arXiv:2305.01069v3 [cs.DS] 9 Jul 2023

通过主划分序列近似次模 k-划分*

Karthekeyan Chandrasekaran Weihang Wang

摘要

在次模 k-划分问题中，输入是一个次模函数 $f\colon 2^V \to \mathbb{R}_{\geq 0}$（由评估预言机给出）、一个正整数 k，目标是将底集 V划分为k个非空部分 $V_1, V_2, \ldots, V_k$，以最小化 $\sum_{i=1}^{k} f(V_i)$。Narayanan、Roy 和 Patkar [17]基于主划分序列设计了一种用于次模 k-划分的算法，并证明对于图割函数这一特殊情形，该算法的近似比为 2（这一结果随后由 Ravi 和 Sinha [21]重新发现）。在本研究中，我们考察了他们的算法在三个次模函数子族——即单调、对称和正交子模函数——上的近似比，并得出了以下结果：

1. 他们的算法对于单调次模 k-划分的近似比为 $4/3$。这一结果改进了此前通过其他算法已知可实现的单调次模 k-划分的 2 近似比。此外，我们的 $4/3$ 上界与近期在多项式数量的函数求值查询 [22]条件下证明的下界相匹配。我们的 $4/3$ 上界也是某个图分区问题首次取得超过 2 的改进，该问题是单调次模 k-划分的一个特例。2. 他们的算法对于对称次模 k-划分的近似比为 2。这一结果将他们的近似比分析推广到了图割函数之外。3. 他们的算法对于正交子模次模 k-划分的近似比为 2。

我们还构造了一个例子，说明他们的算法对于任意次模函数的近似比为 $\Omega(n/k)$。

1 引言

若对每个 $A,\ B \subseteq V$，集合函数 $f\colon 2^V \to \mathbb{R}$ 满足 $f(A) + f(B) \geq f(A \cap B) + f(A \cup B)$，则称其为次模函数。函数 $f\colon 2^V \to \mathbb{R}$ 的评估预言机以子集 $A \subseteq V$ 为输入，并返回 $f(A)$。我们考虑如下定义的次模 k-划分 问题：输入包括定义在有限底集 V 上的非负次模函数 $f\colon 2^V \to \mathbb{R}_{\geq 0}$（通过评估预言机给出）以及一个整数 $k \geq 2$。目标是将 V 划分为 k 个非空部分 $V_1, V_2, \ldots, V_k$，使 $\sum_{i=1}^{k} f(V_i)$ 最小。也就是说，目标是计算

$$\min \left\{ \sum_{i \in [k]} f(V_i) : \ V_1, V_2, \ldots, V_k \text{ is a partition of } V,\ V_i \neq \emptyset \ \forall i \in [k] \right\}.$$

全文中，我们假设输入的次模函数非负，并用 n表示底集 V的大小。如果 $k = 2$，则该问题可化简为经典的次模最小化问题。我们强调，我们关注的是输入中包含 k 时的次模 k-分区问题（关于固定常数 k下的问题讨论，见 [4]）。次模 k-划分刻画了若干有趣的划分问题，我们将在下文讨论其中一些特殊情形。对于任意次模函数，该问题是 NP难的 [9], ; 在多项式数量的函数求值查询下，无法获得 $(2-\epsilon)$-近似

*伊利诺伊大学厄巴纳-香槟分校。电子邮件：{karthe, weihang3}@illinois.edu。部分经费由美国国家科学基金会资助，项目编号为 CCF-1814613 和 CCF-1907937。

查询 [22],；假设指数时间假设 [5], 成立，对于任意常数 c，都无法获得 $n^{1/(\log\log n)^c}$-近似，而目前已知的最佳近似比为 $O(k)$ [18, 25]。

在本文中，我们关注针对次模函数子族的次模 k-划分问题，即单调、对称和正交子模次模函数。集合函数 $f\colon 2^V \to \mathbb{R}$是

1. 若对任意 $A \subseteq B \subseteq V$， $f(B) \geq f(A)$，则称其为单调函数；2. 若对任意 $A \subseteq V$， $f(A) = f(V - A)$，则称其为对称函数；3. 若对任意 A， $B \subseteq V$，$f(A) + f(B) \geq f(A - B) + f(B - A)$，则称其为正交子模函数。

如果输入的次模函数是单调/对称/正交子模的，我们就将相应的次模 k-划分问题称为单调/对称/正交子模次模 k-划分。需要指出的是，单调次模函数和对称次模函数也都是正交子模的[1]。因此，正交子模次模 k-划分问题同时推广了单调次模 k-划分和对称次模 k-划分问题。下面我们将讨论对称/单调/正交子模次模 k-划分的近似情况，以及其中一些广为人知的特殊情形（不同方法所达到的对称/单调/正交子模次模 k-划分近似因子汇总见表 1）。

单调次模 k-划分。单调次模 k-划分问题的特殊情形包括拟阵 k-划分和覆盖 k-划分——这里关注的次模函数分别是拟阵秩函数和覆盖函数。拟阵 k-划分涵盖了若干有趣的问题：例如，(1) 将给定矩阵的列划分为 k 个非空部分，使这些部分所张成子空间的维数之和最小；(2) 将给定图的边划分为 k 个非空部分，使这些部分形成的连通分量数之和最大化。覆盖 k-划分也涵盖了若干有趣的问题：例如，(3) 将给定图的顶点划分为 k 个非空部分 V_1，V_2，……，V_k 以最小化$\sum_{i=1}^{k} f(V_i)$，其中 $f(P)$是与顶点子集 $P \subseteq V$相邻的边数。为了更好地理解求解/近似单调次模 k-划分的难度，我们鼓励读者在继续阅读之前，先简要思考矩阵列划分问题这一具体特殊情形（即上文所述的问题(1)，它乍看起来是一个线性代数问题）。

单调次模 k-划分是 NP难的 [22]。此外，它存在一种简单（且快速）的 $(2 - 1/k)$-近似算法，以下称为成本最低的单元素集合划分算法：返回划分 $V_1 := \{v_1\}$，$V_2 := \{v_2\}$，……，$V_{k-1} := \{v_{k-1}\}$，$V_k := V - \{v_1, \ldots, v_{k-1}\}$，其中底集的$n$个元素按 v_1，……，v_n排序，使得 $f(\{v_1\}) \leq f(\{v_2\}) \leq \ldots \leq f(\{v_n\})$。Santiago [22] 证明了这是一个 2-近似算法。其分析可以扩展为证明它实际上是一个 $(2-1/k)$-近似算法[2]，而且这是该算法所能达到的最佳近似比[3]。另外，[25]中提出的贪心分裂算法可达到 $(2-2/k)$-近似。在不可近似性方面，Santiago [22] 证明，不存在任何算法能够通过多项式数量的函数求值查询，对每个常数 $\epsilon > 0$都获得 $(4/3 - \epsilon)$-近似。

对称次模 k-划分。对称次模 k-划分问题的著名特例包括图 k-割和超图 k-划分——这里关注的次模函数分别是显式给定的图和超图的割函数。图 k-割是 NP完全问题 [9]，并且

[1]事实上，由于对于每个 $A, B \subseteq V$， $f(A) \geq f(A - B)$ 且 $f(B) \geq f(B - A)$，单调函数是正交子模的。对称次模函数也是正交子模的：对于每个 $A\ B \subseteq V$ $f\ A + f\ B = f\ V - A + f\ B \geq f\ V - A \cup B +,,$ 我 $f\ V - A \cap B = f\ V - A - B + f\ B - A = f\ A - B + f\ B - A$ 们有 ( ) ( ) ( ) ( ) (( )) (( )) (( )) ( ) ( ) ( )。[2]最廉价单元素划分算法返回的解，其成本为 $f(V - V_k) + \sum_{i=1}^{k-1} f(\{v_i\}) \leq f(V) + \sum_{i=1}^{k-1} f(\{v_i\}) \leq f(V) + (1 - 1/k) \sum_{i=1}^{k} f(\{v_i\}) \leq (2 - 1/k) \max\{f(V) \sum_{i=1}^{k} f(\{v_i\})\}$，而最优 k-划分的成本至少为 $\max\{f(V) \sum_{i=1}^{k} f(\{v_i\})\}$。最优 k-划分 $V_1^*, \ldots, V_k^*$ 的成本下界来自非负性和次模性，即 $\sum_{i=1}^{k} f(V_i^*) \geq f(V)$；此外，若对最优划分的编号方式满足对于所有 $i \in [k-1]$，$\min\{j \in [n] : v_j \in V_i^*\} \leq \min\{j \in [n] : v_j \in V_{i+1}^*\}$，则由单调性可得 $f(V_i^*) \geq f(\{v_i\})$，因此，$\sum_{i=1}^{k} f(V_i^*) \geq \sum_{i=1}^{k} f(\{v_i\})$。[3]如以下例子所示，最廉价单元素划分算法所能达到的最佳近似比为 $2 - 1/k$：设 f为底集 S 上关于 k-划分 $\{S_1, \ldots, S_k\}$ 的划分拟阵的秩函数，其中对于所有 $i \in [k]$，$|S_i| \geq 2$。那么，该算法可能返回 $\{\{s_1\}, \{s_2\}, \ldots, \{s_{k-1}\}, S - \{s_1, \ldots, s_{k-1}\}\}$，其中对于所有 $i \in [k-1]$，$s_i \in S_i$；这个 k-划分的目标值为 $2k - 1$，而划分 $\{S_1, \ldots, S_k\}$ 的目标值为 k。

根据小集合扩展假设 [11]，对于任意常数 $\epsilon > 0$，都不存在多项式时间的（$2-\epsilon$）-近似算法。针对图 k-割，已有几种实现 2-近似的方法——(i) 贪心分裂方法 [23]，(ii) 基于 Gomory-Hu 树的方法 [24]，(iii) 基于极端集合的方法[12]，(iv) 基于主划分序列的方法 [2,17,21]，以及 (v) 基于覆盖线性规划的方法 [6,14,20]。贪心分割、Gomory-Hu 树和基于极端集合的方法可实现（$2-2/k$）-近似，而基于主划分序列和覆盖线性规划的方法可为图 k-割实现（$2-2/n$）-近似。研究还表明，针对图 k-割的主划分序列方法与覆盖线性规划方法彼此相关 [6]。基于主划分序列的方法是我们主要关注的算法，我们将在第 2 节详细讨论。

对于更一般的对称次模 k-划分问题，上一段讨论的图 k-割问题的两种方法已被推广，可得到 2-近似——贪心分裂方法 [25] 以及 Gomory-Hu 树方法可得到 $(2-2/k)$-近似。分析基于主划分序列的方法在对称次模 k-划分问题中的近似比，是推动我们开展这项工作的动机之一。在不可近似性方面，Santiago [22]证明，不存在一种算法能够通过多项式数量的函数求值查询，对每个常数 $\epsilon > 0$都得到 $(2-\epsilon)$-近似。

正模次模 k-划分。我们所熟悉的正模次模函数中，唯一自然的一类是对称次模函数和单调次模函数，以及它们的正线性组合。如前所述，正模次模 k-划分是对称次模 k-划分和单调次模 k-划分的统一推广。据作者所知，文献此前尚未研究过正模次模 k-划分，也没有针对正模次模 k-划分的专门算法或对现有算法的近似比分析。对 [25] 中提出的贪心分裂算法的分析稍作修改即可证明，该算法对正模次模k-划分可达到 $(3-2/k)$-近似——为简洁起见，我们不再给出这一分析。在不可近似性方面，由于对称次模函数也属于正模次模函数，对称次模 k-划分的下界同样适用于正模次模 k-划分；也就是说，不存在一种针对正模次模 k-划分的算法，能够通过多项式数量的函数求值查询，对每个常数 $\epsilon > 0$都得到 $(2-\epsilon)$-近似。

1.1 我们的结果

在这项工作中，我们研究 Narayanan、Roy 和 Patkar 的 [17]基于主划分序列的算法，用于次模 k-划分。他们证明，该算法对图 k-割可达到 2-近似（Ravi 和 Sinha [21]随后也独立得出了这一结果）。我们得到以下结果：

1. 他们的算法对单调次模 k-划分可达到 $4/3$-近似。这一结果改进了已知可通过两种不同算法达到的 2 倍近似：最廉价单元素划分算法和贪心分裂算法。此外，我们的上界 $4/3$ 与 Santiago [22]证明的下界相匹配。我们稍后将讨论这一上界结果的意义。2. 他们的算法对对称次模 k-划分可达到 2-近似。这一近似因子与已知可通过另外两种算法达到的 2 倍近似相匹配：贪心分裂算法和基于 Gomory-Hu 树的算法；同时也与下界 [11,22]相匹配。我们在此的贡献是将 [17,21] 的分析推广到图割函数之外。3. 他们的算法对正交子模次模 k-划分可达到 2-近似。这一结果改进了已知可通过贪心分裂算法达到的 3 倍近似，并与 Santiago [22]证明的 2 倍下界相匹配。

比较结果见表 1。图 k-割是对称次模 k-划分/正交子模次模 k-划分中一个研究充分的特殊情形，因此我们将其列为表格的最后一列以供比较。主划分序列对应行中的近似因子是我们工作的主要结果。表格最后一行列出了已知的近似比下界

以供比较。图割函数的下界基于小集合扩展假设 [11]，其余下界则基于多项式数量的函数求值查询。表格中的短横线表示该方法要么无法推广，要么尚未分析该方法对该子类的近似比。

我们为每个次模函数子族构造了匹配的下界，从而补充了其算法近似比的上界。我们的结果表明，对于广泛的次模函数子族，基于主划分序列的算法达到了可能的最佳近似比，从而展示了该算法的强大能力和适用性。另一方面，通过一个下界构造，我们证明了其针对任意次模函数的算法的近似比为 $\Omega(n/k)$。该构造表明，其基于主划分序列的算法无法直接将次模 k-划分的近似比提升到当前最佳 $O(k)$以上。

我们简要讨论我们的 $4/3$-近似结果对于单调次模 k-划分的重要意义。首先，在我们的结果之前，尚无已知的次模函数族能够使次模k-划分问题达到优于 2 的近似比。我们关于单调次模函数的结果打破了这一 2 近似比障碍，适用于广泛的次模函数族。其次，我们关于单调次模 k-划分的结果，即使对于我们现在描述的一个图分区问题，也带来了新的近似结果。对于一个边权为 w 的图 $G = (V, E)$：$E \to \mathbb{R}_+$，考虑函数d, f：$2^V \to \mathbb{R}_+$，其定义为 $d(S) := w(\delta(S))$ 和 $f(S) := w(E[S]) + w(\delta(S))$，对每个 $S \subseteq V$ 均成立。其中，$\delta(S)$ 表示恰有一个端点属于 S 的边的集合，$E[S]$表示两个端点均属于 S 的边的集合，并且 $w(F) := \sum_{e \in F} w(e)$ 对每个 $F \subseteq E$ 均成立。函数 d 是该图的切割函数，且为对称次模函数。函数 f 是该图的覆盖函数，且为单调次模函数。对于函数 d 的次模 k-划分称为图 k-切割，并且已知在小集合扩展假设 [11] 下，图 k-切割不允许 $(2-\epsilon)$-近似。相比之下，我们的结果表明，图中的覆盖 k-划分，即对于函数 f 的次模 k-划分，允许达到 $4/3$-近似。我们注意到，图中的覆盖 k-划分是 NP难的 [22]，其可近似性是一个引人关注的开放问题。

	单调次模函数	对称次模函数	正交子模次模函数	图割函数
贪心分割	$2-2/k$ [25]	$2-2/k$ [25]	$3-2/k$ [25]*	$2-2/k$ [23]
极端集合	—	—	—	$2-2/k$ [12]
Gomory-Hu 树	—	$2-2/k$ [民间常识]	—	$2-2/k$ [24]
覆盖线性规划	—	—	—	$2-2/k$ [6, 14]
成本最低的单元素集合分区	$2-1/k$ [22]	—	—	—
主分割序列	$4/3-4/(9n+3)$ (定理 3.1)	$2-2/n$ （定理 3.2）	$2-2/(n+1)$ （定理 3.3）	$2-2/n$ [17, 21]
下界	$4/3-\epsilon$ [22]	$2-\epsilon$ [22]	$2-\epsilon$ [22]	$2-\epsilon$ [11]

表 1：使用不同方法进行对称/单调/正交子模次模 k 划分的近似因子。第一行中带星号的结果可通过对 [25] 中给出的对称次模函数近似比的已知分析稍作修改得到。

结构安排。 我们在第 2 节讨论基于主划分序列的算法。在第 3 节中，我们针对次模函数的三个子族，分析该算法的近似比，并给出与之匹配的下界构造。在第 4 节中，我们展示了一个次模 k-划分实例，在该实例中，该算法所达到的近似比为 $\Omega(n/k)$。

1.2 相关工作

Narayanan、Roy 和 Patkar 提出了基于主划分序列的算法，用于次模 k-划分 [17]。我们将在第 2 节正式定义次模函数的主划分序列，并介绍他们的算法。他们分析了该算法针对以下问题的近似比：

超图中两种 k-划分问题的变体。这两种变体并非对称/单调/正交子模次模 k-划分的特殊情况，也与我们的工作没有直接关系。不过，我们介绍这两种变体，以彰显基于主划分序列的方法的通用性，并进一步阐明 Narayanan、Roy 和 Patkar 的研究成果；这些成果似乎尚未得到学界广泛认识。给定一个超图 $H = (V, E)$、一个超边成本函数 $c : E \to \mathbb{R}_+$以及一个整数 k，目标是找到 V的一个划分 $\mathcal{P} := \{V_1, V_2, \ldots, V_k\}$，将其划分为 k 个非空部分，使感兴趣的目标值最小：

1. 如果目标值是 $\mathcal{P}$ 中至少与两个部分相交的超边的成本之和，那么该问题称为超图 k-割。2. 如果目标值是相对于划分 $\mathcal{P}$ 的超边成本之和，其中超边e 相对于 $\mathcal{P}$ 的成本为 $c(e)(\ell - 1)$，其中 ℓ 是 $\mathcal{P}$ 中与 e相交的部分数，那么该问题称为归一化覆盖 k-划分[4]。

Narayanan、Roy 和 Patkar [17] 证明，他们的基于主划分序列的算法对超图 k-割可达到 $r(1 - 1/n)$ 近似，其中 r 是最大超边的大小， n 是输入超图中的顶点数；对归一化覆盖 k-划分可达到 $(2 - 2/n)$ 近似。由他们的两个结果均可推出，基于主划分序列的算法对图 k-割可达到 $(2 - 2/n)$ 近似。他们针对图 k-割提出的基于主划分序列的算法，等价于 Barahona [2]提出的拉格朗日松弛方法。Ravi 和 Sinha [21] 重新发现了基于主划分序列的算法对图 k-割的近似比至多为 2；Baïou 和 Barahona [1]则重新发现了该算法对超图 k-割的近似比至多为 r 。

我们指出，对 [25] 中提出的贪心分裂算法的分析稍作修改即可证明，该算法对于归一化覆盖 k-划分和超图 k-划分可达到 $(2 - 2/k)$-近似。需要注意的是，超图 k-划分是对称次模 k-划分的一个特例，且不同于超图 k-割（关于超图 k-划分的定义，见引言开头对对称次模 k-划分的讨论）。关于不可近似性，已知在指数时间假设 [5]成立的前提下，超图 k-割无法达到 $n^{1/(\log \log n)^c}$ 的近似比，其中 c 是一个常数。更一般的次模 k-划分问题（引言中提及）的最佳不可近似性结果，源自超图 k-割的这一不可近似性结果。

2 基于主划分序列的算法

本节回顾 Narayanan、Roy 和 Patkar [17] 为次模 k-划分设计的基于主划分序列的算法。首先介绍一些记号。在本文中，集合 S 的划分 定义为由 S 的两两不相交非空 子集组成的集合，且这些子集的并集为 S；集合 S 的k-划分 定义为 S 的一个划分，且恰有 k 个部分。对于集合 S 的两个不同划分 $\mathcal{P}$ 和 $\mathcal{Q}$，若 $\mathcal{Q}$ 的每个部分都完全包含于 $\mathcal{P}$ 的某个部分中，则称 $\mathcal{Q}$ 细化 $\mathcal{P}$（等价地， $\mathcal{P}$ 是 $\mathcal{Q}$ 的粗化）。对于集合 S 的两个不同划分 $\mathcal{P}$ 和 $\mathcal{Q}$，若存在一个部分 $P \in \mathcal{P}$，使得 $P \notin \mathcal{Q}$，且 $Q \in \mathcal{Q}$ 的每个部分都满足 $Q \subsetneq P$ 或 $Q \in \mathcal{P}$（即， Q 是部分 P 的真子集，或 Q 是划分 $\mathcal{P}$ 的一个部分），则称 $\mathcal{Q}$ 是通过仅细化 $\mathcal{P}$ 的一个部分而由 $\mathcal{P}$ 得到的；我们将这样的部分 $P \in \mathcal{P}$ 称为被 $\mathcal{Q}$ 细化的部分。

设 $f : 2^V \to \mathbb{R}$ 是定义在底集 V 上的集合函数。对于 V 的子集构成的集合族 $\mathcal{P}$，我们记作$f(\mathcal{P}) := \sum_{P \in \mathcal{P}} f(P)$。如果对于每个 V 的 k-划分 $\mathcal{Q}$，都有 $f(\mathcal{P}) \leq f(\mathcal{Q})$，那么我们称划分 $\mathcal{P} = \{P_1, \ldots, P_k\}$ 为最优的 k-划分。对于底集 V 的一个划分 $\mathcal{P}$ 以及函数 $g_f : \mathbb{R}_{\geq 0} \to \mathbb{R}$，我们将函数 $g_{f,\mathcal{P}} : \mathbb{R}_{\geq 0} \to \mathbb{R}$ 定义如下：

$$g_{f,\mathcal{P}}(b) := f(\mathcal{P}) - b|\mathcal{P}| \text{ and}$$

[4]我们引入这一术语，是因为该问题等价于寻找底集 V的一个划分 V_1， V_2， ……， V_k，以最小化 $\sum_{i=1}^{k} f(V_i) - f(V)$，其中 $f : 2^V \to \mathbb{R}_+$ 是一个显式给定的覆盖函数（每个覆盖函数都可以用一个超图 [3]唯一表示）。我们将减去 $f(V)$视为目标归一化，因为它是各部分函数值之和的一个平凡下界：对于每个 k-划分 V_1， V_2， ……， V_k，都有 $\sum_{i=1}^{k} f(V_i) \geq f(V)$，因为 f是覆盖函数。

$$g_f(b) := \text{最小}\{g_{f,\mathcal{P}}(b) : \mathcal{P}\text{是 } V\}\text{的一个划分}$$

如果函数 f 的含义可由上下文明确，我们就省略下标 f，分别改记为 $g_{\mathcal{P}}$ 和 g。根据定义，函数 g_f 是分段线性的。可以证明，g_f 至多有 $|V|-1$ 个断点。下一个定理表明，如果函数 $f\colon 2^V \to \mathbb{R}$ 是次模函数，那么存在一个划分序列，它在断点处取得 g_f 函数值，并且具有良好的结构；此外，只要能够访问次模函数 f 的评估预言机，就可以在多项式时间内计算出这些断点及这样的划分序列。我们强调，该定理适用于任意次模函数（其取值不一定非负）。

定理 2.1 ([15,17])。设 $f\colon 2^V \to \mathbb{R}$ 是底集 V 上的次模函数。那么，存在一个划分序列 $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_r$，其元素均为 V 的划分，以及一系列值 $b_1, b_2, \ldots, b_{r-1}$，使得

1. $\mathcal{P}_1 = \{V\}$ 和 $\mathcal{P}_r = \{\{v\} : v \in V\}$，*2.* 对于每个 $j \in [r-1]$，划分 $\mathcal{P}_{j+1}$ 是通过仅细化 $\mathcal{P}_j$ 的一个部分，从 $\mathcal{P}_j$ 得到的，*3.* $b_1 < b_2 < \ldots < b_{r-1}$，*4.* 对每个 $j \in [r-1]$，$g(b_j) = g_{\mathcal{P}_j}(b_j) = g_{\mathcal{P}_{j+1}}(b_j)$，以及 *5.* 对所有 $b \in$，$g(b) = g_{\mathcal{P}_1}(b)$（$(-\infty, b_1]$，$g(b) = g_{\mathcal{P}_{j+1}}(b)$ 对每个 $j \in [r-2]$均成立，且 $b \in [b_j, b_{j+1}]$），并且对所有 $b \in [b_{r-1}, \infty$，$g(b) = g_{\mathcal{P}_r}(b))$。

此外，给定对次模函数 f 的评估预言机访问权限，可以在多项式时间内计算出这样一组划分 $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_r$ 和相应的值 $b_1, b_2, \ldots, b_{r-1}$，其中这些划分是对 V 的划分。

对于次模函数 $f\colon 2^V \to \mathbb{R}$，我们将满足定理 2.1 中所给条件的划分序列 $\mathcal{P}_1$，$\mathcal{P}_2$，$\ldots$，$\mathcal{P}_r$ 以及值序列 b_1，b_2，$\ldots$，b_{r-1} 分别称为 f 的**主划分序列**和**临界值序列**。我们注意到，由于采用了反向索引顺序并施加了条件 2，此定义与 [15,17] 中的定义不同——这些论文中的证明也表明条件 2 成立（另见 [21]）。文献中将次模函数的主划分序列称为次模函数的划分主格，因为与该划分序列相关联的结构是一个格。本文选择称其为主划分序列，因为该序列足以满足我们的需要。关于次模函数的划分主格及其计算，读者可参阅 [2,7,8,10,13,15–17,19]。

现在我们讨论 Narayanan、Roy 和 Patkar [17] 提出的基于主划分序列的次模 k-划分算法。该算法计算一个满足定理 2.1 中所有条件的主划分序列。如果该序列中包含一个恰好具有 k 个部分的划分，则算法返回这一 k-划分。否则，算法通过细化序列中部分数量小于 k 且数量最大的划分，返回一个 k-划分。该细化过程以序列中部分数量大于 k 且数量最少的划分为基础。细化过程的正式描述见算法 1。由于满足定理 2.1 条件的序列 $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_r$ 可以在多项式时间内计算出来，因此算法 1 确实可以在多项式时间内实现。按照设计，该算法返回一个 k-划分。本文余下部分将重点分析该算法的近似比。

为了构造能够体现算法 1 近似比紧下界的示例，我们需要下面的命题。该命题确定了一种特殊情形：在此情形下，主分割是唯一的，并且仅由两个划分组成——即单元素集合划分和仅包含一个部分的划分。

命题 2.1。令 $f\colon 2^V \to \mathbb{R}_{\geq 0}$ 为一个非负次模函数。假设对于每个划分 $\mathcal{P} \neq \mathcal{Q}, \{V\}$，其中 $\mathcal{Q} := \{\{v\} : v \in V\}$，函数 f 满足

$$\frac{f(\mathcal{P}) - f(V)}{|\mathcal{P}| - 1} > \frac{f(\mathcal{Q}) - f(V)}{|V| - 1}.$$

然后，f 的主划分序列是 $\{V\}, \mathcal{Q}$。

算法 1 基于主划分序列的次模 k-划分算法

输入： 一个通过求值预言机给出的次模函数 $f\colon\ 2^V \to \mathbb{R}$ 和一个整数 $k \geq 2$。**输出：** 将 V 划分为 k 个部分的 $\mathcal{P}$。使用定理 2.1 计算次模函数 f 的主划分序列 $\mathcal{P}_1$、$\mathcal{P}_2$、.....、$\mathcal{P}_r$，使其满足该定理中所述的所有条件。**如果** $|\mathcal{P}_j| = k$ 对某个 $j \in [r]$，**则**返回 $\mathcal{P}\colon = \mathcal{P}_j$。**结束如果**令 $i \in \{2, 3, \ldots, r\}$，使得 $|\mathcal{P}_{i-1}| < k < |\mathcal{P}_i|$。令 $S \in \mathcal{P}_{i-1}$ 为由 $\mathcal{P}_i$ 细化的部分，并令 $\mathcal{P}'$ 为包含于 S 中的 $\mathcal{P}_i$ 的部分。令 $\mathcal{P}' = \{B_1, \ldots, B_{|\mathcal{P}'|}\}$ 使得 $f(B_1) \leq f(B_2) \leq \ldots \leq f(B_{|\mathcal{P}'|})$。返回 $\mathcal{P}\colon = (\mathcal{P}_{i-1} \backslash \{S\}) \cup \left\{B_1, B_2, \ldots, B_{k-|\mathcal{P}_{i-1}|}\right\} \cup \left\{\bigcup_{j=k-|\mathcal{P}_{i-1}|+1}^{|\mathcal{P}'|} B_j\right\}$。

证明。令 $n := |V|$。只需证明，对于 V 的每个满足 $\mathcal{P} \neq \{V\}, \mathcal{Q}$ 的划分 $\mathcal{P}$，我们有

$$f(\mathcal{P}) - b|\mathcal{P}| > \min\{f(V) - b, f(\mathcal{Q}) - b \cdot n\} \quad \forall b \in \mathbb{R} \tag{1}$$

这足以确保除 $\{V\}$ 和 $\mathcal{Q}$ 之外，没有其他划分满足定理 2.1 中的条件 4 和 5。根据假设，我们有

$$\frac{f(\mathcal{P}) - f(V)}{|\mathcal{P}| - 1} > \frac{f(\mathcal{Q}) - f(V)}{n - 1},$$

这等价于

$$f(\mathcal{P}) - b'|\mathcal{P}| > f(V) - b', \tag{2}$$

其中，对于每个划分 $\mathcal{P} \neq \{V\}, \mathcal{Q}$，$b' = (f(\mathcal{Q}) - f))/(n-1)$。现在，假设不等式 (1) 对某个划分 $\mathcal{P}$ 和某个 $b \in \mathbb{R}$ 不成立，则

$$f(\mathcal{P}) - b|\mathcal{P}| \leq f(V) - b \text{ and} \tag{3}$$
$$f(\mathcal{P}) - b|\mathcal{P}| \leq f(\mathcal{Q}) - bn. \tag{4}$$

因此，我们有

$$\begin{aligned}
(b - b')|\mathcal{P}| &= (f(\mathcal{P}) - b'|\mathcal{P}|) - (f(\mathcal{P}) - b|\mathcal{P}|) \\
&> (f(V) - b') - (f(\mathcal{P}) - b|\mathcal{P}|) \qquad \text{(using inequality (2))} \\
&\geq (f(V) - b') - (f(V) - b) \qquad \text{(using inequality (3))} \\
&= b - b'.
\end{aligned}$$

这意味着 $b - b' > 0$。此外，我们还有

$$\begin{aligned}
(b - b')|\mathcal{P}| &= (f(\mathcal{P}) - b'|\mathcal{P}|) - (f(\mathcal{P}) - b|\mathcal{P}|) \\
&> (f(V) - b') - (f(\mathcal{P}) - b|\mathcal{P}|) \qquad \text{(using inequality (2))} \\
&\geq (f(V) - b') - (f(\mathcal{Q}) - bn) \qquad \text{(using inequality (4))} \\
&= (f(\mathcal{Q}) - b'n) - (f(\mathcal{Q}) - bn) \qquad \text{(by definition of } b') \\
&= (b - b')n.
\end{aligned}$$

由于 $b - b' > 0$，可得 $|\mathcal{P}| > n$，这与 $\mathcal{P}$ 是 V 的一个划分这一事实矛盾。因此，通过反证法可得不等式 (1) 成立。 □

3 近似因子分析

在本节中，我们分析算法 1 对各种次模函数子族的近似比。我们陈述并证明一些对所有次模函数都有用的引理（引理

3.1 和 3.2）。我们在第 3.1 节中分析算法 1 针对单调次模函数的近似比，在第 3.2 节中分析其针对对称次模函数的近似比，并在第 3.3 节中分析其针对正模次模函数的近似比。每个小节的结尾，我们都会对该算法近似比的紧确性作出说明。

我们的第一个引理指出了算法 1 返回最优 k-划分的一种特殊情形。Narayanan、Roy 和 Patkar 也指出了这一特殊情形。值得注意的是，以下引理对任意次模函数（可能非非负）均成立。

引理 3.1 ([17]). 设 $k \geq 2$ 为整数， $f: 2^V \to \mathbb{R}$ 为底集 V上的次模函数，且 $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_r$ 为满足定理 *2.1* 条件的次模函数 f的主划分序列。若存在 $j \in [r]$ 使得 $|\mathcal{P}_j| = k$，则 $\mathcal{P}_j$ 是最优的 k-划分。

证明。 设 $\mathcal{P}^*$是使 $f(\mathcal{P}^*)$ 最小的 k-划分，且设 b_j为满足 $g(b_j) = g_{\mathcal{P}_j}(b_j)$ 的值。那么，

$$f(\mathcal{P}^*) - b_j \cdot k = g_{\mathcal{P}^*}(b_j) \geq g(b_j) = g_{\mathcal{P}_j}(b_j) = f(\mathcal{P}_j) - b_j|\mathcal{P}_j| = f(\mathcal{P}_j) - b_j \cdot k,$$

因此， $f(\mathcal{P}^*) \geq f(\mathcal{P}_j)$。所以， $\mathcal{P}_j$确实是最优的 k-划分。 □

为了解决不存在满足 $j \in [r]$ 的 $|\mathcal{P}_j| = k$ 这一情况，我们需要下面的引理，它给出了最优值的两个下界。引理中的第一个下界适用于任意次模函数（可能为非非负函数），而第二个下界适用于非负次模函数。

引理 3.2。 设 $k \geq 2$ 为整数， $f: 2^V \to \mathbb{R}$ 是底集 V 上的次模函数， $\mathcal{P}^*$ 是使 $f(\mathcal{P}^*)$ 最小化的 V 的 k-划分，且 $\mathcal{P}_1$、$\mathcal{P}_2$、.....、$\mathcal{P}_r$ 是满足定理 *2.1* 条件的次模函数 f 的主划分序列。假设对所有 $j \in [r]$, 均有 $|\mathcal{P}_j| \neq k$ 。令 $\mathcal{P}_{i-1}$、$\mathcal{P}_i$ 为满足 $|\mathcal{P}_{i-1}| < k < |\mathcal{P}_i|$ 的划分。那么，

(i) $f(\mathcal{P}^*) \geq \frac{|\mathcal{P}_i| - k}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|} f(\mathcal{P}_{i-1}) + \frac{k - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|} f(\mathcal{P}_i)$ *and*

(ii) $f(\mathcal{P}^*) \geq f(\mathcal{P}_{i-1})$ *if* f *is non-negative.*

证明。 我们证明下面两个下界。

(i) 令 b_{i-1} 为满足 $g_{\mathcal{P}_{i-1}}(b_{i-1}) = g_{\mathcal{P}_i}(b_{i-1})$ 的值。于是，我们有

$$f(\mathcal{P}_{i-1}) - b_{i-1}|\mathcal{P}_{i-1}| = f(\mathcal{P}_i) - b_{i-1}|\mathcal{P}_i| \implies b_{i-1} = \frac{f(\mathcal{P}_i) - f(\mathcal{P}_{i-1})}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}. \tag{5}$$

根据定理 2.1 的条件 4，我们还有

$$\begin{aligned} f(\mathcal{P}^*) - b_{i-1} \cdot k = g_{\mathcal{P}^*}(b_{i-1}) \geq g(b_{i-1}) = g_{\mathcal{P}_i}(b_{i-1}) = f(\mathcal{P}_i) - b_{i-1}|\mathcal{P}_i| \\ \implies f(\mathcal{P}^*) \geq f(\mathcal{P}_i) + b_{i-1}(k - |\mathcal{P}_i|). \end{aligned} \tag{6}$$

结合 (5) 和 (6)，可得

$$\begin{aligned} f(\mathcal{P}^*) &\geq f(\mathcal{P}_i) + \frac{f(\mathcal{P}_i) - f(\mathcal{P}_{i-1})}{|P_i| - |P_{i-1}|}(k - |\mathcal{P}_i|) \\ &= \frac{|\mathcal{P}_i| - k}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|} f(\mathcal{P}_{i-1}) + \frac{k - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|} f(\mathcal{P}_i). \end{aligned}$$

(ii) 令 P_1^*、$\ldots$、P_k^* 为 $\mathcal{P}^*$ 的各个部分（顺序任意）。令 $k' := |\mathcal{P}_{i-1}|$。我们知道 $k' < k$。考虑由 $Q_1 := P_1^*$、$Q_2 := P_2^*$、$\ldots$、$Q_{k'-1} := P_{k'-1}^*$、$Q_{k'} := \cup_{j=k'}^{k} P_j^*$ 得到的 k'-划分 $\mathcal{Q}$。于是，

$$f(\mathcal{P}^*) = \sum_{i=1}^{k} f(P_i^*) \geq \left(\sum_{i=1}^{k'-1} f(P_i^*) \right) + f(\cup_{j=k'}^{k} P_j^*) = \sum_{i=1}^{k'} f(Q_i) = f(\mathcal{Q}).$$

上面的不等式源于次模性和非负性。划分 $\mathcal{Q}$ 是一个 k'-划分。根据引理 3.1，划分 $\mathcal{P}_{i-1}$ 是一个最优的 k'-划分。因此，

$$f(\mathcal{Q}) \geq f(\mathcal{P}_{i-1}).$$

上述两个不等式共同推出 $f(\mathcal{P}^*) \geq f(\mathcal{P}_{i-1})$。

□

3.1 单调次模函数

在本节中，我们界定单调次模 k划分的算法 1 的近似比。以下是本节的主要定理。

定理 3.1. 对于非负单调次模 k-划分，算法 *1* 的近似比为 $\frac{4}{3} - \frac{4}{9n+3}$，其中 n 是底集的大小。

算法 1 实现的 $4/3$ 渐近近似因子对于非负单调次模 k-划分而言是最佳可能的：对于每个常数 $\epsilon > 0$，不存在能够使用多项式数量的函数求值查询 [22] 实现 $(4/3 - \epsilon)$-近似的算法。证明该定理后，我们还将给出示例，以展示算法 1 的近似比的紧确性。定理 3.1 的证明可由下文给出的引理 3.1 和引理 3.3 得出。

引理 3.3. 设 $k \geq 2$ 为一个整数，$f : 2^V \to \mathbb{R}_{\geq 0}$ 为定义在大小为 n 的底集 V 上的非负单调次模函数，$\mathcal{P}^*$ 为使 $f(\mathcal{P}^*)$ 最小的 k-划分，$\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_r$ 为满足定理 *2.1* 条件的次模函数 f 的主划分序列。假设对所有 $j \in [r]$，均有 $|\mathcal{P}_j| \neq k$。那么，算法 *1* 返回的划分 $\mathcal{P}$ 满足

$$f(\mathcal{P}) \leq \left(\frac{4}{3} - \frac{4}{9n+3}\right) f(\mathcal{P}^*).$$

证明。设 $\mathcal{P}_{i-1}, \mathcal{P}_i$ 为满足 $|\mathcal{P}_{i-1}| < k < |\mathcal{P}_i|$的划分。令 S 和 $\mathcal{P}' = \{B_1, B_2, \ldots, B_{|\mathcal{P}'|}\}$ 如算法 1 中所定义。

首先，由于 $\cup_{j=k-|\mathcal{P}_{i-1}|+1}^{|\mathcal{P}'|} B_j \subseteq S$ 和 f 是单调的，我们有

$$f\left(\cup_{j=k-|\mathcal{P}_{i-1}|+1}^{|\mathcal{P}'|} B_j\right) \leq f(S).$$

其次，根据我们对 $B_1, B_2, \ldots, B_{k-|\mathcal{P}_{i-1}|}$的选择，我们知道

$$\sum_{j=1}^{k-|\mathcal{P}_{i-1}|} f(B_j) \leq \frac{k - |\mathcal{P}_{i-1}|}{|\mathcal{P}'|} f(\mathcal{P}').$$

因此，

$$\begin{aligned}
f(\mathcal{P}) &= f(\mathcal{P}_{i-1}) - f(S) + \sum_{j=1}^{k-|\mathcal{P}_{i-1}|} f(B_j) + f\left(\bigcup_{j=k-|\mathcal{P}_{i-1}|+1}^{|\mathcal{P}'|} B_j\right) \\
&\leq f(\mathcal{P}_{i-1}) + \sum_{j=1}^{k-|\mathcal{P}_{i-1}|} f(B_j) \\
&\leq f(\mathcal{P}_{i-1}) + \frac{k - |\mathcal{P}_{i-1}|}{|\mathcal{P}'|} f(\mathcal{P}') \\
&= f(\mathcal{P}_{i-1}) + \frac{k - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}(f(\mathcal{P}_i) + f(S) - f(\mathcal{P}_{i-1})),
\end{aligned}$$

其中，最后一个等式是由以下事实得出的：$f(\mathcal{P}_{i-1}) - f(S) + f(\mathcal{P}') = f(\mathcal{P}_i)$ 和 $|\mathcal{P}'| = |\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1$。重新整理可得

$$\begin{aligned}
f(\mathcal{P}) &\le \frac{f(\mathcal{P}_{i-1})}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}(|\mathcal{P}_i| - k + 1) + \frac{f(\mathcal{P}_i)}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}(k - |\mathcal{P}_{i-1}|) \\
&\quad + \frac{f(S)}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}(k - |\mathcal{P}_{i-1}|) \\
&= \left(\frac{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right)\left(\frac{f(\mathcal{P}_{i-1})}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}(|\mathcal{P}_i| - k + 1) + \frac{f(\mathcal{P}_i) + f(S)}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}(k - |\mathcal{P}_{i-1}|)\right) \\
&\le \left(\frac{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right)\left(f(\mathcal{P}^*) + \frac{f(\mathcal{P}_{i-1})}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|} + \frac{f(S)}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}(k - |\mathcal{P}_{i-1}|)\right) \\
&\le \left(\frac{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right)\left(f(\mathcal{P}^*) + \frac{f(\mathcal{P}_{i-1})}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}(1 + k - |\mathcal{P}_{i-1}|)\right). \qquad (7)
\end{aligned}$$

其中，上述第二个不等式由引理 3.2(i) 得出，而上述第三个不等式是因为 $f(S) \le f(\mathcal{P}_{i-1})$。不等式 (7) 表明

$$\begin{aligned}
f(\mathcal{P}) &\le \left(\frac{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right)\left(f(\mathcal{P}^*) + f(\mathcal{P}_{i-1}) + \frac{f(\mathcal{P}_{i-1})}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}(1 + k - |\mathcal{P}_i|)\right) \\
&\le \left(1 - \frac{1}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right)(f(\mathcal{P}^*) + f(\mathcal{P}_{i-1})) \quad (\text{since } k < |\mathcal{P}_i|) \\
&\le 2f(\mathcal{P}^*),
\end{aligned}$$

其中，最后一个不等式是因为根据引理 3.2(ii)，$f(\mathcal{P}_{i-1}) \le f(\mathcal{P}^*)$。上述分析表明，近似比至多为 2。现在我们进一步改进分析。根据上述不等式，我们可以假设 $f(\mathcal{P}^*) \neq 0$，因为如果 $f(\mathcal{P}^*) = 0$，那么返回的 k-划分 $\mathcal{P}$ 也满足 $f(\mathcal{P}) = 0$，因此是最优的。令 $c := f(\mathcal{P}_{i-1})/f(\mathcal{P}^*)$。根据引理 3.2(ii)，有 $f(\mathcal{P}_{i-1}) \le f(\mathcal{P}^*)$，因此，$c \in [0, 1]$。为方便起见，我们定义 $A := k - |\mathcal{P}_{i-1}|$ 和 $B := |\mathcal{P}_i| - k$，并注意到 $A, B \ge 1$。使用这一记法，我们可以将不等式 (7) 改写为

$$\begin{aligned}
f(\mathcal{P}) &\le \left(\frac{A+B}{A+B+1}\right)\left(f(\mathcal{P}^*) + \frac{1+A}{A+B}f(\mathcal{P}_{i-1})\right) \\
&= \left(\frac{A+B}{A+B+1}\right)\left(1 + \frac{1+A}{A+B} \cdot c\right) f(\mathcal{P}^*). \qquad (8)
\end{aligned}$$

由引理 3.2(i)，我们有

$$f(\mathcal{P}^*) \ge \left(\frac{B}{A+B}\right) f(\mathcal{P}_{i-1}) + \left(\frac{A}{A+B}\right) f(\mathcal{P}_i) = \left(\frac{B}{A+B}\right) c f(\mathcal{P}^*) + \left(\frac{A}{A+B}\right) f(\mathcal{P}_i).$$

重新整理可得

$$f(\mathcal{P}_i) \le \left(1 - \frac{B}{A+B} \cdot c\right)\left(\frac{A+B}{A}\right) f(\mathcal{P}^*) = \left(\frac{A+B}{A} - \frac{B}{A} \cdot c\right) f(\mathcal{P}^*).$$

由于 $\mathcal{P}$ 是通过粗化 $\mathcal{P}_i$ 得到的，根据 f 的次模性和非负性，我们有 $f(\mathcal{P}) \le f(\mathcal{P}_i)$。这意味着

$$f(\mathcal{P}) \le \left(\frac{A+B}{A} - \frac{B}{A} \cdot c\right) f(\mathcal{P}^*). \qquad (9)$$

结合不等式 (8) 和 (9)，我们有

$$\frac{f(\mathcal{P})}{f(\mathcal{P}^*)} \le \max_{c \in [0,1]} \min\left\{\left(\frac{A+B}{A+B+1}\right)\left(1 + \frac{1+A}{A+B} \cdot c\right),\ \frac{A+B}{A} - \frac{B}{A} \cdot c\right\}. \qquad (10)$$

因此，为了给近似比提供上界，只需对不等式 (10) 的右端提供上界。由于 $\binom{A+B}{A+B+1}\left(1+\frac{1+A}{A+B}\cdot c\right)$ 和 $\frac{A+B}{A}-\frac{B}{A}\cdot c$ 都是关于 c 的线性函数，其中前者随 c 的增加而增加，后者则随 c 的增加而减小，因此该值

$$\max_{c\in\mathbb{R}}\min\left\{\binom{A+B}{A+B+1}\left(1+\frac{1+A}{A+B}\cdot c\right),\ \frac{A+B}{A}-\frac{B}{A}\cdot c\right\}$$

当这两项相等时取到。令 $\binom{A+B}{A+B+1}\left(1+\frac{1+A}{A+B}\cdot c^*\right)=\frac{A+B}{A}-\frac{B}{A}\cdot c^*$ 并求解c^*，得到

$$c^*=\frac{\frac{A+B}{A}-\frac{A+B}{A+B+1}}{\frac{1+A}{A+B+1}+\frac{B}{A}}=\frac{\frac{B}{A}+\frac{1}{A+B+1}}{\frac{1+A}{A+B+1}+\frac{B}{A}}.$$

将 $c=c^*$ 代入 $\frac{A+B}{A}-\frac{B}{A}\cdot c$，得到

$$\begin{aligned}
\frac{f(\mathcal{P})}{f(\mathcal{P}^*)} &\le \max_{c\in[0,1]}\min\left\{\frac{A+B}{A+B+1}\left(1+\frac{1+A}{A+B}\cdot c\right),\ \frac{A+B}{A}-\frac{B}{A}\cdot c\right\}\\
&\le \frac{A+B}{A}-\frac{B}{A}\cdot c^*\\
&= 1+\frac{B}{A}(1-c^*)\\
&= 1+\frac{B}{A}\left(1-\frac{\frac{B}{A}+\frac{1}{A+B+1}}{\frac{1+A}{A+B+1}+\frac{B}{A}}\right)\\
&= 1+\frac{\frac{B}{A+B+1}}{\frac{1+A}{A+B+1}+\frac{B}{A}}\\
&= 1+\frac{AB}{A+A^2+AB+B^2+B}\\
&\le 1+\frac{AB}{3AB+A+B}\quad(\text{since } A^2+B^2\ge 2AB)\\
&= \frac{4}{3}-\frac{1}{3}\cdot\frac{A+B}{3AB+A+B}\\
&= \frac{4}{3}-\frac{1}{3}\cdot\frac{A+B}{3AB/2+3AB/2+A+B}\\
&\le \frac{4}{3}-\frac{1}{3}\cdot\frac{A+B}{3AB/2+3(A^2+B^2)/4+A+B}\quad(\text{since } AB\le\frac{A^2+B^2}{2})\\
&= \frac{4}{3}-\frac{1}{3}\cdot\frac{A+B}{3(A+B)^2/4+A+B}\\
&= \frac{4}{3}-\frac{1}{3}\cdot\frac{1}{3(A+B)/4+1}\\
&\le \frac{4}{3}-\frac{4}{9n+3}.
\end{aligned}$$

上面的最后一个不等式是因为 $A+B=|\mathcal{P}_i|-|\mathcal{P}_{i-1}|\le n-1$。 □

注 3.1. 算法 *1* 对非负单调次模函数的近似比至少为 $4/3-4/(9n+3)$。我们通过以下示例证明这一点：令 $V=\{a,b,c\}$、$k=2$ 和f：$2^V\to\mathbb{R}_{\ge 0}$ 定义如下

$$\begin{aligned}
&f(\emptyset)=0,\ f(\{a\})=1,\ f(\{b\})=f(\{c\})=1+\epsilon,\\
&f(\{a,b\})=f(\{a,c\})=\frac{3}{2}+\epsilon,\ f(\{b,c\})=f(V)=2+2\epsilon.
\end{aligned}$$

f 的次模性和单调性可以通过考察所有可能的子集来验证。此外，该实例的主划分序列为 $\{V\},\{\{a\},\{b\},\{c\}\}$。因此，算法 *1* 返回了2-划分

$\{\{a\},\{b,c\}\}$，其目标值为 $3+2\epsilon$。一个最优的2-划分是 $\{\{c\},\{a,b\}\}$，其目标值为 $5/2+\epsilon$。因此，近似比为 $\frac{3+2\epsilon}{5/2+\epsilon} \to \frac{6}{5}$，因为 $\epsilon \to 0$。注意，对于 $n=3$，定理 *3.1* 保证的近似比为 $\frac{4}{3} - \frac{4}{9n+3} = \frac{6}{5}$。

我们通过证明存在某些单调次模函数，使得算法 1 的近似比渐近地至少为 $4/3$（即当 $n \to \infty$ 时），来结束本节。

引理 3.4。对于每个奇正整数 $n \geq 5$，都存在一个函数 $k = k(n)$（即 k 是 n 的函数）以及一个定义在含有 n 个元素的底集上的非负单调次模 k- 划分实例，使得算法 *1* 在该实例上的近似比可以任意接近 $4/3 - 4/(3n+3)$。

证明。令 $n \geq 5$ 为任意奇数，$k = \frac{n+1}{2}$，并令 $V = \{v_1, \ldots, v_n\}$ 为底集。此外，令 $U := \{v_1, \ldots, v_{\frac{n-1}{2}}\}$ 和 $D := \{v_{\frac{n+1}{2}}, \ldots, v_n\}$，使得 $V = U \uplus D$。令 $g\colon 2^U \to \mathbb{R}_{\geq 0}$ 为定义在底集 U 上的函数，其定义如下：

$$g(S) = \begin{cases} \frac{1}{2} + \frac{1}{2} \cdot |S| & \text{if } \emptyset \neq S \subseteq U, \\ 0 & \text{if } S = \emptyset. \end{cases}$$

且 $f\colon 2^V \to \mathbb{R}_{\geq 0}$ 定义为

$$f(S) := \min\left\{ g(S \cap U) + (1+\epsilon)|S \cap D|,\ \frac{n+1}{2} \right\} \ \forall\ S \subseteq V,$$

其中 $\epsilon > 0$ 是无穷小量。函数 f 满足 $f(\emptyset) = 0$, $f(V) = \frac{n+1}{2}$, $f(U) = \frac{n+1}{4}$, $f(D) = \frac{n+1}{2}$ $f\{v\} = 1$ $v \in U$ $f\{v\} = 1+\epsilon$ $u \in D$ $\mathcal{Q}$, ( ) 对每个 都成立，且 ( ) 对每个 都成立。我们用 来表示将 V 划分为 n 个单元素集合所得的划分，由此可得 $f(\mathcal{Q}) = n + (n+1)_2{}^{\epsilon}$。为方便起见，对于所有 $S \subseteq V$，我们将写作 $h(S) := g(S \cap U) + |S \cap D|$，使得在整个证明过程中，对所有 $S \subseteq V$，$f(S) = \min\{h(S), \frac{n+1}{2}\}$。

命题 3.1。函数 f 是次模且单调的。

证明。令 $A, B \subseteq V$ 为任意元素。我们注意到，函数 h 是单调的，并且

$$h(A) + h(B) \geq h(A \cap B) + h(A \cup B).$$

不失一般性，我们可以假设 $h(A) \geq h(B)$，因此 $h(A \cup B) \geq h(A) \geq h(B) \geq h(A \cap B)$。我们考虑以下五种情况：

1. $h(A \cup B) \leq \frac{n+1}{2}$.

$$\implies f(A) + f(B) = h(A) + h(B) \geq h(A \cap B) + h(A \cup B) = f(A \cap B) + f(A \cup B).$$

2. $h(A) \leq \frac{n+1}{2} < h(A \cup B)$】【。

$$\implies f(A) + f(B) = h(A) + h(B) \geq h(A \cap B) + h(A \cup B) > h(A \cap B) + \frac{n+1}{2} = f(A \cap B) + f(A \cup B).$$

3. $h(B) \leq \frac{n+1}{2} < h(A)$

$$\implies f(A) + f(B) = \frac{n+1}{2} + h(B) \geq \frac{n+1}{2} + h(A \cap B) = f(A \cup B) + f(A \cap B).$$

4. $h(A \cap B) \leq \frac{n+1}{2} < h(B)$.

$$\implies f(A) + f(B) = \frac{n+1}{2} + \frac{n+1}{2} \geq \frac{n+1}{2} + h(A \cap B) = f(A \cup B) + f(A \cap B).$$

5. $\frac{n+1}{2} < h(A \cap B)$。

$$\implies f(A) + f(B) = \frac{n+1}{2} + \frac{n+1}{2} = f(A \cup B) + f(A \cap B).$$

由于这五种情况恰有一种成立，我们证明了 f 的次模性。 f 的单调性由 h 的单调性推出。

□

断言 3.2. 对于每个划分 $\mathcal{P} \neq \mathcal{Q}, \{V\}$，函数 f 满足

$$\frac{f(\mathcal{P}) - f(V)}{|\mathcal{P}| - 1} > \frac{f(\mathcal{Q}) - f(V)}{n-1}.$$

证明。 首先，我们注意到 $f(V) = \frac{n+1}{2}$ 且 $\frac{f(\mathcal{Q}) - f(V)}{n-1} = \frac{n + \frac{(n+1)\epsilon}{2} - \frac{n+1}{2}}{n-1} = \frac{1}{2} + \frac{(n+1)\epsilon}{2n-2}$，因此所需的不等式等价于

$$f(\mathcal{P}) \geq \frac{n}{2} + \frac{1}{2}|\mathcal{P}| + \epsilon(|\mathcal{P}| - 1)\frac{n+1}{2n-2}.$$

我们注意到，至多存在一个部分 $P \in \mathcal{P}$ 使得 $h(P) \geq \frac{n+1}{2}$。为证明这一点，假设两个不同的部分 P, $P' \in \mathcal{P}$ 满足 $h(P) \geq \frac{n+1}{2}$ 且 $h(P') \geq \frac{n+1}{2}$。这意味着

$$\begin{aligned} n+1 &\leq h(P) + h(P') \leq 2 \cdot \frac{1}{2} + \frac{1}{2}|(P \cup P') \cap U| + (1+\epsilon)|(P \cup P') \cap D| \\ &\leq 1 + \frac{1}{2} \cdot \frac{n-1}{2} + (1+\epsilon)\frac{n+1}{2} = \frac{3n+1}{4} + 1 + \frac{(n+1)\epsilon}{2} < n+1, \end{aligned}$$

因为 ϵ 是无穷小量且 $n \geq 5$，从而导出矛盾。因此，我们可以考虑两种情形： $\mathcal{P}$不包含任何满足 $h(P) \geq \frac{n+1}{2}$的部分 P，以及 $\mathcal{P}$恰好包含一个满足 $h(P) \geq \frac{n+1}{2}$的部分 P。

假设 $\mathcal{P}$ 不含任何满足 $h(P) \geq \frac{n+1}{2}$ 的部分 P。我们用 t 表示与 U 均有非空交集的 $\mathcal{P}$ 的部分数。于是， $|\mathcal{P}| - t$ 表示分别包含于 D 的 $\mathcal{P}$ 的部分数，并且有 $|\mathcal{P}| - t \leq |D| = \frac{n+1}{2}$。由此可得

$$\begin{aligned} f(\mathcal{P}) - \frac{n}{2} - \frac{|\mathcal{P}|}{2} - \epsilon(|\mathcal{P}| - 1)\frac{n+1}{2n-2} &= \frac{1}{2} \cdot t + \frac{1}{2} \cdot \frac{n-1}{2} + (1+\epsilon)\frac{n+1}{2} \\ &\quad - \frac{n}{2} - \frac{|\mathcal{P}|}{2} - \epsilon(|\mathcal{P}| - 1)\frac{n+1}{2n-2} \\ &= \frac{n+1}{4} - \frac{|\mathcal{P}| - t}{2} + \epsilon\left(\frac{n+1}{2} - \frac{(|\mathcal{P}| - 1)(n+1)}{2n-2}\right) \\ &\geq \frac{n+1}{4} - \frac{n+1}{4} + \epsilon\left(\frac{n+1}{2} - \frac{(|\mathcal{P}| - 1)(n+1)}{2n-2}\right) \\ &\geq \epsilon\left(\frac{n+1}{2} - \frac{(n-2)(n+1)}{2n-2}\right) \quad (\text{since } |\mathcal{P}| \leq n-1) \\ &= \epsilon \cdot \frac{n+1}{2n-2} \\ &> 0. \end{aligned}$$

现在，假设 $\mathcal{P}$ 恰好包含一个部分 $P \in \mathcal{P}$，使得 $h(P) \geq \frac{n+1}{2}$。 $\mathcal{P}$中的其他每个部分 $P' \neq P$ 均满足 $f(P') \geq 1$。由此可得

$$f(\mathcal{P}) \geq \frac{n+1}{2} + 1 \cdot (|\mathcal{P}| - 1) = \frac{n}{2} + |\mathcal{P}| - \frac{1}{2} > \frac{n}{2} + \frac{1}{2} \cdot |\mathcal{P}| + \epsilon(|\mathcal{P}| - 1)\frac{n+1}{2n-2}$$

因为 $|\mathcal{P}| \geq 2$ 和 ϵ 之间的差距可以任意小。

□

根据命题 2.1，算法 1 返回划分 $\{\{v_1\}, \ldots, \{v_{\frac{n-1}{2}}\}, D\}$（因为 $\{v_1\}$、……、$\{v_{\frac{n-1}{2}}\}$ 是使 f 值最小的 $k-1$ 单元素集合），其目标值为 $\frac{n-1}{2} + \frac{n+1}{2} = n$。划分$\{\{v_1, \ldots, v_{\frac{n+1}{2}}\}, \{v_{\frac{n+3}{2}}\}, \ldots, \{v_n\}\}$的目标值为 $\frac{n+1}{4} + (1+\epsilon) + (1+\epsilon)\frac{n-1}{2} = \frac{3n+3}{4} + (n+1)\frac{\epsilon}{2}$。近似比至少为

$$\frac{n}{\frac{3n+3}{4} + \frac{(n+1)\epsilon}{2}} \to \frac{4}{3} - \frac{4}{3n+3} \ (\text{as } \epsilon \to 0).$$

引理 3.4 的证明至此完成。

□

3.2 对称次模函数

本节将界定用于对称次模 k 分区的算法 1 的近似比。以下是本节的主要定理。

定理 3.2. 算法 *1* 对非负对称次模 k划分的近似比为 $2(1-1/n)$，其中 n 为底集的大小。

算法 1 达到的渐近近似因子 2 对于非负对称次模 k-划分是最优的：对于每个常数 $\epsilon > 0$，都不存在一个算法能够使用多项式数量的函数求值查询 [22] 实现 $(2-\epsilon)$-近似。证明该定理后，我们将讨论算法 1 的近似比的紧确性。定理 3.2 的证明由下文给出的引理 3.1 和引理 3.5 推出。

引理 3.5. 设 $k \geq 2$ 为一个整数，$f\colon 2^V \to \mathbb{R}$ 为定义在大小为 n的底集 V上的非负对称次模函数，$\mathcal{P}^*$ 为 V的一个使 $f(\mathcal{P}^*)$最小化的 k-划分，且 $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_r$ 为满足定理 *2.1* 条件的次模函数 f 的一个主划分序列。假设对于所有 $j \in [r]$，均有 $|\mathcal{P}_j| \neq k$。那么，算法 *1* 返回的划分 $\mathcal{P}$ 满足

$$f(\mathcal{P}) \leq 2\left(1 - \frac{1}{n}\right) f(\mathcal{P}^*).$$

证明。令 $\mathcal{P}_{i-1}$，$\mathcal{P}_i$ 为满足 $|\mathcal{P}_{i-1}| < k < |\mathcal{P}_i|$ 的划分。令 S 和 $\mathcal{P}' = \{B_1, B_2, \ldots, B_{|\mathcal{P}'|}\}$ 如算法 1 中所述。

首先，我们注意到，对于每个 $T \subseteq S$，f 的对称性、次模性和非负性共同意味着

$$f(T) = f(V-T) = f((V-S) \cup (S-T)) \leq f(V-S) + f(S-T) = f(S) + f(S-T). \tag{11}$$

其次，根据我们对 B_1，B_2，…，$B_{k-|\mathcal{P}_{i-1}|}$ 的选择，我们知道

$$\sum_{j=1}^{k-|\mathcal{P}_{i-1}|} f(B_j) \leq \frac{k-|\mathcal{P}_{i-1}|}{|\mathcal{P}'|} f(\mathcal{P}'). \tag{12}$$

由于 $\cup_{j=k-|\mathcal{P}_{i-1}|+1}^{|\mathcal{P}'|} B_j \subseteq S$，我们有

$$\begin{aligned}
f\left(\bigcup_{j=k-|\mathcal{P}_{i-1}|+1}^{|\mathcal{P}'|} B_j\right) &\leq f(S) + f\left(\bigcup_{j=1}^{k-|\mathcal{P}_{i-1}|} B_j\right) \quad \text{(by inequality (11))} \\
&\leq f(S) + \sum_{j=1}^{k-|\mathcal{P}_{i-1}|} f(B_j) \quad \text{(by submodularity and non-negativity)} \\
&\leq f(S) + \frac{k-|\mathcal{P}_{i-1}|}{|\mathcal{P}'|} f(\mathcal{P}'). \quad \text{(by inequality (12))}
\end{aligned}$$

因此，我们有

$$\begin{aligned}
f(\mathcal{P}) &= f(\mathcal{P}_{i-1}) - f(S) + \sum_{j=1}^{k-|\mathcal{P}_{i-1}|} f(B_j) + f\left(\bigcup_{j=k-|\mathcal{P}_{i-1}|+1}^{|\mathcal{P}'|} B_j\right) \\
&\leq f(\mathcal{P}_{i-1}) - f(S) + \frac{k-|\mathcal{P}_{i-1}|}{|\mathcal{P}'|} f(\mathcal{P}') + \left(f(S) + \frac{k-|\mathcal{P}_{i-1}|}{|\mathcal{P}'|} f(\mathcal{P}')\right) \\
&= f(\mathcal{P}_{i-1}) + 2 \cdot \frac{k-|\mathcal{P}_{i-1}|}{|\mathcal{P}'|} f(\mathcal{P}') \\
&= f(\mathcal{P}_{i-1}) + 2 \cdot \frac{k-|\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1} (f(\mathcal{P}_i) + f(S) - f(\mathcal{P}_{i-1})),
\end{aligned}$$

其中，最后一个等式由以下事实得出： $f(\mathcal{P}_{i-1}) - f(S) + f(\mathcal{P}') = f(\mathcal{P}_i)$，且 $|\mathcal{P}'| = |\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1$。重新整理后，得到

$$f(\mathcal{P}) \leq \frac{f(\mathcal{P}_{i-1})}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}(|\mathcal{P}_i| + |\mathcal{P}_{i-1}| - 2k + 1) + \frac{f(\mathcal{P}_i)}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}(2k - 2|\mathcal{P}_{i-1}|) + \frac{k - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1} \cdot 2f(S).$$

现在，我们注意到，根据对称性、次模性和非负性，有 $2f(S) = f(S) + f(V - S) \leq f(\mathcal{P}_{i-1})$。因此，我们有

$$\begin{aligned}
f(\mathcal{P}) &\leq \frac{f(\mathcal{P}_{i-1})}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}(|\mathcal{P}_i| + |\mathcal{P}_{i-1}| - 2k + 1) + \frac{f(\mathcal{P}_i)}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}(2k - 2|\mathcal{P}_{i-1}|) \\
&\quad + \frac{k - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1} \cdot f(\mathcal{P}_{i-1}) \\
&= \left(\frac{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right)\left(\frac{f(\mathcal{P}_{i-1})}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}(|\mathcal{P}_i| - k + 1) + \frac{f(\mathcal{P}_i)}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}(2k - 2|\mathcal{P}_{i-1}|)\right) \\
&\leq \left(\frac{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right)\left(\frac{f(\mathcal{P}_{i-1})}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}(2|\mathcal{P}_i| - 2k) + \frac{f(\mathcal{P}_i)}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}(2k - 2|\mathcal{P}_{i-1}|)\right) \\
&\leq \left(\frac{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right) \cdot 2f(\mathcal{P}^*) \quad \text{(by Lemma 3.2(i))} \\
&= \left(1 - \frac{1}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right) \cdot 2f(\mathcal{P}^*) \\
&\leq 2\left(1 - \frac{1}{n}\right) f(\mathcal{P}^*),
\end{aligned}$$

其中，第二个不等式是因为 $1 \leq |\mathcal{P}_i| - k$，最后一个不等式是因为 $|\mathcal{P}_i| - |\mathcal{P}_{i-1}| \leq n - 1$。 □

注释 3.2. 与基于贪心分割和 *Gomory-Hu* 树的方法不同，算法 *1* 中给出的主划分序列方法对于非负对称次模函数并不能达到 $2 - 2/k$ 的近似比：对于图 k-割这一特殊情形，*[21]*的命题 *5.2* 给出了一族图割函数，在这些函数上，算法 *1* 返回的解的近似比可以任意接近2。

3.3 正交子模函数

在本节中，我们界定算法 1 用于正交子模 k 分区的近似比。以下是本节的主要定理。

定理 3.3。 算法 *1* 对非负正交子模 k-分区的近似比为$2(1 - \frac{1}{n+1})$，其中 n 是底集的大小。

——

算法 1 达到的渐近近似因子 2 对于非负正交子模 k-划分是可能达到的最佳结果：对于每个常数 $\epsilon > 0$，不存在能够使用多项式数量的函数求值查询 [22] 实现 $(2 - \epsilon)$-近似的算法。引理 3.1 和 3.6 完成了定理 3.3 的证明。定理 3.3 的证明如下述引理 3.1 和引理 3.6 所示。

引理 3.6. 设 $k \geq 2$ 为整数， $f\colon 2^V \to \mathbb{R}_{\geq 0}$ 为定义在大小为 n 的底集 V 上的非负正交子模次模函数，$\mathcal{P}^*$ 为使 $f(\mathcal{P}^*)$ 最小的 k-划分，且 $\mathcal{P}_1$、$\mathcal{P}_2$、.....、$\mathcal{P}_r$ 为次模函数 f 的主划分序列，并满足定理 *2.1* 的条件。假设对所有 $j \in [r]$， $|\mathcal{P}_j| \neq k$ 均成立。那么，算法 *1* 返回的划分 $\mathcal{P}$ 满足

$$f(\mathcal{P}) \leq 2\left(1 - \frac{1}{n+1}\right) f(\mathcal{P}^*).$$

证明。设 $\mathcal{P}_{i-1}, \mathcal{P}_i$ 为满足 $|\mathcal{P}_{i-1}| < k < |\mathcal{P}_i|$ 的划分。令 S 和 $\mathcal{P}' = \{B_1, B_2, \ldots, B_{|\mathcal{P}'|}\}$ 如算法 **1** 中所述。

首先，我们注意到，对于每个 $T \subseteq S$，非负性和正交子模性意味着

$$f(T) \leq f(T) + f(\emptyset) \leq f(S) + f(S - T). \tag{13}$$

其次，根据我们对 B_1，B_2、.....、$B_{k-|\mathcal{P}_{i-1}|}$ 的选择，我们知道

$$\sum_{j=1}^{k-|\mathcal{P}_{i-1}|} f(B_j) \leq \frac{k - |\mathcal{P}_{i-1}|}{|\mathcal{P}'|} f(\mathcal{P}'). \tag{14}$$

由于 $\cup_{j=k-|\mathcal{P}_{i-1}|+1}^{|\mathcal{P}'|} B_j \subseteq S$，我们有

$$\begin{aligned}
f\left(\bigcup_{j=k-|\mathcal{P}_{i-1}|+1}^{|\mathcal{P}'|} B_j\right) &\leq f(S) + f\left(\bigcup_{j=1}^{k-|\mathcal{P}_{i-1}|} B_j\right) \quad \text{(by inequality (13))} \\
&\leq f(S) + \sum_{j=1}^{k-|\mathcal{P}_{i-1}|} f(B_j) \quad \text{(by submodularity and non-negativity)} \\
&\leq f(S) + \frac{k - |\mathcal{P}_{i-1}|}{|\mathcal{P}'|} f(\mathcal{P}'). \quad \text{(by inequality (14))}
\end{aligned}$$

因此，我们得到

$$\begin{aligned}
f(\mathcal{P}) &= f(\mathcal{P}_{i-1}) - f(S) + \sum_{j=1}^{k-|\mathcal{P}_{i-1}|} f(B_j) + f\left(\bigcup_{j=k-|\mathcal{P}_{i-1}|+1}^{|\mathcal{P}'|} B_j\right) \\
&\leq f(\mathcal{P}_{i-1}) - f(S) + 2 \sum_{j=1}^{k-|\mathcal{P}_{i-1}|} f(B_j) + f(S) \\
&= f(\mathcal{P}_{i-1}) + 2 \cdot \sum_{j=1}^{k-|\mathcal{P}_{i-1}|} f(B_j) \\
&\leq f(\mathcal{P}_{i-1}) + 2 \cdot \frac{k - |\mathcal{P}_{i-1}|}{|\mathcal{P}'|} f(\mathcal{P}') \\
&= f(\mathcal{P}_{i-1}) + \frac{2k - 2|\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1} (f(\mathcal{P}_i) + f(S) - f(\mathcal{P}_{i-1})),
\end{aligned}$$

其中，最后一个等号是根据 $f(\mathcal{P}_{i-1}) - f(S) + f(\mathcal{P}') = f(\mathcal{P}_i)$ 和 $|\mathcal{P}'| = |\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1$ 得出的。利用 $f(S) \leq f(\mathcal{P}_{i-1})$ 这一事实，我们得到

$$\begin{aligned}
f(\mathcal{P}) &\leq \left(\frac{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right) \left(\frac{f(\mathcal{P}_{i-1})}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|} (|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1)\right) \\
&\qquad + \left(\frac{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right) \left(\frac{f(\mathcal{P}_i)}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|} (2k - 2|\mathcal{P}_{i-1}|)\right) \\
&= \left(\frac{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right) \left(\frac{f(\mathcal{P}_{i-1})}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|} ((2|\mathcal{P}_i| - 2k) + (2k - |\mathcal{P}_{i-1}| - |\mathcal{P}_i| + 1))\right) \\
&\qquad + \left(\frac{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right) \left(\frac{2f(\mathcal{P}_i)}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|} (k - |\mathcal{P}_{i-1}|)\right) \\
&\leq \left(\frac{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right) \left(2f(\mathcal{P}^*) + \frac{f(\mathcal{P}_{i-1})}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|} (2k - |\mathcal{P}_{i-1}| - |\mathcal{P}_i| + 1)\right),
\end{aligned} \tag{15}$$

其中，最后一个不等式由引理 3.2(i) 得出。不等式 (15) 表明

$$\begin{aligned} f(\mathcal{P}) &\le \left(\frac{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right)\left(2f(\mathcal{P}^*) + f(\mathcal{P}_{i-1}) + \frac{f(\mathcal{P}_{i-1})}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}|}(2k - 2|\mathcal{P}_i| + 1)\right) \\ &\le \left(1 - \frac{1}{|\mathcal{P}_i| - |\mathcal{P}_{i-1}| + 1}\right)(2f(\mathcal{P}^*) + f(\mathcal{P}_{i-1})) \quad (\text{since } k < |\mathcal{P}_i|) \\ &\le 3f(\mathcal{P}^*), \end{aligned}$$

其中，最后一个不等式是因为根据引理 3.2(ii)，$f(\mathcal{P}_{i-1}) \le f(\mathcal{P}^*)$。上述分析已经表明，近似比至多为 3。现在我们进一步收紧该近似比。作为上述不等式的推论，我们可以假设 $f(\mathcal{P}^*) \neq 0$，因为如果 $f(\mathcal{P}^*) = 0$，那么返回的 k-划分 $\mathcal{P}$ 也满足$f(\mathcal{P}) = 0$，因此是最优的。令 $c := f(\mathcal{P}_{i-1})/f(\mathcal{P}^*)$。根据引理 3.2(ii)，我们有 $f(\mathcal{P}_{i-1}) \le f(\mathcal{P}^*)$，从而有 $c \in [0,1]$。为方便起见，我们定义 $A := k - |\mathcal{P}_{i-1}|$ 和 $B := |\mathcal{P}_i| - k$，并注意到 $A,\ B \ge 1$。使用这些记号，我们可以将不等式 (15) 重写为

$$\begin{aligned} f(\mathcal{P}) &\le \left(\frac{A+B}{A+B+1}\right)\left(2f(\mathcal{P}^*) + \frac{A-B+1}{A+B}f(\mathcal{P}_{i-1})\right) \\ &= \left(\frac{A+B}{A+B+1}\right)\left(2 + \frac{A-B+1}{A+B}\cdot c\right)f(\mathcal{P}^*). \end{aligned} \tag{16}$$

根据引理 3.2(i)，我们有

$$f(\mathcal{P}^*) \ge \left(\frac{B}{A+B}\right)f(\mathcal{P}_{i-1}) + \left(\frac{A}{A+B}\right)f(\mathcal{P}_i) = \left(\frac{B}{A+B}\right)cf(\mathcal{P}^*) + \left(\frac{A}{A+B}\right)f(\mathcal{P}_i).$$

重新整理后，得到

$$f(\mathcal{P}_i) \le \left(1 - \frac{B}{A+B}\cdot c\right)\left(\frac{A+B}{A}\right)f(\mathcal{P}^*) = \left(\frac{A+B}{A} - \frac{B}{A}\cdot c\right)f(\mathcal{P}^*).$$

由于 $\mathcal{P}$ 是通过对 $\mathcal{P}_i$ 进行粗化得到的，根据 f 的次模性和非负性，我们有 $f(\mathcal{P}) \le f(\mathcal{P}_i)$。因此，

$$f(\mathcal{P}) \le \left(\frac{A+B}{A} - \frac{B}{A}\cdot c\right)f(\mathcal{P}^*). \tag{17}$$

结合不等式 (16) 和 (17)，我们得到

$$\frac{f(\mathcal{P})}{f(\mathcal{P}^*)} \le \max_{c\in[0,1]} \min\left\{\left(\frac{A+B}{A+B+1}\right)\left(2 + \frac{A-B+1}{A+B}\cdot c\right),\ \frac{A+B}{A} - \frac{B}{A}\cdot c\right\}. \tag{18}$$

因此，为了上界化近似比，只需对不等式 (18) 的右侧进行上界化。两项 $\binom{A+B}{A+B+1}\left(2 + \frac{A-B+1}{A+B}\cdot c\right)$ 和 $\frac{A+B}{A} - \frac{B}{A}\cdot c$ 都是关于 c 的线性项，后者作为 c的函数是递减的。接下来，我们考虑两种情况：$A - B + 1 \le 0$ 和 $A - B + 1 > 0$。

假设 $A - B + 1 \le 0$，那么项 $\binom{A+B}{A+B+1}\left(2 + \frac{A-B+1}{A+B}\cdot c\right)$ 是关于 c 的线性非增函数。最大值

$$\max_{c\in[0,1]} \min\left\{\left(\frac{A+B}{A+B+1}\right)\left(2 + \frac{A-B+1}{A+B}\cdot c\right),\ \frac{A+B}{A} - \frac{B}{A}\cdot c\right\}$$

在 $c = 0$ 处达到。因此，我们有

$$\begin{aligned} \frac{f(\mathcal{P})}{f(\mathcal{P}^*)} &\le \min\left\{\frac{A+B}{A+B+1}\cdot 2,\ \frac{A+B}{A}\right\} \\ &\le \frac{A+B}{A+B+1}\cdot 2 \end{aligned}$$

$$= 2\left(1 - \frac{1}{A+B+1}\right)$$
$$\leq 2\left(1 - \frac{1}{n}\right), \tag{19}$$

其中，最后一个不等式是由 $A+B = |\mathcal{P}_i| - |\mathcal{P}_{i-1}| \leq n-1$ 这一事实得出的。

现在，我们考虑 $A-B+1 > 0$ 的情况。项 $\binom{A+B}{A+B+1}\left(2 + \frac{A-B+1}{A+B} \cdot c\right)$ 关于 c 是线性递增的，因此最大值

$$\max_{c\in\mathbb{R}} \min\left\{\binom{A+B}{A+B+1}\left(2 + \frac{A-B+1}{A+B} \cdot c\right), \frac{A+B}{A} - \frac{B}{A} \cdot c\right\}$$

当两项相等时达到。令 $\binom{A+B}{A+B+1}\left(2 + \frac{A-B+1}{A+B} \cdot c^*\right) = \frac{A+B}{A} - \frac{B}{A} \cdot c^*$ 并解出 c^*，得到

$$c^* = \frac{\frac{A+B}{A} - 2 \cdot \frac{A+B}{A+B+1}}{\frac{A-B+1}{A+B+1} + \frac{B}{A}}.$$

将 $c = c^*$ 代入 $\frac{A+B}{A} - \frac{B}{A} \cdot c$，我们得到

$$\begin{aligned}
\frac{f(\mathcal{P})}{f(\mathcal{P}^*)} &\leq \max_{c\in[0,1]} \min\left\{\left(\frac{A+B}{A+B+1}\right)\left(2 + \frac{A-B+1}{A+B} \cdot c\right), \frac{A+B}{A} - \frac{B}{A} \cdot c\right\} \\
&\leq \frac{A+B}{A} - \frac{B}{A} \cdot c^* \\
&= 1 + \frac{B}{A}(1 - c^*) \\
&= 1 + \frac{B}{A}\left(1 - \frac{\frac{A+B}{A} - 2 \cdot \frac{A+B}{A+B+1}}{\frac{A-B+1}{A+B+1} + \frac{B}{A}}\right) \\
&= 1 + \frac{\frac{2B}{A+B+1}}{\frac{A-B+1}{A+B+1} + \frac{B}{A}} \\
&= 1 + \frac{2AB}{A^2 + A + B^2 + B} \\
&\leq 1 + \frac{2AB}{2AB + A + B} \quad (\text{since } A^2 + B^2 \geq 2AB) \\
&= 2 - \frac{A+B}{2AB + A + B} \\
&= 2 - \frac{A+B}{AB + AB + A + B} \\
&\leq 2 - \frac{A+B}{AB + (A^2 + B^2)/2 + A + B} \quad (\text{since } AB \leq \frac{A^2+B^2}{2}) \\
&= 2 - \frac{A+B}{(A+B)^2/2 + A + B} \\
&= 2 - \frac{1}{(A+B)/2 + 1} \\
&\leq 2 - \frac{2}{n+1}.
\end{aligned}$$

上面的最后一个不等式是因为 $A+B = |\mathcal{P}_i| - |\mathcal{P}_{i-1}| \leq n-1$。 □

注释 3.3。非负偏模次模函数的算法 *1* 的近似比至少为 $2 - 2/(n+1)$。我们通过以下示例证明这一点，其中 $n = 3, k = 2$ 设定如下：令 $V = \{a, b, c\}$、$k = 2$ 和 f：$2^V \to \mathbb{R}_{\geq 0}$ 定义为

$$f(\emptyset) = 0,\ f(a) = f(b) = 1,\ f(c) = 1 + \epsilon,$$

$$f(\{a,b\}) = 1+\epsilon,\ f(\{b,c\}) = f(\{a,c\}) = 2,\ f(\{a,b,c\}) = 1+\epsilon.$$

通过考虑所有可能的子集，可以验证 f 的次模性和正模性。此外，该实例的主划分序列为 $\{V\}$, $\{\{a\},\{b\},\{c\}\}$。因此，该算法返回的 2-划分为$\{\{a\},\{b,c\}\}$，而最优的 2-划分为 $\{\{c\},\{a,b\}\}$。因此，当 $\epsilon \to 0$ 时，近似比趋近于 $3/2$。我们注意到，对于 $n=3$，定理 *3.3* 所保证的近似比为 $2-2/(n+1) = 3/2$。

4 任意次模函数的下界

在本节中，我们给出一个次模 k-划分实例，其中算法 1 达到 $\Omega(n/k)$ 的近似因子。我们强调，该实例中的次模函数并非对称/单调/正交子模。

令 $V = \{v_0, v_1, \ldots, v_{n-1}\}$ 为底集。我们定义一个有向图 $D = (V, E(D))$ 和一个超图$H = (V, E(H))$，它们具有相同的顶点集 V，如下所示（见图 1）：

$$E(D) = \{v_0 v_i : i \in [n-1]\} \text{ and}$$
$$E(H) = \{\{v_1, v_2, \ldots, v_{n-1}\}\}.$$

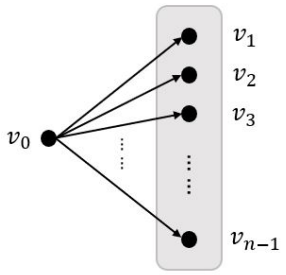


图 1：第 4 节中的示例。弧属于有向图 D，而超边 $\{v_1, \ldots, v_{n-1}\}$ 属于超图 H。

对于每个子集 $S \subseteq V$，我们用 $d_D^{in}(S)$ 表示 D 中尾端位于 $\bar{S}$ 且头端位于 S 的弧的数量。我们用 $d_H(S)$ 表示 H 中至少有一个顶点属于 S 且至少有一个顶点属于 $\bar{S}$ 的超边数量。接下来，我们定义一个集合函数 $f\colon 2^V \to \mathbb{R}_{\geq 0}$，如下所示:

$$f(S) := a \cdot d_D^{in}(S) + d_H(S) \quad \forall S \subseteq V,$$

其中 $a \gg 1$ 是一个很大的常数。注意，f 是次模的，因为它是两个次模函数的正线性组合（且它既不单调，也不对称，也不是正交子模）。

命题 4.1。 f 的主划分序列为 $\{V\}$、$\{\{v_i\}\colon \ i \in \{0,1,\ldots,n-1\}\}$。

证明。 为方便起见，我们用 $\mathcal{Q}$ 表示将 V 划分为单元素集合所得的划分。根据命题 2.1 以及 $f(V) = 0$ 这一事实，只需证明：对于 V 的任意划分 $\mathcal{P}$，只要 $\mathcal{P}$ 不是 $\mathcal{Q}$ 或 $\{V\}$，就有

$$\frac{f(\mathcal{P})}{|\mathcal{P}|-1} > \frac{f(\mathcal{Q})}{n-1}. \tag{20}$$

令 $P_0 \in \mathcal{P}$ 为包含 v_0 的部分。于是，当 $P_0 \neq \{v_0\}$ 时，$f(P_0) \geq 1$，否则 $f(P_0) = 0$。对于每个不包含 v_0 的部分 $P \in \mathcal{P}$，当 $|P| = 1$ 时，$f(P) \geq 1+a$，若$|P| \geq 2$则 $f(P) \geq 2+a$。由于 $\mathcal{P} \neq \mathcal{Q}$，因此 $P_0 \neq \{v_0\}$，或者至少有一个部分 $P \in \mathcal{P} \setminus \{P_0\}$ 的大小为$|P| \geq 2$。因此，$f(\mathcal{P}) = \sum_{P\in\mathcal{P}} f\ P(\geq) 1+a$ $(|\mathcal{P}|-1)+1$。此外，我们有 $f(\mathcal{Q}) = 1+a(n-1)$，从而

$$\frac{f(\mathcal{P})}{|\mathcal{P}|-1} \geq \frac{(1+a)(|\mathcal{P}|-1)+1}{|\mathcal{P}|-1} = 1+a+\frac{1}{|\mathcal{P}|-1} > 1+a = \frac{(1+a)(n-1)}{n-1} = \frac{f(\mathcal{Q})}{n-1}.$$

这证明了不等式（20）。 □

命题 4.2. 算法 *1* 在输入 (f, k) 上的近似比为 $\Omega(n/k)$。

证明。 我们注意到，对于所有 $i \in [n-1]$， $f(\{v_0\}) = 0$ 且 $f(\{v_i\}) = 1 + a$。根据命题 **4.1**，对于输入 (f, k)，算法 **1** 返回一个划分 $\mathcal{P}$，该划分由使 f 最小的 $k-1$ 个单元素部分，以及这些 $k-1$ 个单元素部分的并集的补集组成。因此，返回的划分 $\mathcal{P}$ 包含 $\{v_0\}$ 作为一个部分，从而

$$f(\mathcal{P}) = \sum_{P \in \mathcal{P}: P \neq \{v_0\}} f(P) \geq f(V - \{v_0\}) = a(n-1).$$

第二个不等式由函数 f 的次模性得出。考虑目标值为 $(1+a)(k-1)+1$ 的 k-划分 $\{\{v_1\}, \{v_2\}, \ldots, \{v_{k-1}\}, V - \{v_1, \ldots, v_{k-1}\}\}$。这意味着最优 k-划分 $\mathcal{P}^*$ 满足 $f(\mathcal{P}^*) \leq (1+a)(k-1)+1$。因此，算法 **1** 返回的解的近似比为

$$\frac{f(\mathcal{P})}{f(\mathcal{P}^*)} \geq \frac{a(n-1)}{(1+a)(k-1)+1} \to \frac{n-1}{k-1} \text{ as } a \to \infty.$$

□

5 结论

Narayanan [15] 证明了次模函数的主划分序列存在。由于划分序列与次模函数相关联的格结构，文献中将次模函数的主划分序列称为次模函数的划分主格[2,7,8,10,13,16,19]。本研究中，我们选择称之为主划分序列，因为该序列已足以满足我们的目的。Narayanan、Roy 和 Patkar [17] 利用主划分序列设计了一种用于次模 k-划分的算法。他们分析了该算法在某些由超图导出的次模函数子族上的近似比。在本研究中，我们考察了该算法在三个广泛的次模函数子族上的近似比，即单调、对称和正模次模函数。结果表明，基于主划分序列的算法在这三个子族上均达到了最佳可能的渐近近似因子。我们的一项创新贡献是将单调次模 k-划分的可近似性从 2 提高到 $4/3$，从而达到不可近似阈值。值得进一步确定单调次模 k-划分中特殊情形的可近似性，例如拟阵 k-划分和覆盖 k-划分；这些问题本身就很有趣，因为它们涵盖了若干有趣的划分问题。

致谢。Karthekeyan Chandrasekaran 谨此感谢 Chandra Chekuri 提出关于基于主划分序列的方法用于对称次模 k 划分时的近似比问题。

参考文献

[1] M. Baïou 和 F. Barahona，超树打包与超图中的 *k*-割问题，《国际学习与智能优化会议》，LION，2023 年，第 521–534 页。

[2] F. Ba rahona，关于 *k*-割问题，《运筹学快报》**26**（2000 年），第 3 期，99–105 页 .

[3] D. Chakrabarty 和 Z. Huang，覆盖函数的测试，《自动机、语言与编程》，ICALP，2012年，第170–181页。

[4] K. Chandrasekaran 和 C. Chekuri，超图 k-固定 k 的确定性多项式时间算法，《运筹学数学》（2022），FOCS 2020 预备版本：第 810–821 页。

[5] C. Chekuri 和 S. Li，关于近似 k路超图割问题的困难性，Theory Comput. 16（2020），1–8。[6] C. Chekuri、K. Quanrud 和 C. Xu，最小 k割的线性规划松弛与树打包，SIAM离散数学杂志 34（2020），第 2 期，1334–1353。[7] W. Cunningham，网络的最优攻击与加固，ACM杂志 32（1985），第 3 期，549—561。[8] M. P. Desai、Narayanan 和 S. B. Patkar，通过分解和次模函数的划分主格实现有限状态机，Discrete Appl. Math. 131（2003），299–310。[9] O. Goldschmidt 和 D. Hochbaum，固定 k 时 k割问题的多项式算法，Mathematics of Operations Research 19（1994），第 1 期，24–37。[10] V. Kolmogorov，计算图的划分主序列的更快算法，Algorithmica（2010），394–412。[11] P. Manurangsi，基于小集合扩展假设的最大二分团问题、最小 k割和最稠密至少-k-子图的不可近似性，Algorithms 11(1)（2018），10。[12] H. Nagamochi 和 Y. Kamidoi，图中的最小成本子划分，信息处理快报 102（2007），第 2 期，79–84。[13] K. Nagano、Y. Kawahara 和 S. Iwata，最小平均成本聚类，Advances in Neural Information Processing Systems，NIPS，第 23 卷，2010。[14] J. Naor 和 Y. Rabani，树打包与 k割近似，第 12 届 ACM-SIAM离散算法研讨会论文集，SODA，2001，26—27 页。[15] H. Narayanan，次模函数的划分主格，Linear Algebra and its Applications 144（1991），179–216。[16] H. Narayanan，次模函数与电网络（第 2 版），1997。[17] H. Narayanan、S. Roy 和 S. Patkar，使用划分主格方法的最小k重叠问题近似算法，Journal of Algorithms 21（1996），第 2 期，306–330。[18] K. Okumoto、T. Fukunaga 和 H. Nagamochi，用于近似划分超图和次模系统的分治算法，Algorithmica 62（2012），第 3 期，787–806。[19] S. B. Patkar 和 H. Narayanan，使用次模函数改进图划分，Discrete Appl. Math. 131（2003），535–553。[20] K. Quanrud，k割的快速确定性近似，《近似、随机化与组合优化：算法与技术》论文集，APPROX，2019，23:1–23:20 页。[21] R. Ravi 和 A. Sinha，使用网络强度作为拉格朗日松弛近似 k割，欧洲运筹学杂志 186（2008），第 1 期，77–90。[22] R.Santiago，次模划分问题的新近似与困难性结果，《组合算法国际研讨会》论文集，IWOCA，2021，516–530 页。[23] H. Saran 和 V. Vazirani，在最优值两倍以内求 k割，SIAM计算杂志 24（1995），第 1 期，101–108。[24] V. Vazirani，近似算法，Springer，柏林，德国，2003 - 2003（ger;eng）。[25] L. Zhao、H. Nagamochi 和 T. Ibaraki，用于近似多路划分问题的贪心分裂算法，数学规划 102（2005），第 1 期，167–183。