北京大学

本科生毕业论文

题目：二元矩阵的低秩近似

On Low Rank Approximation of Binary Matrices

姓　　名：丹晨

学　　号：1200012743

院　　系：信息科学技术学院

专　　业：智能科学与技术

导　　师：王立威教授

2016年5月

摘要

我们考虑二元矩阵的低秩近似问题。这里我们给定 $d \times n$ 的二元矩阵 $\mathbf{A}$以及一个小整数 $k < d$。我们的目标是寻找两个大小分别为$d \times k$ 和 $k \times n$ 的二元矩阵 $\mathbf{U}$ 和 $\mathbf{V}$，使得 $\mathbf{A} - \mathbf{UV}$ 的 Frobenius 范数最小化。依赖于二元矩阵乘积的不同定义，这个问题有两种不同的形式： GF(2) 形式和布尔形式。在之前，这个问题仅有的结果是对于特例$k = 1$ 的 2-近似算法 [1, 2] (此时两种形式是等价的)。

在本文中，我们给出 首个 对于一般情形 $k > 1$ 的结果，同时包含了GF(2) 模型和布尔模型。对于 GF(2) 形式，我们给出了一个简单的列选择算法，并证明其达到了$O(k)$的近似比。对于布尔形式，我们给出了另一个算法，并证明其达到了 $O(2^k)$的近似比。对于常数 k, 两个算法的时间复杂度均为矩阵大小的多项式时间。我们同时证明了二元矩阵的低秩近似问题即使在$k = 1$的特例下也是 NP-难的，解决了 [3] 中的一个猜想。

关键词：低秩近似，二元矩阵，近似算法

On Low Rank Approximation of Binary Matrices

Chen Dan (Machine Intelligence)
Directed by Prof. Liwei Wang

ABSTRACT

We consider the problem of low rank approximation of binary matrices. Here we are given a $d \times n$ binary matrix $\mathbf{A}$ and a small integer $k < d$. The goal is to find two binary matrices $\mathbf{U}$ and $\mathbf{V}$ of sizes $d \times k$ and $k \times n$ respectively, so that the Frobenius norm of $\mathbf{A} - \mathbf{U}\mathbf{V}$ is minimized. There are two models of this problem, depending on the definition of the product of binary matrices: The $\mathrm{GF}(2)$ model and the Boolean semiring model. Previously, the only known results are 2-approximation algorithms for the special case $k = 1$ [1, 2] (where the two models are equivalent).

In this paper, we give the *first* results for the general case $k > 1$ for both $\mathrm{GF}(2)$ and Boolean model. For the $\mathrm{GF}(2)$ model, we show that a simple column-selection algorithm achieves $O(k)$-approximation. For the Boolean model, we develop a new algorithm and show that it is $O(2^k)$-approximation. For constant k, both algorithms run in polynomial time in the size of the matrix. We also show that the low rank binary matrix approximation problem is NP-hard even for $k = 1$, solving a conjecture in [3].

KEYWORDS: Low Rank Approximation, Binary Matrices, Approximation Algorithms

目录

序言

低秩近似是一个经典问题。给定一个大小为$d \times n$的矩阵$\mathbf{A}$，我们的目标是找到一个秩为k的矩阵，使得它是$\mathbf{A}$的近似。具体地说，是求解如下的优化问题：

$$\min_{\mathbf{U},\mathbf{V}} \|\mathbf{A} - \mathbf{U}\mathbf{V}\|_F^2, \tag{1}$$

其中$\mathbf{U}, \mathbf{V}$的大小分别是$d \times k$ 和 $k \times n$，k通常是一个较小的整数，即想要得到的秩。这里，近似的误差用Frobenius范数$\|\cdot\|_F$来衡量.

在许多应用中，$\mathbf{A}$是数据矩阵.$\mathbf{A}$的每一列是一个d维向量，而$\mathbf{A}$的每一列对应着一种属性或者特征。$\mathbf{A}$的低秩近似通常被称为因子分析和降维：矩阵$\mathbf{U}$的k列是“因子”，即低维空间的一组基，而矩阵$\mathbf{V}$的每一列则表示数据在新的基上的线性组合系数。

若矩阵$\mathbf{A}$，$\mathbf{U}$，$\mathbf{V}$是实数矩阵，那么低秩近似可以使用奇异值分解(SVD)有效求解。这个问题已经提出并被研究了超过一百年，被不同文献分别称作主成分分析(PCA)[4]，Karhunen-Loève变换[5]，等等。

在本文中，我们考虑二元矩阵的低秩近似问题。研究这个问题的动机来源于在许多现实应用中，数据是二元（类别）形式，而非连续形式的。在二元形式的问题中，我们要求$\mathbf{A}$和秩-k矩阵$\mathbf{U}, \mathbf{V}$都是二元的，这与经典的连续形式问题不同。依赖于向量内积的不同定义，在二元矩阵的低秩近似问题有两种模型。第一种被称为GF(2)模型，在这个模型中$\mathbf{u}, \mathbf{v}$的内积被定义为$\mathbf{u}^T\mathbf{v} := \oplus_i u_i v_i$。另一种模型被称为布尔模型，这里内积被定义为$\mathbf{u}^T\mathbf{v} := \bigvee_i (u_i \wedge v_i)$。

布尔模型通常被称为布尔因子分析(Boolean Factor Analysis (BFA))。在机器学习和数据挖掘中，这个模型具有许多应用，包括隐变量分析，话题模型(topic models)，关系规则挖掘(association rule mining)，聚类，以及数据库覆盖 (database tiling) [6–10]. Belohlavek and Vychodil [6] 研究了如下问题：求$\mathbf{A}$尽可能低秩的精确分解，即$\mathbf{A} = \mathbf{U}\mathbf{V}$，使得$\mathbf{U}, \mathbf{V}$ 的布尔秩尽可能小。他们给出了利用形式概念分析(Formal Concept Analysis)给出此问题的一个等价表述，并设计了一个贪心算法。对于在实际应用中更为流行的低秩近似问题，已有的工作包含了许多经验性算法，例如[7, 11–13]。然而，据我们所知，这些算法并没有任何理论保证。

对于低秩近似的GF(2)模型，我们没有发现任何已有的工作。一个相关的研究是GF(2)上的独立成分分析(Independent Component Analysis, ICA)，它在信号处理领域受到了越来越广泛的关注[14–16]。

在本文中，我们所考虑的二元秩-k近似问题的形式表述如下。给定一个二元矩阵$\mathbf{A} \in \{0,1\}^{d\times n}$,

$$\min_{\mathbf{U}\in\{0,1\}^{d\times k},\mathbf{V}\in\{0,1\}^{k\times n}} \|\mathbf{A} - \mathbf{UV}\|_F^2. \tag{2}$$

其中，矩阵乘积$\mathbf{UV}$可以分别定义在GF(2)和布尔半环上，对应问题的两种不同模型。

此前，这个问题仅有的理论结果是研究了$k = 1$的特例(此时GF(2)模型和布尔模型是等价的)，给出了2-近似算法。我们将在一节中给出他们的具体表述。

此外，二元低秩近似问题的计算复杂性的研究此前也较为有限。对于秩-1近似问题，Tan证明了与本问题等价的最大加权二分团(Maximum Edge Weight Biclique) 若有多项式时间算法，则$NP = RP$。当秩 k是输入的一部分时，在布尔半环上判断是否存在矩阵$\mathbf{U}$和 $\mathbf{V}$使得$\mathbf{A} = \mathbf{UV}$等价于NP-完全的最小基集合(Minimal Set Basis) 问题[17]，这也推出了此时给出低秩近似问题的任意近似比的算法均为NP-难的，参见[18]。另一方面，这并不能推出当$k \ll d, n$时问题的困难性。事实上， 最小基集合(Minimal Set Basis)在参数k给定时，可以通过一个简单的核化(kernelization)算法求解[19]。同时，注意到在 GF(2)模型下，寻找使得$\mathbf{A} = \mathbf{UV}$的$\mathbf{U}$和 $\mathbf{V}$可以通过高斯消元法有效求解，即使k是输入的一部分。从这个角度可以看出，本问题的GF(2)模型相比布尔半环模型，其计算复杂性相对更低一些。

主要成果

在本文中，我们给出首个对于一般情形$k > 1$的结果，同时包含了GF(2)模型和布尔模型。对于GF(2)形式，我们给出了一个简单的列选择算法，并证明其达到了$O(k)$的近似比。对于布尔形式，我们给出了另一个算法，并证明其达到了$O(2^k)$的近似比.对于常数k，两个算法的时间复杂度均为矩阵大小的多项式时间。

GF(2)模型:我们首先给出GF(2)模型下的近似算法。我们证明我们的算法是$(\frac{k}{2} + 1 + \frac{k}{2(2^k-1)})$-近似算法(定理 1)。当$k = 1$时，这个算法是2-近似，与这个特例下的现有结果近似比相同[1, 2]。我们同时证明了这个近似比是紧的，即对所有的k和$\epsilon > 0$，存在一个二元矩阵$\mathbf{A}$，使得算法在这个矩阵上取得的近似比不小于$\frac{k}{2} + 1 + \frac{k}{2(2^k-1)} - \epsilon$(定理2)。

这个近似算法十分简洁。我们只需简单地取出$\mathbf{A}$的k列构成基矩阵$\mathbf{U}$，然后计算$\mathbf{A}$在$\mathbf{U}$上的投影$\mathbf{V}$，选择使得近似误差最小的解即可。

对于本算法，主要的难度在于近似比的证明，即证明算法的近似比为$(\frac{k}{2} + 1 + \frac{k}{2(2^k-1)})$。这里，我们给出一个简单的证明概要。考虑方程(2)中的问题，在本文中，我们称$\mathbf{U}$为基矩阵，因为其列向量是低维空间的一组基；并称$\mathbf{V}$为系数矩阵，因为其列向量包含了线性组合系数。对于最优基矩阵$\mathbf{U}$的每一列$\mathbf{u}_i$，考虑其在$\mathbf{A}$的列向

量中的最近邻。令$\mathbf{a}_1,\ldots,\mathbf{a}_n$为$\mathbf{A}$的$n$列。记$\mathbf{a}_{(\mathbf{u}_i)}$为$\mathbf{u}_i$在$\mathbf{A}$的列向量中的最近邻。给定最优基矩阵$\mathbf{U}$，我们便可以得到由$\mathbf{A}$的列向量组成的矩阵$\mathbf{A}_{(\mathbf{U})} := (\mathbf{a}_{(\mathbf{u}_1)},\ldots,\mathbf{a}_{(\mathbf{u}_k)})$。注意到方程(2)的最优解并不唯一。事实上，固定最优基矩阵$\mathbf{U}$，对于每个矩阵
$\mathbf{B} = (\mathbf{b}_1,\ldots,\mathbf{b}_k)$,$\mathbf{b}_i \in \{0,1\}^k$，只要$\mathbf{B}$在GF(2)上的秩为$k$，$(\mathbf{UB},\mathbf{B}^{-1}\mathbf{V})$就一定也是一组最优解。每个最优基矩阵 $\mathbf{UB}$诱导出了一个最近邻矩阵$\mathbf{A}_{(\mathbf{UB})}$。我们将证明一定存在一个秩为$k$的矩阵$\mathbf{B}$，当以其诱导出的最近邻矩阵$\mathbf{A}_{(\mathbf{UB})}$为基矩阵时，近似误差一定不超过最优解 $(\mathbf{UB},\mathbf{B}^{-1}\mathbf{V})$的$(\frac{k}{2}+1+\frac{k}{2(2^k-1)})$倍。令$\mathrm{Err}(\mathbf{b}_1,\ldots,\mathbf{b}_k)$为$\mathbf{B} = (\mathbf{b}_1,\ldots,\mathbf{b}_k)$时基矩阵$\mathbf{A}_{(\mathbf{UB})}$对应的近似误差。我们的目标是给出下面式子的界：

$$\min_{\mathbf{b}_1,\ldots,\mathbf{b}_k} \mathrm{Err}(\mathbf{b}_1,\ldots,\mathbf{b}_k), \tag{3}$$

其中$\mathbf{b}_i \in \{0,1\}^k$对所有$i \in [k]$成立。

直接给出(3)的界无疑是非常困难的。为了解决这一难题，我们考虑一系列误差最小化问题，共k+1个。对于第r个$(0 \le r \le k)$最小化问题，我们只最优化$\mathbf{b}_1,\ldots,\mathbf{b}_k$中$r$个向量变动，固定其余$k-r$个向量不动的情形。给定$\mathbf{b}_1,\ldots,\mathbf{b}_k$, 令

$$\begin{aligned}
\mathrm{Err}^{(0)}(\mathbf{b}_1,\ldots,\mathbf{b}_k) &:= \mathrm{Err}(\mathbf{b}_1,\ldots,\mathbf{b}_k), && (4)\\
\mathrm{Err}^{(r)}(\mathbf{b}_1,\ldots,\mathbf{b}_{k-r}) &:= \min_{\mathbf{b}\in\{0,1\}^k} \mathrm{Err}^{(r-1)}(\mathbf{b}_1,\ldots,\mathbf{b}_{k-r},\mathbf{b}), && (5)\\
\mathrm{Err}^{(k)}() &:= \min_{\mathbf{b}\in\{0,1\}^k} \mathrm{Err}^{(k-1)}(\mathbf{b}), && (6)
\end{aligned}$$

则$\mathrm{Err}^{(k)}()$等价于方程(3)。

我们的最终目标是给出$\mathrm{Err}^{(k)}()$与方程(2)的最优解所产生的误差之比的上界，考虑其推广问题：对所有$0 \le r \le k$，$\mathrm{Err}^{(r)}(\mathbf{b}_1,\ldots,\mathbf{b}_{k-r})$的上界。为使描述更加精确，令$\mathrm{OPT}_k$为方程(2)的最优解所产生的误差，我们将证明$\mathrm{Err}^{(r)}(\mathbf{b}_1,\ldots,\mathbf{b}_{k-r})$的大小可以被$\mathrm{OPT}_k$加上一个依赖于$r$的量以及$\mathbf{b}_1,\ldots,\mathbf{b}_{k-r}$限制住(定理4)。重点在于$r = k$时，这个上界变成了$\mathrm{OPT}_k$的倍数，这正是我们想要的。之所以引入$\mathrm{Err}^{(0)},\ldots,\mathrm{Err}^{(k-1)}$是因为我们希望利用$\mathrm{Err}^{(r)}$与$\mathrm{Err}^{(r-1)}$的关系去证明我们希望得到的上界。事实上，推广问题的上界可以对r做数学归纳法证明。

尽管$\mathrm{Err}^{(r)}$与$\mathrm{Err}^{(r-1)}$的关系非常清楚：

$$\mathrm{Err}^{(r)}(\mathbf{b}_1,\ldots,\mathbf{b}_{k-r}) = \min_{\mathbf{b}} \mathrm{Err}^{(r-1)}(\mathbf{b}_1,\ldots,\mathbf{b}_{k-r},\mathbf{b}),$$

但是直接对$\mathbf{b}$做最优化是非常困难的。我们采用*加权平均*的方法。因为对每个
$\mathbf{b} \in \{0,1\}^k$, $\mathrm{Err}^{(r)}(\mathbf{b}_1,\ldots,\mathbf{b}_{k-r}) \le \mathrm{Err}^{(r-1)}(\mathbf{b}_1,\ldots,\mathbf{b}_{k-r},\mathbf{b})$, 所以对任意$w_{\mathbf{b}}$满足$w_{\mathbf{b}} \ge$

0并且$\sum_{\mathbf{b}} w_{\mathbf{b}} = 1$, $\mathrm{Err}^{(r)}(\mathbf{b}_1,\ldots,\mathbf{b}_{k-r}) \le \sum_{\mathbf{b}} w_{\mathbf{b}} \mathrm{Err}^{(r-1)}(\mathbf{b}_1,\ldots,\mathbf{b}_{k-r},\mathbf{b})$. 我们谨慎地选取权重$w_{\mathbf{b}}$去得到一个小的上界。我们进行两层加权平均。考虑商空间$\mathrm{GF}(2)^k/\mathrm{span}(\mathbf{b}_1,\ldots,\mathbf{b}_{k-r})$和陪集$[\mathbf{b}] := \mathbf{b} + \mathrm{span}(\mathbf{b}_1,\ldots,\mathbf{b}_{k-r})$。在第一层中，我们对陪集$[\mathbf{b}]$做加权平均，并且得到一个依赖于陪集的$\mathrm{Err}^{(r)}$的上界。在第二层中，我们用另一系列权重对所有陪集做加权平均。在陪集$[\mathbf{b}]$内，我们按照如下方法挑选权重：令$\mathbf{U},\mathbf{V}$为已经确定的方程(2)的最优解。对每个$\mathbf{c} \in [\mathbf{b}]$，记$n_{\mathbf{c}}$为$\mathbf{V}$的所有列向量中等于$\mathbf{c}$的个数。我们对$\mathbf{c}$赋予的权重正比于$n_{\mathbf{c}}$。对于第二层，令$n_{[\mathbf{b}]} := \sum_{\mathbf{c}\in[\mathbf{b}]}$为$\mathbf{V}$的所有列中属于陪集$[\mathbf{b}]$的总数目。我们按如下方法对陪集$[\mathbf{b}]$赋予权重：如果$[\mathbf{b}] = \mathrm{span}(\mathbf{b}_1,\ldots,\mathbf{b}_{k-r})$，则其权重赋为零。对于其他情形，我们对$[\mathbf{b}]$分配的权重正比于$\frac{n_{[\mathbf{b}]}}{\sum_{[\mathbf{b}]} n_{[\mathbf{b}]} - \lambda n_{[\mathbf{b}]}}$，其中$\lambda$是一个依赖于$r$的常量。利用两层加权平均，我们便得到了推广问题的上界，我们需要的近似比也随之得出。

我们通过构造证明该上界是最优的。即，我们构造一个近似低秩矩阵，其由两个秩为k的矩阵的乘积加上一个非常稀疏的矩阵构成，通过给予这两个秩为k的矩阵特殊的结构使得我们算法的近似比充分接近$\frac{k}{2} + 1 + \frac{k}{2(2^k-1)}$。

布尔模型:对于布尔秩-k近似问题，我们提出了一种新算法。事实上，布尔模型更加困难。我们算法可以达到$(2^{k-1}+1)$的近似比，运行时间为$O((2^k+2)!(n^{2^k})d)$。与GF(2)情形相比，此算法更为复杂。主要困难在于布尔半环并不具有GF(2)那样好的结构。对于布尔模型，给定一个最优基矩阵$\mathbf{U}$和一个布尔秩为k的矩阵$\mathbf{B} = (\mathbf{b}_1,\ldots,\mathbf{b}_k)$ ($\mathbf{b}_i \in \{0,1\}^k$)，一般而言$\mathbf{UB}$不再是一个最优基矩阵（这里$\mathbf{UB}$是布尔积）。详细地说，随着$\mathbf{A}$大小的增加，以$\mathbf{UB}$为基矩阵的近似比可以达到任意大。这是因为$\mathbf{u}_2$甚至不可以在已知$\mathbf{u}_1$和$\mathbf{u}_1 \vee \mathbf{u}_2$（逐位或运算）的情况下近似获得。

我们的想法是利用$\mathbf{U}$的列的所有2^k种线形组合来构造k个基向量。令$(\mathbf{U},\mathbf{V})$为布尔低秩矩阵近似问题的一个最优解。令$\mathbf{b}_1,\ldots,\mathbf{b}_{2^k-1}$为$\{0,1\}^k$中全部$2^k-1$个非零向量。我们考虑$\mathbf{U}\mathbf{b}_1,\ldots,\mathbf{U}\mathbf{b}_{2^k-1}$在$\mathbf{A}$的列向量中的最近邻，分别记为$\mathbf{a}_{(\mathbf{U}\mathbf{b}_1)},\ldots,\mathbf{a}_{(\mathbf{U}\mathbf{b}_{2^k-1})}$。基矩阵的每一列都通过$\mathbf{a}_{(\mathbf{U}\mathbf{b}_1)},\ldots,\mathbf{a}_{(\mathbf{U}\mathbf{b}_{2^k-1})}$的布尔组合来构造。我们的构造有如下分解性质：令$\mathbf{b}_{(1)},\ldots,\mathbf{b}_{(2^k-1)}$为$\mathbf{b}_1,\ldots,\mathbf{b}_{2^k-1}$的一个排列，满足$n_{\mathbf{b}_{(1)}} \ge n_{\mathbf{b}_{(2)}} \ge \ldots \ge n_{\mathbf{b}_{(2^k-1)}}$，令$\mathbf{U}_{\mathbf{alg}} = (\mathbf{u}_{\mathbf{alg}_1},\ldots,\mathbf{u}_{\mathbf{alg}_k})$为我们构造的基矩阵，则对任意$i \in [2^k-1]$，$\mathbf{U}\mathbf{b}_{(i)}$与$\mathbf{U}_{\mathbf{alg}}\mathbf{b}_{(i)}$之间的距离上界可以由对所有$j \ge i$，$\mathbf{U}\mathbf{b}_{(j)}$与$\mathbf{a}_{(\mathbf{U}\mathbf{b}_{(j)})}$的距离之和给出上界。利用这条性质，我们可以给出近似比的上界。具体的构造我们将在第二章中给出。

最后我们来讨论二元低秩近似的计算复杂性。我们证明这个问题即使在$k=1$时也是NP-难的，从而改进了Tan [20]的结果，并解决了[3]中的猜想。

相关工作

据我们所知，所有已知的低秩近似问题的理论结果都是关于秩为1的特殊情形($k=1$)，此时即寻找二元向量$\mathbf{u}$, $\mathbf{v}$使得$\|\mathbf{A}-\mathbf{u}\mathbf{v}^T\|_F$最小化。

Shen等人[1]将秩为1时的问题转化为整数线性规划(ILP)。他们证明了解决这个线性规划的松弛可以得到一个2-近似。他们还利用[21]中提出的将线性规划表述成等价的最大流问题的方法提升了效率。

Jiang等人[2]注意到在秩为1的情形，从$\mathbf{A}$中简单地选取最优列即可得到2-近似。为了得到这个结论，令$\mathbf{u}^*$，$\mathbf{v}^*$为使得$\|\mathbf{A}-\mathbf{u}\mathbf{v}^T\|_F$最小化的解。令$i^*=\min_{i\in[n]}\|\mathbf{A}_i-\mathbf{u}^*\|_F$，其中$\mathbf{A}_i$ 是$\mathbf{A}$的第i列。因此，$\mathbf{A}_{i^*}$是$\mathbf{A}$的列向量中离$\mathbf{u}^*$最近的一个。于是对任意$i\in[n]$, $\|\mathbf{A}_i-\mathbf{A}_{i^*}\|_F\leq\|\mathbf{A}_i-\mathbf{u}^*\|_F+\|\mathbf{A}_{i^*}-\mathbf{u}^*\|_F\leq 2\|\mathbf{A}_i-\mathbf{u}^*\|_F$。从而以$\mathbf{A}_{i^*}$ 作为基向量$\mathbf{u}^*$. Jiang等人也将这个方法推广到了其他矩阵分解问题在$k>1$的情形，即将$\mathbf{A}$分解为$\mathbf{U}\mathbf{V}$，但要求$\mathbf{V}$的每一列包含至多一个1。这就是说，它们试图只用用$\mathbf{U}$的一列（而不是它们的线性组合）去近似$\mathbf{A}$的每一列。然而他们的结果并不能为我们的问题在$k>1$的情形提供任何理论结果。

第一章　GF(2)下的矩阵低秩近似问题

在本章中，我们考虑 GF(2)模型。我们将描述整个算法，并给出近似比的分析。

GF(2) 模型的近似算法十分简洁。以下我们将这个算法称为列选择算法。算法的流程如下：令 k 为所需要的秩。对矩阵$\mathbf{A}$的任意k列，我们将其组成一个基矩阵 $\mathbf{P}$，然后计算其对应的最优系数矩阵$\mathbf{Q}$。$\mathbf{Q}$ 的每一列可以通过枚举$\{0,1\}^k$中的 2^k 个向量得到。(计算矩阵$\mathbf{Q}$的一列被称为 最近码问题(Nearest Codeword Problem)① 最后，算法输出使得近似误差最小的$(\mathbf{P},\mathbf{Q})$作为解。容易看出算法的时间复杂度是 $\binom{n}{k}n2^kd = O(n^{k+1}d)$.

以下定理指出了列选择算法是 $\left(\frac{k}{2}+1+\frac{k}{2(2^k-1)}\right)$-近似算法。

定理 1. *对 $\forall k \geq 1$,令 OPT_k 为方程(2)最优解的误差。那么，列选择算法输出的解的误差不超过 $\left(\frac{k}{2}+1+\frac{k}{2(2^k-1)}\right)\cdot \mathrm{OPT}_k$.*

更进一步地，我们证明近似比 $\left(\frac{k}{2}+1+\frac{k}{2(2^k-1)}\right)$ 对于列选择算法不可改进。

定理 2. *对 $\forall k \geq 1$ 和 $\forall \epsilon > 0$，存在矩阵 $\mathbf{A}$ 使得列选择算法的输出的误差至少为 $\left(\frac{k}{2}+1+\frac{k}{2(2^k-1)}-\epsilon\right)\cdot \mathrm{OPT}_k$.*

如同序言部分所描述的，证明定理 1 的思路是考虑所有等价的最优基矩阵。这就是说，对任意一个给定的最优基矩阵 $\mathbf{U}$, 我们考虑所有的 $\mathbf{UB}$，其中 $\mathbf{B} = (\mathbf{b}_1,\ldots,\mathbf{b}_k)$ 是满秩矩阵，以及$\mathbf{Ub}_1,\ldots,\mathbf{Ub}_k$在$\mathbf{A}$中的最近邻矩阵 $\mathbf{P} = (\mathbf{a}_{(\mathbf{Ub})_1},\ldots,\mathbf{a}_{(\mathbf{Ub})_k})$。$\mathbf{P}$ 导出的近似误差记作 $\mathrm{Err}(\mathbf{b}_1,\ldots,\mathbf{b}_k)$。我们将要证明 $\min_{\mathbf{b}_1,\ldots,\mathbf{b}_k}\mathrm{Err}(\mathbf{b}_1,\ldots,\mathbf{b}_k)$ 不超过 $\left(\frac{k}{2}+1+\frac{k}{2(2^k-1)}\right)\mathrm{OPT}_k$，其中 OPT_k是最优解的误差。这个定理的证明方法，是在固定末尾$k-r$个 $\mathbf{b}_i$，优化r个 $(0 \leq r \leq k)$ $\mathbf{b}_i$，从而建立$k+1$个误差的加性上界。当$r = k$，加性上界和乘性上界相同，从而完成对近似比的证明。

在定理 4，我们给出加性上界。这个定理的证明是整个证明中的主要技术部分。以下我们首先给出定理 4中使用的记号。这些记号在证明中被频繁使用。为了清晰起见，我们将这些记号的含义写在如下两个表格中。

定义 3. *令 $\mathbf{A}$ 为所要近似的矩阵，$(\mathbf{U},\mathbf{V})$为问题方程(2)一组固定的最优解。对于 $\mathbf{u} \in \{0,1\}^d$, $\mathbf{c} \in \{0,1\}^k$，及 $\mathcal{X} \subset \{0,1\}^k$，约定记号如下：*

① 最近码问题(Nearest Codeword Problem) 的任意常数近似比近似算法都是NP-难的 [22].已知的关于多项式时间近似算法的结果有Berman 和 Karpinski 提出的$O(k/\log d)$-随机近似算法 [23]， Alon, Panigrahy, 及 Yekhanin [24]给出了相同近似比的近多项式时间近似算法。我们的算法中，使用简单的穷举法计算$\mathbf{Q}$的每一列。这是因为在低秩近似问题中，k 通常是一个小整数，故 2^k 的时间复杂度可以接受。

符号	含义
$\mathbf{a}_{(\mathbf{u})}$	$\mathbf{u}$ 在 $\mathbf{A}$ 的列向量中的最近邻 （如果有超过一个，则任取一个）
$\mathcal{J}_{\mathbf{c}} := \{j \in [n] : \mathbf{V}_j = \mathbf{c}\}$	$\mathbf{V}$的列向量中与 $\mathbf{c}$ 相等的列构成的集合
$n_{\mathbf{c}} := \lvert\mathcal{J}_{\mathbf{c}}\rvert$	$\mathbf{V}$ 的列向量中与 $\mathbf{c}$ 相等的列的个数
$L_{\mathbf{c}} := \sum_{j\in\mathcal{J}_{\mathbf{c}}} \lvert\mathbf{a}_j - \mathbf{U}\mathbf{c}\rvert$	在 $\mathcal{J}_{\mathbf{c}}$中的列的近似误差
$N_{\mathcal{X}} := \sum_{\mathbf{c}\in\mathcal{X}} n_{\mathbf{c}}$	$\mathbf{V}$ 的列向量中属于 $\mathcal{X}$的个数
$M_{\mathbf{c}} = \begin{cases} \frac{L_{\mathbf{c}}}{n_{\mathbf{c}}} & n_{\mathbf{c}} > 0 \\ d & n_{\mathbf{c}} = 0 \end{cases}$	$\mathcal{J}_{\mathbf{c}}$中的列向量的平均近似误差的上界

表 1.1　误差与最近邻的定义

对于所有 $1 \le r \le k$ 及线性无关的向量 $\mathbf{b}_1, \ldots, \mathbf{b}_r$ *in* $\{0,1\}^k$:

符号	含义
$\mathrm{span}^c(\mathbf{b}_1, \ldots, \mathbf{b}_r) := \{0,1\}^k \setminus \mathrm{span}(\mathbf{b}_1, \ldots, \mathbf{b}_r)$	$\mathrm{span}(\mathbf{b}_1, \ldots, \mathbf{b}_r)$ 关于全空间的补
$\mathrm{span}^{\setminus i}(\mathbf{b}_1, \ldots, \mathbf{b}_r) := \mathrm{span}(\mathbf{b}_1, \ldots, \mathbf{b}_{i-1}, \mathbf{b}_{i+1}, \ldots, \mathbf{b}_r)$	除了第 i个向量之外， 所有$\mathbf{b}_j$张成的线性子空间

表 1.2　线性子空间的定义

以下我们给出加性上界的结论。

定理 4. 令 $\mathbf{b}_1, \ldots, \mathbf{b}_k$ 为$\{0,1\}^k$ 中 k 个线性无关的向量。那么对每个 $0 \le r \le k$，

$$\mathrm{Err}^r(\mathbf{b}_1, \ldots, \mathbf{b}_{k-r}) \le \mathrm{OPT}_k + \lambda_r \cdot \sum_{\mathbf{c}\in\mathrm{span}^c(\mathbf{b}_1,\ldots,\mathbf{b}_{k-r})} L_{\mathbf{c}} + \sum_{i=1}^{k-r} f_i(\mathbf{b}_1, \ldots, \mathbf{b}_{k-r}) M_{\mathbf{b}_i}, \tag{1.1}$$

其中 $M_{\mathbf{b}_i}$ 的含义参见定义 *3*，

$$\lambda_r = \begin{cases} 0 & r = 0 \\ \frac{r}{2}\left(1 + \frac{1}{2^r - 1}\right), & 1 \le r \le k \end{cases}$$

以及

$$f_i(\mathbf{b}_1, \ldots, \mathbf{b}_{k-r}) = N_{\mathcal{X}} + \frac{1}{2} N_{\mathcal{Y}}, \tag{1.2}$$

这里 $\mathcal{X} = \mathbf{b}_i + \mathrm{span}^{\setminus i}(\mathbf{b}_1, \ldots, \mathbf{b}_{k-r})$， $\mathcal{Y} = \mathrm{span}^c(\mathbf{b}_1, \ldots, \mathbf{b}_{k-r})$.

定理 1 可以由定理 4 立即推出。

证明. （**定理 1** ）

在定理4 取$r = k$。那么方程(1.1) 右侧的最后一项此时为0。方程(1.1) 右侧的第二项变为 $\lambda_k \cdot \sum_{\mathbf{c}\in\{0,1\}^k} L_{\mathbf{c}}$。注意到 $\sum_{\mathbf{c}\in\{0,1\}^k} L_{\mathbf{c}} = \mathrm{OPT}_k$ 以及 $1 + \lambda_k = \frac{k}{2} + 1 + \frac{k}{2(2^k-1)}$，定理得证。 □

以下我们证明定理 4。首先我们需要一个简单的引理：

引理 5. *设 $a_1, \cdots, a_n$ 和 λ 为非负实数。记 $S := \sum_{i=1}^n a_i$。若 $S > \lambda a_i$ 对所有 $i \in [n]$均成立，那么*

$$\sum_{i=1}^n \frac{a_i}{S - \lambda a_i} \geq \frac{n}{n - \lambda}. \tag{1.3}$$

证明. 由 Cauchy-Schwarz 不等式，我们有 $\sum_{i=1}^n (S - \lambda a_i) \sum_{i=1}^n \frac{1}{S-\lambda a_i} \geq n^2$，注意到 $\sum_{i=1}^n (S - \lambda a_i) = (n - \lambda)S$，有 $\sum_{i=1}^n \frac{S}{S-\lambda a_i} \geq \frac{n^2}{n-\lambda}$。而由于 $\frac{S}{S-\lambda a_i} = 1 + \lambda \frac{a_i}{S-\lambda a_i}$，故进一步可得 $n + \lambda \sum_{i=1}^n \frac{a_i}{S-\lambda a_i} \geq \frac{n^2}{n-\lambda}$，引理得证。 □

同时我们指出 $M_{\mathbf{c}}$ 是将 $\mathbf{a}_{(\mathbf{Uc})}$ 近似为 $\mathbf{Uc}$所得到的误差的上界，如以下引理所述。

引理 6. *对任意 $\mathbf{c} \in \{0,1\}^k$, $|\mathbf{a}_{(\mathbf{Uc})} - \mathbf{Uc}| \leq M_{\mathbf{c}}$*

证明. 分两种情况讨论。若 $n_{\mathbf{c}} = 0$，那么由于 $(\mathbf{a}_{(\mathbf{Uc})} - \mathbf{Uc}) \in \{0,1\}^d$，引理显然成立。而若 $n_{\mathbf{c}} > 0$，因 $\mathbf{a}_{(\mathbf{Uc})}$ 是 $\mathbf{Uc}$ 在 $\mathbf{A}$的列向量中的最近邻， $L_{\mathbf{c}}$ 是 $n_{\mathbf{c}}$ 列的总误差。根据最小数不超过平均数的原理， $|\mathbf{a}_{(\mathbf{Uc})} - \mathbf{Uc}| \leq \frac{L_{\mathbf{c}}}{n_{\mathbf{c}}}$，引理得证。 □

在给出上述两个引理之后，我们可以开始定理4的证明了。

证明. （**定理4**）我们对于 r使用数学归纳法证明本定理。在整个证明中，我们固定一组二元低秩近似问题(2) 的最优解$(\mathbf{U}, \mathbf{V})$。

归纳奠基

我们首先证明不等式(1.1) 在 $r = 0$的情形。注意到此时， $\lambda_0 = 0$ 以及 $f_i(\mathbf{b}_1, \ldots, \mathbf{b}_k) = N_{\mathbf{b}_i + \mathrm{span}^{\setminus i}(\mathbf{b}_1,\ldots,\mathbf{b}_k)}$。因此，我们只需证明

$$\mathrm{Err}^{(0)}(\mathbf{b}_1, \cdots, \mathbf{b}_k) \leq \mathrm{OPT}_k + \sum_{i=1}^k M_{\mathbf{b}_i} N_{\mathbf{b}_i + \mathrm{span}^{\setminus i}(\mathbf{b}_1,\cdots,\mathbf{b}_k)}.$$

回顾 $\mathrm{Err}^{(0)}(\mathbf{b}_1, \cdots, \mathbf{b}_k) = \mathrm{Err}(\mathbf{b}_1, \cdots, \mathbf{b}_k)$，以及 $\mathrm{Err}(\mathbf{b}_1, \cdots, \mathbf{b}_k)$是以 $\mathbf{A}_{(\mathbf{UB})} = (\mathbf{a}_{(\mathbf{Ub}_1)}, \ldots, \mathbf{a}_{(\mathbf{Ub}_k)})$ 作为基矩阵的近似误差，其中 $\mathbf{a}_{(\mathbf{Ub}_i)}$ 是 $\mathbf{Ub}_i$ 在 $\mathbf{A}$的列向量中的最近邻。

总近似误差 $\mathrm{Err}(\mathbf{b}_1,\cdots,\mathbf{b}_k)$ 是所有列向量 $\mathbf{a}_j$的近似误差之和。任取 $\mathbf{b}\in\{0,1\}^k$。由于 $\mathbf{b}_1,\cdots,\mathbf{b}_k$ 在 $\mathrm{GF}(2)^k$中线性无关，因此它们是$\mathrm{GF}(2)^k$的一组基，因而$\mathbf{b}$ 可以表为 $\mathbf{b}_1,\cdots,\mathbf{b}_k$的线性组合。从而我们可设 $\mathbf{b}=\sum_{i=1}^k x_i\mathbf{b}_i$， $x_1,\ldots,x_k\in\{0,1\}$。(这里的求和号 $\sum$ 使用的是 GF(2)上的加法。) 记 $I_{\mathbf{b}}=\{i:x_i=1\}$ 。列向量 $\mathbf{a}_j$的近似误差可以表为 $\min_{\mathbf{b}}|\mathbf{a}_j-\sum_{i\in I_{\mathbf{b}}}\mathbf{a}_{(\mathbf{U}\mathbf{b}_i)}|$ 。回顾 $\mathcal{J}_{\mathbf{c}}$的定义： $\mathbf{V}$ 的列向量中与 $\mathbf{c}$相等的下标集合。我们通过划分 $[n]=\cup_{\mathbf{c}\in\{0,1\}^k}\mathcal{J}_{\mathbf{c}}$ 来分解总近似误差。首先，我们有：

$$\begin{aligned}\sum_{j\in\mathcal{J}_{\mathbf{c}}}\left|\mathbf{a}_j-\sum_{i\in I_{\mathbf{c}}}\mathbf{a}_{(\mathbf{U}\mathbf{b}_i)}\right| &\le \sum_{j\in\mathcal{J}_{\mathbf{c}}}\left(\left|\mathbf{a}_j-\sum_{i\in I_{\mathbf{c}}}\mathbf{U}\mathbf{b}_i\right|+\sum_{i\in I_{\mathbf{c}}}|\mathbf{a}_{(\mathbf{U}\mathbf{b}_i)}-\mathbf{U}\mathbf{b}_i|\right)\\ &\le L_{\mathbf{c}}+n_{\mathbf{c}}\sum_{i\in I_{\mathbf{c}}}M_{\mathbf{b}_i},\end{aligned}\tag{1.4}$$

其中，最后一个不等号可以由 $L_{\mathbf{c}},n_{\mathbf{c}}$的定义和引理 6得到。由于 $L_{\mathbf{c}}$ 是最优解 $(\mathbf{U},\mathbf{V})$ 在所有 $\mathcal{J}_{\mathbf{c}}$中的列向量的近似误差。方程(1.4) 导出了如下的加性误差上界：

$$\begin{aligned}\mathrm{Err}^{(0)}(\mathbf{b}_1,\cdots,\mathbf{b}_k) &= \sum_{j=1}^{n}\min_{\mathbf{b}\in\{0,1\}^k}\left|\mathbf{a}_j-\sum_{i\in I_{\mathbf{b}}}\mathbf{a}_{(\mathbf{U}\mathbf{b}_i)}\right|\\ &= \sum_{\mathbf{c}\in\{0,1\}^k}\sum_{j\in\mathcal{J}_{\mathbf{c}}}\min_{\mathbf{b}\in\{0,1\}^k}\left|\mathbf{a}_j-\sum_{i\in I_{\mathbf{b}}}\mathbf{a}_{(\mathbf{U}\mathbf{b}_i)}\right|\\ &\le \sum_{\mathbf{c}\in\{0,1\}^k}\sum_{j\in\mathcal{J}_{\mathbf{c}}}\left|\mathbf{a}_j-\sum_{i\in I_{\mathbf{c}}}\mathbf{a}_{(\mathbf{U}\mathbf{b}_i)}\right|\\ &\le \sum_{\mathbf{c}\in\{0,1\}^k}\left(L_{\mathbf{c}}+n_{\mathbf{c}}\sum_{i\in I_{\mathbf{c}}}M_{\mathbf{b}_i}\right)\\ &= \mathrm{OPT}_k+\sum_{\mathbf{c}\in\{0,1\}^k}\sum_{i\in I_{\mathbf{c}}}n_{\mathbf{c}}M_{\mathbf{b}_i},\end{aligned}\tag{1.5}$$

其中最后一个不等号用到了 $\mathrm{OPT}_k=\sum_{\mathbf{c}}L_{\mathbf{c}}$。考虑方程(1.5)的第二项。我们有：

$$\begin{aligned}\sum_{\mathbf{c}\in\{0,1\}^k}\sum_{i\in I_{\mathbf{c}}}n_{\mathbf{c}}M_{\mathbf{b}_i} &= \sum_{i=1}^{k}\sum_{\mathbf{c}\in\{0,1\}^k}n_{\mathbf{c}}M_{\mathbf{b}_i}I[i\in I_{\mathbf{c}}]\\ &= \sum_{i=1}^{k}M_{\mathbf{b}_i}\left(\sum_{\mathbf{c}\in\mathbf{b}_i+\mathrm{span}^{\setminus i}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r})}n_{\mathbf{c}}\right)\\ &= \sum_{i=1}^{k}M_{\mathbf{b}_i}N_{\mathbf{b}_i+\mathrm{span}^{\setminus i}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r})},\end{aligned}\tag{1.6}$$

其中最后一个不等号用到了 $N_{\mathcal{X}}$的定义（见定义 3）。联立方程(1.5) 和方程(1.6) 就完成了 $r=0$情况的证明。

归纳过渡

假设方程(1.1) 对所有 $r' \le r$均成立，我们接下来证明

$$\mathrm{Err}^{(r+1)}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) \le \mathrm{OPT}_k + \lambda_{r+1} \sum_{\mathbf{c}\in \mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{c}} + \sum_{i=1}^{k-r-1} f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) M_{\mathbf{b}_i}. \tag{1.7}$$

由于 $\mathrm{Err}^{(r+1)}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) = \min_{\mathbf{b}} \mathrm{Err}^{(r)}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b})$，我们有对所有加权平均 $w_{\mathbf{b}}$，其中 $w_{\mathbf{b}} \ge 0$， $\sum_{\mathbf{b}\in\{0,1\}^k} w_{\mathbf{b}} = 1$，均有

$$\mathrm{Err}^{(r+1)}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) \le \sum_{\mathbf{b}\in\{0,1\}^k} w_{\mathbf{b}} \mathrm{Err}^{(r)}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b}). \tag{1.8}$$

我们将分两个层次先后进行加权平均。考虑商空间 $\mathrm{GF}(2)^k/\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})$ 以及其对应的陪集。将所有 2^{r+1} 个陪集记作 $\mathbf{p}_0+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}),\cdots,\mathbf{p}_{2^{r+1}-1}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})$。不失一般性，可设 $\mathbf{p}_0 \in \mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})$。第一层加权平均将在每个陪集$\mathbf{p}_i+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})$中进行，我们将会最终得到一个 $\mathrm{Err}^{(r+1)}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})$ 的上界，这个上界依赖于 $\mathbf{p}_i$。第二层加权平均则在所有陪集上进行，并最终得到想要的上界。两层加权平均使用不同的方式来选取权重。

第一层加权平均 (在陪集内部):

对于 $\mathbf{p} \in \{\mathbf{p}_1,\ldots,\mathbf{p}_{2^{r+1}-1}\}$，定义 $Z(\mathbf{p}) := N_{\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})}$。那么 $Z(\mathbf{p})$ 是 $\mathbf{V}$ 的列向量中属于 $\mathbf{p}$所对应陪集的数量。我们只考虑使得$Z(\mathbf{p}) > 0$的那些$\mathbf{p}$。我们按照以下方式选取权值：假如$\mathbf{b} \in \mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})$，则取 $w_{\mathbf{b}} = n_{\mathbf{b}}/Z(\mathbf{p})$，否则取为0。对于每个 $\mathbf{b} \in \mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})$，我们有 $\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b}) = \mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{p})$。结

合归纳假设可得：

$$\begin{aligned}
& \mathrm{Err}^{(r+1)}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) \\
\leq\ & \frac{1}{Z(\mathbf{p})}\sum_{\mathbf{b}\in\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} n_{\mathbf{b}}\mathrm{Err}^{(r)}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b}) \\
\leq\ & \mathrm{OPT}_k+\lambda_r\sum_{\mathbf{c}\in\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{p})} L_{\mathbf{c}}+ \\
& \frac{1}{Z(\mathbf{p})}\sum_{\mathbf{b}\in\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} n_{\mathbf{b}}[f_{k-r}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b})M_{\mathbf{b}}+ \\
& \sum_{i=1}^{k-r-1} f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b})M_{\mathbf{b}_i}]
\end{aligned} \tag{1.9}$$

我们的下一个目标是分别给出方程(1.9)最后一行中的两项的上界估计。

方程(1.9)最后一行中第一项的上界估计

对于第一项，我们以下证明：

$$\begin{aligned}
& \frac{1}{Z(\mathbf{p})}\sum_{\mathbf{b}\in\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} n_{\mathbf{b}}f_{k-r}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b})M_{\mathbf{b}} \\
=\ & \sum_{\mathbf{b}\in\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{b}}\cdot\left(1+\frac{1}{2Z(\mathbf{p})}N_{\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{p})}\right).
\end{aligned} \tag{1.10}$$

由$M_{\mathbf{b}}$的定义，容易看出 $n_{\mathbf{b}}M_{\mathbf{b}} = L_{\mathbf{b}}$。因此，我们只需分析 $f_{k-r}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b})$即可。事实上，对所有 $\mathbf{b}\in\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})$，我们有

$$\mathrm{span}^{\backslash(k-r)}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b}) = \mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})$$

因此，

$$\mathbf{b}+\mathrm{span}^{\backslash(k-r)}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b}) = \mathbf{b}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) = \mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})$$

另一方面，

$$\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b}) = \mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{p})$$

从而对任意 $\mathbf{b} \in \mathbf{p} + \mathrm{span}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1})$ 我们均有

$$\begin{aligned} f_{k-r}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{b}) &= N_{\mathbf{b}+\mathrm{span}^{\backslash(k-r)}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{b})} + \frac{1}{2} N_{\mathrm{span}^c(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{b})} \\ &= N_{\mathbf{p}+\mathrm{span}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1})} + \frac{1}{2} N_{\mathrm{span}^c(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{p})} \end{aligned}$$

综上所述，

$$\begin{aligned} & \frac{1}{Z(\mathbf{p})} \sum_{\mathbf{b} \in \mathbf{p}+\mathrm{span}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1})} n_{\mathbf{b}} f_{k-r}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{b}) M_{\mathbf{b}} \\ = & \sum_{\mathbf{b} \in \mathbf{p}+\mathrm{span}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1})} L_{\mathbf{b}} \cdot \frac{1}{Z(\mathbf{p})} \cdot \left(N_{\mathbf{p}+\mathrm{span}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1})} + \frac{1}{2} N_{\mathrm{span}^c(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{p})} \right) \\ = & \sum_{\mathbf{b} \in \mathbf{p}+\mathrm{span}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1})} L_{\mathbf{b}} \cdot \left(1 + \frac{1}{2Z(\mathbf{p})} N_{\mathrm{span}^c(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{p})} \right) \end{aligned}$$

于是方程(1.10)得证。

方程(1.9)最后一行中第二项的上界估计

对于第二项，我们以下证明：

$$\begin{aligned} & \frac{1}{Z(\mathbf{p})} \sum_{\mathbf{b} \in \mathbf{p}+\mathrm{span}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1})} n_{\mathbf{b}} \sum_{i=1}^{k-r-1} f_i(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{b}) M_{\mathbf{b}_i} \\ = & \frac{1}{Z(\mathbf{p})} \sum_{i=1}^{k-r-1} M_{\mathbf{b}_i} \sum_{\mathbf{b} \in \mathbf{p}+\mathrm{span}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1})} n_{\beta} f_i(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{b}) \\ \leq & \sum_{i=1}^{k-r-1} M_{\mathbf{b}_i} f_i(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}) \end{aligned} \tag{1.11}$$

考虑 $f_i(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{b})$。注意到对任意 $1 \leq i \leq k-r-1$ 和 $\mathbf{b} \in \mathbf{p} + \mathrm{span}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1})$，我们有

$$\begin{aligned} f_i(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{b}) &= N_{\mathbf{b}_i+\mathrm{span}^{\backslash i}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{b})} + \frac{1}{2} N_{\mathrm{span}^c(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{b})} \\ &= N_{\mathbf{b}_i+\mathrm{span}^{\backslash i}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{b})} + \frac{1}{2} N_{\mathrm{span}^c(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{p})}. \end{aligned}$$

因此

$$\begin{aligned} \mathbf{b}_i + \mathrm{span}^{\backslash i}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{b}) &= [\mathbf{b}_i + \mathrm{span}^{\backslash i}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1})] \cup \\ & \quad [\mathbf{b}_i + \mathbf{b} + \mathrm{span}^{\backslash i}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1})] \end{aligned} \tag{1.12}$$

我们有

$$
\begin{aligned}
f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b}) \quad = \quad & N_{\mathbf{b}_i+\mathbf{b}+\mathrm{span}^{\setminus i}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})}+N_{\mathbf{b}_i+\mathrm{span}^{\setminus i}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})}+ \\
& \frac{1}{2}N_{\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{p})}. \qquad (1.13)
\end{aligned}
$$

记

$$
\begin{aligned}
X_1(i,\mathbf{p}) &= \mathbf{p}+\mathrm{span}^{\setminus i}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) \\
X_2(i,\mathbf{p}) &= \mathbf{p}+\mathbf{b}_i+\mathrm{span}^{\setminus i}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) \\
X_3(i,\mathbf{p}) &= \mathbf{b}_i+\mathrm{span}^{\setminus i}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) \\
X_4(i,\mathbf{p}) &= \mathrm{span}^{c}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{p}) \qquad (1.14)
\end{aligned}
$$

容易看出 $\forall j_1,j_2\in\{1,2,3,4\},j_1\neq j_2$均有 $X_{j_1}(i,\mathbf{p})\cap X_{j_2}(i,\mathbf{p})=\varnothing$ 以及 $X_1(i,\mathbf{p})\cup X_2(i,\mathbf{p})=\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})$.
以下分两种情况讨论：假如 $\mathbf{b}\in X_1(i,\mathbf{p})$，那么 $\mathbf{b}_i+\mathbf{b}+\mathrm{span}^{\setminus i}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})=\mathbf{b}_i+\mathbf{p}+\mathrm{span}^{\setminus i}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})$。综合方程(1.13) 和方程(1.14) 可得

$$
f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b})=N_{X_2(i,\mathbf{p})}+N_{X_3(i,\mathbf{p})}+\frac{1}{2}N_{X_4(i,\mathbf{p})} \qquad (1.15)
$$

假如 $\mathbf{b}\in X_2(i,\mathbf{p})$，那么 $\mathbf{b}_i+\mathbf{b}+\mathrm{span}^{\setminus i}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})=\mathbf{p}+\mathrm{span}^{\setminus i}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})$. 综合方程(1.13)和方程(1.14) 可得

$$
f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b})=N_{X_1(i,\mathbf{p})}+N_{X_3(i,\mathbf{p})}+\frac{1}{2}N_{X_4(i,\mathbf{p})} \qquad (1.16)
$$

以下我们给出方程(1.11)中内层求和的上界。

$$
\begin{aligned}
& \sum_{\mathbf{b}\in\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} n_{\mathbf{b}} f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b}) \\
= & \sum_{\mathbf{b}\in X_1(i,\mathbf{p})\cup X_2(i,\mathbf{p})} n_{\mathbf{b}} f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b}) \\
= & \sum_{\mathbf{b}\in X_1(i,\mathbf{p})} n_{\mathbf{b}} f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b}) + \sum_{\mathbf{b}\in X_2(i,\mathbf{p})} n_{\mathbf{b}} f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b}) \\
= & \sum_{\mathbf{b}\in X_1(i,\mathbf{p})} n_{\mathbf{b}} \left(N_{X_2(i,\mathbf{p})} + N_{X_3(i,\mathbf{p})} + \frac{1}{2}N_{X_4(i,\mathbf{p})}\right) \\
& + \sum_{\mathbf{b}\in X_2(i,\mathbf{p})} n_{\mathbf{b}} \left(N_{X_1(i,\mathbf{p})} + N_{X_3(i,\mathbf{p})} + \frac{1}{2}N_{X_4(i,\mathbf{p})}\right) \\
= & N_{X_1(i,\mathbf{p})}\left(N_{X_2(i,\mathbf{p})} + N_{X_3(i,\mathbf{p})} + \frac{1}{2}N_{X_4(i,\mathbf{p})}\right) \\
& + N_{X_2(i,\mathbf{p})}\left(N_{X_1(i,\mathbf{p})} + N_{X_3(i,\mathbf{p})} + \frac{1}{2}N_{X_4(i,\mathbf{p})}\right) \\
= & 2N_{X_1(i,\mathbf{p})}N_{X_2(i,\mathbf{p})} + (N_{X_1(i,\mathbf{p})} + N_{X_2(i,\mathbf{p})})\left(N_{X_3(i,\mathbf{p})} + \frac{1}{2}N_{X_4(i,\mathbf{p})}\right) \\
\le & \frac{1}{2}(N_{X_1(i,\mathbf{p})} + N_{X_2(i,\mathbf{p})})^2 + (N_{X_1(i,\mathbf{p})} + N_{X_2(i,\mathbf{p})})\left(N_{X_3(i,\mathbf{p})} + \frac{1}{2}N_{X_4(i,\mathbf{p})}\right) \\
= & (N_{X_1(i,\mathbf{p})} + N_{X_2(i,\mathbf{p})})\left[\frac{1}{2}(N_{X_1(i,\mathbf{p})} + N_{X_2(i,\mathbf{p})} + N_{X_4(i,\mathbf{p})}) + N_{X_3(i,\mathbf{p})}\right],
\end{aligned} \tag{1.17}
$$

其中，第三个等号是因为方程(1.15) 和方程(1.16)。由方程(1.14) 和方程(1.15)，我们有：

$$
Z(\mathbf{p}) = N_{\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} = N_{X_1(i,\mathbf{p})} + N_{X_2(i,\mathbf{p})}, \tag{1.18}
$$

以及

$$
X_1(i,\mathbf{p}) \cup X_2(i,\mathbf{p}) \cup X_4(i,\mathbf{p}) = \mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}). \tag{1.19}
$$

因此，

$$
N_{X_1(i,\mathbf{p})} + N_{X_2(i,\mathbf{p})} + N_{X_4(i,\mathbf{p})} = N_{\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})}. \tag{1.20}
$$

从而由方程(1.14), 方程(1.18), 方程(1.20) 及$f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})$的定义，我们有

$$
\begin{aligned}
& (N_{X_1(i,\mathbf{p})} + N_{X_2(i,\mathbf{p})})\left[\frac{1}{2}(N_{X_1(i,\mathbf{p})} + N_{X_2(i,\mathbf{p})} + N_{X_4(i,\mathbf{p})}) + N_{X_3(i,\mathbf{p})}\right] \\
= & Z(\mathbf{p})\left(\frac{1}{2}N_{\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} + N_{\mathbf{b}_i+\mathrm{span}^{\setminus i}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})}\right) \\
= & Z(\mathbf{p}) f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}).
\end{aligned} \tag{1.21}
$$

将方程(1.21) 代入方程(1.17)可得

$$\sum_{\mathbf{b}\in\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} n_{\mathbf{b}} f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b}) \le Z(\mathbf{p}) f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}).$$

现在我们可以证明方程(1.11)了：

$$\begin{aligned}
&\frac{1}{Z(\mathbf{p})}\sum_{\mathbf{b}\in\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} n_{\mathbf{b}} \sum_{i=1}^{k-r-1} f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b}) M_{\mathbf{b}_i} \\
= &\frac{1}{Z(\mathbf{p})}\sum_{i=1}^{k-r-1} M_{\mathbf{b}_i} \sum_{\mathbf{b}\in\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} n_{\mathbf{b}} f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{b}) \\
\le &\frac{1}{Z(\mathbf{p})}\sum_{i=1}^{k-r-1} M_{\mathbf{b}_i} Z(\mathbf{p}) f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) \\
= &\sum_{i=1}^{k-r-1} M_{\mathbf{b}_i} f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}).
\end{aligned} \tag{1.22}$$

这就给出了方程(1.9)中第二项的上界。

最后，结合方程(1.9)，方程(1.10) 和方程(1.11)，我们有对任意满足$Z(\mathbf{p}) > 0$的$\mathbf{p}\in\{\mathbf{p}_1,\cdots,\mathbf{p}_{2^{r+1}-1}\}$，均有：

$$\begin{aligned}
&\mathrm{Err}^{(r+1)}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) \\
\le\ &\mathrm{OPT}_k + \lambda_r \sum_{\mathbf{c}\in\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{p})} L_{\mathbf{c}} + \sum_{i=1}^{k-r-1} f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) M_{\mathbf{b}_i} \\
&+ \left(1+\frac{1}{2Z(p)} N_{\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{p})}\right) \sum_{\mathbf{b}\in\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{b}}.
\end{aligned} \tag{1.23}$$

注意到 $\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) = \mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{p}) \cup [\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})]$，我们有：

$$\sum_{\mathbf{c}\in\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{p})} L_{\mathbf{c}} = \sum_{\mathbf{c}\in\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{c}} - \sum_{\mathbf{b}\in\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{b}}. \tag{1.24}$$

将方程(1.24) 代入方程(1.23)，可得

$$\begin{aligned}
&\mathrm{Err}^{(r+1)}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) \\
\le\ &\mathrm{OPT}_k + \lambda_r \sum_{\mathbf{c}\in\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{c}} + \sum_{i=1}^{k-r-1} f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) M_{\mathbf{b}_i} \\
&+ \left(1+\frac{1}{2Z(\mathbf{p})} N_{\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{p})} - \lambda_r\right) \sum_{\mathbf{b}\in p+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{b}}.
\end{aligned} \tag{1.25}$$

于是我们完成了第一层的加权平均。方程(1.25) 就是我们通过在 $\mathbf{p}$ 所对应的陪集中进行加权平均而得到的上界。这个上界将在第二层加权平均中继续被用到。

在开始第二层的加权平均之前，我们首先指出当 $r = 0$ 时，归纳过渡 $r \to r+1$已经完成了。因此在这个特例下，我们不再需要进行第二轮的加权平均。为了看出这一点，我们只需注意到如下结论：当 $r = 0$时， $\mathrm{span}^c(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{p}) = \emptyset$ and $\mathbf{p} + \mathrm{span}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}) = \mathrm{span}^c(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1})$。将这两个等式代入方程(1.23) 即得：

$$\begin{aligned}\mathrm{Err}^{(1)}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-1}) &= \mathrm{OPT}_k + \sum_{i=1}^{k-1} f_i(\mathbf{b}_1, \cdots, \mathbf{b}_{k-1}) M_{\mathbf{b}_i} + \sum_{\mathbf{b}\in\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-1})} L_{\mathbf{b}} \\ &= \mathrm{OPT}_k + \sum_{i=1}^{k-1} f_i(\mathbf{b}_1, \cdots, \mathbf{b}_{k-1}) M_{\mathbf{b}_i} + \lambda_1 \sum_{\mathbf{b}\in\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-1})} L_{\mathbf{b}},\end{aligned}$$

最后一步用到了 $\lambda_{r+1} = \lambda_1 = 1$。因此当 $r = 0$ 时，归纳过渡已经完成。于是在接下来的证明中，我们可以假设 $r \geq 1$ 。

以下我们进行第二层的加权平均。

第二层加权平均（在陪集之间）

在第二层的加权平均中，我们将在所有非平凡的陪集之间进行加权平均。称一个陪集 $\mathbf{p}_i + \mathrm{span}(\mathbf{b}_1, \ldots, \mathbf{b}_{k-r-1})$ 是非平凡的，当且仅当 $Z(\mathbf{p}_i) > 0$ 且 $i \neq 0$ (即这个陪集不同于 $\mathrm{span}(\mathbf{b}_1, \ldots, \mathbf{b}_{k-r-1})$)。在方程(1.25)中我们已经通过在每个非平凡的陪集内部做加权平均，而得到了 Err^{r+1}的一个上界。在这一部分中，我们将通过在所有非平凡的陪集之间对方程(1.25)做加权平均而证明原定理。注意在方程(1.25)的上界中，只有最后一项与陪集有关，即无论陪集怎么选取，其余项都是一样的。因此，我们只需关注这一项即可。为了记号的清晰起见，我们把这一项记作 H:

$$H := \left(1 + \frac{1}{2Z(\mathbf{p})} N_{\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{p})} - \lambda_r\right) \sum_{\mathbf{b}\in\mathbf{p}+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{b}}. \tag{1.26}$$

注意到

$$[\mathbf{p}_i + \mathrm{span}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1})] \cup \mathrm{span}^c(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}, \mathbf{p}_i) = \mathrm{span}^c(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}),$$

以及

$$\bigcup_{i=1}^{2^{r+1}-1} \mathbf{p}_i + \mathrm{span}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}) = \mathrm{span}^c(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}).$$

我们有

$$N_{\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{p}_i)} = N_{\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} - N_{\mathbf{p}_i+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} = \sum_{j=1}^{2^{r+1}-1} Z(\mathbf{p}_j) - Z(\mathbf{p}_i).$$

因此对于 $\mathbf{p}_i$ 所对应的非平凡陪集，我们有：

$$1 + \frac{1}{2Z(\mathbf{p}_i)} N_{\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1},\mathbf{p}_i)} - \lambda_r = \frac{1}{2Z(\mathbf{p}_i)} \left[\sum_{j=1}^{2^{r+1}-1} Z(\mathbf{p}_j) - (2\lambda_r - 1) Z(\mathbf{p}_i) \right] \tag{1.27}$$

由方程(1.26)和方程(1.27)可得对任意满足$Z(\mathbf{p}_i) > 0$的$i \in [2^{r+1}-1]$，我们均有

$$H = \frac{1}{2Z(\mathbf{p}_i)} \left[\sum_{j=1}^{2^{r+1}-1} Z(\mathbf{p}_j) - (2\lambda_r - 1) Z(\mathbf{p}_i) \right] \sum_{\mathbf{b}\in\mathbf{p}_i+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{b}} \tag{1.28}$$

在开始第二层的加权平均之前，我们首先需要对一种特殊情况作单独处理。考虑如下情形：存在一个$\mathbf{p}_i$对应的陪集，使得$\sum_{j=1}^{2^{r+1}-1} Z(\mathbf{p}_j) - (2\lambda_r - 1) Z(\mathbf{p}_i) \le 0$。在这种情况下，由方程(1.28)可以看出$H \le 0$。结合方程(1.25) 以及 $\lambda_r \le \lambda_{r+1}$的事实，我们有

$$\begin{aligned}
&\mathrm{Err}^{(r+1)}(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}) \\
\le\ & \mathrm{OPT}_k + \lambda_r \sum_{\mathbf{c}\in\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{c}} + \sum_{i=1}^{k-r-1} f_i(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}) M_{\mathbf{b}_i} \\
\le\ & \mathrm{OPT}_k + \lambda_{r+1} \sum_{\mathbf{c}\in\mathrm{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{c}} + \sum_{i=1}^{k-r-1} f_i(\mathbf{b}_1, \cdots, \mathbf{b}_{k-r-1}) M_{\mathbf{b}_i}
\end{aligned}$$

从而在这种特殊情况下，定理已经成立。

以下我们可以在假设所有非平凡陪集均满足

$$\sum_{j=1}^{2^{r+1}-1} Z(\mathbf{p}_j) - (2\lambda_r - 1) Z(\mathbf{p}_i) > 0.$$

的条件下，进行第二轮加权平均。我们按照如下方式选取权值。首先，我们只给非平凡的陪集分配正的权值。对于每个 $\mathbf{p}_i$ 所对应的陪集，其权重 w_i 取为

$$w_i = \frac{2Z(\mathbf{p}_i)}{\sum_{j=1}^{2^{r+1}-1} Z(\mathbf{p}_j) - (2\lambda_r - 1) Z(\mathbf{p}_i)}.$$

由方程(1.28) 我们有：

$$w_i H = \sum_{\mathbf{b}\in\mathbf{p}_i+\mathrm{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{b}}.$$

因此，

$$\sum_{i=1}^{2^{r+1}-1} w_i H = \sum_{i=1}^{2^{r+1}-1} \sum_{\mathbf{b}\in\mathbf{p}_i+\text{span}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{b}} = \sum_{\mathbf{b}\in\text{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{b}}. \tag{1.29}$$

另一方面，由引理5可得

$$\begin{aligned}\sum_{i=1}^{2^{r+1}-1} w_i &= 2\sum_{i=1}^{2^{r+1}-1} \frac{Z(\mathbf{p}_i)}{\sum_{j=1}^{2^{r+1}-1} Z(\mathbf{p}_j) - (2\lambda_r - 1)Z(\mathbf{p}_i)} \\ &\geq \frac{2(2^{r+1}-1)}{2^{r+1}-1-(2\lambda_r-1)} \\ &= \frac{2^{r+1}-1}{2^r-\lambda_r}\end{aligned} \tag{1.30}$$

注意到 $(2^r-1)\lambda_r = r2^{r-1}$，因此，

$$(2^{r+1}-1)\lambda_{r+1} - 2(2^r-1)\lambda_r = (r+1)2^r - r2^r = 2^r,$$

进而

$$(2^{r+1}-1)\lambda_{r+1} - (2^{r+1}-1)\lambda_r = 2^r - \lambda_r. \tag{1.31}$$

结合方程(1.31)和方程(1.30)可得

$$\sum_{i=1}^{2^{r+1}-1} w_i \geq \frac{1}{\lambda_{r+1}-\lambda_r}. \tag{1.32}$$

由方程(1.29)和方程(1.32)有：

$$H \leq (\lambda_{r+1}-\lambda_r) \sum_{\mathbf{b}\in\text{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{b}}. \tag{1.33}$$

将方程(1.33)与方程(1.25)及方程(1.26)中H的定义，我们最终得到：

$$\text{Err}^{(r+1)}(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) \leq \text{OPT}_k + \lambda_{r+1} \sum_{\mathbf{c}\in\text{span}^c(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1})} L_{\mathbf{c}} + \sum_{i=1}^{k-r-1} f_i(\mathbf{b}_1,\cdots,\mathbf{b}_{k-r-1}) M_{\mathbf{b}_i}.$$

这样，我们就完成了归纳过渡，并最终完整地证明了定理。

□

证明. （**定理 2**）

我们给出一个构造性的证明。对任意k及$\epsilon > 0$，我们将构造一个矩阵$\mathbf{A}$，当它是算法的输入时，列选择算法的误差至少是 $\left(\frac{k}{2}+1+\frac{k}{2(2^k-1)}-\epsilon\right)\cdot \text{OPT}_k$ 为了简明起见，我们所构造的矩阵是方阵，即 $d = n$。这个矩阵是一个近似低秩的矩阵，即$\mathbf{A}$是两个

秩-k 矩阵的乘积加上一个稀疏的噪声矩阵。

我们假设 **A** 的大小满足 $k|n$ 及 $(2^k-1)|n$。设 $p := n/k$ ，$q := n/(2^k-1)$ 。**A** 的构造如下：

$$\mathbf{A} := \mathbf{LR} + \mathbf{I}_n, \tag{1.34}$$

其中，$\mathbf{I}_n$ 是单位矩阵，**L** 是一个如下定义的$n \times k$矩阵：:

$$\mathbf{L} := \begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \ldots & \mathbf{c}_k \end{pmatrix},$$

这里，$\mathbf{c}_i = \begin{pmatrix} \underbrace{0\ldots0}_{(i-1)p}\underbrace{1\ldots1}_{p}\underbrace{0\ldots0}_{(k-i)p} \end{pmatrix}^T$ 。对每个 $\mathbf{c}_i$ ，只有p 个非零元素。**R** 是一个如下定义的 $k \times n$ 矩阵：

$$\mathbf{R} := \begin{pmatrix} \mathbf{b}_1 \otimes \mathbf{1}_q & \mathbf{b}_2 \otimes \mathbf{1}_q & \ldots & \mathbf{b}_{2^k-1} \otimes \mathbf{1}_q \end{pmatrix},$$

这里，$\mathbf{1}_q = (11\ldots1)^T$ 是长度为q的全1向量；$\otimes$ 是矩阵的Kronecker 乘积；而k维列向量 $\mathbf{b}_i$是i的二进制表示，例如 $\mathbf{b}_1 = (0\ldots001)^T, \mathbf{b}_2 = (0\ldots010)^T, \mathbf{b}_3 = (0\ldots011)^T$，等等。以下是$A$的可视化形式：

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}_{n\times k} \begin{pmatrix} 0 & & 0 & 0 & & 0 & & 1 & & 1 \\ \vdots & \ldots & \vdots & \vdots & \ldots & \vdots & \ldots & \vdots & \ldots & \vdots \\ 0 & & 0 & 1 & & 1 & & 1 & & 1 \\ 1 & & 1 & 0 & & 0 & & 1 & & 1 \end{pmatrix}_{k\times n} + \mathbf{I}_n$$

我们将用到矩阵**L**和**R**的如下性质：

1) **L** 中不同列的1互不重叠。**L** 的列向量的任何非零线性组合至少包含q个1。

2) **R** 的列向量包含所有的非零k维向量，并重复q次。

为了证明结论，我们将说明无论基矩阵中**A**的k列怎样选取，其导出的近似误差在n充分大时，至少是我们想要的下界。

我们将分两种情况讨论。在第一种情况中，我们假设基矩阵中选取的k个$\mathbf{A}$的列向量具有如下性质：它们所对应的 $\mathbf{A}-\mathbf{I}_n$中的k列线性无关。反过来，在第二种情况中，我们假设它们线性相关。

以下我们考虑第一种情况。令 $(\mathbf{U},\mathbf{V})$为列选择算法在输入矩阵$\mathbf{A}$的输出。$\mathbf{U}$由$\mathbf{A}$的k列组成。记这k列的下标为$i_1,\ldots,i_k$。设$\mathbf{U}^-$ 为$\mathbf{A}-\mathbf{I}_n$的第$i_1,\ldots,i_k$列所组成的矩阵。我们考虑$\mathbf{A}-\mathbf{I}_n$的秩k 近似。 $\mathbf{U}^-$ 一定是最优基矩阵，因为$\mathbf{A}-\mathbf{I}_n=\mathbf{LR}$ 的秩是 k 而根据我们的假设，$\mathbf{U}^-$ 包含了 k 个线性无关的向量，它们中的每一个都是$\mathbf{L}$中列向量的线性组合。因此，存在一个 $k\times n$ 矩阵 $\mathbf{V}^-$ 使得$\mathbf{A}-\mathbf{I}_n=\mathbf{U}^-\mathbf{V}^-$。

我们的下一个目标是证明当n充分大时，$\mathbf{V}=\mathbf{V}^-$。考虑$\mathbf{A}$ 用$\mathbf{UV}$近似之后，在每列上的误差。令 $\mathbf{A}=(\mathbf{a}_1,\ldots,\mathbf{a}_n)$, $\mathbf{V}=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$，以及$\mathbf{V}^-=(\mathbf{v}_1^-,\ldots,\mathbf{v}_n^-)$。为了简明起见，我们用 $|\cdot|$ 替代 $\|\|_F$。对于第i列，一方面我们有：

$$
\begin{aligned}
|\mathbf{U}\mathbf{v}_i-\mathbf{a}_i| &= |\mathbf{U}\mathbf{v}_i-\mathbf{U}\mathbf{v}_i^-+\mathbf{U}\mathbf{v}_i^--\mathbf{U}^-\mathbf{v}_i^-+\mathbf{U}^-\mathbf{v}_i^--\mathbf{a}_i| \\
&\geqslant |\mathbf{U}(\mathbf{v}_i-\mathbf{v}_i^-)|-|(\mathbf{U}-\mathbf{U}^-)\mathbf{v}_i^-|-|\mathbf{U}^-\mathbf{v}_i^--\mathbf{a}_i|. \qquad (1.35)
\end{aligned}
$$

而另一方面，

$$
\begin{aligned}
|\mathbf{U}\mathbf{v}_i-\mathbf{a}_i| &= |\mathbf{U}\mathbf{v}_i-\mathbf{U}\mathbf{v}_i^-+\mathbf{U}\mathbf{v}_i^--\mathbf{U}^-\mathbf{v}_i^-+\mathbf{U}^-\mathbf{v}_i^--\mathbf{a}_i| \\
&\leq |\mathbf{U}(\mathbf{v}_i-\mathbf{v}_i^-)|+|(\mathbf{U}-\mathbf{U}^-)\mathbf{v}_i^-|+|\mathbf{U}^-\mathbf{v}_i^--\mathbf{a}_i|. \qquad (1.36)
\end{aligned}
$$

以下我们分别对方程(1.35)和方程(1.36)右边的三项进行分析。假如$\mathbf{V}\neq\mathbf{V}^-$，那么一定存在某一列$i$使得$\mathbf{v}_i\neq\mathbf{v}_i^-$。因此，第一项$\mathbf{U}(\mathbf{v}_i-\mathbf{v}_i^-)$ 是$\mathbf{U}$的列向量的一组非零线性组合。由于$\mathbf{U}$包含了$\mathbf{LR}+\mathbf{I}_n$的k列，根据$\mathbf{L}$ 和$\mathbf{R}$的构造，不难看出$\mathbf{v}_i\neq\mathbf{v}_i^-, |\mathbf{U}(\mathbf{v}_i-\mathbf{v}_i^-)|\geq p-1$。对于第二项$|(\mathbf{U}-\mathbf{U}^-)\mathbf{v}_i^-|$，由于$\mathbf{U}-\mathbf{U}^-$ 的每一列具有形式$\mathbf{e}_j=(0\ldots010\ldots0)^T$(存在一个$j$使得只有第$j$个元素是1)，我们有 $|(\mathbf{U}-\mathbf{U}^-)\mathbf{v}_i^-|=|\mathbf{v}_i^-|\leq k$。对于第三项$\mathbf{U}^-\mathbf{v}_i^--\mathbf{a}_i$，由于$\mathbf{U}^-\mathbf{V}^-=\mathbf{LR}$，我们有 $|\mathbf{U}^-\mathbf{v}_i^--\mathbf{a}_i|=|\mathbf{e}_i|=1$.

综上所述，假如$\mathbf{v}_i\neq\mathbf{v}_i^-$，那么$|\mathbf{U}\mathbf{v}_i-\mathbf{a}_i|\geq p-k-2$。假如$\mathbf{v}_i=\mathbf{v}_i^-$，那么 $|\mathbf{U}\mathbf{v}_i-\mathbf{a}_i|\leq k+1$。因此，当$n$充分大，使得$p=n/k$大于$2k+3$，我们必有$\mathbf{V}=\mathbf{V}^-$.

现在我们可以计算每列的近似误差了。对于使得$\mathbf{a}_i$是$\mathbf{U}$中一列的i，这一列的近似

误差是0。若不然，我们有：

$$\begin{aligned}|\mathbf{U}\mathbf{v}_i - \mathbf{a}_i| &= |\mathbf{U}\mathbf{v}_i - \mathbf{U}\mathbf{v}_i^- + \mathbf{U}\mathbf{v}_i^- - \mathbf{U}^-\mathbf{v}_i^- + \mathbf{U}^-\mathbf{v}_i^- - \mathbf{a}_i| \\ &= |(\mathbf{U} - \mathbf{U}^-)\mathbf{v}_i^- + \mathbf{U}^-\mathbf{v}_i^- - \mathbf{a}_i| \\ &= |(\mathbf{U} - \mathbf{U}^-)\mathbf{v}_i^- + \mathbf{e}_i|.\end{aligned}$$

如上所述，$\mathbf{U} - \mathbf{U}^-$ 的每一列具有$\mathbf{e}_j$的形式。注意到i和j一定不同，因为$\mathbf{a}_i$不是$\mathbf{U}$的一列。因此我们有 $|(\mathbf{U} - \mathbf{U}^-)\mathbf{v}_i^- + \mathbf{e}_i| = |\mathbf{v}_i^-| + 1$。从而，

$$|\mathbf{U}\mathbf{V} - \mathbf{A}| = \sum_{i=1}^{n} |\mathbf{U}\mathbf{v}_i - \mathbf{a}_i| = \sum_{i=1}^{n}(1 + |\mathbf{v}_i^-|) - 2k = n - 2k + \sum_{i=1}^{n} |\mathbf{v}_i^-|, \tag{1.37}$$

其中第二个等式成立是因为对于使得$\mathbf{a}_i$是$\mathbf{U}$中一列的那些i，我们有$|\mathbf{v}_i^-| = |\mathbf{v}_i| = 1$，而这样的$i$的总数为$k$。

接下来我们考察$\mathbf{v}_i^-$。由于$\mathbf{A} - \mathbf{I}_n = \mathbf{L}\mathbf{R}$ 包含了所有$2^k - 1$ 中 $\mathbf{L}$中列向量的非零线性组合，并回顾如下事实：$\mathbf{U}^-$ 包含 k 个线性无关的向量，其中每一个都是$\mathbf{L}$的列向量的线性组合，因此， $\mathbf{V}^-$ 一定是 $\mathbf{R}$经过列之间的打乱而形成的，即它们只有列的顺序不同。根据这个结论以及$\mathbf{R}$的构造不难看出

$$\sum_{i=1}^{n} |\mathbf{v}_i^-| = qk2^{k-1}. \tag{1.38}$$

因此，列选择算法的近似误差是$|\mathbf{A} - \mathbf{U}\mathbf{V}| = n + qk2^{k-1} - 2k$。我们将这个误差与最优误差进行比较。注意到$(\mathbf{L}, \mathbf{R})$ 的近似误差为n，因此，最优误差至多为n。于是我们有列选择算法的近似比至少是$\frac{n+qk2^{k-1}-2k}{n} = 1 + \frac{k}{2} + \frac{k}{2(2^k-1)} - \frac{2k}{n}$。取$n$ 大于 $2k/\epsilon$ ，我们就证明了结论。

最后我们考虑第二种情况，即从$\mathbf{A}$ 中选取的列向量具有如下性质：它们对应的$\mathbf{A} - \mathbf{I}_n$中的$k$列线性相关。不难验证，根据$\mathbf{L}$和$\mathbf{R}$的构造，$\mathbf{L}$中无法被$\mathbf{A} - \mathbf{I}_n$表出的$k$列至少导致$pq - n$的误差。因此，用这种列向量作为基矩阵，近似比至少是$\frac{pq-n}{n} = \frac{n}{k(2^k-1)} - 1$。当$n$充分大之后，这个近似比趋于无穷大，其表现很明显不如第一种情况。这就完成了我们的证明。 □

第二章　布尔半环上的低秩近似

在本章中，我们为布尔半环上的低秩近似涉及一种不同的算法。这里，采用集合而非向量的记号更为方便。对于$\mathbf{A}$的列向量$\mathbf{a}_i$，记$\mathcal{A}_i := \{j \in [d] : (\mathbf{a}_i)_j = 1\}$，即$\mathbf{a}_i$是集合$\mathcal{A}_i$的特征向量。类似地，对于问题的一组最优解$(\mathbf{U}, \mathbf{V})$，记$\mathcal{U}_i := \{j \in [d] : (\mathbf{u}_i)_j = 1\}$。于是在本节中，我们总是把$\mathbf{A}$，$\mathbf{U}$或$\mathbf{V}$的列向量看作集合。给定$\mathcal{S} \subset [k]$，令$\mathcal{J}_{\mathcal{S}} := \{j \in [n] : \mathcal{V}_j = \mathcal{S}\}$，以及$n_{\mathcal{S}} := |\mathcal{J}_{\mathcal{S}}|$。利用以上记号，$\mathbf{U}$ 与集合$\mathcal{S}$的特征向量的布尔乘积记作$\mathcal{U}_{\mathcal{S}} := \bigcup_{i\in\mathcal{S}} \mathcal{U}_i$. (为了简明起见，以下用$\mathcal{U}_i$ 代替$\mathcal{U}_{\{i\}}$) 与上一章类似，$\mathcal{U}_{\mathcal{S}}$ 在$\mathbf{A}$ 的列向量中的最近邻记作$\mathbf{a}_{(\mathcal{U}_{\mathcal{S}})}$。由于在本章中采用了集合记号，为了简洁起见，记$\mathcal{D}_{\mathcal{S}} \subset [d]$为向量$\mathbf{a}_{(\mathcal{U}_{\mathcal{S}})}$所对应的集合，即 $\mathcal{D}_{\mathcal{S}} := \{i \in [d] : \mathbf{a}_{(\mathcal{U}_{\mathcal{S}})i} = 1\}$。

我们将构造一组秩-k 近似的解 $\mathcal{B}_1, \ldots, \mathcal{B}_k$，其中$\mathcal{B}_i \subset [d]$ 是基矩阵的列向量的集合表述。基矩阵确定之后，系数矩阵可以用和上一章同样的方法确定。算法的详细步骤见算法1。

设 $\mathcal{S}_1, \ldots, \mathcal{S}_{2^k-1}$ 是 $2^k - 1$ 个 $[k]$ 的非空子集的一个排序，使得 $n_{\mathcal{S}_1} \leq \ldots \leq n_{\mathcal{S}_{2^k-1}}$，并且 $\mathcal{S}_0 = \emptyset$。如序言部分所提到的，我们将会按照如下方式构造$\mathcal{B}_1, \ldots, \mathcal{B}_k$：1) $\mathcal{B}_i$ 是所有$\mathcal{D}_{\mathcal{S}_l}$，$\ell \in [2^k - 1]$的布尔组合；2) 对任意 $\ell \in [2^k - 1]$

$$\mathcal{U}_{\mathcal{S}_\ell} \triangle \left(\bigcup_{i\in S_\ell} \mathcal{B}_i \right) \subseteq \left(\bigcup_{\ell'\geq\ell} \left(\mathcal{U}_{\mathcal{S}_{\ell'}} \triangle \mathcal{D}_{\mathcal{S}_{\ell'}} \right) \right). \tag{2.1}$$

引理 7. *假如方程(2.1)对所有 $\ell \in [2^k - 1]$成立，那么$\mathcal{B}_1, \ldots, \mathcal{B}_k$ 导出的近似误差不超过最优解的2^k倍。*

证明. 首先我们有对所有 ℓ

$$\begin{aligned}\left| \mathcal{A}_j \triangle \left(\bigcup_{i\in\mathcal{S}_\ell} \mathcal{B}_i \right) \right| &\leq |\mathcal{A}_j \triangle \mathcal{U}_{\mathcal{S}_\ell}| + |\mathcal{U}_{\mathcal{S}_\ell} \triangle (\bigcup_{i\in S_\ell} \mathcal{B}_i)| \\ &\leq |\mathcal{A}_j \triangle \mathcal{U}_{\mathcal{S}_\ell}| + \sum_{\ell'\geq\ell} |\mathcal{D}_{\mathcal{S}_{\ell'}} \triangle \mathcal{U}_{\mathcal{S}_{\ell'}}|.\end{aligned}$$

两边对所有$j \in \mathcal{J}_{\mathcal{S}_\ell}$求和可得：

$$\sum_{j\in\mathcal{J}_{\mathcal{S}_\ell}} \left| \mathcal{A}_j \triangle \left(\bigcup_{i\in S_\ell} \mathcal{B}_i \right) \right| \leq (\sum_{j\in\mathcal{J}_{\mathcal{S}_\ell}} |\mathcal{A}_j \triangle \mathcal{U}_{\mathcal{S}_\ell}|) + \sum_{\ell'\geq\ell} n_{\mathcal{S}_\ell} |\mathcal{D}_{\mathcal{S}_{\ell'}} \triangle \mathcal{U}_{\mathcal{S}_{\ell'}}|.$$

记 $\mathrm{Err}(\mathcal{B}_1, \ldots, \mathcal{B}_k)$ 为 $\mathcal{B}_1, \ldots, \mathcal{B}_k$导出的近似误差，$\mathrm{OPT}_k$为最优解的误差。在上述不等

Algorithm 1 布尔半环上的矩阵低秩近似算法

for 所有 $2^k - 1$ 个列向量组 $\mathcal{A}_{j_1}, \mathcal{A}_{j_2}, \ldots, \mathcal{A}_{j_{2^k-1}}$ **do**
    for 所有双射$\pi : [2^k - 1] \to (2^{[k]} \setminus \{\emptyset\})$ **do**
        for $i \in [k]$ 和 $\ell \in [2^k - 1]$ **do**
            计算

$$\mathcal{E}_i^\ell := \bigcap_{\substack{\ell' \geq \ell: \\ i \in \mathcal{S}_{\ell'}}} D_{S_l}$$

            ，其中 $D_{\mathcal{S}} = \mathcal{A}_{j_{\pi^{-1}(\mathcal{S})}}$，$\emptyset \neq \mathcal{S} \subseteq [k]$.
        end for
        for $1 \leq \ell_1 < \ell_2 \leq 2^k - 1$ 满足$i \in \mathcal{S}_{\ell_1} \cap \mathcal{S}_{l_2}$ **do**
            计算

$$\mathcal{F}_i^{\ell_1,\ell_2} := \mathcal{E}_i^{\ell_1+1} \setminus \left[\bigcup_{i' \in \mathcal{S}_{\ell_2}} \mathcal{E}_{i'}^{\ell_1} \right].$$

        end for
        for $i \in [k]$ **do**
            通过如下方法计算解向量$\{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k\}$：

$$\mathcal{B}_i := \mathcal{E}_i^1 \cup \left(\bigcup_{\substack{\ell_1 < \ell_2: \\ i \in \mathcal{S}_{\ell_1} \cap \mathcal{S}_{\ell_2}}} \mathcal{F}_i^{\ell_1,\ell_2} \right).$$

        end for
    end for
    计算解向量的近似误差。
    if 假如当前解向量的近似误差比之前都更低 **then**
        保存$\{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k\}$作为输出。
    end if
end for

式两边对所有$\ell \in [2^k-1]$ 求和，再两边加上 $(\sum_{j\in\mathcal{J}_{\mathcal{S}_0}}|\mathcal{A}_j|)$，我们得出：

$$\mathrm{Err}(\mathcal{B}_1,\ldots,\mathcal{B}_k) \le \mathrm{OPT}_k + \sum_{\ell=1}^{2^k-1}\sum_{\ell'\ge\ell} n_{\mathcal{S}_\ell}|\mathcal{D}_{\mathcal{S}_{\ell'}} \triangle \mathcal{U}_{\mathcal{S}_{\ell'}}|$$

由于 $n_{\mathcal{S}_\ell} \le n_{\mathcal{S}_{\ell'}}$ 对所有 $\ell' \ge \ell$成立，最后一个不等式变为：

$$\mathrm{Err}(\mathcal{B}_1,\ldots,\mathcal{B}_k) \le \mathrm{OPT}_k + \sum_{\ell=1}^{2^k-1}\sum_{\ell'\ge\ell} n_{\mathcal{S}_{\ell'}}|\mathcal{D}_{\mathcal{S}_{\ell'}} \triangle \mathcal{U}_{\mathcal{S}_{\ell'}}|$$

交换 ℓ 和 ℓ' 的求和顺序，并注意到以下事实： $n_{\mathcal{S}_{\ell'}}|\mathcal{D}_{\mathcal{S}_{\ell'}} \triangle \mathcal{U}_{\mathcal{S}_{\ell'}}| \le \sum_{j\in\mathcal{J}_{\mathcal{S}_{\ell'}}}|\mathcal{A}_j \triangle \mathcal{U}_{\mathcal{S}_{\ell'}}|$，我们最终得到：

$$\begin{aligned}\mathrm{Err}(\mathcal{B}_1,\ldots,\mathcal{B}_k) &\le \mathrm{OPT}_k + \sum_{\ell'=1}^{2^k-1}\ell'\sum_{j\in\mathcal{J}_{\mathcal{S}_{\ell'}}}|\mathcal{A}_j \triangle \mathcal{U}_{\mathcal{S}_{\ell'}}| \\ &\le \mathrm{OPT}_k + (2^k-1)\sum_{\ell'=1}^{2^k-1}\sum_{j\in\mathcal{J}_{\mathcal{S}_{\ell'}}}|\mathcal{A}_j \triangle \mathcal{U}_{\mathcal{S}_{\ell'}}| \\ &\le 2^k\,\mathrm{OPT}_k\,.\end{aligned}$$

这就完成了引理的证明。 □

首先叙述一些有用的简单性质。设$\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D}$ 为集合，则有以下结论成立：

引理 8.
1. $(\mathcal{A}\cup\mathcal{B})\setminus\mathcal{C} = (\mathcal{A}\setminus\mathcal{C})\cup(\mathcal{B}\setminus\mathcal{C})$
2. $(\mathcal{A}\cap\mathcal{B})\setminus\mathcal{C} = (\mathcal{A}\setminus\mathcal{C})\cap(\mathcal{B}\setminus\mathcal{C})$
3. $\mathcal{A}\setminus(\mathcal{B}\cup\mathcal{C}) = (\mathcal{A}\setminus\mathcal{B})\cap(\mathcal{A}\setminus\mathcal{C})$
4. $\mathcal{A}\setminus(\mathcal{B}\cap\mathcal{C}) = (\mathcal{A}\setminus\mathcal{B})\cup(\mathcal{A}\setminus\mathcal{C})$

推论 9. 由上一个引理，
1. $(\mathcal{A}\cup\mathcal{B})\setminus(\mathcal{C}\cup\mathcal{D}) \subseteq (\mathcal{A}\setminus\mathcal{C})\cup(\mathcal{B}\setminus\mathcal{D})$
2. $(\mathcal{A}\cap\mathcal{B})\setminus(\mathcal{C}\cap\mathcal{D}) \subseteq (\mathcal{A}\setminus\mathcal{C})\cup(\mathcal{B}\setminus\mathcal{D})$

类似地，
1. $(\mathcal{A}\cup\mathcal{B})\triangle(\mathcal{C}\cup\mathcal{D}) \subseteq (\mathcal{A}\triangle\mathcal{C})\cup(\mathcal{B}\triangle\mathcal{D})$
2. $(\mathcal{A}\cap\mathcal{B})\triangle(\mathcal{C}\cap\mathcal{D}) \subseteq (\mathcal{A}\triangle\mathcal{C})\cup(\mathcal{B}\triangle\mathcal{D})$

引理 10. 三角形不等式：
1. $\mathcal{A}\setminus\mathcal{B} \subseteq (\mathcal{A}\setminus\mathcal{C})\cup(\mathcal{C}\setminus\mathcal{B})$.

2. $\mathcal{A} \triangle \mathcal{B} \subseteq (\mathcal{A} \triangle \mathcal{C}) \cup (\mathcal{C} \triangle \mathcal{B})$.

以下我们开始构造 $\mathcal{B}_1, \ldots, \mathcal{B}_{2^k-1}$ 使得它们满足方程 (2.1)。第一步是定义集合 $\mathcal{E}_i^\ell$。对于$i \in [k]$和$\ell \in [2^k-1]$，定义：

$$\mathcal{E}_i^\ell := \bigcap_{\substack{\ell' \geq \ell: \\ i \in \mathcal{S}_{\ell'}}} \mathcal{D}_{\mathcal{S}_{\ell'}}$$

显然由定义我们有$\mathcal{E}_i^\ell \subseteq \mathcal{E}_i^{\ell'}$ 对所有 $1 \leq \ell \leq \ell' \leq 2^k - 1$成立。

我们基于$\mathcal{D}_{\mathcal{S}_1}, \ldots, \mathcal{D}_{\mathcal{S}_{2^k-1}}$来构造$\mathcal{B}_1, \ldots, \mathcal{B}_{2^k-1}$。这个构造的主要思路是为了得到对任意$\ell$，

$$\bigcup_{i \in \mathcal{S}_\ell} \mathcal{B}_i = \left(\bigcup_{i \in \mathcal{S}_\ell} \mathcal{E}_i^\ell \right) \cup \mathcal{R}_\ell \ , \tag{2.2}$$

，其中R_ℓ 可以表示 $\mathcal{U}_{\mathcal{S}_{\ell'}} \triangle \mathcal{D}_{\mathcal{S}_{\ell'}}$ 之并($\ell' \geq \ell$) 的子集。以下引理说明了(2.2)的构造蕴含了方程(2.1).

引理 11. *假设方程(2.2)成立，那么*

$$\mathcal{U}_{\mathcal{S}_\ell} \triangle \left(\bigcup_{i \in \mathcal{S}_\ell} \mathcal{B}_i \right) \subseteq \left(\bigcup_{\substack{\ell' \geq \ell: \\ \mathcal{S}_\ell \cap \mathcal{S}_{\ell'} \neq \emptyset}} \left(\mathcal{U}_{\mathcal{S}_{\ell'}} \triangle \mathcal{D}_{\mathcal{S}_{\ell'}} \right) \right) \cup \mathcal{R}_\ell. \tag{2.3}$$

证明. 利用(2.2)，我们有：

$$\begin{aligned} \mathcal{U}_{\mathcal{S}_\ell} \triangle \left(\bigcup_{i \in \mathcal{S}_\ell} \mathcal{B}_i \right) &\subseteq \left(\mathcal{U}_{\mathcal{S}_\ell} \setminus \left[\bigcup_{i \in \mathcal{S}_\ell} \mathcal{E}_i^\ell \right] \right) \cup \left(\left[\bigcup_{i \in \mathcal{S}_\ell} \mathcal{E}_i^\ell \right] \setminus \mathcal{U}_{\mathcal{S}_\ell} \right) \cup \mathcal{R}_\ell \\ &\subseteq \left(\bigcup_{i \in \mathcal{S}_\ell} \left(\mathcal{U}_i \setminus \mathcal{E}_i^\ell \right) \right) \cup (\mathcal{D}_{\mathcal{S}_\ell} \setminus \mathcal{U}_{\mathcal{S}_\ell}) \cup \mathcal{R}_\ell \end{aligned}$$

其中我们用到了推论9以及$\bigcup_{i \in \mathcal{S}_\ell} \mathcal{E}_i^\ell \subseteq \mathcal{D}_{\mathcal{S}_\ell}$。以下我们将要证明对任意$i \in S_\ell$,

$$\mathcal{U}_i \setminus \mathcal{E}_i^\ell \subseteq \bigcup_{\substack{\ell' \geq \ell: \\ i \in \mathcal{S}_{\ell'}}} \left(\mathcal{U}_{\mathcal{S}_{\ell'}} \setminus \mathcal{D}_{\mathcal{S}_{\ell'}} \right).$$

为了证明这个结论，首先注意到

$$\mathcal{U}_i \subseteq \bigcap_{\substack{\ell' \geq \ell: \\ i \in \mathcal{U}_{\ell'}}} \mathcal{U}_{\mathcal{S}_{\ell'}}.$$

因此，

$$\mathcal{U}_i \setminus \mathcal{E}_i^\ell \subseteq \left(\bigcap_{\substack{\ell' \geq \ell: \\ i \in \mathcal{S}_{\ell'}}} \mathcal{U}_{\mathcal{S}_{\ell'}}\right) \setminus \left(\bigcap_{\substack{\ell' \geq \ell: \\ i \in \mathcal{S}_{\ell'}}} \mathcal{D}_{\mathcal{S}_{\ell'}}\right) \subseteq \bigcup_{\substack{\ell' \geq \ell: \\ i \in \mathcal{S}_{\ell'}}} \left(\mathcal{U}_{\mathcal{S}_{\ell'}} \setminus \mathcal{D}_{\mathcal{S}_{\ell'}}\right).$$

从而

$$\begin{aligned} \mathcal{U}_{\mathcal{S}_\ell} \triangle \left(\bigcup_{i \in \mathcal{S}_\ell} \mathcal{B}_i\right) &\subseteq \left(\bigcup_{i \in \mathcal{S}_\ell} \bigcup_{\substack{\ell' \geq \ell: \\ i \in \mathcal{S}_{\ell'}}} \left(\mathcal{U}_{\mathcal{S}_{\ell'}} \setminus \mathcal{D}_{\mathcal{S}_{\ell'}}\right)\right) \cup (\mathcal{D}_{\mathcal{S}_\ell} \setminus \mathcal{U}_{\mathcal{S}_\ell}) \cup \mathcal{R}_\ell \\ &\subseteq \left(\bigcup_{\substack{\ell' \geq \ell: \\ \mathcal{S}_\ell \cap \mathcal{S}_{\ell'} \neq \emptyset}} \left(\mathcal{U}_{\mathcal{S}_{\ell'}} \triangle \mathcal{D}_{\mathcal{S}_{\ell'}}\right)\right) \cup \mathcal{R}_\ell. \end{aligned} \tag{2.4}$$

□

为了满足方程(2.2)，我们按照如下方式构造：首先对所有的$i \in [k]$，取$\mathcal{B}_i = \mathcal{E}_i^1$。接下来，对每个$\ell \in [2^k - 1]$，我们只需向$\bigcup_{i \in S_\ell} \mathcal{B}_i$加上

$$\left(\bigcup_{i \in \mathcal{S}_\ell} \mathcal{E}_i^\ell\right) \setminus \left(\bigcup_{i \in \mathcal{S}_\ell} \mathcal{E}_i^1\right)$$

，使得方程(2.2) 被满足。但现在，为了能够对所有ℓ'同时限制$\mathcal{R}_{\ell'}$，我们需要利用$\mathcal{S}_1, \ldots, \mathcal{S}_{2^k-1}$的序关系仔细处理每一项。

构造中最主要的一步如下：对满足$i \in \mathcal{S}_{\ell_1} \cap \mathcal{S}_{\ell_2}$的$\ell_1 < \ell_2$，定义集合$\mathcal{F}_i^{\ell_1,\ell_2}$如下：

$$\mathcal{F}_i^{\ell_1,\ell_2} := \mathcal{E}_i^{\ell_1+1} \setminus \left[\bigcup_{i' \in \mathcal{S}_{\ell_2}} \mathcal{E}_{i'}^{\ell_1}\right].$$

现在，可以定义秩-k近似的一组解$(\mathcal{B}_1, \ldots, \mathcal{B}_k)$：

$$\mathcal{B}_i := \mathcal{E}_i^1 \cup \left(\bigcup_{\substack{\ell_1 < \ell_2: \\ i \in \mathcal{S}_{\ell_1} \cap \mathcal{S}_{\ell_2}}} \mathcal{F}_i^{\ell_1,\ell_2}\right). \tag{2.5}$$

固定ℓ，则有

$$\bigcup_{i \in \mathcal{S}_\ell} \mathcal{B}_i = \left(\bigcup_{i \in \mathcal{S}_\ell} \mathcal{E}_i^1\right) \cup \left(\bigcup_{i \in S_\ell} \bigcup_{\substack{\ell_1 < \ell_2: \\ i \in \mathcal{S}_{\ell_1} \cap \mathcal{S}_{\ell_2}}} \mathcal{F}_i^{\ell_1,\ell_2}\right).$$

以下引理给出了方程(2.2)中$\mathcal{R}_\ell$的表达式。

引理 12.

$$\bigcup_{i\in\mathcal{S}_\ell}\mathcal{B}_i=\left(\bigcup_{i\in\mathcal{S}_\ell}\mathcal{E}_i^\ell\right)\cup\left(\bigcup_{i\in\mathcal{S}_\ell}\bigcup_{\substack{\ell\leq\ell_1<\ell_2:\\ i\in\mathcal{S}_{\ell_1}\cap\mathcal{S}_{\ell_2}}}\mathcal{F}_i^{\ell_1,\ell_2}\right).$$

于是我们可以令

$$\mathcal{R}_\ell=\bigcup_{i\in\mathcal{S}_\ell}\bigcup_{\substack{\ell\leq\ell_1<\ell_2:\\ i\in\mathcal{S}_{\ell_1}\cap\mathcal{S}_{\ell_2}}}\mathcal{F}_i^{\ell_1,\ell_2}\quad.$$

证明. 当 $\ell_1<\ell$，我们有 $\mathcal{F}_i^{\ell_1,\ell_2}\subseteq\mathcal{E}_i^{\ell_1+1}\subseteq\mathcal{E}_i^\ell$, 于是，

$$\bigcup_{\substack{\ell_1<\ell\\ \ell_1<\ell_2:\\ i\in\mathcal{S}_{\ell_1}\cap\mathcal{S}_{\ell_2}}}\mathcal{F}_i^{\ell_1,\ell_2}\subseteq\mathcal{E}_i^\ell$$

进而

$$\bigcup_{i\in S_\ell}\bigcup_{\substack{\ell_1<\ell_2:\\ i\in\mathcal{S}_{\ell_1}\cap\mathcal{S}_{\ell_2}}}\mathcal{F}_i^{\ell_1,\ell_2}\subseteq\left(\bigcup_{i\in\mathcal{S}_\ell}\mathcal{E}_i^\ell\right)\cup\left(\bigcup_{i\in S_\ell}\bigcup_{\substack{\ell\leq\ell_1<\ell_2:\\ i\in\mathcal{S}_{\ell_1}\cap\mathcal{S}_{\ell_2}}}\mathcal{F}_i^{\ell_1,\ell_2}\right)$$

另一方面，

$$\bigcup_{i\in S_\ell}\bigcup_{\substack{\ell\leq\ell_1<\ell_2:\\ i\in\mathcal{S}_{\ell_1}\cap\mathcal{S}_{\ell_2}}}\mathcal{F}_i^{\ell_1,\ell_2}\subseteq\bigcup_{i\in S_\ell}\bigcup_{\substack{\ell_1<\ell_2:\\ i\in\mathcal{S}_{\ell_1}\cap\mathcal{S}_{\ell_2}}}\mathcal{F}_i^{\ell_1,\ell_2}$$

因此，

$$\begin{aligned}&\left(\bigcup_{i\in\mathcal{S}_\ell}\mathcal{E}_i^\ell\right)\cup\left(\bigcup_{i\in S_\ell}\bigcup_{\substack{\ell\leq\ell_1<\ell_2:\\ i\in\mathcal{S}_{\ell_1}\cap\mathcal{S}_{\ell_2}}}\mathcal{F}_i^{\ell_1,\ell_2}\right)\\ =&\left(\bigcup_{i\in\mathcal{S}_\ell}\mathcal{E}_i^\ell\right)\cup\left(\bigcup_{i\in S_\ell}\bigcup_{\substack{\ell_1<\ell_2:\\ i\in\mathcal{S}_{\ell_1}\cap\mathcal{S}_{\ell_2}}}\mathcal{F}_i^{\ell_1,\ell_2}\right)\\ =&\bigcup_{i\in\mathcal{S}_\ell}\left[\mathcal{E}_i^\ell\cup\left(\bigcup_{\substack{\ell_1<\ell_2:\\ i\in\mathcal{S}_{\ell_1}\cap\mathcal{S}_{\ell_2}}}\mathcal{F}_i^{\ell_1,\ell_2}\right)\right]\end{aligned}$$

由于

$$\mathcal{B}_i = \mathcal{E}_i^1 \cup \left(\bigcup_{\substack{\ell_1<\ell_2:\\ i\in\mathcal{S}_{\ell_1}\cap\mathcal{S}_{\ell_2}}} \mathcal{F}_i^{\ell_1,\ell_2} \right) \subseteq \mathcal{E}_i^{\ell} \cup \left(\bigcup_{\substack{\ell_1<\ell_2:\\ i\in\mathcal{S}_{\ell_1}\cap\mathcal{S}_{\ell_2}}} \mathcal{F}_i^{\ell_1,\ell_2} \right)$$

我们有：

$$\bigcup_{i\in\mathcal{S}_\ell} \mathcal{B}_i \subseteq \bigcup_{i\in\mathcal{S}_\ell} \left[\mathcal{E}_i^{\ell} \cup \left(\bigcup_{\substack{\ell_1<\ell_2:\\ i\in\mathcal{S}_{\ell_1}\cap\mathcal{S}_{\ell_2}}} \mathcal{F}_i^{\ell_1,\ell_2} \right) \right]$$

以下只需证明：

$$\bigcup_{i\in\mathcal{S}_\ell} \left[\mathcal{E}_i^{\ell} \cup \left(\bigcup_{\substack{\ell_1<\ell_2:\\ i\in\mathcal{S}_{\ell_1}\cap\mathcal{S}_{\ell_2}}} \mathcal{F}_i^{\ell_1,\ell_2} \right) \right] \subseteq \bigcup_{i\in\mathcal{S}_\ell} \mathcal{B}_i$$

注意到由方程(2.5),

$$\bigcup_{\substack{\ell_1<\ell_2:\\ i\in\mathcal{S}_{\ell_1}\cap\mathcal{S}_{\ell_2}}} \mathcal{F}_i^{\ell_1,\ell_2} \subseteq \mathcal{B}_i$$

于是只要说明对所有 ℓ，均有

$$\bigcup_{i\in\mathcal{S}_\ell} \mathcal{E}_i^{\ell} \subseteq \bigcup_{i\in\mathcal{S}_\ell} \mathcal{B}_i. \tag{2.6}$$

我们通过归纳证明对所有$\ell' \leq \ell$，均有

$$\bigcup_{i\in\mathcal{S}_\ell} \mathcal{E}_i^{\ell'} \subseteq \bigcup_{i\in\mathcal{S}_\ell} \mathcal{B}_i \tag{2.7}$$

来证明方程(2.6)。归纳奠基$\ell' = 1$ 成立是因为由$\mathcal{B}_i$的定义，$\mathcal{E}_i^1 \subseteq \mathcal{B}_i$对所有$i$均成立。现在设 $\ell' < \ell$，$i \in \mathcal{S}_\ell$。分两种情况讨论。假如 $i \notin \mathcal{S}_{\ell'}$，我们有$\mathcal{E}_i^{\ell'+1} = \mathcal{E}_i^{\ell'}$，从而由归纳假设 $\mathcal{E}_i^{\ell'+1} \subseteq \bigcup_{i'\in\mathcal{S}_\ell} \mathcal{B}_{i'}$。以下可设 $i \in \mathcal{S}_{\ell'}$。由归纳假设，我们有

$$\bigcup_{i'\in\mathcal{S}_\ell} \mathcal{E}_{i'}^{\ell'} \subseteq \bigcup_{i'\in\mathcal{S}_\ell} \mathcal{B}_{i'} \ ,$$

由于 $i \in \mathcal{S}_{\ell'} \cap \mathcal{S}_\ell$ ，我们有：

$$\mathcal{F}_i^{\ell',\ell} = \mathcal{E}_i^{\ell'+1} \setminus \left[\bigcup_{i'\in\mathcal{S}_\ell} \mathcal{E}_{i'}^{\ell'} \right] \subseteq \mathcal{B}_i \ ,$$

于是我们可以断言在这种情况下， $\mathcal{E}_i^{\ell'+1} \subseteq \bigcup_{i'\in\mathcal{S}_\ell} \mathcal{B}_{i'}$同样成立，从而完成了证明。 □

最后，当$\ell' \geq \ell$ 时，可以通过以下引理来证明$\mathcal{R}_\ell$ 是 $\mathcal{U}_{\mathcal{S}_{\ell'}} \triangle \mathcal{D}_{\mathcal{S}_{\ell'}}$之并的子集。

引理 13. *设* $\ell_1 < \ell_2$ *满足* $i \in \mathcal{S}_{\ell_1} \cap \mathcal{S}_{\ell_2}$*，那么*

$$\mathcal{F}_i^{\ell_1,\ell_2} \subseteq \left(\mathcal{D}_{\mathcal{S}_{\ell_2}} \setminus \mathcal{U}_{\mathcal{S}_{\ell_2}}\right) \cup \left(\bigcup_{i' \in \mathcal{S}_{\ell_2}} \bigcup_{\substack{\ell' \geq \ell_1: \\ i' \in \mathcal{S}_{\ell'}}} \left(\mathcal{U}_{\mathcal{S}_{\ell'}} \setminus \mathcal{D}_{\mathcal{S}_{\ell'}}\right)\right).$$

证明. 只需注意到如下放缩：

$$\begin{aligned}
\mathcal{F}_i^{\ell_1,\ell_2} &\subseteq \left(\mathcal{E}_i^{\ell_1+1} \setminus \mathcal{U}_{\mathcal{S}_{\ell_2}}\right) \cup \left(\mathcal{U}_{\mathcal{S}_{\ell_2}} \setminus \left[\bigcup_{i' \in \mathcal{S}_{\ell_2}} \mathcal{E}_{i'}^{\ell_1}\right]\right) \\
&\subseteq \left(\mathcal{D}_{\mathcal{S}_{\ell_2}} \setminus \mathcal{U}_{\mathcal{S}_{\ell_2}}\right) \cup \left(\left[\bigcup_{i' \in \mathcal{S}_{\ell_2}} \bigcap_{\substack{\ell' \geq \ell_1: \\ i' \in \mathcal{S}_{\ell'}}} \mathcal{U}_{\mathcal{S}_{\ell'}}\right] \setminus \left[\bigcup_{i' \in \mathcal{S}_{\ell_2}} \bigcap_{\substack{\ell' \geq \ell_1: \\ i' \in \mathcal{S}_{\ell'}}} \mathcal{D}_{\mathcal{S}_{\ell'}}\right]\right) \\
&\subseteq \left(\mathcal{D}_{\mathcal{S}_{\ell_2}} \setminus \mathcal{U}_{\mathcal{S}_{\ell_2}}\right) \cup \left(\bigcup_{i' \in \mathcal{S}_{\ell_2}} \bigcup_{\substack{\ell' \geq \ell_1: \\ i' \in \mathcal{S}_{\ell'}}} \left(\mathcal{U}_{\mathcal{S}_{\ell'}} \setminus \mathcal{D}_{\mathcal{S}_{\ell'}}\right)\right),
\end{aligned}$$

这里，我们用到了 $\mathcal{E}_i^{\ell_1} \subseteq \mathcal{D}_{\mathcal{S}_{\ell_2}}$ ，以及 $\mathcal{U}_{i'} \subseteq \mathcal{U}_{\mathcal{S}_{\ell'}}$ 对所有使得$i' \in \mathcal{S}_{\ell'}$，$i' \in \mathcal{S}_{\ell_2}$ 的$\ell' \geq \ell_1$成立。 □

综合方程(2.4) , 引理 12 和引理 13，我们最终得到

$$\mathcal{U}_{\mathcal{S}_\ell} \triangle \left(\bigcup_{i \in \mathcal{S}_\ell} \mathcal{B}_i\right) \subseteq \left(\bigcup_{\ell' \geq \ell} \left(\mathcal{U}_{\mathcal{S}_{\ell'}} \triangle \mathcal{D}_{\mathcal{S}_{\ell'}}\right)\right)$$

对所有 ℓ成立，从而证明了算法的近似比。

定理 14. *以上算法的近似比为*2^k*。*

以下我们分析算法的时间复杂度。我们枚举了所有 n^{2^k-1} 种选择列向量的方式，以及 $(2^k-1)!$ 个双射。对它们中的每一个，我们可以在$O(k^2 2^k)$的时间内计算$\mathcal{E}_i^\ell$，然后用$O(k^2 2^{2k})$的时间计算 $\mathcal{F}_i^{\ell_1,\ell_2}$ ，最后用$O(k 2^{2k})$的时间计算$\mathcal{B}_i$。计算$\mathrm{Err}(\mathcal{B}_1, \ldots, \mathcal{B}_k)$ 需要 $O(2^{2k} dn)$的时间。因此，总的时间复杂度为 $O(n^{2^k-1}(2^k-1)! k^2 2^{2k} dn) = O((2^k+2)! dn^{2^k})$。

注记 15. *使用类似于*GF(2)*模型中的加权平均技巧，我们事实上可以将近似比改进到*$2^{k-1}+1$*。这里我们略去这个结论的证明。*

第三章 计算复杂性

在本章中，我们证明秩-1 二元矩阵近似问题是NP-完全的。我们首先定义两个相关问题。

设H为带权完全二部图，邻接矩阵 $\mathbf{W} = (w_{ij})$ 的大小为$d \times n$，包含了其所有边的权重。最大边权二部团问题(MAXIMUM EDGE WEIGHT BICLIQUE) 是指寻找H的一个二部团子图，使得其包含的所有边权值之和最大。它可以被等价表述为如下的优化问题：$\mathbf{x}^\mathsf{T}\mathbf{W}\mathbf{y}$，其中$\mathbf{x} \in \{0,1\}^d$ ，$\mathbf{y} \in \{0,1\}^n$。最大二部割(BIPARTITE MAX-CUT)问题是指寻找H的一个割，使得这个割的所有边权值之和最大。它也可以被等价表述为如下的优化问题：$\mathbf{x}^\mathsf{T}\mathbf{W}\mathbf{y}$，其中$\mathbf{x} \in \{-1,1\}^d$，$\mathbf{y} \in \{-1,1\}^n$。可以看出，这两个问题的区别只在于$\mathbf{x}$和$\mathbf{y}$的定义域。

Shen, Ji 和 Ye [1] 发现，秩-1二元矩阵近似问题 等价于最大边权二部团问题 在所有边权均取值于$\{-1,1\}$的特殊情况。具体地，若$\mathbf{A}$为一个$d \times n$ 二元矩阵，$\mathbf{u} \in \{0,1\}^d$，$\mathbf{v} \in \{0,1\}^n$，令$\mathbf{J}_{d,n}$为$d \times n$ 全1矩阵，我们有

$$\|\mathbf{A} - \mathbf{u}\mathbf{v}^\mathsf{T}\|_F^2 = \|\mathbf{A}\|_F^2 - 2\mathbf{u}^\mathsf{T}\mathbf{A}\mathbf{v} + \|\mathbf{u}\mathbf{v}^\mathsf{T}\|_F^2 = \|\mathbf{A}\|_F^2 - \mathbf{u}^\mathsf{T}(2\mathbf{A} - \mathbf{J}_{d,n})\mathbf{v}.$$

于是，最小化$\|\mathbf{A}-\mathbf{u}\mathbf{v}^\mathsf{T}\|_F^2$ 等价于最大化$\mathbf{u}^\mathsf{T}(2\mathbf{A}-\mathbf{J}_{d,n})\mathbf{v}$。同时，注意到$(2\mathbf{A}-\mathbf{J}_{d,n})$ 的所有元素均取值于$\{-1,1\}$，因此，边权取值在$\{-1,1\}$的最大边权二部团问题的NP-困难性蕴含了秩-1 二元矩阵近似问题 的NP-困难性。为了证明最大边权二部团问题的NP-困难性，我们只需建立从最大二部割问题到这个问题的归约。

Roth 和 Viswanathan 证明了最大二部割问题即使在所有权值均取值于$\{-1,1\}$的特例也是NP-难的 [25]。它们首先证明了当所有权值均取值于$\{-1,0,1\}$时的NP-困难性，然后归约到了权值在$\{-1,1\}$的情形。

Tan 证明了最大边权二部团问题当权值取值于 $\{-1,0,1\}$时是NP-难的 [20] ，并给出了该情形到$\{-1,1\}$的一个随机归约，即证明了问题没有多项式算法，除非 NP=RP。他将本问题的NP-困难性留作了一个开放问题。本问题的计算复杂性也被Amit [26]留作了开放问题。

Roth , Viswanathan 以及 Tan 给出的从$\{-1,0,1\}$-权值到$\{-1,1\}$-权值的归约是类似的。其主要想法是将$n \times n$ $\{-1,0,1\}$-权值矩阵$\mathbf{W}$转换成一个$nm \times nm$ $\{-1,1\}$-权值矩阵$\mathbf{W}'$，其中$\mathbf{W}'$包含$m \times m$分块矩阵，每个分块矩阵对应 $\mathbf{W}$的一个元素。每个(-1)元素被转换为一个全-(-1) $m\times m$ 矩阵，类似地，每个 1元素被转换为全1 $m\times m$ 矩阵。他

们的不同之处在于，Tan 将每个0元素转化为一个随机 $m \times m$ $\{-1,1\}$-矩阵，而 Roth 和 Viswanathan 则将每个0元素转换成一个$m \times m$ *Hadamard* 矩阵。我们将会证明，这个转换为*Hadamard* 矩阵的技巧在最大边权二部团问题问题中同样适用，从而证明其NP-困难性。

定理 16. 秩-1的二元矩阵近似问题 是NP-难的。

我们给出从权值在$\{-1,0,1\}$上的最大边权二部团问题 到权值在$\{0,1\}$上的最大边权二部团问题的多项式时间归约，从而证明原定理。我们的证明需要如下三个引理。

以下引理与 [25, Lemma 4.2] 类似。区别在于原文中的定义域是$\{-1,1\}$，而本文中是$\{0,1\}$。

引理 17. 设$\mathbf{W}$为$n \times n$矩阵，$m \geq 1$。定义$\mathbf{W}' = \mathbf{W} \otimes \mathbf{J}_m$，其中$\mathbf{J}_m := \mathbf{J}_{m,m}$。那么，

$$\max_{\mathbf{u},\mathbf{v}} \mathbf{u}^\mathsf{T}\mathbf{W}'\mathbf{v} = m^2 \cdot \max_{\mathbf{x},\mathbf{y}} \mathbf{x}^\mathsf{T}\mathbf{W}\mathbf{y} \ ,$$

这里，$\mathbf{u},\mathbf{v} \in \{0,1\}^{mn}$，$\mathbf{x},\mathbf{y} \in \{0,1\}^n$。进一步地，若$\mathbf{x}$和$\mathbf{y}$使得$\mathbf{x}^\mathsf{T}\mathbf{W}\mathbf{y}$最大化，那么$\mathbf{u} = \mathbf{x} \otimes \mathbf{1}_m$和$\mathbf{v} = \mathbf{y} \otimes \mathbf{1}_m$也使得$\mathbf{u}^\mathsf{T}\mathbf{W}'\mathbf{v}$最大化。

证明. 首先，当$\mathbf{u} = \mathbf{x} \otimes \mathbf{1}_d$，$\mathbf{v} = \mathbf{y} \otimes \mathbf{1}_m$时，

$$\begin{aligned}\mathbf{u}^\mathsf{T}(\mathbf{W} \otimes \mathbf{J}_m)\mathbf{v} &= (\mathbf{x} \otimes \mathbf{1}_m)^\mathsf{T}(\mathbf{W} \otimes \mathbf{J}_m)(\mathbf{y} \otimes \mathbf{1}_m) \\ &= (\mathbf{x}^\mathsf{T}\mathbf{W}\mathbf{y}) \otimes (\mathbf{1}_m^\mathsf{T}\mathbf{J}_m\mathbf{1}_m) = m^2 \cdot (\mathbf{x}^\mathsf{T}\mathbf{W}\mathbf{y}) \ .\end{aligned}$$

接下来，我们考虑使得$\mathbf{u}^\mathsf{T}\mathbf{W}'\mathbf{v}$最大的$\mathbf{u}$和$\mathbf{v}$。我们证明$\mathbf{u}$ 和$\mathbf{v}$可以取成$\mathbf{u} = \mathbf{x} \otimes \mathbf{1}_m$，$\mathbf{v} = \mathbf{y} \otimes \mathbf{1}_m$的形式而不减小$\mathbf{u}^\mathsf{T}\mathbf{W}'\mathbf{v}$的值。首先固定$\mathbf{v}$，让$\mathbf{u}$ 变为想要的形式，然后类似地让$\mathbf{v}$变为想要的形式。

先固定$\mathbf{v}$，令$\mathbf{z} = \mathbf{W}'\mathbf{v}$。注意到使得$\mathbf{u}^\mathsf{T}\mathbf{z}$ 最大化的$\mathbf{u}$必须满足：$\mathbf{z}_i > 0$时$\mathbf{u}_i = 1$，而$\mathbf{z}_i < 0$时$\mathbf{u}_i = 0$。由于 $\mathbf{W}' = \mathbf{W} \otimes \mathbf{J}_m$我们有$\mathbf{z}_{jm+1} = \mathbf{z}_{jm+2} = \cdots = \mathbf{z}_{(j+1)m}$对所有$j = 0,1,\ldots,n-1$成立。因此，我们可以取一个满足$\mathbf{u}_{jm+1} = \mathbf{u}_{jm+2} = \cdots = \mathbf{u}_{(j+1)m}$对所有 $j = 0,1,\ldots,n-1$成立的$\mathbf{u}$，即$\mathbf{u} = \mathbf{x} \otimes \mathbf{1}_m$对某个$\mathbf{x} \in \{0,1\}^n$成立，使得它同时最大化$\mathbf{u}^\mathsf{T}\mathbf{z}$。然后我们可以类似地固定$\mathbf{u}$并将$\mathbf{v}$变为$\mathbf{v} = \mathbf{y} \otimes \mathbf{1}_m$的形式，从而完成证明。
□

以下引理是[25, Lemma 4.3]在$\{0,1\}$的类似形式。

引理 18. 设$\mathbf{H}$为$m \times m$*Hadamard* 矩阵。对所有$\mathbf{x}, \mathbf{y} \in \{0,1\}^m$，

$$|\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{y}| \leq m^{3/2} \ .$$

证明. 首先注意到

$$\|\mathbf{H}\mathbf{y}\|^2 = \mathbf{y}^\mathsf{T}(\mathbf{H}^\mathsf{T}\mathbf{H})\mathbf{y} = \mathbf{y}^\mathsf{T}(m\mathbf{I})\mathbf{y} = m \cdot \|\mathbf{y}\|^2 \ .$$

于是由 Cauchy-Schwartz 不等式，

$$|\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{y}| \leq \|\mathbf{x}^\mathsf{T}\| \cdot \|\mathbf{H}\mathbf{y}\| = \sqrt{m} \cdot \|\mathbf{x}\| \cdot \|\mathbf{y}\| \leq m^{3/2} \ .$$

证毕。 □

引理 19. 设$\mathbf{W} = (w_{ij})$为$n \times n$ $\{-1,0,1\}$-矩阵，并设$\mathbf{H}$为 $m \times m$ *Hadamard* 矩阵。定义$(mn) \times (mn)$ $\{-1,1\}$-分块矩阵$\widetilde{\mathbf{W}} = (\widetilde{\mathbf{W}}_{ij})$，其中块 $\widetilde{\mathbf{W}}_{ij}$ 如下定义：

$$\widetilde{\mathbf{W}}_{ij} = \begin{cases} w_{ij}\mathbf{J}_m & \text{if } w_{ij} \neq 0 \\ \mathbf{H} & \text{if } w_{ij} = 0 \end{cases} .$$

记$\mathbf{W}' = \mathbf{W} \otimes \mathbf{J}_m$，那么对任意$\mathbf{u}, \mathbf{v} \in \{0,1\}^{mn}$,

$$\left|\mathbf{u}^\mathsf{T}\widetilde{\mathbf{W}}\mathbf{v} - \mathbf{u}^\mathsf{T}\mathbf{W}'\mathbf{v}\right| \leq n^2 \cdot m^{3/2} \ .$$

证明. 只需进行一些简单的估计：

$$\begin{aligned} \left|\mathbf{u}^\mathsf{T}\widetilde{\mathbf{W}}\mathbf{v} - \mathbf{u}^\mathsf{T}\mathbf{W}'\mathbf{v}\right| &= \left|\mathbf{u}^\mathsf{T}(\widetilde{\mathbf{W}} - \mathbf{W}')\mathbf{v}\right| \\ &\leq n^2 \cdot \max_{\mathbf{x},\mathbf{y} \in \{0,1\}^m} \left|\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{y}\right| \\ &\leq n^2 \cdot m^{3/2} \ , \end{aligned}$$

最后一个不等式是引理 18的推论。 □

证明. （**定理 16**）假设$\mathbf{W}$是$n \times n$ $\{-1,0,1\}$-矩阵。设$m = 2^\ell$ 是最小的比$4n^4$大的2的幂，$\mathbf{H}$ 为$m \times m$ Sylvester Hadamard 矩阵。我们接下来按照引理 19的方式定义 $\widetilde{\mathbf{W}}$和$\mathbf{W}'$。

那么

$$
\begin{aligned}
&\left|\max_{\mathbf{u},\mathbf{v}\in\{0,1\}^{mn}} \mathbf{u}^{\mathsf{T}}\widetilde{\mathbf{W}}\mathbf{v} - m^2\cdot \max_{\mathbf{x},\mathbf{y}\in\{0,1\}^{n}} \mathbf{x}^{\mathsf{T}}\mathbf{W}\mathbf{y}\right| \\
=&\left|\max_{\mathbf{u},\mathbf{v}\in\{0,1\}^{mn}} \mathbf{u}^{\mathsf{T}}\widetilde{\mathbf{W}}\mathbf{v} - \max_{\mathbf{u},\mathbf{v}\in\{0,1\}^{mn}} \mathbf{u}^{\mathsf{T}}\mathbf{W}'\mathbf{v}\right| \\
\leq& n^2\cdot m^{3/2} \leq \frac{m^{1/2}}{2}\cdot m^{3/2} = \frac{m^2}{2} \ ,
\end{aligned}
$$

其中第一个等式由引理17得到，第一个不等式由引理19得到。

由于 $m^2\cdot\max_{\mathbf{x},\mathbf{y}\in\{0,1\}^n}\mathbf{x}^{\mathsf{T}}\mathbf{W}\mathbf{y}$ 是m^2的整数倍，故$\max_{\mathbf{u},\mathbf{v}\in\{0,1\}^{mn}}\mathbf{u}^{\mathsf{T}}\widetilde{\mathbf{W}}\mathbf{v}$ 唯一确定了 $\max_{\mathbf{x},\mathbf{y}\in\{0,1\}^n}\mathbf{x}^{\mathsf{T}}\mathbf{W}\mathbf{y}$的值。这样我们就完成了归约。 □

结论与展望

在本文中，我们给出二元矩阵近似问题的首个 对于一般情形 $k > 1$ 的结果，同时包含了GF(2) 模型和布尔模型。对于 GF(2) 形式，我们给出了一个简单的列选择算法，并证明其达到了$(\frac{k}{2}+1+\frac{k}{2(2^k-1)})$的近似比。我们同时证明了这个近似比是紧的，即我们的分析不可改进。对于布尔形式，我们给出了另一个算法，并证明其达到了 $2^{k-1}+1$的近似比。当$k = 1$时，两个算法均是2-近似，与这个特例下的现有结果近似比相同[1, 2]。对于常数 k, 两个算法的时间复杂度均为矩阵大小的多项式时间。我们同时证明了二元矩阵的低秩近似问题即使在$k = 1$的特例下也是 NP-难的，解决了 [3] 中的一个猜想。

本文中的算法均为相应问题的首个具有近似比上界分析的算法，具有较为显著的理论意义。由于$k = 1$无法满足低秩近似在实际应用中的需求，因此，我们在$k > 1$的推广是很有价值的。

但是同时我们也要注意到，对于稍大的k，我们的算法在时间复杂度上还有不足。为了使算法在实际工程中应用，还需要进一步降低时间复杂度。如果仅仅是为了实际使用，可以采用一些经验性方法或随机算法，将时间复杂度降低到可以接受的范围内，但这样将失去所有的理论保证。因此，在将来的工作中，寻求既具有近似比保证，时间复杂度又满足实际应用需求的算法将是一个重要的方向。

参考文献

[1] Ed. by J. F. E. IV *et al.* "*Mining discrete patterns via binary matrix factorization*". In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, **2009**: 757–766.

[2] P. Jiang *et al.* "*A clustering approach to constrained binary matrix factorization*". In: *Data Mining and Knowledge Discovery for Big Data*. Springer, **2014**: 281–303.

[3] M. Koyutürk and A. Grama. "*PROXIMUS: a framework for analyzing very high dimensional discrete-attributed datasets*". In: *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, **2003**: 147–156.

[4] H. Hotelling. "*Analysis of a complex of statistical variables into principal components.*" *Journal of educational psychology*, **1933**, *24*(6): 417.

[5] K. K. Karhunen. "*Über lineare Methoden in der Wahrscheinlichkeitsrechnung*". *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.* **1947**, *37*: 1–79.

[6] R. Belohlavek and V. Vychodil. "*Discovery of optimal factors in binary data via a novel method of matrix decomposition*". *Journal of Computer and System Sciences*, **2010**, *76*(1): 3–20.

[7] P. Miettinen *et al.* "*The discrete basis problem*". *Knowledge and Data Engineering, IEEE Transactions on*, **2008**, *20*(10): 1348–1362.

[8] J. Vaidya, V. Atluri and Q. Guo. "*The role mining problem: finding a minimal descriptive set of roles*". In: *Proceedings of the 12th ACM symposium on Access control models and technologies*, **2007**: 175–184.

[9] J. K. Seppänen, E. Bingham and H. Mannila. "*A simple algorithm for topic identification in 0–1 data*". In: *Knowledge Discovery in Databases: PKDD 2003*. Springer, **2003**: 423–434.

[10] T. Šingliar and M. Hauskrecht. "*Noisy-or component analysis and its application to link analysis*". *The Journal of Machine Learning Research*, **2006**, *7*: 2189–2213.

[11] A. A. Frolov *et al.* "*Boolean factor analysis by attractor neural network*". *Neural Networks, IEEE Transactions on*, **2007**, *18*(3): 698–707.

[12] C. Lucchese, S. Orlando and R. Perego. "*Mining Top-K Patterns from Binary Datasets in Presence of Noise*". In: *SDM*, **2010**: 165–176.

[13] M. Frank *et al.* "*Multi-assignment clustering for boolean data*". *The Journal of Machine Learning Research*, **2012**, *13*(1): 459–489.

[14] A. Yeredor. "*Independent component analysis over galois fields of prime order*". *Information Theory, IEEE Transactions on*, **2011**, *57*(8): 5342–5359.

[15] H. W. Gutch *et al.* "*ICA over finite fields: separability and algorithms*". *Signal Processing*, **2012**, *92*(8): 1796–1808.

[16] A. Painsky, S. Rosset and M. Feder. "*Generalized Independent Component Analysis Over Finite Alphabets*". *Information Theory, IEEE Transactions on*, **2015**.

[17] L. Stockmeyer. *The minimal set basis problem is NP-complete* [IBM Research Report], **1975**.

[18] P. Miettinen *et al.* "*The Discrete Basis Problem*". *IEEE Trans. Knowl. Data Eng*, **2008**, *20*(10): 1348–1362.

[19] H. Fleischner *et al.* "*Covering graphs with few complete bipartite subgraphs*". *Theor. Comput. Sci*, **2009**, *410*(21-23): 2045–2053.

[20] Ed. by M. Agrawal *et al.* "*Inapproximability of Maximum Weighted Edge Biclique and Its Applications*". In: *TAMC 2008*. Springer, **2008**: 282–293.

[21] D. S. Hochbaum and A. Pathria. "*Forest harvesting and minimum cuts: a new approach to handling spatial constraints*". *Forest Science*, **1997**, *43*(4): 544–554.

[22] S. Arora *et al.* "*The hardness of approximate optima in lattices, codes, and systems of linear equations*". In: *Foundations of Computer Science, 1993. Proceedings., 34th Annual Symposium on*, **1993**: 724–733.

[23] P. Berman and M. Karpinski. "*Approximating minimum unsatisfiability of linear equations*". In: *Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms*, **2002**: 514–516.

[24] N. Alon, R. Panigrahy and S. Yekhanin. "*Deterministic approximation algorithms for the nearest codeword problem*". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, **2009**: 339–351.

[25] R. M. Roth and K. Viswanathan. "*On the Hardness of Decoding the Gale-Berlekamp Code*". *IEEE Transactions on Information Theory*, **2008**, *54*(3): 1050–1060.

[26] N. Amit. *The Bicluster Graph Editing Problem* [M.Sc. Thesis], **2004**.

作者本科阶段发表论文

Chen Dan, Kristoffer Arnsfelt Hansen, He Jiang, Liwei Wang, Yuchen Zhou, *"On Low Rank Approximation of Binary Matrices"*, Submitted to NIPS 2016, arXiv preprint 1511.01699.5

致谢

燕园四年，如白驹过隙，匆匆而过。即将毕业之际，在此我要感谢每一个在学习，生活与科研中给予我帮助的人。

首先，要感谢我的导师王立威教授。非常荣幸能在本科阶段受到王老师的指导，他带领我走进了机器学习理论研究的大门，使我得以看到这门美丽的学科的冰山一角。正是在他的帮助下，我完成了自己的第一篇论文的推导、写作与投稿。除了科研和教学之外，王老师还在留学与未来发展上给了我许多有价值的建议。可以说他是我在北大四年中，对我帮助最多的人。

此外，也要感谢封举富教授和林宙辰教授，从他们的课程中我学到了很多。封老师的模式识别、林老师的数据分析和王老师的信息论是我在信科学习的课程中，最为精彩的三门课。这些课程让我完成了对机器学习和优化理论的入门，给我的科研打下了良好的基础。

本文的写作离不开我的合作者的讨论和帮助。感谢周宇宸同学，本文中最重要的定理1的证明是我们共同完成的。这个定理十分不容易，本文中的长篇证明没有他的合作很难最终成型。感谢丹麦Aarhus大学的Kristoffer Arnsfelt Hansen教授，他帮助我们完成了本文第二、第三章中一些定理的证明。同时，也要感谢姜和同学在本文写作、投稿过程中的协助。

加入实验室两年来，我在科研方面有了很大的进步。特别感谢张佳琦师兄，郑凯师兄，谭子涵师兄和牟文龙师弟，他们给我的建议让我增长了见识，我们之间的讨论也十分有益。也要感谢实验室的其他所有同学，和你们一起度过了难忘的两年时光。

在本科期间，我在微软亚洲研究院进行了为期半年的科研实习。要感谢微软亚洲研究院的陈薇老师和贺笛师兄给我的指导和帮助。

在出国申请阶段，我得到了许多师兄和同学的帮助。感谢金驰师兄、张弛丞师兄、严松柏师兄在申请期间给我的指导和建议。感谢刘垚、周宇宸、姜和、刘径舟、刘易等许多同学在申请期间的讨论和帮助。

感谢孙泽远师兄，他一直给我有价值的经验和建议，帮助我在来到信科的初期完成了顺利的过渡。感谢我的好朋友们，和你们一起度过了愉快的时光。特别感谢两位高中至今的老朋友，史杨勍惟和尤之一，感谢你们长期以来的支持。

感谢北京大学，希望北大早日成为真正的世界一流大学。

最后，感谢家人对我长期的关心，我在北大一切都好。

北京大学学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

论文作者签名：　　　　　　日期：　　年　　月　　日

学位论文使用授权说明

（必须装订在提交学校图书馆的印刷本）

本人完全了解北京大学关于收集、保存、使用学位论文的规定，即：

- 按照学校要求提交学位论文的印刷本和电子版本；
- 学校有权保存学位论文的印刷本和电子版，并提供目录检索与阅览服务，在校园网上提供服务；
- 学校可以采用影印、缩印、数字化或其它复制手段保存论文；
- 因某种特殊原因需要延迟发布学位论文电子版，

（保密论文在解密后遵守此规定）

论文作者签名：　　　　　　导师签名：　　　　　　日期：　　年　　月　　日

版权声明