U(i)$ 如下：如果 $i \in V(U)$，则它是满足 $i \in V(A)$ 的唯一指标形状片段 $A \in U$ 的幂。否则，它为 0。

- 对于指标形状 U，定义 $deg(U) = \sum_{i\in V(U)} deg^U(i)$。这也是单项式 p_U 的度。

- 对于 α 中的形状 α 和顶点 i，令 $deg^\alpha(i) = \sum_{i\in e\in E(\alpha)} l_e$。

- 对于任意形状 α，令 $deg(\alpha) = deg(U_\alpha) + deg(V_\alpha)$。

- 对于指标形状 $U \in \mathcal{I}_{mid}$ 和类型 $t \in \{1,2\}$，如果存在，则令 $U_t \in U$ 表示 U 中类型为 t 的指标形状片段；否则，定义 U_t 为 $\emptyset$。注意，这是定义良好的，因为对于每种类型 t，U 中至多有一个类型为 t 的指标形状片段，因为 $U \in \mathcal{I}_{mid}$。此外，用 $|U|_t$ 表示元组 U_t 的长度。

现在我们将描述矩矩阵 Λ 的分解。

定义 6.4.3。 如果一个形状 α 满足以下性质：

- U_α 和 V_α 都只包含类型为 1 的指标形状片段，
- $deg^\alpha(i) + deg^{U_\alpha}(i) + deg^{V_\alpha}(i)$ 对所有 $i \in V(\alpha)$ 都是偶数，
- α 是正规的，
- α 满足截断参数 D_{sos}、D_V、D_E。

然后定义

$$\lambda_\alpha = \left(\frac{1}{\sqrt{k}}\right)^{deg(\alpha)} \left(\frac{k}{d}\right)^{|\alpha|_1} \Delta^{|\alpha|_2} \prod_{j \in V_2(\alpha)} (deg^\alpha(j) - 1)!! \prod_{e \in E(\alpha)} \frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}}$$

否则，定义 $\lambda_\alpha = 0$。

推论 6.4.4。 $\Lambda = \sum \lambda_\alpha M_\alpha$。

6.4.2 定性工具界

在本节中，我们将证明主要的半正定质量条件，并获得另外两个条件的定性界，以便在完整验证中重复使用。与前几节一样，我们将先陈述这些界，引入记号，然后以统一的方式证明它们。

引理 6.4.5（半正定质量）。对于所有 $U \in \mathcal{I}_{mid}$， $H_{Id_U} \succeq 0$

我们定义以下量，以刻画 τ， γ 内的顶点对傅里叶系数的贡献。

定义 6.4.6. 对于 $U \in \mathcal{I}_{mid}$和 $\tau \in \mathcal{M}_U$，若对于所有顶点 $i \in V(\tau) \setminus U_\tau \setminus V_\tau$， $deg^\tau(i)$ 均为偶数，则定义

$$S(\tau) = \left(\frac{k}{d}\right)^{|\tau|_1 - |U_\tau|_1} \Delta^{|\tau|_2 - |U_\tau|_2} \prod_{j \in V_2(\tau) \setminus U_\tau \setminus V_\tau} (deg^\tau(j) - 1)!! \prod_{e \in E(\tau)} \frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}}$$

否则，定义 $S(\tau) = 0$。对于所有满足 $w(U) > w(V)$ 且 $\gamma \in \Gamma_{U,V}$ 的 $U, V \in \mathcal{I}_{mid}$，若对于 $V(\gamma) \setminus U_\gamma \setminus V_\gamma$ 中的所有顶点 i，$deg^\gamma(i)$ 均为偶数，则定义

$$S(\gamma) = \left(\frac{k}{d}\right)^{|\gamma|_1 - \frac{|U_\gamma|_1 + |V_\gamma|_1}{2}} \Delta^{|\gamma|_2 - \frac{|U_\gamma|_2 + |V_\gamma|_2}{2}} \prod_{j \in V_2(\gamma) \setminus U_\gamma \setminus V_\gamma} (deg^\gamma(j) - 1)!! \prod_{e \in E(\gamma)} \frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}}$$

否则，定义 $S(\gamma) = 0$。

为了得到最佳界，对正态分布进行离散化会很方便。以下事实可由高斯求积的标准结果推出，参见例如 [61,引理 4.3]。

事实 6.4.7 (正态分布的离散化**)**。存在一个绝对常数 C_{disc}，使得对于任意正整数 D，都存在一个定义在实数上的分布 $\mathcal{E}$，其支撑集包含 D 个点 p_1、...、p_D，使得 $|p_i| \leq C_{disc}\sqrt{D}$ 对所有 $i \leq D$ 和 $\mathbb{E}_{g \sim \mathcal{E}}[g^t] = \mathbb{E}_{g \sim \mathcal{N}(0,\ 1)}[g^t]$ 均成立，其中 $t = 0$ 为 1、... 、 $2D - 1$。

定义 6.4.8。对于任意形状 τ，假设 $U' = (U_\tau)_2$、$V' = (V_\tau)_2$ 分别是 U_τ、V_τ 中的 *2* 型顶点。定义 $R(\tau) = (C_{disc}\sqrt{D_E})^{\sum_{j \in U' \cup V'} deg^\tau(j)}$。

现在我们可以陈述定性界了。

引理 6.4.9（定性中间形状界）。对于所有 $U \in \mathcal{I}_{mid}$ 和 $\tau \in \mathcal{M}_U$，

$$\begin{bmatrix} \frac{S(\tau)R(\tau)}{|Aut(U)|} H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)R(\tau)}{|Aut(U)|} H_{Id_U} \end{bmatrix} \succeq 0$$

我们再次使用第 6.1.4 节中 H'_γ 的典范定义。

引理 6.4.10 (定性交集项界**)**。对于所有满足 U、$V \in \mathcal{I}_{mid}$ 的 $w(U) > w(V)$ 以及所有 $\gamma \in \Gamma_{U、V}$，

$$\frac{|Aut(V)|}{|Aut(U)|} \cdot \frac{1}{S(\gamma)^2 R(\gamma)^2} H_{Id_V}^{-\gamma,\gamma} \preceq H'_\gamma$$

半正定质量条件的证明

这里的大部分记号和分析与张量主成分分析中的情形相似；由于稀疏主成分分析的应用中存在两种顶点类型，我们只需对其进行适当修改。当我们从定义 6.4.3 中复合形状 σ、σ' 时，为使 $\lambda_{\sigma\circ\sigma'}$ 非零，注意到 $\lambda_{\sigma\circ\sigma'}$中的所有顶点 i都应使 $deg^{\sigma\circ\sigma'}(i)+deg^{U_{\sigma\circ\sigma'}}(i)+deg^{V_{\sigma\circ\sigma'}}(i)$ 为偶数。为方便地表达这一概念，我们再次使用奇偶向量这一概念。

定义 6.4.11. 定义奇偶向量 ρ 为其矩阵项属于 $\{0,1\}$的向量。对于$U\in\mathcal{I}_{mid}$，定义 $\mathcal{P}_U$ 为奇偶向量 ρ 的集合，其坐标由 U_1后接 U_2进行索引。

定义 6.4.12. 对于左形状 σ，定义 $\rho_\sigma\in\mathcal{P}_{V_\sigma}$（称为 σ的奇偶向量）为满足以下条件的奇偶向量：对于每个顶点 $i\in V_\sigma$，ρ_σ 的第 i 个矩阵项等于$deg^{U_\sigma}(i)+deg^{\sigma}(i)$的奇偶性，也就是说，$(\rho_\sigma)_i\equiv deg^{U_\sigma}(i)+deg^{\sigma}(i)$（模 2）。对于 $U\in\mathcal{I}_{mid}$和 $\rho\in\mathcal{P}_U$，令 $\mathcal{L}_{U,\rho}$ 为满足 $\rho_\sigma=\rho$的所有左形状 $\sigma\in\mathcal{L}_U$ 的集合，也就是奇偶向量为 ρ的所有左形状的集合。

对于形状 τ、τ系数矩阵 H_τ 和奇偶向量 $\rho\in\mathcal{P}_{U_\tau}$，$\rho'\in\mathcal{P}_{V_\tau}$，定义 τ-系数矩阵 $H_{\tau,\rho,\rho'}$ 如下：当 $H_{\tau,\rho,\rho'}(\sigma,\sigma')=H_\tau(\sigma,\sigma')$ 时，取该值；否则取 0。这立即推出以下命题。

命题 6.4.13。对于任意形状 τ和 τ-系数矩阵 H_τ，等式成立$H_\tau=\sum_{\rho\in\mathcal{P}_{U_\tau},\rho'\in\mathcal{P}_{V_\tau}}H_{\tau,\rho,\rho'}$

命题 6.4.14。对于任意 $U\in\mathcal{I}_{mid}$，$H_{Id_U}=\sum_{\rho\in\mathcal{P}_U}H_{Id_U,\rho,\rho}$

证明。对于任意 σ，$\sigma'\in\mathcal{L}_U$，根据定义 6.4.3，注意到要使 $H_{Id_U}(\sigma,\sigma')$ 非零，必须有 $\rho_\sigma=\rho_{\sigma'}$。

■

现在我们将正态分布离散化，同时匹配前 $2D_E-1$ 个矩。

定义 6.4.15。 令 $\mathcal{D}$ 为实数上的一个分布，它通过在事实 *6.4.7* 中令 $D = D_E$而得到。因此，特别地，对于从 $\mathcal{D}$中采样的任意 x，我们有 $|x| \leq C_{disc}\sqrt{D_E}$；并且对于 $t \leq 2D_E - 1$，$\mathbb{E}_{x\sim\mathcal{D}}[x^t] = (t-1)!!$。

我们定义以下量，以捕捉 σ 内顶点对傅里叶系数的贡献。

定义 6.4.16。 对于形状 $\sigma \in \mathcal{L}$，若对所有顶点 $i \in V(\sigma) \setminus V_\sigma$，$deg^\sigma(i) + deg^{U\sigma}(i)$ 均为偶数，则定义

$$T(\sigma) = \left(\frac{1}{\sqrt{k}}\right)^{deg(U_\sigma)} \left(\frac{k}{d}\right)^{|\sigma|_1 - \frac{|V_\sigma|_1}{2}} \Delta^{|\sigma|_2 - \frac{|V_\sigma|_2}{2}} \prod_{j\in V_2(\sigma)\setminus V_\sigma} (deg^\sigma(j)-1)!! \prod_{e\in E(\sigma)} \frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}}$$

否则，定义 $T(\sigma) = 0$。

定义 6.4.17。 设 $U \in \mathcal{I}_{mid}$。设 x_i 对于 $i \in U_2$ 为变量。将它们统记为 x_{U_2}。对于 $\rho \in \mathcal{P}_U$，定义 $v_{\rho,}{}^{x_{U_2}}$ 为以左形状 $\sigma \in \mathcal{L}$ 为索引的向量，其第 σ 项为 $T(\sigma)\prod_{i\in U_2} x_i^{deg^\sigma(i)}$，若 $\sigma \in \mathcal{L}_{U,\rho}$；否则为 0。

以下命题显而易见，并立即推出半正定质量条件。

命题 6.4.18。 对于任意 $U \in \mathcal{I}_{mid}, \rho \in \mathcal{P}_U$，假设 x_i 对于 $i \in U_2$ 是从 $\mathcal{D}$ 中抽取的随机变量。那么，$H_{Id_U,\rho,\rho} = \frac{1}{|Aut(U)|}\underline{\mathbb{E}_x[v_{\rho,x_{U_2}} v_{\rho,x_{U_2}}^T]}$。

证明。 注意，对于 σ，$\sigma' \in \mathcal{L}_U$，$\rho$和 $t \in \{1,2\}$，$(|\sigma|_t - \frac{|V_\sigma|_t}{2}) + (|\sigma'|_t - \frac{|V_{\sigma'}|_t}{2}) = |\sigma\circ\sigma'|_t$。验证定义 6.4.3 的条件并使用定义 6.4.15，即可得出结论。

■

半正定质量条件引理 *6.4.5* 的证明。 根据上述命题，我们有 $H_{Id_U} = \sum_{\rho\in\mathcal{P}_U} H_{Id_U,\rho,}$ $\rho \succeq 0$。

■

定性中间形状界

下一个命题阐明了这样一个事实：当我们复合形状 σ, τ, σ'^T 时，要使$\lambda_{\sigma\circ\tau\circ\sigma'^T}$ 非零，合并顶点的度的奇偶性必须相应地相加。

命题 6.4.19. 对于所有 $U \in \mathcal{I}_{mid}$ 和 $\tau \in \mathcal{M}_U$，都存在两组奇偶向量P_τ、$Q_\tau \subseteq \mathcal{P}_U$以及一个双射 $\pi\colon P_\tau \to Q_\tau$，使得 $H_\tau = \sum_{\rho\in P_\tau} H_{\tau,\ \rho,\ \pi(\rho)}$。

证明. 根据定义 6.4.3，要使 $H_\tau(\sigma, \sigma')$ 非零，我们必须满足：在$\sigma \circ \tau \circ \sigma'$中，对于所有 $i \in U_\tau \cup V_\tau$，$deg^{U_\sigma}(i) + deg^{U_{\sigma'}}(i) + deg^{\sigma\circ\tau\circ\sigma'^T}(i)$ 都必须是偶数。换句话说，对于任意 $\rho \in \mathcal{P}_U$，至多存在一个 $\rho' \in \mathcal{P}_U$，使得若我们取 $\sigma \in \mathcal{L}_{U,\ \rho,\ \sigma'} \in \mathcal{L}_U$，且 $H_\tau(\sigma, \sigma')$ 非零，那么 σ'的奇偶性为 ρ'。此外，注意到 ρ'决定了 ρ。接着，我们令 P_τ 为所有使得 ρ' 存在的 ρ 的集合，令 Q_τ 为 ρ' 的集合，并在此情形下定义 $\pi(\rho) = \rho'$。

■

命题 6.4.20。对于任意 $U \in \mathcal{I}_{mid}$和 $\tau \in \mathcal{M}_U$，假设我们取 $\rho \in P_\tau$。令 π为命题 *6.4.19* 中的双射，使得 $\pi(\rho) \in Q_\tau$。令 $U' = (U_\tau)_2$、$V' = (V_\tau)_2$分别为 U_τ、V_τ 中的 *2* 型顶点。令 x_i（$i \in U' \cup V'$）为从 $\mathcal{D}$ 中独立抽样的随机变量。定义 $x_{U'}$（分别为 $x_{V'}$）为变量 x_i 的子集，其中 $i \in U'$（分别为 $i \in V'$）。那么，

$$H_{\tau,\rho,\pi(\rho)} = \frac{1}{|Aut(U)|^2} S(\tau) \mathop{\mathbb{E}}_{x} \left[v_{\rho,x_{U'}} \left(\prod_{i\in U'\cup V'} x_i^{deg^\tau(i)} \right) v^T_{\pi(\rho),x_{V'}} \right]$$

证明。对于$\sigma \in L_{U,\rho,}\sigma' \in \mathcal{L}_{U,\pi(\rho)}$和 $t \in \{1, 2\}$，我们有 $(|\tau|_t - |U_\tau|_t) + (|\sigma|_t - \frac{|V_\sigma|_t}{2}) + (|\sigma'|_t - \frac{|V_{\sigma'}|_t}{2}) = |\sigma \circ \tau \circ \sigma'|_t$。然后，利用定义 6.4.15 对定义 6.4.3 中条件进行直接验证，即可得出结论。

■

我们现在可以证明定性中间形状界。

定性中间形状界引理 *6.4.9* 的证明。令 P_τ, Q_τ, π 来自命题 6.4.19。令 $U' = (U_\tau)_2, V' = (V_\tau)_2$ 分别为 U_τ, V_τ 中的类型 2 顶点。令 x_i 为关于 $i \in U' \cup V'$ 的随机变量，它们从 $\mathcal{D}$ 中独立抽样。定义 $x_{U'}$（分别地，$x_{V'}$）为变量 x_i 关于 $i \in U'$（分别地，$i \in V'$）的子集。

对于 $\rho \in \mathcal{P}_U$，定义 W_ρ，$\rho = \mathbb{E}_{y_{U_2} \sim \mathcal{D}^{U_2}}[v_{\rho, y_{U_2}} v_{\rho, y_{U_2}}^T]$ 使得 H_{Id_U}，ρ，$\rho = \frac{1}{|Aut}(U)|W_\rho$，$\rho$。注意到 W_ρ，$\rho = \mathbb{E}[v_{\rho, x_{U'}} v_{\rho, x_{U'}}^T] = \mathbb{E}[v_{\rho, x_{V'}} v_{\rho, x_{V'}}^T]$，因为 $x_{U'}$ 和 $x_{V'}$ 也是从 $\mathcal{D}$ 中抽取的变量集合，并且 U'、V' 与 U_2 大小相同，因为 $U_\tau = V_\tau = U$。

对于 ρ，$\rho' \in \mathcal{P}_U$，定义 Y_ρ，$\rho' = \mathbb{E}\left[v_{\rho, x_{U'}} \left(\prod_{i \in U' \cup V'} x^{deg^\tau(}{}_i\right) v_\pi^T)(\rho)\, x_{V'}\right]$。那么，$H_\tau = \sum_{\rho \in P_\tau} H_\tau$，$\rho$，$\pi(\rho) = {}_{|Aut}{}^1(U)|^2 S(\tau) \sum_{\rho \in P_\tau} Y_\rho$，$\pi(\rho)$。我们有

$$\begin{bmatrix} \frac{S(\tau)R(\tau)}{|Aut(U)|} H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)R(\tau)}{|Aut(U)|} H_{Id_U} \end{bmatrix} = \frac{S(\tau)}{|Aut(U)|^2} \begin{bmatrix} R(\tau) \sum_{\rho \in \mathcal{P}_U} W_{\rho,\rho} & \sum_{\rho \in P_\tau} Y_{\rho,\pi(\rho)} \\ \sum_{\rho \in P_\tau} Y_{\rho,\pi(\rho)}^T & R(\tau) \sum_{\rho \in \mathcal{P}_U} W_{\rho,\rho} \end{bmatrix}$$

由于 $S(\tau)|Aut(U)|^2 \geq 0$，只需证明 $\begin{bmatrix} R(\tau) \sum_{\rho \in \mathcal{P}_U} W_{\rho,\rho} & \sum_{\rho \in P_\tau} Y_{\rho,\pi(\rho)} \\ \sum_{\rho \in P_\tau} Y_{\rho,\pi(\rho)}^T & R(\tau) \sum_{\rho \in \mathcal{P}_U} W_{\rho,\rho} \end{bmatrix} \succeq 0$。考虑

$$\begin{aligned} \begin{bmatrix} R(\tau) \sum_{\rho \in \mathcal{P}_U} W_{\rho,\rho} & \sum_{\rho \in P_\tau} Y_{\rho,\pi(\rho)} \\ \sum_{\rho \in P_\tau} Y_{\rho,\pi(\rho)}^T & R(\tau) \sum_{\rho \in \mathcal{P}_U} W_{\rho,\rho} \end{bmatrix} =& R(\tau) \begin{bmatrix} \sum_{\rho \in \mathcal{P}_U \setminus P_\tau} W_{\rho,\rho} & 0 \\ 0 & \sum_{\rho \in \mathcal{P}_U \setminus Q_\tau} W_{\rho,\rho} \end{bmatrix} \\ &+ \begin{bmatrix} R(\tau) \sum_{\rho \in P_\tau} W_{\rho,\rho} & \sum_{\rho \in P_\tau} Y_{\rho,\pi(\rho)} \\ \sum_{\rho \in P_\tau} Y_{\rho,\pi(\rho)}^T & R(\tau) \sum_{\rho \in P_\tau} W_{\pi(\rho),\pi(\rho)} \end{bmatrix} \end{aligned}$$

我们有 $\sum_{\rho \in \mathcal{P}_U \setminus P_\tau} W_\rho,\ \rho = \sum_{\rho \in \mathcal{P}_U \setminus P_\tau} \mathbb{E}[v_{\rho, x_{U'}} v_{\rho, x_{U'}}^T] \succeq 0$。类似地，$\sum_{\rho \in \mathcal{P}_U \setminus Q_\tau} W_\rho,\ \rho \succeq 0$。此外，$R(\tau) \geq 0$ 因此，我们有上述表达式中的第一项，

$R(\tau)\begin{bmatrix}\sum_{\rho\in\mathcal{P}_U\setminus P_\tau} W_{\rho,\rho} & 0\\ 0 & \sum_{\rho\in\mathcal{P}_U\setminus Q_\tau} W_{\rho,\rho}\end{bmatrix}$, is positive semidefinite. For the second term,

$$\begin{aligned}
&\begin{bmatrix}R(\tau)\sum_{\rho\in P_\tau} W_{\rho,\rho} & \sum_{\rho\in P_\tau} Y_{\rho,\pi(\rho)}\\ \sum_{\rho\in P_\tau} Y^T_{\rho,\pi(\rho)} & R(\tau)\sum_{\rho\in P_\tau} W_{\pi(\rho),\pi(\rho)}\end{bmatrix}\\
&\quad=\sum_{\rho\in P_\tau}\begin{bmatrix}R(\tau)\,\mathbb{E}[v_{\rho,x_{U'}}v^T_{\rho,x_{U'}}] & \mathbb{E}\left[v_{\rho,x_{U'}}\left(\prod_{i\in U'\cup V'} x_i^{deg^\tau(i)}\right)v^T_{\pi(\rho),x_{V'}}\right]\\ \mathbb{E}\left[v^T_{\rho,x_{U'}}\left(\prod_{i\in U'\cup V'} x_i^{deg^\tau(i)}\right)v_{\pi(\rho),x_{V'}}\right] & R(\tau)\,\mathbb{E}[v_{\pi(\rho),x_{V'}}v^T_{\pi(\rho),x_{V'}}]\end{bmatrix}\\
&\quad=\sum_{\rho\in P_\tau}\mathbb{E}\begin{bmatrix}R(\tau)v_{\rho,x_{U'}}v^T_{\rho,x_{U'}} & v_{\rho,x_{U'}}\left(\prod_{i\in U'\cup V'} x_i^{deg^\tau(i)}\right)v^T_{\pi(\rho),x_{V'}}\\ v^T_{\rho,x_{U'}}\left(\prod_{i\in U'\cup V'} x_i^{deg^\tau(i)}\right)v_{\pi(\rho),x_{V'}} & R(\tau)v_{\pi(\rho),x_{V'}}v^T_{\pi(\rho),x_{V'}}\end{bmatrix}
\end{aligned}$$

我们将证明，对于每个 $\rho\in P_\tau$以及从 $\mathcal{D}$ 中对 x_i 的每次抽样，期望中的项都是半正定的，由此即可完成证明。固定 $\rho\in P_\tau$，并对从 $\mathcal{D}$ 中抽取的 x_i 进行任意一次抽样。令 $w_1=v_{\rho,x_{U'}}$，$w_2=v_{\pi(\rho),x_{V'}}$。令 $E=\prod_{i\in U'\cup V'} x_i^{deg^\tau(i)}$。我们希望证明 $\begin{bmatrix}R(\tau)w_1w_1^T & Ew_1w_2^T\\ Ew_1^Tw_2 & R(\tau)w_2w_2^T\end{bmatrix}\succeq 0$。对于从 $\mathcal{D}$ 中抽取的所有 y，$|y|\le C_{disc}\sqrt{D_E}$ 因此，$|E|\le(C_{disc}\sqrt{D_E})^{\sum_{j\in U'\cup V'} deg^\tau(j)}=R(\tau)$。

如果 $E\ge 0$，那么

$$\begin{aligned}
\begin{bmatrix}R(\tau)w_1w_1^T & Ew_1w_2^T\\ Ew_1^Tw_2 & R(\tau)w_2w_2^T\end{bmatrix}&=(R(\tau)-E)\begin{bmatrix}w_1w_1^T & 0\\ 0 & w_2w_2^T\end{bmatrix}+E\begin{bmatrix}w_1w_1^T & w_1w_2^T\\ w_1^Tw_2 & w_2w_2^T\end{bmatrix}\\
&=(R(\tau)-E)\left(\begin{bmatrix}w_1\\ 0\end{bmatrix}\begin{bmatrix}w_1 & 0\end{bmatrix}+\begin{bmatrix}0\\ w_2\end{bmatrix}\begin{bmatrix}0 & w_2\end{bmatrix}\right)+E\begin{bmatrix}w_1\\ w_2\end{bmatrix}\begin{bmatrix}w_1 & w_2\end{bmatrix}\\
&\succeq 0
\end{aligned}$$

由于 $R(\tau) - E \geq 0$ 而且如果 $E < 0$,

$$\begin{aligned}
\begin{bmatrix} R(\tau)w_1w_1^T & Ew_1w_2^T \\ Ew_1^Tw_2 & R(\tau)w_2w_2^T \end{bmatrix} &= (R(\tau)+E)\begin{bmatrix} w_1w_1^T & 0 \\ 0 & w_2w_2^T \end{bmatrix} - E\begin{bmatrix} w_1w_1^T & -w_1w_2^T \\ -w_1^Tw_2 & w_2w_2^T \end{bmatrix} \\
&= (R(\tau)+E)\left(\begin{bmatrix} w_1 \\ 0 \end{bmatrix}\begin{bmatrix} w_1 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ w_2 \end{bmatrix}\begin{bmatrix} 0 & w_2 \end{bmatrix}\right) - E\begin{bmatrix} w_1 \\ -w_2 \end{bmatrix}\begin{bmatrix} w_1 & -w_2 \end{bmatrix} \\
&\succeq 0
\end{aligned}$$

因为 $R(\tau) + E \geq 0$。 ■

定性交集项界

与命题 6.4.19 一样，下一个命题说明，当我们复合形状 σ、γ、γ^T、σ'^T 时，为了使 $\lambda_{\sigma\circ\gamma\circ\gamma^T\circ\sigma'^T}$ 非零，合并顶点的度的奇偶性应相应地相加。与张量主成分分析应用中的做法一样，我们同样定义 $H^{-\gamma}$、Id_V、ρ、ρ' 和 H'_γ、ρ、ρ。以下命题很简单，证明方法也相同。

命题 6.4.21。对于所有满足 $w(U) > w(V)$ 的 U、$V \in \mathcal{I}_{mid}$，对于所有 $\gamma \in \Gamma_{U,V}$，存在一个奇偶向量集合 $P_\gamma \subseteq \mathcal{P}_U$，使得 $H^{-\gamma}$、$Id_V = \sum_{\rho\in P_\gamma} H^{-\gamma}$、$Id_V$、$\rho$、$\rho$。

命题 6.4.22。对于所有 $U, V \in \mathcal{I}_{mid}$，其中 $w(U) > w(V)$，对于 $\gamma \in \Gamma_{U,V}$，$H'_\gamma = \sum_{\rho\in P_\gamma} H'_{\gamma,\rho,\rho}$。

现在我们将定义一些向量，它们是 $v_{\rho,x_{U_2}}$ 的截断。这个定义及其后续命题主要是技术性的，与半正定质量条件分析基本相似。

定义 6.4.23. 令 $U, V \in \mathcal{I}_{mid}$，其中 $w(U) > w(V)$，并令 $\gamma \in \Gamma_{U,V}$。令 x_i 对于 $i \in U_2$ 为变量。将它们统称为 x_{U_2}。对于 $\rho \in \mathcal{P}_U$，定义 $v_\rho^{-\gamma,x_{U_2}}$ 为一个以左形状 $\sigma \in \mathcal{L}$ 为索引的向量，使得第 σ 项为 $v_{\rho,x_{U_2}(\sigma)}$，若 $|V(\sigma\circ\gamma)| \leq D_V$，否则为0。

据此，我们可以将每个切片 $H^{-\gamma,\gamma}_{Id_V,\rho,,\rho}$ 分解。

命题 *6.4.24.* 对任意 $U,V \in \mathcal{I}_{mid}$，其中 $w(U) > w(V)$，以及任意 $\gamma \in \Gamma_{U,V}$，假设我们取 $\rho \in P_\gamma$。将 γ 与 γ^T 复合得到 $\gamma \circ \gamma^T$ 时，令 $U' = (U_{\gamma\circ\gamma^T})2, V' = (V_{\gamma\circ\gamma^T})2$ 分别为 $U_{\gamma\circ\gamma^T}, V_{\gamma\circ\gamma^T}$ 中的类型 *2* 顶点。再令 W' 为 $\gamma \circ \gamma^T$ 中在设置 $V_\gamma = U_\gamma^T$ 时于复合过程中被识别的类型 *2* 顶点集合。令 x_i（$i \in U' \cup W' \cup V'$ 的）为从 $\mathcal{D}$ 中独立抽样得到的随机变量。定义 $x_{U'}$（分别为 $x_{V'}, x_{W'}$）为 x_i 中对应于 $i \in U'$（分别为 $i \in V', i \in W'$）的变量子集。那么，

$$H^{-\gamma,\gamma}_{Id_V,\rho,\rho} = \frac{1}{|Aut(V)|} S(\gamma)^2 \mathop{\mathbb{E}}_{x} \left[(v^{-\gamma}_{\rho,x_{U'}}) \left(\prod_{i \in U' \cup W' \cup V'} x_i^{deg^{\gamma\circ\gamma^T}(i)} \right) (v^{-\gamma}_{\rho,x_{V'}})^T \right]$$

证明。固定 σ，$\sigma' \in \mathcal{L}_U$，$\rho$ 使得 $|V(\sigma \circ \gamma)|$，$|V(\sigma' \circ \gamma)| \leq D_V$。注意，对于 $t \in \{1,2\}$，$|\sigma|_t - \frac{|V_\sigma|_t}{2} + |\sigma'|_t - \frac{|V_{\sigma'}|_t}{2} + 2(|\gamma|_t - \frac{|U_\gamma|_t + |V_\gamma|_t}{2}) = |\sigma \circ \gamma \circ \gamma^T \circ \sigma'^T|_t$。利用定义 6.4.3 和定义 6.4.15，我们可以轻松验证该等式。

■

命题 6.4.25。对于任意 $U,V \in \mathcal{I}_{mid}$，其中 $w(U) > w(V)$，以及任意 $\gamma \in \Gamma_{U,V}$，假设我们取 $\rho \in \mathcal{P}_U$。那么，

$$H'_{\gamma,\rho,\rho} = \frac{1}{|Aut(U)|} \mathop{\mathbb{E}}_{y_{U_2} \sim \mathcal{D}^{U_2}} \left[(v^{-\gamma}_{\rho,y_{U_2}})(v^{-\gamma}_{\rho,y_{U_2}})^T \right]$$

最后，我们可以证明定性交集项界。

定性交集项界引理 *6.4.10* 的证明。设 U', V', W' 如命题 6.4.24 中所定义。我们有

$$\begin{aligned}
\frac{|Aut(V)|}{|Aut(U)|} \cdot \frac{1}{S(\gamma)^2 R(\gamma)^2} H^{-\gamma,\gamma}_{Id_V} &= \sum_{\rho \in P_\gamma} \frac{|Aut(V)|}{|Aut(U)|} \cdot \frac{1}{S(\gamma)^2 R(\gamma)^2} H^{-\gamma,\gamma}_{Id_V,\rho,\rho} \\
&= \sum_{\rho \in P_\gamma} \frac{1}{|Aut(U)|} \cdot \frac{1}{R(\gamma)^2} \mathop{\mathbb{E}}_{x} \left[(v^{-\gamma}_{\rho,x_{U'}}) \left(\prod_{i \in U' \cup W' \cup V'} x_i^{deg^{\gamma\circ\gamma^T}(i)} \right) (v^{-\gamma}_{\rho,x_{V'}})^T \right]
\end{aligned}$$

现在我们将证明，对于所有的 $\rho \in P_\gamma$,

$$\frac{1}{|Aut(U)|} \cdot \frac{1}{R(\gamma)^2} \mathop{\mathbb{E}}_{x} \left[(v_{\rho,x_{U'}}^{-\gamma}) \left(\prod_{i \in U' \cup W' \cup V'} x_i^{deg^{\gamma \circ \gamma^T}(i)} \right) (v_{\rho,x_{V'}}^{-\gamma})^T \right] \preceq H'_{\gamma,\rho,\rho}$$

这归结为证明

$$\begin{aligned} \frac{2}{R(\gamma)^2} \mathop{\mathbb{E}}_{x} \left[(v_{\rho,x_{U'}}^{-\gamma}) \left(\prod_{i \in U' \cup W' \cup V'} x_i^{deg^{\gamma \circ \gamma^T}(i)} \right) (v_{\rho,x_{V'}}^{-\gamma})^T \right] &\preceq 2 \mathop{\mathbb{E}}_{y_{U_2} \sim \mathcal{D}^{U_2}} \left[(v_{\rho,y_{U_2}}^{-\gamma})(v_{\rho,y_{U_2}}^{-\gamma})^T \right] \\ &= \mathop{\mathbb{E}}_{x} \left[(v_{\rho,x_{U'}}^{-\gamma})(v_{\rho,x_{U'}}^{-\gamma})^T + (v_{\rho,x_{V'}}^{-\gamma})(v_{\rho,x_{V'}}^{-\gamma})^T \right] \end{aligned}$$

其中，最后一个等式由期望的线性性以及 $U' \equiv V' \equiv U_2$ 这一事实得出。由于 $H^{-\gamma}$, $Id_V \overset{\gamma}{\rho},,\rho$ 是对称的，我们有

$$\mathop{\mathbb{E}}_{x} \left[(v_{\rho,x_{U'}}^{-\gamma}) \left(\prod_{i \in U' \cup W' \cup V'} x_i^{deg^{\gamma \circ \gamma^T}(i)} \right) (v_{\rho,x_{V'}}^{-\gamma})^T \right] = \mathop{\mathbb{E}}_{x} \left[(v_{\rho,x_{V'}}^{-\gamma}) \left(\prod_{i \in U' \cup W' \cup V'} x_i^{deg^{\gamma \circ \gamma^T}(i)} \right) (v_{\rho,x_{U'}}^{-\gamma})^T \right]$$

因此，只需证明

$$\begin{aligned} &\frac{1}{R(\gamma)^2} \mathop{\mathbb{E}}_{x} \left[(v_{\rho,x_{U'}}^{-\gamma}) \left(\prod_{i \in U' \cup W' \cup V'} x_i^{deg^{\gamma \circ \gamma^T}(i)} \right) (v_{\rho,x_{V'}}^{-\gamma})^T + (v_{\rho,x_{V'}}^{-\gamma}) \left(\prod_{i \in U' \cup W' \cup V'} x_i^{deg^{\gamma \circ \gamma^T}(i)} \right) (v_{\rho,x_{U'}}^{-\gamma})^T \right] \\ &\quad \preceq \mathop{\mathbb{E}}_{x} \left[(v_{\rho,x_{U'}}^{-\gamma})(v_{\rho,x_{U'}}^{-\gamma})^T + (v_{\rho,x_{V'}}^{-\gamma})(v_{\rho,x_{V'}}^{-\gamma})^T \right] \end{aligned}$$

我们将证明，对于从 $\mathcal{D}$ 中对 x_i 进行的每一种抽样，都有

$$\begin{aligned} &\frac{1}{R(\gamma)^2} \left((v_{\rho,x_{U'}}^{-\gamma}) \left(\prod_{i \in U' \cup W' \cup V'} x_i^{deg^{\gamma \circ \gamma^T}(i)} \right) (v_{\rho,x_{V'}}^{-\gamma})^T + (v_{\rho,x_{V'}}^{-\gamma}) \left(\prod_{i \in U' \cup W' \cup V'} x_i^{deg^{\gamma \circ \gamma^T}(i)} \right) (v_{\rho,x_{U'}}^{-\gamma})^T \right) \\ &\quad \preceq (v_{\rho,x_{U'}}^{-\gamma})(v_{\rho,x_{U'}}^{-\gamma})^T + (v_{\rho,x_{V'}}^{-\gamma})(v_{\rho,x_{V'}}^{-\gamma})^T \end{aligned}$$

然后，取期望即可得到结论。具体来说，固定从 $\mathcal{D}$中对 x_i 进行的一次抽样。令 $E = \prod_{i \in U' \cup W' \cup V'} x_i^{deg^{\gamma \circ \gamma^T}(i)}$，并令 $w_1 = v_\rho^{-\gamma,\ x_{U'}}$，$w_2 = v_\rho^{-\gamma,\ x_{V'}}$。那么，我们需要证明的不等式

需要显示的是

$$\frac{E}{R(\gamma)^2}(w_1 w_2^T + w_2 w_1^T) \preceq w_1 w_1^T + w_2 w_2^T$$

现在，由于对所有 i 都有 $|x_i| \leq C_{disc}\sqrt{D_E}$，我们得到 $|E| \leq \prod_{i \in U' \cup W' \cup V'} (C_{disc}\sqrt{D_E})^{deg^{\gamma \circ \gamma^T}(i) - R(\gamma)^2}$。若 $E \geq 0$，利用 $R(\gamma)^2(w_1 - w_2)(w_1 - w_2)^T \succeq 0$ 可得

$$\frac{E}{R(\gamma)^2}(w_1 w_2^T + w_2 w_1^T) \preceq \frac{E}{R(\gamma)^2}(w_1 w_1^T + w_2 w_2^T) \preceq w_1 w_1^T + w_2 w_2^T$$

因为 $0 \leq E \leq R(\gamma)^2$。并且若 $E < 0$，利用 $R(\gamma)^2(w_1 + w_2)(w_1 + w_2)^T \succeq 0$ 可得

$$\frac{E}{R(\gamma)^2}(w_1 w_2^T + w_2 w_1^T) \preceq \frac{-E}{R(\gamma)^2}(w_1 w_1^T + w_2 w_2^T) \preceq w_1 w_1^T + w_2 w_2^T$$

因为 $0 \leq -E \leq R(\gamma)^2$。最后，我们利用这样一个事实：对所有 $\rho \in \mathcal{P}_U$，都有 $H'_{\gamma,\rho,\rho} \succeq 0$，这可以用与引理 6.4.5 的证明相同的方法证明。因此，

$$\frac{|Aut(V)|}{|Aut(U)|} \cdot \frac{1}{S(\gamma)^2 R(\gamma)^2} H^{-\gamma,\gamma}_{Id_V} \preceq \sum_{\rho \in P_\gamma} H'_{\gamma,\rho,\rho} \preceq \sum_{\rho \in \mathcal{P}_U} H'_{\gamma,\rho,\rho} = H'_\gamma$$

■

6.4.3 定量界的直观理解

本节将介绍主定理定理 4.2.3 所需界的一些直观理解，该定理将在第 7.3 节中得到正式证明。非正式地说，该定理指出，当 $m \leq \frac{d}{\lambda^2}$ 和 $m \leq \frac{k^2}{\lambda^2}$ 时， $\Lambda \succeq 0$ 以高概率成立。

我们将尝试理解为什么不等式 $\lambda^2_{\sigma \circ \tau \circ \sigma'^T} \|M_\tau\|^2 \leq \lambda_{\sigma \circ \sigma^T} \lambda_{\sigma' \circ \sigma'^T}$成立。为简化起见，假设 $d < n$，并考察图 6.4 中的形状。本例中采用假设 $d < n$，因为否则，如果 $d > n$，分解就会与图中所示不同。

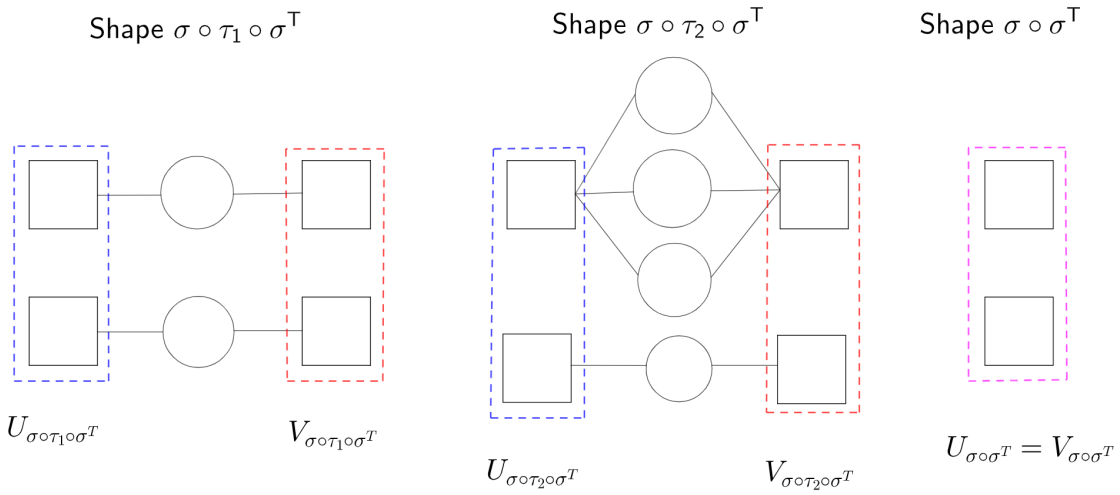


图6.4：形状 $\sigma \circ \tau_1 \circ \sigma^T$，$\sigma \circ \tau_2 \circ \sigma^T$和 $\sigma \circ \sigma^T$。所有边的标签均为 1。

首先，形状 $\sigma \circ \sigma^T$的系数为 $\lambda_{\sigma\circ\sigma^T} \approx \binom{1}{\sqrt{k}}^4 \binom{k}{d}^2$。第一个形状$\sigma \circ \tau_1 \circ \sigma^T$的系数为 $\lambda_{\sigma\circ\tau_1\circ\sigma^T} \approx \binom{1}{\sqrt{k}}^4 \binom{k}{d}^4 \binom{\sqrt{\lambda}}{\sqrt{k}}^4$，并且以高概率，忽略低阶项后，$\|M_{\tau_1}\| \le md$。因此，不等式 $\lambda^2_{\sigma\circ\tau_1\circ\sigma^T} \|M_{\tau_1}\|^2 \le \lambda_{\sigma\circ\sigma^T}\lambda_{\sigma\circ\sigma^T}$可重排为 $m \le \frac{d}{\lambda^2}$。但这恰好是关于 m的假设之一。此外，这也证实了要使我们的策略奏效，就需要关于 m 的这一假设。

第二种形状 $\sigma \circ \tau_2 \circ \sigma^T$的系数为 $\lambda_{\sigma\circ\tau_2\circ\sigma^T} \approx \binom{1}{\sqrt{k}}^4 \binom{k}{d}^4 \binom{\sqrt{\lambda}}{\sqrt{k}}^8$，且以高概率，忽略低阶项后，$\|M_{\tau_2}\| \le m^2 d$。因此，不等式$\lambda^2_{\sigma\circ\tau_2\circ\sigma^T} \|M_{\tau_2}\|^2 \le \lambda_{\sigma\circ\sigma^T}\lambda_{\sigma\circ\sigma^T}$ 可整理为 $m^2 \le \frac{k^2 d}{\lambda^4}$。但这只需将我们关于 m的假设相乘即可，即 $m \le \frac{k^2}{\lambda^2}$ 和 $m \le \frac{d}{\lambda^2}$。

此外，考虑一种形状 $\sigma \circ \tau_3 \circ \sigma^T$，其中 τ_3 与 τ_2 相似，只是它有 t 个（而不是 3 个）不同的圆形顶点，这些顶点是最上方两个方形顶点的公共邻居。分析所需的不等式可得，要使我们的策略成立，m 必须满足 $m \le \frac{k^2}{\lambda^2} \cdot \binom{d}{k^2}^{t+1}$。通过令 t任意增大，我们可以看出，条件$m \le \frac{k^2}{\lambda^2}$ 是必要的。

因此，要使我们的分析成立，假设 $m \le \frac{d}{\lambda^2}$ 和 $m \le \frac{k^2}{\lambda^2}$ 是必要的。我们将证明，事实上，这些假设也是充分的。为此，我们采用一种收费

论证，利用形状 α 在 Λ 的分解中所呈现的特殊结构，以及它们的系数 λ_α；这些正如定义 6.4.3 所述。详情请参见第 7.3 节。

第7章

定量界

本章将以第6章中的定性界为基础，证明平方和下界的主定理 4.4.1、定理 4.3.3 和定理 4.2.3。本章内容改编自 [191], ，不过我们修正了若干排版错误，并改进了技术性论述。

7.1 植入的稍密子图：完整验证

本节将证明关于植入的稍密子图的主定理，即定理 4.4.1。

定理 4.4.1。 设 $C_p > 0$。存在一个常数 $C > 0$ ，使得对于所有充分小的常数 $\varepsilon > 0$, 如果 $k \leq n^{\frac{1}{2}-\varepsilon}$ 且 $p = \frac{1}{2} + \frac{n^{-C_p\varepsilon}}{2}$ ，那么以高概率，由度为 $n^{C\varepsilon}$ 的平方和的伪校准给出的候选解是可行的。

我们将应用这些技术工具。这里，我们选择定理中的 ε ，使其为任意小的常数；注意不要将它与定理 4.4.1 中的 ε 混淆。我们以第6.2节中的定性界为基础（并使用相同的记号）。验证主要条件并应用这些技术工具后，便可得到结果。

7.1.1 中间形状界

引理 7.1.1。 设 $k \leq n^{1/2-\varepsilon}$。对于所有的 $U \in \mathcal{I}_{mid}$ 和 $\tau \in \mathcal{M}_U$,

$$\sqrt{n}^{|V(\tau)|-|U_\tau|} S(\tau) \leq \frac{1}{n^{C_p\varepsilon|E(\tau)|}}$$

证明。 将 $S(\tau)$ 的值代入即可得到这一结果。利用 $k \le n^{1/2-\varepsilon}$,

$$\sqrt{n}^{|V(\tau)|-|U_\tau|} S(\tau) = \sqrt{n}^{|V(\tau)|-|U_\tau|} \left(\frac{k}{n}\right)^{|V(\tau)|-|U_\tau|} (2(\frac{1}{2} + \frac{1}{2n^{C_p\varepsilon}}) - 1)^{|E(\tau)|} \le \frac{1}{n^{C_p\varepsilon|E(\tau)|}}$$

■

推论 7.1.2. 对于所有 $U \in \mathcal{I}_{mid}$ 和 $\tau \in \mathcal{M}_U$，我们有

$$c(\tau) B_{norm}(\tau) S(\tau) \le 1$$

证明. 由于 τ 是适当中间形状，因此我们有 $w(I_\tau) = 0$ 和 $w(S_\tau) = w(U_\tau)$。这意味着 $n^{w(V(\tau))+w(I_\tau)-2w} = \sqrt{n}^{|V(S_\tau)(\tau)|-|U_\tau|}$。由于 τ是正规的，每个顶点 $i \in V(\tau) \setminus U_\tau$ 或 $i \in V(\tau) \setminus V_\tau$都有 $deg^\tau(i) \ge 1$，因此 $|V(\tau) \setminus U_\tau| + |V(\tau) \setminus V_\tau| \le 4|E(\tau)|$。此外，$q = n^{O(1)\cdot\varepsilon C_V}$。我们可以将 C_V 设得足够小，从而利用引理 7.1.1,

$$\begin{aligned}
&c(\tau) B_{norm}(\tau) S(\tau) \\
&\quad = 100(3D_V)^{|U_\tau\setminus V_\tau|+|V_\tau\setminus U_\tau|+2|E(\tau)|} 2^{|V(\tau)\setminus(U_\tau\cup V_\tau)|} \cdot (6D_V \sqrt[4]{2eq})^{|V(\tau)\setminus U_\tau|+|V(\tau)\setminus V_\tau|} \sqrt{n}^{|V(\tau)|-|U_\tau|} S(\tau) \\
&\quad \le n^{O(1)\cdot\varepsilon C_V\cdot|E(\tau)|} \cdot \sqrt{n}^{|V(\tau)|-|U_\tau|} S(\tau) \\
&\quad \le n^{O(1)\cdot\varepsilon C_V\cdot|E(\tau)|} \cdot \frac{1}{n^{C_p\varepsilon|E(\tau)|}} \\
&\quad \le 1
\end{aligned}$$

■

现在我们可以得到中间形状界。

引理 7.1.3. 对于所有 $U \in \mathcal{I}_{mid}$ 和 $\tau \in \mathcal{M}_U$,

$$\begin{bmatrix} \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} & B_{norm}(\tau) H_\tau \\ B_{norm}(\tau) H_\tau^T & \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} \end{bmatrix} \succeq 0$$

证明. 我们有

$$\begin{bmatrix} \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} & B_{norm}(\tau) H_\tau \\ B_{norm}(\tau) H_\tau^T & \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} \end{bmatrix}$$

$$= \begin{bmatrix} \left(\frac{1}{|Aut(U)|c(\tau)} - \frac{S(\tau) B_{norm}(\tau)}{|Aut(U)|} \right) H_{Id_U} & 0 \\ 0 & \left(\frac{1}{|Aut(U)|c(\tau)} - \frac{S(\tau) B_{norm}(\tau)}{|Aut(U)|} \right) H_{Id_U} \end{bmatrix}$$

$$+ B_{norm}(\tau) \begin{bmatrix} \frac{S(\tau)}{|Aut(U)|} H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)}{|Aut(U)|} H_{Id_U} \end{bmatrix}$$

由引理 6.2.7, $\begin{bmatrix} S & \\ (\tau)|Aut(U & |^{H_{Id_U}} \\ \hline & H_\tau^T \end{bmatrix} \begin{bmatrix} & H_\tau \\ )()|Aut(U)|^{H_{Id_U}} & \\ S\ \tau & \end{bmatrix} \succeq 0$，因此上面的第二项是半正定的。对于第一项，由引理 6.2.5，$H_{Id_U} \succeq 0$ 并由推论 7.1.2，$|Aut^{1}(U)|c(\tau)$ $_ \frac{S(\tau) B_{norm}(\tau)}{|Aut(U)|} \geq 0$，这证明第一项也是半正定的。

■

7.1.2 交项界

引理 7.1.4。 假设 $k \leq n^{1/2-\varepsilon}$。对于所有 $U, V \in \mathcal{I}_{mid}$，其中 $w(U) > w(V)$ 以及对于所有$\gamma \in \Gamma_{U,V}$,

$$n^{w(V(\gamma) \setminus U_\gamma)} S(\gamma)^2 \leq \frac{1}{n^{B\varepsilon(|V(\gamma) \setminus (U_\gamma \cap V_\gamma)| + |E(\gamma)|)}}$$

其中某个常数 B 仅取决于 C_p。特别地，它与 C_V 无关。

证明。 由于 γ 是左形状，我们有 $|U_\gamma| \geq |V_\gamma|$，因为 V_γ 是 γ 的唯一最小顶点分隔集，因此 $n^{w(V(\gamma) \setminus U_\gamma)} = n^{|V(\gamma)| - |U_\gamma|} \leq n^{|V(\gamma)| - \frac{|U_\gamma| + |V_\gamma|}{2}}$。此外，注意 $2|V(\gamma)| - |U_\gamma| - |V_\gamma| = |U_\gamma \setminus V_\gamma| + |V_\gamma \setminus U_\gamma| + 2|V(\gamma) \setminus U_\gamma \setminus V_\gamma| \geq |V(\gamma) \setminus (U_\gamma \cap V_\gamma)|$。

因此，

$$\begin{aligned} n^{w(V(\gamma)\setminus U_\gamma)}S(\gamma)^2 &= n^{|V(\gamma)\setminus U_\gamma)|}\left(\frac{k}{n}\right)^{2|V(\gamma)|-|U_\gamma|-|V_\gamma|}(2(\frac{1}{2}+\frac{1}{2n^{C_p\varepsilon}})-1)^{2|E(\gamma)|} \\ &\leq n^{|V(\gamma)|-\frac{|U_\gamma|+|V_\gamma|}{2}}\left(\frac{1}{n^{1/2+\varepsilon}}\right)^{2|V(\gamma)|-|U_\gamma|-|V_\gamma|}\left(\frac{1}{n^{2C_p\varepsilon}}\right)^{|E(\gamma)|} \\ &\leq \left(\frac{1}{n^{\varepsilon}}\right)^{2|V(\gamma)|-|U_\gamma|-|V_\gamma|}\left(\frac{1}{n^{2C_p\varepsilon}}\right)^{|E(\gamma)|} \\ &\leq \frac{1}{n^{B\varepsilon(|V(\gamma)\setminus(U_\gamma\cap V_\gamma)|+\sum_{e\in E(\gamma)}l_e)}} \end{aligned}$$

对于一个仅依赖于 C_p 的常数 B。 ■

我们得到交项界。

引理 7.1.5. 对于所有 U、$V\in\mathcal{I}_{mid}$，其中 $w(U)>w(V)$，以及所有 $\gamma\in\Gamma_{U,V}$，

$$c(\gamma)^2N(\gamma)^2B(\gamma)^2H_{Id_V}^{-\gamma,\gamma}\preceq H'_\gamma$$

证明. 由引理 6.2.8，我们有

$$c(\gamma)^2N(\gamma)^2B(\gamma)^2H_{Id_V}^{-\gamma,\gamma}=c(\gamma)^2N(\gamma)^2B(\gamma)^2S(\gamma)^2\frac{|Aut(U)|}{|Aut(V)|}H'_\gamma$$

使用与引理 6.2.5 相同的证明，我们可以看出 $H'_\gamma\succeq 0$。因此，只需证明 $c(\gamma)^2N(\gamma)^2B(\gamma)^2S(\gamma)^2\frac{|Aut(U)|}{|Aut(V)|}\leq 1$。由于 $U,V\in\mathcal{I}_{mid}$，$|Aut(U)|=|U|!$ $|Aut(V)|=|V|!$。因此，$\frac{|Aut(U)|}{|Aut(V)|}=\frac{|U|!}{|V|!}\leq D_V^{|U_\gamma\setminus V_\gamma|}$。另外，$q=n^{O(1)\cdot\varepsilon C_V}$。令 B为引理 7.1.4 中的常数。我们可以将 C_V设得足够小，从而利用引理 7.1.4，

$$
\begin{aligned}
c(\gamma)^2N(\gamma)^2B(\gamma)^2S(\gamma)^2\frac{|Aut(U)|}{|Aut(V)|} &\le 100^2(3D_V)^{2|U_\gamma\setminus V_\gamma|+2|V_\gamma\setminus U_\gamma|+4|E(\alpha)|}4^{|V(\gamma)\setminus(U_\gamma\cup V_\gamma)|}\\
&\quad\cdot(3D_V)^{4|V(\gamma)\setminus V_\gamma|+2|V(\gamma)\setminus U_\gamma|}(6D_V\sqrt[4]{2eq})^{2|V(\gamma)\setminus U_\gamma|+2|V(\gamma)\setminus V_\gamma|}\\
&\quad\cdot n^{w(V(\gamma)\setminus U_\gamma)}S(\gamma)^2\cdot D_V^{|U_\gamma\setminus V_\gamma|}\\
&\le n^{O(1)\cdot\varepsilon C_V\cdot(|V(\gamma)\setminus(U_\gamma\cap V_\gamma)|+\sum_{e\in E(\gamma)}l_e)}\cdot n^{w(V(\gamma)\setminus U_\gamma)}S(\gamma)^2\\
&\le n^{O(1)\cdot\varepsilon C_V\cdot(|V(\gamma)\setminus(U_\gamma\cap V_\gamma)|+\sum_{e\in E(\gamma)}l_e)}\cdot\frac{1}{n^{B\varepsilon(|V(\gamma)\setminus(U_\gamma\cap V_\gamma)|+\sum_{e\in E(\gamma)}l_e)}}\\
&\le 1
\end{aligned}
$$

■

7.1.3 截断误差界

本节将证明截断误差界。我们采用 [191, 节 10]中的策略和记号。首先，我们需要下面得到的关于 $B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma')$ 的一个界。

引理 7.1.6。 假设 $k\le n^{1/2-\varepsilon}$。对于所有 $U\in\mathcal{I}_{mid}$ 和 $\sigma,\ \sigma'\in\mathcal{L}_U$，

$$
B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma')\le\frac{1}{n^{0.5\varepsilon|V(\alpha)|+C_p\varepsilon|E(\alpha)|}}\left(\frac{k}{n}\right)^{|U|}
$$

证明。 设 $\alpha=\sigma\circ\sigma'$。注意到 $|V(\sigma)|+|V(\sigma')|=|V(\alpha)|+|U|$。通过选取足够小的 C_V，

$$
\begin{aligned}
B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma') &= (6D_V\sqrt[4]{2eq})^{|V(\sigma)\setminus U_\sigma|+|V(\sigma)\setminus V_\sigma|}n^{\frac{w(V(\sigma))-w(U)}{2}}\\
&\quad\cdot(6D_V\sqrt[4]{2eq})^{|V(\sigma')\setminus U_{\sigma'}|+|V(\sigma')\setminus V_{\sigma'}|}n^{\frac{w(V(\sigma'))-w(U)}{2}}\\
&\quad\cdot\frac{1}{|Aut(U)|}\left(\frac{k}{n}\right)^{|V(\alpha)|}(2(\frac{1}{2}+\frac{1}{2n^{C_p\varepsilon}})-1)^{|E(\alpha)|}\\
&\le n^{O(1)\cdot\varepsilon C_V\cdot|V(\alpha)|}\sqrt{n}^{|V(\sigma)|-|U|}\sqrt{n}^{|V(\sigma')|-|U|}\left(\frac{k}{n}\right)^{|V(\alpha)|}\frac{1}{n^{C_p\varepsilon|E(\alpha)|}}\\
&\le\frac{1}{n^{0.5\varepsilon|V(\alpha)|+C_p\varepsilon|E(\alpha)|}}\left(\frac{k}{n}\right)^{|U|}
\end{aligned}
$$

■

现在，我们可以应用这一策略了。

引理 7.1.7。 对于所有 $\alpha \in \mathcal{M}'$，当 $\|M_\alpha\| \leq B_{norm}(\alpha)$ 时，

$$\sum_{U \in \mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) \succeq \frac{1}{n^{K_1 D_{sos}^2}} Id_{sym}$$

其中 $K_1 > 0$ 为常数。

证明。 对于 $V \in \mathcal{I}_{mid}$，我们有 $\lambda_V = \binom{k}{n}^{|V|}$。现在，我们选取 $w_V = \binom{k}{n}^{D_{sos}-|V|}$。于是，对于所有 $\sigma \in \mathcal{L}_V$，我们有 $w_V \leq \frac{w_{U_\sigma}\lambda_{U_\sigma}}{|\mathcal{I}_{mid}|B_{norm}{}^{\sigma}{}^{2}c\ \sigma}()()^2 H_{Id_V}(\sigma,\sigma)$，利用引理 7.1.6 可轻易验证这一点。结果遂得证。

■

引理 7.1.8。

$$\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} \leq \frac{n^{K_2 D_{sos}}}{2^{D_V}}$$

其中 $K_2 > 0$ 为常数。

证明。我们有

$$\begin{aligned}
&\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} \\
&= \sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{1}{|Aut(U)|c(\gamma)} \sum_{\substack{\sigma,\sigma' \in \mathcal{L}_{U_\gamma} : |V(\sigma)| \leq D_V, |V(\sigma')| \leq D_V, \\ |V(\sigma \circ \gamma)| > D_V \text{ or } |V(\sigma' \circ \gamma)| > D_V}} B_{norm}(\sigma) B_{norm}(\sigma') H_{Id_{U_\gamma}}(\sigma, \sigma')
\end{aligned}$$

上述求和中可能出现的 σ, σ' 集合必定是非平凡的，因此， $\sigma, \sigma' \in \mathcal{L}'_U$。接下来，

$$\begin{aligned}
&\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} \\
&= \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma,\sigma' \in \mathcal{L}'_U} B_{norm}(\sigma) B_{norm}(\sigma') H_{Id_U}(\sigma, \sigma') \sum_{\gamma \in \Gamma_{U,*} : |V(\sigma \circ \gamma)| > D_V \text{ or } |V(\sigma' \circ \gamma)| > D_V} \frac{1}{|Aut(U)|c(\gamma)}
\end{aligned}$$

对于 $\sigma \in \mathcal{L}'_U$，定义 $m_\sigma = D_V + 1 - |V(\sigma)| \geq 1$。该集合的定义恰好保证，对于所有 $\gamma \in \Gamma_{U,*}$，当且仅当 $|V(\gamma)| \geq |U| + m_\sigma$ 时，我们有 $|V(\sigma \circ \gamma)| > D_V$。因此，对于 $\sigma, \sigma' \in \mathcal{L}'_U$，

$$\begin{aligned}
&\sum_{\gamma \in \Gamma_{U,*}: |V(\sigma\circ\gamma)|>D_V \text{ or } |V(\sigma'\circ\gamma)|>D_V} \frac{1}{|Aut(U)|c(\gamma)} \\
&\qquad = \sum_{\gamma \in \Gamma_{U,*}: |V(\gamma)| \geq |U| + \min(m_\sigma, m_{\sigma'})} \frac{1}{|Aut(U)|c(\gamma)} \\
&\qquad \leq \frac{1}{2^{\min(m_\sigma, m_{\sigma'})-1}}
\end{aligned}$$

此外，对于 σ，$\sigma' \in \mathcal{L}'_U$，我们有 $|V(\sigma \circ \sigma')| + min(m_\sigma, m_{\sigma'}) - 1 \geq D_V$。因此 e,

$$\begin{aligned}
&\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} \\
&\qquad \leq \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} B_{norm}(\sigma) B_{norm}(\sigma') H_{Id_U}(\sigma, \sigma') \frac{1}{2^{\min(m_\sigma, m_{\sigma'})-1}} \\
&\qquad \leq \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{n^{O(1)D_{sos}}}{n^{0.5\varepsilon|V(\sigma\circ\sigma')|} 2^{\min(m_\sigma, m_{\sigma'})-1}}
\end{aligned}$$

其中用到了引理 7.1.6。利用 $n^{0.5\varepsilon|V(\sigma\circ\sigma')|} \geq n^{0.1\varepsilon|V(\sigma\circ\sigma')|} 2^{|V(\sigma\circ\sigma')|}$，

$$\begin{aligned}
\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} &\leq \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{n^{O(1)D_{sos}}}{n^{0.1\varepsilon|V(\sigma\circ\sigma')|} 2^{|V(\sigma\circ\sigma')|} 2^{\min(m_\sigma, m_{\sigma'})-1}} \\
&\leq \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{n^{O(1)D_{sos}}}{n^{0.1\varepsilon|V(\sigma\circ\sigma')|} 2^{D_V}} \\
&\leq \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{n^{O(1)D_{sos}}}{D_{sos}^{D_{sos}} n^{0.1\varepsilon|V(\sigma\circ\sigma')|} 2^{D_V}}
\end{aligned}$$

最后一步是证明 $\sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{1}{D_{sos}^{D_{sos}} n^{0.1\varepsilon|V(\sigma\circ\sigma')|}} \leq 1$，这样就完成了证明。但只要我们将 C_V 设得足够小，这一点就能成立。

■

由此我们得出以下结论。

引理 7.1.9。 只要对所有 $\alpha \in \mathcal{M}'$, $\|M_\alpha\| \le B_{norm}(\alpha)$,

$$\sum_{U \in \mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) \succeq 6 \left(\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)|c(\gamma)} \right) Id_{sym}$$

证明。 取 C_{sos} 足够小，使得 $\frac{1}{n^{K_1 D_{sos}^2}} \ge 6 \frac{n^{K_2 D_{sos}}}{2^{D_V}}$；通过令 $C_{sos} < K_3 C_V$ 为足够小的常数 $K_3 > 0$，即可满足这一条件。然后，由于 $Id_{Sym} \succeq 0$，利用引理 7.1.7 和引理 7.1.8,

$$\begin{aligned}
\sum_{U \in \mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) &\succeq \frac{1}{n^{K_1 D_{sos}^2}} Id_{sym} \\
&\succeq 6 \frac{n^{K_2 D_{sos}}}{2^{D_V}} Id_{sym} \\
&\succeq 6 \left(\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)|c(\gamma)} \right) Id_{sym}
\end{aligned}$$

■

7.2 张量主成分分析：完整验证

In 本节将证明定理 4.3.3 所需的所有界 .

定理 4.3.3。 设 $k \ge 2$ 为一个整数。存在常数 C、$C_\Delta > 0$ ，使得对于所有足够小的常数 $\varepsilon > 0$，若 $\lambda \le n^{\frac{k}{4}-\varepsilon}$ 且 $\Delta = n^{-C_\Delta \varepsilon}$, _则以高概率，由伪校准得到的、度为 $n^{C\varepsilon}$ 的平方和解是可行的。

我们沿用第 6.3 节中的记号和定性界。验证条件后，该定理便可直接由这些技术工具推出。

7.2.1 中间形状界

引理 7.2.1。 假设 $\lambda \leq n^{\frac{k}{4}-\varepsilon}$。对于所有 $U \in \mathcal{I}_{mid}$和 $\tau \in \mathcal{M}_U$，假设 $deg^\tau(i)$ 对所有 $i \in V(\tau) \setminus U_\tau \setminus V_\tau$ 都是偶数，则

$$\sqrt{n}^{|V(\tau)|-|U_\tau|} S(\tau) \leq \frac{1}{n^{0.5\varepsilon \sum_{e\in E(\tau)} l_e}}$$

证明。 首先，我们声称 $\sum_{e\in E(\tau)} kl_e \geq 2(|V(\tau)|-|U_\tau|)$。对于任意顶点 $i \in V(\tau) \setminus U_\tau \setminus V_\tau$，$deg^\tau(i)$ 是偶数且不为 0，因此，$deg^\tau(i) \geq 2$。任意顶点 $i \in U_\tau \setminus V_\tau$不能有$deg^\tau(i) = 0$，否则 $U_\tau \setminus \{i\}$ 是一个权重严格小于 U_τ 的顶点分隔集，这是不可能的，因此，$deg^\tau(i) \geq 1$。因此，

$$\begin{aligned}
\sum_{e\in E(\tau)} kl_e = \sum_{i\in V(\tau)} deg^\tau(i) &\geq \sum_{i\in V(\tau)\setminus U_\tau\setminus V_\tau} deg^\tau(i) + \sum_{i\in U_\tau\setminus V_\tau} deg^\tau(i) + \sum_{i\in V_\tau\setminus U_\tau} deg^\tau(i) \\
&\geq 2|V(\tau) \setminus U_\tau \setminus V_\tau| + |U_\tau \setminus V_\tau| + |V_\tau \setminus U_\tau| \\
&= 2(|V(\tau)| - |U_\tau|)
\end{aligned}$$

通过选择足够小的 C_Δ，我们有

$$\begin{aligned}
\sqrt{n}^{|V(\tau)|-|U_\tau|} S(\tau) &= \sqrt{n}^{|V(\tau)|-|U_\tau|} \Delta^{|V(\tau)|-|U_\tau|} \prod_{e\in E(\tau)} \left(\frac{\lambda}{(\Delta n)^{\frac{k}{2}}}\right)^{l_e} \\
&\leq \sqrt{n}^{|V(\tau)|-|U_\tau|} \Delta^{|V(\tau)|-|U_\tau|} \prod_{e\in E(\tau)} n^{(-\frac{k}{4}-0.5\varepsilon)l_e} \\
&= \sqrt{n}^{|V(\tau)|-|U_\tau|-\frac{\sum_{e\in E(\tau)} kl_e}{2}} \Delta^{|V(\tau)|-|U_\tau|} \prod_{e\in E(\tau)} n^{-0.5\varepsilon l_e} \\
&= \Delta^{|V(\tau)|-|U_\tau|} \prod_{e\in E(\tau)} n^{-0.5\varepsilon l_e} \\
&\leq \frac{1}{n^{0.5\varepsilon \sum_{e\in E(\tau)} l_e}}
\end{aligned}$$

■

推论 7.2.2。 对于所有 $U \in \mathcal{I}_{mid}$ 和 $\tau \in \mathcal{M}_U$，我们有

$$c(\tau)B_{norm}(\tau)S(\tau) \leq 1$$

证明。由于 τ 是适当中间形状，我们有 $w(I_\tau) = 0$，且 $w(S_{\tau,min}) = w(U_\tau)$。这意味着 $n^{\frac{(v(\tau))+w(I_\tau)-w(S_{\tau,min})}{2}} = \sqrt{n}^{|V(\tau)|-|U_\tau|}$。如果对于任意顶点 $i \in$, $deg^\tau(i)$ 都是奇数，则 $V(\tau) \setminus U_\tau \setminus V_\tau$，且不等式成立。因此，假设对于所有 $i \in V(\tau) \setminus U_\tau \setminus V_\tau$, $deg^\tau(i)$ 都是偶数。正如引理 7.2.1 的证明中所观察到的，每个顶点 $i \in V(\tau) \setminus U_\tau$ 或 $i \in V(\tau) \setminus V_\tau$ 都有 $deg^\tau(i) \geq 1$，因此 $|V(\tau) \setminus U_\tau| + |V(\tau) \setminus V_\tau| \leq 4\sum_{e \in E(\tau)} l_e$。此外，$|E(\tau)| \leq \sum_{e \in E(\tau)} l_e$，且 $q = n^{O(1)\cdot\varepsilon(C_V+C_E)}$。我们可以将 C_V、C_E 取为充分小，使得利用引理 7.2.1，

$$\begin{aligned}
c(\tau)B_{norm}(\tau)S(\tau) &= 100(3D_V)^{|U_\tau \setminus V_\tau|+|V_\tau \setminus U_\tau|+k|E(\tau)|}2^{|V(\tau)\setminus(U_\tau \cup V_\tau)|} \\
&\quad \cdot 2e(6qD_V)^{|V(\tau)\setminus U_\tau|+|V(\tau)\setminus V_\tau|} \prod_{e \in E(\tau)} (400D_V^2 D_E^2 q)^{l_e} \sqrt{n}^{|V(\tau)|-|U_\tau|} S(\tau) \\
&\leq n^{O(1)\cdot\varepsilon(C_V+C_E)\cdot\sum_{e \in E(\tau)} l_e} \cdot \sqrt{n}^{|V(\tau)|-|U_\tau|} S(\tau) \\
&\leq n^{O(1)\cdot\varepsilon(C_V+C_E)\cdot\sum_{e \in E(\tau)} l_e} \cdot \frac{1}{n^{0.5\varepsilon\sum_{e \in E(\tau)} l_e}} \\
&\leq 1
\end{aligned}$$

■

现在我们可以证明中间形状界。

引理 7.2.3. 对于所有 $U \in \mathcal{I}_{mid}$ 和 $\tau \in \mathcal{M}_U$,

$$\begin{bmatrix} \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} & B_{norm}(\tau)H_\tau \\ B_{norm}(\tau)H_\tau^T & \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} \end{bmatrix} \succeq 0$$

证明。我们有

$$\begin{bmatrix} \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} & B_{norm}(\tau) H_\tau \\ B_{norm}(\tau) H_\tau^T & \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} \end{bmatrix}$$

$$= \begin{bmatrix} \left(\frac{1}{|Aut(U)|c(\tau)} - \frac{S(\tau) B_{norm}(\tau)}{|Aut(U)|} \right) H_{Id_U} & 0 \\ 0 & \left(\frac{1}{|Aut(U)|c(\tau)} - \frac{S(\tau) B_{norm}(\tau)}{|Aut(U)|} \right) H_{Id_U} \end{bmatrix}$$

$$+ B_{norm}(\tau) \begin{bmatrix} \frac{S(\tau)}{|Aut(U)|} H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)}{|Aut(U)|} H_{Id_U} \end{bmatrix}$$

根据引理 6.3.8, $\begin{bmatrix} S & \\ (\tau)|Aut(U & |H_{Id_U} \\ \hline & H_\tau^T \end{bmatrix}$ $\begin{matrix} H_{\tau)()|Aut(U)|} H_{Id_U} \\ S\ \tau \end{matrix}$ $\succeq 0$，因此上面的第二项是半正定的。对于第一项，根据引理 6.3.6，$H_{Id_U} \succeq 0$ 并且根据推论 7.2.2，$\frac{|Aut^{1}(U)|}{|c(\tau)_\ S(\tau) B_{norm}(\tau)|Aut(U)|} \geq 0$，这证明第一项也是半正定的。

■

7.2.2 交项界

引理 7.2.4。设 $\lambda \leq n^{\frac{k}{4}-\varepsilon}$。对于所有满足 $w(U) > w(V)$ 的 $U, V \in \mathcal{I}_{mid}$，以及所有 $\gamma \in \Gamma_{U,V}$，

$$n^{w(V(\gamma) \setminus U_\gamma)} S(\gamma)^2 \leq \frac{1}{n^{B\varepsilon(|V(\gamma) \setminus (U_\gamma \cap V_\gamma)| + \sum_{e \in E(\gamma)} l_e)}}$$

其中存在一个常数 B，它仅依赖于 C_Δ。特别地，它与 C_V 和 C_E 无关。

证明。假设存在一个顶点 $i \in V(\gamma) \setminus U_\gamma \setminus V_\gamma$，使得 $deg^\gamma(i)$ 为奇数，那么 $S(\gamma) = 0$，不等式成立。因此，假设对于所有顶点 $i \in V(\gamma) \setminus U_\gamma \setminus V_\gamma$，$deg^\gamma(i)$ 都是偶数。我们首先断言 $k \sum_{e \in E(\gamma)} l_e \geq 2|V(\gamma) \setminus U_\gamma|$。由于 γ 是左形状，$V(\gamma) \setminus U_\gamma$ 中的所有顶点 i 都有 $deg^\gamma(i) \geq 1$。特别地，所有顶点 $i \in V_\gamma \setminus U_\gamma$ 都有 $deg^\gamma(i) \geq 1$。此外，如果 $i \in V(\gamma) \setminus U_\gamma \setminus V_\gamma$，由于 $deg^\gamma(i)$ 是偶数，我们必有 $deg^\gamma(i) \geq 2$。

令 S' 为具有 $deg^{\gamma}(i) \geq 1$ 的顶点 $i \in U_\gamma \setminus V_\gamma$ 的集合。然后注意，$|S'| + |U_\gamma \cap V_\gamma| \geq |V_\gamma| \Longrightarrow |S'| \geq |V_\gamma \setminus U_\gamma|$；否则 $S' \cup (U_\gamma \cap V_\gamma)$ 将是 γ 的一个权重严格小于 V_γ 的顶点分隔集，这是不可能的。然后，

$$\begin{aligned}
\sum_{e \in E(\gamma)} kl_e &= \sum_{i \in V(\gamma)} deg^{\gamma}(i) \\
&\geq \sum_{i \in V(\gamma) \setminus U_\gamma \setminus V_\gamma} deg^{\gamma}(i) + \sum_{i \in U_\gamma \setminus V_\gamma} deg^{\gamma}(i) + \sum_{i \in V_\gamma \setminus U_\gamma} deg^{\gamma}(i) \\
&\geq 2|V(\gamma) \setminus U_\gamma \setminus V_\gamma| + |S'| + |V_\gamma \setminus U_\gamma| \\
&\geq 2|V(\gamma) \setminus U_\gamma \setminus V_\gamma| + 2|V_\gamma \setminus U_\gamma| \\
&= 2|V(\gamma) \setminus U_\gamma|
\end{aligned}$$

最后请注意，$2|V(\gamma)| - |U_\gamma| - |V_\gamma| = |U_\gamma \setminus V_\gamma| + |V_\gamma \setminus U_\gamma| + 2|V(\gamma) \setminus U_\gamma \setminus V_\gamma| \geq |V(\gamma) \setminus (U_\gamma \cap V_\gamma)|$。通过将 C_Δ 选得足够小，我们有

$$\begin{aligned}
n^{w(V(\gamma) \setminus U_\gamma)} S(\gamma)^2 &= n^{|V(\gamma) \setminus U_\gamma|} \Delta^{2|V(\gamma)| - |U_\gamma| - |V_\gamma|} \prod_{e \in E(\gamma)} \left(\frac{\lambda^2}{(\Delta n)^k} \right)^{l_e} \\
&\leq n^{|V(\gamma) \setminus U_\gamma|} \Delta^{2|V(\gamma)| - |U_\gamma| - |V_\gamma|} \prod_{e \in E(\gamma)} n^{-(\frac{k}{2} + \varepsilon) l_e} \\
&\leq \Delta^{2|V(\gamma)| - |U_\gamma| - |V_\gamma|} \prod_{e \in E(\gamma)} n^{-\varepsilon l_e} \\
&\leq \frac{1}{n^{B\varepsilon(|V(\gamma) \setminus (U_\gamma \cap V_\gamma)| + \sum_{e \in E(\gamma)} l_e)}}
\end{aligned}$$

其中 B 是一个仅依赖于 C_Δ 的常数。 ∎

注 7.2.5. 在上述界中，请注意，对于$V(\gamma) \setminus (U_\gamma \cap V_\gamma)$中的每个顶点，都存在 $n^{B\varepsilon}$ 的衰减。在植入分布中引入松弛参数C_Δ 的一个主要技术原因，就是为了引入这种衰减；当前的技术工具需要这种衰减。

现在我们可以得到交项界。

引理 7.2.6。 对于所有 U、$V \in \mathcal{I}_{mid}$，其中 $w(U) > w(V)$，以及所有 $\gamma \in \Gamma_{U,V}$，

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{Id_V}^{-\gamma,\gamma} \preceq H'_\gamma$$

证明。 根据引理 6.3.9，我们有

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{Id_V}^{-\gamma,\gamma} \preceq c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} H'_\gamma$$

采用与引理 6.3.6 相同的证明，我们可以看到 $H'_\gamma \succeq 0$。因此，只需证明 $c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} \leq 1$。由于 $U, V \in \mathcal{I}_{mid}$，$|Aut(U)| = |U|!$ $|Aut(V)| = |V|!$。因此，$\frac{|Aut(U)|}{|Aut(V)|} = \frac{|U|!}{|V|!} \leq D_V^{|U_\gamma \setminus V_\gamma|}$。此外，$|E(\gamma)| \leq \sum_{e \in E(\gamma)} l_e$ 且 $q = n^{O(1)\cdot\varepsilon(C_V+C_E)}$。令 B 为引理 7.2.4 中的常数。我们可以将 C_V, C_E 设得足够小，从而利用引理 7.2.4，

$$\begin{aligned}
&c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} \\
&\quad \leq 100^2 (3D_V)^{2|U_\gamma \setminus V_\gamma| + 2|V_\gamma \setminus U_\gamma| + 2k|E(\alpha)|_4 |V(\gamma) \setminus (U_\gamma \cup V_\gamma)|} \\
&\qquad \cdot (3D_V)^{4|V(\gamma)\setminus V_\gamma| + 2|V(\gamma)\setminus U_\gamma|} (6qD_V)^{2|V(\gamma)\setminus U_\gamma| + 2|V(\gamma)\setminus V_\gamma|} \prod_{e \in E(\gamma)} (400 D_V^2 D_E^2 q)^{2l_e} \\
&\qquad \cdot n^{w(V(\gamma)\setminus U_\gamma)} S(\gamma)^2 \cdot D_V^{|U_\gamma \setminus V_\gamma|} \\
&\quad \leq n^{O(1)\cdot\varepsilon(C_V+C_E)\cdot(|V(\gamma)\setminus(U_\gamma \cap V_\gamma)| + \sum_{e\in E(\gamma)} l_e)} \cdot n^{w(V(\gamma)\setminus U_\gamma)} S(\gamma)^2 \\
&\quad \leq n^{O(1)\cdot\varepsilon(C_V+C_E)\cdot(|V(\gamma)\setminus(U_\gamma \cap V_\gamma)| + \sum_{e\in E(\gamma)} l_e)} \cdot \frac{1}{n^{B\varepsilon(|V(\gamma)\setminus(U_\gamma \cap V_\gamma)| + \sum_{e\in E(\gamma)} l_e)}} \\
&\quad \leq 1
\end{aligned}$$

■

7.2.3 截断误差界

在本节中，我们将采用 [191]第 10 节中概述的策略，得到截断误差界。我们还会沿用该记号。首先，我们需要关于 $B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\ \sigma')$ 的以下界。

引理 7.2.7. 假设 $\lambda = n^{\frac{k}{4}-\varepsilon}$。对于所有 $U \in \mathcal{I}_{mid}$和 $\sigma, \sigma' \in \mathcal{L}_U$，

$$B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma') \leq \frac{1}{n^{0.5\varepsilon C_\Delta |V(\sigma\circ\sigma')|_\Delta D_{sos} n|U|}}$$

证明。 假设存在一个顶点 $i \in V(\sigma) \setminus V_\sigma$ 使得 $deg^\sigma(i) + deg^{U_\sigma}(i)$ 为奇数，那么$H_{Id_U}(\sigma, \sigma') = 0$，且该不等式成立。因此，假设对于所有 $i \in V(\sigma) \setminus V_\sigma$， $deg^\sigma(i) + deg^{U_\sigma}(i)$ 都是偶数。类似地，假设对于所有 $i \in V(\sigma') \setminus V_{\sigma'}$， $deg^{\sigma'}(i) + deg^{U_{\sigma'}}(i)$ 都是偶数。此外，如果 $\rho_\sigma \neq \rho_{\sigma'}$，我们就有 $H_{Id_U}(\sigma, \sigma') = 0$，这样就完成证明了。因此，假设 $\rho_\sigma = \rho_{\sigma'}$。

设 $\alpha = \sigma \circ \sigma'$。我们首先证明 $\sum_{e\in E(\alpha)} kl_e + 2deg(\alpha) \geq 2|V(\alpha)| + 2|U|$。首先，注意所有顶点 $i \in V(\alpha) \setminus (U_\alpha \cup V_\alpha)$ 的 $deg^\alpha(i)$ 都必须是偶数且非零，因此至少为 2。此外，在集合 $U_\alpha \setminus (U_\alpha \cap V_\alpha)$ 和 $V_\alpha \setminus (U_\alpha \cap V_\alpha)$ 中，至少有$|U| - |U_\alpha \cap V_\alpha|$ 个度至少为 1 的顶点，因为 U是一个最小顶点分隔集。还需注意， $deg(\alpha) \geq |U_\alpha| + |V_\alpha|$。这意味着

$$\begin{aligned}
&\sum_{e\in E(\alpha)} kl_e + 2deg(\alpha) \\
&\qquad \geq 2|V(\alpha) \setminus (U_\alpha \cup V_\alpha)| + 2(|U| - |U_\alpha \cap V_\alpha|) + 2(|U_\alpha| + |V_\alpha|) \\
&\qquad = 2(|V(\alpha)| - |U_\alpha \cup V_\alpha|) + 2(|U| - |U_\alpha \cap V_\alpha|) + 2(|U_\alpha \cup V_\alpha| + |U_\alpha \cap V_\alpha|) \\
&\qquad = 2|V(\alpha)| + 2|U|
\end{aligned}$$

这里我们用到了这样一个事实： $U_\alpha \cap V_\alpha \subseteq U$。最后，通过将 C_V, C_E选得足够小，

$$
\begin{aligned}
&B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma') \\
&= 2e(6qD_V)^{|V(\sigma)\setminus U_\sigma|+|V(\sigma)\setminus V_\sigma|}\prod_{e\in E(\sigma)}(400D_V^2D_E^2q)^{l_e}n^{\frac{w(V(\sigma))-w(U)}{2}} \\
&\quad\cdot 2e(6qD_V)^{|V(\sigma')\setminus U_{\sigma'}|+|V(\sigma')\setminus V_{\sigma'}|}\prod_{e\in E(\sigma')}(400D_V^2D_E^2q)^{l_e}n^{\frac{w(V(\sigma'))-w(U)}{2}} \\
&\quad\cdot\frac{1}{|Aut(U)|}\Delta^{|V(\alpha)|}\left(\frac{1}{\sqrt{\Delta n}}\right)^{deg(\alpha)}\prod_{e\in E(\alpha)}\left(\frac{\lambda}{(\Delta n)^{\frac{k}{2}}}\right)^{l_e} \\
&\le n^{O(1)\cdot\varepsilon(C_V+C_E)\cdot(|V(\alpha)|+\sum_{e\in E(\alpha)}l_e)}\Delta^{|V(\alpha)|}\left(\frac{1}{\sqrt{\Delta}}\right)^{deg(\alpha)} \\
&\quad\cdot\sqrt{n}^{|V(\alpha)|-|U|}\left(\frac{1}{\sqrt{n}}\right)^{deg(\alpha)}\prod_{e\in E(\alpha)}n^{(-\frac{k}{4}-0.5\varepsilon)l_e} \\
&\le \frac{n^{O(1)\cdot\varepsilon(C_V+C_E)\cdot(|V(\alpha)|+\sum_{e\in E(\alpha)}l_e)}}{n^{\varepsilon C_\Delta|V(\alpha)|}n^{0.5\varepsilon\sum_{e\in E(\alpha)}l_e}}\cdot\frac{1}{\Delta^{D_{sos}n|U|}}\sqrt{n}^{|V(\alpha)|+|U|-deg(\alpha)-\frac{1}{2}\sum_{e\in E(\alpha)}kl_e} \\
&\le \frac{1}{n^{0.5\varepsilon C_\Delta|V(\alpha)|}\Delta^{D_{sos}n|U|}}
\end{aligned}
$$

其中我们使用了事实 $\Delta \le 1$， $deg(\alpha) \le 2D_{sos}$。 ■

现在我们通过证明以下界来应用这一策略。

引理 7.2.8. 只要对所有 $\alpha \in \mathcal{M}'$ 都有 $\|M_\alpha\| \le B_{norm}(\alpha)$，

$$\sum_{U\in\mathcal{I}_{mid}} M_{Id_U}^{fact}(H_{Id_U}) \succeq \frac{\Delta^{2D_{sos}^2}}{n^{D_{sos}}}Id_{sym}$$

证明。 对于 $V \in \mathcal{I}_{mid}$, $\lambda_V = \frac{1}{n^{|V|}}$。然后我们选择 $w_V = \left(\frac{1}{n}\right)^{D_{sos}-|V|}$。对于所有左形状 $\sigma \in \mathcal{L}_V$，很容易利用引理 7.2.7 验证 $w_V \le \frac{w_{U_\sigma}\lambda_{U_\sigma}}{|\mathcal{I}_{mid}|B_{norm}(\sigma)^2 c(\sigma)}()()^2 H_{Id_V}(\sigma,\sigma)$。至此证明完毕。

■

引理 7.2.9。

$$\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}}\frac{d_{Id_U}(H_{Id_U},H'_\gamma)}{|Aut(U)|c(\gamma)}\leq\frac{1}{\Delta^{2D_{sos}}2^{D_V}}$$

证明。 我们采用与引理 7.1.8 相同的论证和记号。代入各个界后，得到

$$\begin{aligned}
&\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}}\frac{d_{Id_U}(H_{Id_U},H'_\gamma)}{|Aut(U)|c(\gamma)}\\
&\qquad\leq\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma')\frac{1}{2^{\min(m_\sigma,m_{\sigma'})-1}}\\
&\qquad\leq\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{1}{n^{0.5\varepsilon C_\Delta|V(\sigma\circ\sigma')|}\Delta^{D_{sos}}n|U|2^{\min(m_\sigma,m_{\sigma'})-1}}\\
&\qquad\leq\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{1}{n^{0.5\varepsilon C_\Delta|V(\sigma\circ\sigma')|}\Delta^{D_{sos}}2^{\min(m_\sigma,m_{\sigma'})-1}}
\end{aligned}$$

其中用到了引理 7.2.7。利用 $n^{0.5C_\Delta|V(\sigma\circ\sigma')|}\geq n^{0.1\varepsilon C_\Delta|V(\sigma\circ\sigma')|}2^{|V(\sigma\circ\sigma')|}$，

$$\begin{aligned}
&\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}}\frac{d_{Id_U}(H_{Id_U},H'_\gamma)}{|Aut(U)|c(\gamma)}\\
&\qquad\leq\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{1}{n^{0.1\varepsilon C_\Delta|V(\sigma\circ\sigma')|}\Delta^{D_{sos}}2^{|V(\sigma\circ\sigma')|}2^{\min(m_\sigma,m_{\sigma'})-1}}\\
&\qquad\leq\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{1}{n^{0.1\varepsilon C_\Delta|V(\sigma\circ\sigma')|}\Delta^{D_{sos}}2^{D_V}}\\
&\qquad\leq\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{1}{D_{sos}^{D_{sos}}n^{0.1\varepsilon C_\Delta|V(\sigma\circ\sigma')|}\Delta^{2D_{sos}}2^{D_V}}
\end{aligned}$$

其中我们将 C_{sos} 设得足够小，使得 $D_{sos}=n^{\varepsilon C_{sos}}\leq n^{c\varepsilon C_\Delta}=\frac{1}{\Delta^\circ}$—最后一步是论证 $\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{1}{D_{sos}^{D_{sos}}n^{0.1C_\Delta\varepsilon|V(\sigma\circ\sigma')|}}\leq 1$，从而完成证明。而只要将 C_V,C_E 设得足够小，就能得到这一点。 ■

我们终于可以完成对截断误差的分析了。

引理 7.2.10。 每当 $\|M_\alpha\| \le B_{norm}(\alpha)$ 对所有 $\alpha \in \mathcal{M}'$ 都成立时，

$$\sum_{U \in \mathcal{I}_{mid}} M_{Id_U}^{fact}(H_{Id_U}) \succeq 6 \left(\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)|c(\gamma)} \right) Id_{sym}$$

证明。 选择足够小的 C_{sos} ，使得 $\frac{\Delta^{2D_{sos}^2}}{n^{D_{sos}}} \ge \frac{6}{\Delta^{2D_{sos}} 2^{D_V}}$ ；令 $C_{sos} < 0.5 C_V$ 即可满足这一条件。然后，由于 $Id_{Sym} \succeq 0$，利用引理 7.2.8 和引理 7.2.9，

$$\begin{aligned}
\sum_{U \in \mathcal{I}_{mid}} M_{Id_U}^{fact}(H_{Id_U}) &\succeq \frac{\Delta^{2D_{sos}^2}}{n^{D_{sos}}} Id_{sym} \\
&\succeq \frac{6}{\Delta^{2D_{sos}} 2^{D_V}} Id_{sym} \\
&\succeq 6 \left(\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)|c(\gamma)} \right) Id_{sym}
\end{aligned}$$

■

7.3 稀疏主成分分析：完整验证

本节将完整证明定理 4.2.3。

定理 4.2.3。 存在一个常数 $C > 0$ ，使得对于所有充分小的常数 $\varepsilon > 0$，若 $m \le \frac{d^{1-\varepsilon}}{\lambda^2}$ 、$m \le \frac{k^{2-\varepsilon}}{\lambda^2}$ ，且存在一个常数 A，使得 $0 < A < \frac{1}{4}$、$d^{4A} \le k \le d^{1-A\varepsilon}$ 和 $\frac{\sqrt{\lambda}}{\sqrt{k}} \le d^{-A\varepsilon}$，则以高概率，由伪校准得到的、度为 $d^{C\varepsilon}$ 的平方和解是可行的。

我们已经在第 6.4 节中给出了相关的定性界。我们将使用该节中的界和记号，并应用相应的技术工具。

定义 7.3.1。定义 $n =$ 最大值$(d, \ m)$。

上述定义　　定义符合技术工具中使用的记号。因此，我们可以使　　e

用那里所述的界。一旦我们验证了条件，定理就会立即由该技术工具推出。

7.3.1 中间形状界

引理 7.3.2. 假设 $0 < A < \frac{1}{4}$ 是一个常数，使得 $\frac{\sqrt{\lambda}}{\sqrt{k}} \leq d^{-A\varepsilon}$ 和 $\frac{1}{\sqrt{k}} \leq d^{-2A}$ 对于所有满足 $m \leq \frac{d^{1-\varepsilon}}{\lambda^2}$ 的 m，$m \leq \frac{k^{2-\varepsilon}}{\lambda^2}$；对于所有 $U \in \mathcal{I}_{mid}$ 和 $\tau \in \mathcal{M}_U$，假设对于所有 $i \in V(\tau) \setminus U_\tau \setminus V_\tau$，$deg^\tau(i)$ 都是偶数，则

$$\sqrt{d}^{|\tau|_1 - |U_\tau|_1} \sqrt{m}^{|\tau|_2 - |U_\tau|_2} S(\tau) \leq \prod_{j \in V_2(\tau) \setminus U_\tau \setminus V_\tau} (deg^\tau(j) - 1)!! \cdot \frac{1}{d^{A\varepsilon \sum_{e \in E(\tau)} l_e}}$$

Proof. 令 $r_1 = |\tau|_1 - |U_\tau|_1, r_2 = |\tau|_2 - |U_\tau|_2$。由于 $\Delta \leq 1$，只需证明

$$E := \sqrt{d}^{r_1} \sqrt{m}^{r_2} \left(\frac{k}{d}\right)^{r_1} \left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{\sum_{e \in E(\tau)} l_e} \leq \frac{1}{d^{A\varepsilon \sum_{e \in E(\tau)} l_e}}$$

我们需要以下断言。

断言 7.3.3. $\sum_{e \in E(\tau)} l_e \geq 2 \max(r_1, r_2)$.

证明。我们首先证明 $\sum_{e \in E(\tau)} l_e \geq 2r_1$。对于任意顶点 $i \in V_1(\tau) \setminus U_\tau \setminus V_\tau$，$deg^\tau(i)$ 是偶数且不为 0，因此，$deg^\tau(i) \geq 2$。任何顶点 $i \in U_\tau \setminus V_\tau$ 都不能有 $deg^\tau(i) = 0$，否则 $U_\tau \setminus \{i\}$ 就是一个权重严格小于 U_τ 的顶点分隔集，这是不可能的，因此，$deg^\tau(i) \geq 1$。同理，对于 $i \in V_\tau \setminus U_\tau$，$deg^\tau(i) \geq 1$。此外，由于 H_τ 是二分的，我们有 $\sum_{i \in V_1(\tau)} deg^\tau(i) = \sum_{j \in V_2(\tau)} deg^\tau(j) = \sum_{e \in E(\tau)} l_e$。考虑

$$\begin{aligned}\sum_{e\in E(\tau)} l_e &= \sum_{i\in V_1(\tau)} deg^{\tau}(i)\\ &\geq \sum_{i\in V_1(\tau)\setminus U_\tau\setminus V_\tau} deg^{\tau}(i) + \sum_{i\in (U_\tau)_1\setminus V_\tau} deg^{\tau}(i) + \sum_{i\in (V_\tau)_1\setminus U_\tau} deg^{\tau}(i)\\ &\geq 2|V_1(\tau)\setminus U_\tau\setminus V_\tau| + |(U_\tau)_1\setminus V_\tau| + |(V_\tau)_1\setminus U_\tau|\\ &= 2r_1\end{aligned}$$

我们可以类似地证明 $\sum_{e\in E(\tau)} l_e \geq 2r_2$ ■

为阐明主要思路，我们将首先证明较弱的界 $E \leq 1$。注意到，我们的假设蕴含 $m \leq \frac{d}{\lambda^2}$，$m \leq \frac{k^2}{\lambda^2}$并且利用事实 $\frac{\sqrt{\lambda}}{\sqrt{k}} \leq d^{-A\varepsilon} \leq 1$，可得 $E \leq \sqrt{d}^{r_1}\sqrt{m}^{r_2}\binom{k}{d}^{r_1}\binom{\sqrt{\lambda}}{\sqrt{k}}^{2\max(r_1,r_2)}$。

命题 7.3.4。对于整数 r_1、$r_2 \geq 0$，若 $m \leq \frac{d}{\lambda^2}$且 $m \leq \frac{k^2}{\lambda^2}$，则，

$$\sqrt{d}^{r_1}\sqrt{m}^{r_2}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2\max(r_1,r_2)} \leq 1$$

证明。我们将分别考虑情形 $r_1 \geq r_2$ 和 $r_1 < r_2$。如果 $r_1 \geq r_2$，则有

$$\begin{aligned}\sqrt{d}^{r_1}\sqrt{m}^{r_2}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_1} &\leq \sqrt{d}^{r_1}\left(\frac{\sqrt{d}}{\lambda}\right)^{r_2}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_1}\\ &= \left(\frac{\lambda}{\sqrt{d}}\right)^{r_1-r_2}\\ &\leq \left(\frac{1}{\sqrt{m}}\right)^{r_1-r_2}\\ &\leq 1\end{aligned}$$

如果 $r_1 < r_2$，则有

$$\begin{aligned}\sqrt{d}^{r_1}\sqrt{m}^{r_2}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_2} &= \sqrt{d}^{r_1}\sqrt{m}^{r_2-r_1}\sqrt{m}^{r_1}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_2}\\ &\le \sqrt{d}^{r_1}\left(\frac{k}{\lambda}\right)^{r_2-r_1}\left(\frac{\sqrt{d}}{\lambda}\right)^{r_1}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_2}\\ &= 1\end{aligned}$$

■

为了得到所需的界，我们仿照这一论证，同时仔细跟踪 d^{ε} 的因子。

命题 7.3.5。 对于整数 r_1、$r_2 \ge 0$，以及整数 $r \ge 2\max(r_1、\ r_2)$，若 $m \le \frac{d^{1-\varepsilon}}{\lambda^2}$ 且 $m \le \frac{k^{2-\varepsilon}}{\lambda^2}$，则，

$$\sqrt{d}^{r_1}\sqrt{m}^{r_2}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r} \le \left(\frac{1}{d^{A\varepsilon}}\right)^{r}$$

证明。如果 $r_1 \ge r_2$,

$$\begin{aligned}E &= \sqrt{d}^{r_1}\sqrt{m}^{r_2}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r-2r_1}\\ &\le \sqrt{d}^{r_1}\left(\frac{\sqrt{d}^{1-\varepsilon}}{\lambda}\right)^{r_2}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r-2r_1}\\ &= \left(\frac{\lambda}{\sqrt{d}^{1-\varepsilon}}\right)^{r_1-r_2}\left(\frac{1}{\sqrt{d}}\right)^{\varepsilon r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r-2r_1}\\ &\le \left(\frac{1}{\sqrt{m}}\right)^{r_1-r_2}\left(\frac{1}{\sqrt{d}}\right)^{\varepsilon r_1}\left(\frac{1}{d^{A\varepsilon}}\right)^{r-2r_1}\\ &\le \left(\frac{1}{d^{2A}}\right)^{\varepsilon r_1}\left(\frac{1}{d^{A\varepsilon}}\right)^{r-2r_1}\\ &= \left(\frac{1}{d^{A\varepsilon}}\right)^{r}\end{aligned}$$

并且如果 $r_1 < r_2$,

$$
\begin{aligned}
E &= \sqrt{d}^{r_1}\sqrt{m}^{r_2-r_1}\sqrt{m}^{r_1}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_2}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r-2r_2} \\
&\le \sqrt{d}^{r_1}\left(\frac{\sqrt{k}^{2-\varepsilon}}{\lambda}\right)^{r_2-r_1}\left(\frac{\sqrt{d}^{1-\varepsilon}}{\lambda}\right)^{r_1}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_2}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r-2r_2} \\
&= \left(\frac{\sqrt{k}}{\sqrt{d}}\right)^{\varepsilon r_1}\left(\frac{1}{\sqrt{k}}\right)^{\varepsilon r_2}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r-2r_2} \\
&\le \left(\frac{1}{\sqrt{k}}\right)^{\varepsilon r_2}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r-2r_2} \\
&\le \left(\frac{1}{d^{2A}}\right)^{\varepsilon r_2}\left(\frac{1}{d^{A\varepsilon}}\right)^{r-2r_2} \\
&\le \left(\frac{1}{d^{A\varepsilon}}\right)^{\sum_{e\in E(\tau)} l_e}
\end{aligned}
$$

■

将上述断言中的 $r = \sum_{e\in E(\tau)} l_e$ 设定为此值即可得到该结果。 ■

推论 7.3.6。 对于所有 $U \in \mathcal{I}_{mid}$ 和 $\tau \in \mathcal{M}_U$，我们有

$$c(\tau)B_{norm}(\tau)S(\tau)R(\tau) \le 1$$

证明。首先注意，如果对于任意顶点 $i \in V(\tau) \setminus U_\tau \setminus V_\tau$， $deg^\tau(i)$ 为奇数，那么 $S(\tau) = 0$，不等式成立。因此，假设对于所有 $i \in V(\tau) \setminus U_\tau \setminus V_\tau$， $deg^\tau(i)$ 都是偶数。由于 τ 是适当中间形状，我们有 $w(I_\tau) = 0$，且 $w(S_{\tau,min}) = w(U_\tau)$。这意味着 $n^{|V(\tau)|+w(I_\tau)/2} = \sqrt{d}^{|\tau|_1-|U_\tau|_1}\sqrt{m}^{|\tau|_2-|U_\tau|_2-w(S_{\tau,min})}$。正如引理 7.3.2 的证明中所观察到的，每个顶点 $i \in V(\tau) \setminus U_\tau$ 或 $i \in V(\tau) \setminus V_\tau$ 都有 $deg^\tau(i) \ge 1$，因此，$|V(\tau) \setminus U_\tau| + |V(\tau) \setminus V_\tau| \le 4\sum_{e\in E(\tau)} l_e$。此外， $q = d^{O(1)\cdot\varepsilon(C_V+C_E)}$。我们可以令 C_V, C_E

足够小，使得

$$\begin{aligned}
c(\tau)B_{norm}(\tau)S(\tau)R(\tau) &= 100(6D_V)^{|U_\tau\setminus V_\tau|+|V_\tau\setminus U_\tau|+2|E(\tau)|}4^{|V(\tau)\setminus(U_\tau\cup V_\tau)|} \\
&\quad \cdot 2e(6qD_V)^{|V(\tau)\setminus U_\tau|+|V(\tau)\setminus V_\tau|}\prod_{e\in E(\tau)}(400D_V^2D_E^2q)^{l_e} \\
&\quad \cdot \sqrt{d}^{|\tau|_1-|U_\tau|_1}\sqrt{m}^{|\tau|_2-|U_\tau|_2}S(\tau)(C_{disc}\sqrt{D_E})^{\sum_{j\in(U_\tau)_2\cup(V_\tau)_2}deg^\tau(j)} \\
&\le d^{O(1)\cdot(C_V+C_E)\cdot\varepsilon\sum_{e\in E(\tau)}l_e}\cdot\prod_{j\in V_2(\tau)\setminus V_2(U_\tau)\setminus V_2(V_\tau)}(deg^\tau(j)-1)!!\cdot\frac{1}{d^{A\varepsilon\sum_{e\in E(\tau)}l_e}} \\
&\le d^{O(1)\cdot(C_V+C_E)\cdot\varepsilon\sum_{e\in E(\tau)}l_e}\cdot(D_VD_E)^{\sum_{e\in E(\tau)}l_e}\cdot\frac{1}{d^{A\varepsilon\sum_{e\in E(\tau)}l_e}} \\
&\le 1
\end{aligned}$$

■

现在我们可以得到所需的中间形状界。

引理 7.3.7。对于所有 $U\in\mathcal{I}_{mid}$ 和 $\tau\in\mathcal{M}_U$，

$$\begin{bmatrix}\frac{1}{|Aut(U)|c(\tau)}H_{Id_U} & B_{norm}(\tau)H_\tau\\ B_{norm}(\tau)H_\tau^T & \frac{1}{|Aut(U)|c(\tau)}H_{Id_U}\end{bmatrix}\succeq 0$$

证明。我们有

$$\begin{aligned}
&\begin{bmatrix}\frac{1}{|Aut(U)|c(\tau)}H_{Id_U} & B_{norm}(\tau)H_\tau\\ B_{norm}(\tau)H_\tau^T & \frac{1}{|Aut(U)|c(\tau)}H_{Id_U}\end{bmatrix} \\
&\quad=\begin{bmatrix}\left(\frac{1}{|Aut(U)|c(\tau)}-\frac{S(\tau)R(\tau)B_{norm}(\tau)}{|Aut(U)|}\right)H_{Id_U} & 0\\ 0 & \left(\frac{1}{|Aut(U)|c(\tau)}-\frac{S(\tau)R(\tau)B_{norm}(\tau)}{|Aut(U)|}\right)H_{Id_U}\end{bmatrix} \\
&\qquad+B_{norm}(\tau)\begin{bmatrix}\frac{S(\tau)R(\tau)}{|Aut(U)|}H_{Id_U} & H_\tau\\ H_\tau^T & \frac{S(\tau)R(\tau)}{|Aut(U)|}H_{Id_U}\end{bmatrix}
\end{aligned}$$

由引理 6.4.9, $\begin{bmatrix} S(\tau)R(\tau)|Aut(U)| H_{Id_U} & H_\tau \\ H_\tau^T & S(\tau)R(\tau)|Aut(U)| H_{Id_U} \end{bmatrix} \succeq 0$，因此上面的第二项是半正定的。对于第一项，由引理 6.4.5，$H_{Id_U} \succeq 0$ 并由推论 7.3.6，$|Aut(U)|c(\tau) - S(\tau)R(\tau)B_{norm}(\tau)|Aut(U)| \geq 0$，这证明第一项也是半正定的。

■

7.3.2 交项界

引理 7.3.8. 假设 $0 < A < \frac{1}{4}$ 是一个常数，使得 $\sqrt{\lambda} \leq d^{-A\varepsilon}\sqrt{k} \leq d^{-2A}\sqrt{k}$ 和 $k \leq d^{-A\varepsilon}$。对于所有满足 $m \leq \frac{d^{1-\varepsilon}}{\lambda^2}$ 的 m、所有 U，其中 $w(U) > w(V)$，以及所有 $\gamma \in \Gamma_{U,V}$，

$$n^{w(V(\gamma)\setminus U_\gamma)} S(\gamma)^2 \leq \left(\prod_{j \in V_2(\gamma)\setminus U_\gamma \setminus V_\gamma} (deg^\gamma(j) - 1)!! \right)^2 \frac{1}{d^{B\varepsilon(|V(\gamma)\setminus(U_\gamma \cap V_\gamma)| + \sum_{e \in E(\gamma)} l_e)}}$$

其中某个常数 $B > 0$ 仅依赖于 C_Δ。特别地，它与 C_V 和 C_E 无关。

证明. 假设存在一个顶点 $i \in V(\gamma) \setminus U_\gamma \setminus V_\gamma$，使得 $deg^\gamma(i)$ 为奇数，那么 $S(\gamma) = 0$，且不等式成立。因此，假设对于所有顶点 $i \in V(\gamma) \setminus U_\gamma \setminus V_\gamma$，$deg^\gamma(i)$ 均为偶数。我们有 $n^{w(V(\gamma)\setminus U_\gamma)} = d^{|\gamma|_1 - |U_\gamma|_1} m^{|\gamma|_2 - |U_\gamma|_2}$。代入 $S(\gamma)$，可知我们需要证明

$$E := d^{|\gamma|_1 - |U_\gamma|_1} m^{|\gamma|_2 - |U_\gamma|_2} \left(\frac{k}{d}\right)^{2|\gamma|_1 - |U_\gamma|_1 - |V_\gamma|_1} \Delta^{2|\gamma|_2 - |U_\gamma|_2 - |V_\gamma|_2} \prod_{e \in E(\gamma)} \frac{\lambda^{l_e}}{k^{l_e}} \leq \frac{1}{d^{B\varepsilon(|V(\gamma)\setminus(U_\gamma \cap V_\gamma)| + \sum_{e \in E(\gamma)} l_e)}}$$

令 S' 为具有 $deg^\gamma(i) \geq 1$ 的顶点集合 $i \in U_\gamma \setminus V_\gamma$。令 e, f 分别表示 S' 中类型 1 顶点和类型 2 顶点的数量。注意，$S' \cup (U_\gamma \cap V_\gamma)$ 是 γ 的一个顶点分隔集。令 $g = |V_\gamma \setminus U_\gamma|_1$（分别为 $h = |V_\gamma \setminus U_\gamma|_2$）表示 $V_\gamma \setminus U_\gamma$ 中类型 1 顶点（分别为类型 2 顶点）的数量。我们首先断言 $d^e m^f \geq d^g m^h$。为此，注意顶点分隔集 $S' \cup (U_\gamma \cap V_\gamma)$ 的权重为 $\sqrt{d}^{e + |U_\gamma \cap V_\gamma|_1} \sqrt{m}^{f + |U_\gamma \cap V_\gamma|_2}$。另一方面，

权重为 $\sqrt{d}^{g+|U_\gamma\cap V_\gamma|_1}\sqrt{m}^{h+|U_\gamma\cap V_\gamma|_2}$。由于 γ 是左形状，V_γ 是唯一的最小顶点分隔集，因此有不等式 $\sqrt{d}^{e+|U_\gamma\cap V_\gamma|_1}\sqrt{m}^{f+|U_\gamma\cap V_\gamma|_2}\geq\sqrt{d}^{g+|U_\gamma\cap V_\gamma|_1}\sqrt{m}^{h+|U_\gamma\cap V_\gamma|_2}$，这意味着 $d^e m^f \geq d^g m^h$。令 $p = |V(\gamma)\setminus(U_\gamma\cup V_\gamma)|_1$（分别为 $q = |V(\gamma)\setminus(U_\gamma\cup V_\gamma)|_2$）表示 $V(\gamma)\setminus(U_\gamma\cup V_\gamma)$ 中类型 1 顶点（分别为类型 2 顶点）的数量。为说明主要思路，我们先证明较弱的不等式 $E\leq 1$。由于 $\Delta\leq 1$，只需证明

$$d^{|\gamma|_1-|U_\gamma|_1}m^{|\gamma|_2-|U_\gamma|_2}\left(\frac{k}{d}\right)^{2|\gamma|_1-|U_\gamma|_1-|V_\gamma|_1}\prod_{e\in E(\gamma)}\frac{\lambda^{l_e}}{k^{l_e}}\leq 1$$

我们有 $d^{|\gamma|_1-|U_\gamma|_1}m^{|\gamma|_2-|U_\gamma|_2} = d^{p+g}m^{q+h}\leq n^{p+\frac{e+g}{2}}m^{q+\frac{f+h}{2}}$因为 $d^e m^f\geq d^g m^h$。此外，$2|\gamma|_1-|U_\gamma|_1-|V_\gamma|_1 = 2p+e+g$。因此，只需证明

$$n^{p+\frac{e+g}{2}}m^{q+\frac{f+h}{2}}\left(\frac{k}{d}\right)^{2p+e+g}\prod_{e\in E(\gamma)}\left(\frac{\lambda}{k}\right)^{l_e}\leq 1$$

我们需要以下命题。

命题 7.3.9。$\sum_{e\in E(\gamma)} l_e\geq\max(2p+e+g,\ 2q+f+h)$

证明。由于 H_γ是二分的，我们有$\sum_{e\in E(\gamma)} l_e = \sum_{i\in V_1(\gamma)} deg^\gamma(i) = \sum_{i\in V_2(\gamma)} deg^\gamma(i)$。注意，所有顶点 $i\in V(\gamma)\setminus U_\gamma\setminus V_\gamma$ 都有 $deg^\gamma(i)$ 非零且为偶数，因此，$deg^\gamma(i)\geq 2$。于是，

$$\begin{aligned}\sum_{e\in E(\gamma)} l_e &= \sum_{i\in V_1(\gamma)} deg^\gamma(i)\\ &\geq \sum_{i\in V_1(\gamma)\setminus U_\gamma\setminus V_\gamma} deg^\gamma(i)+\sum_{i\in(U_\gamma)_1\setminus V_\gamma} deg^\gamma(i)+\sum_{i\in(V_\gamma)_1\setminus U_\gamma} deg^\gamma(i)\\ &\geq 2p+e+g\end{aligned}$$

同理，

$$\begin{aligned}\sum_{e\in E(\gamma)} l_e &= \sum_{i\in V_2(\gamma)} deg^\gamma(i)\\ &\geq \sum_{i\in V_2(\gamma)\setminus U_\gamma\setminus V_\gamma} deg^\gamma(i) + \sum_{i\in (U_\gamma)_2\setminus V_\gamma} deg^\gamma(i) + \sum_{i\in (V_\gamma)_2\setminus U_\gamma} deg^\gamma(i)\\ &\geq 2q+f+h\end{aligned}$$

因此，$\sum_{e\in E(\gamma)} l_e \geq \max(2p+e+g, 2q+f+h)$。 ■

现在，设 $r_1 = p + \frac{e+g}{2}, r_2 = q + \frac{f+h}{2}$。那么，$\sum_{e\in E(\gamma)} l_e \geq 2\max(r_1, r_2)$，我们希望证明 $d^{r_1}m^{r_2}\left(\frac{k}{d}\right)^{2r_1-\cdots}\left(\frac{\lambda}{k}\right)^{2-\cdots}{}^{\max(r_1,r_2)} \leq 1$ 这个表达式直接由命题 7.3.4 两边平方得到。

现在，为了证明 $E \leq \frac{1}{d^{B\varepsilon|V_\gamma\setminus U_\gamma\cap V_\gamma(()())|+\sum_{e\in E(\gamma)} l_e)}}$，我们沿用这一论证，同时仔细追踪因子 d^ε。再次利用 $d^e m^f \geq d^g m^h$，只需证明

$$\begin{aligned}d^{p+\frac{e+g}{2}}m^{q+\frac{f+h}{2}}\left(\frac{k}{d}\right)^{2|\gamma|_1-|U_\gamma|_1-|V_\gamma|_1}&\Delta^{2|\gamma|_2-|U_\gamma|_2-|V_\gamma|_2}\prod_{e\in E(\gamma)}\frac{\lambda^{l_e}}{k^{l_e}}\\ &\leq \frac{1}{d^{B\varepsilon(|V(\gamma)\setminus(U_\gamma\cap V_\gamma)|+\sum_{e\in E(\gamma)} l_e)}}\end{aligned}$$

其思想是，边的 $d^{B\varepsilon}$ 衰减来自于对m的更强假设，即 $m \leq d^{1-\varepsilon}_{\lambda^2}$，$m \leq k^{2-\varepsilon}_{\lambda^2}$。而 $d^{B\varepsilon}$ 型顶点的 $V(\gamma)\setminus(U_\gamma\cap V_\gamma)$衰减，一方面来自于对 m 的更强假设，另一方面来自于 $\frac{k}{d}$中的因子，后者对于零度顶点尤其有用。最后，$d^{B\varepsilon}$ 型顶点的$V(\gamma)\setminus(U_\gamma\cap V_\gamma)$衰减来自于 Δ中的因子。事实上，注意到对于一个取决于 C_Δ, $\Delta^{2|\gamma|_2-|U_\gamma|_2-|V_\gamma|_2} \leq d^{-B\varepsilon|V_{(\gamma)}\setminus(\overline{U_\gamma\cap V_\gamma})|_2}$的常数 B。因此，只要证明

$$d^{p+\frac{e+g}{2}}m^{q+\frac{f+h}{2}}\left(\frac{k}{d}\right)^{2|\gamma|_1-|U_\gamma|_1-|V_\gamma|_1}\left(\frac{\lambda}{k}\right)^{\sum_{e\in E(\gamma)} l_e} \leq \frac{1}{d^{B\varepsilon(|V(\gamma)\setminus(U_\gamma\cap V_\gamma)|_1+\sum_{e\in E(\gamma)} l_e)}}$$

设 c_0 为 $V(\gamma) \setminus (U_\gamma \cap V_\gamma)$ 中满足 $deg^\gamma(i) = 0$的 i 型顶点的数量。由于它们的度为 0，因此必定位于 $(U_\gamma)_1 \setminus V_\gamma$ 中。此外，我们有 $2|\gamma|_1 - |U_\gamma|_1 - |V_\gamma|_1 = 2p+e+g+c_0$，从而 $\binom{k}{d}^{2|\gamma|_1-|U_\gamma|_1-|V_\gamma|_1} = \binom{k}{d}^{2p+e+g+c_0}$。对于这些零度顶点，$\frac{k}{d} \leq d^{-A\varepsilon}$ 中的因子会带来 $\frac{1}{d^{B\varepsilon}}$的衰减。因此，只需证明

$$d^{p+\frac{e+g}{2}} m^{q+\frac{f+h}{2}} \left(\frac{k}{d}\right)^{2p+e+g} \left(\frac{\lambda}{k}\right)^{\sum_{e\in E(\gamma)} l_e} \leq \frac{1}{d^{B\varepsilon(p+q+e+f+g+h)+\sum_{e\in E(\gamma)} l_e)}}$$

对于某个常数 $B > 0$。注意到 $p + q + e + f + g + h \leq 2(\sum_{e\in E(\gamma)} l_e)$。因此，使用记号 $r_1 = p + \frac{e+g}{2}, r_2 = q + \frac{f+h}{2}$，只需证明

$$d^{r_1} m^{r_2} \left(\frac{k}{d}\right)^{2r_1} \left(\frac{\lambda}{k}\right)^{\sum_{e\in E(\gamma)} l_e} \leq \frac{1}{d^{B\varepsilon \sum_{e\in E(\gamma)} l_e}}$$

对于某个常数 $B > 0$。但这可通过对命题 7.3.5 平方得到，其中我们令 $r = \sum_{e\in E(\gamma)} l_e$。

■

注 7.3.10. 在上述界中，请注意，对于 $V(\gamma) \setminus (U_\gamma \cap V_\gamma)$ 中的每个顶点，都会有 $d^{B\varepsilon}$ 的衰减。在植入分布中引入松弛参数 C_Δ 的一个主要技术原因，就是为了引入这种衰减；当前的技术工具需要这种衰减。

由此，我们得到交项界。

引理 7.3.11. 对于所有满足 $w(U) > w(V)$ 的 $U, V \in \mathcal{I}_{mid}$，以及所有 $\gamma \in \Gamma_{U,V}$，

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H^{-\gamma,\gamma}_{Id_V} \preceq H'_\gamma$$

证明。 由引理 6.4.10，我们有

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H^{-\gamma,\gamma}_{Id_V} \preceq c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 R(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} H'_\gamma$$

使用与引理 6.4.5 中相同的证明，我们可以看出 $H'_\gamma \succeq 0$。因此，只需证明 $c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 R(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} \leq 1$。由于 $U, V \in \mathcal{I}_{mid}$，$Aut(U) = |U|_1! |U|_2! \; Aut(V) = |V|_1! |V|_2!$. 因此，$\frac{|Aut_V U|}{|Aut\ V|}(\ )(\ ) = \frac{|U|_1}{|V|_1! |U|_2! \; ! |V|_2!} \leq D_V^{|U_\gamma \setminus V_\gamma|}$。此外，$|E(\gamma)| \leq \sum_{e \in E(\gamma)} l_e$ 以及 $q = d^{O(1) \cdot \varepsilon (C_V + C_E)}$。注意 $R(\gamma)^2 = (C_{disc} \sqrt{D_E})^2 \sum_{j \in (U_\gamma)_2 \cup (V_\gamma)_2} deg^\gamma(j) \leq d^{O(1) \cdot \varepsilon C_E \cdot \sum_{e \in E(\gamma)} l_e}$ 以及

$$\left(\prod_{j \in V_2(\gamma) \setminus U_\gamma \setminus V_\gamma} (deg^\gamma(j) - 1)!! \right)^2 \leq (D_V D_E)^{2 \sum_{e \in E(\tau)} l_e} \leq d^{O(1) \cdot \varepsilon (C_V + C_E) \cdot \sum_{e \in E(\gamma)} l_e}$$

令 B 为引理 7.3.8 中的常数。我们可以将 C_V, C_E 设得足够小，从而利用引理 7.3.8，

$$\begin{aligned}
& c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 R(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} \\
& \quad \leq 100^2 (6 D_V)^{2|U_\gamma \setminus V_\gamma| + 2|V_\gamma \setminus U_\gamma| + |E(\alpha)|} 16^{|V(\gamma) \setminus (U_\gamma \cup V_\gamma)|} \\
& \quad \quad \cdot (3 D_V)^{4|V(\gamma) \setminus V_\gamma| + 2|V(\gamma) \setminus U_\gamma|} (6 q D_V)^{2|V(\gamma) \setminus U_\gamma| + 2|V(\gamma) \setminus V_\gamma|} \prod_{e \in E(\gamma)} (400 D_V^2 D_E^2 q)^{2 l_e} \\
& \quad \quad \cdot n^{w(V(\gamma) \setminus U_\gamma)} S(\gamma)^2 d^{O(1) \cdot \varepsilon C_E \cdot \sum_{e \in E(\gamma)} l_e} \cdot D_V^{|U_\gamma \setminus V_\gamma|} \\
& \quad \leq d^{O(1) \cdot \varepsilon (C_V + C_E) \cdot (|V(\gamma) \setminus (U_\gamma \cap V_\gamma)| + \sum_{e \in E(\gamma)} l_e)} \cdot n^{w(V(\gamma) \setminus U_\gamma)} S(\gamma)^2 \\
& \quad \leq d^{O(1) \cdot \varepsilon (C_V + C_E) \cdot (|V(\gamma) \setminus (U_\gamma \cap V_\gamma)| + \sum_{e \in E(\gamma)} l_e)} \cdot \frac{1}{d^{B \varepsilon (|V(\gamma) \setminus (U_\gamma \cap V_\gamma)| + \sum_{e \in E(\gamma)} l_e)}} \\
& \quad \leq 1
\end{aligned}$$

■

7.3.3 截断误差界

在本节中，我们将使用 [191,节 10] 中概述的策略来获得截断误差界。我们还会沿用该记号。为此，我们首先需要对量 $B_{norm}(\sigma) B_{norm}(\sigma') H_{Id_U}(\sigma, \sigma')$ 得到一个界。

引理 7.3.12。设 $0 < A < \frac{1}{4}$ 是一个满足 $\frac{\sqrt{\lambda}}{\sqrt{k}} \leq d^{-A\varepsilon}$ 和 $\frac{1}{\sqrt{k}} \leq d^{-2A}$ 的常数。设 m 满足 $m \leq \frac{d^{1-\varepsilon}}{\lambda^2}$、$m \leq \frac{k^{2-\varepsilon}}{\lambda^2}$。对于所有 $U \in \mathcal{I}_{mid}$ 和 σ、$\sigma' \in \mathcal{L}_U$，

$$B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma') \leq \frac{1}{d^{0.5A\varepsilon(|V(\sigma\circ\sigma')|+\sum_{e\in E(\alpha)}l_e}} \cdot \frac{1}{d^{|U_\sigma|_1+|U_{\sigma'}|_1}m^{|U_{\sigma'}|_2+|U_{\sigma'}|_2}}$$

证明。假设存在一个顶点 $i \in V(\sigma) \setminus V_\sigma$，使得 $deg^\sigma(i) + deg^{U_\sigma}(i)$ 为奇数，则$H_{Id_U}(\sigma, \sigma') = 0$，且该不等式成立。因此，假设对于所有 $i \in V(\sigma) \setminus V_\sigma$，$deg^\sigma(i) + deg^{U_\sigma}(i)$ 都是偶数。同理，假设对于所有 $i \in V(\sigma') \setminus V_{\sigma'}$，$deg^{\sigma'}(i) + deg^{U_{\sigma'}}(i)$ 都是偶数。此外，如果 $\rho_\sigma \neq \rho_{\sigma'}$，则有 $H_{Id_U}(\sigma, \sigma') = 0$，结论便得证。因此，假设 $\rho_\sigma = \rho_{\sigma'}$。

设 $V(\sigma) \setminus U_\sigma \setminus V_\sigma$ 中有 e 个类型 1 顶点（分别有 f 个类型 2 顶点）。那么，$n^{w(V(\sigma))-w(U_\sigma)} \sqrt{d}^{|V(\sigma)|_1-|U|_1}\sqrt{m}^{|V(\sigma)|_2-|U|_2} = \sqrt{d}^{|U_\sigma|_1}\sqrt{m}^{|U_\sigma|_2}\sqrt{d}^{e}\sqrt{m}^{f}$。设 $V(\sigma') \setminus U_{\sigma'} \setminus V_{\sigma'}$ 中有 g 个类型 1 顶点（分别有 h 个类型 2 顶点）。那么，同理，$n^{w(V(\sigma'))-w(U)} \leq \sqrt{d}^{|U_{\sigma'}|_1}\sqrt{m}^{|U_{\sigma'}|_2}\sqrt{d}^{g}\sqrt{m}^{h}$。

令 $\alpha = \sigma \circ \sigma'$。由于 $V(\alpha) \setminus U_\alpha \setminus V_\alpha$ 中所有顶点的度数至少为 2，因此有$\sum_{e\in E(\alpha)} l_e \geq \sum_{i\in V_1(\alpha)\setminus U_\alpha\setminus V_\alpha} deg^\alpha(i) \geq 2(e+g) + |U_\sigma|_1 + |U_\sigma|_2$。类似地，$\sum_{e\in E(\alpha)} l_e \geq 2(f+h) + |U_{\sigma'}|_1 + |U_{\sigma'}|_2$。因此，在命题 7.3.5 中令 $r_1 = e+g$，$r_2 = f+h$，可得

$$\sqrt{d}^{e+g}\sqrt{m}^{f+h}\left(\frac{k}{d}\right)^{e+g}\prod_{e\in E(\alpha)}\frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}} \leq \frac{1}{d^{A\varepsilon\sum_{e\in E(\alpha)}l_e}}$$

此外，

$$\left(\frac{k}{d}\right)^{|\alpha|_1} \leq \left(\frac{k}{d}\right)^{e+g+|U_\sigma|_1+|U_{\sigma'}|_1}$$

and

$$\prod_{j\in V_2(\alpha)}(deg^\alpha(j)-1)!! \leq d^{\varepsilon C_V\sum_{e\in E(\alpha)}l_e}$$

因此,

$$
\begin{aligned}
&n^{\frac{w(V(\sigma))-w(U)}{2}} n^{\frac{w(V(\sigma'))-w(U)}{2}} H_{Id_U}(\sigma,\sigma') \\
&\le d^{O(1)D_{sos}} \sqrt{d}^{e} \sqrt{m}^{f} d^{O(1)D_{sos}} \sqrt{d}^{g} \sqrt{m}^{h} \\
&\quad \cdot \frac{1}{|Aut(U)|} \left(\frac{1}{\sqrt{k}}\right)^{deg(\alpha)} \left(\frac{k}{d}\right)^{|\alpha|_1} \Delta^{|\alpha|_2} \prod_{j\in V_2(\alpha)} (deg^{\alpha}(j)-1)!! \prod_{e\in E(\alpha)} \frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}} \\
&\le d^{O(1)D_{sos}} d^{\varepsilon C_V \sum_{e\in E(\alpha)} l_e} \sqrt{d}^{e+g} \sqrt{m}^{f+h} \left(\frac{k}{d}\right)^{e+g} \prod_{e\in E(\alpha)} \frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}} \cdot \frac{1}{d^{|U_\sigma|_1+|U_{\sigma'}|_1} m^{|U_{\sigma'}|_2+|U_{\sigma'}|_2}} \\
&\le \frac{d^{\varepsilon C_V \sum_{e\in E(\alpha)} l_e}}{d^{A\varepsilon \sum_{e\in E(\alpha)} l_e}} \cdot \frac{1}{d^{|U_\sigma|_1+|U_{\sigma'}|_1} m^{|U_{\sigma'}|_2+|U_{\sigma'}|_2}}
\end{aligned}
$$

通过将 C_V, C_E 设得足够小，并代入 $B_{norm}(\sigma)$ $B_{norm}(\sigma')$ 的表达式，即可得到所需结果。

■

现在可以应用该策略了。

引理 7.3.13。 只要对所有 $\alpha \in \mathcal{M}'$，都有 $\|M_\alpha\| \le B_{norm}(\alpha)$，

$$
\sum_{U\in\mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) \succeq \frac{1}{d^{K_1 D_{sos}^2}} Id_{sym}
$$

其中常数 $K_1 > 0$ 可以依赖于 C_Δ。

证明。 对于 $V \in \mathcal{I}_{mid}$, $\lambda_V = \frac{\Delta^{|V|_2}}{d^{|V|_1} k^{|V|_2}}$。令该量在所有 V 上的最小值为 N。接着我们选择 $w_V = N/\lambda_V$, 使得对于所有左形状 $\sigma \in \mathcal{L}_V$，由引理 7.3.12 可得 $w_V \le \frac{|\mathcal{I}_{mid}| B_{norm}\ \sigma\ ^2 c\ \sigma\ ()()^2 H_{Id_V}(\sigma,\sigma)}{w_{U_\sigma}\lambda_{U_\sigma}}$，证明完毕。

■

引理 7.3.14。

$$
\sum_{U\in\mathcal{I}_{mid}} \sum_{\gamma\in\Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)| c(\gamma)} \le \frac{d^{K_2 D_{sos}}}{2^{D_V}}
$$

对于一个可能依赖于 C_Δ 的常数 $K_2 > 0$ 。

证明。 我们进行与引理 7.1.8 的证明相同的计算，直到

$$\begin{aligned}
&\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}}\frac{d_{Id_U}(H_{Id_U},H'_\gamma)}{|Aut(U)|c(\gamma)}\\
&\qquad\le\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma')\frac{1}{2^{\min(m_\sigma,m_{\sigma'})-1}}\\
&\qquad\le\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{d^{O(1)D_{sos}}}{d^{0.5A\varepsilon|V(\sigma\circ\sigma')|}2^{\min(m_\sigma,m_{\sigma'})-1}}
\end{aligned}$$

这里用到了引理 7.3.12。利用 $d^{0.5A\varepsilon|V(\sigma\circ\sigma')|}\ge d^{0.1A\varepsilon|V(\sigma\circ\sigma')|}2^{|V(\sigma\circ\sigma')|}$，

$$\begin{aligned}
\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}}\frac{d_{Id_U}(H_{Id_U},H'_\gamma)}{|Aut(U)|c(\gamma)}&\le\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{d^{O(1)D_{sos}}}{d^{0.1A\varepsilon|V(\sigma\circ\sigma')|}2^{|V(\sigma\circ\sigma')|}2^{\min(m_\sigma,m_{\sigma'})-1}}\\
&\le\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{d^{O(1)D_{sos}}}{d^{0.1A\varepsilon|V(\sigma\circ\sigma')|}2^{D_V}}\\
&\le\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{d^{O(1)D_{sos}}}{D_{sos}^{D_{sos}}d^{0.1A\varepsilon|V(\sigma\circ\sigma')|}2^{D_V}}
\end{aligned}$$

最后一步是证明 $\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{1}{D_{sos}^{D_{sos}}d^{0.1A\varepsilon|V(\sigma\circ\sigma')|}}\le 1$，这样就完成了证明。但只要将 C_V,C_E 设得足够小，这一点就能成立。 ■

最后，我们可以证明截断误差是可以处理的。

引理 7.3.15。 只要对所有 $\alpha\in\mathcal{M}'$都有 $\|M_\alpha\|\le B_{norm}(\alpha)$，

$$\sum_{U\in\mathcal{I}_{mid}}M_{Id_U}^{fact}(H_{Id_U})\succeq 6\left(\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}}\frac{d_{Id_U}(H'_\gamma,H_{Id_U})}{|Aut(U)|c(\gamma)}\right)Id_{sym}$$

证明。 选择足够小的 C_{sos}，使得 $\frac{1}{d^{K_1D_{sos}^2}}\ge 6\frac{d^{K_2D_{sos}}}{2^{D_V}}$；只要将 $C_{sos}<K_3C_V$设为足够小的常数 $K_3>0$，就可以满足这一条件。然后，由于 $Id_{Sym}\succeq 0$，利用

引理 7.3.13 和引理 7.3.14,

$$\begin{aligned}\sum_{U\in\mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) &\succeq \frac{1}{d^{K_1 D^2_{sos}}} Id_{sym}\\ &\succeq 6\frac{d^{K_2 D_{sos}}}{2^{D_V}} Id_{sym}\\ &\succeq 6\left(\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)|c(\gamma)}\right) Id_{sym}\end{aligned}$$

■

第8章

后续工作与未来研究

本章将回顾一些本学位论文未涉及的后续研究，并提出未来工作的方向。最后，我们将简要讨论本研究对计算机科学更广泛的影响，以此作为本学位论文的结语。

8.1 非乘积分布的非线性集中性

第2章中的技术适用于一组彼此独立抽取的随机变量。一个自然的问题是：能否将其推广到这些变量并非独立的情形？例如，相较于分析 Erd˝os- R´enyi 随机图，我们可能希望分析均匀的 d-正则图，此时这种推广就很有用。这种推广似乎很有可能实现，因为我们的证明技术本质上只需要一个能够快速混合到给定分布的马尔可夫链，随后便可递归应用 Poincar´e 不等式。我们将此留待未来研究。

8.2 平方和下界

在这篇学位论文中，我们介绍了若干平方和下界。尽管这些下界建立在一些基本的概念性基石之上，例如我们证明的非线性集中结果，以及伪校准这类简单启发式方法，但当前证明中的一个重要技术障碍是证明本身技术性很强，涉及许多相互关联的环节。弄清楚能否简化这些证明，是一个重要的研究问题。除了有助于更好地理解平方和层级之外，这也将帮助我们理解计算机科学中若干基本问题所面临的计算障碍。以下是这类问题的一些例子。

8.2.1 稀疏独立集

在一项后续工作中， [119], 我们证明了稀疏 Erd˝os-R´enyi 随机图上最大独立集这一重要问题的平方和下界。

本学位论文研究的平方和下界，其输入均取自乘积分布，其中每个分布要么是拉ademacher分布，要么是高斯分布。许多先前关于 SoS 下界的工作也采用了这种情形。回想一下，第2章将其称为稠密情形 。同样重要的是研究引人入胜的平均情形稀疏情形 问题，其中输入分布可能具有较高的奥尔利奇范数；例如，输入可以是从$G_{n,p}$ 而非$G_{n,12}$ 中抽样得到的 Erd˝os-R´enyi 随机图，其中 $p = o(1)$。本工作及先前工作中为证明高阶平方和下界而发展的技术，并不能轻易推广到这种情形。该工作[119] 开启了这一研究方向，聚焦于随机稀疏图上的基本问题——最大独立集问题。

考虑图$G \sim G_{n,\frac{d}{n}}$ 上的独立集问题，其中 d 是平均度。如果 $d = \frac{n}{2}$，那么这与最大团问题相同，而 [20]中已经得到了平方和下界。现在我们聚焦于情形 $d \ll n$。我们先给出真实最优值的大小。

事实 8.2.1 ([53, 55, 63])。 以高概率， G 中的最大独立集大小为$(1+o_d(1)) \cdot \frac{2\ln d}{d} \cdot n$。

著名的 Lov´asz ϑ 函数可以高效地计算该值的上界，并且在此类随机图中，其值是众所周知的。

事实 8.2.2 ([52])。 以高概率， $\vartheta(G) = \Theta(\frac{n}{\sqrt{d}})$。

该 ϑ 函数的值也是此问题的二次平方和松弛的输出。因此，存在约为 $\sqrt{d}$ 的整数间隙。因此，我们自然会问，更高阶 SoS 是否能取得更好的效果，还是这一间隙会持续存在。在我们的工作中，我们证明了，这一 $\sqrt{d}$ 整数间隙在更高阶的 SoS 中同样持续存在

我们证明了两个主要结果，一个是在 $(\log n)^2 \le d \le \sqrt{n}$ 的情形下，另一个是在 $n^{\Omega(1)} \le d \le \frac{n}{2}$ 的情形下。请注意，我们尚未涵盖平均度 d 为常数的情形。这是未来工作中一个有趣的研究方向。

在第一种情形（$\log n$）$^2 \le d \le \sqrt{n}$中，我们展示了 SoS 松弛的度 D_{sos} 与整数间隙之间的权衡。

定理 8.2.3. 存在一个绝对常数 $c_0 \in \mathbb{N}$，使得对于充分大的 $n \in \mathbb{N}$和 $d \in [(\log n)^2$，以及满足 $k \le \frac{n^{c_0}}{D_{SoS} \cdot \log n \cdot d^{1/2}}$ 的参数 k、$DSoS$，以高概率，对于$G \sim G_{n,\, d/n}$，存在一个用于最大独立集问题的度为 D 的 SoS 伪期望，其目标值为 $(1 - o(1))k$。

特别地，当 $d \in [n^{\Omega(1)}\sqrt{n}]$ 时，这给出了针对多项式度 $n^{\Omega(1)}$ SoS 的平方和下界。在第二种情形 $n^{\Omega(1)} \le d \le \frac{n}{2}$中，我们给出了针对对数次数 SoS 的平方和下界。

定理 8.2.4。对于任意 $\varepsilon_1, \varepsilon_2 > 0$，存在 $\delta > 0$，使得对于 $d \in [n^{\varepsilon_1}, n/2]$ 和 $k \le \frac{n}{d^{1/2+\varepsilon_2}}$，以 $G \sim G_{n,\, d/n}$ 为条件时，以高概率存在一个度为 $(\delta \log d)$ 的伪期望，其目标值为 $(1 - o(1))k$。

我们指出，这些定理排除了对任意 $d \ge$polylog(n) 进行多项式时间认证（即常数次数 SoS）的可能性。

总体而言，我们采用类似的技术来证明这些结果，即伪校准、图矩阵和近似半正定分解。不过，这种方法并不能直接奏效，我们通过若干新思路和新技术克服了这些困难。下面总结其中一些。

- 我们克服的第一个概念性难题是，由于缺乏合适的候选植入分布，我们无法应用伪校准。对于大多数自然的植入分布选择，简单统计量都能区分随机分布和植入分布。虽然很可能存在一种合适的植入分布，使我们能够使用伪校准，但我们尚未找到。

相反，在这项工作中，我们直接采用朴素的植入分布，但修改了伪校准的启发式方法（我们称之为带连通截断的伪校准），以构造候选矩矩阵。

- 第二个概念上的难点是，基于稀疏图构造的图矩阵缺乏良好的范数界。在那项工作中，我们通过仔细分析运用迹方法，得到了更好的范数界。此外，正如我们在第2章中看到的，我们也能够利用一般递归定理，在不使用迹方法的情况下得到类似的范数界。

除了上述进展之外，我们还开发了若干技术工具，例如条件化、交集权衡引理的推广等。更多详情请参见 [119]。

8.2.2 植入仿射平面与最大割

对于第5章中的植入仿射平面问题，我们从 $\mathcal{N}(0, I_n)$中独立地采样了 m 个向量 $d_1, \dots, d_m$，并证明了 $m \le n^{3/2-\varepsilon}$的平方和下界。然而，根据注 5.4.9 中对 $\widetilde{\mathbb{E}}[1]$ 的分析，我们预期 $m \ll n^{2-\varepsilon}$也成立下界。这是因为，正如我们在第3章中所见，并将在下一节重新讨论的，分析$\sim\mathbb{E}[1]$ 是推测平方和之幂的一种成熟方法。因此，我们猜想**猜想 8.2.5.** 定理 *4.1.4* 对采样向量数量的界仍然成立，m 只需将其放宽为 $m \le n^{2-\varepsilon}$。

与此对偶（事实上，我们在第5章的证明中利用了这种对偶性），我们猜想，只要满足 $p \ge n^{1/2+\varepsilon}$，植入布尔向量问题就存在平方和下界。

猜想 8.2.6. 当随机子空间的维数上界放宽至 $p \ge n^{1/2+\varepsilon}$ 时，定理 *4.1.5* 仍然成立 p 。

我们指出，近期工作 [238] 展示了针对 $m \geq n+1$ 的植入仿射平面搜索变体的多项式时间算法，而此前已知的算法则需要$m \gg n^2$ [161]。 [161] 中的算法采用谱算法，对噪声具有鲁棒性，而且很可能可由平方和捕获。另一方面， [238] 中的算法基于格，并不对噪声具有鲁棒性（即它假设所有向量都必须精确地位于这两个平面上），也无法由平方和捕获。

在我们针对植入布尔向量问题和植入仿射平面问题的平方和下界中，我们假设输入项是独立同分布地从高斯分布或布尔分布中选取的。事实上，我们的证明技术很可能在分布“足够随机”时仍然适用，例如从球面上的均匀分布中抽样。由此直觉可以得到一个可能的推广：在植入布尔向量问题中，如果子空间是随机邻接矩阵底部特征向量所张成的特征子空间，那么这个实例仍然应该很难。后一种情形出现在最大割问题中；对此，我们提出如下猜想。**猜想 8.2.7.** 设 $d \geq 3$，并令 G 为一个随机 d-正则图，其有 n 个顶点。对于某个 $\delta > 0$，以高概率存在一个度为 n^{δ} 的伪期望算子 $\widetilde{\mathbb{E}}$，其作用于布尔变量 x_i，且最大割值至少为

$$\frac{1}{2} + \frac{\sqrt{d-1}}{d}(1 - \mathrm{o}_{d,n}(1))$$

上式以高概率等于最大割的谱松弛值，因此，从定性的角度看，这个猜想表达的是，度为 n^{δ} 的平方和方法无法显著收紧基本的谱松弛。

值得指出的是，就证明平方和无法显著优于 Goemans-Williamson 松弛这一目标而言，随机实例并不是整数间隙实例。将（甚至是 4 阶）平方和与 Goemans-Williamson 算法进行比较时，主要困难似乎在于缺乏一种候选困难输入分布。

支持这一猜想的证据在于，对

随机输入 $d_1, \ldots, d_m$ 所要求的唯一性质，是带有厄米特多项式项的图矩阵满足范数界。当变量 $\{d_{u,i}\}$ 是从其他某种分布中独立同分布抽取时，如果我们使用该分布下正交归一多项式对应的图矩阵，并假设该分布的矩满足适当的界，那么同样的范数界也成立 [1]。当d_u 从球面或其他坐标非独立同分布的分布中均匀抽取时，类似的范数界似乎也很可能成立。此外，正如上一节所述，第2章中的技巧可能有助于得到这样的范数界。

8.2.3 唯一博弈

著名的唯一博弈猜想（UGC） [125] 断言，一种称为唯一博弈问题的图论问题是 NP 难的。由于这一猜想带来了众多推论（例如 [125, 127, 195]），并且与度量几何 [129] 和离散傅里叶分析 [128]等多个领域存在联系，因此在学界引起了极大关注。近期一系列令人振奋的研究成果 [65, 24, 222]表明，一种与唯一博弈密切相关、称为 2对2博弈的问题是 NP 难的。这是证明 UGC 的重要一步，也为 UGC 成立提供了证据。

在算法方面，人们曾尝试通过多种方式（例如参见 [226, 43, 9]）证伪 UGC。特别是，Barak等人 [22] 证明，度为 8 的平方和方法能够高效求解此前针对线性规划和半正定规划提出的唯一博弈问题整数间隙实例。这项工作引起了学界的浓厚兴趣，因为它表明平方和方法或许可以用来证伪 UGC。

因此，理解平方和在唯一博弈问题上的表现至关重要。一个很好的第一步，是理解平方和在最大割问题上的表现；最大割是唯一博弈问题的一个特例。事实上，我们还可以进一步明确问题，探究平方和在以下问题上的表现：

随机图上的最大割，更准确地说，是猜想 8.2.7。 [170, 165] 中给出了 2 度和 4 阶的下界，而将他们的分析推广到高阶 SoS，是一项虽非易事但十分重要的开放问题。

8.3 低次似然比假设

如第3章所述，对于足够良好的分布，低次似然比假设能够从解析角度预测有界时间假设检验中的计算障碍。更多详情请参见 [110, 146, 98]及其中的参考文献。由于完整证明这一假设很可能比证明例如 $P \neq NP$更困难，因此超出了当前技术所能达到的范围。尽管如此，在受限证明系统中证实这一假设，仍是一个引人入胜且重要的未来研究方向。具体而言，基于第3章的记号，我们希望证明：对于足够良好的分布 ν， μ，经过伪校准后，$\widetilde{\text{如}}$果 $\mathbb{E}[1] = 1 + o(1)$，那么就存在一个平方和下界。事实上，在本研究中，我们对若干基本问题证实了这一点。一般性地证明这一结论，将大大增进我们对有界时间算法之幂的理解。

8.4 技术改进

在概述了未来研究的一些总体方向后，我们现在具体说明几个改进研究结果中某些技术方面的方向。

改进参数依赖关系

在我们的许多下界中，都要求傅里叶系数呈多项式衰减。例如，对于每个新的傅里叶特征，我们要求其衰减为 n^{ε}，其中 n 是输入规模。这样做是为了处理在进行收费论证时范数界中出现的各种其他因子。在证明中，我们将这些称为顶点衰减或边衰减，分别对应于

它们在我们使用的图矩阵论证中的编码方式。通过这种做法，我们得到的下界会稍弱一些。例如，对于张量主成分分析，我们得到的是 $n^{1/4}$ 下界（忽略多对数因子），而不是 a $n^{1/4}$ 下界；对于任意 $\varepsilon > 0$，我们得到 $n^{1/4-\varepsilon}$ 下界。一般来说，尽管这类衰减有助于证明，但尚不清楚它是否确有必要；如何进行更精细的分析，以将已知上界与下界之间的差距缩小到一个多对数因子以内，仍是一个开放问题。

与上述讨论相关的另一个开放问题，是在我们的下界中提高 SoS 的度。例如，对于 Sherrington-Kirkpatrick 下界，将 SoS次数从 n^{ε} 提高到 Ω(n) 仍是一个开放问题。我们目前的技术无法处理这一问题，但我们预期下界仍然成立。

精确满足约束

在我们的一些下界中，植入分布仅近似满足诸如存在大小为 k 的子图、存在一个单位向量 u，以及 u 为 k-稀疏等约束。虽然我们希望使用能够精确满足此类约束的植入分布，但矩矩阵会变得难以分析得多。

对于 Sherrington-Kirkpatrick 下界，我们通过使用一种舍入技术 [86] 解决了这个问题。植入团的平方和下界中也出现了同样的问题 [20]，Pang 最近的一篇论文解决了这一问题 [180]。我们将其一般情形留待未来工作解决。

8.5 对计算机科学的影响

正如我们在引言中所看到的，理论计算机科学研究当前的状况似乎是要理解各种问题的计算极限。尽管可能存在诸如解决 P 与 NP 问题这样的终极目标，或诸如解决唯一博弈猜想这样的相对适度目标，但这一过程中仍有许多内容值得学习和发掘

。例如，对于各种问题，信息理论上可能实现的事情与计算上可行的事情之间似乎存在明显的差距。我们的工作进一步揭示了这一引人入胜的现象，即所谓的信息—计算权衡。然而，还有其他问题有待回答。例如，是什么使认证看起来比估计或恢复更困难？我们能否刻画出这样一种问题的精确性质：对于平方和等不同类别的算法而言，这些性质可能使问题变得困难或容易？尽管在这一总体探索过程中，许多丰富的结构正在逐渐被揭示，但我们仍未能真正理解它。尽管如此，迄今为止我们所发现成果的应用，无论在技术上还是在哲学上，都已经遍及数学和科学的各个分支，因此，这一领域的研究并不只是出于单纯的好奇。我们希望我们的工作能够成为朝着这一宏伟目标迈进的有意义的一步。

参考文献

[1] Kwangjun Ahn、Dhruv Medarametla 和 Aaron Potechin。图矩阵：范数界及其应用。abs/1604.03423，2020年。[2] Genevera I Allen 和 Mirjana Maleti´c-Savati´c。用于代谢组学的稀疏非负广义主成分分析。*Bioinformatics*，27(21):3029–3035，2011年。[3]Moustafa Alzantot、Bharathan Balaji 和 Mani Srivastava。你听见了吗？针对自动语音识别的对抗样本。*arXiv preprintarXiv:1801.00554*，2018年。[4] Arash A Amini 和 Martin J Wainwright。稀疏主成分半正定松弛的高维分析。载于*2008*年 *IEEE* 信息论国际研讨会，第2454–2458页。IEEE，2008年。[5] Anima Anandkumar、Yuan Deng、Rong Ge 和 Hossein Mobahi。张量主成分分析的同伦分析。载于学习理论会议，第79–104页。PMLR，2017年。[6] Anima Anandkumar、Rong Ge 和 Majid Janzamin。张量幂方法动态分析：在学习过完备潜变量模型中的应用。*arXiv*预印本 *arXiv:1411.1488*，2014年。[7] Animashree Anandkumar、Rong Ge、Daniel Hsu、Sham M Kakade 和 Matus Telgar-sky。用于学习潜变量模型的张量分解。机器学习研究杂志，15:2773–2832，2014年。[8] Animashree Anandkumar、Rong Ge 和 Majid Janzamin。通过交替秩一更新实现有保证的非正交张量分解。*arXiv*预印本 *arXiv:1402.5180*，2014年。[9] Sanjeev Arora、Boaz Barak 和 David Steurer。用于唯一博弈及相关问题的亚指数算法。美国计算机协会会刊 *(JACM)*，62(5):1–25，2015年。[10] Sanjeev Arora、Satish Rao 和 Umesh Vazirani。扩张流与最稀疏割的 $\sqrt{\log n}$-近似。载于第*36*届 *ACM*计算理论研讨会论文集，2004年。[11] Gerard Ben Arous、Reza Gheissari 和 Aukosh Jagannath。张量主成分分析的算法阈值。《概率年刊》，48(4):2052–2087，2020年。[12] Gerard Ben Arous、Song Mei、Andrea Montanari 和 Mihai Nica。尖峰张量模型的景观。纯粹与应用数学通讯，72(11):2282–2330，2019年。[13] G´erard Ben Arous、Alexander S Wein 和 Ilias Zadik。稀疏主成分分析中的自由能井与重叠间隙性质。载于学习理论会议，第479–482页。PMLR，2020年。

[14] Jinho Baik、G´erard Ben Arous、Sandrine P´ech´e 等。非零复样本协方差矩阵最大特征值的相变。《概率年刊》，33(5):1643–1697，2005年。[15] Ainesh Bakshi、Ilias Diakonikolas、He Jia、Daniel M Kane、Pravesh K Kothari 和 Santosh S Vempala。稳健学习任意高斯混合模型。*arXiv*预印本 *arXiv:2012.02119*，2020年。[16] Ainesh Bakshi 和 Pravesh Kothari。非球形混合分布的抗离群点聚类。*arXiv*预印本 *arXiv:2005.02970*，2020年。[17] Ainesh Bakshi 和 Pravesh K Kothari。列表可解码子空间恢复：多项式时间内实现维度无关误差。*arXiv*预印本 *arXiv:2002.05139*，2020年。[18] Ainesh Bakshi 和 Adarsh Prasad。稳健线性回归：多项式时间内的最优速率。载于第*53*届 *ACM SIGACT*计算理论研讨会论文集，第102–115页，2021年。[19] Afonso S. Bandeira、Dmitriy Kunisky 和 Alexander S. Wein。证明约束主成分分析问题中界的计算困难性，2019年。[20] B. Barak、S. B. Hopkins、J. Kelner、P. Kothari、A. Moitra 和 A. Potechin。植入团问题的近乎紧平方和下界。载于第*57*届 *IEEE*计算机科学基础研讨会论文集，第428–437页，2016年。[21] Boaz Barak、Fernando G. S. L. Brand˜ao、Aram Wettroth Harrow、Jonathan A. Kelner、David Steurer 和 Yuan Zhou。超收缩性、平方和证明及其应用。$CoRR$, abs/1205.4484，2012年。[22]Boaz Barak、Fernando GSL Brandao、Aram W Harrow、Jonathan Kelner、David Steurer 和 Yuan Zhou。超收缩性、平方和证明及其应用。载于第四十四届 *ACM*计算理论研讨会论文集，第307–326页，2012年。[23] Boaz Barak、Jonathan A Kelner 和 David Steurer。通过平方和方法进行字典学习与张量分解。载于第四十七届 *ACM*计算理论研讨会论文集，第143–151页，2015年。[24] Boaz Barak、Pravesh K Kothari 和 David Steurer。短码图中的小集合扩张与2对2猜想。*arXiv*预印本 *arXiv:1804.08662*，2018年。[25] Boaz Barak 和 David Steurer。平方和证明与追求最优算法。载于国际数学家大会，2014年。[26] Avraham Ben-Aroya、Oded Regev 和 Ronald De Wolf。矩阵值函数的超收缩不等式及其在量子计算和局部可解码码中的应用。载于第*49*届 *IEEE*计算机科学基础研讨会论文集，第477–486页。IEEE，2008年。

[27] Florent Benaych-Georges、Charles Bordenave 和 Antti Knowles。稀疏随机矩阵的谱半径。载于*Annales de l'Institut Henri Poincar´e, Probabilit´es et Statis-tiques*，第56卷，第2141–2161页。Institut Henri Poincar´e，2020年。[28] Quentin Berthet 和 Philippe Rigollet。稀疏主成分检测的复杂性理论下界。载于学习理论会议，第1046–1066页，2013年。[29] Quentin Berthet、Philippe Rigollet 等。高维稀疏主成分的最优检测。《统计年鉴》，41(4):1780–1815，2013年。[30] Aditya Bhaskara、Moses Charikar、Eden Chlamtac、Uriel Feige 和 Aravindan Vi- jayaraghavan。检测高对数密度：针对最稠密 k-子图的 o (n 1/4) 近似算法。载于第四十二届*ACM*计算理论研讨会论文集，第201–210页，2010年。[31] Aditya Bhaskara、Moses Charikar、Venkatesan Guruswami、Aravindan Vijayaraghavan 和 Yuan Zhou。最稠密 k-子图强半定规划松弛的多项式整数性间隙。载于第二十三届*ACM-SIAM*离散算法年度研讨会论文集，第388–405页。SIAM，2012年。[32] Vijay Bhattiprolu、Venkatesan Guruswami 和 Euiwoong Lee。球面上随机张量最大值的平方和证书。*arXiv*预印本 *arXiv:1605.00903*，2016年。[33] K-H Borgwardt。单纯形法所需的平均枢轴步数为多项式级。*Zeitschrift f¨ur Operations Research*，26(1):157–177，1982年。[34] Karl Heinx Borgwardt。单纯形法的概率分析。载于*DGOR/NSOR*，第564–575页。施普林格，1988年。[35] St´ephane Boucheron、Olivier Bousquet、G´abor Lugosi 和 Pascal Massart。独立随机变量函数的矩不等式。《概率年刊》，33(2):514–560，2005年。[36] Mark Braverman、Young Kun Ko、Aviad Rubinstein 和 Omri Weinstein。具有完美完备性的最稠密-k-子图问题的ETH困难性。载于第二十八届*ACM-SIAM*离散算法年度研讨会论文集，第1326–1341页。SIAM，2017年。[37] Matthew Brennan 和 Guy Bresler。稀疏主成分分析的最优平均情形归约：从弱假设到强困难性。*arXiv*预印本 *arXiv:1902.07380*，2019年。[38] Matthew Brennan 和 Guy Bresler。秘密泄漏导致的可归约性与统计-计算间隙。载于学习理论会议，第648–847页。PMLR，2020年。

[39] Matthew Brennan、Guy Bresler、Samuel B Hopkins、Jerry Li 和 Tselil Schramm。统计查询算法与低次测试几乎等价。*arXiv*预印本 *arXiv:2009.06107*，2020年。[40] Matthew Brennan、Guy Bresler 和 Wasim Huleihel。植入稀疏结构问题的可归约性与计算下界。载于学习理论会议，第48–166页。PMLR，2018年。[41] Matthew Brennan、Guy Bresler 和 Wasim Huleihel。子矩阵检测计算下界的普适性。载于学习理论会议，第417–468页。PMLR，2019年。[42] Jonah Brown-Cohen 和 Prasad Raghavendra。反驳随机约束满足问题的扩展形式下界。载于第十四届*ACM-SIAM*离散算法年度研讨会论文集，第305–324页。SIAM，2020年。[43] Moses Charikar、Konstantin Makarychev 和 Yury Makarychev。唯一博弈的近最优算法。载于第*38*届*ACM*计算理论研讨会论文集，2006年。[44] Sourav Chatterjee。基于可交换对的集中不等式。斯坦福大学，2005年。[45]Sourav Chatterjee。用于集中不等式的Stein方法。*arXiv*预印本 *math/0604352*，2006年。[46] Wei-Kuo Chen。具有Rademacher先验的尖峰随机张量中的相变。《统计年鉴》，47(5):2734–2756，2019年。[47] Wei Kuo Chen、Madeline Handschy 和 Gilad Lerman。多重尖峰随机张量中的相变。*Scanning Electron Microsc Meet at*，2018年。[48] Yudong Chen 和 Jiaming Xu。植入问题以及簇和子矩阵数量递增时子矩阵定位中的统计-计算权衡。*arXiv*预印本 *arXiv:1402.1267*，2014年。[49] Eden Chlamt´aˇc 和 Pasin Manurangsi。达到对数密度阈值的Sherali-Adams整数性间隙。*arXiv*预印本 *arXiv:1804.07842*，2018年。[50]Agniva Chowdhury、Petros Drineas、David P Woodruff 和 Samson Zhou。稀疏主成分分析的近似算法。*arXiv*预印本 *arXiv:2006.12748*，2020年。[51] Hyonho Chun 和 S¨und¨uz Kele¸s。使用多变量稀疏偏最小二乘回归进行表达数量性状基因座定位。《遗传学》，182(1):79–90，2009年。[52] Amin Coja-Oghlan。随机图的Lov´asz数。《组合学、概率与计算》，14(4):439–465，2005年。

[53] Amin Coja-Oghlan 和 Charilaos Efthymiou。随机图中的独立集。*Random Structures & Algorithms*，47(3):436–486，2015年。[54] Andrea Crisanti 和 Tommaso Rizzo。Sherrington-Kirkpatrick模型的 ∞副本对称性破缺解分析。*Physical Review E*，65(4):046137，2002年。[55] Varsha Dani 和 Cristopher Moore。采用加权二阶矩方法求解随机图中的独立集。载于近似、随机化与组合优化，第6845卷计算机科学讲义，第472–482页。施普林格，海德堡，2011年。[56] George Dantzig。线性规划及其扩展。普林斯顿大学出版社，2016年。[57] Alexandre d' Aspremont、Laurent Ghaoui、Michael Jordan 和 Gert Lanckriet。利用半定规划直接表述稀疏主成分分析。神经信息处理系统进展，17:41–48，2004年。[58] Amir Dembo、Andrea Montanari 和 Subhabrata Sen。稀疏随机图的极值割。《概率年刊》，45(2):1190–1217，2017年。[59] Yash Deshpande 和 Andrea Montanari。隐藏团和隐藏子矩阵问题的改进平方和下界。载于学习理论会议，第523–562页。PMLR，2015年。[60] Yash Deshpande 和 Andrea Montanari。通过协方差阈值化实现稀疏主成分分析。*TheJournal of Machine Learning Research*，17(1):4913–4953，2016年。[61] Ilias Diakonikolas、Daniel M Kane 和 Alistair Stewart。高维高斯分布和高斯混合分布鲁棒估计的统计查询下界。载于*2017 IEEE*第*58*届计算机科学基础年会（*FOCS*），第73–84页。IEEE，2017年。[62] Nat Dilokthanakul、Pedro AM Mediano、Marta Garnelo、Matthew CH Lee、Hugh Salimbeni、Kai Arulkumaran 和 Murray Shanahan。基于高斯混合变分自编码器的深度无监督聚类。*arXiv*预印本 *arXiv:1611.02648*，2016年。[63] Jian Ding、Allan Sly 和 Nike Sun。随机正则图上的最大独立集。*Acta Math.*，217(2):263–340，2016年。[64]Yunzi Ding、Dmitriy Kunisky、Alexander S Wein 和 Afonso S Bandeira。稀疏主成分分析的亚指数时间算法。*arXiv*预印本 *arXiv:1907.11635*，2019年。[65] Irit Dinur、Subhash Khot、Guy Kindler、Dor Minzer 和 Muli Safra。朝着证明2对1游戏猜想迈进？载于第*50*届*ACM SIGACT*计算理论研讨会论文集，第376–389页，2018年。

[66] Tommaso d'Orsi、Pravesh K Kothari、Gleb Novikov 和 David Steurer。稀疏主成分分析：算法、对抗扰动与证书。载于*2020 IEEE*第*61*届计算机科学基础年会（*FOCS*），第553–564页。IEEE，2020年。[67] Tommaso d'Orsi、Pravesh K. Kothari、Gleb Novikov 和 David Steurer。稀疏主成分分析：算法、对抗扰动与证书。载于*2020 IEEE*第*61*届计算机科学基础年会（*FOCS*），2020年。[68] Olivier Duchenne、Francis Bach、In-So Kweon 和 Jean Ponce。一种基于张量的高阶图匹配算法。*IEEE*模式分析与机器智能汇刊，33(12):2383–2395，2011年。[69] Rishabh Dudeja 和 Daniel Hsu。张量主成分分析的统计查询下界。机器学习研究杂志，22(83):1–51，2021年。[70] G´abor Elek 和 G´abor Lippner。博雷尔预言机：一种分析方法，用于构造常数时间算法。美国数学会会刊，138(8):2939–2947，2010年。[71] Zhou Fan 和 Andrea Montanari。局部算法求解半定规划的效果如何？载于第*49*届*ACM SIGACT*计算理论年会论文集，第604–614页，2017年。[72] Uriel Feige 和 Robert Krauthgamer。在半随机图中寻找并证明一个大型隐藏团。*Random Structures & Algorithms*，16(2):195–208，2000年。[73] Uriel Feige、David Peleg 和 Guy Kortsarz。稠密k子图问题。*Algorithmica*，29(3):410–421，2001年。[74] Uriel Feige、Michael Seltser 等。关于最稠密 *k*-子图问题。Citeseer，1997年。[75] Vitaly Feldman。统计查询学习，第2090–2095页。施普林格纽约，纽约，NY，2016年。[76] Vitaly Feldman、Elena Grigorescu、Lev Reyzin、Santosh S Vempala 和 Ying Xiao。用于检测植入团的统计算法及其下界。美国计算机协会会刊（*JACM*），64(2):1–37，2017年。[77] Vitaly Feldman、Cristobal Guzman 和 Santosh Vempala。均值向量估计和随机凸优化的统计查询算法。运筹学数学，2021年。[78] Vitaly Feldman、Will Perkins 和 Santosh Vempala。具有植入解的随机可满足性问题的复杂性。*SIAM*计算杂志，47(4):1294–1338，2018年。

[79]Volker Fischer、Mummadi Chaithanya Kumar、Jan Hendrik Metzen 和 Thomas Brox。用于语义图像分割的对抗样本。*arXiv* 预印本 *arXiv:1703.01101*，2017年。[80] N. Fleming、P. Kothari 和 T. Pitassi。半代数证明与高效算法设计。2019年。[81] PJ Forrester、NC Snaith 和 JJM Verbaarschot。随机矩阵理论的发展。《物理学杂志 *A*：数学与通论》，36(12):R1，2003年。[82] Zolt´an F¨uredi 和 J´anos Koml´os。随机对称矩阵的特征值。*Combinatorica*，1(3):233–241，1981年。[83] David Gamarnik 和 Ilias Zadik。植入团问题的图景：稠密子图与重叠间隙性质。*arXiv* 预印本 *arXiv:1904.07174*，2019年。[84] Chao Gao、Zongming Ma 和 Harrison H Zhou。稀疏典型相关分析：自适应估计与计算障碍。《统计年鉴》，45(5):2074–2101，2017年。[85] Rong Ge 和 Tengyu Ma。使用平方和算法分解超完备三阶张量。*arXiv* 预印本 *arXiv:1504.05287*，2015年。[86] Mrinalkanti Ghosh、Fernando Granha Jeronimo、Chris Jones、Aaron Potechin 和 Goutham Rajendran。通过植入仿射平面证明 Sherrington-Kirkpatrick 模型的平方和下界。载于*2020*年 *IEEE* 第*61*届计算机科学基础年会 *(FOCS)*，第954–965页。IEEE，2020年。[87] Mrinalkanti Ghosh、Fernando Granha Jeronimo、Chris Jones、Aaron Potechin 和 Goutham Rajendran。通过植入仿射平面证明 Sherrington-Kirkpatrick 模型的平方和下界。载于*2020*年 *IEEE* 第*61*届计算机科学基础年会 *(FOCS)*，第954–965页。IEEE，2020年。[88] M. X. Goemans 和 D. P. Williamson。用于最大割和最大 2SAT 的 .878-近似算法。载于第*26*届 *ACM*计算理论研讨会论文集，第422–431页，1994年。[89] M.X. Goemans 和 D.P. Williamson。使用半定规划改进最大割和可满足性问题的近似算法。美国计算机协会会刊，42(6):1115–1145，1995年。初步版本发表于*STOC'94* 论文集。[90] Ian J Goodfellow、Jonathon Shlens 和 Christian Szegedy。解释并利用对抗样本。*arXiv* 预印本 *arXiv:1412.6572*，2014年。[91] Dima Grigoriev。背包问题正定性证明的复杂度。计算复杂性，10(2):139–154，2001年。[92] Dima Grigoriev。奇偶性正定理演算证明的次数线性下界。《理论计算机科学》，259(1-2):613–622，2001年。

[93] Francesco Guerra。平均场自旋玻璃模型中的破缺复制对称性界。《数学物理通讯》，233(1):1–12，2003年。[94] Venkatesan Guruswami 和 Ali Kemal Sinop。Lasserre 层级、高阶特征值，以及用于图划分和具有半正定目标函数的二次整数规划的近似方案。载于*FOCS*，第482–491页，2011年。[95] Bruce Hajek、Yihong Wu 和 Jiaming Xu。随机图社群检测的计算下界。载于学习理论会议，第899–928页。PMLR，2015年。[96] Frank Hansen 和 Gert K Pedersen。詹森算子不等式。《伦敦数学学会通报》，35(4):553–564，2003年。[97] Dan Hendrycks、Kevin Zhao、Steven Basart、Jacob Steinhardt 和 Dawn Song。自然对抗样本。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第15262–15271页，2021年。[98] Justin Holmgren 和 Alexander S Wein。低次猜想的反例。*arXiv* 预印本 *arXiv:2004.08454*，2020年。[99] Shlomo Hoory、Nathan Linial 和 Avi Wigderson。扩展图及其应用。《美国数学会通报》，43(04):439–562，2006年8月。[100] Samuel Hopkins。统计推断与平方和方法。博士论文，康奈尔大学，2018年。[101] Samuel B Hopkins。多项式时间内实现次高斯速率的均值估计。《统计年鉴》，48(2):1193–1213，2020年。[102] Samuel B Hopkins、Pravesh Kothari、Aaron Henry Potechin、Prasad Raghavendra 和 Tselil Schramm。植入团问题中四次平方和的整数间隙。《*ACM*算法汇刊》*(TALG)*，14(3):1–31，2018年。[103] Samuel B Hopkins、Pravesh K Kothari 和 Aaron Potechin。平方和与植入团：对所有度数下 MPW 矩的紧致分析及四次时的最优下界。*arXiv* 预印本 *arXiv:1507.05230*，2015年。[104] Samuel B Hopkins、Pravesh K Kothari、Aaron Potechin、Prasad Raghavendra、Tselil Schramm 和 David Steurer。平方和方法在检测隐藏结构方面的威力。载于第*58*届 *IEEE* 计算机科学基础研讨会论文集，第720–731页。IEEE，2017年。[105] Samuel B Hopkins、Tselil Schramm 和 Jonathan Shi。用于过完备张量分解的鲁棒谱算法。载于学习理论会议，第1683–1722页。PMLR，2019年。

[106] Samuel B Hopkins、Tselil Schramm、Jonathan Shi 和 David Steurer。基于平方和证明的快速谱算法：张量分解与植入稀疏向量。载于第四十八届年度 *ACM*计算理论研讨会论文集，第178–191页，2016年。[107] Samuel B Hopkins、Jonathan Shi 和 David Steurer。通过平方和证明进行张量主成分分析。载于学习理论会议，第956–1006页。PMLR，2015年。[108] Samuel B Hopkins、Jonathan Shi 和 David Steurer。通过平方和证明进行张量主成分分析。载于学习理论会议，第956–1006页，2015年。[109] Samuel B Hopkins 和 David Steurer。利用少量样本进行高效贝叶斯估计：社区检测及相关问题。载于*2017*年 *IEEE* 第*58*届计算机科学基础年度研讨会（*FOCS*），第379–390页。IEEE，2017年。[110] Samuel Brink Klevit Hopkins。统计推断与平方和方法。2018年。[111] Carlos Hoppen 和 Nicholas Wormald。局部算法、大围长正则图与随机正则图。*Combinatorica*，38(3)：619–664，2018年。[112] Daniel Hsu、Sham M Kakade 和 Tong Zhang。用于学习隐马尔可夫模型的谱算法。计算机与系统科学杂志，78(5)：1460–1480，2012年。[113] Wei-Ning Hsu、Anuroop Srira m、Alexei Baevski、Tatiana Likhomanenko、Qiantong Xu、Vineel Pratap、Jacob Kahn、Ann Lee、Ronan Collobert、Gabriel Synnaeve 等。稳健的 wav2vec 2.0：分析自监督预训练中的领域偏移。*arXiv*预印本 *arXiv:2104.01027*，2021年。[114] De Huang 和 Joel A. Tropp。从 Poincar´é 不等式到非线性矩阵集中。*Bernoulli*，27(3)：1724–1744，2021年。[115] Aukosh Jagannath、Patrick Lopatto 和 Leo Miolane。张量主成分分析的统计阈值。应用概率年刊，30(4)：1910–1933，2020年。[116] Zhuxi Jiang、Yin Zheng、Huachun Tan、Bangsheng Tang 和 Hanning Zhou。变分深度嵌入：一种无监督生成式聚类方法。*arXiv*预印本 *arXiv:1611.05148*，2016年。[117] Iain M Johnstone 和 Arthur Yu Lu。稀疏主成分分析。*arXiv*预印本 *arXiv:0901.4392*，2009年。[118] Ian T Joliffe 和 BJT Morgan。主成分分析与探索性因子分析。医学研究中的统计方法，1(1)：69–95，1992年。

[119] Chris Jones、Aaron Potechin、Goutham Rajendran、Madhur Tulsiani 和 Jeff Xu。稀疏独立集的平方和下界。*IEEE第62届计算机科学基础年度研讨会（FOCS）*，2021年。[120] Sushrut Karmalkar、Adam R Klivans 和 Pravesh K Kothari。列表可解码线性回归。*arXiv预印本 arXiv:1905.05679*，2019年。[121] Richard M Karp。组合问题之间的可归约性。载于《计算机计算复杂性》，第85–103页。施普林格，1972年。[122] R.M. Karp。组合问题之间的可归约性。载于 R.E. Miller 和 J.W. Thatcher 编，《计算机计算复杂性》，第85–103页。普莱enum出版社，1972年。[123] Michael Kearns。基于统计查询进行高效抗噪学习。美国计算机协会会刊（*JACM*），45(6)：983–1006，1998年。[124] Ilyes Khemakhem、Diederik Kingma、Ricardo Monti 和 Aapo Hyvarinen。变分自编码器与非线性独立成分分析：一个统一框架。载于国际人工智能与统计会议，第2207–2217页。PMLR，2020年。[125] Subhash Khot。论唯一性双证明者单轮博弈的幂。载于第*34*届 *ACM*计算理论研讨会论文集，第767–775页，2002年。[126] Subhash Khot。排除图最小二等分、稠密k子图和二分团的多项式时间近似方案。*SIAM*计算杂志，36(4)：1025–1071，2006年。[127] Subhash Khot、Guy Kindler、Elchanan Mossel 和 Ryan O'Donnell。最大割及其他二变量约束满足问题的最优不可近似性结果？载于第*45*届 *IEEE*计算机科学基础研讨会论文集，第146–154页，2004年。[128] Subhash Khot 和 Oded Regev。顶点覆盖可能难以近似到$2-\varepsilon$以内。载于第*18*届 *IEEE*计算复杂性会议论文集，2003年。[129] Subhash Khot 和 Nisheeth Vishnoi。唯一博弈猜想、割问题的整数间隙，以及负型度量嵌入 ℓ_1的性质。载于第*46*届 *IEEE*计算机科学基础研讨会论文集，第53–63页，2005年。[130] Chiheon Kim、Afonso S Bandeira 和 Michel X Goemans。超图中的社区检测、尖峰张量模型与平方和。载于*2017*年国际抽样理论与应用会议（*SampTA*），第124–128页。IEEE，2017年。[131] Diederik P Kingma 和 Max Welling。自编码变分贝叶斯。*arXiv预印本 arXiv:1312.6114*，2013年。[132] Bohdan Kivva、Goutham Rajendran、Pradeep Ravikumar 和 Bryon Aragam。通过混合预言机学习潜在因果图。神经信息处理系统进展，34，2021年。

[133] Bohdan Kivva、Goutham Rajendran、Pradeep Ravikumar 和 Bryon Aragam。无辅助信息且采用混合先验的深度生成模型的可识别性。*arXiv* 预印本 *arXiv:2206.10044*，2022年。[134] Bohdan Kivva、Goutham Rajendran、Pradeep Kumar Ravikumar 和 Bryon Aragam。无辅助信息的深度生成模型的可识别性。载于神经信息处理系统进展，第35卷，2022。[135] Victor Klee 和 George J Minty。单纯形算法的效果究竟如何？《不等式》，3(3):159–175，1972年。[136] Adam Klivans、Pravesh K Kothari 和 Raghu Meka。用于离群值鲁棒回归的高效算法。载于学习理论会议，第1420–1430页。PMLR，2018年。[137] Pravesh Kothari、Ryuhei Mori、Ryan O'Donnell 和 David Witmer。用于反驳任意约束满足问题（CSP）的平方和下界。载于第*49*届*ACM*计算理论研讨会论文集，2017年。[138] Pravesh K Kothari 和 Peter Manohar。用于着色问题的无压力平方和下界。*arXiv* 预印本 *arXiv:2105.07517*，2021年。[139] Pravesh K Kothari 和 Ruta Mehta。平方和与纳什均衡相遇：求解任意均衡的下界。载于第*50*届*ACM SIGACT*计算理论年会论文集，第1241–1248页，2018年。[140] Pravesh K Kothari 和 David Steurer。基于平方和的鲁棒离群值矩估计。*arXiv* 预印本 *arXiv:1711.11581*，2017年。[141] Robert Krauthgamer、Boaz Nadler、Dan Vilenchik 等。半定松弛能否将稀疏主成分分析求解至信息极限？《统计年鉴》，43(3):1300–1322，2015年。[142]Jean-Louis Krivine。预序环。数学分析杂志 ´，12(1):307–326，1964年。[143] Dmitriy Kunisky。超立方体上平方和伪矩的保正扩展。*arXiv* 预印本 *arXiv:2009.07269*，2020年。[144] Dmitriy Kunisky。认证问题中的谱障碍。博士论文，纽约大学，2021年。[145] Dmitriy Kunisky 和 Afonso S. Bandeira。Sherrington-Kirkpatrick 哈密顿量的紧致 4 阶平方和下界。abs/1907.11686，2019年。[146] Dmitriy Kunisky、Alexander S Wein 和 Afonso S Bandeira。关于假设检验计算困难性的札记：基于低次似然比的预测。*arXiv* 预印本 *arXiv:1907.11636*，2019年。

[147] Alexey Kurakin、Ian Goodfellow 和 Samy Bengio。大规模对抗机器学习。*arXiv* 预印本 *arXiv:1611.01236*，2016年。[148] Jean B Lasserre。利用多项式和矩问题进行全局优化。*SIAM* 优化期刊，11(3):796–817，2001年。[149] Massimo Lauria。平方和与整数规划松弛。瑞典皇家理工学院课程。[150] James R Lee、Prasad Raghavendra 和 David Steurer。半定规划松弛规模的下界。载于第四十七届*ACM*计算理论研讨会论文集，第567–576页，2015年。[151] Thibault Lesieur、L´eo Miolane、Marc Lelarge、Florent Krzakala 和 Lenka Zdeborov´a。尖峰张量估计中的统计与计算相变。载于*2017*年 *IEEE* 信息论国际研讨会（*ISIT*），第511–515页。IEEE，2017年。[152] Jinyu Li、Li Deng、Reinhold Haeb-Umbach 和 Yifan Gong。鲁棒自动语音识别：通往实际应用的桥梁。2015年。[153] Nan Li 和 Baoxin Li。用于高光谱图像机载压缩的张量补全。载于*2010*年 *IEEE* 图像处理国际会议，第517–520页。IEEE，2010年。[154] Allen Liu 和 Ankur Moitra。解决高斯混合的鲁棒可学习性问题。载于第*53*届*ACM SIGACT*计算理论年会论文集，第518–531页，2021年。[155] L. Lov´asz 和 A. Schrijver。矩阵锥、集合函数与 0-1 优化。*SIAM* 优化期刊，1(12):166–190，1991年。[156] Tengyu Ma 和 Avi Wigderson。稀疏主成分分析的平方和下界。载于神经信息处理系统进展，第1612–1620页，2015年。[157] Zongming Ma。稀疏主成分分析与迭代阈值化。《统计年鉴》，41(2):772–801，2013年。[158] Michael W Mahoney 和 Petros Drineas。用于改进数据分析的 CUR 矩阵分解。美国国家科学院院刊，106(3):697–702，2009年。[159] Angshul Majumdar。曲波域中的稀疏主成分分析编码图像压缩。信号、图像与视频处理，3(1):27–34，2009年。[160] Pasin Manurangsi。逼近最稠密 k-子图问题的近多项式比率的 ETH 困难性。载于第*49*届*ACM SIGACT*计算理论年会论文集，第954–961页，2017年。

[161] Cheng Mao 和 Alexander S Wein。植入子空间中向量的最优谱恢复。*arXiv* 预印本 *arXiv:2105.15081*，2021年。[162] David W Matula。随机图中的最大团大小。南卫理公会大学计算机科学系，得克萨斯州达拉斯……，1976年。[163] Dhruv Medarametla 和 Aaron Potechin。均匀低阶图矩阵的范数界。*RANDOM*，2016年。[164] Raghu Meka、Aaron Potechin 和 Avi Wigderson。植入团的平方和下界。载于第四十七届 *ACM* 计算理论研讨会论文集，第 87–96 页，2015年。[165] Sidhanth Mohanty、Prasad Raghavendra 和 Jeff Xu。平方和下界提升：从 2 阶到 4 阶。载于第 *52* 届 *ACM SIGACT* 计算理论研讨会论文集，第 840–853 页，2020 年。[166] Ankur Moitra。理论计算机科学中的平方和。平方和：理论与应用，77:83，2020年。[167] Ankur Moitra 和 Alexander S Wein。张量网络中的谱方法。载于第 *51* 届 *ACM SIGACT* 计算理论研讨会论文集，第 926–937 页，2019年。[168] A. Montanari。Sherrington-Kirkpatrick 哈密顿量的优化。载于第 *60* 届 *IEEE*计算机科学基础研讨会论文集，第 1417–1433 页，2019年。[169] Andrea Montanari、Daniel Reichman 和 Ofer Zeitouni。谱方法的局限性：从高斯隐藏团问题到高斯张量的秩一扰动。神经信息处理系统进展，28，2015年。[170] Andrea Montanari 和 Subhabrata Sen。稀疏随机图上的半定规划及其在社区检测中的应用。载于第 *48* 届 *ACM* 计算理论研讨会论文集，第 814–827 页，2016年。[171] Nikhil Naikal、Allen Y Yang 和 S Shankar Sastry。通过稀疏主成分分析进行目标识别的信息性特征选择。载于*2011*年国际计算机视觉会议，第 818–825 页。IEEE，2011年。[172]Paarth Neekhara、Shehzeen Hussain、Prakhar Pandey、Shlomo Dubnov、Julian McAuley 和 Farinaz Koushanfar。语音识别系统的通用对抗扰动。*arXiv* 预印本 *arXiv:1905.03828*，2019年。[173] Yurii Nesterov。平方函数系统与优化问题。载于高性能优化，第 405–440 页。施普林格，2000年。

[174] R. O'Donnell。布尔函数分析中的若干主题。载于$STOC$，第 569–578 页，2008年。[175] Ryan O'Donnell。布尔函数分析。剑桥大学出版社，2014年。[176] Ryan O'Donnell。即使是近似意义下，平方和也显然无法自动化。载于第八届理论计算机科学创新会议（*ITCS 2017*）。Schloss Dagstuhl-莱布尼茨信息学中心，2017年。[177] Raphael Olivier 和 Bhiksha Raj。对抗攻击下自动语音识别模型的最新改进。*arXiv* 预印本 *arXiv:2203.16536*，2022。[178] Sean O'Rourke、Van Vu 和 Ke Wang。随机矩阵的特征向量。组合理论杂志 *A* 辑，144(C):361–442，2016年11月。[179]Dmitry Panchenko。混合 p自旋模型的帕里西公式。概率年刊，42(3):946–958，2014年5月。[180] Shuo Pang。精确植入团的平方和下界。载于第 *36* 届计算复杂性会议，第 200 卷，*LIPIcs.* 莱布尼茨信息学国际会议论文集，第 26 篇文章，63 页。Schloss Dagstuhl。莱布尼茨信息学中心，瓦德恩，2021年。[181] C.H. Papadimitriou 和 K. Steiglitz。组合优化：算法与复杂性。普伦蒂斯-霍尔，1982年。[182] Dimitris Papailiopoulos、Alexandros Dimakis 和 Stavros Korokythakis。通过低秩近似进行稀疏主成分分析。载于国际机器学习会议，第 747–755 页。PMLR，2013年。[183] G. Parisi。自旋玻璃的无穷多个序参量。物理评论快报，43:1754–1756，1979年12月。[184] Giorgio Parisi。自旋玻璃 SK 模型的一系列近似解。《物理学杂志 *A*：数学与通论》，13(4):L115，1980年。[185] Pablo A Parrilo。鲁棒性与优化中的结构化半定规划和半代数几何方法。博士论文，加州理工学院，2000年。[186] Pablo A Parrilo。半代数问题的半定规划松弛。《数学规划》，96(2):293–320，2003年。[187] Debashis Paul。高维尖峰协方差模型中样本特征结构的渐近性质。中华统计学志，第 1617–1642 页，2007年。[188] Daniel Paulin、Lester Mackey 和 Joel A. Tropp。随机矩阵的 Efron–Stein 不等式。概率年刊，44(5):3431–3473，2016年9月。[189] Amelia Perry、Alexander S Wein 和 Afonso S Bandeira。尖峰张量模型的统计极限。*arXiv* 预印本 *arXiv:1612.07728*，2016年。

[190]D´enes Petz. 若干迹不等式综述。*Banach Center Publications*，30(1):287–298，1994年。[191] Aaron Potechin 和 Goutham Rajendran。用于证明认证问题平方和下界的技术工具。*arXiv*预印本 *arXiv:2011.04253*，2020年。[192] Aaron Potechin 和 Goutham Rajendran。主成分分析的亚指数时间平方和下界。载于神经信息处理系统进展，2022。[193] Aaron Potechin 和 David Steurer。利用平方和进行精确张量补全。*arXiv*预印本 *arXiv:1702.06237*，2017年。[194] Mihai Putinar。紧半代数集上的正多项式。*Indiana University Mathematics Journal*，42(3):969–984，1993。[195] Prasad Raghavendra。针对每个约束满足问题，是否都有最优算法和不可近似性结果？载于第*40*届*ACM*计算理论研讨会论文集，第245–254页，2008年。[196] Prasad Raghavendra、Satish Rao 和 Tselil Schramm。在谱阈值以下强力反驳随机约束满足问题。载于第*49*届*ACM SIGACT*计算理论年会论文集，第121–131页，2017年。[197] Prasad Raghavendra 和 Tselil Schramm。4阶平方和程序中植入团的紧下界。*arXiv*预印本 *arXiv:1507.05136*，2015年。[198] Prasad Raghavendra 和 Benjamin Weitz。平方和证明的比特复杂度。载于第*44*届自动机、语言与程序设计国际研讨会论文集。Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik，2017年。[199] Goutham Rajendran。通过平方和层级进行组合优化。*arXiv*预印本 *arXiv:2208.04374*，2022。[200] Goutham Rajendran、Bohdan Kivva、Ming Gao 和 Bryon Aragam。多项式时间内的结构学习：贪心算法、Bregman信息与指数族。神经信息处理系统进展，34，2021年。[201] Goutham Rajendran 和 Madhur Tulsiani。利用 Efron-Stein 不等式分析多项式随机矩阵的集中性。载于*2023*年*ACM-SIAM*离散算法研讨会（*SODA*）论文集，第3614–3653页。SIAM，2023年。[202] Goutham Rajendran 和 Wei Zou。分析端到端神经模型在自动语音识别中的鲁棒性。*arXiv*预印本 *arXiv:2208.08509*，2022。[203] Mirco Ravanelli、Jianyuan Zhong、Santiago Pascual、Pawel Swietojanski、Joao Monteiro、Jan Trmal 和 Yoshua Bengio。用于鲁棒语音识别的多任务自监督学习。载于*ICASSP 2020-2020 IEEE*声学、语音与信号处理国际会议（*ICASSP*），第6989–6993页。IEEE，2020年。

[204] Danilo Jimenez Rezende、Shakir Mohamed 和 Daan Wierstra。深度生成模型中的随机反向传播与近似推断。载于国际机器学习会议，第1278–1286页。PMLR，2014年。[205] Bruce Reznick。希尔伯特第十七问题的一些具体方面。*Contemporary Mathematics*，253:251–272，2000年。[206] Emile Richard 和 Andrea Montanari。张量主成分分析的统计模型。载于神经信息处理系统进展，第2897–2905页，2014年。[207] Benjamin Rossman。检测团的平均情形复杂度。博士论文，马萨诸塞理工学院，2010年。[208] Benjamin Rossman。随机图上 k 团的单调复杂度。*SIAM*计算杂志，43(1):256–279，2014年。[209] Gian-Carlo Rota 和 Timothy C. Wallstrom。随机积分：一种组合方法。概率年刊，25(3):1257–1283，1997年。[210] Grant Schoenebeck。某些 k-CSP 的线性层级 Lasserre 下界。载于第*49*届*IEEE*计算机科学基础研讨会论文集，2008年。[211] Tselil Schramm 和 David Steurer。快速且鲁棒的张量分解及其在字典学习中的应用。载于学习理论会议，第1760–1793页。PMLR，2017年。[212] Warren Schudy 和 Maxim Sviridenko。独立随机变量多项式的 Bernstein 型集中性与矩不等式：多线性情形。*arXiv*预印本 *arXiv:1109.5193*，2011年。[213] Nicu Sebe 和 Michael S Lew。鲁棒计算机视觉：理论与应用，第26卷。施普林格科学与商业媒体，2013年。[214] Hanif D. Sherali 和 Warren P. Adams。零一规划问题中介于连续表示与凸包表示之间的松弛层级。*SIAM*离散数学杂志，3(3):411–430，1990年。[215] David Sherrington 和 Scott Kirkpatrick。可解的自旋玻璃模型。物理评论快报，35:1792–1796，1975年12月。[216] Naum Zuselevich Shor。求解多项式数学规划问题全局极值的一种方法。《控制论》，23(5):695–700，1987年。[217] Steve Smale。线性规划单纯形法的平均步数。《数学规划》，27(3):241–262，1983年。[218] Daniel A Spielman 和 Shang-Hua Teng。算法的平滑分析：为什么单纯形算法通常能在多项式时间内完成。美国计算机协会会刊（*JACM*），51(3):385–463，2004年。

[219] Charles Stein. 相依随机变量之和的分布正态近似误差界。载于第六届伯克利数学统计与概率研讨会论文集，第*2*卷：概率论，第583–602页。加利福尼亚大学出版社，1972年。[220] Charles Stein. 期望的近似计算。IMS，1986年。[221] Gilbert Stengle. 半代数几何中的零点定理与正定性定理。《数学年刊》，207(2):87–97，1974年。[222] Khot Subhash、Dor Minzer 和 Muli Safra. 格拉斯曼图中的伪随机集合具有近乎完美的展开性。载于*2018*年*IEEE*第*59*届计算机科学基础年度研讨会（*FOCS*），第592–601页。IEEE，2018年。[223] Christian Szegedy、Wojciech Zaremba、Ilya Sutskever、Joan Bruna、Dumitru Erhan、Ian Goodfellow 和 Rob Fergus. 神经网络的引人注目特性。*arXiv*预印本 *arXiv:1312.6199*，2013年。[224] Michel Talagrand. 帕里西公式。《数学年刊》，第221–263页，2006年。[225] Kean Ming Tan、Ashley Petersen 和 Daniela Witten. RNA测序数据分类。载于下一代测序数据的统计分析，第219–246页。施普林格，2014年。[226] Luca Trevisan. 唯一博弈的近似算法。载于第*46*届*IEEE*计算机科学基础研讨会论文集，2005年。[227] Joel A Tropp. 矩阵集中不等式导论。《基础与趋势》® 机器学习，8(1-2):1–230，2015年。[228] Lieven Vandenberghe 和 Stephen Boyd. 半定规划。《*SIAM*评论》，38(1):49–95，1996年。[229] Van H Vu. 随机矩阵的谱范数。载于第三十七届*ACM*计算理论研讨会论文集，第423–430页，2005年。[230] Dong Wang、Huchuan Lu 和 Ming-Hsuan Yang. 使用稀疏原型进行在线目标跟踪。《*IEEE*图像处理汇刊》，22(1):314–325，2012年。[231] Tengyao Wang、Quentin Berthet、Richard J Samworth 等. 稀疏主成分估计中的统计与计算权衡。《统计年鉴》，44(5):1896–1930，2016年。[232]Yiming Wang、Jinyu Li、Heming Wang、Yao Qian、Chengyi Wang 和 Yu Wu. Wav2vec-switch：从原始-噪声语音对中进行对比学习，以实现鲁棒语音识别。载于*ICASSP 2022-2022 IEEE*国际声学、语音与信号处理会议（*ICASSP*），第7097–7101页。IEEE，2022年。

[233] Alexander S Wein、Ahmed El Alaoui 和 Cristopher Moore. Kikuchi层级与张量主成分分析。载于*2019*年*IEEE*第*60*届计算机科学基础年度研讨会（*FOCS*），第1446–1468页。IEEE，2019年。[234] Eugene P Wigner. 无限维边界矩阵的特征向量 I。载于《*Eugene Paul Wigner*文集》，第524–540页。施普林格，1993年。[235] Matthew Willetts 和 Brooks Paige. 我不需要 u：无需侧信息的可辨识非线性独立成分分析。*arXiv*预印本 *arXiv:2106.05238*，2021年。[236] Cihang Xie、Mingxing Tan、Boqing Gong、Jiang Wang、Alan L Yuille 和 Quoc V Le. 对抗样本提升图像识别能力。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第819–828页，2020年。[237] Cihang Xie、Yuxin Wu、Laurens van der Maaten、Alan L Yuille 和 Kaiming He. 通过特征去噪提升对抗鲁棒性。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第501–509页，2019年。[238] Ilias Zadik、Min Jae Song、Alexander S. Wein 和 Joan Bruna. 基于格的方法在聚类中胜过平方和，2021年。[239] Lenka Zdeborov´a 和 Stefan Boettcher. 关于随机正则图中最大割与二分宽度的猜想。《统计力学杂志：理论与实验》，2010(02):P02020，2010年。[240] Lenka Zdeborov´a 和 Florent Krzakala. 推断的统计物理：阈值与算法。《物理学进展》，65(5):453–552，2016年。[241] Anru Zhang 和 Dong Xia. 张量奇异值分解：统计与计算极限。《*IEEE*信息论汇刊》，64(11):7311–7338，2018年。[242] Qinqing Zheng 和 Ryota Tomioka. 通过子空间范数插值凸与非凸张量分解。《神经信息处理系统进展》，28，2015年。