arXiv:2305.03110v1 [cs.DS] 4 May 2023

一种稀疏的 使用快速哈希的 Johnson-Lindenstrauss 变换

Jakob Bæk TejsHouen*
BARC，哥本哈根大学
jakob@tejs.dk

Mikkel Thorup*BARC，哥本哈根大学
mikkel2thorup@gmail.com

摘要

Kane 和 Nelson 提出的稀疏 Johnson-Lindenstrauss 变换（SODA 2012）给出了一种线性降维映射 $A \in \mathbb{R}^{m\times u}$，将其映射到 ℓ_2，并以 $1-\delta$ 的概率将距离畸变保留在 $1+\varepsilon$ 以内，其中 $m = O(\varepsilon^{-2}\log 1/\delta)$，且 A 的每一列都包含 $O(\varepsilon m)$ 个非零项。此前对稀疏 Johnson-Lindenstrauss 变换的分析都假设可以使用一个 $\Omega(\log 1/\delta)$-路独立的哈希函数。本文的主要贡献是，在对哈希函数要求较少的条件下，对稀疏 Johnson-Lindenstrauss 变换进行了更一般的分析。我们还证明，Dahlgaard、Knudsen、Rotenberg 和 Thorup 提出的混合制表哈希函数（FOCS 2015）满足我们的分析条件，从而首次对一种可与实用哈希函数配合使用的稀疏 Johnson-Lindenstrauss 变换进行了分析。

*本研究得到 VILLUM 基金会哥本哈根基础算法研究中心（BARC）研究员资助项目 16582 的支持。

1 引言

降维是一种常用技术，可在处理高维数据时提升速度。其基本思想是将一组点 $X \subseteq \mathbb{R}^u$ 映射到较低维空间，同时近似保留其几何结构。Johnson-Lindenstrauss 引理 [24] 是这方面的一项基础性成果。

引理 1（[24]）。对于任意 $0 < \varepsilon < 1$、整数 n、u 以及满足 $|X| = n$ 的 $X \subseteq \mathbb{R}^u$，存在一个映射 $f\colon X \to \mathbb{R}^m$，满足 $m = O(\varepsilon^{-2} \log n)$，使得

$$\forall w, w' \in X,\ \left| \left\| f(w) - f(w') \right\|_2 - \left\| w - w' \right\|_2 \right| \le \varepsilon \left\| w - w' \right\|_2 .$$

[6, 30] 中已经证明，对于几乎整个 n、u、ε 的取值范围，目标维数 m 都是最优的。更准确地说，对于任意 n、u、ε，存在一个点集 $X \subseteq \mathbb{R}^u$，满足 $|X| = n$，使得对于任意映射 $f\colon X \to \mathbb{R}^m$，若欧几里得范数的畸变至多为 $(1 \pm \varepsilon)$，则必须满足 $m = \Omega(\min\{u, n, \varepsilon^{-2}\log(\varepsilon^2 n)\})$。

所有已知的 Johnson–Lindenstrauss 引理证明都构造了一个线性映射 f。Johnson 和 Lindenstrauss 最初的证明 [24] 选择了 $f(x) = \Pi x$，其中 $\Pi \in \mathbb{R}^{m \times u}$ 是投影到随机 m 维子空间上的适当缩放的正交投影。另一种简单的构造是令 $f(x) = \frac{1}{\sqrt{m}} A x$，其中 $A \in \mathbb{R}^{m \times u}$，且每个条目都是相互独立的 Rademacher 变量。[1]在这两种情况下，都可以证明，只要 $m = \Omega(\varepsilon^{-2} \log 1/\delta)$，那么

$$\forall w \in \mathbb{R}^u,\ \Pr\left[\left| \|f(w)\|_2^2 - \|w\|_2^2 \right| \ge \varepsilon \|w\|_2^2 \right] \le \delta. \tag{1}$$

通过令 $\delta < 1/\binom{n}{2}$，并对所有成对的 z、$z' \in X$ 取 $w = z - z'$，再结合并集界，即可得到 Johnson-Lindenstrauss 引理。(1) 也称为分布式 Johnson-Lindenstrauss 引理。已有研究证明，目标维数 m 是紧的；更准确地说，m 必须至少为 $\Omega(\min\{u, \varepsilon^{-2}\log\ 1/\delta\})$ [23, 26]。

稀疏 Johnson-Lindenstrauss 变换。加快嵌入时间的一种方法，是用稀疏矩阵替换上述构造中的稠密 A。Achlioptas 于 [3] 首次在这方面取得进展，他证明 A 可以选为具有独立同分布条目的矩阵，其中 $A_{ij} = 0$ 以概率 $2/3$ 取值，否则 A_{ij} 从 $\pm\sqrt{\frac{3}{m}}$ 中均匀选取。他证明，这种构造能够达到对 Johnson-Lindenstrauss 引理的最佳分析所得到的相同 m。因此，这种方法基本上实现了 3 倍加速，但渐近嵌入时间仍为 $O(m\|x\|_0)$，其中 $\|x\|_0$ 是 x 中非零元素的数量。

为了改善渐近嵌入时间，Kane 和 Nelson 在 [28], 遵循 [14, 27, 8], 中的工作，引入了稀疏 Johnson-Lindenstrauss 变换，将维度降至本质上最优的维度 $m = O(\varepsilon^{-2}\log n)$，且每列仅有 $s = O(\varepsilon^{-1}\log n)$ 个非零项。这样可将嵌入时间加快至 $O(\varepsilon^{-1}\log\ n\|x\|_0) = O(\varepsilon m\|x\|_0)$，从而使嵌入时间提升 ε^{-1} 倍。该变换几乎达到了 Nelson 和 Nguyen 给出的稀疏度下界 [31]：他们证明，任何稀疏矩阵每列至少需要 $s = \Omega(\varepsilon^{-1}\log(n)/\log(1/\varepsilon))$ 个非零项。

[1]Rademacher 变量 X ± 1 $[X = 1] =$ $[X = -1] =$ $\frac{1}{2}$ 是一个从 中均匀选取的随机变量，即 Pr Pr。

使用哈希。当输入维度 u 很大时，显式存储矩阵 A 并不可行。作为替代，我们使用哈希函数来计算 A 的非零项。遗憾的是，先前对稀疏 Johnson-Lindenstrauss 变换 [28, 10] 的分析假设可以访问一个 $\Omega(\log 1/\delta)$-wise 独立哈希函数，而这种函数效率低下。这引出了一个自然的问题：

> 对于稀疏 Johnson-Lindenstrauss 变换，要使其正常工作，哈希函数需要具备哪些充分性质?

本工作的目标是推动这一问题的研究进展。具体而言，我们针对一种稀疏 Johnson-Lindenstrauss 变换提出了一种新的分析方法，该方法对哈希函数的假设更少。这一改进的分析使我们能够得出结论：存在一种采用混合制表哈希且具有高效率的稀疏 Johnson-Lindenstrauss 变换。

混合制表哈希。在介绍混合制表哈希之前，我们先讨论由 Zobrist [40] 提出的简单制表哈希。简单制表哈希接受一个整数参数 $c > 1$，并将键 $x \in [u] = \{0, \ldots, u-1\}$ 视为由 c 个字符组成的向量，x_0、...、$x_{c-1} \in \Sigma = [u^{1/c}]$。对于每个字符，我们初始化一个完全随机的表 T_i: $\Sigma \to [2^r]$，然后按如下方式计算 x 的哈希值。

$$h(x) = T_0[x_0] \oplus \ldots \oplus T_{c-1}[x_{c-1}],$$

其中 $\oplus$ 是按位异或运算。我们称 h 为具有c 个字符的简单制表哈希函数。

现在我们可以定义混合制表哈希，它是简单制表哈希的一种变体，于 [11]中提出。与简单制表哈希一样，混合制表哈希以$c > 1$ 为参数，并且还需要一个整数参数 $d \geq 1$。同样，我们将键 $x \in [u]$视为由 c个字符组成的向量： x_0、.....、$x_{c-1} \in \Sigma = [u^{1/c}]$。然后令 h_1: $\Sigma^c \to [2^r]$、 h_2: $\Sigma^c \to \Sigma^d$和 h_3: $\Sigma^d \to [2^r]$ 为相互独立的简单制表哈希函数。混合制表哈希定义如下

$$h(x) = h_1(x) \oplus h_3(h_2(x)).$$

我们称 h 为具有 c个字符和 d 个派生字符的混合制表哈希函数。我们将 h_2（x） $\in \Sigma^d$ 称为派生字符。通过将 h_1 和 h_2 存储为一个表（其条目位于 $[2^r] \times \Sigma^d$中），可以高效地实现混合制表哈希，因此整个哈希函数只需进行 $c + d$ 次查表即可计算。

我们的贡献。我们的主要贡献是对一种稀疏 Johnson-Lindenstrauss 变换提出了一种新的分析方法，该方法不依赖于哈希函数的高阶独立性。相反，我们证明，只要哈希函数支持去耦分解，并结合强集中界，就足以满足要求。

我们证明，混合制表哈希满足这些条件。这首次实现了一种实用的哈希函数，使其能够支持稀疏 Johnson-Lindenstrauss 变换。

1.1 稀疏 Johnson-Lindenstrauss 变换

如前文所述，稀疏 Johnson-Lindenstrauss 变换由 Kane 和 Nelson [28]提出，他们给出了两种稀疏度相同的不同构造。后来，一种更简单的

分析见于 [10]，该分析还将结果推广到了更一般的构造类。在本文中，我们将只关注其中一种构造，具体描述如下。

在讨论稀疏 Johnson-Lindenstrauss 变换的构造之前，我们先考虑相关的 CountSketch。CountSketch 在 [9] 中被提出，并在 [36]中针对降维进行了分析。在 CountSketch 中，我们按如下方式构造矩阵 A：选取一个两两独立的哈希函数 h: $[u] \to [m]$，以及一个 4-路独立的符号函数 σ: $[u] \to \{-1, 1\}$。对于每个 $x \in [u]$，令 $A_{h(x),x} = \sigma(x)$，并将第 x 列的其余元素设为 0。显然，这种构造使每列恰好有 1 个非零元素。[36] 中证明了，若 $m = \Omega(\varepsilon^{-2}\delta^{-1})$，则该构造满足分布式 Johnson-Lindenstrauss 引理，即式 (1)。该结论可通过对任意 $x \in \mathbb{R}^d$ 界定 $\|Ax\|_2^2 - \|x\|_2^2$ 的二阶矩，再应用切比雪夫不等式得出。

目标维数 m 对失败概率 δ 的依赖关系较差，这是因为我们只使用了二阶矩。因此，人们可能希望通过考察更高阶矩来改善这种依赖关系。遗憾的是，对于一般的 $x \in \mathbb{R}^d$，这种依赖关系无法改善；只有当 $\|x\|_\infty^2 / \|x\|_2^2$ 较小时，才有可能改善。$\|x\|_\infty^2 / \|x\|_2^2$ 究竟需要小到什么程度，[17] 已经给出了证明。因此，为了改善对 δ 的依赖关系，我们需要增加每列中的非零元素数量。

现在我们可以描述稀疏 Johnson-Lindenstrauss 变换的构造了。该构造将 s 个 CountSketch 矩阵拼接起来，并将所得矩阵缩放 $\frac{1}{\sqrt{s}}$。显然，这种构造的每列有 s 个非零项；并且如 [28, 10] 中所示，如果 $s = \Omega(\varepsilon^{-1}\log(1/\delta))$，那么我们就可以得到最优目标维数$m = O(\varepsilon^{-2}\log(1/\delta))$。更形式化地，我们按如下方式构造矩阵 A：

1. 我们选取一个 哈希函数 h: $[s] \times [u] \to [m/s]$ 和符号函数 σ: $[s] \times [u] \to \{-1, 1\}$。

2. 对每个 $x \in [u]$，我们将 $A_{i \cdot m/s + h(i,x), x} = \sigma_{(i,x)}\sqrt{s}$ 对每个 $i \in [s]$ 设为该值，并将第 x 列的其余元素设为 0。

在先前的分析中 [28, 10], ，已经证明，如果 h 和 σ 是 $\Omega(\log 1/\delta)$-路独立的，那么该构造就能奏效。遗憾的是，实际中无法使用 $\Omega(\log 1/\delta)$-路独立的哈希函数，因此本文旨在以更少的哈希函数假设，对稀疏 Johnson-Lindenstrauss 变换进行分析。具体而言，我们放宽对哈希函数 h 和符号函数 σ 的假设，只要求它们满足去耦分解和强集中性质。形式化定理见第3节。

我们还证明了混合制表满足这些性质，因此可以使用混合制表实现稀疏 Johnson-Lindenstrauss 变换。下面更正式地说明我们所说的“混合制表可以实现稀疏 Johnson-Lindenstrauss 变换”是什么意思。令 h_1: $\Sigma^c = [u] \to [m/s]$, h_2: $\Sigma^c \to \Sigma^d$，且 h_3: $\Sigma^d \to [m/s]$ 为实现混合制表哈希函数 $h_1(x) \oplus h_3(h_2(x))$ 的独立简单制表哈希函数。然后，我们将其扩展到定义域 $[s] \times [u]$，如下所示：

1. 将 h_2': $[s] \times \Sigma^c \to \Sigma^d$ 定义为 $h_2'(i, x) = h_2(x) \oplus (\underbrace{i, \ldots, i}_{d \text{ 次}})$，也就是说，每个派生字符都与 i 进行异或。

2. 接下来，我们通过 $h(i, x) = h_1(x) \oplus h_3(h_2'(i, x))$ 和 $\sigma(i, x) = \sigma_1(x) \cdot \sigma_3(h_2'(i, x))$ 定义 h: $[s] \times [u] \to [m/s]$ 和 σ: $[s] \times [u] \to \{-1, 1\}$，其中 h_1 和 h_3 是上述描述的简单制表哈希函数，而 σ_1: $\Sigma^c \to \{-1, 1\}$ 和 σ_3: $\Sigma^d \to \{-1, 1\}$ 是相互独立的简单制表函数。

1.2 哈希速度

当我们使用制表方案时，它通常是 Ω(log n)-独立哈希的一种快速替代方案。通常，我们在 $O(q)$ 时间内使用 $q-1$ 次多项式实现 q-独立哈希函数，而 Siegel [34] 已证明，除非使用较大的空间，否则这是可能达到的最佳结果。更准确地说，对于某个键域 $[u]$，如果我们希望进行 $t<q$ 次内存访问，那么所需空间至少为 $u^{1/t}$。因此，如果我们希望独立性高于常数级，同时仍保持常数级求值时间，那么确实需要 $u^{\Omega(1)}$ 的空间。在我们的应用中，我们必须同时计算许多哈希值，因此一种替代策略是使用多点求值来计算该多项式。这样可以将每个哈希值所需的时间降至 $O(\log^2 q)$，但这仍然是超常数时间。

使用制表哈希时，我们使用大小为 $O(|\Sigma|)$ 的表，其中 $|\Sigma|=u^{1/c}$ 和 $c=O(1)$。如果表能放入缓存，查表速度就很快；对于 8 位字符，这一点很容易实现。每次查表时，我们还会执行少量非常快速的 AC^0 操作：类型转换、按位异或和移位。就快速缓存访问与乘法运算的对比而言，这两者无法直接比较；不过，[1, 表 1] 中的实验发现，简单制表哈希比计算 2-独立多项式哈希更快。

比较制表方案时，最简单的方法是看查表次数。将 h_1 和 h_2 存储在同一张表中时，混合制表哈希需要查表 $c+d$ 次。对于 $d=c$，[1] 中的实验发现，混合制表哈希比简单制表哈希慢两倍多一些；而 [12] 中的实验发现，混合制表哈希的速度与 3-独立多项式哈希大致相当。这支持了我们的观点：混合制表哈希具有实用性。

理论上，我们也可以使用一种需要大量空间的高度独立哈希函数，但我们不知道有任何高效的构造方法。Siegel 谈到他的构造时称其“对于任何实际应用来说都慢得多” [34], ；而 Thorup [35] 提出了一种比 Siegel 的构造更简单的方法，但它仍然效率不高。[1] 中的实验发现，它比混合制表哈希慢了一个数量级以上。

2 相关工作

更稀疏的 Johnson-Lindenstrauss 变换。如前所述，Nelson 和 Nguyen [31] 给出了一个下界，排除了显著改进的可能性，但尽管如此，研究人员仍在探索更稀疏的嵌入。极端情况下，[39]的特征哈希考虑的是 $s=1$。这个下界表明，特征哈希无法适用于所有向量，但在 [17]中，作者根据度量 $\|w\|_\infty^2/\|w\|_2^2$，给出了特征哈希适用的向量范围的紧确界。后来，[21] 将这一结果推广为对 s 与度量 $\|w\|_\infty^2/\|w\|_2^2$ 之间权衡关系的完整理解。本文将只关注 $s=\Theta(\varepsilon^{-1}\log 1/\delta)$ 和$m=\Theta(\varepsilon^{-2}\log 1/\delta)$ 的情形

快速 Johnson-Lindenstrauss 变换。加速 Johnson-Lindenstrauss 变换求值的另一种方向，是利用能够快速进行矩阵-向量乘法的稠密矩阵。Ailon 和 Chazelle [4] 首次采用了这一方法，[4] 他们提出了快速 Johnson-Lindenstrauss 变换。他们最初的构造最近被 [16] 证明可实现如下嵌入时间$O(u\log u+m(\log 1/\delta+\varepsilon\log^2(1/\delta)/\log(1/\varepsilon)))$。

这促生了大量后续研究，旨在将运行时间改进到一个简洁的

$O(u\log u)$。部分研究牺牲了最优目标维数 $m = O(\varepsilon^{-2}\log 1/\delta)$，以加快构造速度，并接受次优的 $m = O(\varepsilon^{-2}\log^3 n)$ [15] $m = O(\varepsilon^{-1}\log^{3/2} n\log^{3/2} u+\varepsilon^{-2}\log^4 n\log^4 u)$ [29] $m = O(\varepsilon^{-2}\log^2 n)$ [19, 38, 18], $\log^4 u)$ [29]、(log)、(log log log log)、(log) 和 $m = O(\varepsilon^{-2}\log n\log^2(\log n)\log^3 u)$ [22]。另一项进展则假设目标维数 m 远小于起始维数 u。在假设$m = o(u^{1/2})$的情况下，文献 [5, 7] 中的工作实现了 $O(u\log m)$ 的嵌入时间。唯一在某些情形下优于原始快速 Johnson-Lindenstrauss 变换的构造，是近期对 Kac Johnson-Lindenstrauss 变换的分析 [22]；该变换使用 Kac 随机游走 [25]。他们证明，该方法可实现 $O(u\log u+\min\{u\log n, m\log n\log^2(\log n)\log^3 u\})$ 的嵌入时间。

关于制表哈希的既有研究。Patrascu 和 Thorup [33] 的工作开创了对基于表查的哈希的研究，其成果超出了构造所具备的 3-独立性所能预示的范围。大量论文表明，基于表查的哈希适用于最小值独立哈希 [32, 13], 、用于 k-统计量的哈希 [11], 以及非空箱数 [2]。此外，还有多篇论文致力于证明基于表查的哈希具有强集中性结果 [33, 32, 1, 20]。也有研究通过实验考察基于表查的哈希，结果显示其性能出色 [12, 1]。

2.1 预备知识

在本节中，我们将介绍本文中通篇使用的符号。首先，我们介绍 p-范数。

定义 2（p-范数）。设 $p \geq 1$ 和 X 为一个随机变量，且 $\mathrm{E}[|X|^p] < \infty$。我们随后将 X 的 p-范数定义为 $\|X\|_p = \mathrm{E}[|X|^p]^{1/p}$。

在本文中，我们将反复处理价值函数 $v\colon U \times [m] \to \mathbb{R}$。有时我们会将它们视为向量，特别是，我们将写作

$$\|v\|_2 = \sqrt{\sum_{x\in U}\sum_{j\in[m/s]} v(x,j)^2},$$
$$\|v\|_\infty = \max_{x\in U, j\in[m/s]} |v(x,j)|.$$

我们还将使用 [20] 中引入的 Ψ_p 函数。

定义 3. 对于 $p \geq 2$，我们将函数 $\Psi_p\colon \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ 定义如下，

$$\Psi_p(M,\sigma^2) = \begin{cases} \left(\frac{\sigma^2}{pM^2}\right)^{1/p} M & \text{if } p < \log\frac{pM^2}{\sigma^2} \\ \frac{1}{2}\sqrt{p}\sigma & \text{if } p < e^2\frac{\sigma^2}{M^2} \\ \frac{p}{e\log\frac{pM^2}{\sigma^2}} M & \text{if } \max\left\{\log\frac{pM^2}{\sigma^2}, e^2\frac{\sigma^2}{M^2}\right\} \leq p \end{cases}.$$

已有研究表明，[20] $\Psi_p(1,\lambda)$ 与参数为 λ 的泊松分布随机变量的 p-范数在常数因子范围内相当。他们还证明，$\Psi_p(M,\sigma^2)$ 可以用于

涉及完全随机哈希函数 h 的上界表达式： $U \to [m]$。设 v：$U \times [m] \to \mathbb{R}$ 为一个价值函数，那么他们证明了以下结论。

$$\left\| \sum_{x \in U} v(x, h(x)) \right\|_{\leq} C\Psi_p(\|v\|_\infty, \|v\|_2^2/m) ,$$

其中 C 是一个普适常数。

3 新分析概述

我们的主要技术贡献是对稀疏 Johnson-Lindenstrauss 变换进行了一种新的分析，放宽了对哈希函数 h的假设。我们证明，如果 h 满足解耦分解性质和强集中性质，那么我们可以为稀疏 Johnson-Lindenstrauss 变换获得相同的界。如果 h 是 Ω(log $1/\delta$)-wise 独立的，那么 h 满足这两种性质，因此我们的假设比以往分析中的假设更弱。

在本节中，我们将非正式地概述这一新方法。技术细节和结果的正式表述将在第 4 节给出。

为了描述我们的方法，我们考察随机变量

$$Z = \|Aw\|_2^2 - 1 = \frac{1}{s} \sum_{i \in [s]} \sum_{x \neq y \in [u]} \sigma(i,x)\sigma(i,y) \left[h(i,x) = h(i,y)\right] w_x w_y. \tag{2}$$

这里 $w \in \mathbb{R}^u$ 是一个单位向量。用这一记号，目标就变成界定 $\Pr[|Z| \geq \varepsilon]$。

我们分析的第一步是对式（2）进行解耦。在 [10], 中的一项证明里也使用了解耦，但由于我们希望针对更一般的哈希函数证明该结果，因此无法直接使用标准的解耦不等式。相反，我们将假设我们的哈希函数允许进行去耦分解。我们将在第 4 节中对此进行形式化定义；目前，我们先假设我们的哈希函数允许使用标准的解耦不等式。如果我们应用马尔可夫不等式以及针对完全随机哈希的标准解耦不等式，就可以得到如下表达式。

$$\begin{aligned} \Pr[|Z| \geq \varepsilon] &\leq \varepsilon^{-p} \operatorname{E}[|Z|^p] \\ &\leq \left(\varepsilon^{-1}\frac{4}{s}\right)^p \operatorname{E}\left[\left| \sum_{i \in [s]} \sum_{x,y \in [u]} \sigma(i,x)\sigma'(i,y) \left[h(i,x) = h'(i,y)\right] w_x w_y \right|^p\right] \end{aligned} \tag{3}$$

其中，(h', σ') 是 (h, σ) 和 $p \geq 2$ 的相互独立副本。解耦的作用在于打破了部分依赖关系，从而能够进行更简单的分析。

现在的目标是分析 $\left\| \sum_{i \in [s]} \sum_{x,\ y \in [u]} \sigma(i,x)\sigma'(i,y) \left[h(i,x) = h'(i,y)\right] w_x w_y \right\|_p$。具体而言，我们首先固定 (h', σ')，并利用 (h, σ) 的随机性来界定 $\left\| \sum_{i \in [s],\ j \in [m/s]} \sum_{x \in [u]} \sigma(i,x) \left[h(i,x) = j\right] w_x a_{ij} \right\|_p$，其中 $a_{ij} = \sum_{y \in [u]} \sigma'(i,y) \left[h'(i,y) = j\right] w_y$。为此，我们将假设该对 (h, σ) 具有强集中性。其形式化定义将在第 4 节中给出，但非正式地说，如果该对具有与完全随机哈希类似的集中性结果，我们就称其具有强集中性。

现在我们将 $|a_{ij}|$ 视为箱 $(i,j) \in [s] \times [m/s]$ 的负载。接下来，我们将 $[s] \times [m/s]$ 划分为重箱和轻箱，并分别进行处理。我们选取一个参数 k，并令 I 为最重的 k 个箱。利用三角不等式，我们得到

$$\left\| \sum_{i\in[s], j\in[m/s]} \sum_{x\in[u]} \sigma(i,x)\,[h(i,x)=j]\,w_x a_{ij} \right\|_p \le \left\| \sum_{(i,j)\in I} \sum_{x\in[u]} \sigma(i,x)\,[h(i,x)=j]\,w_x a_{ij} \right\|_p + \left\| \sum_{(i,j)\in[s]\times[m/s]\setminus I} \sum_{x\in[u]} \sigma(i,x)\,[h(i,x)=j]\,w_x a_{ij} \right\|_p .$$

我们证明，轻箱的贡献就如同碰撞彼此独立时一样。这一点应该不难理解，因为如果每个箱中的碰撞数量都很少，那么这些碰撞的行为就仿佛彼此独立。相比之下，我们证明重箱的贡献由最重的箱主导。事实证明，这正是完成分析所需要的。

4 技术结果

在本节　　中，我们将进一步阐述第3节中的描述，并将这些想法形式化。

解耦。理想情况下，我们希望使用标准的解耦不等式，即式（3）。遗憾的是，我们不能指望更一般的哈希函数也支持如此简洁的解耦。因此，我们引入去耦分解的概念。

定义 4（去耦分解）。设 $p \ge 2$、 $L \ge 1$ 和 $0 \le \gamma \le 1$。如果存在哈希函数 h_α: $[s]\times U_\alpha \to [m/s]$ 和符号函数 s: $[s]\times U_\alpha \to \{-1,1\}$ ，对所有 α 均成立，使得.....那么我们称一族可能经过随机化的集合（U_α）是关于一对（h、σ）的性质 P 的（p、L、γ）-去耦分解。

$$\begin{aligned} &\Pr[|Z| \ge \varepsilon] \\ &\quad \le \left(\varepsilon^{-1} \sum_\alpha \frac{L}{s} \left\| \sum_{i\in[s]} \sum_{x,y\in U_\alpha} \sigma_\alpha(i,x)\sigma'_\alpha(i,y)\left[h_\alpha(i,x)=h'_\alpha(i,y)\right] w_x w_y \right\|_p \right)^p + \gamma \end{aligned} \tag{4}$$

其中（h_α，σ_α）和（h'_α，σ'_α）服从相同的分布，而（h_α，σ_α）在给定（h'_α，σ'_α）和 U_α 的条件下满足性质 P。

读者应将完全随机哈希的式（3）与式（4）进行比较。这些表达式之间主要有 3 处区别。

1. 首先要注意，在去耦分解中，我们对不同的集合（U_α）求和，而完全随机哈希则无需这样做。我们允许去耦分解在每个集合 U_α 上采用不同的解耦。这一点非常强大，因为一般的哈希函数在输入域上不一定均匀分布。

2. 对于去耦分解，我们允许加性错误概率为 γ。如果哈希函数大多数时候都能实现解耦，只在发生某些罕见事件时例外，这会很有用。

3. 最后一处差异在于，与式 (3) 相比，去耦分解允许的损失因子大得多。在完全随机哈希的情况下，我们只损失 4 倍；但对于更一般的哈希函数，这一损失可能更大。

最后，我们注意到，式 (3) 意味着，如果 (h,σ) 对某个整数 $p \geq 2$ 是 $2p$-元独立的，那么 $[u]$ 就是 (h,σ) 的去耦分解，且任意满足 (h,σ) 的性质 P 都成立。

强集中性。我们需要的第二个性质是，该哈希函数具有强集中性。

定义 5（强集中）。设 h: $[s]\times U \to [m/s]$ 为哈希函数，设 σ: $[s]\times U \to \{-1,1\}$ 为符号函数。如果数对 (h,σ) 是 (p,L)-强集中，则称其为...

1. 对于所有值函数， v: $[s]\times[m/s]\to\mathbb{R}$，以及所有向量， $w\in\mathbb{R}^U$，

$$\left\|\sum_{i\in[s]}\sum_{x\in U}\sigma(i,x)v(i,h(i,x))w_x\right\|_p \leq \Psi_p\left(L\|v\|_\infty\|w\|_\infty, L\frac{s}{m}\|v\|_2^2\|w\|_2^2\right), \tag{5}$$

$$\left\|\sum_{i\in[s]}\sum_{x\in U}\sigma(i,x)v(i,h(i,x))w_x\right\|_p \leq \sqrt{L\frac{p}{\log(m/s)}\|v\|_2^2\|w\|_2^2}. \tag{6}$$

2. 对于所有向量， $w\in\mathbb{R}^U$，

$$\left\|\sum_{i\in[s]}\sum_{j\in[m/s]}\left(\sum_{x\in U}\sigma(i,x)\left[h(i,x)=j\right]w_x\right)^2\right\|_{p/2} \leq L\max\left\{s\|w\|_2^2, \frac{p}{\log m/s}\|w\|_2^2\right\}. \tag{7}$$

3. 如果 $p\leq\log m$，

$$\left\|\max_{i\in[s],j\in[m/s]}\left|\sum_{x\in U}\sigma(i,x)\left[h(i,x)=j\right]w_x\right|\right\|_p \leq e\sqrt{L\frac{\log m}{\log m/s}}\|w\|_2. \tag{8}$$

我们基本上需要哈希函数具备 3 种不同的性质，才能说它具有很强的集中性。

1. 第一种性质是关于随机变量的集中性结果

$$\sum_{i\in[s]}\sum_{x\in U}\sigma(i,x)v(i,h(i,x))w_x.$$

这里我们需要两个不同的集中性结果：第一个集中性结果，即式 (5)，大致对应于将 Bennett 不等式应用于完全随机的哈希函数后得到的 p-范数版本；而第二个集中性结果，即式 (5)，则对应于对参数为 s/m 的独立 Bernoulli-Rademacher 变量的加权和所能得到的最佳超收缩结果。[2]

[2]参数为 α 的伯努利–拉德马赫变量是随机变量 $X\in\{-1,0,1\}$，其中 $\Pr[X=1]=\Pr[X=-1]=\alpha/2$，且 $\Pr[X=0]=1-\alpha$。

2. 第二个性质给出了平方和的界

$$W = \sum_{i\in[s]} \sum_{j\in[m/s]} \left(\sum_{x\in U} \sigma(i,x)\,[h(i,x)=j]\,w_x \right)^2 .$$

条件（式 (7)）通过两种情况中的最大值对 $\|W\|_{p/2}$ 进行界定。第一种情况对应于 $\mathrm{E}[W]$，第二种情况则源于将式 (6) 应用于

$$\sup_{\substack{z\in\mathbb{R}^{[s]\times[m/s]},\\ \|z\|_2=1}} \left\| \sum_{i\in[s]} \sum_{j\in[m]} \sum_{x\in U} \sigma(i,x) z_{i,h(i,x)} w_x \right\|_p^2 .$$

虽然乍看之下这似乎有些奇怪，但这大致已经是所能做到的最佳结果，因为可以证明娱乐主管

$$\max\left\{ \mathrm{E}[W],\ \sup_{\substack{z\in\mathbb{R}^{[s]\times[m/s]},\\ \|z\|_2=1}} \left\| \sum_{i\in[s]} \sum_{j\in[m]} \sum_{x\in U} \sigma(i,x) z_{i,h(i,x)} w_x \right\|_p^2 \right\} \le \|W\|_{p/2} .$$

3. 最后一个性质是对最大坐标 $\max_{i\in[s],j\in[m/s]} \left|\sum_{x\in U} \sigma(i,x)\,[h(i,x)=j]\,w_x\right|$ 的一个界。对于完全随机哈希，这个界是公式 (6) 的一个自然推论。具体而言，对于完全随机哈希，我们得到的是

$$\begin{aligned}
&\left\| \max_{i\in[s],j\in[m/s]} \left| \sum_{x\in U} \sigma(i,x)\,[h(i,x)=j]\,w_x \right| \right\|_p \\
&\qquad \le \left\| \max_{i\in[s],j\in[m/s]} \left| \sum_{x\in U} \sigma(i,x)\,[h(i,x)=j]\,w_x \right| \right\|_{\log m} \\
&\qquad \le e \max_{i\in[s],j\in[m/s]} \left\| \left| \sum_{x\in U} \sigma(i,x)\,[h(i,x)=j]\,w_x \right| \right\|_{\log m} \\
&\qquad \le e\sqrt{L\frac{\log m}{\log m/s}}\,\|w\|_2 .
\end{aligned}$$

这种推导对于一般的哈希函数并不成立，但哈希函数仍然可以满足式（8）。

[20] 的结果表明，如果哈希函数 $h\colon [s]\times U \to [m/s]$ 和符号函数$\sigma\colon [s]\times U \to [m/s]$ 对于整数 $p \ge 2$ 是 p-路独立的，那么这一对 (h,σ) 是 (p,K)-强集中，其中 K 是一个普适常数。

主要结果。现在我们可以陈述主要结果：我们对稀疏 Johnson-Lindenstrauss 变换进行了新的分析，该分析仅假设哈希函数针对强集中性质具有一种解耦分解。

定理 6。设 $h\colon [s]\times[u]\to[m/s]$ 是一个哈希函数，$\sigma\colon [s]\times[u]\to\{-1,1\}$ 是一个符号函数。此外，设给定 $0<\varepsilon<1$ 和 $0<\delta<1$，并定义 $p=\log 1/\delta$。

假设存在常数 L_1、L_2、L_3 和 $0\le\gamma<1$，它们仅取决于 $(h$、$\sigma)$ 和 p，使得

1. 对于 $(h$、$\sigma)$ 的 $(p$、$L_2)$-强集中性质，存在一个 $(p$、L_1、$\gamma)$-去耦分解 (U_α)

2. 对所有向量 $w\in\mathbb{R}^u$、$\sum_\alpha\sum_{x\in U_\alpha}w_x^2\le L_3\|w\|_2^2$。

3. $m\ge\left(16e^7L_1^2L_2^3L_3^2\right)\cdot\varepsilon^{-2}\log(1/\delta)$。

4. $s\ge\left(64e^3L_1L_2^{3/2}L_3\right)\cdot\varepsilon^{-1}\log(1/\delta)$.

则以下结论成立

$$\Pr[|Z|\ge\varepsilon]\le\delta+\gamma.$$

如前所述，完全随机哈希函数满足该定理所需的所有性质，因此为完全随机哈希下的稀疏 Johnson-Lindenstrauss 变换提供了一种新的分析。我们稍后还将证明，混合制表满足该定理的假设，从而首次对一种可行且实用的哈希函数下的稀疏 Johnson-Lindenstrauss 变换进行了分析。

定理 6 的分析中，主要难点在于下面这个技术引理。定理 6 的证明思路是使用去耦分解，并将下面的引理应用于每个部分。

引理 7. 设 $h,\overline{h}\colon [s]\times U\to[m/s]$ 是哈希函数，$\sigma,\overline{\sigma}\colon [s]\times U\to\{-1,1\}$ 是符号函数。设 $p\ge2$，并假设存在常数 L，使得在以 $(\overline{h},\overline{\sigma})$ 为条件时，(h,σ) 是 (p,L)-强集中；类似地，在以 (h,σ) 为条件时，$(\overline{h},\overline{\sigma})$ 是 (p,L)-强集中。那么，对所有向量 $w\in\mathbb{R}^U$，

$$\left\|\sum_{i\in[s]}\sum_{x,y\in U}\sigma(i,x)\overline{\sigma}(i,y)\left[h(i,x)=\overline{h}(i,y)\right]w_xw_y\right\|_p \le\Psi_p\left(32e^3L^{3/2}\|w\|_2^2,32e^6L^3\frac{s^2}{m}\|w\|_2^4\right)+36e^3L\frac{p}{\log m/s}\|w\|_2^2.$$

这个引理表明，该表达式分为两个不同的区域。第一个区域 $\Psi_p\left(32e^3L^{3/2}\|w\|_2^2,32e^6L^3\frac{s^2}{m}\|w\|_2^4\right)$，基本上就是我们在每次碰撞 $\left[h(i,x)=\overline{h}(i,y)\right]$ 彼此独立时所预期的情形。另一个区域 $36e^3L\quad^{p}_{\text{对数}\,m/s}\|w\|_2^2$，基本上就是最大坐标所产生的贡献。

我们的分析受到这两种情形的启发，并尝试明确利用它们。我们首先固定 (h,σ)，然后将坐标分为重坐标和轻坐标。接着我们证明，轻坐标的贡献为 $\Psi_p\left(32e^3L^{3/2}\|w\|_2^2,32e^6L^3\frac{s^2}{m}\|w\|_2^4\right)$，这与以下直觉一致：如果每个坐标上只有少量碰撞，那么这些碰撞的行为就好像彼此独立。类似地，我们证明，重坐标的贡献由最重的坐标主导，即该贡献为 $36e^3L\quad^{p}_{\log m/s}\|w\|_2^2$。

4.1 混合制表哈希

我们关于混合制表哈希的主要结果如下。

定理 8. 设 $h\colon\ [s]\times[u]\to[m/s]$ 和 $\sigma\colon\ [s]\times[u]\to\{-1,1\}$ 是第 1.1 节所述的混合制表函数。此外，设给定 $0<\varepsilon<1$ 和 $0<\delta<1$，并定义 $p=\log 1/\delta$。

若 $m\geq\gamma_p^{3c}\varepsilon^{-2}\log(1/\delta)$ 且 $s\geq\gamma_p^{3/2c}\varepsilon^{-1}\log(1/\delta)$，其中 $\gamma_p=Kc\max\left\{1,\ {}^{p}\!/_{\log|\Sigma|}\right\}$，这里的普适常数为 K。

则以下结论成立

$$\Pr[|Z|\geq\varepsilon]\leq\delta+\varepsilon 3^c\left|\Sigma\right|^{-d}.$$

该结果可通过证明混合制表哈希具有 $(p,4^{c+2},4\varepsilon^{-2}3^c\frac{s}{m}\left|\Sigma\right|^{-d})$-去耦分解，且混合制表具有强集中性质得出。主要的新内容是证明这种去耦分解，而强集中性质的分析则是对 [20] 中分析的修改。

5 稀疏 Johnson-Lindenstrauss 变换的分析

我们先说明引理 7 如何推出主要结果——定理 6。

定理 6 的证明。我们先使用解耦分解的式 (4)，得到

$$\begin{aligned}&\Pr[|Z|\geq\varepsilon]\\&\quad\leq\left(\varepsilon^{-1}\sum_{\alpha}\frac{L_1}{s}\left\|\sum_{i\in[s]}\sum_{x,y\in U_\alpha}\sigma_\alpha(i,x)\sigma'_\alpha(i,y)\left[h_\alpha(i,x)=h'_\alpha(i,y)\right]v_xv_y\right\|_p\right)^p+\gamma\end{aligned}$$

现在固定 α，并在固定 U_α 的条件下应用引理 7

$$\begin{aligned}&\left\|\sum_{i\in[s]}\sum_{x,y\in U_\alpha}\sigma_\alpha(i,x)\sigma'_\alpha(i,y)\left[h_\alpha(i,x)=h'_\alpha(i,y)\right]v_xv_y\right\|_p\\&\qquad\leq\Psi_p\left(32e^3L_2^{3/2},32e^6L^3\frac{s^2}{m}\right)\sum_{x\in U_\alpha}w_x^2+36e^3L_2\frac{p}{\log m/s}\sum_{x\in U_\alpha}w_x^2\end{aligned}$$

由此可得

$$\begin{aligned}&\sum_{\alpha}\frac{L_1}{s}\left\|\sum_{i\in[s]}\sum_{x,y\in U_\alpha}\sigma_\alpha(i,x)\sigma'_\alpha(i,y)\left[h_\alpha(i,x)=h'_\alpha(i,y)\right]v_xv_y\right\|_p\\&\qquad\leq\sum_{\alpha}\frac{L_1}{s}\left(\Psi_p\left(32e^3L_2^{3/2},32e^6L^3\frac{s^2}{m}\right)\sum_{x\in U_\alpha}w_x^2+36e^3L_2\frac{p}{\log m/s}\sum_{x\in U_\alpha}w_x^2\right)\end{aligned}$$

现在我们利用 $\sum_\alpha \sum_{x\in U_\alpha} w_x^2 \le L_3 \|w\|_2^2$，得到

$$\sum_\alpha \frac{L_1}{s}\left(\Psi_p\left(32e^3L_2^{3/2}, 32e^6L^3\frac{s^2}{m}\right)\sum_{x\in U_\alpha} w_x^2 + 36e^3L_2\frac{p}{\log m/s}\sum_{x\in U_\alpha} w_x^2\right)$$
$$\le \frac{L_3L_1}{s}\left(\Psi_p\left(32e^3L_2^{3/2}, 32e^6L^3\frac{s^2}{m}\right) + 36e^3L_2\frac{p}{\log m/s}\right)\|w\|_2^2$$

现在可以验证，若 m 和 s 满足所述假设，则

$$\frac{L_3L_1}{s}\left(\Psi_p\left(32e^3L_2^{3/2}, 32e^6L^3\frac{s^2}{m}\right) + 36e^3L_2\frac{p}{\log m/s}\right)\|w\|_2^2 \le e^{-1}\varepsilon$$

综合所有事实，我们得到

$$\Pr[|Z| \ge \varepsilon] \le \left(\varepsilon^{-1}(e^{-1}\varepsilon)\right)^p + \gamma = \delta + \gamma.$$

证明至此完成。 □

本节余下部分将证明我们的主要技术引理——引理 7。首先，我们需要 [20]中的以下两个引理。

引理 9. 设 $f\colon \mathbb{R}_{\ge 0}^n \to \mathbb{R}_{\ge 0}$ 是一个非负函数，且在每个自变量上单调递增，并假设存在正实数 $(\alpha_i)_{i\in[n]}$ 和 $(t_i)_{i\in[n]}$，使得对所有$\lambda \ge 0$,

$$f(\lambda^{\alpha_0}t_0, \ldots, \lambda^{\alpha_{n-1}}t_{n-1}) \le \lambda f(t_0, \ldots, t_{n-1}).$$

设 $(X_i)_{i\in[n]}$ 是非负随机变量。那么，对所有 $p \ge 1$，有

$$\|f(X_0, \ldots, X_{n-1})\|_p \le n^{1/p}\max_{i\in[n]}\left(\frac{\|X_i\|_{p/\alpha_i}}{t_i}\right)^{1/\alpha_i} f(t_0, \ldots, t_{n-1}).$$

引理10。令 $p \ge 2$、 $M > 0$和 $\sigma^2 > 0$，那么

$$\frac{1}{2}\sqrt{p}\sigma \le \Psi_p(M, \sigma^2) \le \max\left\{\frac{1}{2}\sqrt{p}\sigma, \frac{1}{2e}pM\right\}.$$

现在我们已准备好证明引理 7。

引理 7 的证明。我们首先定义 $v_h, v_{\bar{h}}\colon [s]\times[m/s] \to \mathbb{R}$ 为,

$$v_h(i,j) = \sum_{x\in U}\sigma(i,x)w_x\left[h(i,x) = j\right],$$
$$v_{\bar{h}}(i,j) = \sum_{y\in U}\overline{\sigma}(i,y)w_y\left[\overline{h}(i,y) = j\right].$$

接下来我们要证明

$$\left\| \sum_{i\in[s],j\in[m/s]} v_h(i,j)v_{\bar{h}}(i,j) \right\|_p \leq \Psi_p\left(32e^3L^{3/2}\left\|w\right\|_2^2, 32e^6L^3\left\|w\right\|_2^4\right) + 4e^3L\frac{p}{\log m/s}\left\|w\right\|_2^2 .$$

首先我们考虑 $^{p}_{m/s}\left\|w\right\|_2^2 \geq s\left\|w\right\|_{2\text{对数}}^2$的情形。由柯西-施瓦茨不等式和式（7）可得,

$$\left\| \sum_{i\in[s],j\in[m/s]} v_h(i,j)v_{\bar{h}}(i,j) \right\|_p \leq \left\| \sum_{i\in[s],j\in[m/s]} v_h(i,j)^2 \right\|_p \leq L\frac{p}{\log m/s}\left\|w\right\|_2^2 .$$

现在我们关注 $^{p}_{m/s}\left\|w\right\|_2^2 < s\left\|w\right\|_{2\log}^2$的情形。我们定义 $\pi\colon [m] \to [s]\times[m/s]$ 为满足以下条件的双射。

$$|v_h(\pi(0))| \geq |v_h(\pi(1))| \geq \ldots \geq |v_h(\pi(m-1))| .$$

我们注意到 π 是一个随机函数，但我们可以定义 π ，使其仅依赖于 h和 σ的随机性。我们定义 $k = \lfloor p/\log(m/p)\rfloor$, $I = \{\pi(i) \mid i\in[k]\}$，以及随机函数$v'_h, v''_h\colon [s]\times[m/s]\to\mathbb{R}$为

$$v'_h(i,j) = v_h(i,j)\left[(i,j)\in I\right],$$
$$v''_h(i,j) = v_h(i,j)\left[(i,j)\notin I\right].$$

再次注意到， v'_h 和 v''_h 仅取决于 h 和 σ的随机性。于是，我们可以将表达式写为

$$\left\| \sum_{i\in[s],j\in[m/s]} v_h(i,j)v_{\bar{h}}(i,j) \right\|_p = \left\| \sum_{i\in[s]}\sum_{y\in U} \overline{\sigma}(i,y)v_h(i,\overline{h}(i,y))w_y \right\|_p$$
$$\leq \left\| \sum_{i\in[s]}\sum_{y\in U} \overline{\sigma}(i,y)v'_h(i,\overline{h}(i,y))w_y \right\|_p + \left\| \sum_{i\in[s]}\sum_{y\in U} \overline{\sigma}(i,y)v''_h(i,\overline{h}(i,y))w_y \right\|_p .$$

我们将分别对每一项进行界定。首先界定$\left\|\sum_{i\in[s]}\sum_{y\in U}\overline{\sigma}(i,y)v'_h(i,\overline{h}(i,y))w_y\right\|_p$。固定 h和 σ，并利用公式 (6) 得到

$$\left\| \sum_{i\in[s]}\sum_{y\in U} \overline{\sigma}(i,y)v'_h(i,\overline{h}(i,y))w_y \right\|_p \leq \left\| \sqrt{L\frac{p}{\log m/s}\left\|w\right\|_2^2\left\|v'_h\right\|_2^2} \right\|_p$$
$$= \sqrt{L\frac{p}{\log m/s}}\left\|w\right\|_2 \left\| \sqrt{\sum_{(i,j)\in I} v'_h(i,j)^2} \right\|_p .$$

注意到$\sum_{(i,j)\in I} v_h'(i,j)^2 = \text{最大值}_{J\subseteq[s]\times[m/s],|J|=k} \sum_{(i,j)\in J} v_h(i,j)^2$。于是得到

$$\begin{aligned}\left\|\sqrt{\sum_{(i,j)\in I} v_h'(i,j)^2}\right\|_p &= \left\|\sqrt{\max_{J\subseteq[s]\times[m/s],|J|=k}\sum_{(i,j)\in J} v_h(i,j)^2}\right\|_p \\ &\le \left(\sum_{J\subseteq[s]\times[m/s],|J|=k}\left\|\sqrt{\sum_{(i,j)\in J} v_h(i,j)^2}\right\|_p^p\right)^{1/p} \\ &\le \binom{ms}{k}^{1/p}\max_{J\subseteq[s]\times[m/s],|J|=k}\left\|\sqrt{\sum_{(i,j)\in J} v_h(i,j)^2}\right\|_p\end{aligned}$$

我们使用斯特林界，得到 $\binom{ms}{k}^{1/p} \le \left(\frac{ems}{k}\right)^{k/p} \le \left(e^{ms\log(\frac{ms/p}{p})}\right)^{1/(ms/p)\log(ms/p)} \le e^3$。因此得到

$$\left\|\sqrt{\sum_{(i,j)\in I} v_h'(i,j)^2}\right\|_p \le e^3 \max_{J\subseteq[s]\times[m/s],|J|=k}\left\|\sqrt{\sum_{(i,j)\in J} v_h(i,j)^2}\right\|_p$$

标准的体积论证表明，存在一个 $1/4$-网 $Z \subseteq \mathbb{R}^J$，满足 $|Z| \le 9^k$，使得

$$\begin{aligned}\left\|\sqrt{\sum_{(i,j)\in J} v_h(i,j)^2}\right\|_p &= \left\|\sup_{z\in\mathbb{R}^J,\|z\|_2=1}\sum_{(i,j)\in J} z_{i,j}v_h(i,j)\right\|_p \\ &\le \left\|\sup_{z\in Z}\sum_{(i,j)\in J} z_{i,j}v_h(i,j)\right\|_p \\ &\quad + \left\|\sup_{z\in\mathbb{R}^J,\|z\|_2=1}\sum_{(i,j)\in J}(z_{i,j}-z'_{i,j})v_h(i,j)\right\|_p \\ &\le \left\|\sup_{z\in Z}\sum_{(i,j)\in J} z_{i,j}v_h(i,j)\right\|_p \\ &\quad + \sup_{z\in\mathbb{R}^J,\|z\|_2=1}\|z-z'\|_2\left\|\sqrt{\sum_{(i,j)\in J} v_h(i,j)^2}\right\|_p\end{aligned}$$

其中，$z' \in Z$ 是最接近 z的元素，因此 $\|z-z'\|_2 \le 1/4$。由于 Z中至多有 9^k个元素，因此 $\left\|\sup_{z\in Z}\sum_{(i,j)\in J} z_{i,j}v_h(i,j)\right\|_p \le 9\sup_{z\in Z}\left\|\sum_{(i,j)\in J} z_{i,j}v_h(i,j)\right\|_p$，这里我们用到了 $k \le p$。综合以上事实可得

$$\left\|\sqrt{\sum_{(i,j)\in J} v_h(i,j)^2}\right\|_p \le 36\sup_{z\in Z}\left\|\sum_{(i,j)\in J} z_{i,j}v_h(i,j)\right\|_p$$

由此可得

$$
\begin{aligned}
e^3 \max_{J\subseteq[s]\times[m/s],|J|=k} & \left\| \sqrt{\sum_{(i,j)\in J} v_h(i,j)^2} \right\|_p \\
&\le 36e^3 \max_{J\subseteq[s]\times[m/s],|J|=k} \max_{z\in Z} \left\| \sum_{(i,j)\in J} z_{i,j} v_h(i,j) \right\|_p \\
&= 36e^3 \max_{J\subseteq[s]\times[m/s],|J|=k} \max_{\substack{z\in\mathbb{R}^{s\times m/s},\\ \|z\|_2=1}} \left\| \sum_{i\in[s]} \sum_{x\in U} \sigma(i,x) z_{i,h(i,x)} \left[(i,h(i,x))\in J\right] w_x \right\|_p
\end{aligned}
$$

然后，我们可以利用公式（6）得到：

$$
\begin{aligned}
36e^3 \max_{J\subseteq[s]\times[m/s],|J|=k} \max_{\substack{z\in\mathbb{R}^{s\times m/s},\\ \|z\|_2=1}} & \left\| \sum_{i\in[s]} \sum_{x\in U} \sigma(i,x) z_{i,h(i,x)} \left[(i,h(i,x))\in J\right] w_x \right\|_p \\
&\le 36e^3 \max_{J\subseteq[s]\times[m/s],|J|=k} \max_{\substack{z\in\mathbb{R}^{s\times m/s},\\ \|z\|_2=1}} \sqrt{L\frac{p}{\log m/s}} \|w\|_2 \|z\|_2 \\
&= 36e^3 \sqrt{L\frac{p}{\log m/s}} \|w\|_2
\end{aligned}
$$

结合这些事实，我们得到 $\left\|\sum_{i\in[s]}\sum_{y\in U}\overline{\sigma}(i,y)v_h'(i,\overline{h}(i,y))w_y\right\|_p \le 36e^3 L \quad {}^{p}\text{对数}\,\underline{m/s}\,\|w\|_2^2$。

现在我们将界定 $\left\|\sum_{i\in[s]}\sum_{y\in U}\overline{\sigma}(i,y)v_h''(i,\overline{h}(i,y))w_y\right\|_p$。我们固定 h 和 ε，并使用公式 (5) 得到

$$
\begin{aligned}
\left\| \sum_{i\in[s]} \sum_{y\in U} \overline{\sigma}(i,y) v_h''(i,\overline{h}(i,y)) w_y \right\|_p &\le \left\| \Psi_p\left(L\|w\|_\infty \left\|v_h'\right\|_\infty, L\frac{s}{m}\|w\|_2^2 \left\|v_h'\right\|_2^2 \right) \right\|_p \\
&\le \left\| \Psi_p\left(L\|w\|_\infty \left|v_h'(\pi(k+1))\right|, L\frac{s}{m}\|w\|_2^2 \|v_h\|_2^2 \right) \right\|_p .
\end{aligned}
$$

现在我们使用引理 9 得到,

$$
\begin{aligned}
&\left\| \Psi_p\left(L\|w\|_\infty \left|v_h'(\pi(k+1))\right|, L\frac{s}{m}\|w\|_2^2 \|v_h\|_2^2 \right) \right\|_p \\
&\qquad \le \sqrt{2}\Psi_p\left(L\|w\|_\infty \left\| \left|v_h'(\pi(k+1))\right| \right\|_p, L\frac{s}{m}\|w\|_2^2 \left\| \|v_h\|_2^2 \right\|_{p/2} \right).
\end{aligned}
$$

由于我们假设 ${}^{p}\log m\,\|w\|_2^2 < s\|w\|_2^2$，那么由式（7）可得 $\left\|\|v_h\|_2^2\right\|_{p/2} \le Ls\|w\|_2^2$。

现在我们将界定 $\|v_h'(\pi(k+1))\|_p$。为此，我们将区分两种情况：要么$p \ge \log m$，要么 $p < \log\ m$。我们先考虑 $p \ge \log\ m$ 的情况。我们将利用 $|v_h'(\pi(k+1))| \le$

$\sum_{i\in[k+1]}\left|v_h'(\quad\pi\ i\ \right|())$。由此可得

$$\begin{aligned}
&\left\|v_h'(\pi(k+1))\right\|_p \\
&\le \left\|\frac{\sum_{i\in[k+1]}|v_h'(\pi(i))|}{k+1}\right\|_p \\
&\le \left(\binom{m}{k+1}2^{k+1}\max_{\substack{J\subseteq[s]\times[m/s],\,(\sigma_{i,j})_{(i,j)\in J}\in\{-1,1\}^J\\ |J|=k+1}}\left(\frac{\left\|\sum_{(i,j)\in J}\sigma_{i,j}v_h(i,j)\right\|_p}{k+1}\right)^p\right)^{1/p} \\
&\le \max_{\substack{J\subseteq[s]\times[m/s],\,(s_{i,j})_{(i,j)\in J}\in\{-1,1\}^J\\ |J|=k+1}}2\binom{m}{k+1}^{1/p}\frac{\left\|\sum_{(i,j)\in J}\sigma_{i,j}v_h(i,j)\right\|_p}{k+1}
\end{aligned}$$

注意，$\left\|\sum_{(i,j)\in J}\sigma_{i,j}v_h(i,j)\right\|_p=\left\|\sum_{x\in U}\sum_{(i,j)\in J}\sigma(i,x)s_{i,j}\left[h(i,x)=j\right]w_x\right\|_p$。由于我们有 $p\ge\log m$，因此 $k\ge 1$，这意味着 $k+1\le 2\quad^p\log(m/p)$。由此我们得到 $\binom{m}{k+1}^{1/p}\le\left(2p/\right.^{em}\left.\log(m/p)\right)^{2/}\log(m/p)\le 2e^3$。现在我们使用等式 (6) 得到，

$$\begin{aligned}
\left\|\sum_{x\in U}\sum_{(i,j)\in J}\sigma(i,x)s_{i,j}\left[h(i,x)=j\right]w_x\right\|_p &= \left\|\sum_{i\in[s]}\sum_{x\in U}\sigma(i,x)\left[(i,h(i,x))\in J\right]s_{i,h(i,x)}w_x\right\|_p \\
&\le \sqrt{L\frac{p}{\log m/s}}\sqrt{|J|}\left\|w\right\|_2 \\
&= \sqrt{L\frac{p}{\log m/s}}\sqrt{k+1}\left\|w\right\|_2
\end{aligned}$$

综合以上可得，$\left\|v_h'(\pi(k+1))\right\|_p\le 4e^3\sqrt{L\quad^p\text{对数 }m/s}\frac{\|w\|_2}{\sqrt{k+1}}$。于是我们得到，

$$\begin{aligned}
&\left\|\sum_{i\in[s]}\sum_{y\in U}\overline{\sigma}(i,y)v_h''(i,\overline{h}(i,y))w_y\right\|_p \\
&\qquad\le \sqrt{2}\Psi_p\left(4e^3L\sqrt{L\frac{p}{(k+1)\log m/s}}\left\|w\right\|_\infty\left\|w\right\|_2, L^2\frac{s^2}{m}\left\|w\right\|_2^4\right) \\
&\qquad\le \sqrt{2}\Psi_p\left(4e^3L\sqrt{L\frac{\log m/p}{\log m/s}}\left\|w\right\|_2^2, L^2\frac{s^2}{m}\left\|w\right\|_2^4\right)
\end{aligned}$$

如果 $\log\ m/p\le 4\log\ m/s$，那么我们得到：

$$\begin{aligned}
\left\|\sum_{i\in[s]}\sum_{y\in U}\overline{\sigma}(i,y)v_h''(i,\overline{h}(i,y))w_y\right\|_p &\le \sqrt{2}\Psi_p\left(16e^3L^{3/2}\left\|w\right\|_2^2, L^2\frac{s^2}{m}\left\|w\right\|_2^4\right) \\
&\le \Psi_p\left(32e^3L^{3/2}\left\|w\right\|_2^2, 2L^2\frac{s^2}{m}\left\|w\right\|_2^4\right)
\end{aligned}$$

如果 $\log\ m/p > 4\log\ m/s$，那么 $m/p > (m/s)^4$，这意味着 $\frac{pm}{s^2} \le \sqrt{p}m$。利用这一点，我们得到 $\frac{p16e^6L^3\log\ \frac{m/p}{m/s}\|w\|_2^4}{L^2\frac{s^2}{m}\|w\|_2^4} \le 16e^6L\sqrt{p}m^{\log}\log\ m/p \le 16e^6L$。这里我们用到了 $\sqrt{s}\log\ 1/x \le 1$。现在我们使用引理10，得到

$$\begin{aligned}\left\|\sum_{i\in[s]}\sum_{y\in U}\overline{\sigma}(i,y)v_h''(i,\overline{h}(i,y))w_y\right\|_p &\le \sqrt{2}\Psi_p\left(32e^3L^{3/2}\left\|w\right\|_2^2, L^2\frac{s^2}{m}\left\|w\right\|_2^4\right)\\ &\le \sqrt{2}\sqrt{pL^3 16e^6\frac{s^2}{m}\left\|w\right\|_2^4}\\ &\le \Psi_p\left(8e^3L^{3/2}\left\|w\right\|_2, 32e^6L^3\frac{s^2}{m}\left\|w\right\|_2^4\right)\end{aligned}$$

现在考虑 $p < \log m$ 的情况。由式（8），可得

$$\left\|v_h'(\pi(k+1))\right\|_p \le \left\|\max_{i\in[s],j\in[m/s]}\left|\sum_{x\in U}\sigma(i,x)\left[h(i,x)=j\right]w_x\right|\right\|_p \le e\sqrt{L\frac{\log m}{\log m/s}}\left\|w\right\|_2$$

于是我们得到,

$$\begin{aligned}\left\|\sum_{i\in[s]}\sum_{y\in U}\overline{\sigma}(i,y)v_h''(i,\overline{h}(i,y))w_y\right\|_p &\le \sqrt{2}\Psi_p\left(eL\sqrt{L\frac{\log m}{\log m/s}}\left\|w\right\|_\infty\left\|w\right\|_2, L\frac{s^2}{m}\left\|w\right\|_2^4\right)\\ &\le \sqrt{2}\Psi_p\left(eL\sqrt{L\frac{\log m}{\log m/s}}\left\|w\right\|_2^2, L\frac{s^2}{m}\left\|w\right\|_2^4\right)\end{aligned}$$

如果 $s \le m^{3/4}$，那么我们得到 $\log\ m \le 4\log\ m/s$，并且

$$\sqrt{2}\Psi_p\left(eL\sqrt{L\frac{\log m}{\log m/s}}\left\|w\right\|_2^2, L\frac{s^2}{m}\left\|w\right\|_2^4\right) \le \sqrt{2}\Psi_p\left(4eL^{3/2}\left\|w\right\|_2^2, L\frac{s^2}{m}\left\|w\right\|_2^4\right)$$

如所需。如果 $s \ge m^{3/4}$，那么我们得到 $\frac{peL^3\log\ m}{L^2s^2/m} \le \frac{eLm\log\frac{s^2}{m}}{s^2} \le eL\log_{m^{1/2}}{}^2 m \le 16/eL$，其中用到了 $\sqrt{x}\log^2 1/x \le 16/e^2$。再次应用引理10，可得

$$\begin{aligned}\sqrt{2}\Psi_p\left(4eL^{3/2}\left\|w\right\|_2^2, L\frac{s^2}{m}\left\|w\right\|_2^4\right) &\le \sqrt{p32/eL^3\frac{s^2}{m}}\left\|w\right\|_2^2\\ &\le \Psi_p\left(8L^{3/2}\left\|w\right\|_2^2, 32L^3\frac{s^2}{m}\left\|w\right\|_2^4\right).\end{aligned}$$

综合以上结果，我们得到

$$\begin{aligned}\left\|\sum_{i\in[s]}\sum_{y\in U}\overline{\sigma}(i,y)v_h''(i,\overline{h}(i,y))w_y\right\|_p &\le \Psi_p\left(32e^3L^{3/2}\left\|w\right\|_2, 32e^6L^3\frac{s^2}{m}\left\|w\right\|_2^4\right),\\ \left\|\sum_{i\in[s]}\sum_{y\in U}\overline{\sigma}(i,y)v_h'(i,\overline{h}(i,y))w_y\right\|_p &\le 4e^3L\frac{p}{\log m/s}\left\|w\right\|_2^2.\end{aligned}$$

现在我们得出结论

$$\begin{aligned}
&\left\| \sum_{i\in[s],j\in[m/s]} v_h(i,j)v_{\bar{h}}(i,j) \right\|_p \\
&\qquad \le \left\| \sum_{i\in[s]} \sum_{y\in U} \overline{\sigma}(i,y) v'_h(i,\overline{h}(i,y)) w_y \right\|_p + \left\| \sum_{i\in[s]} \sum_{y\in U} \overline{\sigma}(i,y) v''_h(i,\overline{h}(i,y)) w_y \right\|_p \\
&\qquad \le \Psi_p\left(32e^3L^{3/2}\left\|w\right\|_2^2, 32e^6L^3\left\|w\right\|_2^4\right) + 4e^3L\frac{p}{\log m/s}\left\|w\right\|_2^2 .
\end{aligned}$$

由此完成证明。 □

6 混合制表哈希分析

本节的目标是证明关于混合制表哈希的主要结果。主要的新结果在第 6.2 节，我们证明了混合制表具有去耦分解。在第 6.3 节，我们证明混合制表也具有强集中性质。第 6.3 节中的证明是对 [20] 中混合制表哈希相关证明的修改。

6.1 制表哈希的符号约定和先前结果

我们需要讨论键 $x \in \Sigma^c$ 的各个字符，因此需要引入一些符号约定。

定义 11（位置字符）。设 Σ 是一个字母表，$c > 0$ 是一个正整数。我们称元素 (i, y) 为 Σ^c 的一个位置字符。

我们将键 $x = (y_0, \ldots, y_{c-1}) \in \Sigma^c$ 视为由 c 个位置字符组成的集合，$\{(0, y_0) \ldots (c-1, y_{c-1})\} \subseteq [c] \times \Sigma$。我们定义集合 $P_{partial}$ 和 P_{prefix}，它们包含部分键。

$$P_{partial} = \left\{ \{(i,\alpha_i)\}_{i\in I} \,\middle|\, \emptyset \neq I \subseteq [c], \forall i \in I\colon\ \alpha_i \in \Sigma \right\} P_{prefix} = \left\{ \{(i,\alpha_i)\}_{i\in I} \,\middle|\, \emptyset \neq I \subseteq [c], \right.$$
$$\left. I \text{是一个包含 0 的区间，且 } \forall i \in I\colon\ \alpha_i \in \Sigma \right\}$$

对于部分键 $\pi = \{(i_0,\alpha_0)\ldots,(i_{k-1},\alpha_{k-1})\} \in P_{partial}$，我们将 $I_\pi = \{i_0, \ldots, i_{k-1}\}$ 定义为已用位置的集合。我们也将写作 $|\pi| = |I_\pi|$。

对于简单制表哈希函数 $h\colon \Sigma^c \to R$，我们将 h_I for $I \subseteq [c]$ 定义为只考察 I中各位置的哈希函数，即 $h_I(x) = \bigoplus_{i\in I} T_i[x_i]$。类似地，对于 $h_2\colon \Sigma^c \to \Sigma^d$，我们将 h^I for $I \subseteq [d]$ 定义为限制在位置 I上的部分键。

对于 $\pi \in P_{partial}$ 和 $\tau \in P_{prefix}$，我们将随机集合 $U_{\pi,\ \tau}$ 定义如下

$$U_{\pi,\tau} = \{x \in \Sigma^c \,|\, \pi \subseteq x, \tau \subseteq h_2(x)\}$$

我们还需要以下来自 [20] 的引理。

引理 12. 设 h: $U \to [m]$ 为一个均匀随机函数，设 v: $U \times [m] \to \mathbb{R}$ 为一个固定值函数，并假设对于所有键 $x \in U$， $\sum_{j\in[m]} v(x,j) = 0$。定义随机变量 $X_v = \sum_{x\in U} v(x,h(x))$。则对于所有 $p \geq 2$,

$$\|X_v\|_p \leq L\Psi_p\left(M_v, \sigma_v^2\right) ,$$

其中 $L \leq 16e$ 是一个普适常数。

引理 13. 设 h: $U \to [m]$ 是一个均匀随机函数， ε: $U \to \{-1,1\}$ 是一个均匀随机符号函数，且 v: $U \times [m] \to \mathbb{R}$ 是一个固定值函数。则对于所有 $p \geq 2$,

$$\left\|\sum_{x\in U} \varepsilon(s)v(x,h(x))\right\|_p \leq L\sqrt{\frac{p}{\log(m)}}\|v\|_2 ,$$

其中 $L \leq e$ 是一个普适常数。

引理 14. 设 h: $\Sigma^c \to [m]$ 是一个简单制表哈希函数， v: $\Sigma^c \times [m] \to \mathbb{R}$ 是一个价值函数，并假设对于所有键 $x \in U$， $\sum_{j\in[m]} v(x,j) = 0$。则对于所有 $p \geq 2$,

$$\left\|\sum_{x\in\Sigma^c} v(x,h(x))\right\|_p \leq L_1\Psi_p\left(K_c\gamma_p^{c-1}\|v\|_\infty, K_c\gamma_p^{c-1}\frac{\|v\|_2^2}{m}\right) ,$$

其中 $K_c = (L_2 c)^{c-1}$、 L_1 和 L_2 是普适常数，并且

$$\gamma_p = \frac{\max\left\{\log(m) + \log\left(\frac{\sum_{x\in\Sigma^c}\|v[x]\|_2^2}{\max_{x\in\Sigma^c}\|v[x]\|_2^2}\right)/c, p\right\}}{\log\left(e^2 m\left(\max_{x\in\Sigma^c}\frac{\|v[x]\|_1^2}{\|v[x]\|_2^2}\right)^{-1}\right)}$$

引理 15. 设 h: $\Sigma^c \to [m]$ 是一个简单制表函数， ε: $\Sigma^c \to \{-1,1\}$ 是一个简单制表符号函数，且 v_i: $\Sigma^c \times [m] \to \mathbb{R}$ 是针对 $i \in [k]$ 的价值函数。对于每个实数 $p \geq 2$,

$$\left\|\sum_{j\in[m]}\left(\sum_{x\in\Sigma^c}\varepsilon(x)v(x,h(x)\oplus j)\right)^2\right\|_p \leq \left(\frac{Lc\max\{p,\log(m)\}}{\log\left(\frac{e^2 m\sum_{x\in\Sigma^c}\sum_{j\in[m]}v(x,j)^2}{\sum_{x\in\Sigma^c}(\sum_{j\in[m]}|v(x,j)|)^2}\right)}\right)^c \|v\|_2^2 ,$$

其中 L 是一个普适常数。

6.2 去耦

在证明混合制表哈希的去耦分解之前，我们先给出一个一般的去耦引理。我们从陈述 [37] 中的一个标准去耦结果开始。

引理 16. 设 f_{ij}: $T \times T \to \mathbb{R}$ 是定义在 $i,j \in [n]$ 上的函数。设 $(X_i)_{i\in[n]}$ 是取值于 T 的独立随机变量，且对所有 $i \neq j \in [n]$，均有 $\mathrm{E}[f_{ij}(X_i,X_j) \mid X_j] = \mathrm{E}[f_{ij}(X_i,X_j) \mid X_i] = 0$。那么，对每个凸函数 F: $\mathbb{R} \to \mathbb{R}$，都有

$$\mathrm{E}\left[F\left(\sum_{i\neq j\in[n]} f_{ij}(X_i,X_j)\right)\right] \leq \mathrm{E}\left[F\left(4\sum_{i,j\in[n]} f_{ij}(X_i,X_j')\right)\right] ,$$

其中 $(X_i')_{i\in[n]}$ 是 $(X_i)_{i\in[n]}$ 的一个独立副本。

我们将解耦结果略作推广如下。

引理 17. 设 $f_{i_0,\dots,i_{c-1},j_0,\dots,j_{c-1}}\colon T^c\times T^c\to\mathbb{R}$ 是关于 $i_0\dots,i_{c-1},j_0,\dots,j_{c-1}\in[n]$的函数。设 $(X_i^{(k)})_{i\in[n],k\in[c]}$ 是取值于 T 的独立随机变量，满足

$$\begin{aligned}\mathrm{E}\Big[f_{i_0,\dots,i_{c-1},j_0,\dots,j_{c-1}}((X_{i_0}^{(0)},\dots,X_{i_{c-1}}^{(c-1)}),(X_{j_0}^{(0)},\dots,X_{j_{c-1}}^{(c-1)}))\,\Big|\,(X_{i_k}^{(k)})_{k\in[c]\setminus\{l\}},(X_{j_k}^{(k)})_{k\in[c]}\Big]=0\,,\\ \mathrm{E}\Big[f_{i_0,\dots,i_{c-1},j_0,\dots,j_{c-1}}((X_{i_0}^{(0)},\dots,X_{i_{c-1}}^{(c-1)}),(X_{j_0}^{(0)},\dots,X_{j_{c-1}}^{(c-1)}))\,\Big|\,(X_{i_k}^{(k)})_{k\in[c]},(X_{j_k}^{(k)})_{k\in[c]\setminus\{l\}}\Big]=0\,,\end{aligned}\tag{9}$$

对所有 $i_0\dots,i_{c-1},j_0,\dots,j_{c-1}\in[n]$ 以及所有满足 $i_k\neq j_k$的 $l\in[c]$，对所有 $k\in[c]$均成立。于是，对于每个凸函数 $F\colon\mathbb{R}\to\mathbb{R}$，有

$$\begin{aligned}&\mathrm{E}\left[F\left(\sum_{\substack{i_0\dots,i_{c-1},j_0,\dots,j_{c-1}\in[n]\\ \forall k\in[c]:i_k\neq j_k}} f_{i_0,\dots,i_{c-1},j_0,\dots,j_{c-1}}((X_{i_0}^{(0)},\dots,X_{i_{c-1}}^{(c-1)}),(X_{j_0}^{(0)},\dots,X_{j_{c-1}}^{(c-1)}))\right)\right]\\ &\le\mathrm{E}\left[F\left(4^c\sum_{i_0\dots,i_{c-1},j_0,\dots,j_{c-1}\in[n]} f_{i_0,\dots,i_{c-1},j_0,\dots,j_{c-1}}((X_{i_0}^{(0)},\dots,X_{i_{c-1}}^{(c-1)}),(\hat{X}_{j_0}^{(0)},\dots,\hat{X}_{j_{c-1}}^{(c-1)}))\right)\right]\,,\end{aligned}\tag{10}$$

其中，$(\hat{X}_i^{(k)})_{i\in[n],k\in[c]}$ 是 $(X_i^{(k)})_{i\in[n],k\in[c]}$的独立副本。

证明。通过对 c使用归纳法，归纳起始情形 $c=1$恰好就是引理 16。现在假设 $c>1$，且式 (10) 对 $c-1$成立。

We s首先定义函数，$g_{i_{c-1},j_{c-1}}\colon T\times T\to\mathbb{R}$ 对于所有 $i_{c-1},j_{c-1}\in[n]$，通过

$$g_{i_{c-1},j_{c-1}}(x,y)=\sum_{\substack{i_0\dots,i_{c-2},j_0,\dots,j_{c-2}\in[n]\\ \forall k\in[c-1]:i_k\neq j_k}} f_{i_0,\dots,i_{c-1},j_0,\dots,j_{c-1}}((X_{i_0}^{(0)},\dots,X_{i_{c-2}}^{(c-2)}x),(X_{j_0}^{(0)},\dots,X_{j_{c-2}}^{(c-2)},y)).$$

注意，式 (9) 意味着

$$\begin{aligned}\mathrm{E}\Big[g_{i_{c-1},j_{c-1}}(X_{i_{c-1}}^{(c-1)},X_{j_{c-1}}^{(c-1)})\,\Big|\,X_{i_{c-1}}^{(c-1)},(X_i^{(k)})_{i\in[n],k\in[c-1]}\Big]=0\,,\\ \mathrm{E}\Big[g_{i_{c-1},j_{c-1}}(X_{i_{c-1}}^{(c-1)},X_{j_{c-1}}^{(c-1)})\,\Big|\,X_{j_{c-1}}^{(c-1)},(X_i^{(k)})_{i\in[n],k\in[c-1]}\Big]=0\,,\end{aligned}$$

对于所有 $i_{c-1}\neq j_{c-1}\in[n]$。这意味着我们可以在以.....为条件的情况下使用引理 16

$(X_i^{(k)})_{i\in[n],\ k\in[c-1]}$ 以获得该结果

$$
\begin{aligned}
&\mathrm{E}\left[F\left(\sum_{\substack{i_0\ldots,i_{c-1},j_0,\ldots,j_{c-1}\in[n]\\ \forall k\in[c]:i_k\neq j_k}} f_{i_0,\ldots,i_{c-1},j_0,\ldots,j_{c-1}}((X_{i_0}^{(0)},\ldots,X_{i_{c-1}}^{(c-1)}),(X_{j_0}^{(0)},\ldots,X_{j_{c-1}}^{(c-1)}))\right)\right]\\
&=\mathrm{E}\left[\mathrm{E}\left[F\left(\sum_{i_{c-1}\neq j_{c-1}\in[n]} g_{i_{c-1},j_{c-1}}(X_{i_{c-1}}^{(c-1)},X_{c-1}^{(c-1)})\right)\,\middle|\,(X_i^{(k)})_{i\in[n],k\in[c-1]}\right]\right]\\
&\le\mathrm{E}\left[\mathrm{E}\left[F\left(4\sum_{i_{c-1},j_{c-1}\in[n]} g_{i_{c-1},j_{c-1}}(X_{i_{c-1}}^{(c-1)},\hat{X}_{c-1}^{(c-1)})\right)\,\middle|\,(X_i^{(k)})_{i\in[n],k\in[c-1]}\right]\right]\\
&=\mathrm{E}\left[F\left(4\sum_{\substack{i_0\ldots,i_{c-1},j_0,\ldots,j_{c-1}\in[n]\\ \forall k\in[c-1]:i_k\neq j_k}} f_{i_0,\ldots,i_{c-1},j_0,\ldots,j_{c-1}}((X_{i_0}^{(0)},\ldots,X_{i_{c-1}}^{(c-1)}),(X_{j_0}^{(0)},\ldots,X_{j_{c-2}}^{(c-2)},\hat{X}_{j_{c-1}}^{(c-1)}))\right)\right].
\end{aligned}
$$

接下来我们定义函数 $h_{i_0,\ldots,i_{c-2},j_0,\ldots,j_{c-2}}\colon\ T^{c-1}\times T^{c-1}\to\mathbb{R}$ 对所有$i_0\ldots,i_{c-2},j_0,\ldots,j_{c-2}\in[n]$，通过

$$
\begin{aligned}
&h_{i_0,\ldots,i_{c-2},j_0,\ldots,j_{c-2}}((x_0,\ldots,x_{c-2}),(y_0,\ldots,x_{c-2}))\\
&\qquad=\sum_{i_{c-1},j_{c-1}\in[n]} f_{i_0,\ldots,i_{c-1},j_0,\ldots,j_{c-1}}((x_0,\ldots,x_{c-2},X_{i_{c-1}}^{(c-1)}),(y_0,\ldots,y_{c-2},\hat{X}_{j_{c-1}}^{(c-1)}))\,.
\end{aligned}
$$

现在令 $i_0\ldots,i_{c-2},j_0,\ldots,j_{c-2}\in[n]$ 和 $l\in[c-1]$ 满足 $i_k\neq j_k$ 对所有 $k\in[c-1]$ 均成立，并令$H=h_{i_0,\ldots,i_{c-2},j_0,\ldots,j_{c-2}}$。那么，再次利用等式 (9)，我们得到，对于

$$
\begin{aligned}
&\mathrm{E}\Big[H((X_{i_0}^{(0)},\ldots,X_{i_{c-2}}^{(c-2)}),(X_{j_0}^{(0)},\ldots,X_{j_{c-2}}^{(c-2)}))\,\Big|\,(X_{i_k}^{(k)})_{k\in[c-1]\setminus\{l\}},(X_{j_k}^{(k)})_{k\in[c-1]},(X_i^{(c-1)},\hat{X}_i^{(c-1)})_{i\in[n]}\Big]=0,\\
&\mathrm{E}\Big[H((X_{i_0}^{(0)},\ldots,X_{i_{c-2}}^{(c-2)}),(X_{j_0}^{(0)},\ldots,X_{j_{c-2}}^{(c-2)}))\,\Big|\,(X_{i_k}^{(k)})_{k\in[c-1]},(X_{j_k}^{(k)})_{k\in[c-1]\setminus\{l\}},(X_i^{(c-1)},\hat{X}_i^{(c-1)})_{i\in[n]}\Big]=0.
\end{aligned}
$$

然后，我们可以在给定 $(X_i^{(c-1)}, \hat{X}_i^{(c-1)})_{i\in[n]}$ 的条件下应用归纳假设，从而得到[3]

$$
\begin{aligned}
&\mathrm{E}\left[F\left(4\sum_{\substack{i_0\ldots,i_{c-1},j_0,\ldots,j_{c-1}\in[n]\\ \forall k\in[c-1]:i_k\neq j_k}} f_{i_0,\ldots,i_{c-1},j_0,\ldots,j_{c-1}}((X_{i_0}^{(0)},\ldots,X_{i_{c-1}}^{(c-1)}),(X_{j_0}^{(0)},\ldots,X_{j_{c-2}}^{(c-2)},\hat{X}_{j_{c-1}}^{(c-1)}))\right)\right]\\
&=\mathrm{E}\left[\mathrm{E}\left[F\left(4\sum_{\substack{i_0\ldots,i_{c-2},j_0,\ldots,j_{c-2}\in[n]\\ \forall k\in[c-1]:i_k\neq j_k}} h_{\mathbf{i},\mathbf{j}}((X_{i_0}^{(0)},\ldots,X_{i_{c-2}}^{(c-2)}),(X_{j_0}^{(0)},\ldots,X_{j_{c-2}}^{(c-2)}))\right)\middle|(X_i^{(c-1)},\hat{X}_i^{(c-1)})_{i\in[n]}\right]\right]\\
&\le\mathrm{E}\left[\mathrm{E}\left[F\left(4^c\sum_{i_0\ldots,i_{c-2},j_0,\ldots,j_{c-2}\in[n]} h_{\mathbf{i},\mathbf{j}}((X_{i_0}^{(0)},\ldots,X_{i_{c-2}}^{(c-2)}),(\hat{X}_{j_0}^{(0)},\ldots,\hat{X}_{j_{c-2}}^{(c-2)}))\right)\middle|(X_i^{(c-1)},\hat{X}_i^{(c-1)})_{i\in[n]}\right]\right]\\
&=\mathrm{E}\left[F\left(4^c\sum_{i_0\ldots,i_{c-1},j_0,\ldots,j_{c-1}\in[n]} f_{i_0,\ldots,i_{c-1},j_0,\ldots,j_{c-1}}((X_{i_0}^{(0)},\ldots,X_{i_{c-1}}^{(c-1)}),(\hat{X}_{j_0}^{(0)},\ldots,\hat{X}_{j_{c-1}}^{(c-1)}))\right)\right].
\end{aligned}
$$

至此，归纳步骤完成，证明也随之结束。 □

现在我们可以证明混合制表的一个去耦引理。

引理 18。

$$
\left\|\sum_{i\in[s]}\sum_{x,y\in[u]}\left[h_2(x)\neq h_2(y)\right]\sigma(i,x)\sigma(i,y)\left[h(i,x)=h(i,y)\right]w_xw_y\right\|_p \tag{11}
$$

$$
\le\sum_{\pi\in P_{partial}}\sum_{\tau\in P_{prefix}}4^{c+1}\left\|\sum_{i\in[s]}\sum_{x,y\in U_{\pi,\tau}}\sigma_{I_\pi,|\tau|}(i,x)\overline{\sigma_{I_\pi,|\tau|}}(i,y)\left[h_{I_\pi,|\tau|}(i,x)=\overline{h_{I_\pi,|\tau|}}(i,y)\right]w_xw_y\right\|_p \tag{12}
$$

其中

$$
\begin{aligned}
h_{I,r}(i,x)&=\left(h_{1,I}(x)\oplus h_{3,\{r\}}(h_{2,I}^{\{r\}}(x)\oplus i\oplus h_{2,I^c}^{\{r\}}(x))\right)\oplus\left(h_{1,I^c}(x)\oplus h_{3,\{r\}^c}(h_2^{\{r\}^c}(x)\oplus i^{\otimes(d-1)})\right)\\
\sigma_{I,r}(i,x)&=\left(\sigma_{1,I}(x)\oplus\sigma_{3,\{r\}}(h_{2,I}^{\{r\}}(x)\oplus i\oplus h_{2,I^c}^{\{r\}}(x))\right)\oplus\left(\sigma_{1,I^c}(x)\oplus\sigma_{3,\{r\}^c}(h_2^{\{r\}^c}(x)\oplus i^{\otimes(d-1)})\right)\\
\overline{h_{I,r}}(i,y)&=\left(\overline{h_{1,I}}(y)\oplus\overline{h_{3,\{r\}}}(\overline{h_{2,I}^{\{r\}}}(y)\oplus i\oplus h_{2,I^c}^{\{r\}}(y))\right)\oplus\left(h_{1,I^c}(y)\oplus h_{3,\{r\}^c}(h_2^{\{r\}^c}(y)\oplus i^{\otimes(d-1)})\right)\\
\overline{\sigma_{I,r}}(i,y)&=\left(\overline{\sigma_{1,I}}(y)\oplus\overline{\sigma_{3,\{r\}}}(\overline{h_{2,I}^{\{r\}}}(y)\oplus i\oplus h_{2,I^c}^{\{r\}}(y))\right)\oplus\left(\sigma_{1,I^c}(y)\oplus\sigma_{3,\{r\}^c}(h_2^{\{r\}^c}(y)\oplus i^{\otimes(d-1)})\right)
\end{aligned}
$$

此外，我们有

$$
\sum_{\pi}\sum_{\tau}\sum_{x\in U_{\pi,\tau}}w_x^2\le d2^c\left\|w\right\|_2^2 \tag{13}
$$

对所有 $w\in\mathbb{R}^u$ 均成立。

[3]在 在接下来的计算中，我们分别用 i 和 j 简记 $i_0,\ldots,i_{c-2}$ 和 $j_0,\ldots,j_{c-2}$ ly.

证明。我们先来证明等式（13）。令 $x \in [u]$，现在很容易看出，存在 2^c,$\pi \in P_{partial}$ 使得 $\pi \subseteq x$；类似地，存在 $d, \tau \in P_{prefix}$ 使得 $\tau \subseteq h_2(x)$。因此，存在 $d2^c$ 对 $(\pi, \tau) \in P_{partial} \times P_{prefix}$，使得 $x \in U_{\pi,\ \tau}$，从而等式（13）得证。

现在我们将着重证明等式（11）。首先定义事件 $A_{x,y,\tau}$ for $x, y \in [u]$ and $\tau \in P_{prefix}$ by $[A_{x,y,\tau}] = \left[h_{2,[|\tau|]}(x) = h_{2,[|\tau|]}(y) = \tau\right]\left[h_{2,\{|\tau|\}}(x) \neq h_{2,\{|\tau|\}}(y)\right]$. 然后我们可以写出

$$\begin{aligned}
&\left\| \sum_{i\in[s]} \sum_{x,y\in[u]} \left[h_2(x) \neq h_2(y)\right] \sigma(i,x)\sigma(i,y) \left[h(i,x) = h(i,y)\right] w_x w_y \right\|_p \\
&= \left\| \sum_{\tau\in P_{prefix}} \sum_{i\in[s]} \sum_{x,y\in[u]} \left[A_{x,y,\tau}\right] \sigma(i,x)\sigma(i,y) \left[h(i,x) = h(i,y)\right] w_x w_y \right\|_p \\
&\leq \sum_{\tau\in P_{prefix}} \left\| \sum_{i\in[s]} \sum_{x,y\in[u]} \left[A_{x,y,\tau}\right] \sigma(i,x)\sigma(i,y) \left[h(i,x) = h(i,y)\right] w_x w_y \right\|_p
\end{aligned}$$

然后我们定义函数 $h_r, \overline{h_r}\colon [s] \times [u] \to [m/s]$ 和 $\sigma_r, \overline{\sigma_r}\colon [s] \times [u]$，其中

$$\begin{aligned}
h_r(i,x) &= h_1(x) \oplus h_{3,\{r\}}(h_2^{\{r\}}(x) \oplus i) \oplus h_{3,\{r\}^c}(h_2^{\{r\}^c}(x) \oplus i^{\otimes(d-1)}) \\
\sigma_r(i,x) &= \sigma_1(x)\sigma_{3,\{r\}}(h_2^{\{r\}}(x) \oplus i) \cdot \sigma_{3,\{r\}^c}(h_2^{\{r\}^c}(x) \oplus i^{\otimes(d-1)}) \\
\overline{h_r}(i,y) &= h_1(y) \oplus \overline{h_{3,\{r\}}}(h_2^{\{r\}}(y) \oplus i) \oplus h_{3,\{r\}^c}(h_2^{\{r\}^c}(y) \oplus i^{\otimes(d-1)}) \\
\overline{\sigma_r}(i,y) &= \sigma_1(x)\overline{\sigma_{3,\{r\}}}(h_2^{\{r\}}(x) \oplus i) \cdot \sigma_{3,\{r\}^c}(h_2^{\{r\}^c}(x) \oplus i^{\otimes(d-1)})
\end{aligned}$$

对于固定的 $\tau \in P_{prefix}$，我们可以看到该表达式

$$\sum_{i\in[s]} \sum_{x,y\in[u]} \left[A_{x,y,\tau}\right] \sigma(i,x)\sigma(i,y) \left[h(i,x) = h(i,y)\right] w_x w_y$$

满足随机变量 $(h_{3,\ \{|\tau|\}}(\alpha)\sigma_{3,\ \{|\tau|\}}(\alpha))_{\alpha\in\Sigma}$ 的引理 16 要求。因此应用引理 16，我们得到

$$\begin{aligned}
&\sum_{\tau\in P_{prefix}} \left\| \sum_{i\in[s]} \sum_{x,y\in[u]} \left[A_{x,y,\tau}\right] \sigma(i,x)\sigma(i,y) \left[h(i,x) = h(i,y)\right] w_x w_y \right\|_p \\
&\leq \sum_{\tau\in P_{prefix}} 4 \left\| \sum_{i\in[s]} \sum_{x,y\in[u]} \left[x \neq y\right] \sigma_{|\tau|}(i,x)\overline{\sigma_{|\tau|}}(i,y) \left[h_{|\tau|}(i,x) = \overline{h_{|\tau|}}(i,y)\right] w_x w_y \right\|_p
\end{aligned}$$

现在我们可以将该表达式改写如下,

$$\sum_{\tau\in P_{prefix}} 4\left\|\sum_{i\in[s]}\sum_{x,y\in[u]}[x\neq y]\,\sigma_{|\tau|}(i,x)\overline{\sigma_{|\tau|}}(i,y)\left[h_{|\tau|}(i,x)=\overline{h_{|\tau|}}(i,y)\right]w_xw_y\right\|_p$$

$$=\sum_{\tau\in P_{prefix}} 4\left\|\sum_{\pi\in P_{partial}}\sum_{i\in[s]}\sum_{x,y\in[u]}[x\cap y=\pi]\,\sigma_{|\tau|}(i,x)\overline{\sigma_{|\tau|}}(i,y)\left[h_{|\tau|}(i,x)=\overline{h_{|\tau|}}(i,y)\right]w_xw_y\right\|_p$$

$$=\sum_{\pi\in P_{partial}}\sum_{\tau\in P_{prefix}} 4\left\|\sum_{i\in[s]}\sum_{x,y\in[u]}[x\cap y=\pi]\,\sigma_{|\tau|}(i,x)\overline{\sigma_{|\tau|}}(i,y)\left[h_{|\tau|}(i,x)=\overline{h_{|\tau|}}(i,y)\right]w_xw_y\right\|_p$$

对于固定的 $\pi\in P_{partial}$ 和 $\tau\in P_{prefix}$，我们看到该表达式

$$\sum_{i\in[s]}\sum_{x,y\in[u]}[x\cap y=\pi]\,\sigma_{|\tau|}(i,x)\overline{\sigma_{|\tau|}}(i,y)\left[h_{|\tau|}(i,x)=\overline{h_{|\tau|}}(i,y)\right]w_xw_y$$

满足引理 17 对随机变量 $(h_{1,I_\pi^c}(x)\,h_{2,I_\pi^c}(x)\,\sigma_{1,I_\pi^c}(x)\,\sigma_{2,I_\pi^c}(x))^{x\in[u]}$ 的要求。因此，应用引理 16 可得

$$\sum_{\pi\in P_{partial}}\sum_{\tau\in P_{prefix}} 4\left\|\sum_{i\in[s]}\sum_{x,y\in[u]}[x\cap y=\pi]\,\sigma_{|\tau|}(i,x)\overline{\sigma_{|\tau|}}(i,y)\left[h_{|\tau|}(i,x)=\overline{h_{|\tau|}}(i,y)\right]w_xw_y\right\|_p$$

$$\leq\sum_{\pi\in P_{partial}}\sum_{\tau\in P_{prefix}} 4^{c+1}\left\|\sum_{i\in[s]}\sum_{x,y\in[u]}[x\cap y=\pi]\,\sigma_{I_\pi,|\tau|}(i,x)\overline{\sigma_{I_\pi,|\tau|}}(i,y)\left[h_{I_\pi,|\tau|}(i,x)=\overline{h_{I_\pi,|\tau|}}(i,y)\right]w_xw_y\right\|_p$$

这就完成了等式（11）的证明，从而也完成了该引理的证明。 □

最后，我们可以证明混合制表具有 $(p,4^{c+2},4\varepsilon^{-2}3^c\frac{s}{m}\left|\Sigma\right|^{-d})$-解耦分解。

引理 19。

$$\Pr[|Z|\geq\varepsilon]$$

$$\leq\left(\varepsilon^{-1}\sum_{\pi\in P_{partial}}\sum_{\tau\in P_{prefix}}\frac{4^{c+4}}{s}\left\|\sum_{i\in[s]}\sum_{x,y\in U_{\pi,\tau}}\sigma_{I_\pi,|\tau|}(i,x)\overline{\sigma_{I_\pi,|\tau|}}(i,y)\left[h_{I_\pi,|\tau|}(i,x)=\overline{h_{I_\pi,|\tau|}}(i,y)\right]w_xw_y\right\|_p\right)^p$$
$$+4\varepsilon^{-2}3^c\frac{s}{m}\left|\Sigma\right|^{-d},$$

其中

$$h_{I,r}(i,x)=\left(h_{1,I}(x)\oplus h_{3,\{r\}}(h_{2,I}^{\{r\}}(x)\oplus i\oplus h_{2,I^c}^{\{r\}}(x))\right)\oplus\left(h_{1,I^c}(x)\oplus h_{3,\{r\}^c}(h_2^{\{r\}^c}(x)\oplus i^{\otimes(d-1)})\right)$$

$$\sigma_{I,r}(i,x)=\left(\sigma_{1,I}(x)\oplus\sigma_{3,\{r\}}(h_{2,I}^{\{r\}}(x)\oplus i\oplus h_{2,I^c}^{\{r\}}(x))\right)\oplus\left(\sigma_{1,I^c}(x)\oplus\sigma_{3,\{r\}^c}(h_2^{\{r\}^c}(x)\oplus i^{\otimes(d-1)})\right)$$

$$\overline{h_{I,r}}(i,y)=\left(\overline{h_{1,I}}(y)\oplus\overline{h_{3,\{r\}}}(\overline{h_{2,I}^{\{r\}}}(y)\oplus i\oplus h_{2,I^c}^{\{r\}}(y))\right)\oplus\left(h_{1,I^c}(y)\oplus h_{3,\{r\}^c}(h_2^{\{r\}^c}(y)\oplus i^{\otimes(d-1)})\right)$$

$$\overline{\sigma_{I,r}}(i,y)=\left(\overline{\sigma_{1,I}}(y)\oplus\overline{\sigma_{3,\{r\}}}(\overline{h_{2,I}^{\{r\}}}(y)\oplus i\oplus h_{2,I^c}^{\{r\}}(y))\right)\oplus\left(\sigma_{1,I^c}(y)\oplus\sigma_{3,\{r\}^c}(h_2^{\{r\}^c}(y)\oplus i^{\otimes(d-1)})\right)$$

此外，我们还有

$$\sum_{\pi}\sum_{\tau}\sum_{x\in U_{\pi,\tau}} w_x^2 \le d2^c \left\|w\right\|_2^2$$

对所有 $w \in \mathbb{R}^u$。

证明。我们使用并集界得到：

$$\begin{aligned}\Pr[|Z| \ge \varepsilon] \le \Pr\left[\left|\sum_{i\in[s]}\sum_{x\neq y\in[u]} \sigma_1(x)\sigma_1(y)\left[(h_1(x),h_2(x))=(h_1(y),h_2(y))\right]w_x w_y\right| \ge \varepsilon/2\right] \\ + \Pr\left[\left|\sum_{i\in[s]}\sum_{x,y\in[u]} \left[h_2(x)\neq h_2(y)\right]\sigma(i,x)\sigma(i,y)\left[h(i,x)=h(i,y)\right]w_x w_y\right| \ge \varepsilon/2\right]\end{aligned}$$

对于第一项，我们利用 $\mathrm{E}\left[(\sum_{i\in[s]}\sum_{x\neq y\in[u]}\sigma_1(x)\sigma_1(y)\left[(h_1(x)\ h_2(x))\ =(h_1(y)\ h_2(y))w_x w_y\right)^2\right] \le 3^c\left\|w\right\|_2^4\frac{s}{m}\left|\Sigma\right|^{-d}$。因此，应用马尔可夫不等式可得：

$$\Pr\left[\left|\sum_{i\in[s]}\sum_{x\neq y\in[u]} \sigma_1(x)\sigma_1(y)\left[(h_1(x),h_2(x))=(h_1(y),h_2(y))\right]w_x w_y\right| \ge \varepsilon/2\right] \le 4\varepsilon^{-2}3^c\frac{s}{m}\left|\Sigma\right|^{-d}.$$

对于第二项，我们对 p 应用马尔可夫不等式，然后由引理 18 可得结论。 □

6.3 集中性

本节旨在证明，混合制表是 (p,γ_p^c)-强集中型的，其中$\gamma_p = Kc\max\left\{1,\ \sqrt[p]{\text{对数}|\Sigma|}\right\}$，且 K为普适常数。接下来的 3~~个~~引理将分别证明强集中的各个部分。它们都遵循与 [20] 中结果相同的证明思路。

引理 20。对于任意价值函数 $v\colon\ [s]\times[m/s]$，以及任意向量 $w\in\mathbb{R}^U$，以下关于混合制表哈希的集中性结果成立

$$\left\|\sum_{i\in[s]}\sum_{x\in U}\sigma(i,x)v(i,h(i,x))w_x\right\|_p \le \Psi_p\left(\gamma_p^c\left\|v\right\|_\infty\left\|w\right\|_\infty, \gamma_p^c\frac{s}{m}normv2^2\left\|w\right\|_2^2\right), \tag{14}$$

$$\left\|\sum_{i\in[s]}\sum_{x\in U}\sigma(i,x)v(i,h(i,x))w_x\right\|_p \le \sqrt{\gamma_p^c\frac{p}{\log m/s}}\left\|v\right\|_2\left\|w\right\|_2. \tag{15}$$

此处 $\gamma_p = Kc\max\left\{1,\ \sqrt[p]{\log|\Sigma|}\right\}$ 对于一个普适常数 K。

证明。我们首先重新改写该表达式。

$$\left\|\sum_{i\in[s]}\sum_{x\in U}\sigma(i,x)v(i,h(i,x))w_x\right\|_p=\left\|\sum_{\alpha\in\Sigma}\sigma_2(\alpha)\sum_{i\in[s]}\sum_{x\in U}\sigma_1(x)v(i,h_1(x)\oplus h_3(x))\left[h_2(x)=\alpha\oplus i\right]w_x\right\|_p$$

我们将价值函数 v' 定义为: $U\times(\Sigma\times[m/s])\to\mathbb{R}$ 由 $v'(x,(i,j))=v(i,j)\left[i\in[s]\right]w_x$ 给出。于是可以将表达式写为

$$\begin{aligned}&\left\|\sum_{\alpha\in\Sigma}\sigma_2(\alpha)\sum_{i\in[s]}\sum_{x\in U}\sigma_1(x)v(i,h_1(x)\oplus h_3(x))\left[h_2(x)=\alpha\oplus i\right]w_x\right\|_p\\&=\left\|\sum_{\alpha\in\Sigma}\sigma_2(\alpha)\sum_{x\in U}v'(x,(\alpha\oplus h_2(x),h_3(\alpha)\oplus h_1(x)))\right\|_p\end{aligned}$$

现在我们先证明式 (15)。由引理 13 可得

$$\begin{aligned}&\left\|\sum_{\alpha\in\Sigma}\sigma_2(\alpha)\sum_{x\in U}v'(x,(\alpha\oplus h_2(x),h_3(\alpha)\oplus h_1(x)))\right\|_p\\&\le e\sqrt{\frac{p}{\log m/s}}\left\|\sum_{\alpha\in\Sigma}\sum_{j\in[m/s]}\left(\sum_{x\in U}v'(x,(\alpha\oplus h_2(x),j\oplus h_1(x)))\right)^2\right\|_{p/2}^{1/2}\end{aligned}$$

现在利用引理 15 可得

$$\begin{aligned}\left\|\sum_{\alpha\in\Sigma}\sum_{j\in[m/s]}\left(\sum_{x\in U}v'(x,(\alpha\oplus h_2(x),j\oplus h_1(x)))\right)^2\right\|_{p/2}&\le\gamma_p^c\sum_{x\in U}\sum_{i\in\Sigma}\sum_{j\in[m/s]}v'(x,(i,j))^2\\&=\gamma_p^c\sum_{x\in U}\sum_{i\in\Sigma}\sum_{j\in[m/s]}v(i,j)^2\left[i\in[s]\right]w_x^2\\&=\gamma_p^c\left\|v\right\|_2^2\left\|w\right\|_2^2\end{aligned}\tag{16}$$

由此可得

$$\begin{aligned}&\left\|\sum_{\alpha\in\Sigma}\sigma_2(\alpha)\sum_{x\in U}v'(x,(\alpha\oplus h_2(x),h_3(\alpha)\oplus h_1(x)))\right\|_p\\&\le e\sqrt{\gamma_p^c\frac{p}{\log m/s}}\left\|v\right\|_2\left\|w\right\|_2\end{aligned}$$

至此，式 (15) 的证明完成。

现在我们关注公式 (14)。利用引理 12 可得

$$\begin{aligned}&\left\|\sum_{\alpha\in\Sigma}\sigma_2(\alpha)\sum_{x\in U}v'(x,(\alpha\oplus h_2(x),h_3(\alpha)\oplus h_1(x)))\right\|_p\\&\le L\left\|\Psi_p\left(A,B\right)\right\|_p\end{aligned}$$

其中

$$A = \max_{\alpha\in\Sigma} \max_{j\in[m/s]} \left| \sum_{x\in U} v'(x, (\alpha \oplus h_2(x), h_3(\alpha) \oplus h_1(x))) \right|$$

$$B = \sum_{\alpha\in\Sigma} \sum_{j\in[m/s]} \left(\sum_{x\in U} v'(x, (\alpha \oplus h_2(x), j \oplus h_1(x))) \right)^2$$

应用引理 9，我们只需界定以下两个表达式

$$\left\| \max_{\alpha\in\Sigma} \max_{j\in[m/s]} \left| \sum_{x\in U} v'(x, (\alpha \oplus h_2(x), h_3(\alpha) \oplus h_1(x))) \right| \right\|_p$$

$$\left\| \sum_{\alpha\in\Sigma} \sum_{j\in[m/s]} \left(\sum_{x\in U} v'(x, (\alpha \oplus h_2(x), j \oplus h_1(x))) \right)^2 \right\|_{p/2}$$

由公式（16）可得

$$\left\| \sum_{\alpha\in\Sigma} \sum_{j\in[m/s]} \left(\sum_{x\in U} v'(x, (\alpha \oplus h_2(x), j \oplus h_1(x))) \right)^2 \right\|_{p/2} \leq \gamma_p^c \|v\|_2^2 \|w\|_2^2$$

令 $\bar{p} = \max\{p, \log(m/s \cdot |\Sigma|\}$，并使用引理 14 可得

$$\begin{aligned}
&\left\| \max_{\alpha\in\Sigma} \max_{j\in[m/s]} \left| \sum_{x\in U} v'(x, (\alpha \oplus h_2(x), j \oplus h_1(x))) \right| \right\|_p \\
&\leq \left\| \max_{\alpha\in\Sigma} \max_{j\in[m/s]} \left| \sum_{x\in U} v'(x, (\alpha \oplus h_2(x), j \oplus h_1(x))) \right| \right\|_{\bar{p}} \\
&\leq \left(\sum_{\alpha\in\Sigma} \sum_{j\in[m/s]} \left\| \sum_{x\in U} v'(x, (\alpha \oplus h_2(x), j \oplus h_1(x))) \right\|_{\bar{p}}^{1/\bar{p}} \right)^{1/\bar{p}} \\
&\leq e\Psi_{\bar{p}} \left(\gamma_p^c \max_{x\in U} \max_{i\in[s]} \max_{j\in[m/s]} |v(i,j)w_x| , \gamma_p^c \frac{s}{m\,|\Sigma|} \sum_{x\in U} \sum_{i\in[s]} \sum_{j\in[m/s]} v(i,j)^2 w_x^2 \right) \\
&\leq e\Psi_{\bar{p}} \left(\gamma_p^c \|v\|_\infty \|w\|_\infty , \gamma_p^c \frac{1}{|\Sigma|} \|v\|_2^2 \|w\|_2^2 \right)
\end{aligned}$$

由此可得

$$\begin{aligned}
&\left\| \sum_{\alpha\in\Sigma} \sigma_2(\alpha) \sum_{x\in U} v'(x, (\alpha \oplus h_2(x), h_3(\alpha) \oplus h_1(x))) \right\|_p \\
&\leq e\Psi_p \left(e\Psi_{\bar{p}} \left(\gamma_p^c \|v\|_\infty \|w\|_\infty , \gamma_p^c \frac{1}{|\Sigma|} \|v\|_2^2 \|w\|_2^2 \right) , \gamma_p^c \frac{s}{m} \|v\|_2^2 \|w\|_2^2 \right)
\end{aligned}$$

由于 $s \le \sqrt{|\Sigma|}$，与 [20, 定理 7] 的证明中相同的分类讨论可以得出，由于

$$\left\| \sum_{\alpha\in\Sigma} \sigma_2(\alpha) \sum_{x\in U} v'(x, (\alpha \oplus h_2(x), h_3(\alpha) \oplus h_1(x))) \right\|_p \le \Psi_p \left(\gamma_p^c \|v\|_\infty \|w\|_\infty, \gamma_p^c \frac{s}{m} \|v\|_2^2 \|w\|_2^2 \right)$$

证明至此完成。 □

引理 21。对于任意向量 $w \in \mathbb{R}^U$，混合制表哈希满足以下集中性结果,

$$\left\| \max_{i\in[s], j\in[m/s]} \left| \sum_{x\in U} \sigma(i,x) [h(i,x) = j] w_x \right| \right\|_p \le \sqrt{\gamma_p^c \frac{\log m}{\log m/s}} \|w\|_2$$

这里 $\gamma_p = Kc \max\left\{1, \sqrt[p]{\text{对数}|\Sigma|}\right\}$，其中 K 是一个普适常数。

证明。我们先重写这个表达式,

$$\begin{aligned} &\left\| \max_{i\in[s], j\in[m/s]} \left| \sum_{x\in U} \sigma(i,x) [h(i,x) = j] w_x \right| \right\|_p \\ &= \left\| \max_{i\in[s], j\in[m/s]} \left| \sum_{\alpha\in\Sigma} \sigma_3(\alpha) \sum_{x\in U} \sigma_1(x) [h_1(x) \oplus h_3(\alpha) = j] [h_2(x) = \alpha \oplus i] w_x \right| \right\|_p . \end{aligned}$$

现在我们固定 h_1, σ_1, h_2 的随机性，并利用引理 13 得到,

$$\begin{aligned} &\left\| \max_{i\in[s], j\in[m/s]} \left| \sum_{\alpha\in\Sigma} \sigma_3(\alpha) \sum_{x\in U} \sigma_1(x) [h_1(x) \oplus h_3(\alpha) = j] [h_2(x) = \alpha \oplus i] w_x \right| \right\|_p \\ &\le \left\| \max_{i\in[s], j\in[m/s]} \left| \sum_{\alpha\in\Sigma} \sigma_3(\alpha) \sum_{x\in U} \sigma_1(x) [h_1(x) \oplus h_3(\alpha) = j] [h_2(x) = \alpha \oplus i] w_x \right| \right\|_{\log m} . \\ &\le e \max_{i\in[s], j\in[m/s]} \left\| \left| \sum_{\alpha\in\Sigma} \sigma_3(\alpha) \sum_{x\in U} \sigma_1(x) [h_1(x) \oplus h_3(\alpha) = j] [h_2(x) = \alpha \oplus i] w_x \right| \right\|_{\log m} \\ &\le e^2 \sqrt{\frac{\log m}{\log m/s}} \max_{i\in[s], j\in[m/s]} \sqrt{\sum_{\alpha\in\Sigma} \sum_{k\in[m/s]} \left(\sum_{x\in U} \sigma_1(x) [h_1(x) = j \oplus k] [h_2(x) = \alpha \oplus i] w_x \right)^2} \end{aligned}$$

我们现在注意到，表达式 ~~$\sqrt{\sum_{\alpha\in\Sigma} \sum_{k\in[m/s]} \left(\sum_{x\in U} \sigma_1(x) [h_1(x) = j \oplus k] [h_2(x) = \alpha \oplus i] w_x \right)^2}$ 不依~~ 赖于 i 和 j，因此不妨考察 $i = j = 0$。由此可得,

$$\begin{aligned} &\left\| \max_{i\in[s], j\in[m/s]} \left| \sum_{x\in U} \sigma(i,x) [h(i,x) = j] w_x \right| \right\|_p \\ &\le e^2 \sqrt{\frac{\log m}{\log m/s}} \left\| \sum_{\alpha\in\Sigma} \sum_{k\in[m/s]} \left(\sum_{x\in U} \sigma_1(x) [h_1(x) = k] [h_2(x) = \alpha] w_x \right)^2 \right\|_{p/2}^{1/2} \end{aligned}$$

然后，我们可以应用引理 15，得到

$$\begin{aligned}&\left\|\max_{i\in[s],j\in[m/s]}\left|\sum_{x\in U}\sigma(i,x)\left[h(i,x)=j\right]w_x\right|\right\|_p\\&\leq\sqrt{\gamma_p^c\frac{\log m}{\log m/s}}\left\|w\right\|_2\end{aligned}$$

至此，证明完毕。 □

引理 22. 对于任意向量 $w\in\mathbb{R}^U$，混合制表哈希满足以下集中性结果,

$$\left\|\sum_{i\in[s]}\sum_{j\in[m/s]}\left(\sum_{x\in U}\sigma(i,x)\left[h(i,x)=j\right]w_x\right)^2\right\|_p\leq\gamma_p^c\max\left\{s\left\|w\right\|_2^2,\frac{p}{\log m/s}\left\|w\right\|_2^2\right\}$$

这里 $\gamma_p=Kc\max\left\{1,\ \sqrt[p]{\log|\Sigma|}\right\}$，其中 K 为某个普适常数。

证明。我们先将该表达式改写为

$$\begin{aligned}&\left\|\sum_{i\in[s]}\sum_{j\in[m/s]}\left(\sum_{x\in U}\sigma(i,x)\left[h(i,x)=j\right]w_x\right)^2\right\|_p\\&\quad=\left\|\sum_{i\in[s]}\sum_{x,y\in U}\sigma(i,x)\sigma(i,y)\left[h(i,x)=h(i,y)\right]w_xw_y\right\|_p\\&\quad\leq\left\|\sum_{i\in[s]}\sum_{x,y\in U}\sigma(i,x)\sigma(i,y)\left[(h_1(x),h_2(x))=(h_1(y),h_2(y))\right]w_xw_y\right\|_p\\&\qquad+\left\|\sum_{i\in[s]}\sum_{x,y\in U}\sigma(i,x)\sigma(i,y)\left[h(i,x)=h(i,y)\right]\left[h_2(x)\neq h_2(y)\right]w_xw_y\right\|_p\\&\quad=s\left\|\sum_{z\in\Sigma^d}\sum_{j\in[m/s]}\left(\sum_{x\in U}\sigma_1(x)\left[(h_1(x),h_2(x))=(j,z)\right]w_x\right)^2\right\|_p\\&\qquad+\left\|\sum_{i\in[s]}\sum_{x,y\in U}\sigma(i,x)\sigma(i,y)\left[h(i,x)=h(i,y)\right]\left[h_2(x)\neq h_2(y)\right]w_xw_y\right\|_p\end{aligned}$$

我们将分别界定这两个表达式。利用引理 15，我们得到的是

$$s\left\|\sum_{z\in\Sigma^d}\sum_{j\in[m/s]}\left(\sum_{x\in U}\sigma_1(x)\left[(h_1(x),h_2(x))=(j,z)\right]w_x\right)^2\right\|_p\leq\gamma_p^cs\left\|w\right\|_2^2$$

为了界定另一个表达式，我们使用引理 18 得到，

$$\left\|\sum_{i\in[s]}\sum_{x,y\in U}\sigma(i,x)\sigma(i,y)\left[h(i,x)=h(i,y)\right]\left[h_2(x)\neq h_2(y)\right]w_xw_y\right\|_p$$

$$\leq\sum_{\pi\in P_{partial}}\sum_{\tau\in P_{prefix}}4^{c+1}\left\|\sum_{i\in[s]}\sum_{x,y\in U_{\pi,\tau}}\sigma_{I_\pi,|\tau|}(i,x)\overline{\sigma_{I_\pi,|\tau|}}(i,y)\left[h_{I_\pi,|\tau|}(i,x)=\overline{h_{I_\pi,|\tau|}}(i,y)\right]w_xw_y\right\|_p$$

对于每个 $\pi\in P_{partial}$ 和每个 $\tau\in P_{prefix}$，这对应于一个具有 $c'\leq 2c$ 和 $d'\leq 2$ 的混合制表哈希函数。随后我们可以使用等式（15）得到

$$\sum_{\pi\in P_{partial}}\sum_{\tau\in P_{prefix}}4^{c+1}\left\|\sum_{i\in[s]}\sum_{x,y\in U_{\pi,\tau}}\sigma_{I_\pi,|\tau|}(i,x)\overline{\sigma_{I_\pi,|\tau|}}(i,y)\left[h_{I_\pi,|\tau|}(i,x)=\overline{h_{I_\pi,|\tau|}}(i,y)\right]w_xw_y\right\|_p$$

$$\leq\sum_{\pi\in P_{partial}}\sum_{\tau\in P_{prefix}}4^{c+1}\gamma_p^c\sum_{x\in U_{\pi,\tau}}w_x^2$$

随后我们使用等式（13）得到，

$$\sum_{\pi\in P_{partial}}\sum_{\tau\in P_{prefix}}4^{c+1}\gamma_p^c\sum_{x\in U_{\pi,\tau}}w_x^2\leq\gamma_p^cd8^c\left\|w\right\|_2^2$$

结合这些界，我们得到

$$\left\|\sum_{i\in[s]}\sum_{j\in[m/s]}\left(\sum_{x\in U}\sigma(i,x)\left[h(i,x)=j\right]w_x\right)^2\right\|_p\leq\gamma_p^c+\gamma_p^cd8^c\left\|w\right\|_2^2$$

$$\leq\gamma_p^c\max\left\{s\left\|w\right\|_2^2,\frac{p}{\log m/s}\left\|w\right\|_2^2\right\}$$

正如所需，从而完成证明。 □

参考文献

[1] Anders Aamand、Jakob Bæk Tejs Knudsen、Mathias Bæk Tejs Knudsen、Peter Michael Re- ichstein Rasmussen 和 Mikkel Thorup。具有强集中界的快速哈希。载于第52届 ACM SIGACT 理论计算年会论文集（STOC 2020），第1265–1278页，纽约，纽约州，美国，2020年。计算机协会。[2] Anders Aamand 和 Mikkel Thorup。使用简单制表哈希的非空箱。载于 Timothy M. Chan（编），第三十届 ACM-SIAM 离散算法研讨会论文集（SODA 2019），美国加利福尼亚州圣迭戈，2019年1月6日至9日，第2498–2512页。SIAM，2019年。[3] Dimitris Achlioptas。适用于数据库的随机投影：使用二元随机变量的 Johnson-Lindenstrauss 变换。《计算机与系统科学杂志》，66(4):671–687，2003年。P ODS 2001 专刊。

[4] Nir Ailon 和 Bernard Chazelle。快速 Johnson–Lindenstrauss 变换与近似最近邻。《SIAM 计算杂志》，39(1):302–322，2009年。[5] Nir Ailon 和 Edo Liberty。利用对偶 BCH 码上的 Rademacher 级数快速降维。载于 Shang-Hua Teng（编），第十九届 ACM-SIAM 离散算法研讨会论文集（SODA 2008），美国加利福尼亚州旧金山，2008年1月20日至22日，第1–9页。SIAM，2008年。[6] Noga Alon 和 Bo'az Klartag。近似内积的最优压缩与降维。载于 2017 IEEE 第58届计算机科学基础年会（FOCS），第639–650页，2017年10月。[7] Stefan Bamberger 和 Felix Krahmer。适用于大型数据集的最优快速 Johnson-Lindenstrauss 嵌入。《采样理论、信号处理与数据分析》，19，2021年6月。[8] Vladimir Braverman、Rafail Ostrovsky 和 Yuval Rabani。Rademacher 混沌、随机欧拉图与稀疏 Johnson-Lindenstrauss 变换。CoRR，abs/1011.2590，2010年。[9] Moses Charikar、Kevin Chen 和 Martin Farach-Colton。寻找数据流中的频繁项。《理论计算机科学》，312(1):3–15，2004年。自动机、语言与编程。[10] Michael B. Cohen、T. S. Jayram 和 Jelani Nelson。稀疏 Johnson-Lindenstrauss 变换的简单分析。载于 Raimund Seidel（编），第一届算法简洁性研讨会（SOSA 2018），2018年1月7日至10日，美国路易斯安那州新奥尔良，OASIcs 第61卷，第15:1–15:9页。Schloss Dagstuhl - Leibniz-Zentrum f¨ur Informatik，2018年。[11] S. Dahlgaard、M. B. T. Knudsen、E. Rotenberg 和 M. Thorup。面向 k-划分统计的哈希。载于 2015 IEEE 第56届计算机科学基础年会，第1292–1310页，2015年。[12] Søren Dahlgaard、Mathias Bæk Tejs Knudsen 和 Mikkel Thorup。用于相似度估计与降维的实用哈希函数。载于第31届神经信息处理系统国际会议论文集（NIPS'17），第6618–6628页，美国，2017年。Curran Associates Inc.[13] Søren Dahlgaard 和 Mikkel Thorup。使用扭曲制表实现近似最小值独立性。2014年3月。[14] Anirban Dasgupta、Ravi Kumar 和 Tam´as Sarlos。稀疏 Johnson-Lindenstrauss 变换。STOC '10，第341–350页，纽约，纽约州，美国，2010年。计算机协会。[15] Thong T. Do、Lu Gan、Yi Chen、Nam Nguyen 和 Trac D. Tran。利用结构化随机矩阵实现快速高效的降维。载于 2009 IEEE 声学、语音与信号处理国际会议，第1821–1824页，2009年。[16] Ora Nova Fandina、Mikael Møller Høgsgaard 和 Kasper Green Larsen。更快降维的障碍，2022年。

[17] Casper Freksen、Lior Kamma 和 Kasper Green Larsen。全面理解哈希技巧。载于《第32届神经信息处理系统国际会议论文集》，NIPS'18，第5394–5404页，纽约州红钩，美国，2018年。Curran Associates Inc.[18] Casper Benjamin Freksen 和 Kasper Green Larsen。关于使用托普利兹矩阵和循环矩阵进行 Johnson-Lindenstrauss 变换。Algorithmica，82(2):338–354，2020年。[19] Aicke Hinrichs 和 Jan Vyb´ıral。循环矩阵的 Johnson-Lindenstrauss 引理。Random Structures & Algorithms，39(3):391–398，2011年。[20] Jakob Bæk Tejs Houen 和 Mikkel Thorup。通过混沌理解制表哈希的矩。载于 Mikolaj Bojanczyk、Emanuela Merelli 和 David P. Woodruff（编），《第49届自动机、语言与编程国际研讨会，ICALP 2022》，2022年7月4日至8日，法国巴黎，LIPIcs第229卷，第74:1–74:19页。Schloss Dagstuhl - Leibniz- Zentrum f¨ur Informatik，2022年。[21] Meena Jagadeesan。理解特征哈希中的稀疏 JL。载于《第33届神经信息处理系统国际会议论文集》，NeurIPS'19，美国纽约州红钩，2019年。Curran Associates Inc.[22] Vishesh Jain、Natesh S. Pillai、Ashwin Sah、Mehtaab Sawhney 和 Aaron Smith。利用 Kac 游走实现快速且内存最优的降维。《应用概率年刊》，32(5):4038–4064，2022年。[23] T. S. Jayram 和 David P. Woodruff。具有次常数误差的 Johnson-Lindenstrauss 变换和流式问题的最优界。ACM 算法汇刊，9(3)，2013 年 6 月。[24] William Johnson 和 Joram Lindenstrauss。将 Lipschitz 映射扩展到希尔伯特空间。《当代数学》，26:189–206，1984年1月。[25] Mark Kac。动力学理论基础。载于《第三届伯克利数学统计与概率研讨会论文集》，第3卷，第171–197页，1956年。[26] Daniel Kane、Raghu Meka 和 Jelani Nelson。近乎最优的显式 Johnson-Lindenstrauss 族。载于 Leslie Ann Goldberg、Klaus Jansen、R. Ravi 和 Jos´e D. P. Rolim（编），《近似、随机化与组合优化：算法与技术》，第628–639页，柏林、海德堡，2011年。Springer Berlin Heidelberg。[27] Daniel M. Kane 和 Jelani Nelson。一种去随机化的稀疏 Johnson-Lindenstrauss 变换，2010年。[28] Daniel M. Kane 和 Jelani Nelson。更稀疏的 Johnson-Lindenstrauss 变换。《ACM学报》，61(1)，2014年1月。[29] Felix Krahmer 和 Rachel Ward。通过受限等距性质实现改进的 Johnson–Lindenstrauss 嵌入。《SIAM 数学分析杂志》，43(3):1269–1281，2011年。[30] Kasper Green Larsen 和 Jelani Nelson。Johnson-Lindenstrauss 引理的最优性。载于《2017 IEEE 第58届计算机科学基础年会（FOCS）》，第633–638页，2017年。

[31] Jelani Nelson and Huy L. Nguy ˚An。降维映射的稀疏度下界。载于第 45 届 ACM 理论计算年会论文集，STOC ’13，第 101–110 页，美国纽约州纽约，2013。计算机协会。[32] Mihai Patrascu 和 Mikkel Thorup。扭曲制表哈希。载于 Sanjeev Khanna（编），第二十四届 ACM-SIAM 离散算法研讨会论文集，SODA 2013，2013 年 1 月 6–8 日，美国路易斯安那州新奥尔良，第 209–228 页。SIAM，2013。[33] Mihai P˘atra¸scu 和 Mikkel Thorup。简单制表哈希的威力。《ACM 杂志》，59(3)，2012 年 6 月。[34] Alan Siegel。关于极其随机且具有常数时间复杂度的哈希函数的通用类。《SIAM 计算杂志》，33(3):505–543，2004。曾在 FOCS’89 上报告。[35] Mikkel Thorup。简单制表、快速扩展器、双重制表和高独立性。载于第 54 届计算机科学基础年会（FOCS），第 90–99 页，2013。[36] Mikkel Thorup 和 Yin Zhang。基于制表的 5-独立哈希及其在线性探测和二阶矩估计中的应用。《SIAM 计算杂志》，41(2):293–331，2012。[37] Roman Vershynin。《高维概率：数据科学中的导论与应用》。剑桥统计与概率数学丛书。剑桥大学出版社，2018。[38] Jan Vyb´ıral。循环矩阵的 Johnson–Lindenstrauss 引理变体。《泛函分析杂志》，260(4):1096–1105，2011。[39] Kilian Weinberger、Anirban Dasgupta、John Langford、Alex Smola 和 Josh Attenberg。用于大规模多任务学习的特征哈希。载于第 26 届国际机器学习年会论文集，ICML ’09，第 1113–1120 页，美国纽约州纽约，2009。计算机协会。[40] Albert Lindsey Zobrist。一种用于棋类游戏的新哈希方法。技术报告 88，威斯康星大学麦迪逊分校计算机科学系，威斯康星州麦迪逊，1970。