

机器学习的算法问题

Ankur Moitra

© 草稿日期为 2014 年 3 月 30 日

目录

目录	i
前言	1
1 引言	3
2 非负矩阵分解	5
2.1 引言	5
2.2 代数算法	10
2.3 稳定性和可分离性	15
2.4 主题模型	19
3 张量方法	25
3.1 基础知识	25
3.2 扰动界限	30
3.3 系统发育树和隐马尔可夫模型	34
3.4 社区检测	40
3.5 混合模型的扩展	44
3.6 独立成分分析	50
4 稀疏恢复	53
4.1 基础知识	53
4.2 唯一性和不确定性原理	56
4.3 追踪算法	59

4.4 Prony方法	61
4.5 压缩感知	64
5 字典学习	71
5.1 背景	71
5.2 全秩字典	73
5.3 超完备字典	77
6 高斯混合模型	83
6.1 历史	83
6.2 基于聚类的算法	86
6.3 密度估计的讨论	89
6.4 无聚类算法	91
6.5 一元算法	96
6.6 代数几何的视角	101
7 矩阵补全	105
7.1 背景	105
7.2 核范数	107
7.3 量子高尔夫	111
参考文献	115

前言

这本专著是基于2013年秋季在MIT教授的“18.S996: 机器学习的算法问题”课程。

感谢记录者Adam Hesterberg、Adrian Vladu、Matt Coudron、Jan-Christian Hutter、Henry Yuen、Yufei Zhao、Hilary Finucane、Matthew Johnson、Kayhan Barmanghelich、Gautam Kamath、GeorgeChen、Pratiksha Thaker、Mohammad Bavarian、Vlad Firoiu、Madalina Persu、CameronMusco、Christopher Musco、Jing Lin、Timothy Chu、Yin-Tat Lee、Josh Alman、Nathan Pinsker和Adam Bouland

。

第一章

介绍

这门课程将围绕机器学习中的算法问题进行组织。算法设计的通常范式是给出一个在所有可能的输入上都成功的算法，但困难在于现代机器学习中的几乎所有优化问题都是计算上难以处理的。然而，实践者使用各种各样的启发式方法在实践中取得了成功。然而，我们经常不理解这些方法何时以及为什么起作用（这是一个如果我们的算法具有可证明的保证就不会出现的问题）。这门课程的核心问题是：

问题1：机器学习中的哪些模型导致可处理的算法问题？

最坏情况分析是舒适的，因为如果一个算法在这个模型中起作用，那么它在实践中肯定也起作用。但是，机器学习系统每天“解决”的优化问题在最坏情况下确实很困难。然而，这些下界并不可怕；许多机器学习问题的困难实例实际上并不是我们想要在实践中解决的！我们将看到许多例子，选择正确的模型将使我们发现具有可证明保证的新算法，我们真正可以理解它们何时以及为什么起作用。在某些情况下，我们甚至能够分析实践者已经使用的方法，并对它们的行为给出新的见解。

问题2新模型 - 能更好地表示我们实际想要解决的实例 - 可以成为开发基本新算法的灵感吗？我们能理解广泛使用的启发式方法何时以及为什么有效吗？

本课程将重点关注

- (a) 非负矩阵分解
- (b) 主题建模
- (c) 张量分解
- (d) 稀疏恢复
- (e) 字典学习
- (f) 学习混合模型
- (g) 矩阵补全

希望随着时间的推移，本课程将添加更多的章节，因为在算法和机器学习的交叉领域还有很多主题需要探索。

第2章

非负矩阵分解

在本章中，我们将探讨非负矩阵分解问题。我们首先回顾一下这个问题的动机。接下来，我们提供了应用于学习主题模型参数的经典问题的新算法。

2.1 引言

为了理解为什么非负矩阵分解在应用中很有用，将其与奇异值分解进行比较将会有所帮助。在本章中，我们将重点关注这两者在文本分析中的应用。

奇异值分解

给定一个 $m \times n$ 矩阵 M ，它的奇异值分解是

$$M = U\Sigma V^T$$

其中 U 和 V 是正交的， Σ 是对角线矩阵，其元素非负。或者我们可以写成

$$M = \sum_{i=1}^r u_i \sigma_i v_i^T$$

其中 u_i 是 U 的第 i 列， v_i 是 V 的第 i 列， σ_i 是 Σ 的第 i 个对角线元素。

每个矩阵都有奇异值分解！实际上，这种表示法在理解线性算子的行为或者一般情况下非常有用。

用于从大型数据矩阵中提取重要的“特征”。我们将重点讨论奇异值分解的后者。这种分解的许多有用性质之一是我们可以立即从中读取对 M 的最佳低秩近似。

定义2.1.1 矩阵 M 的 Frobenius 范数 $\|M\|_F = \sqrt{\sum_{i,j} M_{i,j}^2}$. 或者, 如果 $M = \sum_{i=1}^r u_i \sigma_i v_i^T$, $\|M\|_F = \sqrt{\sum \sigma_i^2}$.

考虑以下优化问题: 设 B 是矩阵 M 在 Frobenius 范数下的最佳秩 k 近似 - 即 B 是所有秩不超过 k 的矩阵中 $\|M - B\|_F$ 的最小值。那么我们可以不失一般性地选择 B 为奇异值分解的前 k 项。

定理2.1.2 (Eckart-Young) 矩阵 M 在 Frobenius 范数下的最佳秩 k 近似是 $B = \sum_{i=1}^k u_i \sigma_i v_i^T$, 其误差为 $\|M - B\|_F = \sqrt{\sum_{i=k+1}^r \sigma_i^2}$.

这是奇异值分解如此广泛应用的原因之一: 如果我们以矩阵 M 的形式给出数据, 但我们相信数据近似低秩, 那么利用这种结构的自然方法是与 M 的最佳秩 k 近似一起工作。即使我们改变如何衡量 B 作为对 M 的近似的好坏的方法, 这个定理也是相当稳健的:

定义2.1.3 矩阵 M 的算子范数是 $\|M\|_2 = \max_{|v|=1} \|Mv\|_2$. 那么如果 $M = \sum_{i=1}^r u_i \sigma_i v_i^T$, $\|M\|_2 = \sigma_1$ (最大奇异值)。

在算子范数下, 最佳逼近 M 的近似值也可以通过 B 来达到 $\sum_{i=1}^k u_i \sigma_i v_i^T$, 此时误差为 $\|M - B\|_2 = \sigma_{k+1}$.

让我们给奇异值分解提供另一种解释。我们可以将 $m \times n$ 矩阵 M 看作是 \mathbb{R}^m 中的一组数据点。我们将这组点与分布 Δ 相关联, 该分布在随机选择点时均匀选择。

进一步假设该分布的期望为零。我们的数据处于高维空间中, 一个自然的问题是: 我们应该如何将数据投影到一维子空间中, 以尽可能地保留信息? 一个具体的目标是找到一个方向 u , 使得在 u 上投影 Δ 的方差最大 (在所有一维投影中)。这个问题导致了奇异向量的另一个特征描述:

$$u_1 = \operatorname{argmax} \frac{\|u^T M\|_2}{\|u\|_2}$$

而最大值为 σ_1 . 类似地, 如果我们想要在一个二维子空间上进行投影, 以最大化投影方差, 我们应该在 $\operatorname{span}(u_1, u_2)$ 上进行投影。

相关地

$$u_2 = \min_{u_1} \operatorname{argmax}_{u \perp u_1} \frac{\|u^T M\|_2}{\|u\|_2}$$

而最大值为 σ_2 . 这被称为奇异向量的变分特征。(在这里, 我们假设奇异值是不同的)。

有高效的算法可以计算奇异值分解。如果 $n = m$, 那么这些算法的运行时间为 $O(mn^2)$ 。基本思想是使用Householder反射将 M 化简为双对角形式, 然后使用QR算法从这个表示中计算奇异值分解。接下来我们将描述一个应用于文本分析的例子。

应用于文本分析

潜在语义索引: [49]

假设我们有一个大型文档集合, 并且我们想要提取出这个集合中的一些隐藏结构(无论是为了进行信息检索还是聚类)。通常的第一步是将数据收集到一个非常大且非常稀疏的矩阵中:

定义2.1.4 术语-文档矩阵 M 是一个 $m \times n$ 矩阵, 其中每一行代表一个单词, 每一列代表一个文档, 第 i 行第 j 列的元素是单词 i 在文档 j 中出现的次数。

显然, 我们丢失了一些信息, 因为这种表示方法没有考虑单词的顺序。然而, 矩阵更容易处理, 底层的假设是仍然可以通过知道每个文档包含哪些单词而不知道它们的顺序来对文档进行聚类。这通常被称为词袋假设。

潜在语义索引背后的思想是计算矩阵 M 的奇异值分解, 并将其用于信息检索和聚类。更加精确地说, 如果我们写成

$$M \approx U^{(k)} \Sigma^{(k)} V^{(k)T}$$

其中 $U^{(k)}$ 是 U 的前 k 列, 等等。那么 $U^{(k)}$ 的列是最大化随机文档的投影方差的 k 个方向。这些

向量被解释为“主题”。更加精确地说，假设我们想要计算文档 i 和文档 j 的“相似度”得分。我们可以通过计算

$$\langle M_i, M_j \rangle$$

其中 M_i 是 M 的第 i 列，等等。这个函数“计算”共同的单词数量。特别是，给定一个查询，我们可以通过计算每个文档中出现的查询词的数量来判断文档与查询的相似程度。这是相当幼稚的。相反，我们可以计算

$$\langle M_i^T U^{(k)}, M_j^T U^{(k)} \rangle$$

直观地，这将每个文档映射到一个长度为 k 的向量，该向量衡量文档中每个主题的存在程度，并通过在这个低维空间中进行内积来计算文档的相似性。在实践中，这是一种更好的相似性度量方式，并由Deerwester等人的开创性论文[49]引入。

然而，它也有自己的弱点。这种方法具有一些相当不理想的特性：

(a) “主题”是正交的

考虑“政治”和“金融”等主题。描述这些主题的词汇集是否不相关？不是的！

(b) “主题”包含负值

这更加微妙，但负面词汇可以用于表示文档与给定主题无关。但是当我们计算相似性时，两个文档基于一个它们都明确不相关的主题被判断为更相似。

这是另一个反直觉和不可取的特性。

非负矩阵分解

由[73]和[98]提出的想法是写

$$M \approx AW$$

其中 A 和 W 分别是 $m \times k$ 和 $k \times n$ 的矩阵，并且需要逐个元素为非负。事实上，让我们假设 M 的每列和为一。很容易看出，如果 D 是一个对角矩阵，其中第 i 个元素是 A 的第 i 列元素之和的倒数

则 $M = \widetilde{A} \widetilde{W}$ 其中 $\widetilde{A} = AD$ 和 $\widetilde{W} = D^{-1}W$ 对数据进行归一化, 使得 \widetilde{A} 和 \widetilde{W} 的每列和为一。因此, 我们正在寻找一组主题 (\widetilde{A} 的列, 每个列都是单词的分布), 以便每个文档都可以通过我们找到的主题的凸组合得到。

这个优化问题在许多机器学习系统中起着关键作用, 如图像分割、文本分析、推荐系统等。但是这个优化问题是 NP 难的[115]。那么现在我们应该怎么办? 放弃吗?

相比之下, 奇异值分解是一个理论和实践一致的问题! 它可以高效地计算, 并且有很多用途。尽管存在这个难解性的结果, 非负矩阵分解确实在实践中被使用。标准的方法是使用交替最小化:

交替最小化: 这个问题是非凸的, 但是假设我们猜测 A 。然后计算最小化 $\|M - AW\|_F$ 的非负 W 是凸的, 并且可以高效地解决。这种方法是猜测 A , 计算最佳 W , 然后将 W 固定, 并计算最佳 A , 依此类推。这个过程会收敛, 但不一定收敛到最优解。

在实践中, 它可能会陷入局部最小值!

我们注意到, 这种方法也被称为期望最大化[50], 并且是标准方法, 不仅适用于非负矩阵分解, 还适用于我们在本课程中研究的许多其他问题, 如字典学习和混合模型学习。

发人深思

但也许像这样的启发式方法正在识别出问题的有趣实例。本课程的目标是在面对棘手问题时不放弃, 并寻找新的解释。这些解释可以是新模型 (避免了问题的某些方面, 使我们能够嵌入困难问题) 或者可以确定已经使用的启发式方法在哪些条件下起作用。这是一个尚未被充分探索的领域。

在下一节中, 我们将询问如果限制主题数量会发生什么。由[115]生成的实例具有与 m 和 n 线性相关的 k , 但当我们寻找解释 300,000 篇《纽约时报》文章的一组主题时, 我们只需要几百个主题。因此, 重新制定问题的一种方式询问其复杂性作为 k 的函数。我们将使用代数技巧基本解决这个问题。然而, 如果我们想要更好的算法, 我们需要更多的东西。

假设。我们将看到，对这个问题的几何解释意味着这些困难实例是不稳定的，并且我们将研究一个条件（可分离性），它强制稳定性，并允许我们给出更好的算法 - 运行时间多项式地依赖于所有参数。

2.2 代数算法

在前一节中，我们介绍了非负矩阵分解问题，并描述了它在文本分析中的应用（它还有许多其他应用）。

Vavasis证明了这个问题在最坏情况下是 NP 难的，但他收缩的实例中的 k -主题数 - 与矩阵大小成线性关系[115]。

在大多数实际应用中， k 远小于 m 或 n ，并且考虑到这一点，我们将问：这个问题的复杂性如何作为 k 的函数？

我们将利用代数工具为任何 $k = O(1)$ 提供多项式时间算法。实际上，我们在这里提出的算法在依赖于 k 的方面几乎是最优的。

定义

让我们正式定义非负矩阵分解问题，因为在前一节中我们只是非正式地进行了定义：假设我们给出了一个大小为 $m \times n$ 的逐元素非负矩阵 M 。

定义 2.2.1 非负矩阵 M 的非负秩（用 $\text{rank}^+(M)$ 表示）是最小的正整数 k ，使得存在非负矩阵 A 和 W ，它们的大小分别为 $m \times k$ 和 $k \times n$ ，并且满足 $M = AW$ 。

等价地， $\text{rank}^+(M)$ 是最小的正整数 k ，使得存在非负秩为一的矩阵 $\{M_i\}$ ，满足 $M = \sum_i M_i$ 。

这两种等价的问题表述在我们的讨论中都很有用。为了对这个参数有一些了解，将其与一个更熟悉的参数进行比较是有帮助的：如果我们省略了 A 和 W 逐元素非负的要求，那么最小的 k 恰好是 M 的秩。因此，下面的关系是显然的：

事实 2.2.2 $\text{rank}^+(M) \geq \text{rank}(M)$

事实上，矩阵的秩和非负秩可能会有很大的不同：

例子。很容易看出 M 的列是由...所张成的

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ \vdots \\ n \end{bmatrix}, \begin{bmatrix} 1^2 \\ 2^2 \\ \vdots \\ n^2 \end{bmatrix} \right\}$$

很容易看出 $\text{rank}(M) = 3$ 。然而， M 的对角线上有零元素，而非对角线上有非零元素。此外，对于任何秩为一的非负矩阵 M_i ，它的零元素和非零元素的模式是一个组合矩形，即某些行和列的交集。通过一个标准的论证可以得出结论， $\text{rank}^+(M) = \Omega(\log n)$ 。

还有更大间隔的例子。

接下来，我们将非负矩阵分解与涉及多项式不等式系统的计算问题相连接。

多项式不等式系统

我们可以将寻找一个证明秩 $^+(M) \leq k$ 的 A 和 W 的问题重新表述为寻找一个特定多项式不等式系统的可行解的问题。更具体地说，我们要解决的问题是：

$$(2.1) \quad \begin{cases} M = AW \\ A \geq 0 \\ W \geq 0 \end{cases}$$

该系统由二次等式约束（每个 M 的每个元素一个约束）和要求 A 和 W 的每个元素非负的线性约束组成。在尝试设计更好的 $k = O(1)$ 算法之前，我们应该提出一个更基本的问题（答案并不明显）：

问题3是否存在任何有限时间的算法？

困难在于即使有解，矩阵 A 和 W 的元素也可能是无理数。这与3-SAT问题完全不同，3-SAT问题有一个简单的暴力算法。相比之下，对于非负矩阵分解来说，设计能在有限时间内运行的算法是非常具有挑战性的。但实际上，在实际的RAM模型中，有一些算法可以在固定的时间内判断多项式不等式系统是否有解。如果有解，这些算法还可以计算出解的隐式表示。

其中一个。输出是一个多项式和一个区间（对于每个变量），在这个区间内只有一个根，这个根是真实解中变量的值。通过对根进行二分搜索，您可以找到尽可能多的解的位。

第一个算法源自Tarski的开创性工作，并且在此基础上有了一系列改进，这些改进基于越来越强大的代数分解。这一系列工作的结果是，算法的运行时间在变量数量上呈指数增长，但在问题的其他参数上呈多项式增长（多项式不等式的数量、最大次数和系数的位复杂度）。运行时间为 $(nD)^{O(r)}$ ，其中 n 是多项式不等式的数量， D 是最大次数， r 是变量的数量[106]。根据指数时间假设，这个运行时间基本上是最优的[78]。特别地，如果存在一个运行时间为 $(pD)^{o(r)}$ 的算法来解决这个问题，那么它将为3-SAT问题提供亚指数时间的算法。

我们可以使用这些算法来解决非负矩阵分解问题。然而，即使对于每个 A 或 W 的条目，我们需要的变量数量为 $k+m$ ，其中 k 和 m 是变量的数量。因此，即使 $k = O(1)$ ，我们仍然需要线性数量的变量，而运行时间将呈指数增长。然而，我们可以希望，即使朴素表示使用了许多变量，也许存在一种更巧妙的表示方法，使用更少的变量。我们能否将多项式不等式系统中的变量数量从 $O(nk + mk)$ 减少到 $f(k)$ ？

如果我们能够做到这一点，那么我们可以在多项式时间内解决非负矩阵分解问题，对于任意的 $k = O(1)$ 。接下来，我们将描述一些关于实数的一阶理论的基本工具。这些结果将有助于形式化我们从上面的直觉，即在推理解决多项式不等式系统的难度时，变量的数量是正确的复杂度度量，但它们的证明超出了本课程的范围。

实数的一阶理论

定义2.2.3 一个集合 S 是半代数的，如果存在多元多项式 p_1, \dots, p_n 使得

$$S = \{x_1, \dots, x_r \mid p_i(x_1, \dots, x_r) \geq 0\}$$

或者如果 S 是这些集合的有限并集或交集。

定义2.2.4 半代数集合 S 的投影定义为

$$\text{proj}_S(X_1, \dots, X_\ell) = \{x_1, \dots, x_\ell \mid \exists x_{\ell+1}, \dots, x_r \text{ such that } p(x_1, \dots, x_r) \in S\}$$

定理2.2.5 (Tarski)半代数集合的投影是半代数的。

这是该领域的基础性结果之一，通常被称为量词消除[110], [107]。为了对这个概念有一些熟悉，考虑代数集合的情况（类似地定义，但是使用多项式等式约束而不是不等式）。事实上，上述定理意味着代数集合的投影本身也是半代数的。它的投影是否也是代数的？

不（例如，考虑一个圆的投影）！

早些时候，我们提到有算法可以在时间 $(nD)^{O(r)}$ 内解决多项式不等式系统的问题，并找到其隐式表示的解（如果存在）。其中， n 是多项式不等式的数量， D 是最大的次数， r 是变量的数量[106]。事实上，这些算法适用于更一般的情况，其中还有一个布尔函数 \mathbb{B} 来约束多项式的符号模式。我们感兴趣的是判断集合

$$S = \{x_1, \dots, x_r \mid \mathbb{B}(p_1(x_1, \dots, x_r), \dots, p_n(x_1, \dots, x_r)) = \text{true}\}$$

是否为空集，并且我们假设我们可以评估 \mathbb{B} （但不能说它有一个简洁的电路）。一个相关的结果是著名的Milnor-Warren界限（参见例如[7]）：

定理2.2.6 (Milnor-Warren) 给定 n 个多项式 p_1, \dots, p_m ，其次数 $\leq D$ ，以及变量 $\mathbf{x} = x_1, \dots, x_r$ ，考虑在 \mathbf{x} 处的符号模式：

$$\mathbf{x} \rightarrow \left(\text{sgn}(p_1(\mathbf{x})), \text{sgn}(p_2(\mathbf{x})), \dots, \text{sgn}(p_m(\mathbf{x})) \right)$$

那么当 \mathbf{x} 在 \mathbb{R}^r 范围内变化时，不同符号模式的数量最多为 $(nD)^{\wedge r}$ 。

事实上，我们可能预期最多有 3^n 个符号模式。事实上，解决多项式不等式系统的算法是基于巧妙地枚举符号模式集合，以便总运行时间由可能存在的不同符号模式的最大数量主导！事实上，Milnor-Warren界可以被看作是Sauer-Shelah引理的类比，后者在监督学习中起到了变量的VC维度的作用。

接下来我们将给出一种减少变量数量的技术。

变量减少

很明显，满足(2.1)的点集是一个半代数集。然而，即使对于 $k=3$ ，这个系统也有线性数量的变量（关于 n 和 m ），因此直接求解(2.1)需要指数时间。

问题4我们能否找到一个替代的多项式不等式系统，表达相同的决策问题但使用更少的变量？

我们将重点研究一种特殊情况，称为单纯因子分解，其中 $\text{rank}(M) = k$ 。在这种情况下，我们询问 $\text{rank}^+(M) = \text{rank}(M) = k$ 是否成立，这简化了问题，因为有以下观察：

声明2.2.7在任何解中， A 和 W 必须分别具有完整的列和行秩。

证明： A 的列空间必须包含 M 的列，同样地 W 的行空间必须包含 M 的行。由于 M 的秩为 k 且 A 和 W 分别具有 k 个列和行，我们可以得出结论， A 和 W 分别具有完全的列秩和行秩。此外，它们的张成空间必须分别是 M 的列空间和行空间。 ■

因此，我们知道 A 和 W 分别具有左伪逆 A^+ 和右伪逆 W^+ 。我们将利用这些伪逆来减少多项式不等式系统中的变量数量：我们有 $A^+A = I_r$ 其中 I_k 是 $k \times k$ 的单位矩阵。因此 $A^+AW = W$ 。

因此，我们可以通过线性变换从 M 的列中恢复 W 的列。这导致了以下备选的多项式不等式系统：

$$(2.2) \quad \begin{cases} MW^+A^+M &= M \\ MW^+ &\geq 0 \\ A^+M &\geq 0 \end{cases}$$

先验地，我们并不清楚我们是否取得了进展，因为这个系统也有与 A^+ 和 W^+ 的条目对应的变量 $nk + mk$ 。然而，考虑一下矩阵 A^+M 。如果我们将 A^+ 表示为一个 $k \times n$ 矩阵，那么我们描述的是它对所有向量的作用，但是关键观察是我们只需要知道 A^+ 对 M 的列的作用，这些列构成了一个 k 维空间。因此，我们可以应用基变换将 M 重新写为 M_R ，其中 M_R 是一个 $k \times m$ 矩阵，然后有一个 $k \times k$ 线性变换 T （由 A^+ 和基变换得到），使得 $TM_R = W$ 。对于 W ，类似的方法也适用，因此我们得到一个新的系统：

$$(2.3) \quad \begin{cases} M_CST M_R &= M \\ M_C S &\geq 0 \\ T M_R &\geq 0 \end{cases}$$

该系统的变量是 S 和 T 中的条目。因此，有 $2k^2$ 个变量。我们需要该系统具备的属性是

- (a) 如果单纯因子分解问题有解，则该系统有解（完备性）
- (b) 如果该系统有任何解，则单纯因子分解有解（正确性）

我们已经证明了第一个属性，第二个属性是因为我们可以设置 $A = M_C S$ 和 $W = T M_R$ ，这是一个有效的因子分解，内部维度为 k 。因此，如果我们将 Renegar 算法应用于这个新系统，算法的运行时间为 $(nm)^{O(k^2)}$ ，并解决了单纯因子分解问题。

上述方法基于 Arora 等人的论文[13]，其中作者还为非负矩阵分解提供了一种变量约简过程（在 A 和 W 不一定具有完全列或行秩的一般情况下）。作者将变量数量从 $(nk + mk)$ 减少到 $f(k) = 2k^2 2^k$ ，这导致了一个关于 k 的双指数时间算法。关键观察是，即使 A 没有完全列秩，我们也可以编写一组多项式不等式系统，该系统的每组列都有一个伪逆矩阵是满秩的（ W 也是如此）。然而， A 可能有多达

关列的最大集合，并且因此得到的多项式不等式系统具有 $f(k)$ 个变量，但 $f(k)$ 本身是 k 的指数级。在[94]中，作者进一步将变量数量减少到 $2k^2$ ，用于非负矩

阵分解，主要思想是即使 A 可能有指数级的线性无关列最大集合，它们的伪逆矩阵在代数上是相关的，并且可以使用 Cramer 法则在一组公共的 k^2 变量上表示。这导致了一个在 $(nm)^{O(k^2)}$ 时间内运行的非负矩阵分解的单指数时间算法，这基本上是最优的，因为任何在时间 $(nm)^{o(k)}$ 内运行的算法都将导致一个次指数时间算法用于 3-SAT [13]。

2.3 稳定性和可分离性

在前一节中，我们采用了代数方法，而在这里，我们将使用一个被称为可分离性的假设[54]，这将使我们能够给出一个在多项式时间内运行的算法（即使对于较大的 r 值）。我们的讨论将围绕中间单纯形问题展开。

中间单纯形问题

让我们定义中间单纯形问题：

我们给定两个多面体 Q 和 P ，其中 $P \subseteq Q$ ，并且 P 由其顶点编码， Q 由其面编码。是否存在一个单纯形 K ，使得 $P \subseteq K \subseteq Q$ ？

我们希望将这个问题与非负矩阵分解联系起来，因为它将帮助我们建立起一个几何视角来解决问题。考虑以下问题：

给定非负矩阵 M 和 A ，是否存在 $W \geq 0$ ，使得 $M = AW$ ？

答案是“是”，当且仅当 M 的每一列都在 A 的列的非负组合所张成的锥体中。此外，如果我们将 M 和 A 的列归一化，使它们的和为一，那么答案是“是”，当且仅当 A 的列的凸包包含 M 的列。回顾一下，在单纯形分解中，我们给定一个秩为 k 的非负矩阵 M ，我们的目标是决定是否秩⁺ $(M) = k$ 。我们将证明单纯形分解问题和中间单纯形问题是等价的[115]。

考虑以下辅助问题，我们称之为 (P0)：

给定 $M = UV$ ，是否存在一个可逆的 $k \times k$ 矩阵 T 使得 UT^{-1} 和 TV 都是非负的？

事实上，Vavasis [115] 证明了 (P0)，中间单纯形和单纯形因子分解问题是多项式时间可互约化的。很容易看出，(P0) 和单纯形因子分解问题是等价的，因为在任何两个分解 $M = UV$ or $M = AW$ （其中内维等于 M 的秩）， M ， U 和 A 的列空间是相同的。同样，行空间也是相同的。 M ， V 和 W 也是相同的。

证明更有趣的方面是 (P0) 和中间单纯形问题之间的等价性。翻译如下：

- (a) 行数 $U \iff$ 顶点数 P
- (b) 行数 $T \iff$ 顶点数 K
- (c) 列数 $V \iff$ 面数 Q

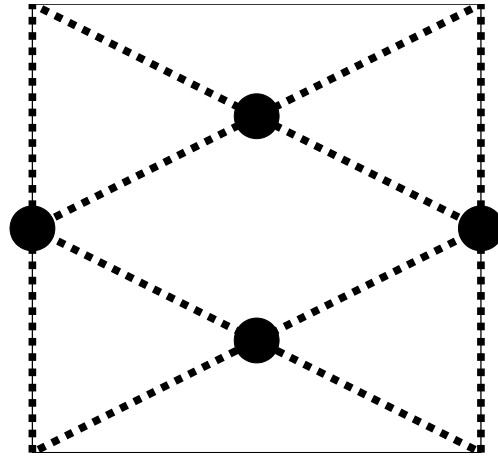


图2.1: 该图取自[115]。中间单纯形问题有两个解，将用于编码变量的真值赋值。

然后，约束条件 UT^{-1} 为非负（大致）约束条件 $P \subseteq K$ 和约束条件 TV 为（大致）约束条件 $K \subseteq Q$ 。由于我们需要小心区分一组向量的凸包和由所有非负组合生成的锥体之间的区别，因此存在一些繁琐的归一化问题。然而，这个等价关系给我们提供了一个有用的几何视角。

Vavasis 利用前一小节的等价关系证明了非负矩阵分解是 NP 难的。考虑图 2.1 中的装置；关键属性是只有两个可能的中间三角形，然后可以用来表示变量 x_i 的真值赋值。完整约简的描述和其正确性的证明比较复杂（见[115]）。

问题在于像图 ?? 中的那些小工具是不稳定的。我们可以通过对问题进行小的扰动来改变解的数量。Donoho 和 Stodden [54] 出于唯一性和鲁棒性的考虑，引入了一个称为可分离性的条件，可以缓解许多这些问题，我们将在下一小节中讨论。

可分离性

定义 2.3.1 如果对于矩阵 A 的每一列，存在一行 A ，其唯一的非零元素在该列中，则称 A 是可分离的。

此外，在可分离的非负矩阵分解问题中，我们给定 M ，并承诺如果存在非负矩阵分解，则存在一个其中 A 是可分离的。Donoho 和 Stodden 使用这个条件（和其他条件）来证明在某些自然条件下，非负矩阵分解是唯一的。Arora, Ge, Kannan 和 Moitra 提供了一个找到它的算法：

定理2.3.2[13] 给定一个非负矩阵 M ，并承诺存在一个非负矩阵分解 $M = AW$ 其中 A 是可分离的，存在一个多项式时间算法来计算最小内维度的分解。

事实上，在文本分析的背景下，可分离性是非常自然的。回想一下，我们将 A 的列解释为主题。我们可以将可分离性看作是这些主题带有锚定词的承诺；简单地说，对于每个主题，存在一个未知的锚定词，如果它出现在一个文档中，那么该文档就是关于给定主题的（部分）内容。例如，*401k* 可能是主题“个人财务”的一个锚定词。

为什么锚定词有帮助呢？很容易看出，如果 A 是可分离的，那么 W 的行将出现在 M 中（经过缩放）。因此，我们只需要确定 M 中哪些行对应于锚定词。根据我们在第2.3节的讨论，我们知道（如果我们缩放 M 、 A 和 W ，使它们的行总和为一）， W 的凸包包含 M 的行。但由于这些行也出现在 M 中，我们可以尝试通过迭代删除不改变其凸包的 M 的行来找到 W 。

让 M^i 表示 M 的第 i 行，让 M^I 表示 M 对于 I 中的行的限制，其中 $I \subseteq [n]$ 。现在我们可以使用以下简单的过程找到锚点词：

找到锚点词[13]

输入：矩阵 $M \in \mathbb{R}^{n \times m}$ 满足定理2.3.2中的条件

输出： $W = M^I$

设 $I = [n]$

对于 $i = 1, 2, \dots, n$

 如果 $M^i \in \text{conv}(\{M^j \mid j \in I, j \neq i\})$ ，则将 $I \leftarrow I - \{i\}$

结束

很容易看出，删除不是锚点词的行不会改变剩余行的凸包，因此上述算法终止

只包含锚点词的集合 I 。此外，在终止时

$$\text{conv}(\{M^i | i \in I\}) = \text{conv}(\{M^j\}_j)$$

或者说凸包与开始时相同。因此，被删除的锚点词是多余的，我们也可以不使用它们。

可分离的NMF [13]

输入：满足定理2.3.2中条件的矩阵 $M \in \mathbb{R}^{n \times m}$

输出： A, W

在 M 上运行Find Anchors，令 W 为输出

求解非负的 A ，使得 $\|M - AW\|_F$ 最小（凸规划）

结束

定理的证明立即从Find Anchors的正确性证明和以下事实得出：如果且仅如果存在一个非负的 A （其行之和为一）使得 $M = AW$ ，则 $\text{conv}(\{M^i\}_i) \subseteq \text{conv}(\{W^i\}_i)$ 。

当直接实现上述算法时，速度会非常慢。

相反，对上述算法进行了许多改进[27]，[84] [65]，并且我们将描述其中一个改进，该改进出现在[12]中。假设我们随机选择一行 M^i 。那么很容易看出，距离 M^i 最远的行将是一个锚词。

同样，如果我们找到了一个锚词，距离它最远的行将是另一个锚词，依此类推。通过这种方式，我们可以贪婪地找到所有的锚行，而且这种方法只依赖于成对距离和投影，因此我们可以在运行这个贪婪算法之前应用维度缩减。这样一来，我们就可以在上述算法的第一步中完全避免线性规划，而且第二步也可以快速实现，因为它涉及将一个点投影到一个 $k-1$ 维的单纯形中。

2.4 主题模型

在这里，我们将考虑一个相关的问题，称为主题建模；详见[28]中的全面介绍。这个问题与非负矩阵分解密切相关，但有两个关键的不同之处。再次，存在一些分解 $M = AW$ 但现在我们无法访问 M ，而是访问到一个非常粗糙的近似 \hat{M} 。直观地说， M 中的每一列都是一个文档，它本身是一个分布。

单词。但现在我们观察到的单词是从这个分布中取样的（所以我们实际上不知道 M 的列）。

第二个不同之处在于我们将 W 视为随机生成的。实际上有很多流行的选择：

- (a) 纯文档：每个文档只涉及一个主题，因此 W 的每一列都只有一个非零值。
- (b) 潜在狄利克雷分配[30]： W 的列是从狄利克雷分布生成的。
- (c) 相关主题模型[29]：某些主题对可以正相关或负相关，并且生成 W 的精确分布是对数正态分布。
- (d) Pachinko分配模型[89]：这是LDA的多层泛化，还允许某些类型的结构化相关性。

还有很多选择。无论如何，我们的目标是从 \widetilde{M} 中学习 A 。为了强调差异，注意即使我们知道 A ，我们也无法精确计算 W 。或者， \widetilde{M} 和 M 可能会非常不同，因为前者可能是稀疏的，而后者是稠密的。是否存在可证明的主题建模算法？

格拉姆矩阵

我们将遵循Arora, Ge和Moitra [14]的方法。起初，这似乎是一个与我们考虑过的问题根本不同的问题，因为在这个模型中，我们不能要求更长的文档，我们只能要求更多的文档。因此，我们增加了 \widetilde{M} 的列数，但每一列与 M 中的相应列并不是那么接近。基本思想是改为使用格拉姆矩阵 G ：

定义2.4.1 让 G 表示一个词 \times 词的矩阵，其在 (a, b) 位置的元素是随机选择的文档中前两个词是 a 和 b 的概率。

定义2.4.2 让 R 表示一个主题 \times 主题的矩阵，其在 (i, j) 位置的元素是随机选择的文档中前两个词（再次是从主题 i 和 j 生成的）的概率。

请注意，我们可以从样本中近似计算 G ，但我们不能直接近似计算 R ，它受我们用于生成 W 列的分布选择控制。更准确地说：

引理 2.4.3 $G = ARA^T$

证明：让 w_1 表示第一个词，让 t_1 表示 w_1 的主题（对于 w_2 也是如此）。我们可以展开 $\mathbb{P}[w_1 = a, w_2 = b]$ 为：

$$\sum_{i,j} \mathbb{P}[w_1 = a, w_2 = b | t_1 = i, t_2 = j] \mathbb{P}[t_1 = i, t_2 = j]$$

引理立即得证。 ■

关键观察是 G 具有可分离的非负矩阵分解，由 A 和 RA^T 给出，因为 A 是可分离的，而后者矩阵是非负的。事实上，如果 RA^T 具有满行秩，则定理 2.3.2 中的算法将找到真正的锚词集合。然而，由于 RA^T 的行不再归一化为和为一，上述分解不一定是唯一的。尽管如此，我们已经取得了一些进展，并且可以采用贝叶斯解释（参见[12]）。

通过贝叶斯规则进行恢复

实际上， A 的条目是条件概率 $\mathbb{P}(w_1 | t_1)$ ，因此我们可以推断关于后验分布 $\mathbb{P}(t_1 | w_1)$ 。实际上，这给我们提供了一个备用特征-化的锚词：如果一个词的后验分布仅支持一个主题，那么它就是一个锚词。特别地

$$\mathbb{P}(t_1 = t | w_1 = w) = \begin{cases} 1, & w \text{ 是 } t \text{ 的锚词,} \\ 0, & \text{否则,} \end{cases}$$

现在我们可以展开：

$$P(w_1 = w' | w_2 = w) = \sum_t \mathbb{P}(w_1 = w' | w_2 = w, t_2 = t) \cdot \mathbb{P}(t_2 = t | w_2 = w),$$

实际上 w_1 与 w_2 是独立的，如果我们在 t_2 的条件下进行调整，则：

$$\begin{aligned} \mathbb{P}(w_1 = w' | w_2 = w, t_2 = t) &= \mathbb{P}(\text{word1} = w' | \text{topic2} = t) \\ &= \mathbb{P}(\text{word1} = w' | \text{word2} = \text{anchor}(t)), \end{aligned}$$

我们可以在确定锚点词后，从 G 中计算出来。因此：

$$\mathbb{P}(w_1 = w' | w_2 = w) = \sum_t \mathbb{P}(\text{word1} = w' | \text{word2} = \text{anchor}(t)) \mathbb{P}(t_2 = t | w_2 = w)$$

我们可以将其视为变量 $\{\mathbb{P}(t_2 = t | w_2 = w)\}$ 的线性系统。不难看出，如果 R 具有满秩，则它有唯一解。最后，我们可以通过贝叶斯规则计算我们最初感兴趣的概率：
 $\mathbb{P}(\text{word } w | \text{topic } t) = \mathbb{P}(\text{topic } t | \text{word } w) \cdot \mathbb{P}(\text{word } w) \mathbb{P}(\text{topic } t)$

$$= \frac{\mathbb{P}(\text{主题 } t | \text{单词 } w) \cdot \mathbb{P}(\text{单词 } w)}{\sum_{w'} \mathbb{P}(\text{主题 } t | \text{单词 } w') \cdot \mathbb{P}(\text{单词 } w')}.$$

我们现在可以陈述算法 **Recover**。令 \widetilde{G} 为经验格拉姆矩阵，其中 $\widetilde{G}_{a,b}$ 是我们样本中第一个单词为 a 且第二个单词为 b 的文档的比例。

假设每个锚定词的概率至少为 p_0 。那么本小节的主要结果是：

定理2.4.4[14] 对于任何可分离的主题模型，其中 R 是满秩的，存在一个多项式时间算法来计算 \widetilde{A} ，使其与 A ε -接近，并且运行时间和样本复杂度（文档数量）是多项式 $(n, 1/p, 1/\varepsilon, 1/\sigma_{\min}(R))$ ，前提是文档长度至少为两个。

在下一小节中，我们将描述一些实验结果。

恢复 [14], [12]

输入：按词-文档矩阵 $M \in \mathbb{R}^{n \times m}$

输出： A, R

计算 \widetilde{G} ，计算 $\mathbb{P}(w_1 = w | w_2 = w')$

运行查找锚点

求解 $\mathbb{P}(\text{主题 } t | \text{词 } w)$ 并使用贝叶斯规则计算 A

结束

实验

我们现在面临一个基本的科学问题：是否真的存在锚词？

以下实验是在[12]中进行的：

- (a) 在一组纽约时报文章上运行 MALLET(一种流行的主题建模工具包), 其输出是一个主题矩阵 A 。
- (b) 使用 A 从一个主题模型生成数据, 在这些数据上运行 MALLET。

重要的是, 这里我们运行的数据实际上来自一个主题模型, 我们可以比较一个算法恢复真实矩阵的效果与另一个算法的效果。然后:

- (c) 在这些数据上运行新算法。

这是一个看似不公平的比较, 因为我们限制了自己只使用了一次MALLET已经找到的主题矩阵 A (所以这是我们对什么构成一个真实主题模型的概念)。然而令人惊讶的是, 前一小节中的算法能够更准确地找到主题矩阵 A , 并且速度快了几个数量级! 这是一个重要的例子, 证明了在我们能够给出可证明算法的条件下, 实际上会得到更好的算法。

第三章

张量方法

在本章中，我们将研究张量分解的算法及其在统计推断中的应用。

3.1 基础知识

在这里，我们将介绍张量的基础知识。矩阵是一个二阶张量-它由一对数字索引。一般来说，张量是按 k 元组索引的， k 被称为张量的阶。我们可以将张量 T 看作是 $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \dots \times \mathbb{R}^{n_k}$ 中的一个点。

在本章中，我们主要关注三阶张量的顺序。如果 T 是一个大小为 $m \times n \times p$ 的三阶张量，我们可以将 T 看作是一组堆叠在一起的大小为 $m \times n$ 的矩阵。

我们可以将线性代数中的许多标准定义推广到张量的情况，但我们要警告读者，虽然这些参数在矩阵中很容易计算，但大多数张量的参数在最坏情况下很难计算。

定义3.1.1 秩一张量是形式为 $T = u \otimes v \otimes w$ 的张量，其中 $T_{i,j,k} = u_i v_j w_k$ 。一般来说，张量 T 的秩是最小的 t ，使得我们可以将 T 写成 t 个秩一张量的和。

问题5 张量在计算上更难处理，那么为什么我们要尝试处理它们呢？

事实上，在下一节中，我们将给出一个激励性的例子，说明了张量方法在统计学和机器学习中的有用性（并且矩阵不足以满足需求）。

案例研究：斯皮尔曼的假设

查尔斯·斯皮尔曼是一位著名的心理学家，他假设存在两种基本的智力：数学智力和语言智力。特别是，他认为学生在各种测试中的表现如何仅取决于他们在这两个维度上的内在能力。为了验证他的理论，他设置了一项研究，让一千名学生参加了十种不同类型的测试。他将这些结果收集到一个矩阵 M 中，其中条目 M_{ij} 用于表示学生 i 在测试 j 上的表现。斯皮尔曼取了矩阵 M 的最佳秩二近似。换句话说，存在向量（不一定是单位向量） $u_1, u_2 \in \mathbb{R}_{1000}$ ， $v_1, v_2 \in \mathbb{R}_{10}$ ，使得

$$M \approx u_1 v_1^T + u_2 v_2^T$$

这被称为因子分析，他的结果在某种程度上证实了他的假设。

但是，这种方法存在一个根本性障碍，通常被称为“旋转问题”。设 $U = [u_1, u_2]$ 和 $V = [v_1, v_2]$ ，并且让 O 是一个正交矩阵。那么

$$UV^T = U O O^T V^T$$

是一种替代的分解方法，可以同样好地近似 M 。然而， UO 的列和 $O^T V^T$ 的行可能会更难解释。总结一下，仅仅因为给定的数据矩阵 M 有一个好的分解，并不意味着因子分析会找到它。

或者，假设我们有一个矩阵 $M = \sum_{i=1}^r x_i y_i^T$ 。

问题6如果我们知道 $\{x_i\}_i$ 和 $\{y_i\}_i$ ，我们能确定 M 吗？

实际上，在只有微不足道的条件下，我们才能唯一确定这些因素。如果 $r = 1$ ，或者我们基于先验原因知道向量 $\{x_i\}_i$ 和 $\{y_i\}_i$ 是正交的，那么我们可以。但是一般情况下，我们可以对 M 进行奇异值分解，得到 $U \Sigma V^T$ ，然后取 $\{\sigma_i u_i\}_i$ 和 $\{v_i\}_i$ 作为解释 M 的另一组因素（如果 $\{x_i\}_i$ 和 $\{y_i\}_i$ 不是正交的，那么这显然是同一个 M 的两组不同因素）。然而，如果我们给定一个张量

$$T = \sum_{i=1}^r x_i \otimes y_i \otimes w_i$$

然后存在一般条件（即如果 $\{x_i\}_i$ ， $\{y_i\}_i$ 和 $\{w_i\}_i$ 都是线性无关的），不仅真正的因式分解是 T 的唯一因式分解 rank r ，而且实际上有简单的算法来找到它！这正是张量方法在统计学和机器学习中无处不在的原因：如果我们

给定一个因子表示统计模型参数的张量，我们可以有效地找到这些因子；然而对于矩阵，因子并不是唯一确定的。

张量问题的复杂性

在前一小节中，我们提到张量方法将提供一种绕过“旋转问题”的方法，这是因子分析中常见的障碍。

那么我们可以直接计算张量的最小秩分解吗？事实上，这个问题在计算上是困难的（在没有进一步的假设的情况下），而且大多数张量问题都是困难的[71]！更糟糕的是，线性代数中的许多标准关系不成立，甚至在某些情况下定义也不明确。

(a) 对于一个矩阵 A ， $\dim(\text{span}(\{A_i\}_i)) = \dim(\text{span}(\{A^j\}_j))$ （列秩等于行秩）。

然而，对于张量来说，没有这样的关系成立。

(b) 对于一个矩阵 A ，最佳的秩 k 逼近可以从其最佳的秩 $k+1$ 逼近中获得。

特别地，如果我们将 $A^{(k+1)}$ 设为 A 的最佳秩 $k+1$ 逼近，则 $A^{(k+1)}$ 的最佳秩 k 逼近是 A 的最佳秩 k 逼近。但是对于张量来说，最佳的秩 k 和秩 $k+1$ 逼近不一定共享任何公共的秩一因子。事实上，从张量 T 中减去最佳的秩一逼近可能会增加其秩。

(c) 对于一个实值矩阵，其在 \mathbb{R} 上的秩和在 \mathbb{C} 上的秩是相同的，但对于张量来说这是错误的。

存在实值张量，其最小秩分解需要复数。

也许最令人担忧的问题是，在某些情况下定义也会失败：

定义3.1.2 张量 T 的边界秩是最小的 r ，使得对于任意的 $\varepsilon > 0$ ，存在一个秩为 r 的张量，其每个元素与 T 的对应元素之间的差距小于 ε 。

我们注意到，在上述定义中使用的范数并不重要。事实上，对于矩阵来说，边界秩等于秩。但对于张量来说，这两者可能不同。

(d) 对于一个张量来说，它的边界秩不一定等于它的秩。

考虑以下的 $2 \times 2 \times 2$ 张量 T ，在 \mathbb{R} 上：

$$T = \left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right).$$

我们将省略 T 具有秩3的证明，但证明 T 可以近似为秩2的任意接近。考虑以下的矩阵

$$S_n = \left(\begin{pmatrix} n & 1 \\ 1 & \frac{1}{n} \end{pmatrix}, \begin{pmatrix} 1 & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n^2} \end{pmatrix} \right) \text{ 和 } R_n = \left(\begin{pmatrix} n & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right).$$

很容易看出 $S_n = n \begin{pmatrix} 1 & 1/n \\ 1/n & 1/n^2 \end{pmatrix}$ ，因此是秩为1的， R_n 也是秩为1的。因此张量 $S_n - R_n$ 的秩为2，但也是一个 $1/n$ 逐元素近似于 T 的。

最后一个问题是很容易看出，一个随机的 $n \times n \times n$ 张量的秩为 $\Omega(n^2)$ ，但如何显式构造任意阶三张量的秩为 $\Omega(n^{1+\varepsilon})$ 是未知的。而任何这样的构造都将为任何明确问题[102]提供第一个超线性的电路下界，这是电路复杂性中一个长期存在的开放问题。

詹里奇算法

虽然我们不能指望在一般情况下找到张量的最小秩分解的算法，但实际上有一些温和的条件下我们可以做到。这个算法在各种应用中已经被多次重新发现，并且经过广泛的搜索，我们发现这个简单的算法最早是在Harshman的一篇工作论文[70]中报道的，作者归功于Robert Jennrich博士。我们将陈述并证明这个结果的一个更一般的版本，遵循Leurgans、Ross和Abel的方法[87]：

定理 3.1.3[70], [87] 考虑一个张量

$$T = \sum_{i=1}^r u_i \otimes v_i \otimes w_i$$

其中每组向量 $\{u_i\}_i$ 和 $\{v_i\}_i$ 是线性独立的，而且每对向量 $\{w_i\}_i$ 也是线性独立的。那么上述分解是唯一的，只是在重新缩放时有所不同，并且存在一种高效的算法来找到它。

我们将看到这个基本结果的各种应用（这可能解释了为什么它被多次重新发现），包括系统发生重建[96]，主题建模[8]和社区检测[9]。这种分解还在学习球形高斯混合模型[75]和独立成分分析[36]中起着关键作用，尽管我们将为后者问题提供一种局部搜索算法。

张量分解 [70], [87]

输入: 张量 $T \in \mathbb{R}^{m \times n \times p}$ satisfying Theorem 3.1.3 中的条件

输出: 因子 $\{u_i\}_i, \{v_i\}_i$ 和 $\{w_i\}_i$

随机选择 \mathbb{S}^{p-1} 中的 a, b uniformly; 设置 $T_a = T(*, *, a)$ 和 $T_b = T(*, *, b)$

计算 $T_a(T_b)^+$ 和 $T_b(T_a)^+$ 的特征分解

令 U 和 V 为特征向量

如果他们的特征值是倒数，则配对 u_i 和 v_i

求解 w_i 在 $T = \sum_{i=1}^r u_i \otimes v_i \otimes w_i$

结束

回想一下 T 是通过 T 的矩阵切片的加权和，每个切片都加权通过 a_i 。很容易看出：

$$\text{声明 3.1.4} \quad T_a = \sum_{i=1}^r \langle w_i, a \rangle u_i v_i^T \quad \text{和} \quad T_b = \sum_{i=1}^r \langle w_i, b \rangle u_i v_i^T$$

或者，令 $D_a = \text{diag}(\{\langle w_i, a \rangle\}_i)$ 并且令 $D_b = \text{diag}(\{\langle w_i, b \rangle\}_i)$ 。然后我们可以写成 $T_a = U D_a V^T$ 和 $T_b = U D_b V^T$ 其中 U 和 V 的列分别是 u_i 和 v_i 。

引理 3.1.5 $T_a(T_b)^+$ 和 $T_b(T_a)^+$ 的特征向量分别是 U 和 V （缩放后）

证明：我们可以使用上述公式计算 T_a 和 T_b 的值

$$T_a(T_b)^+ = U D_a D_b^+ U^+$$

几乎可以肯定地说，在选择 a 和 b 的情况下， $D_a D_b^+$ 的对角线将是不同的 - 这是我们使用每对向量 $\{w_i\}_i$ 线性无关的条件。

因此，上述公式 $T_a(T_b)^+$ 是一种特征分解，而且它是唯一的，因为它的特征值是不同的。我们得出结论， $T_a(T_b)^+$ 的特征向量确实是 U 的列，除了重新缩放， V 也是如此。 ■

现在，为了完成定理的证明，注意到 u_i 和 v_i 作为 $T_a(T_b)^+$ 和 $T_b(T_a)^+$ 的特征向量，其特征值为 $(D_a)_{i,i}(D_b)_{i,i}^{-1}$ 和 $(D_b)_{i,i}(D_a)_{i,i}^{-1}$ 。 (再次， $D_a(D_b)^+$ 的对角线几乎肯定是不同的，所以 v_i 是唯一可能与 u_i 配对的特征向量)。由于我们只有缩放的因子 $u_i \times v_i$ ，我们需要将重新缩放的因子与 w_i 一起推入。然而，我们只需要证明线性系统在 w_i 的解不超过一个（它肯定有一个解）。

定义3.1.6 两个具有相同列数的矩阵 U 和 V 之间的 Khatri-Rao 乘积 \otimes_{KR} 是

$$(U \otimes_{KR} V)_i = u_i \otimes v_i$$

也就是说，大小为 $m \times r$ 和大小为 $n \times r$ 的 U 和 V 的 Khatri-Rao 乘积是一个 $m \times n \times r$ 的张量，其 i th 列是 U 的 i th 列和 V 的 i th 列的张量积。以下引理留给读者作为练习：

引理 3.1.7 如果 U 和 V 的大小为 $m \times r$ 和 $n \times r$ ，并且具有完全列秩和 $r \leq m + n - 1$ ，那么 $U \otimes_{KR} V$ 也具有完全列秩。

这立即意味着线性系统在 w_i 上有唯一解。这完成了定理的证明。

注意，如果 T 的大小为 $m \times n \times p$ ，那么定理的条件只有在 $r \leq \min(m, n)$ 时才成立。上述算法的扩展适用于更高阶的张量，即使 r 大于其因子的任何维度[48]，[66]，[26]，并且在超完备独立分量分析[66]和学习多个高斯混合[26]，[11]中有有趣的应用。

在下一节中，我们将展示上述算法是稳定的 - 在所有学习应用中，我们将从样本中估计 T ，因此我们没有 T 的精确值。

3.2 扰动界限

在最后一节中，我们给出了一个张量分解的算法，当因子是满秩时，它的分解是唯一的（除了缩放）。

然而，在我们的所有应用中，我们将不会精确地给出 T ，而是从我们的样本中计算出一个近似值。本节的主要目标是展示，即使在存在噪声的情况下，定理3.1.3中的算法也能够恢复接近真实因子的因子。在后面的章节中，我们将简单地假设我们给出了真实的张量 T ，而我们在这里提供的内容正是为了证明这种简化的合理性。

本节内容有些技术性，读者可以随意跳过。

回想一下，定理3.1.3中的主要步骤是计算特征分解。因此，我们的第一个目标是建立在哪些条件下特征分解本身是稳定的。更准确地说，设 $M = U D U^{-1}$ ，其中 D 是一个对角矩阵。如果我们给出 $\tilde{M} = M + E$ ，我们何时能够恢复出对 U 的良好估计？

直观地说，如果 D 中的任何对角元素接近或者 U 病态，那么即使是一个小的扰动 E 也会极大地改变特征分解。我们将证明这些是唯一可能出错的事情。这 will 有两个主要步骤。首先，我们需要证明 \tilde{M} 是可对角化的，然后我们可以证明对角化 \tilde{M} 的矩阵必须接近于对角化 M 的矩阵。

条件数

定义3.2.1 矩阵 M 的条件数 *of a matrix* M 被定义为

$$\kappa(M) := \frac{\sigma_{\max}(M)}{\sigma_{\min}(M)},$$

其中 $\sigma_{\max}(M)$ 和 $\sigma_{\min}(M)$ 分别是 M 的最大和最小奇异值。

考虑解方程组 $Mx = b$ 的基本问题。假设我们确切地知道 M ，但我们只知道一个估计值 $\tilde{b} = b + e$ 的 b 。这里 e 是一个误差项。通过使用 \tilde{b} 而不是来解方程 $Mx = b$ ，我们得到一个对 x 的估计值 \tilde{x} 。那么 \tilde{x} 与 x 有多接近？

我们有 $\tilde{x} = M^{-1}\tilde{b} = x + M^{-1}e = x + M^{-1}(\tilde{b} - b)$ 。所以

$$\|x - \tilde{x}\| \leq \frac{1}{\sigma_{\min}(M)} \|b - \tilde{b}\|。$$

由于 $Mx = b$ ，我们也有 $\|b\| \leq \sigma_{\max}(M)\|x\|$ 。因此

$$\frac{\|x - \tilde{x}\|}{\|x\|} \leq \frac{\sigma_{\max}(M)}{\sigma_{\min}(M)} \frac{\|b - \tilde{b}\|}{\|b\|} = \kappa(M) \frac{\|b - \tilde{b}\|}{\|b\|}。$$

换句话说，条件数控制着在解决线性系统时的相对误差。

格什戈林盘定理

回想一下，我们的第一个中间目标是要证明 $M + E$ 是对角化的，并且我们将引用以下定理：

定理3.2.2 矩阵 M 的特征值都包含在复平面上的以下圆盘的并集中：

$$\bigcup_{i=1}^n D(M_{ii}, R_i)$$

其中 $D(a, b) := \{x \mid \|x - a\| \leq b\} \subseteq \mathbb{C}$ and $R_i = \sum_{j \neq i} |M_{ij}|$.

证明：设 (x, λ) 是一个特征向量-特征值对（注意，即使 M 不可对角化，这也是有效的）。设 i 表示 x 的坐标，其绝对值最大。则 $Mx = \lambda x$ 给出 $\sum_j M_{ij}x_j = \lambda x_i$ 。所以 $\sum_{j \neq i} M_{ij}x_j = \lambda x_i - M_{ii}x_i$ 。

我们得出结论：

$$|\lambda - M_{ii}| = \left| \sum_{j \neq i} M_{ij} \frac{x_j}{x_i} \right| \leq \sum_{j \neq i} |M_{ij}| = R_i.$$

因此 $\lambda \in D(M_{ii}, R_i)$. ■

第一部分

现在让我们回到前一节开始时提出的问题: \widetilde{M} 是否可对角化? 考虑

$$U^{-1}\widetilde{M}U = D + U^{-1}EU.$$

证明 \widetilde{M} 可对角化的过程如下：

第(a)部分由于 \widetilde{M} 和 $U^{-1}\widetilde{M}U$ 是相似矩阵，它们具有相同的特征值集合。

部分 (b) 此外，我们可以将定理 3.2.2 应用于 $U^{-1}\widetilde{M}U = D + U^{-1}EU$ 如果 U 是良好条件的， E 足够小，那么半径将远小于 D 中最接近的对角元素对。因此，我们得出结论， $U^{-1}\widetilde{M}U$ 和 \widetilde{M} 的特征值是不同的，因此后者可以被对角化。

感谢 Santosh Vempala 指出原始分析中的错误；参见 [66] 以获取更详细的证明。

第二部分

现在我们可以转向我们的第二个中间目标，即如何将其与实际对角化 $M = U D U^{-1}$ 进行比较？

更具体地说，如果 $(\widetilde{u}_i, \widetilde{\lambda}_i)$ 和 (u_i, λ_i) 分别是 \widetilde{M} 和 M 的对应特征向量-特征值对，那么 $(\widetilde{u}_i, \widetilde{\lambda}_i)$ 与 (u_i, λ_i) 有多接近？使用Part 1中的论证，我们知道对于每个 i ， $\widetilde{\lambda}_i \approx \lambda_i$ 。此外，我们假设当 $i = j$ 时， M 的特征值有足够的分离。仍需检查 $\widetilde{u}_i \approx u_i$ 。让

$$\sum_j c_j u_j = \widetilde{u}_i.$$

回想一下 $\widetilde{M} = M + E$ 。将上述方程两边左乘 \widetilde{M} ，我们得到

$$\sum_j c_j \lambda_j u_j + E \widetilde{u}_i = \widetilde{\lambda}_i \widetilde{u}_i. \implies \sum_j c_j (\lambda_j - \widetilde{\lambda}_i) u_j = -E \widetilde{u}_i.$$

设 w_j^T 为 U^{-1} 的第 j 行。将上述方程两边左乘以 w_j^T ，我们得到

$$c_j (\lambda_j - \widetilde{\lambda}_i) = -w_j^T E \widetilde{u}_i.$$

回忆我们假设 M 的特征值是分离的。因此，如果 E 足够小，我们可以保证 $\lambda_j - \widetilde{\lambda}_i$ 与零有一定的距离。然后我们可以限制 c_j 的范围，这意味着 \widetilde{u}_i 和 u_i 是接近的。

我们可以定性地重新陈述如下：让 δ 是 M 的最接近的特征值对之间的间隔，让 κ 是 U 的条件数。那么如果 $\|E\| \leq \text{poly}(1/n, 1/\kappa, \delta)$ ，误差的范数满足 $\|\widetilde{U} - U\| \leq \text{poly}(1/n, 1/\kappa, \delta, \|E\|)$ 。

回到张量分解

我们将引入一些符号来解释张量分解的应用。

让“ \rightarrow ”表示一个矩阵以逆多项式速率收敛到另一个矩阵（作为样本数量的函数）。例如，当 \widetilde{T} 表示一个分布的经验矩（具有有界矩）而 T 表示其真实矩时， $\widetilde{T} \rightarrow T$ 。同样地， $\widetilde{T}_a = \widetilde{T}(*, *, a) \rightarrow T_a$ ， b 也是如此。

我们将其留给读者作为一个练习，来检查在自然条件下 $\widetilde{T}_b^+ \rightarrow T_b^+$ 是否成立。由此可得 $\widetilde{T}_a \widetilde{T}_b^+ \rightarrow T_a T_b^+$ 。我们已经证明了，如果 $E \rightarrow 0$ ，那么 M 和 $M^+ + E$ 的特征分解会收敛。最后，我们得出结论，定理3.1.3中的算法计算出的因子 \widetilde{U} 和 \widetilde{V} 会收敛。

以一个反多项式速率收敛到真实因子 U 和 V ，对于 \widetilde{W} 和 W 也可以使用类似的证明。

开放问题：Kruskal秩

我们用一个开放问题结束本节。

定义3.2.3 向量集合 $\{u_i\}_i$ 的Kruskal秩 \mathbf{o} 是最大的 r ，使得所有 r 个向量的子集线性无关。

我们将在后面看到，计算Kruskal秩是 NP -难的。然而，对于张量分解存在强大的唯一性定理（基于该参数），目前没有已知的算法证明：

定理3.2.4 (Kruskal) 设 $T = \sum_{i=1}^r u_i \otimes v_i \otimes w_i$ ，并且设 k_u ， k_v 和 k_w 分别是 $\{u_i\}_i$ ， $\{v_i\}_i$ 和 $\{w_i\}_i$ 的Kruskal秩。如果 $k_u + k_v + k_w \geq 2r + 2$ ，则 T 的秩为 r ，并且这个分解是唯一的，除了缩放。

开放问题1 是否存在一种在任何自然条件下都能进行张量分解的高效算法，对于 $r = (1 + \varepsilon)n$ ，其中 $\varepsilon > 0$ ？

例如，对于张量分解[26]，自然考虑使用平滑分析模型，其中 T 的因子被扰动，因此不是对手选择的。

上述唯一性定理适用于 $r = 3/2n - O(1)$ ，但在这个模型中，对于 $r = (1 + \epsilon)n$ 的张量分解，目前没有已知的算法（尽管对于高阶张量有更好的算法）。

3.3 系统发育树和隐马尔可夫模型

在这里，我们描述了张量分解在系统发育重建和HMMs中的应用。

模型

系统发育模型具有以下组成部分：

- (a) 一个以 r 为根的二叉树，其中叶子节点的深度不一定相同。

生物学解释是叶子代表现存的物种，内部节点代表物种分化事件。

(b) 一个状态集合 Σ ，例如 $\Sigma = \{A, C, G, T\}$ 。令 $k = |\Sigma|$ 。

(c) 树上的马尔可夫模型；即根节点的状态分布 π_r 和每条边 (u, v) 的转移矩阵 P_{ij}^{uv} 。

我们可以按照以下方式从模型中生成样本：根据 π_r 选择根节点的状态，对于每个父节点 u 的子节点 v ，根据 P^{uv} 的第 i 行定义的分布选择 v 的状态，其中 i 是 u 的状态。或者，我们可以将 $s(\cdot) : V \rightarrow \Sigma$ 视为将状态分配给顶点的随机函数，其中对于 $s(r)$ 的边际分布为 π_r 。

$$P_{ij}^{uv} = \mathbb{P}(s(v) = j | s(u) = i)。$$

请注意，当树中从 v 到 t 的唯一最短路径通过 u 时， $s(v)$ 与 $s(t)$ 是独立的。

我们的目标是从多项式数量的随机样本中学习上述模型 - 包括树和转移矩阵。我们假设转移矩阵是满秩的，这种情况下可以任意选择根节点。为了将其与生物学联系起来，我们假设我们已经对每个现存物种进行了测序，并且这些序列已经被正确地对齐。我们将这些序列中的第 i 个符号视为来自上述模型的独立样本，并且我们希望重建导致我们今天拥有的物种的进化树，以及对这些进化分支的长度有一些理解。我们提到一个警告，计算生物学中最有趣和具有挑战性的问题之一是执行多序列比对，而我们在这里假设一个完全对齐的序列集是我们的起点。此外，我们的进化模型过于简化，只考虑点突变，而不考虑插入、删除和交叉。

这实际上是两个独立的学习目标：我们找到拓扑的方法将遵循Steel [109]和Erdos, Steel, Szekely和Warnow [57]的基础工作。从这个我们可以应用张量方法来找到转移矩阵，按照Chang [36]的方法和后来的Mossel和Roch [96]。

寻找拓扑

这里的基本思想是在树的边上定义一个合适的距离函数[109]，这样我们就可以近似计算样本中叶子之间的距离，然后构建树。

定义树度量

首先假设对于叶子 a 和 b ，我们可以访问 F^{ab} 的真实值，其中

$$F_{ij}^{ab} = \mathbb{P}(s(a) = i, s(b) = j)。$$

在[109]中，Steel以一种允许我们计算叶子节点之间的距离的方式定义了树上的距离度量，给定 F^{ab} 。特别地，让

$$\psi_{ab} := -\ln |\det(F^{ab})|。$$

Steel证明了

$$\psi_{ab} = \sum_{(u,v) \in p_{ab}} \nu_{uv},$$

其中 p_{ab} 是从 a 到 b 的唯一路径，且

$$\nu_{uv} = -\ln |\det(P^{uv})| + \frac{1}{2} \ln \left(\prod_{i \in [k]} \pi_u(i) \right) - \frac{1}{2} \ln \left(\prod_{i \in [k]} \pi_v(i) \right)。$$

然后他证明了 ν_{uv} 始终为非负数（这并不明显），因此 ψ 确实是一个度量。

重要的是，我们可以从样本中估计 F^{ab} ，因此我们可以（近似地）计算出叶子节点上的 ψ_{ab} 。

重建四元组

在这里，我们将使用 ψ 来计算拓扑结构。固定四个叶子 a, b, c 和 d ，并且在这些叶子之间存在着三种可能的诱导拓扑结构，如图3.1所示。（这里的诱导拓扑结构是指删除不在任何最短路径上的边，并在可能的情况下将路径缩减为单个边）。我们的目标是根据两两之间的距离确定哪种诱导拓扑结构是真实的拓扑结构。考虑图3.1左侧的拓扑结构（a）；在这种情况下，我们有

$$\psi(a, b) + \psi(c, d) < \min \{ \psi(a, c) + \psi(b, c), \psi(a, d) + \psi(b, d) \},$$

因此，我们可以通过计算三个值 $\psi(a, b) + \psi(c, d)$ ， $\psi(a, c) + \psi(b, c)$ ，和 $\psi(a, d) + \psi(b, d)$ 来确定真正的诱导拓扑。无论哪个是最小的，都决定了诱导拓扑，因为哪些节点配对在诱导拓扑中有一个共同的父节点（再次在诱导拓扑中）。

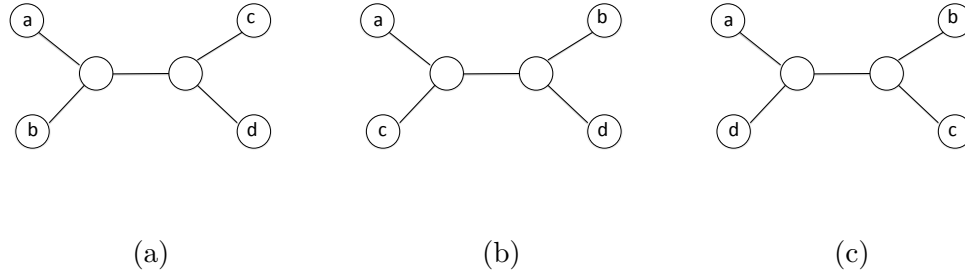


图3.1：可能的四联体拓扑

事实上，仅通过这些四联体测试，我们就可以恢复树的拓扑。例如，一对叶子节点 a, b 如果且仅当这些节点在每个四联体测试中始终具有共同的父节点时，它们才有相同的父节点。因此，我们可以将所有叶子节点配对，使它们具有相同的父节点，并且很容易将这种方法扩展到恢复树的拓扑。

处理噪声

请注意，我们只能从样本中近似计算 F^{ab} 。当 a 和 b 接近时，这意味着对 ψ_{ab} 的近似较好，但当 a 和 b 相距较远时会有噪声。Erdos、Steel、Szekely 和 Warnow 在 [57] 中的方法是仅使用所有距离都很短的四元组。

寻找转移矩阵

在这里，我们假设我们知道树的拓扑结构和所有三元组 a, b, c of leaves 的 T_{abc} ，其中

$$T_{ijk}^{abc} = \mathbb{P}(s(a) = i, s(b) = j, s(c) = k)。$$

（我们可以从随机样本中近似计算）。然后考虑在 a 、 b 和 c 之间的所有最短路径上的唯一节点；由于我们可以任意重新根化树，让这个节点为 r 。然后

$$\begin{aligned} T^{abc} &= \sum_{\ell} \mathbb{P}(s(r) = \ell) \mathbb{P}(s(a) = \cdot | s(r) = \ell) \otimes \mathbb{P}(s(b) = \cdot | s(r) = \ell) \otimes \mathbb{P}(s(c) = \cdot | s(r) = \ell) \\ &= \sum_{\ell} \mathbb{P}(s(r) = \ell) P_{\ell}^{ra} \otimes P_{\ell}^{rb} \otimes P_{\ell}^{rc} \end{aligned}$$

我们使用 P_{ℓ}^{rx} 来表示过渡矩阵 P^{rx} 的第 ℓ 行。

现在我们可以应用第3.1节中的算法来计算一个张量分解，其因子在重新缩放后是唯一的。此外，这些因子是概率分布，因此我们可以计算它们的适当归一化。我们将这个过程称为星形测试。（事实上，在第3.1节中的张量分解算法已经被多次重新发现，也被称为Chang的引理[36]。）

在[96]中，Mossel和Roch使用这种方法来找到一个系统发生树的转移矩阵，给定树的拓扑结构，具体步骤如下。假设 u 和 v 是内部节点， w 是一个叶子节点。此外，假设 v 位于 u 和 w 之间的最短路径上。基本思想是写成

$$P^{uw} = P^{uv} P^{vw}$$

如果我们可以找到 P^{uw} 和 P^{vw} （使用上述星号测试），那么我们可以计算 $P^{uv} = P^{uw} (P^{vw})^{-1}$ ，因为我们假设转移矩阵是可逆的。

然而，存在两个严重的复杂性：

(a) 与找到拓扑结构一样，长路径非常嘈杂。

Mossel和Roch证明了只使用对短路径的查询也可以恢复转移矩阵。

(b) 我们只能恢复到重新标记的张量分解。

在上述星型测试中，我们可以对状态 r 应用任何排列，并相应地对转移矩阵 P^{ra} , P^{rb} 和 P^{rc} 的行进行排列，以使得在 a , b 和 c 上的联合分布保持不变。

然而，Mossel和Roch的方法是在PAC学习[114]的框架下工作，目标是学习一个生成模型，该模型在叶子节点上产生几乎相同的联合分布。特别地，如果有多种方法可以标记内部节点以产生相同的叶子节点上的联合分布，我们对它们是无差别的。

备注3.3.1 *HMM* 是一种特殊情况系统发生树，其中底层拓扑结构是一条毛虫。但请注意，对于上述算法，我们需要转移矩阵和观测矩阵是满秩的。

更准确地说，我们要求转移矩阵是可逆的，并且与隐藏节点对应的行空间以及输出符号对应的列空间的观测矩阵都具有满行秩。

超越满秩?

上述算法假设所有转移矩阵都是满秩的。事实上，如果我們去掉这个假设，那么很容易嵌入一个经典的难学习问题——噪声奇偶问题[31]的实例。让我们首先定义这个无噪声的问题：

令 $S \subset [n]$ ，并且独立均匀地随机选择 $X_j \in \{0, 1\}^n$ ，对于 $j = 1, \dots, m$ 。给定 X_j 和 $b_j = \chi_S(X_j) := \sum_{i \in S} X_j(i) \bmod 2$ ，对于每个 j ，目标是恢复 S 。

这很容易：令 A 为矩阵，其 j 行为 X_j ，令 b 为列向量，其 j 项为 b_j 。很容易看出 1_S 是线性系统 $Ax = b$ 的一个解，其中 1_S 是 S 的指示函数。此外，如果我们选择 $\Omega(n \log n)$ 个样本，那么 A 几乎必然是满列秩的，因此这个解是唯一的。然后我们可以通过在 $GF(2)$ 上求解一个线性系统来找到 S 。然而，上述问题的一个微小变化不会改变样本复杂度，但会使问题变得极其困难。噪声奇偶问题

与上述问题相同，但对于每个 j ，我们独立地给出值 $b_j = \chi_S(X_j)$ ，概率为 $2/3$ ，否则 $b_j = 1 - \chi_S(X_j)$ 。挑战在于我们不知道哪些标签被翻转了。

声明 3.3.2 存在一种暴力算法可以解决噪声奇偶问题
使用 $O(n \log n)$ 样本

证明：对于每个 T ，计算 $\chi_T(X_j)$ 和 b_j 在样本上。实际上，如果且仅如果 $S = T$ ， $\chi_T(X_j)$ 和 b_j 是相关的。 ■

该算法的运行时间为 2^n （大致）。由于 Blum、Kalai 和 Wasserman [31] 的最新技术，运行时间和样本复杂度为 $2^{n/\log n}$ 。广泛认为，即使给出任意多个多项式数量的样本，噪声奇偶问题也没有多项式时间算法。这是一个很好的例子，其样本复杂度和计算复杂度（猜测）是完全不同的。

接下来，我们将展示如何将噪声奇偶问题的样本嵌入到 HMM 中，但是为了做到这一点，我们将使用不是满秩的转移矩阵。考虑一个具有 n 隐藏节点的 HMM，其中第 i 个隐藏节点用于表示 X 的第 i 个坐标和运行奇偶性。

$$\chi_{S_i}(X) := \sum_{i' \leq i, i' \in S} X(i') \bmod 2.$$

我们可以定义以下转移矩阵:

$$\begin{aligned} \text{如果 } i+1 \in S \quad & P^{i+1,i} = \begin{cases} \frac{1}{2} & (0, s_i) \\ \frac{1}{2} & (1, s_i + 1 \pmod{2}) \\ \text{否则为0} \end{cases} \\ \text{如果 } i+1 \notin S \quad & P^{i+1,i} = \begin{cases} \frac{1}{2} & (0, s_i) \\ \frac{1}{2} & (1, s_i) \\ \text{否则为0} \end{cases}. \end{aligned}$$

请注意，这里的观测矩阵肯定不是满秩的，因为我们只观察到两种可能的发射，即使每个内部节点有四种可能的状态！因此，当转移（或观测）矩阵不是满秩时，这些问题变得更加困难！

在这里，我们将张量方法应用于社区检测。有许多情况下，我们希望发现社区-即具有强联系的群体。在这里，我们将重点放在图论方法上，其中我们将社区视为一组节点，这些节点与集合外的节点相比更好地连接在一起。我们可以形式化这个概念的许多方式，每种方式都会导致不同的优化问题，例如最稀疏的割或最密集的子图。

然而，这些优化问题都是 NP -难解的，更糟糕的是，很难近似解决。相反，我们将在一个平均情况模型中制定我们的问题，在这个模型中，存在一个基础的社区结构用于生成一个随机图，我们的目标是以高概率从图中恢复出真实的社区。

块随机模型

在这里，我们介绍了块随机模型，用于生成一个具有 $|V| = n$ 的随机图。此外，该模型由参数 p 和 q 以及由函数 π 指定的分区指定：

- $\pi : V \rightarrow [k]$ 将顶点 V 分成 k 个不相交的组（我们稍后会放宽这个条件）；
- 每个可能的边 (u, v) 都是独立选择的，满足：

$$\Pr[(u, v) \in E] = \begin{cases} q & \pi(u) = \pi(v) \\ p & \text{otherwise} \end{cases}.$$

在我们的设置中，我们将设置 $q > p$ ，但是在某些情况下也研究了该模型 $q < p$ 的情况。（特别是，当 $q = 0$ 时，我们可以要求找到该随机图的 k -着色）。无论如何，我们观察从上述模型生成的随机图，并旨在恢复由 π 描述的分区。

这个信息在理论上是什么时候可能的？事实上，即使对于 $k = 2$ ，其中 π 是一个二分，我们也需要

$$q - p > \Omega\left(\sqrt{\frac{\log n}{n}}\right)$$

为了使真正的二分成为唯一最小的切割，我们需要高概率地将随机图 G 二分。如果 $q - p$ 更小，那么甚至在理论上也不可能找到 π 。事实上，我们还应该要求分区的每个部分都很大，为了简单起见，我们将假设 $k = O(1)$ 和 $|\{u | \pi(u) = i\}| = \Omega(n)$ 。

在块随机模型中，关于分割随机图的工作已经有很长一段时间了，最终由 McSherry [91] 完成：

定理 3.4.1/[91] 存在一种有效的算法，可以恢复 π （重新标记）如果

$$\frac{q - p}{q} > c \sqrt{\frac{\log n / \delta}{qn}}$$

并且至少以概率 $1 - \delta$ 成功。

该算法基于谱聚类，我们将观察到的邻接矩阵视为一个秩 k 矩阵和一个误差项的和。如果误差很小，那么我们可以通过恢复接近真实秩 k 矩阵的东西来

寻找最佳秩 k 近似邻接矩阵。有关详细信息，请参见[91]。

相反，我们将采用Anandkumar等人[9]的方法，使用张量分解。事实上，[9]的算法也适用于混合成员模型，在该模型中，我们允许每个节点成为 $[k]$ 的分布。

然后，如果 π^u 和 π^v 是 u 和 v 的概率分布，则边 (u, v) 的概率为 $\sum_i \pi_i^u \pi_i^v q + p$ 。我们可以将这个概率解释为： u 和 v 根据 π^u 和 π^v 分别选择一个社区，如果他们选择了相同的社区，则以概率 q 存在一条边，否则以概率 p 存在一条边。

请记住，为了应用张量分解方法，我们实际上需要的是条件独立的随机变量！事实上，我们将基于计数三个星星来获得这样的随机变量。

计数三个星星

我们将 V 分成四个集合（任意选择） X, A, B 和 C 。让 $\Pi \in \{0, 1\}^{V \times k}$ 表示将顶点分配给社区的（未知）分配，使得 Π 的每一行都恰好包含一个 1。同时，让 $R \in \mathbb{R}^{k \times k}$ 表示描述每对社区之间存在边的概率的矩阵。特别地，

$$(R)_{ij} = \begin{cases} q & i = j \\ p & i \neq j \end{cases}.$$

考虑乘积 ΠR 。 ΠR 的第 i 列编码了从社区 i 中的顶点到给定其他顶点的边出现的概率：

$$(\Pi R)_{xi} = \Pr[(x, a) \in E | \pi(a) = i].$$

我们将使用 $(\Pi R)^{A_i}$ 来表示矩阵 ΠR 限制在第 i 列和 A 中的行，对于 B 和 C 也是如此。此外，让 p_i 是在 X 中属于社区 i 的节点的比例。然后考虑以下张量

$$T := \sum_i p_i T_x = \sum_i p_i (\Pi R)_i^A \otimes (\Pi R)_i^B \otimes (\Pi R)_i^C.$$

关键主张是：

主张 3.4.2 设 $a \in A, b \in B$ 和 $c \in C$ ；那么 $T_{a,b,c}$ 恰好是一个随机节点 $x \in X$ 与 a, b 和 c 相连的概率。

这是直接从上面的定义得出的。特别是，如果我们观察 (x, a) , (x, b) 和 (x, c) 是否是 G 中的边，这些是条件独立的随机变量。然后我们需要证明：

- (a) 如果 $|X| = \widetilde{\Omega}(|A||B||C|/\epsilon^2)$ ，那么我们可以准确地估计 T
- (b) 因子 $\{(\Pi R)^{Ai}\}_i$, $\{(\Pi R)^{Bi}\}_i$ 和 $\{(\Pi R)^{Ci}\}_i$ 是线性独立的，因此根据定理 3.1.3， T 的张量分解是唯一的。

更准确地说，我们需要这些因素具有良好的条件，以便我们可以从近似 \widetilde{T} 到 T 中近似地计算它们。参见第3.2节。

- (c) 我们可以通过重新标记来恢复 π 从 $\{(\Pi R)^{Ai}\}_i$ 。

部分 (a) 让 $\{X_{a,b,c}\}_{a,b,c}$ 是 X 的几乎相等大小的集合的一个划分，其中一个集合对应于每个 A 中的 a ， B 中的 b 和 C 中的 c 。然后

$$\widetilde{T}_{a,b,c} = \frac{|\{x \in X_{a,b,c} | (x, a), (x, b), (x, c) \in E\}|}{|X_{a,b,c}|}$$

将会在很大概率下接近于 $T_{a,b,c}$ 。然后我们可以使用一个并集约束。

部分 (b) 很容易看出 R 是满秩的，而且如果我们随机选择 A , B 和 C ，如果每个社区足够大，那么很有可能每个社区在 A , B 和 C 中都有很好的代表性，因此因子 $\{(\Pi R)^{Ai}\}_i$, $\{(\Pi R)^{Bi}\}_i$ 和 $\{(\Pi R)^{Ci}\}_i$ 将非常接近线性相关。

部分 (c) 注意，如果我们对 $\{(\Pi R)^{Ai}\}_i$ 有一个很好的近似，那么我们可以将 A 分成社区。反过来，如果 A 足够大，我们可以将这种分区扩展到整个图：如果节点 $x \notin A$ 与 $\pi(a) = i$ 的节点 $a \in A$ 连接的比例接近 q ，那么我们将节点 x 添加到社区 i 中。很有可能，这将恢复出真实的社区。

然而，要完整证明算法，请参见[9]。Anandkumar等人还提供了一个混合成员模型的算法，其中每个 π_u 都是从一个狄利克雷分布中选择的。我们将不涉及后一种扩展，因为我们将在下一节中介绍这些类型的技术。

我们注意到，有一种称为半随机模型的强大扩展可以应用于块随机模型。大致上，这些模型允许“对手”在同一簇中的节点之间添加边缘，并在 G 生成后删除簇之间的边缘。

被生成。如果 π 是 G 的最佳分区，则在更改后这只会更加真实。有趣的是，许多谱算法在这种更灵活的模型中失效，但在这种更一般的情况下，有一些优雅的技术可以恢复 π （参见[60]，[59]）。

3.5 混合模型的扩展

在这里，我们将扩展张量谱模型以适应（一些）混合模型。

纯主题模型

首先，我们描述了将张量方法应用于纯主题模型的简单应用（参见[10]）。请记住，存在一个未知的主题矩阵 A ，并且我们从以下模型中获取样本：

- (a) 以概率 p_i 选择主题 i 用于文档 j
- (b) 根据分布 A_i 选择 N_j 个单词

如果每个文档至少有三个单词，我们可以定义张量 \widetilde{T} ，其中 \widetilde{T} 的 a 、 b 、 c 位置分别表示样本中第一个单词、第二个单词和第三个单词为 a 、 b 和 c 的文档的比例。然后很容易看出，如果文档数量足够大，则 \widetilde{T} 收敛到：

$$T = \sum_{i=1}^r p_i A_i \otimes A_i \otimes A_i$$

为了应用第3.1节中的算法，我们只需要 A 具有完整的列秩。在这种情况下，分解中的因子是唯一的，除了缩放因子，算法将找到它们。最后， A 中的每一列都是一个分布，因此我们可以适当地对这些列进行归一化，并计算值 p_i 。请回顾第3.2节中我们分析了张量分解算法的噪声容忍度。

很容易看出，如果我们至少有多项式 $(n, 1/\epsilon, 1/\sigma_r)$ 长度至少为三文档，那么该算法可以高概率地恢复一个主题矩阵 \widetilde{A} 和一个分布 $\{\widetilde{p}_i\}_i$ ，其中 \widetilde{A} 和 $\{\widetilde{p}_i\}_i$ 分别接近 A 和 $\{p_i\}_i$ 。其中， n 是词汇表的大小， σ_r 是 A 的最小奇异值。

我们将这称为张量方法应用于纯模型的一个例子，因为每个文档只由一个主题描述。类似地，在我们应用于社区检测的情况下，每个节点属于一个且仅属于一个社区。

最后，在我们应用于系统发生重建的情况下，每个隐藏节点处于一个且仅处于一个状态。然而需要注意的是，在主题模型的背景下，更加现实的假设是每个文档本身就是主题的混合，我们将这些称为混合模型。

潜在狄利克雷分配

在这里，我们将为学习一个非常流行的混合模型——潜在狄利克雷分配[30]，提供一个张量谱算法。令 $\Delta := \{x \in \mathbb{R}^r : x \geq 0, \sum_i x_i = 1\}$ 表示 r 维单纯形。然后我们从以下模型中获得样本：

- (a) 根据狄利克雷分布 $\text{Dir}(\{\alpha_i\}_i)$ ，为文档 j 选择一个主题混合 w_j 。
- (b) 重复 N_j 次：从 w_j 中选择一个主题 i ，并根据分布 A_i 选择一个词。

狄利克雷分布的定义如下：

$$p(x) \propto \prod_i x_i^{\alpha_i - 1} \text{ 对于 } x \in \Delta$$

请注意，如果文档很长（比如 $N_j > n \log n$ ），那么在纯主题模型中，文档对的词的经验分布往往非常相似。但在上述混合模型中，情况就不再如此。

将我们的张量谱方法扩展到混合模型的基本问题是计算三个单词的张量 \widetilde{T} 的收敛性

$$T = \sum_{ijk} D_{ijk} A_i \otimes A_j \otimes A_k$$

其中 $D_{i,j,k}$ 是从主题 i, j 和 k 生成的随机文档的前三个单词的概率。在纯主题模型中， $D_{i,j,k}$ 是对角线的，但在混合模型中不是！

定义3.5.1 T 的 *Tucker* 分解是

$$T = \sum_{i,j,k} D_{i,j,k} a_i \otimes b_j \otimes c_k$$

我们称 D 为核心张量。

这与张量分解的标准定义不同，我们只对 $i = j = k$ 求和。好消息是计算张量的Tucker分解很容易。事实上，我们总是可以将 r_1 设置为 $\{T_{i,*,*}\}_i$ 的维度，并且对于 r_2 和 r_3 也是如此。然而，坏消息是Tucker分解通常不是唯一的，所以即使我们有 T ，我们也不能必然计算上述分解，其因子是主题模型中的主题。

我们如何扩展张量谱方法以适应混合模型？

Anandkumar等人[8]的优雅方法基于以下思想：

引理3.5.2

$$\begin{aligned} T &= \sum_{ijk} D_{ijk} A_i \otimes A_j \otimes A_k \\ S &= \sum_{ijk} \widetilde{D}_{ijk} A_i \otimes A_j \otimes A_k \\ \implies T - S &= \sum_{ijk} (D_{ijk} - \widetilde{D}_{ijk}) A_i \otimes A_j \otimes A_k \end{aligned}$$

证明：该证明是多线性代数中的一个简单练习。 ■

因此，如果我们可以访问使用相同因子 $\{A_i\}_i$ 在其Tucker分解的其他张量 S ，我们可以减去 T 和 S ，并希望使核心张量对角化。在我们的设置中，我们可以将 D 看作是Dirichlet分布的三阶矩。我们还可以访问哪些其他张量？

其他张量

我们根据以下实验描述了基于张量 T 的概率：假设 T 为随机文档中前三个词的概率，分别为 a 、 b 和 c 。

但我们也可以考虑其他实验。为了为LDA提供张量谱算法，我们需要进行以下三个实验：

- (a) 随机选择三个文档，并查看每个文档的第一个词
- (b) 随机选择两个文档，并查看第一个文档的前两个词和第二个文档的第一个词
- (c) 随机选择一个文档，并查看其前三个词

这些实验产生的张量的因子相同，但其核在其自然Tucker分解中不同。

定义3.5.3 让 μ 、 M 和 D 为狄利克雷分布的一阶、二阶和三阶矩。

更准确地说，让 μ_i 表示从主题 i 生成的随机文档中第一个词的概率。让 $M_{i,j}$ 表示从主题 i 和 j 生成的随机文档中第一个和第二个词的概率。同样地，让 $D_{i,j,k}$ 表示从主题 i 、 j 和 k 生成的随机文档中前三个词的概率。然后，让 T^1 、 T^2 和 T^3 分别表示第一次、第二次和第三次实验的期望。

$$\text{引理 3.5.4 (a) } T^1 = \sum_{i,j,k} [\mu \otimes \mu \otimes \mu]_{i,j,k} A_i \otimes A_j \otimes A_k$$

$$(b) T^2 = \sum_{i,j,k} [M \otimes \mu]_{i,j,k} A_i \otimes A_j \otimes A_k$$

$$(c) T^3 = \sum_{i,j,k} D_{i,j,k} A_i \otimes A_j \otimes A_k$$

证明:让 w_1 表示第一个词，让 t_1 表示 w_1 的主题（对于其他词也是类似的）。我们可以将 $\mathbb{P}[w_1 = a, w_2 = b, w_3 = c]$ 展开为：

$$\sum_{i,j,k} \mathbb{P}[w_1 = a, w_2 = b, w_3 = c | t_1 = i, t_2 = j, t_3 = k] \mathbb{P}[t_1 = i, t_2 = j, t_3 = k]$$

引理立即得证。 ■

注意 $T_{a,b,c}^2 = T_{a,c,b}^2$ 因为两个词来自同一篇文章。

然而，我们可以以自然的方式对称化 T^2 ：设置 $S_{a,b,c} = T_{a,b,c}^2 + T_{a,c,b}^2 + T_{c,a,b}^2$ 。因此 $S_{a,b,c} = S_{\pi(a),\pi(b),\pi(c)}^2$ 对于任意排列 $\pi: \{a, b, c\} \rightarrow \{a, b, c\}$ 。

我们的主要目标是证明以下恒等式：

$$\alpha_0^2 D + 2(\alpha_0 + 1)(\alpha_0 + 2)\mu^{\otimes 3} - \alpha_0(\alpha_0 + 2)M \otimes \mu \quad (\text{三种方式}) = \text{diag}(\{p_i\}_i)$$

其中 $\alpha_0 = \sum_i \alpha_i$ 。因此我们有

$$\alpha_0^2 T^3 + 2(\alpha_0 + 1)(\alpha_0 + 2)T^1 - \alpha_0(\alpha_0 + 2)S^2 = \sum_i p_i A_i \otimes A_i \otimes A_i$$

重要的一点是，我们可以从我们的样本中估计左边的项（如果我们假设我们知道 α_0 ），并且我们可以应用第3.1节中的算法来恢复右边的张量，前提是 A 具有完整的列秩。实际上，我们可以从我们的样本中计算 α_0 （参见[8]），但我们将专注于证明上述恒等式。

狄利克雷的矩

我们想要建立的主要恒等式只是关于狄利克雷分布的一个陈述。实际上，我们可以将狄利克雷看作是以下组合过程的定义：

(a) 最初，每种颜色 i 的球有 α_i 个

(b) 重复 C 次：随机选择一个球，放回并加上一个相同颜色的球

这个过程给出了狄利克雷分布的另一种特征，从中可以直接计算：

$$(a) \quad \mu = [\frac{\alpha_1}{\alpha_0}, \frac{\alpha_2}{\alpha_0}, \dots, \frac{\alpha_r}{\alpha_0}]$$

$$(b) \quad M_{i,j} = \begin{cases} \frac{\alpha_i(\alpha_i+1)}{\alpha_0(\alpha_0+1)} & i = j \\ \frac{\alpha_i\alpha_j}{\alpha_0(\alpha_0+1)} & \text{否则} \end{cases}$$

$$(c) \quad T_{i,j,k} = \begin{cases} \frac{\alpha_i(\alpha_i+1)(\alpha_i+2)}{\alpha_0(\alpha_0+1)(\alpha_0+2)} & i = j = k \\ \frac{\alpha_i(\alpha_i+1)\alpha_k}{\alpha_0(\alpha_0+1)(\alpha_0+2)} & i = j \neq k \\ \frac{\alpha_i\alpha_j\alpha_k}{\alpha_0(\alpha_0+1)(\alpha_0+2)} & i, j, k \text{不同} \end{cases}$$

例如对于 $T_{i,i,k}$ ，这是第一个和第二个球是颜色 i ，第三个球是颜色 k 的概率。第一个球是颜色 i 的概率是 $\frac{\alpha_i}{\alpha_0}$ 。由于我们放回并加上一个相同颜色的球，第二个球也是颜色 i 的概率是 $\frac{\alpha_i+1}{\alpha_0+1}$ 。第三个球是颜色 k 的概率是 $\frac{\alpha_k}{\alpha_0+2}$ 。在其他情况下，检查上述公式很容易。

请注意，在上述公式中，只考虑分子会更容易思考。如果我们可以证明仅对分子成立以下关系

$$D + 2\mu^{\otimes 3} - M \otimes \mu \quad (\text{三种方式}) = \text{diag}(\{2\alpha_i\}_i)$$

很容易检查，通过乘以 $\alpha_0^3(\alpha_0+1)(\alpha_0+2)$ ，我们可以得到我们想要的公式。

定义3.5.5 令 $R = \text{num}(D) + \text{num}(2\mu^{\otimes 3}) - \text{num}(M \otimes \mu)$ (三种方式)

然后主要引理是：

引理3.5.6 $R = \text{diag}(\{2\alpha_i\}_i)$

我们将通过案例分析来建立这一点：

声明3.5.7 如果 i, j, k 是不同的，则 $R_{i,j,k} = 0$

这是显然的，因为 i, j, k 的分子 D ， $\mu^{\otimes 3}$ 和 $M \otimes \mu$ 都是 $\alpha_i \alpha_j \alpha_k$ 。

声明3.5.8 $R_{i,i,i} = 2\alpha_i$

这也是显然的，因为 i, i, i 的分子 D 是 $\alpha_i(\alpha_i + 1)(\alpha_i + 2)$ ，类似地， $\mu^{\otimes 3}$ 的分子是 α_i^3 。最后， $M \otimes \mu$ 的 i, i, i 的分子是 $\alpha_i^2(\alpha_i + 1)$ 。需要一些小心的情况是：

声明 3.5.9 如果 $i = k$ ， $R_{i,i,k} = 0$

这个案例棘手的原因是因为术语 $M \otimes \mu$ (三种方式) 不完全计数相同。如果我们将 μ 想象成张量的第三个维度，那么在同一篇文档中，第 i 个主题出现两次，但如果我们将 μ 想象成张量的第一或第二个维度，尽管第 i 个主题出现两次，但它并不在同一篇文档中出现两次。因此， $M \otimes \mu$ (三种方式) 的分子是 $\alpha_i(\alpha_i + 1)\alpha_k + 2\alpha_i^2\alpha_k$ 。此外， D 的分子是 $\alpha_i(\alpha_i + 1)\alpha_k$ ， $\mu^{\otimes 3}$ 的分子再次是 $\alpha_i^2\alpha_k$ 。

这三个主张共同证明了上述引理。尽管我们可以立即在纯主题模型中分解的张量 T^3 不再具有混合模型中的对角核张量，在LDA的情况下，我们仍然可以找到一个公式（我们可以从样本中估计每个项），使核心张量对角化。这导致：

定理3.5.10[8] 存在一个多项式时间算法，可以学习一个接近真实矩阵 A 的主题矩阵 \widetilde{A} ，在潜在狄利克雷分配模型中，只要我们至少给出多项式 $(n, 1/\epsilon, 1/\sigma_r, 1/\alpha_{min})$ 长度至少为三文档，其中 n 是词汇表的大小， σ_r 是 A 的最小奇异值， α_{min} 是最小的 α_i 。

同样，在混合模型中也有用于社区检测的算法，其中对于每个节点 u ，我们从狄利克雷分布中选择一个分布 π_u 来表示聚类[9]。然而，这些算法似乎非常依赖于使用狄利克雷分布的假设，并且很难将这些算法推广到任何其他自然分布。

3.6 独立成分分析

我们可以将我们开发的张量方法视为使用高阶矩来学习分布的参数的一种方式（例如用于系统发育树、隐马尔可夫模型、LDA、社区检测等）通过张量分解。在这里，我们将通过应用独立成分分析的方法来给出另一种使用矩方法的风格，这种方法由Comon [42]引入。

这个问题很容易定义：假设我们已经获得了以下形式的样本

$$y = Ax + b$$

其中我们知道变量 x_i 是独立的，线性变换 (A, b) 是未知的。目标是从多项式数量的样本中高效地学习 A, b 。这个问题有着悠久的历史，它的典型动机是考虑一个被称为鸡尾酒会问题的假设情境。

我们有 N 个麦克风和 N 个对话在一个房间里进行。
每个麦克风都听到了由 A 的相应行给出的对话的叠加。如果我们将对话视为独立和无记忆的，我们能分离它们吗？

这类问题通常也被称为盲源分离。我们将遵循 Frieze、Jerrum 和 Kannan [62] 的方法。

步骤 1

我们总是可以将问题转化为 $y = Ax + b$ 的形式，其中 $\mathbb{E}[x_i] = 0$ 并且 $\mathbb{E}[x_i^2] = 1$ ，对于所有的 i ，通过设置 $b = b + A \mathbb{E}[x]$ 和 $A_i = A_i \text{std}(x_i)$ ，其中 $\text{std}(x_i)$ 是 x_i 的标准差。

注意， y 的分布没有改变，但我们已经将 (A, x) 放入了一个规范形式，因为我们无法区分具有相同规范形式的一对线性变换。因此，不失一般性，我们已经将问题简化为学习问题。

$$y = Ax + b$$

对于所有 i ， $\mathbb{E}[x_i] = 0$ ， $\mathbb{E}[x_i^2] = 1$ 。同时，我们可以设置 $b = 0$ ，因为我们可以轻松学习 b 。我们将做出的关键假设是 A 是非奇异的。

步骤 2

$$\mathbb{E}[yy^T] = \mathbb{E}[Axx^T A^T] = AA^T$$

最后一个等式是基于 $E[x_i] = 0$, $E[x_i^2] = 1$ 和每个 x_i 是独立的条件。因此，我们可以获得 $M = AA^T$ ，通过采样足够多的随机样本，我们可以学习到任意精度。但是这告诉我们什么关于 A 呢？我们之前遇到过这个问题： M 不能唯一地定义 A ，我们的方法是考虑更高阶的张量。这次我们将以不同的方式进行。

由于 $M \succ 0$ ，我们可以找到 B 使得 $M = BB^T$ 。 B 和 A 之间有什么关系？

事实上，我们可以写成

$$BB^T = AA^T \Rightarrow B^{-1}AA^T(B^{-1})^T = I$$

这意味着 $B^{-1}A$ 是正交的，因为一个方阵乘以其转置等于单位阵，当且仅当它是正交的。所以我们已经学到了 A ，只是存在一个未知的旋转。我们能希望学到旋转 $R = B^{-1}A$ 吗？提示：如果每个 x_i 都是标准高斯分布呢？

在这种情况下， Rx 也是一个球形高斯分布，因此我们不能指望在没有额外假设的情况下学到 R 。事实上，这是唯一可能出错的情况：只要 x_i 不是高斯分布，我们就能学到 R 和 A 。为了简单起见，让我们假设每个 x_i 是 ± 1 ，因此 $\mathbb{E}[x_i^4] = 1$ ，而高斯分布的四阶矩是三。注意，我们可以将 B^{-1} 应用于我们的样本，因此我们可以想象我们从 $y = Rx$ 中获取样本。我们分析的关键是以下函数 $F(u) := \mathbb{E}[(u^T Rx)^4]$

当 u 在单位球上变化时， $v^T = u^T R$ 也在单位球上变化，因此我们可以通过解决以下等价的优化问题来替代在单位向量 u 上最小化函数 $F(u)$ ：min

$$\mathbb{E}[(v^T x)^4]$$

$\|v\|_2=1$

它的局部最小值是什么？

$$\begin{aligned} \mathbb{E}[(v^T x)^4] &= \mathbb{E}\left[\sum_i (v_i x_i)^4 + 6 \sum_{ij} (v_i x_i)^2 (v_j x_j)^2\right] = \\ &= \sum_i v_i^4 \mathbb{E}(x_i^4) + 6 \sum_{ij} v_i^2 v_j^2 + 3 \sum_i v_i^4 - 3 \sum_i v_i^4 + 3 \left(\sum_i v_i^2\right) \\ &= \sum_i v_i^4 \left(\mathbb{E}[x_i^4] - 3\right) + 3 \end{aligned}$$

因此, $F(v)$ 的局部最小值恰好对应于将 $v_i = \pm 1$, 其中 i 。请记住, $v^T = u^T R$, 因此这个特征说明了 $F(u)$ 的局部最小值对应于将 u 设置为 $\pm R$ 的一列。

该算法通过使用梯度下降 (和对Hessian的下界) 来展示可以快速找到 $F(u)$ 的局部最优解, 然后我们可以对我们找到的向量的正交补进行递归, 以找到 R 的其他列。这个想法需要一些小心, 以确保错误不会累积得太严重, 请参见[62], [16], [16]。

实际上, 我们刚刚计算的是分布的矩的备选基, 即累积量。通常情况下, 这些更容易处理, 因为它们满足一个有吸引力的性质, 即独立变量 X_i 和 X_j 的累积量的和是 X_i 和 X_j 的累积量的和。这正是我们在这里利用的性质。

第四章

稀疏恢复

在本章中，我们将研究稀疏恢复的算法：给定一个矩阵 A 和一个向量 b ，它是其列的稀疏线性组合 - 即 $Ax = b$ 和 x 是稀疏的 - 我们什么时候可以求解 x ？

4.1 基础知识

在本节中，我们只考虑列数多于行数的线性系统 $Ax = b$ 。因此，对于 x 存在不止一个解（如果存在解），我们将对具有最少非零元素的解感兴趣：

定义4.1.1 令 $\|x\|_0$ 为 x 的非零元素个数。

不幸的是，在一般情况下，找到线性方程组的最稀疏解是计算上困难的，但在一些重要的例子中，我们可以高效地求解 x 。

问题7 我们什么时候能找到线性方程组 $Ax = b$ 的最稀疏解？

一个显而易见的观察是，当 A 具有满列秩时，我们可以恢复 x 。在这种情况下，我们可以设置 $x = A^+b$ ，其中 A^+ 是 A 的左伪逆。需要注意的是，无论 x 是否稀疏，这个过程都是有效的。相反，当 A 的列数大于行数时，我们需要利用 x 的稀疏性。我们将证明，在 A 的某些条件下，如果 x 足够稀疏，则它确实是 $Ax = b$ 的唯一最稀疏解。

我们的第一个目标是证明在一个线性系统中找到最稀疏的解是困难的。我们将从相关问题开始：

问题1 (P) 在给定子空间 S 中找到最稀疏的非零向量 x

Khachiyan [81] 证明了这个问题是 NP 难的，这个结果有很多有趣的应用，我们稍后会讨论。

从子集和问题中简化

我们从以下子集和问题的变体简化：

问题2 (S) 给定不同的值 $\alpha_1, \dots, \alpha_m \in \mathbb{R}$, 是否存在一个集合 $I \subseteq [m]$ 使得 $|I| = n$ 且 $\sum_{i \in I} \alpha_i = 0$?

我们将把这个问题的一个实例嵌入到寻找给定子空间中最稀疏的非零向量的问题中。我们将使用以下映射，称为怪异矩阵曲线：

$$\Gamma^w(\alpha_i) \Rightarrow \begin{bmatrix} 1 \\ \alpha_i \\ \alpha_i^2 \\ \vdots \\ \alpha_i^{n-2} \\ \alpha_i^{n-1} \\ \alpha_i^n \end{bmatrix} \in \mathbb{R}^n$$

请注意，这与标准矩阵曲线不同，因为奇怪的矩阵曲线中有 α_i^n 而不是 α_i^{n-1} 。

声明4.1.2 一个具有 $|I| = n$ 的集合 I 具有 $\sum_{i \in I} \alpha_i = 0$ 当且仅当向量集合 $\{\Gamma^w(\alpha_i)\}_{i \in I}$ 线性相关。

证明：考虑矩阵的行是 $\{\Gamma^w(\alpha_i)\}_{i \in I}$ 的行列式。
然后证明基于以下观察：

(a) 行列式是关于变量 α_i 的多项式，总次数为 $\binom{n}{2} + 1$ ，这可以通过将行列式写成其拉普拉斯展开的形式来看出（参见[74]）。

(b) 此外，行列式可被 $\prod_{i < j} (\alpha_i - \alpha_j)$ 整除，因为如果任何 $\alpha_i = \alpha_j$ ，则行列式为零。

因此，我们可以将行列式写成

$$\left(\prod_{\substack{i < j \\ i, j \in I}} (\alpha_i - \alpha_j) \right) \left(\sum_{i \in I} \alpha_i \right)$$

我们假设 α_i 是不同的，因此行列式为零当且仅当 α_i 的和为0。 ■

现在我们可以完成在子空间中找到最稀疏的非零向量的证明：我们可以将 A 设置为由 $\Gamma^w(\alpha_i)$ 组成的 $n \times m$ 矩阵，并且令 $S = \ker(A)$ 。那么存在一个向量 $x \in S$ ，使得 $\|x\|_0 = n$ 当且仅当存在一个子集 I ，使得 $|I| = n$ ，其对应的子矩阵是奇异的。如果没有这样的集合 I ，则任何 $x \in S$ 都有 $\|x\|_0 > n$ 。因此，如果我们能够找到 S 中最稀疏的非零向量，我们就可以解决上述的子集和变体。实际上，这个证明立即得出了计算几何学中一个有趣的结果（在Khachiyan的论文

几年后才“开放”）。

定义4.1.3 \mathbb{R}^n 中的 m 个向量集合在一般位置上，如果最多 m 个向量的任意集合是线性无关的。

从上述规约中，我们得到一个结论，即很难确定 \mathbb{R}^n 中的 m 个向量集合是否处于一般位置上（因为存在一个 I ，其中 $|I| = n$ ，其子矩阵奇异当且仅当向量 $\Gamma^w(\alpha_i)$ 不处于一般位置上）。现在我们回到我们的原始问题：

问题3 (Q) 找到矩阵方程 $Ax = b$ 的最稀疏解 x 。

在 (P) 和 (Q) 之间存在微妙的差别，因为在 (P) 中我们限制了非零向量 x ，但在 (Q) 中对 x 没有这样的限制。然而，从 (Q) 到 (P) 存在一个简单的多对一规约。

引理 4.1.4 寻找最稀疏的解 x 到 $Ax = b$ 是 NP 难的。

证明：假设我们有一个线性系统 $Ax = 0$ ，我们想要找到最稀疏的非零解 x 。令 A^{-i} 等于矩阵 A 删除第 i 列后的矩阵。然后对于每个 i ，令 x^{-i} 是 $A^{-i}x^{-i} = A^i$ 的最稀疏解。令 i^* 为使 x^{-i} 最稀疏的索引，并假设 $\|x^{-i^*}\|_0 = k$ 。我们可以通过将 x 的 i^* 坐标设置为 -1 来构建一个满足 $Ax = 0$ 且 $\|x\|_0 = k + 1$ 的解。事实上，很容易看出 x 是 $Ax = 0$ 的最稀疏解。 ■

4.2 唯一性和不确定性原理

不连贯

在这里，我们将定义一个不连贯矩阵 A 的概念，并证明如果 x 足够稀疏，则它是 $Ax = b$ 的唯一最稀疏解。

定义4.2.1 矩阵 A 的列 $\in \mathbb{R}^{n \times m}$ 如果对于所有的 $i \neq j$ ，它们是 μ -不连贯的：

$$|\langle A_i, A_j \rangle| \leq \mu \|A_i\| \cdot \|A_j\|$$

虽然我们得出的结果可以推广到一般的 A ，但我们将限制我们的注意力在 $\|A_i\| = 1$ 的情况下，因此矩阵是 μ -不连贯的，对于所有的 $i \neq j$ ， $|\langle A_i, A_j \rangle| \leq \mu$ 。

事实上，不连贯矩阵非常常见。假设我们在 \mathbb{R}^n 中随机选择 m 个单位向量；那么很容易证明这些向量将是 $\mu = O(\sqrt{\frac{\log m}{n}})$ 。

因此，即使 $m = n^{100}$ ，这些向量也将是不相关的。实际上，还有更好的不相关向量的构造方法，可以去除对数因子；这几乎是最优的，因为对于任何 $m > n$ ， \mathbb{R}^n 中的任何一组 m 个向量的不相关性至少为 $\frac{1}{\sqrt{n}}$ 。

我们将多次使用以下示例：考虑矩阵 $A = [I, D]$ ，其中 $I \in \mathbb{R}^{n \times n}$ 是单位矩阵， $D \in \mathbb{R}^{n \times n}$ 是DFT矩阵。特别地， $D_{ij} = \frac{w^{(i-1)(j-1)}}{\sqrt{n}}$ 其中 $w = e^{i\frac{2\pi}{n}}$ 。这通常被称为尖峰和正弦矩阵。很容易看出 $\mu = \frac{1}{\sqrt{n}}$ here.

不确定性原理

重要的是，如果 A 是不连贯的，那么如果 x 足够稀疏，它将是 $Ax = b$ 的唯一最稀疏的解。这些类型的结果首先由Donoho和Stark的开创性工作[53]建立，并基于建立一个不确定性原理。

引理4.2.2 假设我们有 $A = [U, V]$ ，其中 U 和 V 是正交的。如果 $b = U\alpha = V\beta$ ，则 $\|\alpha\|_0 + \|\beta\|_0 \geq \frac{2}{\mu}$ 。

对于尖峰和正弦矩阵的解释是，任何信号在标准基础上至少有 \sqrt{n} 个非零元素，或者在傅里叶基础上有 \sqrt{n} 个非零元素。

从形式上来说，一个信号不能在时间和频率域中同时过于局部化！

证明：由于 U 和 V 是正交的，我们有 $\|b\|_2 = \|\alpha\|_2 = \|\beta\|_2$ 。我们可以将 b 重写为 $U\alpha$ 或 $V\beta$ ，因此 $\|b\|_2^2 = |\beta^T (V^T U)\alpha|$ 。由于 A 是不相干的，我们可以得出结论， $V^T U$ 的每个元素的绝对值最多为 $\mu(A)$ ，因此 $|\beta^T (V^T U)\alpha| \leq \mu(A) \|\alpha\|_1 \|\beta\|_1$ 。使用柯西-施瓦茨不等式可得 $\|\alpha\|_1 \leq$

$$\sqrt{\|\alpha\|_0 \|\alpha\|_2} \text{ 因此}$$

$$\|b\|_2^2 \leq \mu(A) \sqrt{\|\alpha\|_0 \|\beta\|_0} \|\alpha\|_2 \|\beta\|_2$$

重新排列，我们有 $\frac{1}{\mu(A)} \leq \sqrt{\frac{\|\alpha\|_0}{\|\beta\|_0}}$ 最后，应用AM-GM不等式
我们得到 $\frac{1}{\mu} \leq \|\alpha\|_0 + \|\beta\|_0$ 这就完成了证明。 ■

这个结果是否紧密？事实上，回到尖峰和正弦的例子，如果选择 b 为组合函数，其中信号在距离为 n 的等距离尖峰处有，那么 b has \sqrt{n} non-zeros 在标准基础上。此外，组合函数是其自身的离散傅里叶变换，因此在使用傅里叶基表示时也有 \sqrt{n} non-zeros。

接下来，我们将上述不确定性原理应用于证明一个唯一性结果：

命题 4.2.3 假设 $A = [U, V]$ 其中 U 和 V 是正交的， A 是 μ -不相干的。如果 $Ax = b$ 且 $\|x\|_0 < \frac{1}{\mu}$ ，那么 x 是唯一最稀疏的解。

证明：考虑任何替代解 $A\tilde{x} = b$ 。设 $y = x - \tilde{x}$ ，此时 $y \in \ker(A)$ 。将 y 写成 $y = [\alpha_y, \beta_y]^T$ ，由于 $Ay = 0$ ，我们有 $U\alpha_y = -V\beta_y$ 。现在我们可以应用不确定性原理并得出结论 $\|y\|_0 = \|\alpha_y\|_0 + \|\beta_y\|_0 \geq 2$ 。
很容易看出 $\|\tilde{x}\|_0 \geq \|y\|_0 - \|x\|_0 > 1$ 因此， \tilde{x} 比 x 有更多的非零元素，这就完成了证明。 ■

事实上，即使 A 是一个任意的不相干矩阵（而不是两个正交基的并集），类似的陈述也是正确的。我们将在下一节进一步讨论这个扩展。

Kruskal秩

我们还可以使用一个更一般的条件来工作，这个条件在证明唯一性时更有力；然而，与不相干性不同，这个条件在计算上很难验证。

定义4.2.4 一组向量 $\{A_i\}_i$ 的Kruskal秩 σ 是最大的 r ，使得所有的 r 个向量的子集都是线性无关的。

事实上，我们已经证明计算给定点集的Kruskal秩是 NP -难的，因为决定Kruskal秩是否为 n 恰好是决定点是否处于一般位置的问题。

然而， A 的Kruskal秩是分析 x 必须有多稀疏的正确参数，以使其成为 $Ax = b$ 的唯一最稀疏解。假设 A 的Kruskal秩是 r 。

定理4.2.5 如果 $\|x\|_0 \leq r/2$ ，则 x 是 $Ax = b$ 的唯一最稀疏解。

证明：考虑任何替代解 $A\tilde{x} = b$ 。同样，我们可以写成 $y = x - \tilde{x}$ in 这种情况下 $y \in \ker(A)$ 。然而 $\|y\|_0 \geq r + 1$ ，因为每组 r 列的 A 是线性独立的，根据假设。然后 $\|\tilde{x}\|_0 \geq \|y\|_0 - \|x\|_0 \geq r/2 + 1$ ，所以 \tilde{x} 比 x 有更多的非零元素，这完成了证明。 ■

事实上，如果 A 不连贯，我们可以对其Kruskal秩进行下界估计（因此，前一节的证明可以看作是这一节的特例）。

定理4.2.6 如果 A 是 μ -不连贯的，则 A 的列的Kruskal秩至少为 $1/\mu$ 。

证明：首先我们注意到，如果存在一组 I 中的 r 列是线性相关的，则 $A^T A$ 的 $I \times I$ 子矩阵必须是奇异的。因此，我们只需证明对于任意大小为 r 的 I ， $A^T A$ 的 $I \times I$ 子矩阵的秩是满秩的，其中 $r = 1/\mu$ 。

所以考虑任意这样的子矩阵。对角线为1，根据假设，非对角线的绝对值最多为 μ 。现在我们可以应用Gershgorin圆盘定理，并得出结论：子矩阵的特征值严格大于零，前提是 $1/\mu \leq r$ （这意味着任意一行的非对角线绝对值之和严格小于1）。证明完成。 ■

因此，我们可以将前一节中的唯一性结果扩展到任意不连贯的矩阵（而不仅仅是由两个正交基组成的矩阵）。请注意，这个界限与我们之前的界限相差两倍。

推论4.2.7 假设 A 是 μ -不连贯的。如果 $Ax = b$ 且 $\|x\|_0 < \frac{1}{2\mu}$ ，那么 x 是唯一最稀疏的解。

有许多算法可以在上述推论中恢复 x ，并且我们将介绍其中一种算法。

4.3 追踪算法

在这里，我们将介绍在 A 不连贯时解决稀疏恢复问题的算法。第一个这样的算法是匹配追踪，由Mallat和Zhang引入[93]；我们将分析正交匹配追踪[99]：

正交匹配追踪

输入：矩阵 $A \in \mathbb{R}^{n \times m}$ ，向量 $b \in \mathbb{R}^n$ ，期望的非零元素个数 $k \in \mathbb{N}$ 。

输出：具有最多 k 个非零元素的解 x 。

初始化： $x^0 = 0$ ， $r^0 = Ax^0 - b$ ， $S = \emptyset$ 。

对于 $\ell = 1, 2, \dots, k$

选择使 $|\langle A_j, r^{\ell-1} \rangle|$ 最大的列 j $\frac{\langle A_j, r^{\ell-1} \rangle}{\|A_j\|_2}$ 。

将 j 添加到 S 中。

设置 $r^\ell = \text{proj}_{U^\perp}(b)$ ，其中 $U = \text{span}(A_S)$ 。

如果 $r^\ell = 0$ ，则跳出循环。

结束

求解 x_S ： $A_S x_S = b$ 。设置 $x_{\bar{S}} = 0$ 。

设 A 为 μ -不相干的，并假设存在一个解 x ，其中 $k < 1/(2\mu)$ 非零条目，因此 x 是线性系统的唯一最稀疏解。

设 $T = \text{supp}(x)$ 。我们将证明正交匹配追踪能够恢复真实的解 x 。我们的分析基于为我们的算法建立以下两个不变量：

- (a) 算法选择的每个索引 j 都在 T 中。
- (b) 每个索引 j 最多被选择一次。

这两个属性立即意味着正交匹配追踪能够恢复真实的解 x ，因为残差误差 r^ℓ 在 $S = T$ 之前都不为零，并且线性系统 $A_T x_T = b$ 有唯一解（否则 x 将不是唯一最稀疏解，这与我们在前一节中证明的唯一性属性相矛盾）。

属性 (b) 很直观，因为一旦 $j \in S$ 在算法的每一步中，我们将有 $r^\ell \perp U$ ，其中 $U = \text{span}(A_S)$ ，所以如果 $j \in S$ ，则 $\langle r^\ell, A_j \rangle = 0$ 。我们的主要目标是建立属性 (a)，我们将通过归纳法证明。
主要引理如下：

引理 4.3.1 如果 $S \subseteq T$ 在一个阶段的开始时，算法会选择 $j \in T$ 。

我们首先证明一个辅助引理：

引理 4.3.2 如果 $r^{\ell-1}$ 在一个阶段的开始时在 T 中有支撑，那么算法会选择 $j \in T$ 。

证明：让 $r^{\ell-1} = \sum$ 然后我们可以重新排列 A 的列，使得前 k' 列对应于 y 的 k' 个非零元素，按照降序排列：

$$\underbrace{|y_1| \geq |y_2| \geq \cdots \geq |y_{k'}|}_{\text{对应于矩阵 } A \text{ 的前 } k' \text{ 列}} > 0, \quad |y_{k'+1}| = 0, |y_{k'+2}| = 0, \dots, |y_m| = 0.$$

因此，支持集 $(y) = \{1, 2, \dots, k'\} \subseteq T$ 。然后为了确保我们选择 $j \in T$ ，一个充分条件是

$$(4.1) \quad |\langle A_1, r^{\ell-1} \rangle| > |\langle A_i, r^{\ell-1} \rangle| \quad \text{对于所有 } i \geq k' + 1.$$

我们可以对(4.1)的左边进行下界估计：

$$|\langle r^{\ell-1}, A_1 \rangle| = \left| \left\langle \sum_{\ell=1}^{k'} y_{\ell} A_{\ell}, A_1 \right\rangle \right| \geq |y_1| - \sum_{\ell=2}^{k'} |y_{\ell}| |\langle A_{\ell}, A_1 \rangle| \geq |y_1| - |y_1| (k' - 1) \mu \geq |y_1| (1 - k' \mu + \mu),$$

在假设 $k' \leq k < 1/(2\mu)$ 的情况下，严格下界为 $|y_1|(1/2 + \mu)$ 。

我们可以对(4.1)的右侧进行上界估计：

$$|\langle r^{\ell-1}, A_i \rangle| = \left| \left\langle \sum_{\ell=1}^{k'} y_{\ell} A_{\ell}, A_i \right\rangle \right| \leq |y_1| \sum_{\ell=1}^{k'} |\langle A_{\ell}, A_i \rangle| \leq |y_1| k' \mu,$$

在假设 $k' \leq k < 1/(2\mu)$ 的情况下，它严格上界受到限制 $|y_1|/2$ 。由于 $|y_1|(1/2 + \mu) > |y_1|/2$ ，我们得出结论，条件(4.1)成立，保证算法选择 $j \in T$ ，从而完成引理的证明。

■

现在我们可以证明主要引理：

证明：假设在一个阶段的开始时， $S \subseteq T$ 。那么残差 $r^{\ell-1}$ 在 T 中有支持，因为我们可以将其写成

$$r^{\ell-1} = b - \sum_{i \in S} z_i A_i, \quad \text{其中 } z = \arg \min \|b - A_S z_S\|_2$$

应用上述引理，我们得出结论，算法选择 $j \in T$ 。 ■

这在正交匹配追踪的正确性证明中归纳地建立了属性 (a)，并完成了证明。请注意，该算法在确保唯一性的阈值上正好起作用。然而，在 $A = [U, V]$ 和 U 和 V 正交的情况下，我们证明了一个比最佳已知唯一性界限好两倍的唯一性结果。目前没有已知的算法能够达到那个最佳唯一性界限，尽管有比上述算法更好的算法（参见[55]）。

匹配追踪

我们注意到匹配追踪与正交匹配追踪在一个关键的方面不同：在后者中，我们在每个阶段的末尾重新计算系数 x_i for $i \in S$ ，因为我们将 b 投影到 U 的垂直方向。然而，我们可以希望当我们向 S 中添加一个新的索引 j 时，这些系数不需要进行太多调整。事实上，匹配追踪不会重新计算这些系数，因此在实践中更快。然而，分析更加复杂，因为我们需要跟踪误差如何累积。

4.4 Prony方法

本节的主要结果是，任何 k 稀疏信号都可以从其离散傅里叶变换的前 $2k$ 个值中恢复出来，这还有一个额外的好处，即我们可以使用FFT快速计算 Ax 。这个算法被称为普罗尼方法，可以追溯到1795年。这是 A 中行数与 x 的稀疏性上界之间的最优关系；然而，这种方法非常不稳定，因为它涉及求逆一个Vandermonde矩阵。

DFT的性质

在Prony的方法中，我们将关键地利用DFT的一些性质。回想一下，DFT矩阵的元素为：

$$F_{a,b} = \left(\frac{1}{\sqrt{n}} \right) \exp \left(\frac{i2\pi(a-1)(b-1)}{n} \right)$$

我们可以写成 $\omega = e^{i2\pi/n}$ ，然后第一行是 $\frac{1}{\sqrt{n}}[1, 1, \dots, 1]$ ；第二行是 $\frac{1}{\sqrt{n}}[1, \omega, \omega^2, \dots]$ ，等等。

我们将利用 F 的以下基本性质：

- (a) F 是正交的: $F^H F = F F^H$, 其中 F^H 是 F 的共轭转置
 (b) F 对卷积操作进行对角化

特别地, 我们将通过相应的线性变换来定义卷积操作:

定义 4.4.1 (循环矩阵) 对于一个向量 $c = [c_1, c_2, \dots, c_n]$, 令

$$M^c = \begin{bmatrix} c_n & c_{n-1} & c_{n-2} & \dots & c_1 \\ c_1 & c_n & c_{n-1} & \dots & c_2 \\ \vdots & & & & \vdots \\ c_{n-1} & \dots & \dots & \dots & c_n \end{bmatrix}.$$

然后我们可以将 $M^c x$ 定义为 c 和 x 的卷积结果, 表示为 $c * x$ 。很容易验证这与卷积的标准定义相符。

事实上, 我们可以使用 F 对 M^c 进行对角化。我们将使用以下事实, 不进行证明:

命题 4.4.2 $M^c = F^H \text{diag}(c) F$, 其中 $c = F c_0$ 。

因此, 我们可以将卷积视为傅里叶表示中的逐坐标乘法:

推论 4.4.3 设 $z = c * x$; 则 $z = c \cdot x$, 其中 \cdot 表示逐坐标乘法。

证明: 我们可以将 $z = M^c x = F^H \text{diag}(c) F x = F^H \text{diag}(c) x = F^H (c \cdot x)$ 写为 $z = F^H (c \cdot x)$, 从而完成了证明。 ■

为了描述 Prony 方法, 我们引入以下辅助多项式:

定义 4.4.4 (辅助多项式)

$$\begin{aligned} p(z) &= \prod_{b \in \text{supp}(x)} \omega^{-b} (\omega^b - z) \\ &= 1 + \lambda_1 z + \dots + \lambda_k z^k. \end{aligned}$$

声明 4.4.5 如果我们知道 $p(z)$, 我们可以找到 $\text{supp}(x)$ 。

证明：实际上，索引 b 在 x 的支持中当且仅当 $p(\omega b) = 0$ 。因此，我们可以在 ω 的幂处评估 p ，并且 p 评估为非零的指数正好是 x 的支持。

Prony 方法的基本思想是使用离散傅里叶变换的前 $2k$ 个值来找到 p ，从而找到 x 的支持。然后，我们可以解一个线性系统来实际找到 x 的值。

寻找辅助多项式

我们的第一个目标是找到辅助多项式。让

$$v = [1, \lambda_1, \lambda_2, \dots, \lambda_k, 0, \dots, 0], \text{ 且 } v = Fv$$

很容易看出，在索引 $b+1$ 处的 v 的值正好是 ω^b 的值。

命题 4.4.6 $\text{supp}(v) = \overline{\text{supp}(x)}$

也就是说， v 的零点对应于 p 的根，因此 x 的非零点。反之， v 的非零点对应于 x 的零点。我们得出结论： $xv = 0$ ，因此：

推论 4.4.7 $M^x v = 0$

证明：我们可以应用命题 4.4.2 将 $xv = 0$ 重写为 $x * v = 0 = 0$ ，从而得出该推论。 ■

让我们明确地写出这个线性系统：

$$M^x = \begin{bmatrix} x_n & x_{n-1} & \dots & x_{n-k} & \dots & x_1 \\ x_1 & x_n & \dots & x_{n-k+1} & \dots & x_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{k+1} & x_k & \dots & x_1 & \dots & x_{k+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \wedge x_{2k} & x_{2k-1} & \dots & x_k & \dots & x_{2k+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

回想一下，我们没有访问矩阵的所有条目，因为我们只有 x 的前 $2k$ 个值的 DFT。然而，考虑一下 $k \times k+1$ 的子矩阵

其左上角的值是 x_{k+1} ，右下角的值是 x_k 。这个矩阵只涉及我们已知的值！考虑一下

$$\begin{bmatrix} x_k & x_{k-1} & \dots & x_1 \\ \vdots & & & \\ x_{2k-1} & x_{2k-2} & \dots & x_k \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{bmatrix} = - \begin{bmatrix} x_{k+1} \\ \vdots \\ x_{2k} \end{bmatrix}$$

事实证明，这个线性系统是满秩的，所以 λ 是线性系统的唯一解（证明留给读者¹）。 λ 中的条目是 p 的系数，所以一旦我们解出 λ ，我们可以在 ω^b 上评估辅助多项式，以找到 x 的支撑。剩下的就是找到 x 的值。实际上，让 M 是 F 在 S 和其前 $2k$ 行的列上的限制。 M 是一个范德蒙德矩阵，所以再次 $M x_S = x_{1,2,\dots,2k}$ 有唯一解，我们可以解这个线性系统找到 x 的非零值。

4.5 压缩感知

在这里，我们将为恢复一个几乎线性的信号 x 提供稳定的算法，该信号在感知矩阵的行数上有一个几乎线性的非零数。回想一下， A 的 Kruskal 秩决定了我们可以允许在 x 中有多少非零值，并且仍然使 x 成为 $Ax = b$ 的唯一最稀疏的解。随机矩阵具有较大的 Kruskal 秩，而对于压缩感知，我们将需要 Kruskal 秩的一个稳健类比：

定义 4.5.1 矩阵 A 如果对于所有 k 稀疏向量 x ，都满足 RIP 常数 δ_k ，则矩阵 A 是 RIP 的。

$$(1 - \delta_k) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_k) \|x\|_2^2$$

如果 A 是一个随机 $m \times n$ 矩阵，其中每个元素都是独立的高斯分布 ($N(0, 1)$) 那么我们可以选择 $m \approx k \log n/k$ 和 $\delta_k = 1/3$ 。接下来，我们将明确稀疏恢复的目标：

定义 4.5.2 $\sigma_k(x) = \min_{\|w\|_0 \leq k} \|x - w\|_1$

即 $\sigma_k(x)$ 是 x 除了 k 个最大元素外的 ℓ_1 和。特别地，如果 $\|x\|_0 \leq k$ 那么 $\sigma_k(x) = 0$ 。

我们的目标是找到一个 w ，使得 $\|x - w\|_1 \leq C\sigma_k(x)$ 从少量 ($\tilde{O}(k)$) 的测量中恢复。请注意，我们不要求 w 是 k 稀疏的。然而，如果 x 恰好是 k 稀疏的，那么满足上述条件的任何 w 必须与 x 完全相等，因此这个新的恢复目标包含了我们之前讲座中的精确恢复目标（并且实际上更强大）。

自然（但难以处理）的方法是：

$$(P0) \quad \min \|w\|_0 \text{ s.t. } Aw = b$$

由于这在计算上很难解决（对于所有 A ），我们将使用 ℓ_1 放松：

$$(P1) \quad \min \|w\|_1 \text{ s.t. } Aw = b$$

我们将证明在这个优化问题的解 w 满足 $w = x$ （或 $\|x - w\|_1 \leq C\sigma_k(x)$ ）的条件。

定理4.5.3[35] 如果 $\delta_{2k} + \delta_{3k} < 1$ ，那么如果 $\|x\|_0 \leq k$ ，我们有 $w = x$ 。

定理4.5.4[34] 如果 $\delta_{3k} + 3\delta_{4k} < 2$ ，那么

$$\|x - w\|_2 \leq \frac{C}{\sqrt{k}} \sigma_k(x)$$

请注意，这将以最佳逼近的 ℓ_1 error 为条件，限制了误差 $x - w$ 的 ℓ_2 范数。

定理4.5.5[40] 如果 $\delta_{2k} < 1/3$ ，那么

$$\|x - w\|_1 \leq \frac{2 + 2\delta_{2k}}{1 - 3\delta_{2k}} \sigma_k(x)$$

我们将遵循[80]的证明，大大简化了分析的类型，并将压缩感知与几乎欧几里得子节-段之间的联系变得明确。从这个角度来看，与纠错码进行类比将更容易。

几乎欧几里得子节

设 $\Gamma = \ker(A)$ 。我们将利用 Γ 的某些几何性质（几乎必然成立）来证明基 Pursuit 的有效性：

定义4.5.6 $\Gamma \subseteq \mathbb{R}^n$ 是一个几乎欧几里得的子集, 如果对于所有的 $v \in \Gamma$,

$$\frac{1}{\sqrt{n}}\|v\|_1 \leq \|v\|_2 \leq \frac{C}{\sqrt{n}}\|v\|_1$$

请注意不等式 $\frac{1}{\sqrt{n}}\|v\|_1 \leq \|v\|_2$ 对于所有的向量成立, 因此第二个不等式是重要的部分。我们要求的是在重新缩放后, ℓ_1 和 ℓ_2 范数几乎等价。

问题8 如果一个向量 v 满足 $\|v\|_0 = o(n)$, 那么 v 可以在 Γ 中吗?

不可以! 任何这样的向量 v 都会有 $\|v\|_1 = o(\sqrt{n})\|v\|_2$, 根据柯西-施瓦茨不等式。让我们从几何角度来思考这些子集。考虑 ℓ_1 范数的单位球:

$$B_1 = \{v \mid \|v\|_1 \leq 1\}$$

这被称为交叉多面体, 是向量 $\{\pm e_i\}_i$ 的凸包, 其中 e_i 是标准基向量。然后 Γ 是一个子空间, 当与 B_1 相交时, 结果是一个接近于球体 B_2 的凸体, 经过重新缩放。

事实上, 自从[63]的工作以来, 已经知道选择 Γ 在随机情况下, $\dim(\Gamma) \geq n - m$, 我们可以选择 $C = O(\sqrt{\log n/m})$ 几乎肯定 (在这种情况下, 它是一个 $m \times n$ 矩阵 A 的核, 它将是我们的感知矩阵)。在讲座的剩余部分中, 我们将建立 Γ 的各种几何属性, 为压缩感知做好准备。

Γ 的属性

在本节中, 令 $S = n/C^2$ 。

定理4.5.7 令 $v \in \Gamma$, 那么要么 $v = 0$, 要么 $|\text{supp}(v)| \geq S$ 。

证明:

$$\|v\|_1 = \sum_{j \in \text{supp}(v)} |v_j| \leq \sqrt{|\text{supp}(v)|} \cdot \|v\|_2 \leq \sqrt{|\text{supp}(v)|} \frac{C}{\sqrt{n}} \|v\|_1$$

其中最后一个不等式使用了 Γ 几乎是欧几里得的性质。最后一个不等式证明了该命题。 ■

现在我们可以将其类比为纠错码。回想一下，我们希望 $C \subseteq \{0, 1\}^n$ 。而速率 R 是 $R = \log |C|/n$ ，相对距离 δ 是

$$\delta = \frac{\min_{x \neq y \in C} d_H(x, y)}{n}$$

其中 d_H 是汉明距离。目标是找到一个编码，其中 $R, \delta = \Omega(1)$ 并且易于编码和解码。在线性码的特殊情况下，例如 $C = \{y | y = Ax\}$ 其中 A 是一个 $n \times Rn$ $\{0, 1\}$ 值矩阵，而 $x \in \{0, 1\}^{Rn}$ 。然后

$$\delta = \frac{\min_{x \neq 0 \in C} \|x\|_0}{n}$$

因此，对于纠错码，我们希望找到具有大（线性）维度的子空间其中每个向量具有线性数量的非零元素。在压缩感知中，我们也希望 Γ 具有这个属性，但更重要的是，我们希望它的 ℓ_1 norm 也均匀分布（例如，大多数非零坐标都很大）。

定义 4.5.8 对于 $\Lambda \subseteq [n]$ ，让 v_Λ 表示 v 在 Λ 坐标上的限制。同样，让 v^Λ 表示 v 在 Λ^c 上的限制。

声明 4.5.9 假设 $v \in \Gamma$ 且 $\Lambda \subseteq [n]$ 且 $|\Lambda| < S/16$ 。那么

$$\|v_\Lambda\|_1 < \frac{\|v\|_1}{4}$$

证明：

$$\|v_\Lambda\|_1 \leq \sqrt{|\Lambda|} \|v_\Lambda\|_2 \leq \sqrt{|\Lambda|} \frac{C}{\sqrt{n}} \|v\|_1$$

■

因此，向量在 Γ 中不仅具有线性数量的非零元素，而且其 ℓ_1 norm 是分散的。现在我们准备证明当 $\|x\|_0$ 足够小（但几乎是线性的）时，(P1) 可以完全恢复 x 。下一节课我们将证明它也是稳定的（使用我们在 Γ 上建立的属性）。

引理 4.5.10 设 $w = x + v$ 且 $v \in \Gamma$ 其中 $\|x\|_0 \leq S/16$ 。那么 $\|w\|_1 > \|x\|_1$ 。

证明：设 $\Lambda = \text{supp}(x)$ 。

$$\|w\|_1 = \|(x + v)_\Lambda\|_1 + \|(x + v)^\Lambda\|_1 = \|(x + v)_\Lambda\|_1 + \|v^\Lambda\|_1$$

现在我们可以使用三角不等式：

$$\|w\|_1 \geq \|x_\Lambda\|_1 - \|v_\Lambda\|_1 + \|v^\Lambda\|_1 = \|x\|_1 - \|v_\Lambda\|_1 + \|v^\Lambda\|_1 = \|x_\Lambda\|_1 - 2\|v_\Lambda\|_1 + \|v\|_1$$

然而 $\|v\|_1 - 2\|v_\Lambda\|_1 \geq \|v\|_1/2 > 0$ 使用上述断言。这意味着引理。 ■

因此，我们可以使用几乎欧几里得子空间来实现精确的稀疏恢复，直到

$$\|x\|_0 = S/16 = \Omega(n/C^2) = \Omega\left(\frac{n}{\log n/m}\right)$$

接下来，我们将考虑稳定恢复。我们的主要定理是：

定理 4.5.11 令 $\Gamma = \ker(A)$ 为具有参数

C 的几乎欧几里得子空间。令 $S = \frac{n}{C^2}$ 。如果 $Ax = Aw = b$ 并且 $\|x\|_1 \leq \|x\|_1$ 我们有

$$\|x - w\|_1 \leq 4\sigma_{\frac{S}{16}}(x).$$

证明： 让 $\Lambda \subseteq [n]$ 是最大化 $\|x_\Lambda\|_1$ 的 $S/16$ 坐标的集合。我们想要上界 $\|x - w\|_1$ 。通过反复应用三角不等式， $\|w\|_1 = \|w^\Lambda\|_1 + \|w_\Lambda\|_1 \leq \|x\|_1$ 和 $\sigma_t(\cdot)$ 的定义，可以得出

$$\begin{aligned} \|x - w\|_1 &= \|(x - w)_\Lambda\|_1 + \|(x - w)^\Lambda\|_1 \\ &\leq \|(x - w)_\Lambda\|_1 + \|x^\Lambda\|_1 + \|w^\Lambda\|_1 \\ &\leq \|(x - w)_\Lambda\|_1 + \|x^\Lambda\|_1 + \|x\|_1 - \|w_\Lambda\|_1 \\ &\leq 2\|(x - w)_\Lambda\|_1 + 2\|x^\Lambda\|_1 \\ &\leq 2\|(x - w)_\Lambda\|_1 + 2\sigma_{\frac{S}{16}}(x). \end{aligned}$$

由于 $(x - w) \in \Gamma$ ，我们可以应用 Claim 4.5.9 得出结论 $\|(x - w)_\Lambda\|_1 \leq \frac{1}{4}\|x - w\|_1$ 。因此

$$\|x - w\|_1 \leq \frac{1}{2}\|x - w\|_1 + 2\sigma_{\frac{S}{16}}(x).$$

证明完成。 ■

请注意，在上述论证中， Γ 的几何特性起到了主要作用。有很多证明表明基础追踪是有效的，但我们在这里提出的优势是它阐明了有限域上的经典误差纠正理论与稀疏恢复之间的联系。

矩阵 A 在这里扮演了纠错码的奇偶校验矩阵的角色，而

因此，它的核对应于码字。然而，在实数情况下存在更多的细微差别：与有限域设置相反，在实数情况下有许多不同范数之间的相互作用，产生了一些有限域情况下不存在的现象。

事实上，该领域的一个核心开放问题是给出一个确定性的构造RIP矩阵的方法：

问题9（开放）是否存在RIP矩阵的明确构造，或者等价地，几乎欧几里德子空间 Γ 的构造？

相比之下，有许多明确的渐近好码的构造方法。目前已知的最佳确定性构造方法是由Guruswami、Lee和Razborov提出的：

定理4.5.12[69] 存在一个多项式时间的确定性算法，用于构造参数为 $C \sim (\log n)^{\log \log n}$ 的几乎欧几里德子空间 Γ 。

我们注意到，实现较弱的保证是很容易的，例如 $\forall v \in \Gamma, \text{supp}(v) = \Omega(n)$ ，但这些对于压缩感知来说是不够的，因为我们还要求 ℓ_1 权重也要分散开。

第五章

字典学习

在本章中，我们将学习字典学习，其中我们给出了许多示例 b_1, b_2, \dots, b_p ，我们希望找到一个字典 A ，使得每个示例都可以写成 A 中列的稀疏线性组合。

5.1 背景

在前一章中，我们研究了用于稀疏恢复的算法。然而，在许多应用中，字典 A 是未知的，必须从示例中学习。另一个动机是经常手动设计一些特征族作为适当的字典，例如正弦波、小波、脊波、曲线波等[94]，每个特征族在不同的环境中都有用：正弦波用于脉冲事件，脊波用于边缘的不连续性，曲线波用于平滑曲线等。但是，对于一个新的数据集，我们能否自动学习到正确的特征族来表示信号？

回想一下，对于任意线性系统，找到最稀疏的解是 NP -难的。然而，对于一些重要的问题类别，问题可以被高效地解决。同样地，我们不能指望对于任意字典都能解决字典学习问题。相反，我们将专注于在我们已经知道可以解决相应的稀疏恢复问题的设置中设计算法来进行字典学习。毕竟，找到一个字典只有在它能够使用它的情况下才有用。我们可以进行稀疏恢复的三个最重要的情况是：

- (a) A 具有满列秩
- (b) A 是不相干的

(c) A 是RIP

我们将介绍Spielman、Wang和Wright [108]的算法，如果 A 是满秩的，并且每个 b_i 是 A 中最多 $\widetilde{O}(\sqrt{n})$ 列的线性组合，那么该算法将成功（在合理的分布假设下）。接下来，我们将给出Arora、Ge和Moitra [15]的算法，用于不相干和过完备的情况，该算法也可以成功地处理 $\widetilde{O}(\sqrt{n})$ 稀疏度。我们注意到，Agarwal等人 [2], [3]最近独立地提出了字典学习的替代算法，其工作范围为 $\widetilde{O}(n^{1/4})$ 的较弱稀疏度界限。

历史

字典学习问题由Olshausen和Field [97]引入，他们对哺乳动物视觉皮层的各种特性感兴趣。神经科学家经常测量神经元的感受野，即神经元对各种刺激的响应方式。事实上，他们的响应被广泛认为是局部化、带通和定向的，Olshausen和Field观察到，如果学习一组自然图像的字典，字典的元素通常具有许多相同的特性。他们的工作表明，理解视觉系统的一个重要步骤可能是确定其用于表示自然图像的基础。

字典学习，通常被称为稀疏编码，是许多机器学习系统的基本构建模块。这种算法原语在去噪、边缘检测、超分辨率和压缩等应用中广泛应用。

字典学习也被用于一些深度学习架构的设计中。

在实践中解决这个问题的流行方法是标准交替最小化方法的变体。假设将 (x_i, b_i) 对收集到矩阵 X 和 B 的列中。然后我们的目标是以一种使 X 的列稀疏的方式从 B 中计算 A 和 X 。

最优方向法[56]：从一个初始猜测的 A 开始，然后交替更新 A 或 X ：

- 给定 A ，计算一个稀疏的 X ，使得 $AX \approx B$ （使用匹配追踪[93]或基追踪[39]等方法）
- 给定 X ，计算最小化 $\|AX - B\|_F$ 的 A

这个算法收敛到一个局部最优解，因为每一步中误差 $\|AX - Y\|_F$ 只会减小。

K-SVD [5] 从一个初始猜测开始 A . 然后重复以下步骤:

- 给定 A , 使用一种追踪方法计算一个稀疏的 X 使得 $AX \approx B$
- 将所有数据点 $B^{(1)}$ 分组, 其中相应的 X 向量在索引 i 处非零
· 减去其他方向上的分量

$$B^{(1)} - \sum_{j=i} A_j X_j^{(1)}$$

- 计算残差矩阵的第一个奇异向量 v_1 , 并更新列 A_i 为 v_1

实际上, 这些算法对于其起始条件非常敏感。在接下来的章节中, 我们将介绍具有可证明保证的字典学习的替代方法。回想一下, 有些字典很容易解决稀疏恢复问题, 而有些则很困难。因此, 在这里更喜欢可证明的算法的一个重要理由是, 如果示例是从一个容易的字典生成的, 我们的算法将真正学习到一个容易的字典 (但对于上述启发式方法来说, 情况可能并非如此)。

5.2 全秩字典

在这里, 我们介绍了 Spielman、Wang 和 Wright [108] 的一个最新算法, 在 A 具有完全列秩且 X 具有 $O(\sqrt{n})$ 个非零元素的特定分布条件下起作用。首先, 我们定义模型:

模型

[108] 中的分布模型实际上比我们在这里使用的模型更一般; 我们这样做只是为了简化分析和符号表示, 但证明几乎相同。让 $\theta \in [1$

成: $\frac{1}{n}, \frac{1}{\sqrt{n \log n}}]$ 是稀疏参数。然后, 向量 x_i 根据以下过程生

(a) x_i 的每个坐标 (独立地) 以概率 θ 选择为非零值

(b) 如果第 j 个坐标是非零的, 则其值从标准高斯分布 $\mathcal{N}(0, 1)$ 中采样 (独立地)

我们观察样本 $b_i = Ax_i$. 字典 A 是一个未知的可逆 $n \times n$ 矩阵, 我们需要从形如 b_i 的多项式数量的样本中恢复 A 。

定理5.2.1[108] 根据上述模型, 存在一个多项式时间算法可以准确学习一个满秩的字典 A , 给定多项式数量的样本。

严格来说, 如果 x_i 的坐标完全独立, 那么我们可以使用可证明的独立成分分析[62]算法来恢复 A , 但是字典学习算法更普遍地适用于坐标不独立的情况。

大纲

基本思想是考虑 B 的行空间, 我们用 $R = \{w^T B\}$ 来表示。注意 $A^{-1}B = X$, 因此 X 的行包含在 R 中。关键观察是:

观察5.2.2 X 的行是 R 中最稀疏的非零向量。

当然, 在子空间中找到最稀疏的非零向量是困难的, 所以我们将考虑类似于稀疏恢复的 ℓ_1 -松弛。

考虑以下四个优化问题:

$$(P0) \quad \text{最小化 } \|w^T B\|_0 \text{ 约束 } w = 0$$

这个优化问题要求在 R 中找到最稀疏 (非零) 的向量。然而, 它是 NP 困难的。我们将考虑以下问题:

$$(P1) \quad \text{最小化 } \|w^T B\|_1 \text{ 约束 } r^T w = 1$$

约束条件 $r^T w$ 固定了一个归一化, 我们将在后面解释如何选择 r 。上述优化问题可以高效地求解, 因为它是一个线性规划问题。我们希望展示它的最优解 $w^T B$ 是 X 的一行 (经过缩放)。实际上, 我们可以将上述线性规划问题转化为一个更简单的问题, 以便更容易分析:

$$(Q1) \quad \text{最小化 } \|z^T X\|_1 \text{ 约束条件 } c^T z = 1$$

这个线性规划的解与 $(P1)$ 的解之间存在一个双射, 由 $z = A^T w$ 和 $c = A^{-1}r$ 给出。让我们将 r 设置为 B 的一列, 因此 c 是对应的 X 的一列, 并且是稀疏的。现在我们可以解释这个直觉:

- 我们将证明 $(Q1)$ 的解是稀疏的，特别是 $\text{supp}(z) \subseteq \text{supp}(c)$ 。
- 对于稀疏的 z ，我们有 $\|z^T X\|_1 \approx \|z\|_1$ （经过适当的缩放）。因此，我们可以分析线性规划问题：

$$(Q1') \quad \min \|z\|_1 \text{ 约束条件 } c^T z = 1$$

注意，如果 c 的坐标中有一个绝对值最大，则 $|\text{supp}(z)| = 1$ 。回想一下，我们选择 c 作为 X 的一列，在我们的分布模型中， X 中的非零元素是高斯分布的。因此，几乎可以确定 c 的绝对值最大的坐标 c_i ，如果我们解 $(P1)$ 目标函数中的向量（即 $w^T B$ ），它将是 X 的第 i 行，只需重新缩放即可。

- 最后，我们可以反复解 $(P1)$ 并找到 X 的所有行，在修正它们的缩放后，我们可以通过计算 $A = BX^+$ 来解 A 。

第一步

假设 z_* 是 $(Q1)$ 的最优解，其中 $c = x_i$ 是 X 的一列。设 $J = \text{supp}(c)$ ，我们可以将 z_* 写成 $z_0 + z_1$ ，其中 z_0 在 J 上支持， z_1 在 J 上支持。我们希望证明 z_0 至少与 z_* 一样好作为 $(\bar{Q}1)$ 的解。特别地，我们想要证明

$$\|z_0^T X\|_1 < \|z_*^T X\|_1$$

定义 5.2.3 如果 R 是一组行， C 是一组列，那么 X_{CR} 是 X 的行和列的交集的子矩阵。

令 S 是 X 中具有 J 中非零元素的列的集合。即 $S = \{j | X_j^J \neq \vec{0}\}$ 。我们现在计算：

$$\begin{aligned} \|z_*^T X\|_1 &= \|z_*^T X_S\|_1 + \|z_*^T X_{\bar{S}}\|_1 \\ &\geq \|z_0^T X_S\|_1 - \|z_1^T X_S\|_1 + \|z_1^T X_{\bar{S}}\|_1 \\ &\geq \|z_0^T X\|_1 - 2\|z_1^T X_S\|_1 + \|z_1^T X\|_1 \end{aligned}$$

还需要证明

$$(5.1) \quad \|z_1^T X\|_1 - 2\|z_1^T X_S\|_1 > 0$$

假设 z_1 是固定的（我们可以对这个空间的 ε -网应用一个联合边界来完成论证）。然后 S 是一个随机集合，我们可以计算：

$$\mathbb{E}[\|z_1^T X_S\|_1] = \frac{|S|}{p} \mathbb{E}[\|z_1^T X\|_1]$$

集合 S 的期望大小是 $p \times \mathbb{E}[\text{supp}(x_i)] \times \theta = \theta^2 np = o(p)$ 。综上所述，这意味着

$$\mathbb{E}[\|z_1^T X\|_1 - 2\|z_1^T X_S\|_1] = \left(1 - \frac{2\mathbb{E}[|S|]}{p}\right) \mathbb{E}[\|z_1^T X\|_1]$$

远离零，并且对于任意固定的 z_1 ，(5.1) 以很高的概率成立。我们可以对所有可能的 z_1 的 ε -网进行并集约束，并得出结论(5.1)对于所有 z_1 成立。这反过来意味着如果 z_1 非零，则 $\|z_0^T X\|_1 < \|z_*^T X\|_1$ ，但这与我们假设 z_* 是 (Q_1) 的最优解矛盾。我们得出结论 z_1 为零，因此 $\text{supp}(z_*) \subseteq \text{supp}(c)$ ，如所需。

第二步

我们希望证明：

$$\|z^T X\|_1 \approx \|z^T\|_1$$

只要 $\text{supp}(z)$ 足够小，就可以得到某个缩放因子。回顾一下前一步，我们知道 $\text{supp}(z) \subseteq \text{supp}(c) = J$ 且 $|J| \leq \widetilde{O}(\theta n)$ 。注意， X^J 的典型列在期望中有 $\theta|J| = \theta^2 n = o(1)$ 个非零元素。这意味着，高概率下， X^J 中的绝大多数列最多只有一个非零元素。很容易看出：

命题 5.2.4 如果 X^J 的每一列最多只有一个非零元素，那么

$$\mathbb{E}[\|z_J^T X^J\|_1] = C \frac{p}{|J|} \|z_J\|_1$$

其中 C 是 X 中非零元素的期望绝对值。

因此，我们可以通过限制 X^J 中具有两个或更多非零元素的列的贡献来建立第二步。因此，这完成了证明。（这一步的更简洁描述是，由于 $|J|$ 为 \sqrt{n} ，且 X 的每一列都有 \sqrt{n} 个非零元素，矩阵 X^J 的作用类似于一个缩放后的单位矩阵）

。

第三步

现在我们可以把所有东西放在一起。由于 $c = x_i$ ，当我们求解 $(P1)$ 时，我们将得到 X 的第 i 行（经过缩放）。如果我们重复求解 $(P1)$ ，我们将找到 X 的所有行（并且可以删除重复项，因为现在 X 的两行将是缩放的副本）。

最后，我们可以计算出正确的缩放（符号上）例如通过使用假设非零条目分布为 $\mathcal{N}(0,1)$ 。因此，我们可以解出 X （行的符号可以翻转），如果 p 足够大（即我们取足够多的样本），则 X 将具有左伪逆，我们可以计算 $A = BX^+$ ，这将恢复 A （最好的情况是对其列进行排列和翻转符号）。

5.3 超完备字典

在这里，我们将介绍 Arora、Ge 和 Moitra [15] 的最新算法，该算法适用于不相干和过完备的字典。该算法背后的关键思想是与重叠聚类问题的联系。我们将考虑一个与 [108] 中类似的模型。设 k 为稀疏参数：

- (a) x_i 的支持是从 $[m]$ 的所有大小为 k 的子集中均匀随机选择的。
- (b) 如果第 j 个坐标非零，则其值独立选择为 $+1$ 或 -1 （概率相等）

[15] 中的主要结果是：

定理 5.3.1/[15] 如果从上述模型中获得样本，则存在一个多项式时间算法可以精确地学习 A ，如果 A 是 μ -不一致的，并且 $k \leq c \min(\sqrt{n/\mu \log n}, m^{1/2-\eta})$ 。运行时间和样本复杂度是多项式 (n, m) 。

回想一下，像 K-SVD 这样的方法依赖于这样的直觉：如果我们有真实的字典 A ，我们可以找到 X ，如果我们有 X ，我们可以找到对 A 的良好近似。然而，分析这些方法的困难在于它们从一个与真实字典相距甚远的字典开始，那么这些算法如何仍然取得进展呢？[15] 中算法的基本思想是通过首先找到 X 的支持集合而不知道字典 A 来打破循环。

我们的第一步是构建一个图，我们将其称为交集图：

定义5.3.2 给定 p 个样本 b_1, b_2, \dots, b_p , 交集图 G 是一个图在 $V = \{b_1, b_2, \dots, b_p\}$ 上, (b_i, b_j) 是一条边, 当且仅当 $|\langle b_i, b_j \rangle| > \tau$

我们应该如何选择 τ ? 我们希望满足以下属性。在本节中, 我们将让 S_i 表示 x_i 的支持。

(a) 如果 (i, j) 是一条边, 则 $S_i \cap S_j = \emptyset$

(b) 如果 $S_i \cap S_j = \emptyset$, 则 (i, j) 是一条边的概率至少为 $\frac{1}{2}$

给定 X 中的某一行, 我们可以将该行非零的所有示例视为属于同一个社区。那么上述图的解释是, 如果存在一条边连接 (i, j) , 我们希望 i 和 j 必然属于同一个社区。此外, 如果 (i, j) 确实属于同一个社区, 那么也应该相当有可能存在一条边。

我们可以根据我们的示例直接计算上述图, 基本思想是通过找到社区来希望学习 X 的支持。关键点在于, 这与标准聚类问题不同, 因为每个样本 x_i 具有 k 个非零值, 所以每个节点属于多项式数量的社区, 因此我们要寻找的是重叠聚类。

交集图

在这里, 我们选择如何设置 τ , 以使上述属性以很高的概率成立。我们可以限制

$$|\langle b_i, b_j \rangle - \langle x_i, x_j \rangle| \leq |S_i| |S_j| \mu = k^2 \mu$$

因此如果 $k^2 \mu < \frac{1}{2}$ 我们选择 $\tau = \frac{1}{2}$ 我们肯定满足条件 (a) 上面。此外, 很容易看出, 如果 S_i 和 S_j 相交, 则 $\langle x_i, x_j \rangle$ 非零的概率至少为 $\frac{1}{2}$ 同样, 这导致条件 (b). 然而, 由于 μ 对于随机字典来说大约是 $1/\sqrt{n}$ (忽略多项式对数因子), 我们需要 $k < n^{1/4}$, 这是一个比解决稀疏恢复问题所需条件更严格的条件。

上述分析是浪费的, 因为它没有利用 X 中非零元素的随机符号. 相反, 我们将依靠以下集中不等式:

引理 5.3.3 (Hanson-Wright) 设 x 为一个分量为独立的亚高斯随机变量的向量, 满足 $E[X_i] = 0$ 和 $Var[X_i] = 1$.

设 M 为一个对称的 $n \times n$ 矩阵。那么，对于每个 $t \geq 0$ 我们有：

$$\mathbb{P}[|x^T M x - \text{trace}(M)| > t] \leq 2 \exp\left(-c \min\left(\frac{t^2}{\|M\|_F^2}, \frac{t}{\|M\|_2}\right)\right)$$

设 S_i 和 S_j 为不相交的集合。令 $N = (A^T A)_{ij}^S$ 和

$$M = \begin{bmatrix} 0 & \frac{1}{2}N \\ \frac{1}{2}N^T & 0 \end{bmatrix},$$

令 y 为将 x_i 限制在 S_i 和 x_j 限制在 S_j 的向量的连接。那么 $y^T M y = \langle b_i, b_j \rangle$ 并且我们可以利用上述引理来限制 $\langle b_i, b_j \rangle$ 与其期望值的偏差。特别地， $\text{trace}(M) = 0$ 和 $\|M\|_F^2 \leq \mu^2 k^2$ 这意味着 $\|M\|_2 \leq \mu k$ 。

因此，如果 $k \leq \frac{1}{\mu \log n}$ 因此，高概率下我们有 $|\langle b_i, b_j \rangle| \leq \frac{1}{3}$ 。
当 S_i 和 S_j 相交时，相同的论证也适用（在这种情况下，我们将在 A 中对应于相同列的条目置零）。因此，如果我们利用 X 的符号随机性，交集图（对于 $\tau = \frac{1}{2}$ ）满足我们所需的属性，前提是 k 是一个比我们解决稀疏恢复所需的对数 n 因子小。即使我们知道 A 。

社区发现

考虑社区 $C_j = \{b_i | S_i \ni j\}$ 。那么对于每对 $b_1, b_2 \in C_j$ ，至少有概率 $\frac{1}{2}$ 此外，我们的交集图可以由 m 个密集社区 $\{C_j\}_j$ 覆盖。我们将通过以下思维实验介绍基本方法：

假设我们在 G 中找到了一个三角形，其中包含 b_1 、 b_2 和 b_3 。我们知道 $S_1 \cap S_2$ 、 $S_1 \cap S_3$ 和 $S_2 \cap S_3$ 都是非空的，但我们如何确定 $S_1 \cap S_2 \cap S_3$ 是否也是非空的呢？

或者，给定三个节点，其中每对节点都属于同一个社区，我们如何知道这三个节点是否都属于同一个社区？直觉告诉我们，如果 b_1 、 b_2 和 b_3 都属于一个共同的社区，那么一个随机的新节点 b_4 与它们都连接的概率要高于它们不连接的概率！我们需要的是在 $S_1 \cap S_2 \cap S_3$ 非空的情况下， b_4 连接到 b_1 、 b_2 和 b_3 的概率的下

界，以及在交集为空的情况下的上界。很容易看出：

声明 5.3.4 如果 $S_1 \cap S_2 \cap S_3 = \emptyset$ 那么 (b_4, b_i) 是 $i = 1, 2, 3$ 的边, 概率至少为 $\Omega(k/m)$

这个声明是正确的, 因为如果 $i \in S_1 \cap S_2 \cap S_3$, 那么 S_4 包含 i 的概率是 k/m 。接下来我们证明上界。令 $a = |S_1 \cap S_2|$, $b = |S_1 \cap S_3|$ 和 $c = |S_2 \cap S_3|$ 。那么:

引理 5.3.5 如果 $S_1 \cap S_2 \cap S_3 = \emptyset$, 那么 (b_4, b_i) 是 $i = 1, 2, 3$ 的边的概率最多为

$$O\left(\frac{k^6}{m^3} + \frac{k^3(a+b+c)}{m^2}\right)$$

证明: 我们知道, 如果 (b_4, b_i) 是一个边, 对于 $i = 1, 2, 3$, 那么 $S_4 \cap S_i$ 必须是非空的。对于 $i = 1, 2, 3$, 我们可以将这个事件分解成子情况:

(a) 要么 S_4 与 S_1 , S_2 和 S_3 不相交 (即它不包含任何在其他集合中的索引)

(b) 要么存在一对 $i = j \in \{1, 2, 3\}$, 使得 $S_i \cap S_j$ 与 S_4 相交, 并且 S_4 与剩余的集合在另一个索引处相交

(a) 的概率最多是 $|S_4 \cap (S_1 \cup S_2 \cup S_3)| \geq 3$, 这个概率最多是 $(3k)^3 \frac{k^3}{m^3}$ 。同样, (b) 的概率对于 $(i = 1, j = 2)$ 最多是 $\frac{ak^3}{m^2}$ 。这意味着引理成立。 ■

我们需要在 Claim 5.3.4 中的下界渐近小于引理 5.3.5 中的上界。很容易看出, 对于任意的 i, j , $|S_i \cap S_j| = O(1)$, 因此我们希望(大致)

$$\frac{k}{m} \gg \frac{k^6}{m^3} + \frac{k^3}{m^2}$$

当 $k < m^{2/5}$ 时, 这是成立的。当这个条件成立时, 对于每个三元组 b_1, b_2, b_3 , 我们将能够通过计数其他多少个节点 b_4 与它们都有边来确定 $S_1 \cap S_2 \cap S_3$ 是否为空的概率很高。

现在我们准备给出一个找到社区的算法。

CommunityFinding [15]输入：交集图 G 输出：社区 $\{C_j\}$ 对于每条边 (b_1, b_2) 设 $C_{1,2} = \{b_3 | S_1 \cap S_2 \cap S_3 = \emptyset\} \cup \{b_1, b_2\}$

结束

移除严格包含另一个集合 $C_{i,j}$ 的集合 $C_{1,2}$

定义 5.3.6 如果 $S_1 \cap S_2 = \{j\}$ 我们将称这对 (b_1, b_2) 为一个识别对
对于社区 j

很容易看出，如果样本数 p 足够大，每个社区都会有一个识别对。然后考虑

$$C_{1,2} = \{b_3 | S_1 \cap S_2 \cap S_3 = \emptyset\} \cup \{b_1, b_2\}$$

如果 (b_1, b_2) 是社区 j 的识别对，则上述集合 $C_{1,2}$ 恰好是 C_j
(我们可以通过上述测试来判断是否
 $S_1 \cap S_2 \cap S_3$ 为空来计算该集合)。

此外，如果 (b_1, b_2) 不是识别对，而是 $S_1 \cap S_2$ 有多个元素，那么我们将有 $C_{1,2} = \bigcup_{j \in S_1 \cap S_2} C_j$ 在这种情况下，集合 $C_{1,2}$ 将在最后一步被删除。如果 $k \leq c \min(\sqrt{n/\mu \log n}, m^{2/5})$ ，则该算法以很高的概率输出正确的社区。在[15]中，作者提供了一个更高阶的社区发现算法，当 $k \leq m^{1/2-\eta}$ (其中 $\eta > 0$) 时，该算法有效，但运行时间是一个多项式，其指数取决于 η 。

恢复真实字典的唯一剩余任务是将社区 C_j 划分为那些第 j 个坐标为 +1 的部分和那些第 j 个坐标为 -1 的部分。实际上，如果 $S_1 \cap S_2 = \{j\}$ ，那么 $\langle b_1, b_2 \rangle$ 的符号告诉我们它们的第 j 个坐标是否具有相同的符号或不同的符号。事实上，很容易证明在一个典型的社区中有足够多的这样的对，我们可以成功地将 C_j 划分为两个集合，其中一个集合中的所有示例的第 j 个坐标等于 +1，而另一个集合中的所有示例的第 j 个坐标等于 -1 (当然我们不知道哪个是哪个)。这反过来又允许我们计算 X ，直到排列或翻转其行的符号，并且我们可以设置 $A = BX^+$ 并精确计算真实字典。

第六章

高斯混合模型

在这一章中，我们将学习高斯混合模型和聚类。基本问题是，给定从 k 个高斯混合中随机采样，我们希望使用少量样本提供一个高效的算法来学习其参数。如果这些参数准确，我们可以对样本进行聚类，我们的误差将几乎与贝叶斯最优分类器一样准确。

6.1 历史

学习高斯混合模型的参数的问题可以追溯到著名的统计学家卡尔·皮尔逊（1894年），他对生物学和进化感兴趣。

事实上，有一种叫做那不勒斯螃蟹的特定物种栖息在他周围的地区。他从这个种群中取了成千上万个样本，并测量了每个样本的某些物理特征。他绘制了发生频率，但得到的密度函数让他感到惊讶。他预期它应该是高斯的，但实际上它甚至不对称于其最大值。参见图6.1。他假设那不勒斯螃蟹可能不是一个物种，而是两个物种，并且他观察到的密度函数可以解释为高斯混合。

在这个卓越的研究中，皮尔逊引入了矩方法。他的基本想法是从样本中计算经验矩，并使用每个经验矩来建立一个关于混合参数的多项式方程组。他亲手解决了这个方程组！实际上，在本单元中我们将回顾他的基本方法。

基础知识

在这里，我们正式描述学习高斯混合模型的问题。回想一下，对于一维高斯分布，其密度函数为：

$$\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}$$

在 \mathbb{R}^n 中，多维高斯分布的密度函数为：

$$\mathcal{N}(\mu, \Sigma) = \frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} \exp\left\{-\frac{(x - \mu)^\top \Sigma^{-1} (x - \mu)}{2}\right\}$$

这里 Σ 是协方差矩阵。如果 $\Sigma = I_n$ 和 $\mu = \vec{0}$ ，那么分布就是：
 $\mathcal{N}(0, 1) \times \mathcal{N}(0, 1) \times \dots \times \mathcal{N}(0, 1)$ 。

两个高斯分布的混合是一个密度函数为：

$$F(x) = w_1 F_1(x) + (1 - w_1) F_2(x)$$

其中 F_1 和 F_2 是高斯分布。我们可以按照以下方式生成随机样本：以概率 w_1 输出一个来自 F_1 的随机样本，否则输出一个来自 F_2 的随机样本。我们的基本问题是从 F 的随机样本中学习描述混合的参数。我们注意到我们将通过其样本复杂度和运行时间来衡量算法的好坏。

矩方法

Pearson 使用矩方法将两个高斯混合分布拟合到他的数据中。高斯混合分布的矩本身是未知参数的多项式，我们用 M_r 表示。

$$\mathbb{E}_{x \leftarrow F_1(x)} [x^r] = M_r(\mu, \sigma^2)$$

然后我们可以写成

$$\mathbb{E}_{x \leftarrow F(x)} [x^r] = w_1 M_r(\mu_1, \sigma_1^2) + (1 - w_1) M_r(\mu_2, \sigma_2^2) = P_r(w_1, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$$

因此，两个高斯混合分布的第 r 阶原始矩本身是一个未知参数的 $r+1$ 次多项式（ P_r ）。

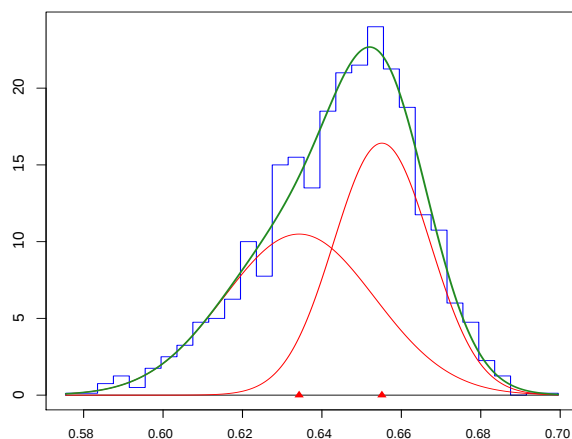


图6.1：对Pearson的那不勒斯螃蟹数据进行两个单变量高斯混合拟合，由Peter Macdonald使用R创建

Pearson的第六时刻检验：我们可以从随机样本中估计 $\mathbb{E}_{x \leftarrow F}[x^r]$ 。让 S 为我们的样本集。然后我们可以计算：

$$\widetilde{M}_r = \frac{1}{|S|} \sum_{x \in S} x^r$$

并且给定多项式数量的样本（对于任何 $r = O(1)$ ）， \widetilde{M}_r 将与 $\mathbb{E}_{x \leftarrow F(x)}[x^r]$ 近似相等。Pearson的方法是：

- 建立一个多项式方程系统

$$\left\{ P_r(w_1, \mu_1, \sigma_1, \mu_2, \sigma_2) = \widetilde{M}_r \right\}, r = 1, 2, \dots, 5$$

- 解决这个系统。每个解都是解释前五个经验矩的所有五个参数的设置。

皮尔逊通过手工解决了上述多项式方程组，并找到了一些候选解。每个解都对应于参数的同时设置，使得混合物的矩与经验矩相匹配。但是我们如何在这些候选解中进行选择呢？其中一些解显然是错误的；有些方差为负值，或者混合权重的值不在 $[0, 1]$ 范围内。但是即使在排除了这些解之后，皮尔逊仍然有多个候选解。他的方法是选择预测与经验第六矩最接近的根

\widetilde{M}_6 . 这被称为第六矩检验。

期望最大化

现代统计学的很大一部分集中在最大似然估计上，该估计会选择将参数设置为最大化混合物生成观测样本的概率。不幸的是，这个估计量的计算是 NP -难的[18]。流行的替代方法被称为期望-最大化，并在Dempster、Laird、Rubin的一篇极具影响力的论文中引入[50]。基本方法是重复以下步骤直到收敛：

- 对于每个 $x \in S$ ，计算后验概率：

$$w_1(x) = \frac{w_1 F_1(x)}{w_1 F_1(x) + (1 - w_1) F_2(x)}$$

- 更新混合权重：

$$w_1 \leftarrow \frac{\sum_{x \in S} w_1(x)}{|S|}$$

- 重新估计参数：

$$\mu_i \leftarrow \frac{\sum_{x \in S} w_i(x) x}{\sum_{x \in S} w_i(x)}, \quad \Sigma_i \leftarrow \frac{\sum_{x \in S} w_i(x) (x - \mu_i)(x - \mu_i)^\top}{\sum_{x \in S} w_i(x)}$$

这种方法容易陷入局部最优解，并且通常对其初始化方式非常敏感（参见例如[105]）。

6.2 基于聚类的算法

我们的基本目标是通过多项式数量的随机样本，给出能够可证地计算高斯混合模型的真实参数的算法。这个问题最早在Dasgupta的开创性论文[45]中提出，第一代算法主要关注混合成分之间几乎没有“重叠”的情况。下一代算法基于代数思想，并且完全避免了聚类。

在我们继续之前，我们将讨论高维高斯分布的一些反直觉特性。为了简化讨论，我们将重点放在球形高斯分布 $\mathcal{N}(\mu, \sigma^2 I)$ 在 \mathbb{R}^n 上。

事实 6.2.1 密度函数的最大值出现在 $x = \mu$ 。

事实 6.2.2 密度函数的几乎所有权重都满足 $\|x - \mu\|^2 = \sigma^2 n \pm \sigma^2 \sqrt{n \log n}$

这些事实似乎是不一致的，但解释是表面积随着半径 R 的增加而增加得更快，而密度函数的值减小，直到达到 $R^2 \approx \sigma^2 n$ 。因此，我们应该将高维球形高斯看作是半径为 $\sigma\sqrt{n}$ 的球体，外面有一层薄壳。

Dasgupta [45] – $\widetilde{\Omega}(\sqrt{n})$ 分离

Dasgupta 提出了第一个可证明的学习高斯混合模型的算法，并要求 $\|\mu_i - \mu_j\|_2 \geq \widetilde{\Omega}(\sqrt{n\sigma_{\max}})$ ，其中 σ_{\max} 是任何方向上的最大方差（例如，如果组件不是球形的）。请注意，分离中的常数取决于 w_{\min} ，我们假设我们知道这个参数（或者它的下界）。

该算法的基本思想是将混合物随机投影到对数 k di-维度上。这种投影将以很高的概率保持每对中心 μ_i 和 μ_j 之间的距离，但会缩小相同组分样本之间的距离，并使每个组分更接近球形，从而更容易进行聚类。然后，我们可以将所有样本聚类到生成它们的组分中，然后对于每个聚类，我们可以选择经验均值和经验协方差，这在很大概率下将是一个很好的估计 μ_i 和 Σ_i 。此外，我们可以通过每个聚类的大小来估计 w_i 。非正式地说，我们可以将这种分离条件理解为：如果我们将每个高斯看作一个球形球体，那么如果组分之间足够远，这些球体将是不相交的。

Arora 和 Kannan [18], Dasgupta 和 Schulman [53] – $\widetilde{\Omega}(n^{1/4})$ 分离

我们将详细描述 [18] 中的方法。基本问题是，如果 \sqrt{n} 分离是我们可以将组件视为不相交的阈值，那么当组件更接近时，我们如何学习？事实上，即使组件只有 $\widetilde{\Omega}(n^{1/4})$ 分离，仍然成立：同一组件中的任意一对样本比不同组件中的任意一对样本更接近。这是怎么回事？解释是，即使表示每个组件的球不再是不相交的，我们仍然很难从它们的重叠区域中采样。

考虑 $x, x' \leftarrow F_1$ 和 $y \leftarrow F_2$ 。

声明6.2.3所有向量 $x - \mu_1, x' - \mu_1, \mu_1 - \mu_2, y - \mu_2$ 几乎正交 (*whp*)

这个声明是显然的，因为向量 $x - \mu_1, x' - \mu_1, y - \mu_2$ 都是从一个球体中均匀选择的，而 $\mu_1 - \mu_2$ 是唯一的固定向量。事实上，任何一组向量，除了一个向量外，都是从一个球体中均匀选择的，它们几乎是正交的。现在我们可以计算：

$$\begin{aligned}\|x - x'\|^2 &\approx \|x - \mu_1\|^2 + \|\mu_1 - x'\|^2 \\ &\approx 2n\sigma^2 \pm 2\sigma^2 \sqrt{n \log n}\end{aligned}$$

同样地：

$$\begin{aligned}\|x - y\|^2 &\approx \|x - \mu_1\|^2 + \|\mu_1 - \mu_2\|^2 + \|\mu_2 - y\|^2 \\ &\approx 2n\sigma^2 + \|\mu_1 - \mu_2\|^2 \pm 2\sigma^2 \sqrt{n \log n}\end{aligned}$$

因此，如果 $\|\mu_1 - \mu_2\| = \widetilde{\Omega}(n^{1/4}, \sigma)$ ，那么 $\|\mu_1 - \mu_2\|^2$ 大于误差项，并且来自同一组分的每对样本将比来自不同组分的每对样本更接近。实际上，我们可以找到合适的阈值 τ 并正确地聚类所有样本。同样，我们可以输出经验均值、经验协方差和每个聚类的相对大小，这些将是真实参数的良好估计。

Vempala和Wang [117] – $\widetilde{\Omega}(k^{1/4})$ 分离

Vempala和Wang [117]消除了对 n 的依赖，并用一个与 k - 组件数量有关的分离条件来替代它。这个想法是，如果我们能够将混合物投影到由 $\{\mu_1, \dots, \mu_k\}$ 张成的子空间 T 中，我们将保持每对组件之间的分离，但降低环境维度。

那么我们如何找到由均值张成的子空间 T 呢？我们将限制我们的讨论在具有公共方差 $\sigma^2 I$ 的球形高斯混合物上。设 $x \sim F$ 为混合物的随机样本，那么我们可以写成 $x = c + z$ 其中 $z \sim N(0, \sigma^2 I_n)$ ， c 是一个随机向量，以概率 w_i 取值为 μ_i ，对于每个 $i \in [k]$ 。所以：

$$\mathbb{E}[xx^T] = \mathbb{E}[cc^T] + \mathbb{E}[zz^T] = \sum_{i=1}^k w_i \mu_i \mu_i^T + \sigma^2 I_n$$

因此， $\mathbb{E}[xx^T]$ 的最大奇异值对应的左奇异向量恰好是 T 的基。我们可以通过足够的随机样本来估计 $\mathbb{E}[xx^T]$ ，计算其奇异值分解并将混合物投影到 T 上，并调用[18]的算法。

Brubaker和Vempala [32] - 分离超平面

如果任何分量的最大方差远大于分量之间的分离距离怎么办？Brubaker和Vempala [32]观察到，现有的算法都无法成功处理并行煎饼的例子，如图??所示。

尽管存在一个超平面将混合物分开，使得几乎所有的一个组分在一边，几乎所有的另一个组分在另一边。[32]提出了一个成功的算法，前提是存在这样一个分离超平面，然而对于两个以上的高斯混合物，条件更加复杂。注意，并不是所有我们希望学习的混合物都有这样一个分离超平面。参见图 ??。

6.3 密度估计的讨论

到目前为止，我们讨论的算法[45]，[53]，[18]，[117]，[1]，[32]都集中在聚类上；即使在聚类不可能的情况下，我们能否给出有效的学习算法？考虑一个由两个高斯分布混合而成的混合物 $F = w_1 F_1 + w_2 F_2$ 。到目前为止，我们考虑的分离条件都暗示着 $d_{TV}(F_1, F_2) = 1 - o(1)$ 。特别地，组分之间的重叠很小。然而，如果 $d_{TV}(F_1, F_2) = 1/2$ ，我们无法确定每个样本是由哪个组分生成的。

更准确地说，两个分布 F 和 G 之间的总变差距离衡量了我们能够将它们耦合的程度：

定义6.3.1 F 和 G 之间的耦合是一个对 (x, y) 的分布，使得 x 的边缘分布是 F ， y 的边缘分布是 G 。错误是 $x = y$ 的概率。

命题6.3.2 F 和 G 之间存在一个错误率为 ε 的耦合，如果且仅如果 $d_{TV}(F, G) \leq \varepsilon$ 。

回到从两个高斯混合分布中聚类样本的问题，如果 $d_{TV}(F_1, F_2) = 1/2$ ，则 F_1 和 F_2 之间存在一个概率为 $1/2$ 的耦合。因此，与其按照通常的方式从两个高斯混合分布中进行采样（选择哪个分量，然后从中随机采样），我们可以采用以下方式进行采样：

- (a) 从 F_1 和 F_2 之间的最佳耦合中选择 (x, y)
- (b) 如果 $x = y$ ，则输出 x

(c) 否则以概率 w_1 输出 x ，否则输出 y

该过程从 F 生成一个随机样本，但对于一半的样本，我们不需要决定是哪个组件生成的！因此，即使我们知道混合物，也没有聚类过程能够正确地将多项式数量的样本分类为哪个组件生成的！因此，在 F_1 和 F_2 的混合物不是 $1 - o(1)$ 的情况下，我们迄今讨论的基本方法不起作用！尽管如此，我们仍然能够给出学习 F 的参数的算法，即使 F_1 和 F_2 的组件几乎完全重叠。

接下来我们将讨论学习算法的一些基本目标类型：

(a) 不恰当的密度估计

在整个过程中，我们假设 $F \in \mathcal{C}$ ，其中 \mathcal{C} 是一些分布类别（例如两个高斯混合）。我们在不恰当的密度估计中的目标是找到任何分布 F ，使得 $d_{TV}(F, F) \leq \varepsilon$ 。这是学习算法的最弱目标。一种常见的方法（特别是在低维度中）是构建一个核密度估计；假设我们从 F 中取得许多样本，并构建一个代表我们样本的点质量分布 G 。然后我们可以设置 $F = G * \mathcal{N}(0, \sigma^2)$ ，如果 F 足够平滑并且我们取足够多的样本，那么 $d_{TV}(F, F) \leq \varepsilon$ 。然而， F 在不了解 F 的组成部分的情况下工作；它之所以有效是因为 F 是平滑的。我们注意到，在高维度中，这种方法在 $F = G * \mathcal{N}(0, \sigma^2 I)$ 接近 F 的情况下会失败得很厉害，即使 F 是平滑的，我们也需要取指数数量的样本才能保证。

(b) 适当的密度估计

在这里，我们的目标是找到一个分布 $F \in \mathcal{C}$ 其中 $d_{TV}(F, F) \leq \varepsilon$ 。请注意，如果 \mathcal{C} 是两个高斯混合物的集合，那么核密度估计不是一个有效的假设，因为它通常是许多高斯混合物的混合物（取决于我们采样的样本数量）。适当的密度估计通常比不适当的密度估计要困难得多。事实上，我们将专注于一个更强的目标：

(b) 参数学习

在这里，我们不仅要求 $d_{TV}(F, F) \leq \varepsilon$ 和 $F \in \mathcal{C}$ ，还希望 F 在逐个组成部分的基础上是一个良好的估计 F 。例如，我们的目标特化到两个高斯混合物的情况是：

定义6.3.3我们将说一个混合 $F = w_1 F_1 + w_2 F_2$ 是 ε -接近（在组件基础上）到 F ，如果存在一个排列 $\pi: \{1, 2\} \rightarrow \{1, 2\}$ ，使得对于所有 $i \in \{1, 2\}$ ：

$$\left| w_i - w_{\pi(i)} \right|, d_{TV}(F_i, F_{\pi(i)}) \leq \varepsilon$$

注意 F 和 F 也必须是混合物： $d_{TV}(F, F) \leq 4\varepsilon$ 。

然而，我们可以有混合物 F 和 F ，它们都是 k 高斯分布的混合物，作为分布是接近的，但在组件基础上不接近。如果可能的话，最好按组件学习 F ，而不仅仅是进行适当的密度估计。请注意，如果 F 与 F 接近，则即使我们无法对样本进行聚类，我们仍然能够近似计算后验概率[79]，这是参数学习相对于一些较弱的学习目标的主要优势之一。

但是需要记住，参数学习的下界并不意味着适当密度估计的下界。我们将为混合高斯模型的参数学习提供最优算法，对于任意 $k = O(1)$ ，运行时间为多项式时间。此外，存在一对混合高斯模型 F 和 F ，它们在组件级别上并不接近，但是有 $d_{TV}(F, F) \leq 2^{-k}$ [95]。因此，不存在一个以 $\text{poly}(n, k, 1/\varepsilon)$ 样本为输入的参数学习算法 - 因为我们至少需要取 2^k 个样本来区分 F 和 F 。但是在适当密度估计的背景下，我们不需要区分这两个混合模型。

开放问题2在 n 维空间中，对于混合高斯模型的适当密度估计，是否存在一个 $\text{poly}(n, k, 1/\varepsilon)$ 时间复杂度的算法？

6.4 无聚类算法

回想一下，我们的目标是学习一个与 F ε -接近的 F 。实际上，相同的定义可以推广到 k 个高斯混合模型：

定义6.4.1如果存在一个排列 $\pi: \{1, 2, \dots, k\} \rightarrow \{1, 2, \dots, k\}$ ，使得对于所有的 $i \in \{1, 2, \dots, k\}$ ，都满足以下条件：

$$\left| w_i - w_{\pi(i)} \right|, d_{TV}(F_i, F_{\pi(i)}) \leq \varepsilon$$

在多项式($n, 1/\varepsilon$)个样本中, 我们何时能够希望学习到一个 ε 接近的估计? 实际上, 有两种微不足道的情况我们无法做到这一点, 但这将是唯一出现问题的情况:

- (a) 如果 $w_i = 0$, 我们永远无法学习到接近 F_i 的 F_i , 因为我们从 F_i 中永远不会得到任何样本。

事实上, 我们需要对每个 w_i 有一个定量的下界, 比如 $w_i \geq \varepsilon$, 这样如果我们取足够多的样本, 我们将至少从每个组件中得到一个样本。

- (b) 如果 $d_{TV}(F_i, F_j) = 0$, 我们永远无法学习到 w_i 或 w_j , 因为 F_i 和 F_j 完全重叠。

同样, 我们需要对 $d_{TV}(F_i, F_j)$ 有一个定量的下界, 比如 $d_{TV}(F_i, F_j) \geq \varepsilon$, 对于每个 $i \neq j$, 这样如果我们取足够多的样本, 我们将至少从各个组件之间的非重叠区域得到一个样本。

定理6.4.2[79], [95] 如果对于每个 i 和 d , $w_i \geq \varepsilon$, 并且对于每个 $i \neq j$, $d_{TV}(F_i, F_j) \geq \varepsilon$ 那么存在一个高效的算法来学习一个 ε -接近的估计 \hat{F} 到 F , 其运行时间和样本复杂度为 $\text{poly}(n, 1/\varepsilon, \log 1/\delta)$, 并且以概率成功为 $1 - \delta$ 。

注意, 多项式的次数多项式地依赖于 k 。Kalai, Moitra 和 Valiant [79] 提供了第一个算法, 用于学习两个高斯混合模型, 没有分离条件。随后, Moitra 和 Valiant [95] 提供了一个算法, 用于学习 k 个高斯混合模型, 同样没有分离条件。

在独立且并发的作品中, Belkin 和 Sinha [23] 也提供了一个多项式时间算法, 用于学习 k 个高斯混合模型, 但是没有明确给出作为 k 函数的运行时间上界 (因为他们的工作依赖于可证明无效的基础定理)。此外, [79] 和 [95] 中的目标是学习 F , 使其组件在总变差距离上接近 F 的组件, 这通常是一个比要求参数加法接近更强的目标, 这是 [23] 中的目标。好处是 [23] 中的算法适用于一维环境中更一般的学习问题, 我们将在本章末尾详细描述这个算法。

在本节中, 我们将重点关注 $k=2$ 的情况, 因为这个算法在概念上更简单。事实上, 我们将关注一个较弱的学习目标: 如果对于所有的 i , 我们将 F 称为 ε -接近 F , 如果 $|w_i - w_{\pi(i)}|, \|\mu_i - \mu_{\pi(i)}\|, \|\Sigma_i - \Sigma_{\pi(i)}\|_F \leq \varepsilon$ 。我们进一步假设 F 已经适当地进行了归一化:

定义6.4.3 一个分布 F 处于各向同性位置，如果

$$(a) \mathbb{E}_{x \leftarrow F}[x] = 0$$

$$(b) \mathbb{E}_{x \leftarrow F}[xx^T] = I$$

或者，我们要求分布的均值为零，并且在每个方向上的方差为一。事实上，这个条件并不像听起来那么强：

声明 6.4.4 如果 $\mathbb{E}_{x \leftarrow F}[xx^T]$ 是满秩的，那么存在一个仿射变换将 F 放置在等距位置

很容易看出 M 是半正定的，并且根据假设，实际上是满秩的。因此，我们可以将 M 写成 BB^T 的形式，其中 B 是可逆的（这通常被称为Cholesky分解[74]）。然后设置 $y = B^{-1}(x - \mu)$ ，很容易看出 $E[y] = 0$ ，并且 $E[yy^T] = B^{-1} M (B^{-1})^T = I$. ■

我们的目标是学习一个对 F 的加法 ε 近似，我们将假设 F 已经经过预处理，使其处于各向同性位置。

大纲

现在我们可以描述算法的基本轮廓，尽管还有很多细节需要填充：

- (a) 考虑一系列投影到一维的过程
- (b) 运行一个单变量学习算法
- (c) 建立一个关于高维参数的线性方程组，并进行反向求解

各向同性投影引理

我们需要克服一些障碍来实现这个计划，但让我们详细讨论一下这个轮廓的细节：

声明 6.4.5 $\text{proj}_r[\mathcal{N}(\mu, \Sigma)] = \mathcal{N}(r^T \mu, r^T \Sigma r)$

或者，高维高斯的投影是一个一维高斯分布，其均值和方差分别为 $r^T \mu$ 和 $r^T \Sigma r$ 。这意味着，如果我们知道单个高斯分量在（已知）方向 r 上的投影的参数，那么我们可以使用这些参数为 μ 和 Σ 设置一个线性约束。如果我们按照这个计划进行，我们需要考虑大约 n^2 个投影来获得足够的线性约束，因为我们需要解决 Σ 中的 $\Theta(n^2)$ 个方差。现在我们将遇到轮廓中的第一个问题。

让我们定义一些符号：

定义 6.4.6 $d_p(\mathcal{N}(\mu_1, \sigma_1^2), \mathcal{N}(\mu_2, \sigma_2^2)) = |\mu_1 - \mu_2| + |\sigma_1^2 - \sigma_2^2|$

我们将把这称为参数距离。最终，我们将提供一个学习高斯混合模型的单变量算法，并希望在 $\text{proj}_r[F]$ 上运行它。

问题 4 但是如果 $d_p(\text{proj}_r[F_1], \text{proj}_r[F_2])$ 非常小怎么办？

这将是一个问题，因为我们需要以指数级的精度运行我们的单变量算法，才能看到是否有两个组件而不是一个！我们如何解决这个问题？事实上，只要 F 处于各向同性位置，几乎肯定不会发生这种情况。为了直观起见，考虑两种情况：

(a) 假设 $\|\mu_1 - \mu_2\| \geq \text{poly}(1/n, \varepsilon)$.

如果 F_1 和 F_2 的均值之间的差异至少是任意固定的倒数多项式，那么很有可能 $\|r^T \mu_1 - r^T \mu_2\|$ 也至少是多项式 $(1/n, \varepsilon)$ 。因此，由于均值的差异， $\text{proj}_r[F_1]$ 和 $\text{proj}_r[F_2]$ 将具有不同的参数。

(b) 假设 $\|\mu_1 - \mu_2\| \leq \text{poly}(1/n, \varepsilon)$.

关键观察是，如果 $d_{TV}(F_1, F_2) \geq \varepsilon$ 并且它们的均值几乎相同，那么它们在随机方向 r 上投影时，它们的协方差 Σ_1 和 Σ_2 必须明显不同。在这种情况下，由于方差的差异， $\text{proj}_r[F_1]$ 和 $\text{proj}_r[F_2]$ 将具有不同的参数。这是以下引理背后的直觉：

引理 6.4.7 如果 F 处于各向同性位置且 $w_i \geq \varepsilon$ 和 $d_{TV}(F_1, F_2) \geq \varepsilon$ ，那么以高概率对于一个随机的 r

$$d_p(\text{proj}_r[F_1], \text{proj}_r[F_2]) \geq 2\varepsilon_3 = \text{poly}(1/n, \varepsilon)$$

请注意，当 F 不处于各向同性位置时，这个引理是不成立的（例如考虑平行的煎饼例子），而且当推广到 $k > 2$ 个高斯混合时，这是失败的关键步骤，因为即使 F 处于各向同性位置，对于几乎所有的 r 选择，投影到 r 上的结果也会指数级地接近于一个 $< k$ 个高斯混合！（[95]中的方法是学习一个 $< k$ 个高斯混合作为真实混合的代理，然后找到一个可以用来将混合聚类成子混合并递归的方向）。

配对引理

接下来我们将遇到第二个问题：假设我们在方向 r 和 s 上进行投影，并学习 $F_1^r = \frac{1}{2}F_1^r + \frac{1}{2}F_2^r$ 和 $F_1^s = \frac{1}{2}F_1^s + \frac{1}{2}F_2^s$ 分别。然后， F_1^r 的均值和方差会对两个高维高斯分布之一产生线性约束， F_1^s 同理。

问题5：我们如何知道它们对相同的高维分量产生约束？

最终，我们希望建立一个线性约束系统来解决 F_1 的参数，但是当我们把 F 投影到不同的方向（比如 r 和 s ）时，我们需要将这两个方向的分量配对。关键观察是，当我们从 r 变化到 s 时，混合模型的参数会连续变化。参见图 ??。因此，当我们投影到 r 上时，根据各向同性投影引理，这两个分量的均值或方差要么明显不同。假设它们的均值相差 ε_3 ；那么如果 r 和 s 接近（与 ε_1 相比），混合模型中每个分量的参数变化不大，并且在 $\text{proj}_r[F]$ 中具有较大均值的分量将对应于在 $\text{proj}_s[F]$ 中具有较大均值的分量。当方差至少相差 ε_3 时，类似的情况也适用。

引理 6.4.8 如果 $\|r - s\| \leq \varepsilon_2 = \text{poly}(1/n, \varepsilon_3)$ 那么

- (a) 如果 $|r^T \mu_1 - r^T \mu_2| \geq \varepsilon_3$ 那么在 $\text{proj}_r[F]$ 和 $\text{proj}_s[F]$ 中具有较大均值的分量对应于相同的高维分量
- (b) 否则，如果 $|r^T \Sigma_1 r - r^T \Sigma_2 r| \geq \varepsilon_3$ 那么在 $\text{proj}_r[F]$ 和 $\text{proj}_s[F]$ 中具有较大方差的分量对应于相同的高维分量

因此，如果我们随机选择 r 并且只在满足 $\|r - s\| \leq \varepsilon_2$ 的方向 s 中进行搜索，我们将能够正确地将不同的一维混合中的分量配对。

条件数引理

现在我们在高维情况下遇到了最后一个问题：假设我们随机选择 r ，并且对于 s_1, s_2, \dots, s_p 我们学习投影 F 的参数在这些方向上并正确地配对组件。我们只能希望在这些投影上学习参数，精度为 ε_1 （我们的单变量学习算法将具有运行时间和样本复杂度 $\text{poly}(1/\varepsilon_1)$ ）。

问题6我们单变量估计中的这些误差如何转化为我们高维估计中的误差对于 $\mu_1, \Sigma_1, \mu_2, \Sigma_2$ ？

回想一下，条件数控制着这个。在高维情况下，我们需要的最后一个引理是：

引理6.4.9 解决 μ_1, Σ_1 的线性系统的条件数是 $\text{poly}(1/\varepsilon_2, n)$ 其中所有方向对 ε_2 分开。

直观地说，如果 r 和 s_1, s_2, \dots, s_p 彼此更接近，那么系统的条件数将更糟糕（因为线性约束更接近冗余），但关键事实是条件数被一个固定的多项式所限制在 $1/\varepsilon^2$ 和 n 的范围内，因此如果我们选择 $\varepsilon_1 = \text{poly}(\varepsilon_2, n)\varepsilon$ ，那么我们对高维参数的估计将在添加 ε 的范围内。请注意，每个参数 $\varepsilon, \varepsilon_3, \varepsilon_2, \varepsilon_1$ 都是之前参数（和 $1/n$ ）的一个固定多项式，因此我们只需要在多项式数量的混合物上运行我们的单变量学习算法，以获得一个 ε -接近的估计 F ！但我们仍然需要设计一个单变量算法，接下来我们回到Pearson的原始问题！

6.5 一元算法

在这里，我们将给出一个单变量算法来学习一个由两个高斯分布混合而成的参数，精度为 ε ，其运行时间和样本复杂度为多项式 $(1/\varepsilon)$ 。请注意，混合分布 $F = w_1 F_1 + w_2 F_2$ 处于各向同性位置（因为各向同性位置的分布的投影仍然处于各向同性位置），并且我们假设 $w_1, w_2 \geq \varepsilon$ 和 $d_{TV}(F_1, F_2) \geq \varepsilon$ 。我们的第一个观察是，所有的参数都是有界的：

声明 6.5.1 (a) $\mu_1, \mu_2 \in [-1/\sqrt{\varepsilon}, 1/\sqrt{\varepsilon}]^d$

$$(b) \sigma_1^2, \sigma_2^2 \in [0, 1/\varepsilon]$$

这个断言是显然的，因为如果上述条件中的任何一个被违反，那么混合物的方差将严格大于 ε （因为 $w_1, w_2 \geq \varepsilon$ 并且混合物的均值为零）。

因此，我们可以尝试使用网格搜索来学习参数：

网格搜索

输入：来自 $F(\Theta)$ 的样本

输出：参数 $\Theta = (w_1, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$

对于所有有效的 Θ ，其中参数是 ε^C 的倍数

使用样本测试 Θ ，如果通过则输出 Θ

结束

例如，我们可以通过计算足够多的随机样本的 $F(\Theta)$ 的前六个矩来测试 Θ ，并且如果其前六个矩的每一个都在观察到的矩的加法 τ 内，则输出 Θ 。（这是对皮尔逊第六矩检验的轻微变体）。

很容易看出，如果我们取足够多的样本并适当地设置 τ ，那么如果我们将真实参数 Θ 四舍五入到任何有效的网格点，其参数为 ε^C 的倍数，则结果 Θ 将以很高的概率通过我们的测试。这被称为完备性。更具挑战性的部分是建立正确性；毕竟，除了接近 Θ 的参数之外，为什么没有其他一组参数 Θ 可以通过我们的测试呢？

或者，我们想要证明任何两个混合 F 和 F ，其参数之间的差异不超过 ε ，必须具有它们的前六个矩 noticeably 不同。主要引理如下：

引理 6.5.2 对于任何不在参数上 ε -接近的 F 和 F ，存在一个 $r \in \{1, 2, \dots, 6\}$ ，其中

$$\left| M_r(\Theta) - M_r(\Theta) \right| \geq \varepsilon^{O(1)}$$

其中 Θ 和 Θ 是 F 和 F 的参数， M_r 是第 r 个原始矩。

令 \widetilde{M}_r 为经验矩。则

$$\left| M_r(\Theta) - M_r(\Theta) \right| \leq \underbrace{\left| \widetilde{M}_r(\Theta) - \widetilde{M}_r \right|}_{\leq \tau} + \underbrace{\left| \widetilde{M}_r - M_r(\Theta) \right|}_{\leq \tau} \leq 2\tau$$

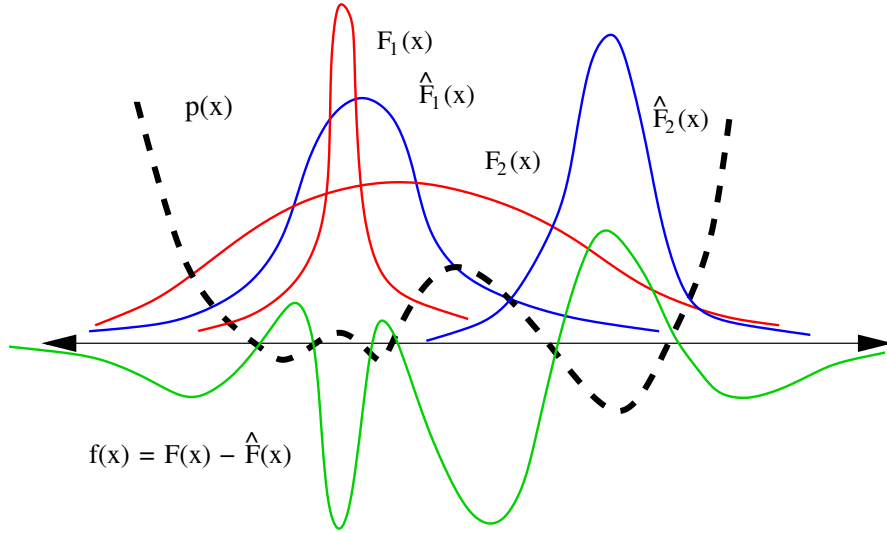


图6.2: 如果 $f(x)$ 最多有六个零点, 我们可以找到最多六次多项式与其符号一致

其中第一项最多为 τ 因为测试通过, 第二项很小, 因为我们可以取足够多的样本 (但仍然是 $\text{poly}(1/\tau)$), 使得经验矩和真实矩接近。因此, 我们可以应用上述引理的逆否命题, 并得出结论: 如果网格搜索输出 Θ , 则 Θ 和 θ 在参数上必须 ε -接近, 这给我们提供了一个高效的单变量算法! 因此, 我们的主要目标是证明如果 F 和 F 不 ε -接近, 则它们的前六个矩中至少有一个明显不同。事实上, 即使 ε

$= 0$ 的情况也具有挑战性: 如果 F 和 F 是两个高斯混合物, 为什么它们的前六个矩必定不同? 我们的主要目标是证明这个陈述, 使用的是热方程。

事实上, 让我们考虑以下思想实验。令 $f(x) = F(x) - F(x)$ 为密度函数 F 和 F 之间的逐点差异。那么, 问题的核心是: 我们能否证明 $f(x)$ 最多在六个点上与 x 轴相交? 见图6.2。

引理6.5.3 如果 $f(x)$ 最多在六个点上与 x 轴相交, 则 F 和 F 的前六个矩不同。

证明：事实上，我们可以构造一个（非零）次数最多为六的多项式 $p(x)$ 与 $f(x)$ 的符号一致 - 即 $p(x)f(x) \geq 0$ 对于所有 x 成立。然后

$$\begin{aligned} 0 < \left| \int_x p(x)f(x)dx \right| &= \left| \int_x \sum_{r=1}^6 p_r x^r f(x)dx \right| \\ &\leq \sum_{r=1}^6 |p_r| \left| M_r(\Theta) - M_r(\Theta) \right| \end{aligned}$$

如果 F 和 F 的前六个矩匹配完全，则右边为零
这是一个矛盾。 ■

所以我们只需要证明 $F(x) - F(x)$ 最多有六个零点。让我们通过归纳法证明一个更强的引理：

引理 6.5.4 令 $f(x) = \sum_{i=1}^k \alpha_i \mathcal{N}(\mu_i, \sigma_i^2, x)$ 是 k 个高斯函数的线性组合（ α_i 可以为负）。那么如果 $f(x)$ 不恒为零，则 $f(x)$ 最多有 $2k - 2$ 个零点。

我们将依赖以下工具：

定理 6.5.5 给定 $f(x) : \mathbb{R} \rightarrow \mathbb{R}$ ，它是解析的并且具有 n 个零点，那么对于任意的 $\sigma^2 > 0$ ，函数 $g(x) = f(x) * \mathcal{N}(0, \sigma^2)$ 最多具有 n 个零点。

这个定理有一个物理解释。如果我们将 $f(x)$ 视为无限一维杆的热分布，则在某个稍后的时间，热分布会是什么样子？实际上，它恰好是 $g(x) = f(x) * \mathcal{N}(0, \sigma^2)$ ，其中 σ^2 是适当选择的值。或者，高斯函数是热方程的 Green 函数。因此，我们对扩散的许多物理直觉对卷积也有影响 - 通过高斯函数进行卷积会使其平滑，并且不会产生新的局部极大值（以及相关的不会产生新的零点）。

最后我们回顾一下基本事实：

事实 6.5.6 $\mathcal{N}(0, \sigma_1^2) * \mathcal{N}(0, \sigma_2^2) = \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$

现在我们准备证明上述引理，并得出结论，如果我们完全知道两个高斯混合的前六个矩，那么我们也会完全知道它的参数。让我们通过归纳法证明上述引理，并假设对于任意线性组合 $k = 3$ 的高斯分布，零交叉点的数量是

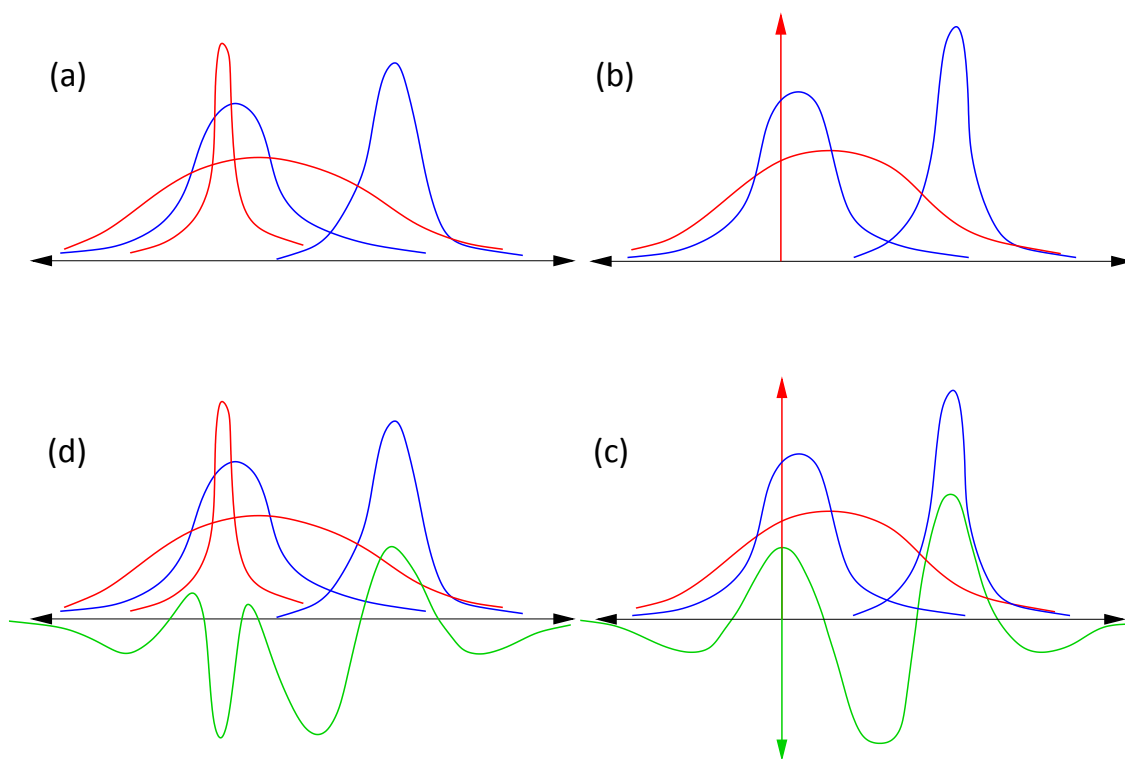


图6.3: (a) 四个高斯分布的线性组合 (b) 从每个方差中减去 σ^2 (c) 加回 δ 函数 (d) 通过 $\mathcal{N}(0, \sigma^2)$ 卷积以恢复原始线性组合

最多四个。现在考虑任意线性组合的四个高斯函数，并且令 σ^2 为任意组成部分的最小方差。参见图6.3(a)。我们可以考虑一个相关的混合物，其中我们从每个组成部分的方差中减去 σ^2 。参见图6.3(b)。

现在，如果我们忽略掉 δ 函数，我们得到一个由三个高斯函数线性组合而成的函数，并且根据归纳法，我们知道它最多有四个零点。但是，当我们加入 δ 函数时，我们能够添加多少个零点呢？我们最多能够添加两个，一个在上升过程中，一个在下降过程中（在这里，为了简化演示，我们忽略了使用 δ 函数时的一些实际分析复杂性）。参见图6.3(c)。现在，我们可以通过 $\mathcal{N}(0, \sigma^2)$ 来卷积函数，以恢复原始的由四个高斯函数线性组合而成的函数，但是这最后一步不会增加零点的数量！参见图6.3(d)。这证明了

$$\left\{ M_r(\Theta) = M_r(\Theta) \right\}, r = 1, 2, \dots, 6$$

只有两个解（真实参数和我们也可以交换哪个是哪个的组件）。事实上，这个多项式方程组也是稳定的，并且对于多项式方程组也有条件数的类似情况，这意味着我们刚刚证明的一个定量版本：如果 F 和 F 不是 ε -接近，则它们的前六个矩明显不同。这给我们提供了我们的一元算法。

6.6 代数几何的视角

在这里，我们将介绍Belkin和Sinha的另一种一元学习算法[23]，它也利用了矩方法，但使用了代数几何的工具进行了更广泛的分析。

多项式族

我们将分析以下分布类别的矩方法：

定义6.6.1分布类别 $F(\Theta)$ 被称为多项式族，如果

$$\forall r, \mathbb{E}_{X \in F(\Theta)} [X^r] = M_r(\Theta)$$

其中 $M_r(\Theta)$ 是一个关于 $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$ 的多项式。

这个定义涵盖了一类广泛的分布，例如其组成部分可以是均匀分布、指数分布、泊松分布、高斯分布或伽马函数。我们需要对分布进行另一个（温和的）条件，以保证它由所有的矩特征化。

事实 6.6.2 如果随机变量 X 的矩生成函数 (*mgf*) 定义为 $\sum \mathbb{E}[X^n] \frac{t^n}{n!}$ 在零的邻域内收敛，它唯一确定了概率分布，即

$$\forall r, M_r(\Theta) = M_r(\Theta) \implies F(\Theta) = F(\Theta).$$

我们的目标是要证明对于任何多项式族，有限数量的矩可以足够。首先，我们介绍相关的定义：

定义 6.6.3 给定一个环 R ，由 $g_1, g_2, \dots, g_n \in R$ 生成的理想 I ，记作 $I = \langle g_1, g_2, \dots, g_n \rangle$ ，定义为

$$I = \left\{ \sum_i r_i g_i \text{ 其中 } r_i \in R \right\}.$$

定义 6.6.4 诺特环是指对于任何理想序列

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \dots,$$

存在 N 使得 $I_N = I_{N+1} = I_{N+2} = \dots$ 。

定理 6.6.5 (希尔伯特的基本定理) 如果 R 是一个诺特环，那么 $R[X]$ 也是一个诺特环。

很容易看出 \mathbb{R} 是一个诺特环，因此我们知道 $\mathbb{R}[x]$ 也是诺特环。现在我们可以证明对于任何多项式族，有限数量的矩可以唯一地确定该族中的任何分布：

定理 6.6.6 让 $F(\Theta)$ 是一个多项式族。如果矩生成函数在零的邻域内收敛，那么存在 N 使得

$$F(\Theta) = F(\Theta) \text{ 当且仅当 } M_r(\Theta) = M_r(\Theta) \quad \forall r \in 1, 2, \dots, N$$

证明： 设 $Q_r(\Theta, \Theta) = M_r(\Theta) - M_r(\Theta)$ 。设 $I_1 = \langle Q_1 \rangle$, $I_2 = \langle Q_1, Q_2 \rangle, \dots$ 。这是我们在 $\mathbb{R}[\Theta, \Theta]$ 中的升链理想。我们可以引用希尔伯特的基础定理

并得出结论： $\mathbb{R}[X]$ 是一个诺特环，因此存在一个 N 使得 $I_N = I_{N+1} = \dots$ 。因此对于所有的 $N + j$ ，我们有

$$Q_{N+j}(\Theta, \Theta) = \sum_{i=1}^N p_{ij}(\Theta, \Theta) Q_i(\Theta, \Theta)$$

对于一些多项式 $p_{ij} \in \mathbb{R}[\Theta, \Theta]$. 因此，如果对于所有的 $r \in 1, 2, \dots, N$ ，有 $M_r(\Theta) = M_r(\Theta)$ ，那么对于所有的 r ，有 $M_r(\Theta) = M_r(\Theta)$ ，根据事实6.6.2，我们得出 $F(\Theta) = F(\Theta)$ 。定理的另一方面是显而易见的。 ■

上述定理对 N 没有给出任何有限的界限，因为基础定理也没有给出。这是因为基础定理是通过反证法证明的，但更根本的是不可能只根据环的选择给出 N 的界限。考虑以下例子

例子1考虑Noether环 $\mathbb{R}[x]$ 。令 $I_i = \langle x^{N-i} \rangle$ 对于 $i = 0, \dots, N$ 。这是一个严格递增的理想链，对于 $i = 0, \dots, N$ 。因此，即使环 $\mathbb{R}[x]$ 被固定，对 N 没有普遍的界限。

像定理6.6.6中的那些界限通常被称为无效的。考虑将上述结果应用于高斯混合模型：根据上述定理，我们知道如果两个混合 F 和 F of k 个高斯分布相同，那么这些混合在它们的第一个 N 矩上是一致的。这里 N 是 k 的一个函数，而 N 是有限的，但我们不能使用上述工具写出 N 作为 k 的一个显式界限。尽管如此，这些工具的应用范围要比我们在前一节中用于证明 $4k - 2$ 个矩足够对于 k 个高斯混合模型的热方程基于的专门工具要广泛得多。

多项式不等式系统

一般来说，我们无法准确获得分布的矩，只能得到噪声近似。我们的主要目标是证明先前结果的定量版本，即任意两个在前 N 个矩上接近的分布 F 和 F ，在其参数上也是接近的。关键事实是我们限制多项式不等式系统的条件数；有多种方法可以做到这一点，但我们将使用量化消除。回顾一下：

定义6.6.7如果存在多元多项式

p_1, \dots, p_n ，使得集合 S 是半代数的。

如果 S 是这样集合的有限并或交

，则 S 是半代数的。

定理6.6.8（塔斯基）半代数集的投影是半代数的。

我们定义以下辅助集合：

$$H(\varepsilon, \delta) = \left\{ \text{对于所有的 } (\Theta, \Theta) : |M_r(\Theta) - M_r(\Theta)| \leq \delta \text{ for } r = 1, 2, \dots, N \Rightarrow \|\Theta - \Theta\| \leq \varepsilon \right\}$$

令 $\varepsilon(\delta)$ 为最小的 ε 作为 δ 的函数：

定理 6.6.9 存在固定常数 C_1, C_2 , such that if $\delta < 1/C_1$ then $\varepsilon(\delta) < C_2 \delta^{1/s}$.

证明：很容易看出我们可以将 $H(\varepsilon, \delta)$ 定义为半代数集的投影，因此使用Tarski定理我们得出 $H(\varepsilon, \delta)$ 也是半代数的。关键观察是因为 $H(\varepsilon, \delta)$ 是半代数的，我们可以选择 ε 作为 δ 的函数，它本身也是 δ 的多项式函数。这里有一些注意事项，因为我们需要证明对于一个固定的 δ ，我们可以选择 ε 严格大于零，而且 ε 和 δ 之间的多项式关系只在 δ 足够小的情况下成立。然而，这些技术问题可以在不太多的工作下解决，参见[23]，主要结果如下。 ■

推论 6.6.10 如果 $|M_r(\Theta) - M_r(\Theta)| \leq \left(\frac{\varepsilon}{C_2}\right)^s$ 那么 $|\Theta - \Theta| \leq \varepsilon$.

因此，存在一个多项式时间算法，可以在多项式时间和样本复杂度为 $\text{poly}(1/\varepsilon)$ 的情况下，学习任何单变量多项式族（其mgf在零的邻域内收敛）的参数，其精度为 ε ；我们可以取足够的样本来估计前 N 个矩时的 ε ，并在参数网格上搜索，与每个矩时匹配的参数集必然在参数距离上接近真实参数。

第7章

矩阵补全

在这里，我们将给出矩阵补全问题的算法，其中我们观察到一个低秩、不相干矩阵 M 的均匀随机条目，并希望设计出能够完全恢复 M 的高效算法。

7.1 背景

矩阵补全问题的常见动机来自推荐系统。具体来说，考虑Netflix问题，我们给出了表示用户 i 对电影 j 评分的评分 $M_{i,j}$ 。我们希望利用这些评分向用户提供好的推荐，标准方法是尝试利用我们对 M 的一些条目的了解来填补 M 的其余部分。让我们更加具体：在上面的例子中，存在一个未知矩阵 $M \in \mathbb{R}^{n \times m}$ ，其行表示用户，列表示电影。对于每个 $(i, j) \in \Omega \subseteq [n] \times [m]$ ，我们给出了值 $M_{i,j}$ 。我们的

目标是精确恢复 M 。

理想情况下，我们希望找到与观测条目 $\{M_{i,j}\}_{(i,j) \in \Omega}$ 一致的最小秩矩阵 X ，但这个问题是 NP 难的。在一些现在的标准假设下，我们将能够提供恢复 M 的有效算法：

- (a) Ω 是均匀随机的
- (b) 矩阵 M 的奇异向量与标准基向量不相关（这样的矩阵被称为不相干的，我们稍后会定义）

事实上，我们将看到，如果 $m \approx mr \log m$ ，那么恢复 M 的算法是高效的，其中 $m \geq n$ 且 $\text{rank}(M) \leq r$ 。这类似

感知, 我们能够从 $O(k \log n/k)$ 线性测量中恢复出一个 k 稀疏信号 x , 这远小于 x 的维度。在这里, 我们也可以从比 M 的维度小得多的观测中恢复出一个低秩矩阵 M 。

让我们检查上面的假设。应该让我们停下来思考的假设是 Ω 是均匀随机的。这有点不自然, 因为如果我们观察到的概率 $M_{i,j}$ 取决于其值本身, 那将更可信。

或者, 如果用户真的喜欢一部电影, 他更有可能给它评分。为了理解第二个假设, 假设 Ω 确实是均匀随机的。考虑

$$M = \Pi \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array} \right] \Pi^T$$

其中 Π 是均匀随机的置换矩阵。 M 是低秩的, 但是除非我们观察到对角线上的所有元素, 否则我们将无法唯一地恢复 M 。事实上, M 的奇异向量包含一些标准基向量; 但是如果我们假设 M 的奇异向量与标准基不相关, 我们可以避免上述问题。

定义 7.1.1 子空间 U 的相干性 $\mu(U)$ (维度为 $\dim(U) = r$) 的

$$\frac{n}{r} \max_i \|P_U e_i\|^2,$$

其中 P_U 表示对 U 的正交投影, e_i 是标准基向量。

很容易看出, 如果我们随机选择 U , 那么 $\mu(U) = \widetilde{O}(1)$ 。同时, 我们有 $1 \leq \mu(U) \leq n/r$, 当 U 包含任何一个 e_i 时, 上界得到满足。现在我们可以看到, 如果我们将 U 设置为上述示例的前奇异向量, 那么 U 具有很高的相干性。我们将需要对 M 的以下条件:

- (a) 令 $M = U \Sigma V^T$, 则 $\mu(U), \mu(V) \leq \mu_0$ 。
- (b) $\|U V^T\|_\infty \leq \frac{\mu_1 \sqrt{r}}{n}$, 其中 $\|\cdot\|_\infty$ 表示任何条目的最大绝对值。

本章的主要结果是:

定理 7.1.2 假设 Ω 是均匀随机选择的。如果存在一个多项式时间算法来恢复 M , 那么它将以很高的概率成功

$$m \geq \max(\mu_1^2, \mu_0) r (n + m) \log^2(n + m)$$

上述定理中的算法基于一种称为矩阵的秩的凸松弛的方法，即核范数。我们将在下一节介绍它，并建立一些性质，但可以将其视为压缩感知中使用的 ℓ_1 最小化方法的类比。这种方法首次在Fazel的论文[58]中引入，Recht、Fazel和Parrilo [104]证明了在矩阵感知的设置中，这种方法可以精确地恢复 M ，这与我们在这里考虑的问题相关。

在一篇具有里程碑意义的论文中，Candes和Recht [33]证明了基于松弛的核范数在矩阵补全中也能成功，并引入了上述假设来证明他们的算法有效。此后，有很多工作改进了对 m 的要求，上述定理和我们的阐述将遵循Recht [103]的一篇最新论文，该论文通过使用Bernstein界的矩阵类比并将其用于一种现在称为“量子高尔夫”的过程中，大大简化了分析，该过程最初由Gross [67]引入。

注7.1.3我们将限制 $M \in \mathbb{R}^{n \times n}$ 并假设 $\mu_0, \mu_1 = \widetilde{O}(1)$ 在我们的分析中，这将减少我们需要跟踪的参数数量。同时，让 $m = n$ 。

7.2 核范数

在这里，我们介绍核范数，它将成为我们矩阵补全算法的基础。我们将按照压缩感知的平行大纲进行。

特别是，一个自然的起点是优化问题：

$$(P0) \quad \min \text{rank}(X) \text{ s.t. } X_{i,j} = M_{i,j} \text{ 对于所有的 } (i, j) \in \Omega$$

这个优化问题是 NP 难的。如果 $\sigma(X)$ 是矩阵 X 的奇异值向量，我们可以将 X 的秩等同于 $\sigma(X)$ 的稀疏度。回想一下，在压缩感知中我们面临着类似的障碍：找到线性方程组的最稀疏解也是 NP 难的，但我们考虑了 ℓ_1 松弛，并证明在各种条件下，这个优化问题可以恢复出最稀疏的解。类似地，我们自然地考虑 $\sigma(X)$ 的 ℓ_1 范数，也称为核范数：

定义 7.2.1 矩阵 X 的核范数，记作 $\|X\|_*$ ，是 $\|\sigma(X)\|_1$ 。

我们将解决凸优化问题：

$$(P1) \quad \text{最小化 } \|X\|_* \text{ 约束 } X_{i,j} = M_{i,j} \text{ 对于所有 } (i, j) \in \Omega$$

我们的目标是证明在条件下, $(P1)$ 的解恰好是 M 。

注意, 这是一个凸优化问题, 因为 $\|X\|_*$ 是一种范数, 而且有多种有效的算法来解决上述问题。

事实上, 对于我们的目的, 一个关键概念是对偶范数。我们不需要完全理解这个概念, 所以我们将其陈述为核范数的特定情况。这个概念给我们一种方法来下界矩阵的核范数:

定义7.2.2 设 $\langle X, B \rangle = \sum_{i,j} X_{i,j} B_{i,j} = \text{trace}(X^T B)$ 表示矩阵内积。

引理 7.2.3 $\|X\|_* = \max_{\|B\| \leq 1} \langle X, B \rangle$ 。

为了对此有所了解, 考虑一种特殊情况, 我们限制 X 和 B 为对角线矩阵。此外, 令 $X = \text{diag}(x)$ 和 $B = \text{diag}(b)$ 。然后 $\|X\|_* = \|x\|_1$ 和约束条件 $\|B\| \leq 1$ (矩阵 B 的谱范数最大为一) 等价于 $\|b\|_\infty \leq 1$ 。因此, 在对角矩阵的特殊情况下, 我们可以得到更熟悉的向量范数的特征描述:

$$\|x\|_1 = \max_{\|b\|_\infty \leq 1} b^T x$$

证明: 我们只证明上述引理的一个方向。我们应该使用什么来证明 X 的核范数。

设 $X = U_X \Sigma_X V_{XT}^T$, 那么我们选择 $B = U_X V_{XT}$ 。然后

$$\langle X, B \rangle = \text{trace}(B^T X) = \text{trace}(V_X U_X^T U_X \Sigma_X V_X^T) = \text{trace}(V_X \Sigma_X V_X^T) = \text{trace}(\Sigma_X) = \|X\|_*$$

这里我们使用了基本事实: $\text{trace}(ABC) = \text{trace}(BCA)$ 。因此, 这证明了 $\|X\|_* \leq \max_{\|B\| \leq 1} \langle X, B \rangle$, 而另一个方向也不难 (见例如[74])。■

我们如何证明 $(P1)$ 的解是 M ? 我们的基本方法将是通过反证法证明。假设不是, 则解为 $M + \frac{Z}{\epsilon}$, 其中 Z 在 Ω 中有支持。我们的目标是构造一个谱范数不超过一的矩阵 B 为

$$\|M + Z\|_* \geq \langle M + Z, B \rangle > \|M\|_*$$

因此 $M + Z$ 不是 $(P1)$ 的最优解。这种策略类似于压缩感知中的一种, 我们假设存在另一个解 w , 它与向量 x 的差向量 y 在感知矩阵 A 的核空间中。我们利用核空间的几何性质证明 w 的 ℓ_1 范数严格大于 x 的范数。

然而，在这里的证明将更加复杂，因为我们的策略是基于 Z 构造 B （而不是依赖于 A 的某些几何性质，无论 y 是什么）。然而，在这里的证明将更加复杂，因为我们的策略是基于 Z 构造 B （而不是依赖于 A 的某些几何性质，无论 y 是什么）。

让我们介绍一些基本的投影算子，在我们的证明中将至关重要。回想一下， $M = U\Sigma V^T$ ，让 u_1, \dots, u_r 是 U 的列，让 v_1, \dots, v_r 是 V 的列。选择 u_{r+1}, \dots, u_n 使得 u_1, \dots, u_n 构成了 \mathbb{R}^n 的正交基——也就是说， u_{r+1}, \dots, u_n 是 U^\perp 的任意正交基。类似地选择 v_{r+1}, \dots, v_n 使得 v_1, \dots, v_n 构成了 \mathbb{R}^n 的正交基。我们将对以下矩阵的线性空间感兴趣：

定义 7.2.4 $T = \text{span}\{u_i v_j^T \mid 1 \leq i \leq r \text{ 或 } 1 \leq j \leq r \text{ 或 两者都是}\}$ 。

然后 $T^\perp = \text{span}\{u_i v_j^T \text{ s.t. } r+1 \leq i, j \leq n\}$ 。我们有 $\dim(T) = r^2 + 2(n-r)r$ and $\dim(T^\perp) = (n-r)^2$ 。此外，我们可以定义线性算子，分别投影到 T 和 T^\perp 上：

$$P_{T^\perp}[Z] = \sum_{i,j=r+1}^n \langle Z, u_i v_j^T \rangle \cdot u_i v_j^T = P_{U^\perp} Z P_{V^\perp}.$$

同样地

$$P_T[Z] = \sum_{(i,j) \in [n] \times [n] - [r+1, n] \times [r+1, n]} \langle Z, u_i v_j^T \rangle \cdot u_i v_j^T = P_U Z + Z P_V - P_U Z P_V.$$

我们现在准备描述定理 7.1.2 的证明大纲。证明将基于以下内容：

- (a) 我们将假设存在一个名为 Y 的辅助矩阵，并证明这足以暗示对于任何在 Ω 中支持的 Z ，有 $\|M + Z\|_* > \|M\|_*$
- (b) 我们将使用量子高尔夫 [67] 来构造这样一个 Y 。

部分 (a)

在这里，我们将陈述我们对辅助矩阵 Y 所需的条件，并证明如果存在这样一个 Y ，那么 M 就是 $(P1)$ 的解。我们要求 Y 在 Ω 中支持

$$(a) \quad \|P_T(Y) - U V^T\|_F \leq \sqrt{r/8n}$$

$$(b) \|P_{T^\perp}(Y)\| \leq 1/2.$$

我们要证明对于任何在 Ω 中支持的 Z , $\|M + Z\|_* > \|M\|_*$ 。回忆一下, 我们想要找到一个谱范数不超过一的矩阵 B , 使得 $\langle M + Z, B \rangle > \|M\|_*$ 。设 U_\perp 和 V_\perp 是 $P_{T^\perp}[Z]$ 的奇异向量。然后考虑

$$B = \begin{bmatrix} U & U_\perp \end{bmatrix} \cdot \begin{bmatrix} V^T \\ V_\perp^T \end{bmatrix} = UV^T + U_\perp V_\perp^T.$$

声明 7.2.5 $\|B\| \leq 1$

证明: 通过构造 $U^T U_\perp = 0$ 和 $V^T V_\perp = 0$, 因此上述表达式对于 B 是其奇异值分解, 因此结论成立。■

因此, 我们可以插入我们对 B 的选择并简化:

$$\begin{aligned} \|M + Z\|_* &\geq \langle M + Z, B \rangle \\ &= \langle M + Z, UV^T + U_\perp V_\perp^T \rangle \\ &= \underbrace{\langle M, UV^T \rangle}_{\|M\|_*} + \langle Z, UV^T + U_\perp V_\perp^T \rangle \end{aligned}$$

在最后一行中, 我们使用了 M 与 $U_\perp V_\perp^T$ 正交的事实。现在, 利用 Y 和 Z 具有不相交的支撑集的事实, 我们可以得出结论:

$$\|M + Z\|_* \geq \|M\|_* + \langle Z, UV^T + U_\perp V_\perp^T - Y \rangle$$

因此, 为了证明本节的主要结果, 只需证明 $\langle Z, UV^T + U_\perp V_\perp^T - Y \rangle > 0$ 。我们可以将这个量展开为其在 T 和 T^\perp 上的投影, 并简化如下:

$$\begin{aligned} \|M + Z\|_* - \|M\|_* &\geq \langle P_T(Z), P_T(UV^T + U_\perp V_\perp^T - Y) \rangle + \langle P_{T^\perp}(Z), P_{T^\perp}(UV^T + U_\perp V_\perp^T - Y) \rangle \\ &\geq \langle P_T(Z), UV^T - P_T(Y) \rangle + \langle P_{T^\perp}(Z), U_\perp V_\perp^T - P_{T^\perp}(Y) \rangle \\ &\geq \langle P_T(Z), UV^T - P_T(Y) \rangle + \|P_{T^\perp}(Z)\|_* - \langle P_{T^\perp}(Z), P_{T^\perp}(Y) \rangle \end{aligned}$$

在最后一行中, 我们使用了 U_\perp 和 V_\perp 是 $P_{T^\perp}[Z]$ 的奇异向量的事实, 因此 $\langle U_\perp V_\perp^T, P_{T^\perp}[Z] \rangle = \|P_{T^\perp}[Z]\|_*$ 。

现在我们可以利用本节中假设的 Y 的性质, 来证明右边的下界。根据 Y 的性质 (a), 我们有 $\|P_T(Y) - UV^T\|_F \leq \sqrt{\frac{r}{2n}}$ 。因此, 我们知道第一项 $\langle P_T(Z), UV^T - P_T(Y) \rangle \geq -\sqrt{\frac{r}{8n}}$ 根据性质 (b) of Y , 我们知道算子范数

of $P_T^\perp(Y)$ 最多为 $1/2$. 因此第三项 $\langle P_{T^\perp}(Z), P_{T^\perp}(Y) \rangle$ 最多为 $\frac{1}{2} \|P_{T^\perp}(Z)\|_*$. 因此

$$\|M + Z\|_* - \|M\|_* \geq -\sqrt{\frac{r}{8n}} \|P_T(Z)\|_F + \frac{1}{2} \|P_{T^\perp}(Z)\|_* \stackrel{?}{\geq} 0$$

我们将证明在 Ω 的选择上, 不等式高概率成立。我们推迟证明这个最后的事实, 因为它和辅助矩阵 Y 的构造都会使用到矩阵伯恩斯坦不等式, 我们将在下一节介绍。

7.3 量子高尔夫

剩下的是构建一个辅助矩阵 Y 并证明在 Ω 上高概率下, 对于任何在 Ω 中支持的矩阵 Z 满足 $\|P_{T^\perp}(Z)\|_* > \frac{r}{2n}$ 为了完成我们在前一节开始的证明, 我们需要计算 $\|P_T(Z)\|_F$ 我们将使用 Gross [67] 提出的方法, 并且我们将遵循 Recht 在 [103] 中的证明, 其中的策略是迭代地构建 Y 在每个阶段, 我们将调用矩阵值随机变量的集中结果来证明 Y 的误差部分几何地减小, 并且我们在构建一个好的辅助矩阵方面取得了快速进展。

首先, 我们将介绍我们将在多个场景中应用的关键集中结果。以下矩阵值 Bernstein 不等式首次出现在 Ahlswede 和 Winter 关于量子信息理论的工作中 [6]。

定理 7.3.1 (非交换伯恩斯坦不等式) 让 $X_1 \dots X_l$ 是大小为 $d \times d$ 的独立均值为 0 的矩阵。让 $\rho_k^2 = \max\{\|\mathbb{E}[X_k X_k^T]\|, \|\mathbb{E}[X_k^T X_k]\|\}$ 并且假设 $\|X_k\| \leq M$ 几乎确定。那么对于 $\tau > 0$,

$$\Pr \left[\left\| \sum_{k=1}^l X_k \right\| > \tau \right] \leq 2d \exp \left\{ \frac{-\tau^2/2}{\sum_k \rho_k^2 + M\tau/3} \right\}$$

如果 $d=1$, 则这是标准的伯恩斯坦不等式。如果 $d>1$ 且矩阵 X_k 是对角矩阵, 则可以通过并集边界和标准伯恩斯坦不等式再次得到该不等式。然而, 为了建立直觉, 请考虑以下玩具问题。让 u_k 是 \mathbb{R}^d 中的随机单位向量, 并且让 $X_k = u_k u_k^T$ 。然后很容易看出 $\rho_k^2 = 1/d$ 。我们需要多少次尝试才能使 $\sum_k X_k$ 在缩放后是否接近单位矩阵? 我们应该预计需要 $\Theta(d \log d)$ 次试验; 即使 u_k 是从标准基向量 $\{e_1 \dots e_d\}$ 中随机选择的, 这个结论仍然成立

这是由于优惠券收集问题。事实上，上述界限证实了我们的直觉，即 $\Theta(d \log d)$ 是必要且充分的。

现在我们将应用上述不等式来构建我们完成证明所需的工具。

定义 7.3.2 令 R_Ω 为将矩阵中除 Ω 之外的所有条目置零的运算符。

引理 7.3.3 如果 Ω 是均匀随机选择的且 $m \geq nr \log n$ ，则高概率下

$$\frac{n^2}{m} \left\| P_T R_\Omega P_T - \frac{m}{n^2} P_T \right\| < \frac{1}{2}$$

注 7.3.4 我们在这里关注的是线性算子在矩阵上的算子范数的界限。设 T 为这样的算子，那么 $\|T\|$ 被定义为

$$\max_{\|Z\|_F \leq 1} \|T(Z)\|_F$$

我们将解释这个界限如何适用于矩阵伯恩斯坦不等式的框架，但完整的证明请参见[103]。注意 $\mathbb{E}[P_T R_\Omega P_T] = P_T \mathbb{E}[R_\Omega] P_T = \frac{m}{n^2} P_T$ 所以我们只需要证明 $P_T R_\Omega P_T$ 与其期望值之间的差距不会太大。设 e_1, e_2, \dots, e_d 为标准基向量。然后我们可以展开：

$$\begin{aligned} P_T(Z) &= \sum_{a,b} \langle P_T(Z), e_a e_b^T \rangle e_a e_b^T \\ &= \sum_{a,b} \langle P_T(Z), e_a e_b^T \rangle e_a e_b^T \\ &= \sum_{a,b} \langle Z, P_T(e_a e_b^T) \rangle e_a e_b^T \end{aligned}$$

因此 $R_\Omega P_T(Z) = \sum_{(a,b) \in \Omega} \langle Z, P_T(e_a e_b^T) \rangle e_a e_b^T$ 最后我们得出结论

$$P_T R_\Omega P_T(Z) = \sum_{(a,b) \in \Omega} \langle Z, P_T(e_a e_b^T) \rangle P_T(e_a e_b^T)$$

我们可以将 $P_T R_\Omega P_T$ 看作是形式为 $\tau_{a,b} : Z \rightarrow$ 的随机算子的总和 $\langle Z, P_T(e_a e_b^T) \rangle P_T(e_a e_b^T)$ ，引理通过将矩阵Bernstein不等式应用于随机算子 $\sum_{(a,b) \in \Omega} \tau_{a,b}$ 。

我们现在可以完成部分(a)的延迟证明：

引理7.3.5如果 Ω 是均匀随机选择的, 并且 $m \geq nr \log n$, 那么对于任何在 Ω 中支持的 Z , 我们有很高的概率

$$\|P_{T^\perp}(Z)\|_* > \sqrt{\frac{r}{2n}} \|P_T(Z)\|_F$$

证明: 利用引理7.3.3和算子范数的定义 (见备注)

我们有

$$\left\langle Z, P_T R_\Omega P_T Z - \frac{m}{n^2} P_T Z \right\rangle \geq -\frac{m}{2n^2} \|Z\|_F^2$$

此外, 我们可以将左边界上界化为:

$$\begin{aligned} \langle Z, P_T R_\Omega P_T Z \rangle &= \langle Z, P_T R_\Omega^2 P_T Z \rangle = \|R_\Omega(Z - P_{T^\perp}(Z))\|_F^2 \\ &= \|R_\Omega(P_{T^\perp}(Z))\|_F^2 \leq \|P_{T^\perp}(Z)\|_F^2 \end{aligned}$$

在最后一行中, 我们使用了 Z 在 Ω 中的支持, 所以 $R_\Omega(Z) = 0$ 。因此, 我们有

$$\|P_{T^\perp}(Z)\|_F^2 \geq \frac{m}{n^2} \|P_T(Z)\|_F^2 - \frac{m}{2n^2} \|Z\|_F^2$$

我们可以利用 $\|Z\|_F^2 = \|P_{T^\perp}(Z)\|_F^2 + \|P_T(Z)\|_F^2$ 并得出结论 $\|P_{T^\perp}(Z)\|_F^2 \geq \frac{m}{4n^2} \|P_T(Z)\|_F^2$ 。现在我们可以完成引理的证明

$$\begin{aligned} \|P_{T^\perp}(Z)\|_*^2 &\geq \|P_{T^\perp}(Z)\|_F^2 \geq \frac{m}{4n^2} \|P_T(Z)\|_F^2 \\ &> \frac{r}{2n} \|P_T(Z)\|_F^2 \end{aligned}$$

■

剩下的就是要证明我们使用的辅助矩阵 Y 实际上确实存在 (高概率下)。

回想一下, 我们要求 Y 在 Ω 中支持, 并且 $\|P_T(Y) - UV^T\|_F \leq \frac{1}{\sqrt{m}}$ 。基本思想是将 Ω 分解为不相交的集合 $\Omega_1, \Omega_2, \dots, \Omega_p$, 其中 $p = \log n$, 并使用每个观测集合来推进剩余的 $P_T(Y) - UV^T$ 。更准确地说, 初始化 $Y_0 = 0$, 此时剩余部分为 $W_0 = UV^T$ 。然后设置

$$Y_{i+1} = Y_i + \frac{n^2}{m} R_{\Omega_{i+1}}(W_i)$$

并更新 $W_{i+1} = UV^T - P_T(Y_{i+1})$ 。很容易看出 $\mathbb{E}[\frac{n^2}{m} R_{\Omega_{i+1}}] = I$ 。直观上, 这意味着在每一步 $Y_{i+1} - Y_i$ 都是 W_i 的无偏估计量, 所以我们应该期望余项迅速减小 (在这里, 我们将依赖于从非交换伯恩斯坦不等式推导出的浓度界)。现在

我们可以解释量子高尔夫的命名方式；在每一步中，我们将高尔夫球击向洞口，但这里我们的目标是近似矩阵 UV^T ，这是由于各种原因在量子力学中出现的问题类型。

很容易看出 $Y = \sum_i Y_i$ 在 Ω 中得到支持，并且 $P_T(W_i) = W_i$ 对于所有 i 。因此我们可以计算

$$\begin{aligned} \|P_T(Y_i) - UV^T\|_F &= \left\| W_{i-1} - P_T \frac{n^2}{m} R_{\Omega_i} W_{i-1} \right\|_F = \left\| P_T W_{i-1} - P_T \frac{n^2}{m} R_{\Omega_i} P_T W_{i-1} \right\|_F \\ &\leq \frac{n^2}{m} \left\| P_T R_{\Omega} P_T - \frac{m}{n^2} P_T \right\| \leq \frac{1}{2} \|W_{i-1}\|_F \end{aligned}$$

最后一个不等式来自引理7.3.3。因此，剩余项的Frobenius范数几何地减小，很容易保证 Y 满足条件(a)。

更技术性的部分是证明 Y 也满足条件(b)。然而，直觉是 $\|P_{T^\perp}(Y_1)\|$ 本身不会太大，而且由于剩余项 W_i 的范数几何地减小，我们应该期望 $\|P_{T^\perp}(Y_i)\|$ 也会减小，因此大部分贡献来自

$$\|P_{T^\perp}(Y)\| \leq \sum_i \|P_{T^\perp}(Y_i)\|$$

来自第一项。有关详细信息，请参见[103]。这完成了证明，即计算凸程序的解确实找到 M ，前提是 M 不相干且 $|\Omega| \geq \max(\mu^2_1, \mu_0) r(n+m) \log^2(n+m)$ 。

参考文献

- [1] D. Achlioptas和F. McSherry。关于混合分布的谱学习。在 *COLT*中，第458-469页，2005年。
- [2] A. Agarwal, A. Anandkumar, P. Jain, P. Netrapalli, R. Tandon通过交替最小化学习稀疏使用过完备字典 *arxiv:1310.7991*, 2013年
- [3] A. Agarwal, A. Anandkumar, P. Netrapalli 稀疏过完备字典的精确恢复 *arxiv:1309.1952*, 2013
- [4] M. Aharon. 用于信号稀疏表示的过完备字典. 博士论文, 2006.
- [5] M. Aharon, M. Elad and A. Bruckstein. K-SVD: 一种用于设计稀疏表示的过完备字典的算法. *IEEE信号处理杂志*, 54(11):4311–4322, 2006.
- [6] R. Ahlswede and A. Winter. 通过量子通道进行识别的强对偶. *IEEE信息论* 48(3):569–579, 2002.
- [7] Noga Alon. 高等代数工具. 在组合数学手册中, 页码 1749–1783, 1996.
- [8] A. Anandkumar, D. Foster, D. Hsu, S. Kakade, Y. Liu. 一种潜在狄利克雷分配的谱算法。在 *NIPS*, 2012年的926-934页。
- [9] A. Anandkumar, R. Ge, D. Hsu和S. Kakade. 一种张量谱方法用于学习混合成员社区模型。在 *COLT*, 2013年的867-881页。
- [10] A. Anandkumar, D. Hsu和S. Kakade. 一种用于隐藏马尔可夫模型和多视图混合模型的矩方法。在 *COLT*, 2012年的33.1-33.34页。

- [11] J. Anderson, M. Belkin, N. Goyal, L Rademacher和J. Voss. 对于学习大型高斯混合模型, 维度越多越好. *arxiv:1311.2891*, 2013.
- [12] S. Arora, R. Ge, Y. Halpern, D. Mimno, A. Moitra, D. Sontag, Y. Wu and M. Zhu. 具有可证明保证的主题建模的实用算法. 在 *ICML*, 2013年的第280-288页。
- [13] S. Arora, R. Ge, R. Kannan and A. Moitra. 在 *STOC*, 2012年的第145-162页计算非负矩阵分解。
- [14] S. Arora, R. Ge and A. Moitra. 学习主题模型-超越奇异值分解. 在 *FOCS*, 2012年的第1-10页。
- [15] S. Arora, R. Ge and A. Moitra. 学习不一致和过完备字典的新算法. *arxiv:1308.6273*, 2013年
- [16] S. Arora, R. Ge, A. Moitra和S. Sachdeva. 可证明的未知高斯噪声ICA, 以及对高斯混合和自编码器的影响. 在MIPS中, 页码2384-2392, 2012年。
- [17] S. Arora, R. Ge, S. Sachdeva和G. Schoenebeck. 在社交网络中找到重叠的社区: 走向严格的方法. 在*EC*中, 2012年。
- [18] S. Arora和R. Kannan. 学习分离的非球形高斯混合物. 应用概率学年刊, 页码69-92, 2005年。
- [19] M. Balcan, A. Blum和A. Gupta. 近似稳定性下的聚类. *ACM杂志*, 2013年。
- [20] M. Balcan, A. Blum和N. Srebro. 关于相似性函数学习的理论. 机器学习, 页码89-112, 2008年。
- [21] M. Balcan, C. Borgs, M. Braverman, J. Chayes 和 S-H Teng. 寻找内生形成的社区. 在 *SODA*, 2013年。
- [22] M. Belkin 和 K. Sinha. 向学习具有任意分离的高斯混合模型靠拢. 在 *COLT*, 2010年的407-419页。
- [23] M. Belkin 和 K. Sinha. 多项式学习分布族. 在 *FOCS*, 2010年的103-112页。
- [24] Q. Berthet 和 P. Rigollet. 稀疏主成分检测的复杂性理论下界. 在 *COLT*, 2013年的1046-1066页。

- [25] A. Bhaskara, M. Charikar 和 A. Vijayaraghavan. 张量分解的唯一性及其在多项式可识别性中的应用. *arxiv:1304.8087*, 2013年.
- [26] A. Bhaskara, M. Charikar, A. Moitra and A. Vijayaraghavan. 张量分解的平滑分析. 在 *STOC*, 2014.
- [27] V. Bittorf, B. Recht, C. Re, and J. Tropp. 用线性规划分解非负矩阵. 在 *NIPS*, 2012.
- [28] D. Blei. 概率主题模型入门. *ACM通信*, 页码77-84, 2012.
- [29] D. Blei and J. Lafferty. 一种相关主题模型的科学模型. *应用统计学年鉴*, 页码17-35, 2007.
- [30] D. Blei, A. Ng and M. Jordan. 潜在狄利克雷分配. *机器学习研究杂志*, 页码993-1022, 2003.
- [31] A. Blum, A. Kalai 和 H. Wasserman. 噪声容忍学习, 奇偶问题和统计查询模型. *ACM杂志* 50: 506-519, 2003年.
- [32] S. C. Brubaker 和 S. Vempala. 同方差PCA和仿射不变聚类. 在 *FOCS*, 第551-560页, 2008年.
- [33] E. Candes 和 B. Recht. 凸优化的精确矩阵补全. *计算数学基础*, 第717-772页, 2008年.
- [34] E. Candes, J. Romberg 和 T. Tao. 从不完全和不准确的测量中稳定信号恢复. *纯粹和应用数学通信*, pp. 1207-1223, 2006年.
- [35] E. Candes 和 T. Tao. 线性规划解码. *IEEE信息论*, 51(12):4203-4215, 2005年.
- [36] J. Chang. 在进化树上完全重建马尔可夫模型: 可识别性和一致性. *数学生物学*, 137(1): 51-73, 1996年.
- [37] K. Chaudhuri和S. Rao. 使用相关性和独立性学习产品分布的混合物. 在 *COLT*, 第9-20页, 2008年.
- [38] K. Chaudhuri和S. Rao. 超越高斯分布: 学习重尾分布的谱方法. 在 *COLT*, 第21-32页, 2008年.

- [39] S. Chen, D. Donoho和M. Saunders. 基 Pursuit 的原子分解。
SIAM J. on Scientific Computing, 20(1): 33-61, 1998年。
- [40] A. Cohen, W. Dahmen和R. DeVore. 压缩感知和最佳 k 项近似。 *Journal of the AMS*, 第211-231页, 2009年。
- [41] J. Cohen and U. Rothblum. 非负矩阵的秩、分解和因子-
化。 *线性代数及其应用*, 第
149-168页, 1993年。
- [42] P. Comon. 独立成分分析: 一个新概念? 信号处理-
学, 第287-314页, 1994年。
- [43] A. Dasgupta. 统计学和概率的渐近理论。 *Springer*, 2008年。
- [44] A. Dasgupta, J. Hopcroft, J. Kleinberg, and M. Sandler. 学习重尾分布的混合
物。 在 *FOCS*, 第491-500页, 2005年。
- [45] S. Dasgupta. 学习高斯混合物。 在 *FOCS*, 第634-644页, 1999年。
- [46] S. Dasgupta and L. J. Schulman. 高斯
混合物的两轮变体EM算法。 在 *UAI*, 第152-159页, 2000年。
- [47] G. Davis, S. Mallat和M. Avellaneda. 贪婪自适应逼近。 *J.*
构造逼近, 13:57-98, 1997年。
- [48] L. De Lathauwer, J. Castaing和J. Cardoso. 基于四阶累积量的盲识别欠定混
合物。 *IEEE信号处理杂志*, 55(6):2965-2973, 2007年。
- [49] S. Deerwester, S. Dumais, T. Landauer, G. Furnas和R. Harshman. 通过潜在语义分析进行索引
。 *JASIS*, 页码391-407, 1990年。
- [50] A.P. Dempster, N.M. Laird和D.B. Rubin. 通过EM算法从不完整数据中获得
最大似然估计。 *Journal of the Royal Statistical Society Series B*, 页码1-38, 1
977年。
- [51] D. Donoho和M. Elad. 通过 ℓ_1 -最小化在一般(非正交)字典中的最优稀疏表示
。 *PNAS*, 100 (5) : 2197-2202, 2003年。
- [52] D. Donoho和X. Huo. 不确定性原理和理想原子分解。
IEEE Trans. on IT, 47 (7) : 2845-2862, 1999年。
- [53] D. Donoho和P. Stark. 不确定性原理和信号恢复。 *SIAM J.*
on Appl. Math., 49 (3) : 906-931, 1989年。

- [54] D. Donoho和V. Stodden. 何时非负矩阵分解能够正确地分解成部分? 在 *NIPS*, 2003年。
- [55] M. Elad. 稀疏和冗余表示。Springer, 2010年。
- [56] K. Engan, S. Aase和J. Hakon-Husoy. 最佳方向法进行框架设计。 *ICASSP*, 5:2443–2446, 1999年。
- [57] P. Erdos, M. Steel, L. Szekely和T. Warnow. 几个对数足以构建(几乎)所有的树。 *I.随机结构和算法*14:153-184, 1997年。
- [58] M. Fazel. 矩阵秩最小化及其应用。 博士论文, 斯坦福大学, 2002年。
- [59] U. Feige和R. Krauthgamer. 在半随机图中找到和证明一个大的隐藏团。 *随机结构和算法*, 页码195–208, 2009年。
- [60] U. Feige和J. Kilian. 半随机图问题的启发式算法。 *JCSS*, 页码639–671, 2001年。
- [61] J. Feldman, R. A. Servedio, and R. O’Donnell. 在没有分离假设的情况下, PAC学习轴对齐高斯混合模型。 在 *COLT*, 2006年的第20-34页。
- [62] A. Frieze, M. Jerrum, R. Kannan. 学习线性变换。 在 *FOCS*, 1996年的第359-368页。
- [63] A. Garnaev and E. Gluskin. 欧几里得球的宽度。 *Sovieth Math. Dokl.*, 1984年的第200-204页。
- [64] A. Gilbert, S. Muthukrishnan and M. Strauss. 使用相干性逼近冗余字典上的函数。 在 *SODA*, 2003年。
- [65] N. Gillis. hotttopixx的鲁棒性分析, 一个用于分解非负矩阵的线性规划模型。 *arxiv:1211.6687*, 2012年。
- [66] N. Goyal, S. Vempala和Y. Xiao. Fourier PCA. 在 *STOC*, 2014年。
- [67] D. Gross. 从任何基础中的少量系数中恢复低秩矩阵。 *arxiv:0910.1879*, 2009年。
- [68] D. Gross, Y-K Liu, S. Flammia, S. Becker和J. Eisert. 通过压缩感知进行量子态重构。 *Physical Review Letters*, 105(15), 2010年。
- [69] V. Guruswami, J. Lee和A. Razborov. 通过扩展码几乎欧几里得子空间 ℓ_1^n 通过 expander codes. *Combinatorica*, 30(1):47–68, 2010年。

- [70] R. Harshman. PARFAC过程的基础: 一个“解释性”多模式因子分析的模型和条件. *UCLA语音学工作论文*, 页码1–84, 1970年.
- [71] J. Håstad. 张量秩是 NP -完全的. *算法杂志*, 11(4):644–654, 1990年.
- [72] C. Hillar 和 L-H. Lim. 大多数张量问题是 NP -难的. *arxiv:0911.1393v4*, 2013年
- [73] T. Hofmann. 概率潜在语义分析. 在 *UAI*, 页码289–296, 1999年.
- [74] R. Horn 和 C. Johnson. 矩阵分析. 剑桥大学出版社, 1990年.
- [75] D. Hsu 和 S. Kakade. 学习球面高斯混合模型: 矩方法和谱分解. 在 *ITCS*, 页码11–20, 2013年.
- [76] P. J. Huber. 投影追踪. *统计学年鉴*13:435–475, 1985年.
- [77] R. A. Hummel 和 B. C. Gidas. 零交叉和热方程. *Courant 数学科学研究所*TR-111, 1984年.
- [78] R. Impagliazzo和R. Paturi. 关于 k -SAT的复杂性. *计算机和系统科学*62(2):pp. 367–375, 2001.
- [79] A. T. Kalai, A. Moitra和G. Valiant. 高效学习两个高斯混合模型. 在 *STOC* 中, 页码为553–562, 2010年.
- [80] B. Kashin和V. Temlyakov. 关于压缩感知的一点备注. 手稿, 2007年.
- [81] L. Khachiyan. 关于在矩阵中逼近极值行列式复杂性的问题. *复杂性杂志*, 页码为138–153, 1995年.
- [82] D. Koller和N. Friedman. 概率图模型. 麻省理工学院出版社, 2009年.
- [83] J. Kruskal. 三维数组: 三线性分解的秩和唯一性及其在算术复杂性和统计学中的应用. *线性代数及其应用*, 页码为95–138, 1997年.
- [84] A. Kumar, V. Sindhvani和P. Kambadur. 快速锥形壳算法用于近可分非负矩阵分解. 在 *ICML*, 第231–239页, 2013年.

- [85] D. Lee和H. Seung。通过非负矩阵分解学习对象的部分。自然, 第788-791页, 1999年。
- [86] D. Lee和H. Seung。非负矩阵分解算法。在 *NIPS*, 第556-562页, 2000年。
- [87] S. Leurgans, R. Ross和R. Abel。三维数组的分解。 *SIAM 矩阵分析和应用杂志*, 第14卷第4期: 1064-1083页, 1993年。
- [88] M. Lewicki和T. Sejnowski。学习过完备表示。神经计算, 第12卷: 337-365页, 2000年。
- [89] W. Li 和 A. McCallum. Pachinko 分配: DAG 结构的主题相关性混合模型. *ICML*, pp. 633-640, 2007.
- [90] B. Lindsay.混合模型: 理论、几何和应用. 数学统计研究所, 1995.
- [91] F. McSherry. 随机图的谱分割. 在 *FOCS*, 页码 529-537, 2001.
- [92] S. Mallat.信号处理的小波导览. Academic-Press, 1998.
- [93] S. Mallat 和 Z. Zhang. 带有时频字典的匹配追踪. *IEEE 信号处理杂志*, 41(12):3397-3415, 1993.
- [94] A. Moitra. 计算非负秩的几乎最优算法. 在 *SODA*, 页码 1454-1464, 2003.
- [95] A. Moitra和G. Valiant。设置高斯混合物的多项式可学习性。在 *FOCS*, 页码 93-102, 2010年。
- [96] E. Mossel和S. Roch。学习非奇异的系统发生和隐藏马尔可夫模型。在 *STOC*, 页码366-375, 2005年。
- [97] B. Olshausen和B. Field。使用过完备基函数进行稀疏编码: V1所采用的策略? 视觉研究, 37(23): 331-3325, 1997年。
- [98] C. Papadimitriou, P. Raghavan, H. Tamaki和S. Vempala。潜在语义索引: 概率分析。 *JCSS*, 页码217-235, 2000年。
- [99] Y. Pati, R. Rezaifar, P. Krishnaprasad。正交匹配追踪: 递归函数逼近及其在小波分解中的应用。 *Asilomar信号、系统和计算机会议*, 页码40-44, 1993年。

- [100] K. Pearson. 对进化的数学理论的贡献。《皇家学会A类哲学交易》, 1894年。
- [101] Y. Rabani, L. Schulman和C. Swamy. 学习大离散域上任意分布的混合。在 *IT CS* 2014年。
- [102] R. Raz. 张量秩和算术公式的下界。在 *STOC*中, 页码659-666, 2010年。
- [103] B. Recht. 矩阵补全的简化方法。《机器学习研究》, 页码3413-3430, 2011年。
- [104] B. Recht, M. Fazel和P. Parrilo. 通过核范数最小化保证最小秩解的矩阵方程。《*SIAM评论*》, 页码471-501, 2010年。
- [105] R. A. Redner和H. F. Walker. 混合密度、最大似然和EM算法。《*SIAM评论*》, 26(2):195-239, 1984年。
- [106] J. Renegar. 关于实数的一阶理论的计算复杂性和几何性质。符号计算杂志, 255-352页, 1991年。
- [107] A. Seidenberg. 一种新的初等代数决策方法。数学年鉴, 365-374页, 1954年。
- [108] D. Spielman, H. Wang和J. Wright. 稀疏字典的精确恢复。机器学习研究杂志, 2012年。
- [109] M. Steel. 在马尔可夫模型下从叶子着色生成树的恢复。应用数学通讯7:19-24, 1994年。
- [110] A. Tarski. 一种初等代数和几何的决策方法。加利福尼亚大学出版社, 1951年。
- [111] H. Teicher. 混合物的可识别性。数理统计年鉴, 244-248页, 1961年。
- [112] J. Tropp. 贪婪是好的: 稀疏逼近的算法结果。 *IEEE IT*, 50(10):2231-2242, 2004.
- [113] J. Tropp, A. Gilbert, S. Muthukrishnan和M. Strauss. 改进的稀疏逼近在准不相干字典上。 *IEEE国际图像处理会议*, 2003.
- [114] L. Valiant. 可学习的理论。 *ACM通讯*, 27(11):1134-1142, 1984.

- [115] S. Vavasis. 非负矩阵分解的复杂性。 *SIAM优化杂志*, 页1364-1377, 2009.
- [116] S. Vempala, Y. Xiao. 从局部最优解中学习：通过高阶PCA学习子空间juntas。
。 *Arxiv:abs/1108.3329*, 2011.
- [117] S. Vempala和G. Wang。 用于学习混合模型的谱算法。
计算机与系统科学杂志, 页码841-860, 2004年。
- [118] M. Wainwright和M. Jordan。 图形模型，指数族和
变分推断。 *机器学习基础与趋势*, 页码
1-305, 2008年。
- [119] P. Wedin。 奇异值分解相关的扰动界限。 *BIT*, 第12卷： 99-111, 1972年。