近似 通过稠密子集和联系背包问题与划分问题 S

Mingyang Deng * MIT
Ce Jin † MIT
Xiao Mao ‡ MIT

2023年1月24日

arXiv:2301.09333v1 [cs.DS] 23 Jan 2023

摘要

背包问题 和划分问题 是两个重要的加法问题，其在 $(1-\varepsilon)$ 近似设定下的细粒度复杂度尚未确定。在这项工作中，我们通过提出改进的算法，推动了这两个问题的研究进展。

- 背包问题 可在 $\tilde{O}(n+(1/\varepsilon)^{2.2})$ 时间内实现 $(1-\varepsilon)$-近似，改进了 Jin 在 ICALP'19 提出的 $\tilde{O}(n+(1/\varepsilon)^{2.25})$ 算法。基于（min $+$）卷积假设，已知其条件下界为 $(n+1/\varepsilon)^{2-o(1)}$。

- 划分问题 可在 $\tilde{O}(n+(1/\varepsilon)^{1.25})$ 时间内实现 $(1-\varepsilon)$-近似，改进了 Bringmann 和 Nakos 在 SODA'21 提出的 $\tilde{O}(n+(1/\varepsilon)^{1.5})$ 算法。基于强指数时间假设，已知其条件下界为 $(1/\varepsilon)^{1-o(1)}$。

我们的两种新算法都应用了 Galil 和 Margalit（SICOMP'91）、Bringmann 和 Wellnitz（SODA'21）关于稠密子集和的加法组合结果。在我们的工作之前，这类技术尚未在背包问题中得到探索。此外，我们还设计了几种新方法，以加速求解加法问题时自然出现的分治步骤。

*dengm@mit.edu
†cejin@mit.edu。部分由美国国家科学基金会 CCF-2129139 项目资助
‡matthew99a@gmail.com

1 引言

1.1 背景

背包问题、子集和和划分问题 是计算机科学和数学优化中的三个基础问题，目前仍在整数规划和细粒度复杂性等领域得到积极研究。在背包问题（有时也称为 0-1 背包问题）中，给定一个由 I 个物品组成的集合，其中每个物品 $i \in I$ 的重量为 $w_i > 0$ 、利润为 $p_i > 0$，以及背包容量 W，我们要选取一个满足重量约束$\sum_{j\in J} w_j \leq W$ 的子集 $J \subseteq I$，使总利润 $\sum_{j\in J} p_j$ 最大化。子集和问题是背包问题的一个特例，其中物品的重量总是等于其利润。划分问题是子集和问题的一个特例，其中容量等于所有物品总重量的一半。换言之，在划分问题中，我们要将输入物品划分为两部分，使两部分的总和尽可能接近。

众所周知，这三个问题都很难：它们出现在 Karp 最初列出的 21 个 NP 难问题中 [Kar72]。为应对 NP 难性，一个自然的方向是研究它们的 近似算法。给定参数 $\varepsilon > 0$，以及一个最优值为 最优值 的输入实例，要求一个 $(1-\varepsilon)$-近似算法输出一个数 SOL，使得 $(1-\varepsilon)\mathrm{OPT} \leq \mathrm{SOL} \leq \mathrm{OPT}$。幸运的是，众所周知，这三个问题都有 全多项式时间近似方案（*FPTAS*），也就是说，对任意 $(1-\varepsilon)$-近似算法，其运行时间为 $\mathrm{poly}(n, 1/\varepsilon)$，其中 $\varepsilon > 0$。

自 20 世纪 70 年代以来，研究者一直致力于为这些问题设计近似方案，并从 n 和 $1/\varepsilon$ [的角度改进其时间复杂度IK75, Law79, GL79, KP99, KMPS03, KP04, Rhe15, JK18, Cha18, MWW19, Jin19, BN21, BC22]。另一方面，近期细粒度复杂性研究指出，在一些广泛认可的困难性假设下，此类改进存在上限 [CMWW19, KPS17, ABHS19, BN21]。下面简要介绍这一方向上最新的结果。

- **背包问题：** Jin 提出的目前已知最佳算法 [Jin19] 的运行时间为 $\tilde{O}(n + \varepsilon^{-2.25})$，它基于 Chan 先前提出的运行时间为 [Cha18] in $\tilde{O}(n + \varepsilon^{-2.4})$ 的算法。 [CMWW19] 和[KPS17] 基于（min +）-卷积假设，证明了条件下界为 $(n+\frac{1}{\varepsilon})^{2-o_{(1)}}$。~

- **子集和：** Bringmann 和 Nakos 提出的目前已知最佳算法[BN21] 的运行时间为 $O(n + \varepsilon^{-2}/2^{\Omega(\sqrt{\log(1/\varepsilon)})}$时间（通过低阶因子改进了 [KMPS03] ）。Bringmann 和 Nakos[BN21] 基于 (min +)-卷积假设证明了匹配的下界。

- **划分问题：** Mucha、Węgrzycki 和 Włodarczyk 首次取得突破 [MWW19]，提出了一种运行时间为 $\tilde{O}(n+1/\varepsilon^{5/3})$ 的随机化算法。随后，Bringmann 和 Nakos [BN21]将其改进为确定性算法，运行时间为 $\tilde{O}(n+1/\varepsilon^{-3/2})$。Abboud、Bringmann、Hermelin 和 Shabtay [ABHS19]证明，在强指数时间假设成立的条件下，划分问题无法在任何 $\mathrm{poly}(n)/\varepsilon^{1-\delta}$ 时间内得到近似，。

我们可以看到，子集和的复杂度问题已经有了定论，但对于背包问题和划分问题，已知的最佳算法与其条件下界之间仍存在差距。

1.2 我们的结果

在这项工作中，我们在这一方向上取得了进展，为背包问题和划分问题提出了改进的近似方案。

定理 1.1。 存在一种随机化算法，可在运行时间 [1] 内对背包问题进行$(1-\varepsilon)$近似

$$\tilde{O}\left(n+\varepsilon^{-11/5}/2^{\Omega(\sqrt{\log(1/\varepsilon)})}\right).$$

定理 1.2。 存在一种确定性算法，可在运行时间内对划分问题进行$(1-\varepsilon)$近似

$$\tilde{O}\left(n+\varepsilon^{-5/4}\right).$$

1.3 技术概述

对于许多与子集和相关的问题，加法组合学中的一个有用结果是 Galil 和 Margalit 提出的结果 [GM91], ，后来 Bringmann 和 Wellnitz 对其进行了改进 [BW21]（基于 Sárközy 的结果 [Sár94]）。这些组合学结果揭示了正整数集合 $\mathcal{S}(X)$的子集和集合 X，其定义为 $\mathcal{S}(X):=\{\sum_{y\in Y} y: Y\subseteq X\}$，当$X$ 是 “稠密的”时。直观而言，该结果表明，如果物品总数很大，那么就能非常高效地计算出子集和中的很大一部分，因此只需近似计算价值不超过某个 λ 的和集的一小段范围。我们将这些组合学结果应用于划分问题，并且出乎意料地，也应用于“二维”背包问题，其中每件物品同时具有重量和价值。

1.3.1 背包问题

利用稠密子集和加速背包问题。我们改进后的背包问题近似方案依赖于 Chan [Cha18] 和 Jin[Jin19] 先前算法中的多个技术组件。不过，使这项改进成为可能的一个关键新思路，是一种此前未在背包算法中探索过的技术：Galil 和 Margalit [GM91] 关于稠密子集和的加法组合学结果。对我们特别有用的一项结果大致指出：当 X 由 n 个属于 $[\ell,2\ell]$ 的不同整数构成，且 $\ell \ll n^2$ 足够小时，就存在一个足够长的区间 $[\lambda,(\sum_{x\in X} x)-\lambda]$，其中稠密地分布着 $\mathcal{S}(X)$ 中的元素；也就是说，任意两个相邻元素之间的间隔都非常小。该结论的正式表述见引理 2.6。我们将看到，这类稠密度结论在 Jin [Jin19] 的框架中很有用。Jin 的背包问题近似算法会分别处理高效率和低效率物品，其中效率定义为利润重量比 p_i/w_i。直观而言，在解中纳入过多低效率物品并不划算。事实上，经过一些技术步骤后，Jin 成功地为任一最优解中低效率物品贡献的总利润设定了上界 B，因此只需计算低效率物品对应的答案，且计算到 B 即可（所需时间会比原来短得多）。Jin 证明这一上界 B 的方法，是采用某种贪心交换论证：给定一个包含过多低效率物品的解集合（这些物品总共占用容量 W_L），将它们移除，再尝试用高效率物品填满腾出的空间 W_L。这样有可能得到更优的解，从而与给定解的最优性矛盾。

[1] 在本文中，我们用 $\tilde{O}(f)$ 表示 $O(f\cdot \operatorname{poly}\log(f))$。

自然，这种交换并非总能带来收益，因为高效率物品可能无法填满整个空间 W_L。Jin 的论证通过额外确保所有物品的大小都落在区间 $[\ell, 2\ell]$ 内来处理这一问题，这样交换后浪费的空间就不会超过 2ℓ （否则总能再放入一个高效率物品）。在我们的新算法中，我们利用关于稠密子集和的组合学结果来改进这一论证：注意，最小化浪费空间的任务等价于一个针对高效率物品大小的子集和问题。通过适当设置参数，我们可以确保稠密子集和结果适用，从而将浪费空间大幅减少。

通过完善这一论证，我们可以改进 Jin [Jin19] ，方法是对低效率物品所贡献的总利润设置更严格的上界，从而缩短运行时间。

1.3.2 划分问题

稠密化分治针对子集和、划分问题、背包问题等问题，已有许多采用分治技术的全多项式时间近似方案算法，例如[Cha18，Jin19、BN21]。与以往的方法不同，我们的改进关键在于对排序后的物品列表进行分治。为了阐明我们的思路，我们注意到，在大多数基于分治的算法中，运行时间的瓶颈出现在底层：此时我们需要合并两个答案集合，而其复杂度往往与顶层的相同。然而，直观而言，如果列表已经排序，那么在底层，物品的价值落在一个较小的区间内，因此这些物品的和集会聚集在若干个较小的区间中，且区间之间存在较大的间隔。为了利用这一性质，我们不再使用常规的一维快速傅里叶变换来合并和集，而是将和集“稠密化”，并使用二维快速傅里叶变换进行合并，使得运行时间仅取决于这些小区间的总长度。

我们注意到，Koiliaris 和 Xu 此前已将二维快速傅里叶变换的思路用于子集和[KX19], ，但用途有所不同：我们使用二维快速傅里叶变换来“稠密化”和集，而在 [KX19] 中，它用于界定每个和对应解的大小。[2]

结合加法组合学结果 Galil 和 Margalit 关于稠密子集和的加法组合学结果 [GM91] 也在一定程度上应用于 $\tilde{O}(n + 1/\varepsilon^{5/3})$算法中，该算法由 [MWW19]提出。在我们的算法中，我们将这一结果与稠密化分治相结合。注意， λ 可能远小于物品总和 σ ，这意味着加性误差相对于更大的 $\varepsilon\sigma$而言也是一种 $\varepsilon'\lambda$ 加性误差，因此我们只需确保在计算过程中达到 $\varepsilon' = \varepsilon\sigma/\lambda(1-\varepsilon')$乘性近似，而不必达到 $(1-\varepsilon)$ 。

1.4 论文结构

我们将在第 2 节给出有用的定义和引理。在第 3 节中，我们将介绍背包问题的算法。在第 4 节中，我们将介绍划分问题的算法。一些标准归约以及先前研究中的已知引理将放在附录中。

[2]事实上，可以验证，将我们进行二维快速傅里叶变换的方法与[KX19] 中的其他技术相结合，可以得到一种确定性解法，在 $\tilde{O}(\sqrt{nt})$ 时间内解决子集和问题。值得探究的是，这两种运用二维快速傅里叶变换的方法能否结合起来，从而确定性地改善子集和问题的运行时间。

2 预备知识

我们写作 $\mathbb{N}=\{0,1,2,\dots\}$ 和 $\mathbb{N}^+=\{1,2,\dots\}$。对于 $n\in\mathbb{N}$，我们写作 $[n]=\{1,2,\dots,n\}$。

2.1 问题陈述

在背包问题中，输入是一列 n 物品 (p_1,w_1) ....., $(p_n,w_n)\in\mathbb{N}\times\mathbb{N}$ 以及背包容量 $W\in\mathbb{N}$，最优值为

$$\mathrm{OPT}:=\max_{J\subseteq[n]}\Big\{\sum_{j\in J}p_j\;\Big|\;\sum_{j\in J}w_j\le W\Big\}.$$

在较简单的划分问题中，输入是一列 n 整数 x_1, ..., $x_n\in\mathbb{N}$，最优值为

$$\mathrm{OPT}:=\max_{J\subseteq[n]}\Big\{\sum_{j\in J}x_j\;\Big|\;\sum_{j\in J}x_j\le\frac{1}{2}\sum_{i\in[n]}x_i\Big\}.$$

给定一个背包问题（或划分问题）实例和参数 $\varepsilon\in(0,1)$，要求一个 $(1-\varepsilon)$-近似算法 输出一个数 SOL，使得 $(1-\varepsilon)$最优值 $\le$ SOL $\le$ 最优值。

对于这两个问题，我们都可以假设 $n=O(\varepsilon^{-4})$ 因此 对数$n=O$(对数 ε^{-1})。对于更大的 n，Lawler 算法 [Law79] 针对背包问题的算法在 $O(n$ 对数$\frac{1}{\varepsilon}+(\frac{1}{\varepsilon})^4)$ 时间内已经接近最优。

有时我们会描述近似比为 $1-O(\varepsilon)$（或 $1-\varepsilon\cdot\mathrm{poly}\log(1/\varepsilon)$）的算法；通过在开始时将 ε 缩小一个常数因子（或对数因子），可以使其达到 $1-\varepsilon$。

2.2 和集与子集和

在多重集 A 中，元素 a 可能出现多次（其出现次数即 A 中 a 的重数）。我们用 $A\uplus B$ 表示不去除重复项的并集（即结果可能是多重集）。

对于多重集 $Y\subset\mathbb{N}$，令 $\Sigma(Y)=\sum_{y\in Y}y$ 表示其元素之和（不去除重复项）。

对于多重集 $X\subset\mathbb{N}$，令 $\mathcal{S}(X)=\{\Sigma(Y)\colon Y\subseteq X\}$ 表示其子集和的集合，并令$\mathcal{S}(X;t)=\mathcal{S}(X)\cap[0,t]$ 表示不超过 t 的子集和集合。

对于一个数 c和一个集合 X，定义 $c\cdot X=\{cx\colon x\in X\}$。对于两个集合 X,Y，定义它们的和集$X+Y=\{x+y\colon x\in X,y\in Y\}$。给定集合 $X\subseteq[n],Y\subseteq[n]$，可使用快速傅里叶变换在 $O(n\log n)$ 时间内计算和集 $X+Y$。这个简单事实可以直接推广到二维情形，陈述如下。

引理 2.1（二维快速傅里叶变换，例如 [Bla10，第 12 章。8]）。给定两个集合 A_1, $A_2\subseteq[n]\times[m]$，可以计算

$$A_1+A_2:=\big\{(x_1+x_2,y_1+y_2):(x_1,y_1)\in A_1,(x_2,y_2)\in A_2\big\}$$

以 $O(nm\log(nm))$ 时间确定性地。

2.3 背包问题与利润函数

在背包问题中，假设对于每个物品 i，都有 $0 < w_i \le W$ 和 $p_i > 0$。那么，最大总利润的一个平凡下界是 $\max_j p_j$。一开始，我们可以丢弃所有满足 $p_i \le \frac{\varepsilon}{n}\max_j p_j$ 的物品 i，这样总利润至多减少 $\varepsilon\max_j p_j$，这仅占最优总利润的 $O(\varepsilon)$。因此，我们可以假设 $\max_j p_j \min_j p_j \le \frac{n}{\varepsilon}$。

对于物品集合 I，我们用 f_I 表示其利润函数，定义如下：

$$f_I(x) = \max\left\{\sum_{i\in J} p_i : \sum_{i\in J} w_i \le x,\ J \subseteq I\right\}$$

在 $x \in [0, +\infty)$。注意，f_I 是单调非递减阶梯函数。采用 Chan [Cha18], 的术语，单调阶梯函数的复杂度指其阶梯数。

设 I_1、I_2 是两个互不相交的物品子集，且 $I = I_1 \uplus I_2$。不难看出 $f_I = f_{I_1} \oplus f_{I_2}$，其中 $\oplus$ 表示（max +）卷积，定义为 $(f\oplus g)(x) = \max_{0\le x'\le x}(f(x')+g(x-x'))$。

2.4 $(1-\delta,\ \Delta)$ 近似至 t

我们的背包问题和划分问题算法都经常使用（$1-\delta$, Δ）-近似至 t 的概念。其定义类似，如下所述。

定义 2.2 (利润函数的近似)。对于函数 $f, \tilde{f}$ 和实数 $t, \Delta \in \mathbb{R}_{\ge 0}, \delta \in [0,1)$，我们称 $\tilde{f}$ 是 $(1-\delta, \Delta)$ 近似于 f，直至 t，如果

$$\tilde{f}(w) \le f(w)$$

对所有 $w \ge 0$ 均成立，且

$$\tilde{f}(w) \ge (1-\delta)f(w) - \Delta$$

当 $f(w) \le t$，$w \ge 0$ 时成立。

以下近似概念对我们的划分问题算法很有用。文献中将类似概念称为“弱近似” [MWW19、BN21], ，与之相对的是近似一般子集和实例所需的“强近似”。

定义 2.3 (整数集的近似)。对于整数集 A，$B \subseteq \mathbb{N}$ 和实数t，$\Delta \in \mathbb{R}_{\ge 0}$，$\delta \in [0,1)$，我们称 A 是 $(1-\delta,\ \Delta)$ 近似于 B 直至 t，如果

1. f对于每个 $b \in B \cap [0,t]$，都存在 $a \in A$ 使得 $(1-\delta)b - \Delta \le a \le b$，并且，
2. 对于每个 $a \in A$，都存在 $b \in B$ 使得 $(1-\delta)b - \Delta \le a \le b$。

在这种情况下，不失一般性，可以假设 $A \subseteq \mathbb{N} \cap [0,t]$。

对于 $t = +\infty$ 的情况，我们只需省略“至多 t”这一短语。

我们也将 $(1, \Delta)$ 近似称为 Δ-加性近似，并将 $(1-\delta, 0)$ 近似称为 $(1-\delta)$-乘性近似，或简称为 $(1-\delta)$ 近似。

关于近似合并和集和利润函数，我们有以下简单事实。

命题 2.4. 对于 $i \in \{1,2\}$，假设 A_i 是一个$(1-\delta, \Delta_i)$近似，在 t范围内近似(X_i) $\mathcal{S}$。那么，$(A_1 + A_2) \cap [0,t]$ 就是一个$(1-\delta, \Delta_1+\Delta_2)$近似，在 t范围内近似$(X_1 \uplus X_2)$ $\mathcal{S}$。

证明。对于任意 $b \in \mathcal{S}(X_1 \uplus X_2) \cap [0,t]$ 其中 $b = b_1 + b_2$对于 $b_i \in \mathcal{S}(X_i) \cap [0,t]$ $(i \in \{1,2\})$，存在 $a_i \in A_i$ 使得 $(1-\delta)b_i - \Delta_i \le a_i \le b_i$。因此，$a_1 + a_2 \le b_1 + b_2 = b \le t$，且$a_1 + a_2 \ge (1-\delta)b_1 - \Delta_1 + (1-\delta)$ $b_2 - \Delta_2 = (1-\delta)b - (\Delta_1 + \Delta_2)$。

反向结论可类似地验证。 □

下面的事实可以类似地证明。

命题 2.5。对于 $i \in \{1,2\}$，假设 f_i 是利润函数，f_{I_i} 达到 t的$(1-\delta,\ \Delta_i)$ 近似。那么，$(f_1 \oplus f_2)$ 是 $f_{I_1 \uplus I_2}$ 达到 t的$(1-\delta,\ \Delta_1+\Delta_2)$ 近似。

沿着 Chan [Cha18]和 Jin [Jin19], 的思路，给定一个单调阶梯函数 f（我们有时也称之为利润函数，尽管它未必等于任何特定物品集合 I的利润函数 f_I），其值域包含于 $\{0\} \cup [A,B]$，可以将其向下舍入到 $1/(1-\varepsilon)$ 的幂次，得到另一个利润函数 $\tilde{f}$，其复杂度仅为 $O(\varepsilon^{-1}\log(B/A))$，且以 $(1-\varepsilon)$-近似 f。在我们的算法中，始终有 $B/A \le \text{poly}(n/\varepsilon)$，因此每次付出 $(1-\varepsilon)$ 的近似因子后，我们总可以假设算法计算出的中间利润函数是复杂度为 $\tilde{O}(\varepsilon^{-1})$ 的单调阶梯函数。

2.5 加法组合学

我们需要若干关于稠密子集和的结果，这些结果由一系列研究逐步建立，其中包括 [Sár94、GM91、Lev03、BW21]。下面的结构引理可由 Bringmann 和 Wellnitz 的定理 4.1 和定理 4.2 推出 [BW21]。

引理 2.6。给定 n 互不相同的正整数 $X = \{x_1, \ldots, x_n\} \subseteq [\ell, 2\ell]$，其中 $\ell = o(n^2/\log n)$。

那么，存在一个普适常数 $c \ge 1$，对于每个 $c\ell^2/n \le t \le \Sigma(X)/2$，都存在 $t' \in \mathcal{S}(X)$ 使得 $0 \le t' - t \le 8\ell/n$。

引理 2.6 的证明见附录 B。

下面的算法引理可由 [BW21], 的主定理推出，并将在我们的 Partition 算法中使用。

引理 *2.7*（由 *[BW21]* 推出）。给定 n 互不相同的正整数 $X = \{x_1, \ldots, x_n\} \subseteq [\ell, 2\ell]$，存在 $\lambda = \tilde{\Theta}(\ell^2/n)$，使得若 $\lambda \le \Sigma(X)/2$，则我们可以在 $\tilde{O}(n)$ 时间内构造一个确定性数据结构，以支持以下查询，查询时间为 $O(1)$：给定 L, R 使得$\lambda \le L \le R \le \Sigma(X)/2$，报告是否存在 $t \in [L,R]$ 使得 $t \in \mathcal{S}(X)$。

备注 2.8。我们指出，*[BW21]* 中陈述的主定理只支持查询给定的 $\lambda \le t \le \Sigma(X)/2$ 是否是一个子集和。在我们的应用中，我们需要支持范围查询的版本。通过在证明 *[BW21]*定理 *4* 时构建一个额外的前缀和数组，即可轻松实现这一点。该版本支持区间和查询。*6]*,

3 背包近似

3.1 已知引理

根据已有的归约（例如 [Cha18、Jin19]），我们可以专注于解决下面这个更简洁的问题，它已经涵盖了背包问题的主要难点。

问题1。 假设 $\varepsilon \in (0,\ 1/2)$ 且 $1/\varepsilon \in \mathbb{N}^+$。给定一个物品列表 $I(p_1,\ w_1) \ldots,\ (p_n,\ w_n)$，其重量 $w_i \in \mathbb{N}$ 和利润 p_i 均为 ε 的倍数，且位于区间 $[1, 2)$内，计算一个利润函数，使其在 $2/\varepsilon$范围内对 f_I 进行$(1-\varepsilon)$近似。

~~引理3.1~~。 如果存在某个 $c \geq 2$，问题 *1* 可以在 $O(n + 1/\varepsilon^c)$ 时间内求解，那么 $(1-\varepsilon)$-~~近似背包问题也可以在 $O(n + 1/\varepsilon^c)$ 时间内求解。

引理3.1 将在附录中证明。

基于 Chan 的技术[Cha18], Jin [Jin19] 得到了以下关于$(1-\varepsilon)$-近似背包问题的引理，其误差较小 B，或者不同价值的数量较少 p_i。

引理 *3.2*（由 [*Jin*19] 的引理*17*推出）。给定一个物品列表 I，其中物品为 $(p_1,\ w_1)$ ..... $(p_n,\ w_n)$，其重量 $w_i \in \mathbb{N}$ 和利润 p_i 是区间 $[1, 2$ 中 ε 的倍数，则可以在 $\tilde{O}\ (n + \varepsilon^{-2}B^{1/3}/2^{\Omega(\sqrt{\log(1/\varepsilon)})})$ 时间内，将利润函数 f_I 近似到 B，近似比为 $(1-\varepsilon)$。

引理 *3.3*（由 [*Cha*18] 的定理*3*推出）。给定一个物品列表 I，其中物品为 $(p_1,\ w_1)$ ..... $(p_n,\ w_n)$，其重量 $w_i \in \mathbb{N}$ 和利润 p_i 是区间 $[1, 2$ 中 ε 的倍数；如果只有 m 个不同的利润值 p_i，则可以在 $\tilde{O}$ $(n+\varepsilon^{-3/2}m/2^{\Omega(\sqrt{\log(1/\varepsilon)})})$ 时间内，将利润函数 f_I 近似到 $1-\varepsilon$，近似比为 $(1-\varepsilon)$。 [3]

下面的有用引理使我们能够合并多个利润函数。该引理由 Chan 使用分治以及改进的 (最小 $+$)-卷积算法 [BCD+14、Wil14、CW16]证明。

引理 *3.4*（[*Cha18*，引理 *2(i)*]）。设 f_1，...，f_m 为单调阶梯函数，其总复杂度为 $O(n)$，且值域包含于 $\{0\} \cup [A, B]$ 中。则我们可以计算出一个复杂度为 $\tilde{O}(\frac{1}{\varepsilon} \log\ B/A)$ 的单调阶梯函数，并在 $O(n)\ +\tilde{O}((\frac{1}{\varepsilon})^2 m/2^{\Omega(\sqrt{\log(1/\varepsilon)})} \log\ B/A)$ 时间内，以 $(1\ -\ O(\varepsilon))$ 近似 $f_1 \oplus \cdots \oplus f_m$。

3.2 通过稠密子集和进行贪心交换论证

本节的目标是证明以下引理 3.5。我们的算法基于一种类似于[Jin19 的贪心交换论证，即引理 20]，，但通过结合关于稠密子集和的数论结果，我们可以获得更好的界。

引理 3.5。 给定一个包含 n 个物品的列表 I，其中 p_i 是区间 $[1, 2)$内 ε 的倍数，以及满足 $m = O(1/\varepsilon)$ 的整数 $1 \leq m \leq n$，可以在 $O(n+\varepsilon^{-11/5}/2^{\Omega(\sqrt{\log(1/\varepsilon)})})$ 时间内计算出一个利润函数，使其以 f_I 为基准，在 $2m$ 范围内实现 $m\varepsilon)$ 加性近似。

[3]在我们论文的会议论文集版本中，我们错误地声称引理 3.3 中的任务可以在$\tilde{O}(n + \varepsilon^{-3/2}m^{3/4}/2^{\Omega(\sqrt{}_{\log(1/\varepsilon)})})$ 时间内完成。此处已更正引理 3.3 的表述。因此，第 3.2 节中的若干参数也相应作了调整。这项更正并未影响主要结果（定理 1.1）的最终时间界，因为在我们的算法中，调用引理 3.3 的步骤并非瓶颈。

引理 3.5 的证明假设以下结论成立，该结论将在后续章节中通过随机划分予以证明。

引理 3.6. 给定一个包含 I 个 $n = O(1/\varepsilon)$ 物品的列表，其中 p_i 是区间 $[1,2)$内 ε 的倍数，可以在 $\tilde{O}(n^{4/5}\varepsilon^{-7/5}/2^{\Omega(\sqrt{\log(1/\varepsilon)})})$ 时间内计算出一个利润函数，使其对f_I具有$(n\varepsilon)$-加性近似。

现在我们描述引理 3.5的算法。给定物品 $(p_1,w_1)\ldots,(p_n,w_n)$，其中 $p_i \in [1,2)$ 是 ε的倍数，我们按效率非递增的顺序对它们排序，$p_1/w_1 \geq p_2/w_2 \geq \cdots \geq p_n/w_n$。然后，我们考虑这一物品序列的各个前缀，并定义如下多样性度量：

定义 3.7 $(D(i))$**.** 对于 $1 \leq i \leq n$，令 $D(i) =$ 最小值$_J\, C([i]\setminus J)$，其中最小化是在所有满足 $|J| \leq 2m$ 的子集 $J \subseteq [i]$ 上进行的；$C([i]\setminus J)$表示 $\{p_j\colon j \in [i]\setminus J\}$ 中不同价值的数量。

关于 $D(i)$，我们有以下直接观察。

观察 3.8. *1.* 对于 $2 \leq i \leq n$，$0 \leq D(i) - D(i-1) \leq 1$。

2. $D(i)$ （以及极小化者 J）可以通过以下贪心算法在 $O(i)$ 时间内计算出来：首先取所有价值 p_1，$p_2,\ldots,p_i$。重复以下操作，最多执行 $2m$ 次：移除重数最小的价值 p_j，并将 j 加入 J。

现在，我们设定参数 $\Delta = \lfloor \varepsilon^{-5/8} \rfloor$。定义 $i \in \{1,2,\ldots,n\}$ 为满足以下条件的最大值：$D(i) \leq \Delta$，可通过观察 3.8 结合二分查找在 $O(n)$ 时间内找到。

以下引理是我们证明引理 3.5 的关键部分。

引理 3.9 （贪心交换引理）。设 $S \subseteq [n]$ 是一个总利润为 $\sum_{s\in S} p_s \leq 2m$的物品集合。设 $B := 9c\varepsilon^{-1}/\Delta$，其中 $c \geq 1$ 是引理 *2.6*中的普适常数。

那么，存在一个物品集合 $S \subseteq [n]$，使得由其中的物品 $[n]\setminus[i]$在 S 贡献的总利润 $\tilde{p}$ 满足

$$\tilde{p} := \sum_{s\in \tilde{S}\cap([n]\setminus[i])} p_s \leq B, \tag{1}$$

and

$$\sum_{s\in\tilde{S}} p_s \geq (1-\varepsilon)\sum_{s\in S} p_s, \tag{2}$$

and

$$\sum_{s\in\tilde{S}} w_s \leq \sum_{s\in S} w_s. \tag{3}$$

证明。如果 $D(i) < \Delta$，那么根据 i 的定义，我们有 $i = n$，并且可以直接令 $S = S$，因为 $\tilde{p} = 0$ 始终成立。因此，接下来我们假设 $D(i) = \Delta$。

我们将 $S \subseteq [n]$ 定义为以下表达式的最大化者

$$\sum_{s\in\tilde{S}\cap[i]} p_s + \sum_{s\in\tilde{S}\cap([n]\setminus[i])} (1-\varepsilon)p_s$$

~ ~在所有 S 满足 $\sum_{s\in\tilde{S}} w_s \le \sum_{s\in S} w_s$ 且 $\sum_{s\in\tilde{S}} p_s \le \sum_{s\in S} p_s$的条件下。我们声称， S 满足以下~性质（1）、（2）、（3）。注意，（2）、（3）可由 S的定义直接推出。主要工作是验证（1）。

~反设（1）不成立。那么，我们可以找到一个子集 $K \subseteq S \cap ([n] \setminus [i])$，其总利润为 $p^* = \sum_{k\in K} p_k \in (B, B+2]$，并且可以通过从~$S \cap ([n] \setminus [i])$ (回想一下，每件物品的利润都在 $[1,2)$).

定义物品集合 $I' := [i] \setminus \tilde{S}$。由于 $S \le \sum_{s\in\tilde{S}} p_s / \min_{s\in\tilde{S}} p_s \le \sum_{s\in\tilde{S}} p_s \le \sum_{s\in S} p_s \le 2m$，根据 $D(i)$的定义，我们知道 $\{p_i : i \in I'\}$ 至少包含 $D(i)$个不同元素。

我们将引理 2.6 应用于整数集合 $X = \{p_i/\varepsilon : i \in I'\} \subseteq [1/\varepsilon, 2/\varepsilon)$，该集合至少包含 Δ 个不同整数，其中引理 $1/\varepsilon = o(\Delta^2/\log \Delta)$ 2.6 中的前提 2.6 由我们对 $\Delta = \lfloor \varepsilon^{-5/8} \rfloor$的选择满足。引理 2.6 指出，对于每个 $t \in [c\varepsilon^{-2}/\Delta, 0.5\Delta/\varepsilon]$，都存在 $t' \in \mathcal{S}(X)$使得 $0 \le t' - t \le 8\varepsilon^{-1}/\Delta$。这里我们令

$$t := \frac{(1-\varepsilon)p^*}{\varepsilon} + \frac{\varepsilon^{-1}}{\Delta},$$

满足 $t > p^*(1-\varepsilon)/\varepsilon > (1-\varepsilon)B/\varepsilon = (1-\varepsilon)(9c\varepsilon^{-1}/\Delta)/\varepsilon > c\varepsilon^{-2}/\Delta$，并且 $t < (B+2)/\varepsilon + \varepsilon^{-1}/\Delta = 9c\varepsilon^{-2}/\Delta + 2/\varepsilon + \varepsilon^{-1}/\Delta \le O(\varepsilon^{-11/8}) \le 0.5\Delta/\varepsilon$。那么引理 2.6 的结论表明，存在一个 $R \subseteq I'$ 物品子集，其总利润为 $\tilde{p} := \varepsilon \cdot t'$，满足

$$1/\Delta \le \tilde{p} - p^*(1-\varepsilon) \le 9/\Delta. \tag{4}$$

注意，(4) 蕴含

$$\begin{aligned} p^* - \tilde{p} &\ge \varepsilon \cdot p^* - 9/\Delta \\ &> \varepsilon \cdot B - 9/\Delta \\ &= \varepsilon \cdot 9c\varepsilon^{-1}/\Delta - 9/\Delta \\ &\ge 0. \end{aligned}$$

~ ~回想一下， $R \subseteq I' = [i] \setminus S$ 和 $K \subseteq S \cap ([n] \setminus [i])$，它们都必须非空。由于物品的效率按非递增顺序排列，因此我们有 $\min_{r\in R} p_r/w_r \ge \max_{k\in K} p_k/w_k$。现在定义物品集合

$$\tilde{S}' := (\tilde{S} \setminus K) \cup R.$$

于是，我们有

$$\begin{aligned} \sum_{s\in\tilde{S}} p_s - \sum_{s\in\tilde{S}'} p_s &= \sum_{k\in K} p_k - \sum_{r\in R} p_r \\ &= p^* - \tilde{p} \\ &\ge 0, \end{aligned}$$

and

$$\begin{aligned}\sum_{s\in\tilde{S}} w_s - \sum_{s\in\tilde{S}'} w_s &= \sum_{k\in K} w_k - \sum_{r\in R} w_r \\ &\geq \frac{\sum_{k\in K} p_k}{\max_{k\in K}(p_k/w_k)} - \frac{\sum_{r\in R} p_r}{\min_{r\in R}(p_r/w_r)} \\ &\geq \frac{1}{\min_{r\in R}(p_r/w_r)} \cdot \left(\sum_{k\in K} p_k - \sum_{r\in R} p_r\right) \\ &= \frac{1}{\min_{r\in R}(p_r/w_r)} \cdot (p^* - \tilde{p}) \\ &\geq 0.\end{aligned}$$

因此，$\sum_{s\in\tilde{S}'} p_s \leq \sum_{s\in\tilde{S}} p_s$ 并且 $\sum_{s\in\tilde{S}'} w_s \leq \sum_{s\in\tilde{S}} w_s$。另一方面，根据（4），我们知道

$$\begin{aligned}&\left(\sum_{s\in\tilde{S}'\cap[i]} p_s + \sum_{s\in\tilde{S}'\cap([n]\setminus[i])} (1-\varepsilon)p_s\right) - \left(\sum_{s\in\tilde{S}\cap[i]} p_s + \sum_{s\in\tilde{S}\cap([n]\setminus[i])} (1-\varepsilon)p_s\right) \\ &= \sum_{r\in R} p_r - \sum_{k\in K} (1-\varepsilon)p_k \\ &= \tilde{p} - (1-\varepsilon)p^* \\ &\geq 1/\Delta > 0,\end{aligned}$$

这与 S 是最大化者的定义相矛盾。

因此，我们已经证明 S 满足（1）。 □

现在我们可以证明引理 3.5。

引理 *3.5* 的证明。回想一下，$i \in \{1, 2, \ldots, n\}$ 是满足以下条件的最大值：$D(i) \leq \Delta$，可以通过观察结果 3.8 利用二分查找在 $O(n)$ 时间内找到。令 $J \subset [i]$ 其中 $|J| \leq 2m$ 是 $D(i)$的最小化者。

现在，我们使用不同的算法近似计算三个物品集合 J、$[i]\setminus J$ 和 $[n]\setminus[i]$ 的利润函数 f_J、$f_{[i]\setminus J}$、$f_{[n]\setminus[i]}$，如下所述：

1. 使用引理 3.6 计算 f_J，使其 $(2m\varepsilon)$-可加性地近似 f_J，时间为 $O(m^{\frac{4}{5}}\varepsilon^{-\frac{7}{5}}/2^{\Omega(\sqrt{\log(1/\varepsilon)})}) \leq O(\varepsilon^{-\frac{7}{5}}/2^{\Omega(\sqrt{\log(1/\varepsilon)})})$ 时间。2. 根据 i 的定义，$[i]\setminus J$ 中的物品最多只有 Δ 个不同的利润值。因此，我们可以使用引理 3.3，在 $\tilde{O}(\Delta\varepsilon^{-3/2}) = \tilde{O}(\varepsilon^{-17/8})$ 时间内计算出 f_2，使其以 $(1-\varepsilon)$ 近似 $f_{[i]\setminus J}$。3. 使用引理 3.2，在 $\tilde{O}(B^{1/3}\varepsilon^{-2}) \leq \tilde{O}(\varepsilon^{-17/8})$ 时间内计算出 f_3，使其以 $(1-\varepsilon)$ 近似 $f_{[n]\setminus[i]}$，精确到 $B = \Theta(\varepsilon^{-1}/\Delta)$（定义见引理 3.9）。

最后，合并 t 这三个部分 f_1、f_2、f_3 使用引理 3.4 在 $\tilde{O}(\varepsilon^{-2})$ 时间内完成，并返回结果 .[4]

[4]尽管第二种和第三种算法的运行时间受第一种算法主导，但由于引理的参数设置存在各种约束，简单的参数重新平衡似乎无法改善复杂度3.6。

在第三部分中，引理 B 3.9 证明了只计算至 的正确性。该引理表明，如果我们只考虑总利润至多为 $2m$的近似集合，那么我们可以假设 $[n] \setminus [i]$ 中的物品所贡献的利润至多为 B (1)，代价仅为引入一个 $(1-\varepsilon)$近似因子（2）。

分析误差时请注意，第一部分会产生 $(2m\varepsilon)$ 的加性误差。第二、第三部分以及最后的合并步骤会产生 $(1-O(\varepsilon))$ 的乘性误差；由于我们只关心近似到 $2m$，这会转化为 $O(m\varepsilon)$ 的加性误差。因此，总体加性误差为 $O(m\varepsilon)$，通过降低 ε 的值，可以使其 $m\varepsilon$ 。

□

现在我们证明，引理 3.5 可用于求解问题 1，这足以证明定理 1.1。

定理 *1.1* 的证明。要解决问题 1，我们将 $[1, 2\varepsilon^{-1})$ 划分为 $O(\log(1/\varepsilon))$ 个区间$[m, 2m)$，其中 m 是 2 的幂，并使用引理 3.5 获得利润函数，使其在截至 $2m$的范围内达到 $m\varepsilon$-加性近似。然后，取这些函数的逐点最小值，即可得到一个 $(1-O(\varepsilon))$ 近似。

□

在接下来的章节中，我们将证明引理 3.6。

3.3 使用小集合 Δ 的 Δ-倍数进行近似

我们首先介绍从先前研究中借鉴的几种额外工具，这些工具将在引理 3.6 的最终证明中用到。

沿用 [Cha18] 的术语，若单调阶梯函数的函数值为 0、p、$2p$、……、lp，其中 l，则称该函数为 p-均匀函数。若一个 p-均匀函数的连续 x-断点之间的间距从左向右单调不减，则称其为伪凹函数。一个 p-均匀且伪凹的函数示例是由具有相同利润 $p_i = p$ 的物品集合 I 所定义的利润函数 f_I，它可以通过简单的贪心算法精确计算：函数 f_I 的取值为 0、p、$2p$、……、np，其 x-断点为 w_1、$w_1 + w_2$、……、$w_1 + \cdots + w_n$，其中 w_i 按非递减顺序排列。

与 [Cha18] 和 [Jin19], 一样，我们将采用通过 Δ-倍数进行近似的方法。对于一个数的Δ 集合，我们称 p 为 Δ-倍数，如果它是某个 δ 的倍数，其中 $\delta \in \Delta$。Chan[Cha18]使用 SMAWK 算法 [AKM+87] 并结合适当的舍入，证明了以下引理:

引理 3.10 ([Cha18，引理 5])。 设 $f_1, \ldots, f_m$ 是值域包含于 $[0, B]$的单调阶梯函数。设 $\Delta \subset [\delta, 8\delta]$。若每个 f_i 对于某个 p_i-均匀且伪凹，其中 $p_i \in [1, 2]$ 是一个 Δ-倍数，则我们可以在 $O(|\Delta|\delta)$时间内，计算出一个$\tilde{\min}\{f_1 \oplus \cdots \oplus f_m, B\}$ 的 $O(Bm/\delta)$ 加法近似的单调阶梯函数。

Chan [Cha18] 给出了一个小集合的构造 Δ 使得 $[1, 2]$ 中的每个实数都可以用一个 Δ的倍数近似。这里，我们给出一种更简洁的构造。

引理 3.11。 对于参数 $0 < \varepsilon < \delta < 1/2$，令 $r = \lceil \log_{1+\varepsilon}(1+2\delta) \rceil = O(\delta/\varepsilon)$，并定义$a_i = \delta(1+\varepsilon)^i$，其中 $0 \le i \le r+1$。令 $\Delta = \{a_i\}$ 为 a_i的集合。那么对于任意 $t \in [1, 2]$，在范围 $[t, t+2\varepsilon]$内存在某个 a_i 的倍数。因此，$[1, 2]$ 中的每个实数都可以用一个 Δ 倍数来近似，误差为 $O(\varepsilon)$ 加性误差，其中 $|\Delta| = r + 2 = O(\delta/\varepsilon)$，且 Δ 中的所有元素都在 $[\delta, 8\delta]$以内。

证明。令 c为满足 $(t+2\varepsilon)/c \geq a_0$的最大整数。由于 c最大，故 $c+1 \geq (t+2\varepsilon)/a_0 \geq 1/\delta$，因此 $(c+1)/c \leq (1/\delta)/(1/\delta-1) = 1+\delta/(1-\delta) \leq 1+2\delta \leq a_{r+1}/a_0$。由于 $(t+2\varepsilon)/(c+1) < a_0$，我们知道 $(t+2\varepsilon)/c < a_{r+1}$。令 k 为 $[0,r]$ 中满足 $a_k \leq (t+2\varepsilon)/c$的最大整数。注意到 $a_{k+1} > (t+2\varepsilon)/c$，因此利用 $t \leq 2$这一事实可得 $a_k = a_{k+1}/(1+\varepsilon) \geq t/c$。由此，$a_k \in [t/c,(t+2\varepsilon)/c]$，所以 $ca_k \in [t,t+2\varepsilon]$。□

3.4 随机划分

假设 $n < 1/\varepsilon$。在本节中，我们将使用随机划分来证明引理 3.6，其重述如下。

引理 3.6。 给定一个包含 I $n = O(1/\varepsilon)$ 个物品的列表，其中 p_i 是区间 $[1,2)$内 ε 的倍数，可以在 $\tilde{O}(n^{4/5}\varepsilon^{-7/5}/2^{\Omega(\sqrt{\log(1/\varepsilon)})})$ 时间内计算出一个利润函数，使其对f_I实现 $(n\varepsilon)$ 加性近似。

证明。令 $\Delta_1 = \Theta(\sqrt{n})$，并令 $\Delta_0 = \Theta(n^{\frac{7}{10}}\varepsilon^{\frac{2}{5}}2^{c\sqrt{\log(1/\varepsilon)}})$，其中 $c > 0$ 是某个较小的常数。不失一般性，假设 Δ_0 是 2 的幂。注意，$\Delta_0 = O(\Delta_1)$，这由 $n = O(1/\varepsilon)$ 得出。

命题 3.12。 我们可以将 I 中的元素划分为 $\Theta(\Delta_1)$ 组 $G_1, G_2, \ldots, G_k$，每组大小为 $O(n/\Delta_1)$，同时每组 G_i 中的所有元素都可由某个 $p_i = \Theta(\Delta_1\varepsilon)$ 的倍数以 p_i 的 $(1+\varepsilon)$ 精度近似。

证明。在引理 3.11中代入 $\delta = \varepsilon\Delta_1$，可得到一个集合 A，其大小为 $O(\Delta_1)$，其中元素的阶为 $\Theta(\Delta_1\varepsilon)$，且 I中的每个物品都可以通过 $(1+\varepsilon)$ 精度由 A 的倍数近似。

我们根据元素在 A 中的除数对 I 中的元素分组。然后，将大小超过 n/Δ_1 的组均匀拆分为两组，重复此过程，直到所有组的大小都不超过 n/Δ_1。□

从现在起，假设 $G_1, G_2, \ldots, G_k$ 是满足命题 3.12 中条件的组。

现在，我们通过将每个 $1 \leq i \leq k$独立且均匀地分配到某个 $I_j(1 \leq j \leq \Delta_0)$中，随机地将 $\{1,2,\ldots,k\}$ 划分为 Δ_0 个部分，I_1、.....、I_{Δ_0}。然后，对每个$1 \leq j \leq \Delta_0$，令 $X_j = \bigcup_{i\in I_j} G_i$。显然，$\{X_j\}$ 是 I的一个划分。

Cl引理 3.13。 以 $\geq 3/4$的概率，$|I_j| = O(\Delta_1/\Delta_0)$，因此 $|X_j| \leq O(n/\Delta_0$).

证明。根据 Chernoff 界[5],，对于某个足够大的常数 $c > 0$，$|I_j| \geq ck/\Delta_0$ 发生的概率至多为 $1/(4\Delta_0)$。因此 $|I_j| = O(\Delta_1/\Delta_0)$ 以概率 j 对所有 $\geq 3/4$ 都成立，这由并集界可得。根据断言 3.12，$|X_j| \leq |I_j|O(n/\Delta_1) = O(m/\Delta_0)$。□

现在假设命题 3.13 中的事件发生。

命题 3.14。 我们可以在 $1 \leq j \leq \Delta_0$ 属于$\tilde{O}(n^2\varepsilon^{-1}/(\Delta_0\Delta_1)) = \tilde{O}(n^{4/5}\varepsilon^{-7/5}/2^{\Omega(\sqrt{\log(1/\varepsilon)})})$ 时，以加性误差 $O(n\varepsilon/\Delta_0)$ 近似 $\bigoplus_{x\in X_j} f_x$。

对于独立随机变量 $x_1, \ldots, x_n \in \{0,1\}$和 $\delta > 0, 0 \leq w_1, \ldots, w_n \leq 1$，令 $X = \sum_{i=1}^n w_i x_i$和$\mu = \mathbb{E}[x]$，则 Pr $[|x-\mu| \geq \delta\mu] \leq 2e^{-\delta^2\mu/3}$。

证明。固定一个 j。根据命题 3.13，$\bigoplus_{x\in X_j} f_x$ 是 $O(n/\Delta_0)$ 个元素的卷积，其中每个元素都是阶为 $\Theta(\Delta_1\varepsilon)$ 的倍数。通过对参数 $B = O(n/\Delta_0)$ $\delta = \Theta(\Delta_1\varepsilon)$ $|\Delta| = |I_j| = O(\Delta_1/\Delta_0)$ 应用引理 3.10，我们可以在时间 $\tilde{O}((n/\Delta_0)^2/(\Delta_1\varepsilon))$ 内近似 $\bigoplus_{x\in X_j} f_x$，加性误差为 $O(\Delta_1^2\varepsilon/\Delta_0) = O(n\varepsilon/\Delta_0)$。

我们可以对所有 $1 \le j \le \Delta_0$ 这样做，运行时间为 $\tilde{O}(n^2/(\Delta_0\Delta_1\varepsilon)) = \tilde{O}(n^{4/5}\varepsilon^{-7/5}/2^{\Omega(\sqrt{\log(1/\varepsilon)})})$。□

现在，我们可以用以下近似值替换 $\bigoplus_{x\in X_j} f_x$，该近似值由断言 3.14 得出，因为造成的总加性误差为 $O(n\varepsilon/\Delta_0)\Delta_0 = O(n\varepsilon)$。

我们使用分治来合并 $\bigoplus_{x\in X_j} f_x$的答案。合并过程可以看作一棵具有 Δ_0 个叶子的完全二叉树。对于 $S \subseteq \{1, 2, \ldots, \Delta_0\}$，定义 $F(S) = \bigoplus_{x\in\cup_{(s\in S)} X_s} f_x$。断言 3.14 使我们能够对所有 $F(S)$ 进行近似 $|S| = 1$。现在，我们有以下关于合并两个子树 S_1 和 S_2的断言。

命题 3.15。假设 $i \le \log_2 \Delta_0$ 且 $|S_1| = |S_2| = 2^i$，其中 $S_1, S_2 \subseteq \{1, 2, \ldots, \Delta_0\}$ 且 $S_1 \cap S_2 = \emptyset$。假设 A_1 是 $F(S_1)$ 的一个近似，加性误差为 err_1；A_2 是 $F(S_2)$ 的一个近似，加性误差为 err_2。那么，以 $\ge 1 - 1/(5\Delta_0)$ 的概率，我们可以在$O(\varepsilon^{-2}\Delta_0^{0.5}/(\Delta_1^{0.5}2^{\Theta(\sqrt{\log(1/\varepsilon)})}))$时间内计算出 $F(S_1 \cup S_2)$ 的一个近似，加性误差为 $err_1 + err_2 + O(2^{0.9i}n\varepsilon/\Delta_0)$。

证明。定义 $\delta_i = 2^{0.9i}n\varepsilon/\Delta_0$。近似 $F(S_1 \cup S_2)$ 的一种朴素方法是将A_1 和 A_2 中的每个值都舍入到 δ_i的倍数，然后像引理 3.4 中那样调用 $(\min +)$ 卷积。接下来我们将介绍一种更好的方法，它利用了 $\{X_i\}$是一个随机划分这一事实。

设全局最优解为选择物品的子集 T。定义 $H_1 = \bigcup_{i\in S_1} X_i, H_2 = \bigcup_{i\in S_2} X_i$。注意，组 $G_1, \ldots, G_k$ 被均匀随机地分配到 $X_1, \ldots, X_{\Delta_0}$ 中。取 $u = Cn\sqrt{2^i/(\Delta_1\Delta_0)\log n}$，其中 $C > 0$为一个足够大的常数。根据 Chernoff 界，$\Pr(|\sum_{x\in T\cap H_1} x - \sum_{x\in T\cap H_2} x| \ge u) \le 1/(5n)$ 的概率为 [6]。

现在假设 $|\sum_{x\in T\cap H_1} x - \sum_{x\in T\cap H_2} x| \le u$，并说明如何在该假设下近似 $F(S_1 \cup S_2)$。在进行 $(\min +)$ 卷积时，我们首先将 $F(S_1)$ $F(S_2)$ 的价值舍入到 δ_i 的整数倍。接下来只需考虑价值相差不超过 u 的数对。然后，我们将数组划分为若干块，使每块中的价值彼此相差不超过 u，并对索引相差至多 1 的块对进行 $(\min +)$-卷积。块大小至多为 $u/\delta_i = \tilde{O}(\varepsilon^{-1}\Delta_0^{0.5}\Delta_1^{-0.5}/2^{0.4i})$，因此，使用 Williams 的 $O(n^2/2^{\Omega(\sqrt{\log(1/\varepsilon)})})$-时间算法计算长度为 n 的 $(\min +)$-卷积 [Wil14]，每次 $(\min +)$-卷积的运行时间为 $O(\varepsilon^{-2}\Delta_0/(\Delta_1 2^{0.8i}2^{\Omega(\sqrt{\log(1/\varepsilon)})}))$。根据命题 3.13，合并后答案中的价值有界于 $\tilde{O}(2^i n/\Delta_0)$，因此总共需要进行 $\tilde{O}(2^i n/(\Delta_0 u))$ 次最小加卷积，总复杂度为 $\tilde{O}(\varepsilon^{-2}2^i n/(\Delta_1 2^{0.8i}2^{\Omega(\sqrt{\log(1/\varepsilon)})}u)) = O(\varepsilon^{-2}\Delta_0^{0.5}/(\Delta_1^{0.5}2^{\Theta(\sqrt{\log(1/\varepsilon)})}))$。□

现在，我们通过将3.15 中的结论应用于分治过程，完成证明。假设所有 $\le \Delta_0$ 次对结论3.15 的调用都得到了正确的近似，其成功概率为

[6]我们使用 Chernoff 界，其中 $w_j = R_j/(2n/\Delta_1)$，且 $R_j = \sum_{x\in T\cap G_j} x$。现在考虑 S_1。若 $j \in \cup_{t\in S_1} I_t$，我们令 $x_j = 1$；否则令 $x_j = 0$。由于划分问题 $\{I_t\}_{1\le t\le\Delta_0}$ 是随机的，$\sum_{j=1}^{k} w_j x_j$ 的期望值为 $\Theta(\Delta_1 2^i/\Delta_0)$。根据 Chernoff 界，以概率 $\le 2e^{\Theta(-(u/(4n/\Delta_1))^2/(\Delta_1 2^i/\Delta_0))} \le 1/(10n)$，该值与期望值的差距为 $u/(4n/\Delta_1)$。根据并集界，$\sum_{x\in T\cap H_2} x$ 和 $\sum_{x\in T\cap H_1} x$ 与期望值的差距都不超过 $u/2$，其概率为 $\ge 1 - 1/(5n)$；在这种情况下，两者之差不超过 u。

概率为 $\geq 3/4$，由并集界可得。

为分析误差项，注意到共有 $O(\Delta_0/2^i)$ 次将两个子树合并，每个子树含有 2^i 个部分，其中断言3.15 对每次此类合并产生 $O(2^{0.9i}n\varepsilon/\Delta_0)$ 加性误差。因此，总加性误差的上界为 $\sum_{2^i\leq\Delta_0}(2^{0.9i}n\varepsilon/\Delta_0)(\Delta_0/2^i) = O(n\varepsilon)$。

现在我们分析时间复杂度。注意，第 i 层的总复杂度为

$$\begin{aligned}&O((\Delta_0/2^i)\cdot\varepsilon^{-2}\Delta_0^{0.5}\Delta_1^{-0.5}/2^{\Omega(\sqrt{\log(1/\varepsilon)})})\\=\ &O(\varepsilon^{-2}\Delta_0^{1.5}\Delta_1^{-0.5}/2^{\Omega(\sqrt{\log(1/\varepsilon)})})\\=\ &O(n^{\frac{4}{5}}\varepsilon^{-\frac{7}{5}}/2^{\Omega(\sqrt{\log(1/\varepsilon)})}).\end{aligned}$$

由于层数为对数级，分治部分的总复杂度为$O(n^{\frac{4}{5}}\varepsilon^{-\frac{7}{5}}/2^{\Omega(\sqrt{\log(1/\varepsilon)})})$。

因此，总复杂度为 $O(n^{\frac{4}{5}}\varepsilon^{-\frac{7}{5}}/2^{\Omega(\sqrt{\log(1/\varepsilon)})})$，并且根据并集界，成功概率为 $\geq 1/2$（可通过重复来放大）。

一个小细节是，当 n 小到使得 $n^{7/10} < \varepsilon^{-2/5}$、$\Delta_0 < 1$ 时，我们的推理就不再成立。在这种情况下，只需令 $\Delta_0 = 1$，运行时间仍然成立。 □

4 划分问题的近似

在本节中，我们将解决以下问题。

问题 2. 假设 $\varepsilon \in (0, 1/2)$ 且 $1/\varepsilon \in \mathbb{N}^+$。给定区间$[1/\varepsilon, 2/\varepsilon)$内由 n 个互不相同的整数构成的集合 X，计算集合 $A \subset \mathbb{N}$，使其对 $\mathcal{S}(X)$进行 n-加性近似。

通过繁琐的归 约，这一方法大量依赖已有技术，可以证明如下结论。

引理 4.1。 若存在某个 $c \geq 1$，问题 2 可在 $O(n + 1/\varepsilon^c)$ 时间内求解，则 $(1-\varepsilon)$-近似 划分问题 也可在 $O(n + 1/\varepsilon^c)$ 时间内求解。

引理 4.1 将在附录中证明。

现在我们开始介绍用于求解问题 2的主要算法。

在下面的引理中，我们合并 $\mathcal{S}(X_1)$ $\mathcal{S}(X_2)$ 的近似，得到 $\mathcal{S}(X_1 \uplus X_2)$ 的一个近似。当 X_1, X_2 来自短区间 $[\ell, \ell + d]$ 时，我们可以通过二维快速傅里叶变换进行稠密化，相较于直接算法实现加速。

引理 4.2。 设 $\delta \in (0, 1/2)$，且 $\ell, d, t, \Delta \in \mathbb{N}^+$ 满足 $d \leq \ell \leq t$。

设 X_1、$X_2 \subseteq \mathbb{N}^+ \cap [\ell, \ell + d]$ 是两个整数集。给定 A_1、$A_2 \subset \mathbb{N}$ 作为输入，其中对于 $i \in \{1, 2\}$，A_i 是 $\mathcal{S}(X_i)$ 的一个 t 近似，误差不超过 $1 - \delta$)，可以在 $O(Z + |A_1| + |A_2|)$ 时间内计算出一个大小为 $|A| \leq Z$ 的集合 $A \subset \mathbb{N}^+$，它是 $\mathcal{S}(X_1 \uplus X_2)$ 的一个 $1 - \delta$、$\Delta - 1$) 近似，误差不超过 t，其中

$$Z \leq O\left(\min\left\{\left\lceil\frac{t}{\Delta}\right\rceil,\ \frac{t}{\ell}\cdot\left\lceil\frac{td}{\ell\Delta}\right\rceil\right\}\right).$$

证明。 令 $\Delta := \lceil\Delta/2\rceil$。我们将运行以下两种算法之一，以使 Z 最小。

算法 1（一维快速傅里叶变换）。对于 $i \in \{1,2\}$，将 A_i 中的整数向下舍入到 Δ 的倍数，我们得到集合 $A_i' \subset \Delta \cdot \mathbb{N}$，它对 A_i 进行了 $(\Delta-1)$-加性近似。然后，由于 $A_i' \subseteq [0,t]$，它们的和集 $A_1'+A_2'$ 可通过快速傅里叶变换在 $O(\lceil t/\Delta \rceil) \le O(\lceil t/\Delta \rceil)$ 时间内计算得出。注意，$A := A_1'+A_2'$ 对 A_1+A_2 的近似加性误差至多为 $2(\Delta-1) \le \Delta-1$，因此，根据命题 2.4，A 是 $\mathcal{S}(X_1 \uplus X_2)$ 的 $(1-\delta, \Delta-1)$-近似，误差不超过 t。

算法 2（使用二维快速傅里叶变换进行稠密化）。对于每个 $a \in A_i$，都存在 $s \in \mathcal{S}(X_i; t)$，使得 $0 \le s-a \le s\delta$。注意，s 是至多 t/ℓ 个来自 $[\ell, \ell+d]$ 的整数之和，因此 s 可以表示为 $s = k\ell + b'$，其中 $k \in \mathbb{N} \cap [0, t/\ell]$ 和 $0 \le b' \le dt/\ell$ 满足条件。因此，$a \in A_i$ 可以表示为 $a = k\ell + b$，其中 $k \in \mathbb{N} \cap [0, t/\ell]$ 和 $-s\delta \le b \le dt/\ell$ 满足条件。接着，将 b 向下舍入到 Δ 的整数倍，我们得到 $A_i' \subset \mathbb{N}$，它以 $(\Delta-1)$-加法方式近似 A_i，使得每个 $a' \in A_i'$ 都可以表示为

$$a' = k\ell + j\bar{\Delta},$$

对于某个 $k \in \mathbb{N} \cap [0, t/\ell]$ 以及 $j \in \mathbb{Z} \cap [-1 - s\delta/\Delta, dt/(\ell\Delta)]$。利用这种二维$(k,\ j)$ 表示法表示A_i'，我们可以使用二维快速傅里叶变换计算 $A_1'+A_2'$（引理2.1）：第一维的大小为$O(t/\ell)$，第二维的大小至多为

$$dt/(\ell\bar{\Delta}) + s\delta/\bar{\Delta} + O(1) \le O\left(\left\lceil \frac{td}{\ell\Delta} \right\rceil\right),$$

其中不等式由 $s \le t$ 和一个假设得出

$$\delta \le O(d/\ell), \tag{5}$$

稍后将证明这一点。因此，该二维快速傅里叶变换的运行时间为

$$\tilde{O}\left(\frac{t}{\ell} \cdot \left\lceil \frac{td}{\ell\Delta} \right\rceil\right).$$

与算法 1 类似，也可以证明，在这种情况下，$A := A_1' + A_2'$ 是 $\mathcal{S}(X_1 \uplus X_2)$ 的 $(1-\delta, \Delta-1)$-近似，误差不超过 t。

为证明假设 (5)，注意，如果 $\delta \ge d/(\ell+d)$ 成立，或者等价地，$(1-\delta)(\ell+d) \le \ell$，那么可以将 $X_1, X_2 \subset [\ell, \ell+d]$ 中的每个整数向下取整为恰好 ℓ，同时仍保证 $(1-\delta)$ 近似，因此可以立即得到一个大小为 $|A| \le \lceil t/\ell \rceil$ 的 $A \subset \ell \cdot \mathbb{N}$，它对 $\mathcal{S}(X_1 \uplus X_2)$ 的近似误差不超过 t，近似因子为 $1-\delta$。 □

然后，我们结合缩放应用引理 4.2，得到以下引理，其近似完全是乘法型的。

引理 4.3. 令 $\delta, \delta_0 \in (0, 1/2)$，且 $\ell, d, T \in \mathbb{N}^+$ 满足 $d \le \ell \le T$。

设 X_1，$X_2 \subseteq \mathbb{N}^+ \cap [\ell, \ell+d]$ 是两个整数集。给定 A_1，$A_2 \subset \mathbb{N}$ 作为输入，其中对于 $i \in \{1,2\}$，A_i 是 $\mathcal{S}(X_i)$ 在 T 范围内的一个 $(1-\delta)$ 近似，则可以在 $\tilde{O}\big(Z + (|A_1|+|A_2|)\log(2T/\ell)\big)$ 时间内，计算出一个大小为 $|A| \le Z$ 的集合 $A \subset \mathbb{N}^+$，使其在 T 范围内对 $\mathcal{S}(X_1 \uplus X_2)$ 进行 $(1-\delta-\delta_0)$-近似，其中

$$Z \le O\left(\min\left\{\frac{\log(2T/\ell)}{\delta_0},\ \frac{T}{\ell} \cdot \left\lceil \frac{d}{\ell\delta_0} \right\rceil\right\}\right).$$

证明。初始化集合 $A=\{0\}$。我们遍历所有满足 $\ell/6 \le r \le T$ 的 2 的整数次幂 r。对于每个 r，应用引理 4.2，将 A_1 和 A_2 以及 $t := 6r$ 和 $\Delta := \lceil \delta_0 r \rceil$ 作为参数，得到一个集合$A_r \subseteq \mathbb{N} \cap [0, 6r]$，使其在 $6r$ 范围内对 $\mathcal{S}(X_1 \uplus X_2)$ 进行 $(1-\delta, \lceil \delta_0 r \rceil - 1)$-近似。然后，我们将 $A_r \cap [r, 6r]$ 中的所有元素插入 A。我们将证明，最终 A 是 $\mathcal{S}(X_1 \uplus X_2)$ 在 T 范围内的 $(1-\delta_0-\delta)$-近似。

注意 到，对于每个 $a \in A_r \cap [r, 6r]$，都存在 $s \in \mathcal{S}(X_1 \uplus X_2)$，使得 $a \le s$ an d

$$\begin{aligned} a &\ge (1-\delta)s - (\lceil \delta_0 r \rceil - 1) \\ &> (1-\delta)s - \delta_0 r \\ &\ge (1-\delta-\delta_0)s, \end{aligned}$$

其中最后一步由 $s \ge a \ge r$ 得出。

反过来，对于每个正的 $s \in \mathcal{S}(X_1 \uplus X_2; T)$（它必须满足 $\ell \le s \le T$），令 r 为一个满足 $3r \le s \le 6r$ 的 2 的幂。于是存在 $a \in A_r$，使得 $a \le s \le 6r$，并且

$$\begin{aligned} a &\ge (1-\delta)s - (\lceil \delta_0 r \rceil - 1) \\ &\ge (1-\delta)s - \delta_0 r \\ &\ge s/2 - r/2 \\ &\ge r, \end{aligned}$$

因此 $a \in A_r \cap [r, 6r]$，从而会被纳入 A；与之前类似，我们有 $a \ge (1-\delta_0-\delta)s$。因此，我们已经证明，$A$ 是 $\mathcal{S}(X_1 \uplus X_2)$ 的一个 $(1-\delta_0-\delta)$-近似，误差不超过 T。

剩下的工作是界定 A 的总运行时间和规模。共有 $O(\log(2T/L))$ 次 r迭代；对于每个 $r \in [\ell/6, T]$ 满足 $t := 6r$ 且 $\Delta := \lceil \delta_0 r \rceil$ 的情况，引理 4.2 给出了上界

$$\begin{aligned} Z_r &\le O\left(\min\left\{\left\lceil \frac{t}{\Delta} \right\rceil,\ \frac{t}{\ell} \cdot \left\lceil \frac{td}{\ell\Delta} \right\rceil\right\}\right) \\ &\le O\left(\min\left\{\left\lceil \frac{r}{\delta_0 r} \right\rceil,\ \frac{r}{\ell} \cdot \left\lceil \frac{rd}{\ell\delta_0 r} \right\rceil\right\}\right) \\ &\le O\left(\min\left\{\left\lceil \frac{1}{\delta_0} \right\rceil,\ \frac{r}{\ell} \cdot \left\lceil \frac{d}{\ell\delta_0} \right\rceil\right\}\right). \end{aligned}$$

因此，对范围 $[\ell/6, T]$ 内的所有二的幂求和，我们得到

$$Z \le \sum_r Z_r \le O\left(\min\left\{\frac{\log(2T/\ell)}{\delta_0},\ \frac{T}{\ell} \cdot \left\lceil \frac{d}{\ell\delta_0} \right\rceil\right\}\right). \qquad \square$$

引理 4.3 去掉上界 。 T 后，立即可得以下推论。

推论 4.4。 设 $\delta, \delta_0 \in (0, 1/2)$，且 $\ell, d \in \mathbb{N}^+$ 满足 $d \le \ell$。

设 X_1， $X_2 \subseteq \mathbb{N}^+ \cap [\ell, \ell+d]$ 为两个总大小为 $|X_1|+|X_2| = n$ 的整数集。给定 A_1， $A_2 \subset \mathbb{N}$ 作为输入，其中对于 $i \in \{1,2\}$， A_i 是 $\mathcal{S}(X_i)$ 的 $(1-\delta)$ 近似，可以在 $\tilde{O}\big(Z + (|A_1|+|A_2|)\log n\big)$ 时间内计算出一个大小为 $|A| \le Z$ 的集合 $A \subset \mathbb{N}^+$，使其对 $\mathcal{S}(X_1 \uplus X_2)$ 实现 $(1-\delta_0-\delta)$ 近似，其中

$$Z \le O\left(\min\left\{\frac{1}{\delta_0},\ \frac{nd}{\ell\delta_0} + n\right\} \cdot \log n\right).$$

证明。 由引理 4.3 中令 $T = n \cdot (\ell + d)$即可立即得出，它是 $\mathcal{S}(X_1 \uplus X_2)$ 中最大元素的上界。 □

现在，我们以分治的方式应用推论 4.4 ，近似计算 $X \subseteq \mathbb{N}^+ \cap [\ell, 2\ell]$ 的子集和。

引理 4.5. 设 $\delta \in (0, 1/2)$ 和 $\ell \in \mathbb{N}^+$。

给定一个含有 n 个整数的整数集 $X \subseteq \mathbb{N}^+ \cap [\ell, 2\ell]$ ，可以在 $O(n + \sqrt{n}/\delta)$ 时间内计算出一个集合 $A \subset \mathbb{N}^+$ ，使其以 $\tilde{}(1-\delta)$ 近似 $\mathcal{S}(X)$。

证明。Let $X = \{x_1, x_2, \ldots, x_n\}$，其中 $\ell \le x_1 < x_2 < \cdots < x_n \le 2\ell$。设 $\delta_0 := \delta / \lceil \log_2 n \rceil$.

我们将使用分治方法，利用推论 X 合并 4.4 中的物品。构建一棵平衡二叉树，其 n 个叶节点从左到右分别表示物品 x_1、.....、 x_n。在每个表示 $x_{[l..r]}$ 的内部节点处，我们使用推论 4.4 合并两个子节点的结果（它们分别表示 $x_{[l..m]}$ 和 $x_{[m+1..r]}$，其中 $m = \lfloor (l+r)/2 \rfloor$ ，并得到 $\mathcal{S}(\{x_l, x_{l+1}, \ldots, x_r\})$ 的近似值。最后，我们在根节点得到 $\mathcal{S}(X)$ 的近似值。

这棵二叉树有 $\lceil \log_2 n \rceil$ 个层级，其中每一层应用推论 4.4 都会使近似因子变差 δ_0。因此， $\mathcal{S}(X)$ 的总体近似因子为 $1 - \delta_0 \cdot \lceil \log_2 n \rceil = 1 - \delta$，符合要求。

还需对推论 4.4的所有调用的总运行时间进行界定。注意，在每次调用中，所述时间复杂度中的 $(|A_1| + |A_2|) \log n$ 项总是可以由两个子节点的输出规模吸收（至多相差 $\log n$ 因子），而这两个子节点的输出规模又分别受其对应子调用的运行时间所界定。因此，只需对推论 Z 4.4 中给出的量求和并界定其上界。

我们分别对二叉树的每个层级进行界定。在第 i 层 $(0 \le i < \lceil \log_2 n \rceil)$，推论 $m_i = 2^i$ 4.4 最多被调用 次，每次调用涉及 $n_i = \lceil n/2^i \rceil$ 个物品，属于 X。注意， $n_i m_i \le 2n$。设这些 m_i 次调用分别涉及 $x_{[1..k_1]}$、 $x_{[k_1+1..k_2]}$、.....、 $x_{[k_{m_i-1}+1..n]}$ 。那么，第 j 次调用的 d价值（推论 4.4中所述）至多为$d_j \le x_{k_j} - x_{k_{j-1}}$。因此，这些 d 价值之和至多为

$$\sum_{j=1}^{m_i} d_j \le \sum_{j=1}^{m_i} (x_{k_j} - x_{k_{j-1}}) \le x_n - x_1 \le \ell. \tag{6}$$

现在我们可以对层级 i（$0 \le i < \lceil \log_2 n \rceil$）上的 m_i 次调用中的 Z quantity 之和进行界定。我们分两种情况讨论。

- **情况 1**： $n_i \le \sqrt{n}$。

  于是，根据推论 4.4,

$$\begin{aligned}
\sum_{j=1}^{m_i} Z_j &\le \sum_{j=1}^{m_i} \left(\frac{n_i d_j}{\ell \delta_0} + n_i \right) \cdot \log n \\
&= \left(\frac{n_i \sum_{j=1}^{m_i} d_j}{\ell \delta_0} + m_i n_i \right) \cdot \log n \\
&\le \left(\frac{n_i}{\delta_0} + m_i n_i \right) \cdot \log n \qquad \text{(by (6))} \\
&\le \tilde{O}\left(\frac{\sqrt{n}}{\delta_0} + n \right).
\end{aligned}$$

- **情况 2：** $n_i > \sqrt{n}$。

  那么，$m_i \le 2n/n_i < 2\sqrt{n}$。根据推论 4.4，

$$\sum_{j=1}^{m_i} Z_j \le m_i \cdot \frac{1}{\delta_0} \cdot \log n$$
$$\le \tilde{O}(\sqrt{n}/\delta_0).$$

因此，在两种情况下，我们都有 $\sum_{j=1}^{m_i} Z_j \le \tilde{O}(n+\sqrt{n}/\delta)$。所以，所有层级上的总运行时间~$0 \le i < \lceil \log_2 n \rceil$ 也为 $O(n+\sqrt{n}/\delta)$。 □

最后，我们结合引理 4.5 与GM91的加法组合学结果[GM91、BW21]，来解决问题 2。

引理 4.6。 我们可以在 2 时间内解决问题 $\tilde{O}\left(n + \text{最小}\{\varepsilon^{-1}n^{1/2},\ \varepsilon^{-1}+\varepsilon^{-2}/n^{3/2}\}\right)$，其至多为 $\tilde{O}(n+1/\varepsilon^{5/4})$。

证明。回顾问题 2中，对于 $\varepsilon > 0$ 其中 $\ell = 1/\varepsilon$ 是一个整数，我们给定一个集合 $X \subseteq \mathbb{N}^+ \cap [\ell, 2\ell)$ 包含 n 个不同整数，并需要计算一个集合 $A \subset \mathbb{N}$，使其 n-加法近似$\mathcal{S}(X)$。

根据参数，我们将选择运行以下两种算法之一。

~**算法 1.** 直接应用引理 4.5，耗时 $\delta := \varepsilon$，耗时 $O(n+\sqrt{n}/\varepsilon)$。当 $n \le \tilde{O}(1/\varepsilon^{1/2})$ 时，算法 1 的运行时间为 $\tilde{O}(n+1/\varepsilon^{5/4})$。

算法 2. 令 $\sigma = \Sigma(X)$，并令 λ 为定理 2.7 中的阈值$\lambda = \tilde{\Theta}(\ell^2/n)$。当 $\lambda \le \sigma/2$时，适用以下算法；特别地，当$1/\varepsilon \ll n^2$。

~初始化 $A = \emptyset$。我们令 $\delta := n/(n+\lambda)$，并在 $O(n+\sqrt{n}/\delta)$ 时间内应用引理 4.5，计算出一个集合 A_δ，使其对 $\mathcal{S}(X)$ 达到 $(1-\delta)$-近似。注意，$A_\delta \cap [0,\lambda]$ 是 $\mathcal{S}(X)$ 的 n-加性近似，误差至多为 λ。因此，我们将 $A_\delta \cap [0,\lambda]$ 中的所有元素插入 A。

然后，利用引理 2.7 中的数据结构，我们计算 n-加性近似$\mathcal{S}(X) \cap [\lambda, \sigma/2]$，并将它们插入 A。为此，我们从区间 $[\lambda, \sigma/2]$ 的左端点 λ 开始，每次使用二分查找（可通过引理 2.7 支持的范围查询来实现）找到该区间内下一个子集和，然后向右跳过 n 阶梯，因为我们允许的加性误差为 n。时间复杂度为 $O(\lceil \frac{\sigma/2-\lambda}{n} \rceil \cdot \log \sigma) \le \tilde{O}(\ell)$。

现在我们已经构造出 A，它对 $\mathcal{S}(X)$ 进行 n-加性近似，误差不超过 $\sigma/2$。利用简单事实：$t \in \Sigma(X)$ 当且仅当 $\sigma - t \in \Sigma(X)$，我们可以对称地使用 A，来获得另一半的近似。具体而言，令 $A' := \{\sigma - a - n : a \in A\}$，不难验证 $A \cup A'$ 是 $\mathcal{S}(X)$ 的 n- 加性近似（误差不超过 σ）。因此，我们返回 $A \cup A'$。

算法 2 的总体时间复杂度为

$$\tilde{O}(n+\ell+\sqrt{n}/\delta) \le \tilde{O}\left(n + 1/\varepsilon + \frac{\sqrt{n}(n+\lambda)}{n}\right)$$
$$\le \tilde{O}\left(n + 1/\varepsilon + \frac{1/\varepsilon^2}{n^{3/2}}\right).$$

当 $n \gg 1/\varepsilon^{1/2}$ 时，运行时间为 $\tilde{O}(n+1/\varepsilon^{5/4})$。 □

结合引理 4.1中的规约，这就证明了我们的主定理 1.2 .

参考文献

[ABHS19] Amir Abboud、Karl Bringmann、Danny Hermelin 和 Dvir Shabtay。子集和与双准则路径的 SETH 下界。载于第 30 届 ACM-SIAM 离散算法研讨会（SODA）论文集，第 41–57 页，2019 年。doi:10.1137/1.9781611975482.3. 1[AKM+87] Alok Aggarwal、Maria M. Klawe、Shlomo Moran、Peter Shor 和 Robert Wilber。矩阵搜索算法的几何应用。Algorithmica，2(1)：195–208，1987 年 11 月。doi:10.1007/BF01840359. 11[BC22] Karl Bringmann 和 Alejandro Cassis。通过有界单调 min-plus 卷积加速背包算法。载于第 49 届国际自动机、语言与编程研讨会（ICALP 2022），2022 年 7 月 4–8 日，法国巴黎，LIPIcs 第 229 卷，第 31:1–31:21 页。Schloss Dagstuhl - Leibniz-Zentrum für Informatik，2022 年。doi:10.4230/LIPIcs.ICALP.2022.31. 1[BCD+14] David Bremner、Timothy M. Chan、Erik D. Demaine、Jeff Erickson、Ferran Hurtado、John Iacono、Stefan Langerman、Mihai P˘atraşcu 和 Perouz Taslakian。项链、卷积与 x+y。Algorithmica，69(2)：294–314，2014 年 6 月。doi:10.1007/s00453-012-9734-3. 7[Bla10] Richard E Blahut。信号处理快速算法。剑桥大学出版社，2010 年。4[BN21] Karl Bringmann 和 Vasileios Nakos。子集和与划分问题近似的细粒度视角。载于 2021 年 ACM-SIAM 离散算法研讨会论文集（SODA 2021），线上会议，2021 年 1 月 10–13 日，第 1797–1815 页。SIAM，2021 年。doi:10.1137/1.9781611976465.108. 1, 3, 5[BW21] Karl Bringmann 和 Philip Wellnitz。关于稠密子集和的近线性时间算法。载于 2021 年 ACM-SIAM 离散算法研讨会论文集（SODA 2021），线上会议，2021 年 1 月 10–13 日，第 1777–1796 页。SIAM，2021 年。doi:10.1137/1.9781611976465.107. 2, 6, 18, 22[Cha18] Timothy M. Chan。0-1 背包问题的近似方案。载于第 1 届算法简洁性研讨会（SOSA）论文集，第 5:1–5:12 页，2018 年。doi:10.4230/OASIcs.SOSA.2018.5. 1, 2, 3, 5, 6, 7, 11[CMWW19] Marek Cygan、Marcin Mucha、Karol Węgrzycki 和 Michał Włodarczyk。关于等价于 (min,+)-卷积的问题。《ACM算法汇刊》，15(1)：14:1–14:25，2019 年 1 月。doi:10.1145/3293465. 1[CW16] Timothy M. Chan 和 Ryan Williams。确定性全点对最短路径、正交向量及更多：快速实现 Razborov-Smolensky 去随机化。载于第 27 届 ACM-SIAM 离散算法年度研讨会（SODA）论文集，第 1246–1255 页，2016 年。doi:10.1137/1.9781611974331.ch87. 7

[GL79] George Gens 和 Eugene Levner。组合问题近似算法的计算复杂性。载于《计算机科学数学基础 1979》论文集，第 8 届研讨会，捷克斯洛伐克奥洛穆茨，1979 年 9 月 3–7 日，计算机科学讲义第 74 卷，第 292–300 页。Springer，1979 年。doi:10.1007/3-540-09526-8_26. 1[GM91] Zvi Galil 和 Oded Margalit。稠密子集和问题的近乎线性时间算法。SIAM J. Comput.，20(6)：1157–1189，1991 年。doi:10.1137/0220072. 2, 3, 6, 18, 22[IK75] Oscar H. Ibarra 和 Chul E. Kim。背包问题和子集和问题的快速近似算法。《ACM学报》（JACM），22(4)：463–468，1975 年 10 月。doi:10.1145/321906.321909. 1[Jin19] Ce Jin。改进的 0-1 背包问题全多项式时间近似方案。载于第 46 届国际自动机、语言与编程研讨会（ICALP 2019），2019 年 7 月 9–12 日，希腊帕特雷，LIPIcs 第 132 卷，第 76:1–76:14 页。Schloss Dagstuhl - Leibniz-Zentrum für Informatik，2019 年。doi:10.4230/LIPIcs.ICALP.2019.76. 1, 2, 3, 6, 7, 11[JK18] Klaus Jansen 和 Stefan E.J. Kraft。无界背包问题的更快全多项式时间近似方案。《欧洲组合学杂志》，68：148–174，2018 年。doi:10.1016/j.ejc.2017.07.016. 1[Kar72] Richard M Karp。组合问题之间的可约性。载于《计算机计算的复杂性》，第 85–103 页。Springer，1972 年。1[KMPS03] Hans Kellerer、Renata Mansini、Ulrich Pferschy 和 Maria Grazia Speranza。子集和问题的高效完全多项式近似方案。《J. Comput. Syst. Sci.》，66(2)：349–370，2003 年。doi:10.1016/S0022-0000(03)00006-0. 1[KP99] Hans Kellerer 和 Ulrich Pferschy。背包问题的一种新型完全多项式时间近似方案。《组合优化杂志》，3(1)：59–71，1999 年 7 月。doi:10.1023/A:1009813105532. 1[KP04] Hans Kellerer 和 Ulrich Pferschy。结合背包问题全多项式时间近似方案的改进动态规划。《组合优化杂志》，8(1)：5–11，2004 年 3 月。doi:10.1023/B:JOCO.0000021934.29833.6b. 1[KPS17] Marvin Künnemann、Ramamohan Paturi 和 Stefan Schneider。关于一维动态规划的细粒度复杂性。载于第 44 届国际自动机、语言与编程研讨会（ICALP）论文集，第 21:1–21:15 页，2017 年。doi:10.4230/LIPIcs.ICALP.2017.21. 1[KX19] Konstantinos Koiliaris 和 Chao Xu。子集和的更快速伪多项式时间算法。《ACM算法汇刊》，15(3)：40:1–40:20，2019 年 6 月。doi:10.1145/3329863. 3, 24[Law79] Eugene L. Lawler。背包问题的快速近似算法。《运筹学数学》，4(4)：339–356，1979 年。doi:10.1287/moor.4.4.339. 1, 4

[Lev03] Vsevolod F Lev。《子集和集合中的块与等差数列》。《算术学报》-华沙-，106(2):123–142，2003。6，22[MWW19] Marcin Mucha、Karol Węgrzycki 和 Michał Włodarczyk。《划分问题的次二次近似方案》。载于第30届 ACM-SIAM 离散算法研讨会（SODA）论文集，第70–88页，2019年。完整版见 https://arxiv.org/abs/1804.02269v2。doi:10.1137/1.9781611975482.5。1，3，5，23，24[Rhe15] Donguk Rhee。《背包问题的更快速全多项式近似方案》。硕士论文，麻省理工学院，2015年。网址：http://hdl.handle.net/1721.1/98564。1[Sár94] A. Sárközy。《精细加法定理 II》。《数论杂志》，48(2):197–218，1994年。2，6，22[Wil14] Ryan Williams。《利用电路复杂性实现更快速的所有点对最短路径》。载于第46届 ACM计算理论研讨会（STOC）论文集，第664–673页，2014年。doi:10.1145/2591796.2591811。7，13

背包问题的一个已知归约

回想一下，我们定义了以下较简单的问题。

问题1. 假设 $\varepsilon \in (0, 1/2)$ 且 $1/\varepsilon \in \mathbb{N}^+$。给定一个包含 I 个物品的列表$(p_1, w_1) \ldots, (p_n, w_n)$，其重量 $w_i \in \mathbb{N}$ 和利润 p_i 均为区间 $[1, 2)$内 ε 的倍数，计算一个利润函数，使其在 $2/\varepsilon$ 以内对 f_I 进行$(1-\varepsilon)$近似。

引理3.1. 如果对于某个 $c \geq 2$，问题 *1* 可以在 $O(n + 1/\varepsilon^c)$ 时间内求解，那么 $(1-\varepsilon)$-近似背包问题也可以在 $O(n + 1/\varepsilon^c)$ 时间内求解。

证明. 首先，我们可以对 ε 进行归约，使 $1/\varepsilon$ 成为一个整数。

我们将利润值限制在较小的区间内，具体做法如下：将物品划分为 $O(\log \max_j p_j \min_j p_j) = O(\log \varepsilon^{-1})$ 组（见第 2.3节），每组包含满足 $p_i \in [2^j, 2^{j+1}]$ 的物品，其中某个 j（可重新缩放为 $[1, 2]$）。最后，使用合并引理引理 3.4 合并所有组的利润函数，总体耗时为 $\tilde{O}(n + \varepsilon^{-2})$。

现在，既然已将利润值限制在 $[1, 2)$ 内，我们可以将每个利润值舍入到 ε 的倍数，这样总共只会带来 $(1 - O(\varepsilon))$ 的近似因子。

最后，以下贪心引理处理总利润大于 $\Omega(\varepsilon^{-1})$ 的情况。

引理 A.1. 假设对于所有 $i \in I$，$p_i \in [1, 2]$。对于 $B = \Omega(\varepsilon^{-1})$，利润函数 f_I 可在 $O(n \log n)$ 时间内，以加性误差 $O(\varepsilon B)$ 进行近似。

证明。只需按效率 p_i/w_i 非增序对物品排序，并定义贪心算法得到的利润函数~f，其函数值为 0、p_1、$p_1 + p_2$、$\ldots$、$p_1 + \cdots + p_n$，而 x 个断点为 0、w_1、$w_1 + w_2$、$\ldots$、$w_1 + \cdots + w_n$。显然，对于 $B = \Omega(\varepsilon^{-1})$，它以 $\max_i p_i \leq 2 \leq O(\varepsilon B)$ 为加性误差近似 f_I。 □

对于较大的利润值，这种贪心方法可实现 $(1 - O(\varepsilon))$ 近似。因此，只需将 f_I 近似到 $2/\varepsilon$。 □

B 引理2.6的证明

我们需要用到一系列研究成果中关于稠密子集和的若干结果，这些研究包括 [Sár94、Lev03、GM91、BW21]。以下定义和定理摘自 [BW21]。此处考虑的集合包含互不相同的正整数。

定义 B.1 (稠密度)。 集合 $X \subset \mathbb{N}^+$ 是 δ-稠密的，若其满足 $|X|^2 \geq \delta \cdot \max X$。

定义 B.2 (近似因子)。 令 $X(d) := X \cap d\mathbb{Z}$ 表示 X 中所有能被 d 整除的数构成的集合。令 $X(d) := X \setminus X(d)$ 表示 X 中所有不能被 d 整除的数构成的集合。若整数 $d > 1$ 满足 $|X(d)| \leq \alpha \cdot \Sigma(X)/|X|^2$，则称其为 X 的 α-近似因子。

定理 B.3 ([BW21，定理 4。1])。 令 $\delta \geq 1$ 和 $\alpha \leq \delta/16$。给定一个 δ-稠密集合 X，其大小为n，则存在一个正整数 d 使得 $X' := X(d)/d$ 是 δ-稠密的，且没有 α-近似因子，并满足以下附加性质：

1. $d \leq 4\Sigma(X)/|X|^2$,
2. $|X'| \geq 0.75|X|$, 3. $\Sigma(X') \geq 0.75\,\Sigma(X)/d$。

定理 B.4 （[BW21，定理 4。2]）。 令 X 为一个多重集，并令

$$C_\delta := 1699200 \cdot \log(2n)\log^2(2),$$
$$C_\alpha := 42480 \cdot \log(2),$$
$$C_\lambda := 169920 \cdot \log(2).$$

如果 X 是 C_δ-稠密的，且没有 C_α-近似因子，那么对于 $\lambda_X := C_\lambda \cdot (\text{最大值 } X) \cdot \Sigma(X)/|X|^2$，我们有

$$\big([\lambda_X, \Sigma(X) - \lambda_X] \cap \mathbb{Z}\big) \subseteq \mathcal{S}(X).$$

现在我们准备好证明引理 2.6。

引理 2.6。 给定 n 个互不相同的正整数 $X = \{x_1, \ldots, x_n\} \subseteq [\ell, 2\ell]$，其中 $\ell = o(n^2/\log n)$。

那么，对于某个普适常数 $c \geq 1$，对于每个 $c\ell^2/n \leq t \leq \Sigma(X)/2$，都存在 $t' \in \mathcal{S}(X)$ 使得 $0 \leq t' - t \leq 8\ell/n$。

证明。 令 $C_\delta = \Theta(\log n)$ $C_\alpha = \Theta(1)$ $C_\lambda = \Theta(1)$ 如定理 B.4 中所定义。于是，X 是 C_δ- 稠密的，因为 $C_\delta \cdot \ell = o(n^2)$。令 d 为定理 B.3 所保证的正整数，使得$X' := X(d)/d$ 是 C_δ-稠密的，且没有 C_α-近似因子。

于是，根据定理 B.4，对于 $\lambda_{X'} := C_\lambda \cdot (\max X') \cdot \Sigma(X')/|X'|^2$ 我们有

$$\big([\lambda_{X'}, \Sigma(X') - \lambda_{X'}] \cap \mathbb{Z}\big) \subseteq \mathcal{S}(X').$$

由于 X' 是 C_δ-稠密的，因此我们有 $\lambda_{X'}/\Sigma(X') = C_\lambda \cdot (\max X')/|X'|^2 \leq C_\lambda/C_\delta < 0.1$。

现在，令 $\lambda := d \cdot \lambda_{X'}$。根据定理 B.3 中的性质 3，我们知道

$$\frac{\Sigma(X)/2}{d} \le \frac{2}{3}\Sigma(X') < \Sigma(X') - \lambda_{X'}.$$

因此，对于任意 $\lambda \le t \le \Sigma(X)/2$，我们有

$$\lambda_{X'} \le \lceil t/d \rceil \le \Sigma(X') - \lambda_{X'}.$$

因此 $\lceil t/d \rceil \in \mathcal{S}(X')$,这意味着 $t' := d \cdot \lceil t/d \rceil \in \mathcal{S}(X)$。由定理 B.3中的性质 1，我们有

$$0 \le t' - t < d \le 4 \cdot |X| \cdot (2\ell)/|X|^2 = 8\ell/n.$$

最后，我们将 λ 上界估计为

$$\begin{aligned}
\lambda &= d \cdot C_\lambda \cdot (\max X') \cdot \Sigma(X')/|X'|^2 \\
&\le d \cdot C_\lambda \cdot (\max X')^2/|X'| \\
&\le d \cdot C_\lambda \cdot (2\ell/d)^2/|X'| \\
&\le d \cdot C_\lambda \cdot (2\ell/d)^2/(0.75n) && \text{(by Property 2)} \\
&= (16C_\lambda/3) \cdot \frac{\ell^2}{dn} \\
&\le O(\ell^2/n).
\end{aligned}$$

□

C 划分问题的已知归约

回顾一下，我们定义了以下更简单的问题。

问题 2. 假设 $\varepsilon \in (0, 1/2)$ 且 $1/\varepsilon \in \mathbb{N}^+$。给定区间 $[1/\varepsilon, 2/\varepsilon)$内由 n 个互不相同的整数构成的集合 X，计算集合 $A \subset \mathbb{N}$，使其以 n 加法近似 $\mathcal{S}(X)$。

接下来，我们将把划分问题归约到这个问题。

根据一个简单的贪心论证，我们可以假设最优值 $\ge t/2$。

引理 C.1 (例如，[MWW19，引理 4.3])。不失一般性，可以假设对于任意子集和˜实例，最优值 $\ge t/2$。否则，该实例可以在 $O(n)$ 时间内精确求解。

接下来，我们有如下关于$(1-\varepsilon)$-近似划分问题的重要引理。该引理背后的关键洞见最早由[MWW19], 提出，指出近似划分问题比近似一般子集和实例容易得多。

引理 C.2 (参见 [MWW19])。设 $X \subset \mathbb{N}^+$ 为一个元素和为 $\sigma = \Sigma(X)$的多重集，并设 $\varepsilon \in (0, 1/2)$。给定一个集合 $A \subset \mathbb{N}$，它以 $\varepsilon\sigma/4$加法近似 $\mathcal{S}(X)$，即可立即求解$(1-\varepsilon)$-近似划分问题，输入为 X。

证明。回想一下，$t = \sigma/2$，且最优值 $= \max\{\Sigma(Y) : \Sigma(Y) \le t, Y \subseteq X\}$。

给定 A，令 $a := \max\{a \in A : a \le t\}$。我们断言

$$(1-\varepsilon)\mathrm{OPT} \le \min\{a, t(1-\varepsilon/2)\} \le \mathrm{OPT},$$

这使我们能够在 X 上求解 $(1-\varepsilon)$-近似划分问题。

我们分别考虑两种情况来证明这一结论。

- 情况 1： $a \le t(1-\varepsilon/2)$。

  根据 A 的定义，存在 $s \in \mathcal{S}(X)$ 使得 $s - \varepsilon\sigma/4 \le a \le s$。我们有 $s \le a + \varepsilon\sigma/4 \le t(1-\varepsilon/2) + \varepsilon\sigma/4 = t$，因此 $s \in \mathcal{S}(X; t)$，从而最优值 $\ge s \ge a$。根据引理 C.1，我们可以假设 $t/2 \le$ 最优值 $\le t$。接着，根据 A 的定义，存在 $a' \in A$ 使得 $a' \le$ 最优值 $\le t$ 且 $a' \ge$ 最优值 $- \varepsilon\sigma/4 \ge$ 最优值 $- \varepsilon$。因此，根据 a 的定义，我们有 $a \ge a' \ge (1-\varepsilon)$最优值。

  因此，我们已经证明

$$(1-\varepsilon)\mathrm{OPT} \le a = \min\{a, t(1-\varepsilon/2)\} \le \mathrm{OPT}.$$

- 情况 2： $a > t(1-\varepsilon/2)$。

  根据 A 的定义，存在 $s \in \mathcal{S}(X)$ 使得 $s - \varepsilon\sigma/4 \le a \le s$。我们有 $s \le a + \varepsilon\sigma/4 \le t + \varepsilon\sigma/4 = t(1+\varepsilon/2)$，且 $s \ge a > t(1-\varepsilon/2)$。

  通过取补，我们也知道 $\sigma - s \in \mathcal{S}(X)$。利用关键事实 $t = \sigma/2$，可知 $\min\{s, \sigma - s\} \in \mathcal{S}(X; t)$，因此 $\mathrm{OPT} \ge \min\{s, \sigma - s\}$。接着，由于 $s \in (t(1-\varepsilon/2)\ t(1+\varepsilon/2)]$，，我们有 $\min\{s, \sigma - s\} \ge t(1-\varepsilon/2)$。

  因此，我们已经证明

$$(1-\varepsilon)\mathrm{OPT} \le t(1-\varepsilon/2) = \min\{a, t(1-\varepsilon/2)\} \le \mathrm{OPT}.$$ □

利用引理 C.2，我们可以通过求得 $\mathcal{S}(X)$的加法近似，来解决$(1-\varepsilon)$-近似划分问题。

我们将进一步简化输入实例 X。首先，我们需要以下引理，它通过将重复元素分组成二的幂，减少输入中的重复元素。该引理的证明见 [MWW19]，，其依据是此前一个类似结论的证明[KX19，引理 2。4]。

引理 C.3 ([MWW19，引理 4。1])。给定一个由 S 个 n 整数构成的多重集 $[t]$，可以在 T 的 $O(n$对数 $n)$ 时间内计算出一个多重集，使得：

- $\mathcal{S}(S; t) = \mathcal{S}(T; t)$.
- $|T| \le |S|$。
- T 中没有任何元素的重数超过二。

对于每个 $y \in T$，都存在相应的 $x \in S$，使得对于某个 $k \in \mathbb{N}$，有 $y = 2^k \cdot x$。

现在我们证明主要引理。

引理 4.1. 若对于某个 $c \ge 1$，问题 *2* 可在 $O(n + 1/\varepsilon^c)$ 时间内求解，则 $(1-\varepsilon)$-近似划分问题也可在 $O(n + 1/\varepsilon^c)$ 时间内求解。

证明。设 $X \subset \mathbb{N}^+$ 为（$1-\varepsilon$）-划分问题的输入多重集。我们不失一般性地假设 $1/\varepsilon$ 是一个整数。

回顾一下，$t = \sigma/2 = \Sigma(X)/2$。我们定义多重集 $Y \subset \mathbb{N}^+$ 如下：对于每个 $x \in X$，向下取整到最接近的 $\lceil \frac{\sigma}{100n/\varepsilon} \rceil$的整数倍，记为 y；如果 y 非零，则将 y 插入 Y。

由于产生的加性损失总计至多为 $n \cdot (\lceil \frac{\sigma}{100n/\varepsilon} \rceil - 1) \leq \frac{\varepsilon\sigma}{100}$，我们知道 $\mathcal{S}(Y)$ 是 $\mathcal{S}(X)$ 的$\varepsilon\sigma/100$-加性近似。

现在，我们可以不失一般性地假设对所有 $y \in Y$， $y \in [1/\varepsilon, 100n/\varepsilon^2] \cap \mathbb{N}^+$；否则，只需对 X、Y 中的所有元素（以及 σ、t）进行缩放即可。

然后，按如下方式定义另一个多重集 $Z \subset \mathbb{N}^+$：对于每个 $y \in Y$，将 y 向下舍入到某个 $k \in \mathbb{N} \cap [0, \log_2(100n/\varepsilon) + 1]$ and $z_0 \in \mathbb{N}^+ \cap [100/\varepsilon, 200/\varepsilon)$，并将 $2^k \cdot z_0$ 插入 Z。注意，每个 $y \in Y$ 在舍入后产生的乘性误差至多为 $\varepsilon/100$。因此， $\mathcal{S}(Z)$ 是 $\mathcal{S}(Y)$ 的一个 $(1 - \varepsilon/100)$ 近似。特别地， $\mathcal{S}(Z)$ 对 $\mathcal{S}(Y)$ 的近似加性误差至多为 $(\varepsilon/100) \cdot \Sigma(Y) \leq \varepsilon\sigma/100$。结合前面的讨论，这意味着 $\mathcal{S}(Z)$ 是 $\mathcal{S}(X)$ 的一个 $\varepsilon\sigma/50$-加性近似。

然后，我们使用引理 C.3处理 Z，并得到另一个集合 $Z' \subset \mathbb{N}^+$，使得 $\mathcal{S}(Z') = \mathcal{S}(Z)$，且 Z' 中任何元素的重数至多为 2。此外，根据引理 C.3 的第四个性质，我们仍有：每个 $z \in Z'$ 都可以表示为 $z = 2^k \cdot z_0$，其中 $k \leq O$是一个非负整数，满足$(\log(n/\varepsilon))$，且 $z_0 \in \mathbb{N}^+ \cap [100/\varepsilon, 200/\varepsilon$是整数。现在，我们可以将 Z' 划分为$O(\log(n/\varepsilon))$组，使得每组都包含来自 $2^k \cdot (\mathbb{N}^+ \cap [100/\varepsilon, 200/\varepsilon))$的不同整数，其中 $k \leq O$是某个非负整数，满足$(\log(n/\varepsilon))$。

选取一个较小的 $\varepsilon' = \Theta\left(\frac{\varepsilon}{\log(n/\varepsilon)}\right)$（假设 $\varepsilon/\varepsilon' \in \mathbb{N}^+$）。对于上文提到的每个组 Z'_j，我们计算一个集合 $A_j \subseteq \mathbb{N}$，以近似 $\mathcal{S}(Z'_j)$，其加性误差为 $\varepsilon'\Sigma(Z'_j)/100$。具体做法如下：回顾一下，$Z'_j$ 包含来自 $2^k \cdot (\mathbb{N}^+ \cap [100/\varepsilon, 200/\varepsilon))$的不同整数；我们将$Z'_j$ 中的整数缩放至$(\varepsilon/\varepsilon') \cdot (\mathbb{N}^+ \cap [100/\varepsilon, 200/\varepsilon])$，然后调用问题 2的算法，该算法以加性误差 $|Z'_j| \leq \frac{\Sigma}{\varepsilon/\varepsilon'(Z'_j)\,()\cdot 100/\varepsilon} = \varepsilon'\Sigma(Z'_j)/100$ 近似 $\mathcal{S}(Z'_j)$，正如所需。这些调用的总~运行时间为（忽略 poly $\log(n/\varepsilon)$ 因子） $\sum_j (|Z_j| + (100/\varepsilon')^c) \leq O(n + 1/\varepsilon^c)$。

现在，利用已计算出的 $A_j \subseteq \mathbb{N}$，它以 $\varepsilon'\Sigma(Z'_j)/100$ 加性误差近似 $\mathcal{S}(Z'_j)$，我们将计算 $\mathcal{S}(Z')$ 的一个近似（回想一下， $Z' = \bigcup_j Z'_j$ 是一个划分问题）。为此，我们首先将每个 A_j 中的每个元素向下取整到 $\lceil \varepsilon'\sigma/100 \rceil$ 的整数倍；取整后的 A'_j 仍以至多 $\varepsilon'\Sigma(Z'_j)/100 + \varepsilon'\sigma/100 \leq \varepsilon'\sigma/50$ 的加性误差近似 $\mathcal{S}(Z'_j)$。最后，我们使用快速傅里叶变换计算所有这些 A'_j 的和集（它们共有 $O(\log(n/\varepsilon))$ 个），并将其作为 $\mathcal{S}(Z')$ 的近似。这里累积的加性误差至多为 $O(\log(n/\varepsilon)) \cdot \varepsilon'\sigma/50 \leq \varepsilon\sigma/50$，这些快速傅里叶变换的运行时间为 $O(\log(n/\varepsilon)) \cdot \tilde{O}\left(\frac{\sigma}{\lceil \varepsilon'\sigma/100 \rceil}\right) \leq \tilde{O}(1/\varepsilon)$。

我们已获得 $\varepsilon\sigma/50$-加性近似的 $\mathcal{S}(Z')$。此前我们已证明 $\mathcal{S}(Z') = \mathcal{S}(Z)$，且 $\mathcal{S}(Z)$ 是 $\varepsilon\sigma/50$-加性近似的 $\mathcal{S}(X)$，因此我们已获得 $\varepsilon\sigma/50$-加性近似的 $\mathcal{S}(X)$。根据引理 C.2，这足以在 X 上求解 $(1-\varepsilon)$-近似划分问题。

~该归约的总运行时间为 $O(n + 1/\varepsilon^c)$。 □