Потпуна полиномијална апроксимациона шема за проблем збир подскупа

Разматра се проблем збир подскупа, односно:

Проблем. Задат је скуп природних бројева $S = \{x_1, x_2, \ldots, x_n\}$ и природни број t. Одредити подскуп скупа S са највећим sбиром мањим или једнаким од t.

Познато је да је варијанта овог проблема када се пита да ли постоји подкупа скупа S са sбиром једнаким t — NP комплетан проблем.

Оптимизациона верзија овог проблема појављује се у практичним применама. На пример, у камион носивости t тона треба утоварити подскуп предмета са највећом могућом тежином, а да се при томе не прекорачи носивост камиона.

У наставку ћемо најпре да размотримо тачан алгоритам експоненцијалне сложености за решавање овог проблема, а онда ћемо видети како се тај алгоритам може изменити тако да постане приближни алгоритам са унапред задатом границом $1 + \epsilon$ односа добијеног и тачног решења. Сложеност тог алгоритма је полином по величини проблема и по $1/\epsilon$, па је тај алгоритам пример *потпуне полиномијалне апроксимационе шеме.*

1 Тачан алгоритам експоненцијалне сложености

Оптимизациона варијанта проблема може се решити једноставним алгоритмом експоненцијалне сложености: довољно је за сваки подскуп $S' \subseteq S$ израчунати збир, и од тих збирова издвојити највећи који је мањи или једнак од t.

Подскупови скупа S могу се проћи итеративно тако што се од збирова бројева $\{x_1, x_2, \ldots, x_{i-1}\}$ формирају збирови бројева $\{x_1, x_2, \ldots, x_i\}$, при чему се избацују (непотребни) збирови који су већи од t (ако збир у неком тренутку постане већи од t, онда ће тим пре после додавања новог сабирка бити већи од t).

Ако је L скуп, а x природни број, нека је

$$L + x = \{l + x \mid l \in L\}.$$

Пример 1.1 *Ако је* $L = \{1, 2, 3, 5, 9\}$, *онда је* $L + 2 = \{3, 4, 5, 7, 11\}$.

Нека $Merge(L, L')$ означава резултат обједињавања два скупа (сортиране листе) L и L'. Ово обједињавање може се извршити алгоритмом сложености $O(|L|+|L'|)$. Итеративни алгоритам експоненцијалне сложености за решавање проблема збир подскупа може се описати следећим једноставним кодом.

$tacanZbirPodskupa(S, t)$
$n \leftarrow |S|$
$L_0 \leftarrow \{0\}$
for $i \leftarrow 1$ **to** n **do**
$\quad L_i \leq Merge(L_{i-1}, L_{i-1} + x_i)$
$\quad$избацити из L_i све бројеве веће од t
return највећи елемент L_n

Да бисмо разумели како овај алгоритам ради, нека је P_i скуп свих збирова бројева из подскупа $\{x_1, x_2, \ldots, x_i\}$.

Пример 1.2 *Ако је* $S = \{1, 4, 5\}$, *онда је* $P_1 = \{0, 1\}$, $P_2 = \{0, 1, 4, 5\}$, $P_3 = \{0, 1, 4, 5, 6, 9, 10\}$.

Пошто је
$$P_i = P_{i-1} \cup (P_{i-1} + x_i),$$
индукцијом се показује да листа L_i садржи све елементе P_i мање или једнаке од t. С обзиром да је дужина L_i у најгорем случају 2^i, сложеност алгоритма је у општем случају експоненцијална.

2 Потпуна полиномијална апроксимациона шема

До алгоритма полиномијалне сложености долази се "проређивањем" сваке листе L_i после њеног формирања. Идеја приближног алгоритма је да ако се у у листи L налазе две блиске вредности, онда ако нам је потребно само приближно решење, можемо од те две вредности да задржимо само једну. Прецизније, проређивањем листе L са параметром δ, $0 < \delta < 1$, добија се листа L' тако што се из L уклања сваки елемент y за који у L' постоји елемент z који покрива (апроксимира) y, тј. за који важе неједнакости
$$z \leq y \leq z(1 + \delta).$$
Сваки уклоњени елемент y је покривен неким необрисаним елементом z који задовољава горњу двоструку неједнакост.

Пример 2.1 *Ако је* $\delta = 0.1$ *и*

$$L = \{10, 11, 12, 15, 20, 21, 22, 23, 24, 29\},$$

онда је

$$L' = \{10, 12, 15, 20, 23, 29\}.$$

Овде је број 11 *покривен бројем* 10*, бројеви* 21 *и* 22 *покривени су бројем* 20*, а број* 24 *покривен је бројем* 23*.*

Пошто је сваки елемент проређене листе истовремено и елемент полазне листе, проређивање може да значајно смањи број елемената који се чувају, чувајући при томе елементе блиске свим обрисаним елементима. Наредни код описује алгоритам проређивања.

$Proredjivanje(L, \delta)$
// $L = \{y_1, y_2, \ldots, y_m\}$ — скуп бројева, сортирана листа
$m \leftarrow |L|$
$L' \leftarrow \{y_1\}$
$poslednji \leftarrow y_1$
for $i \leftarrow 2$ **to** m **do**
  if $y_i > poslednji * (1 + \delta)$ **then**
    $L' \leftarrow L' \cup \{y_i\}$ // додавање на крај новог највећег броја
    $poslednji \leftarrow y_i$
return L'

Алгоритам пролази елементе листе L редом према величини. Нови број се додаје у излазну листу L' само ако је први у листи или ако није покривен последњим бројем додатим у L'.

Уметањем проређивања у петљу тачног алгоритма долази се до приближног алгоритма. Аргументи су скуп $S = \{x_1, x_2, \ldots, x_n\}$ са n природних бројева (произвољним редоследом), задата граница суме t и "параметар апроксимације" ϵ, $0 < \epsilon < 1$. Резултат који алгоритам враћа је, као што ћемо видети, вредност z која је највише $1 + \epsilon$ пута мања од оптималног решења.

$priblizniZbirPodskupa(S, t, \epsilon)$
1 $n \leftarrow |S|$
2 $L_0 \leftarrow \{0\}$
3 **for** $i \leftarrow 1$ **to** n **do**
4 $\quad L_i \leq Merge(L_{i-1}, L_{i-1} + x_i)$
5 $\quad L_i \leq proredjivanje\left(L_i, \frac{\epsilon}{2n}\right)$
6 $\quad$ избацити из L_i све бројеве веће од t
7 нека је z^* највећи елемент у L_n
8 **return** z^*

Пример 2.2 *Нека је* $S = \{104, 102, 201, 101\}$, $t = 308$ *и* $\epsilon = 0.4$. *Параметар проређивања је* $\delta = \epsilon/8 = 0.05$. *Извршавањем алгоритма добија се*

линија 2: $L_0 = \{0\}$,

линија 4: $L_1 = \{0, 104\}$,
линија 5: $L_1 = \{0, 104\}$,
линија 6: $L_1 = \{0, 104\}$,

линија 4: $L_2 = \{0, 102, 104, 206\}$,
линија 5: $L_2 = \{0, 102, 206\}$,
линија 6: $L_2 = \{0, 102, 206\}$,

линија 4: $L_3 = \{0, 102, 201, 206, 303, 407\}$,
линија 5: $L_3 = \{0, 102, 201, 303, 407\}$,
линија 6: $L_3 = \{0, 102, 201, 303\}$,

линија 4: $L_4 = \{0, 101, 102, 201, 203, 302, 303, 404\}$,
линија 5: $L_4 = \{0, 101, 201, 302, 404\}$,
линија 6: $L_4 = \{0, 101, 201, 302\}$.

Алгоритам враћа као одговор $z^* = 302$, *што је само за* 2% *(уместо за захтеваних* $\epsilon = 40\%$*) лошије од оптималног решења* $307 = 104 + 102 + 101$.

Теорема Алгоритам $priblizniZbirPodskupa$ је потпуна полиномијална апроксимациона шема за проблем збир подскупа.

Пре него што пређемо на доказ теореме, доказаћемо четири леме — елементарна помоћна тврђења. Препоручује се да се леме докажу самостално, као задаци за вежбу.

Лема 1. Нека је $q = 1 + \epsilon/(2n)$. За $i = 1, 2, \ldots, n$, за сваки елемент $y \in P_i$ који је мањи или једнак од t, постоји елемент $z \in L_i$, такав да је

$$z \leq y \leq zq^i.$$

Доказ. Доказ се изводи индукцијом по i. За $i = 1$ је $P_1 = L_1$, па је тврђење очигледно тачно. Претпоставимо да је тврђење леме тачно за елементе P_i.

Нека је y' произвољни елемент скупа P_{i+1}. С обзиром на начин формирања скупа P_{i+1}, постоје две могућности:

Случај $y' = y \in P_i$**.** Према индуктивној хипотези постоји $z \in L_i$, такво да је $z \leq y' \leq zq^i$. У вези са елементом z постоје две могућности:

- Ако је елемент z изостављен из L_{i+1} због проређивања, онда је он покривен неким елементом $z' \in L_{i+1}$ и тада је $z' \leq z \leq z'q$.
- Ако елемент z није изостављен из L_{i+1} због проређивања, онда је он "покривен" самим собом, јер за $z' = z$ тривијално важи $z' \leq z \leq z'q$.

Дакле, у оба случаја за елеменат $z \in L_i$ постоји елеменат $z' \in L_{i+1}$ такав да је $z' \leq z \leq z'q$. Комбиновањем са неједнакостима $z \leq y' \leq zq^i$, добија се да је

$$z' \leq z \leq y' \leq zq^i \leq z'q \cdot q^i = z'q^{i+1}.$$

Случај $y' = y + x_{i+1}$**, где је** $y \in P_i$**.** Према индуктивној хипотези постоји $z \in L_i$, такво да је $z \leq y \leq zq^i$. У вези са збиром $z + x_{i+1}$ постоје две могућности:

- Ако је збир $z + x_{i+1}$ изостављен из L_{i+1} због проређивања, онда је он покривен неким елементом $z' \in L_{i+1}$ и тада је $z' \leq z + x_{i+1} \leq z'q$.
- Ако збир $z + x_{i+1}$ није изостављен из L_{i+1} због проређивања, онда је он "покривен" самим собом, јер за $z' = z + x_{i+1}$ тривијално важи $z' \leq z + x_{i+1} \leq z'q$.

Дакле, у оба случаја за збир $z + x_{i+1}$ постоји елеменат $z' \in L_{i+1}$ такав да је $z' \leq z + x_{i+1} \leq z'q$.

Додавањем x_{i+1} двострукој неједнакости $z \leq y \leq zq^i$, добија се

$$z + x_{i+1} \leq y' = y + x_{i+1} \leq zq^i + x_{i+1}.$$

Пошто је $z' \leq z + x_{i+1}$, добија се да је $z' \leq z + x_{i+1} \leq y'$.

Са друге стране, због $z + x_{i+1} \leq z'q$ добија се

$$y' = y + x_{i+1} \leq zq^i + x_{i+1} < zq^i + x_{i+1}q^i = (z + x_{i+1})q^i \leq z'q \cdot q^i = z'q^{i+1}.$$

Тиме је и у овом случају доказано да је

$$z' \leq y' = y + x_{i+1} \leq z'q^{i+1}. \qquad \blacksquare$$

Лема 2. Ако је $0 \leq x \leq 1$, онда важи неједнакост

$$\ln(1 + x) \geq \frac{x}{1 + x}.$$

Доказ. Функција

$$f(x) = \ln(1 + x) - 1 + \frac{1}{1 + x}, \quad 0 \leq x \leq 1,$$

сменом $t = \frac{1}{1+x}$ постаје

$$g(t) = -\ln t - 1 + t, \quad \frac{1}{2} \leq t \leq 1.$$

Извод ове функције је на интервалу $[1/2, 1]$ мањи или једнак од нуле: $g'(t) = -\frac{1}{t} + 1 \leq -\frac{1}{1} + 1 = 0$. Дакле, функција $g(t)$ је опадајућа, па важи неједнакост $g(t) \geq g(1) = 0$. Скупови $\{g(t) \mid \frac{1}{2} \leq t \leq 1\}$ и $\{f(x) \mid 0 \leq x \leq 1\}$ су једнаки, па на интервалу $[0, 1]$ важи неједнакост $f(x) \geq 0$, еквивалентна неједнакости коју треба доказати. $\blacksquare$

Лема 3. Низ $\left(1 + \frac{\epsilon}{2n}\right)^n$, $n = 1, 2, \ldots$ је растући.

Доказ. Из тврђења леме 2 следи да је за $0 < x \leq 1$ негативан извод

$$\left(\frac{1}{x}\ln(1 + x)\right)' = \frac{1}{x^2}\left(\frac{x}{1 + x} - \ln(1 + x)\right) < 0.$$

Дакле, функција $\frac{1}{x}\ln(1+x)$ је опадајућа, функција $x\ln(1+\frac{1}{x})$ је растућа, као и функција

$$e^{x\ln(1+\frac{1}{x})} = (1 + 1/x)^x.$$

Замењујући x са $\frac{2n}{\epsilon}$, закључујемо да је низ $\left(1 + \frac{\epsilon}{2n}\right)^{2n/\epsilon}$, а тиме и низ $\left(1 + \frac{\epsilon}{2n}\right)^n$, $n = 1, 2, \ldots$, растући, што је требало доказати. $\blacksquare$

Лема 4. За свако x, $0 \leq x \leq 1$, важи неједнакост

$$e^{x/2} \leq 1 + x.$$

Доказ. Посматрајмо разлику леве и десне стране ове неједнакости, функцију
$$f(x) = e^{x/2} - 1 - x, \quad 0 \leq x \leq 1.$$
На интервалу $[0, 1]$ извод функције је
$$f'(x) = \frac{1}{2}e^{x/2} - 1 \leq \frac{1}{2}e^{1/2} - 1 < \frac{1}{2} \cdot 2 - 1 = 0.$$
Према томе, функција $f(x)$ је опадајућа и важи неједнакост $f(x) \leq f(0) = 0$, еквивалентна неједнакости коју треба доказати. ■

Доказ теореме. Сви елементи који преостану у листи L_i у линији 6 кода (после проређивања и избацивања бројева већих од t) су елементи одговарајућег скупа P_i (који се добија тачним алгоритмом). Према томе, вредност z^* која се добија у линији 8 кода је збир неког подскупа бројева из S. Нека је $y^* \in P_n$ оптимално решење проблема збир подскупа. Због линије 6 кода важи неједнакост $z^* \leq y^*$. Према томе, да се докаже теорема, потребно је доказати

- да важи неједнакост $y^* \leq z^*(1 + \epsilon)$, и
- да је време извршавања алгоритма полиномијална функција од величине улаза и вредности $1/\epsilon$.

Нека је
$$q = 1 + \frac{\epsilon}{2n}.$$
Пошто је $y^* \in P_n$, на основу леме 1 закључујемо да неједнакост
$$z \leq y^* \leq zq^n$$
важи за неки елемент $z \in L_n$. Пошто је $z^* \geq z$ највећи елемент скупа L_n, добија се да је
$$y^* \leq z^* q^n = z^* \left(1 + \frac{\epsilon}{2n}\right)^n.$$
На основу леме 3 је низ $(1 + \frac{\epsilon}{2n})^n$ растући. Гранична вредност тог низа је
$$\lim_{n \to \infty} \left(1 + \frac{\epsilon}{2n}\right)^n = \left(\lim_{n \to \infty} (1 + \frac{\epsilon}{2n})^{2n/\epsilon}\right)^{\epsilon/2} = e^{\epsilon/2}.$$
Комбиновањем са лемом 4 добија се неједнакост
$$\left(1 + \frac{\epsilon}{2n}\right)^n \leq e^{\epsilon/2} \leq 1 + \epsilon.$$

Према томе, доказана је неједнакост

$$z \leq y^* \leq z(1+\epsilon).$$

Да се докаже други део тврђења теореме, треба одредити горњу границу дужине L_i. После прорeђивања узастопни елементи $z < z'$ у низу L_i задовољавају услов $z' > z(1+\frac{\epsilon}{2n})$. Листа L_i може да садржи бројеве 0, 1 и највише још $\lfloor \log_{1+\epsilon/(2n)} t \rfloor$ других природних бројева, па је

$$|L_i| \leq \log_{1+\epsilon/(2n)} t + 2 = \frac{\ln t}{\ln(1+\epsilon/(2n))} + 2$$

На основу леме 2 може се са доње стране ограничити именилац овог разломка, па је

$$|L_i| \leq \frac{\ln t}{\frac{\epsilon/(2n)}{1+\epsilon/(2n)}} + 2 = \frac{2n(1+\epsilon/(2n))\ln t}{\epsilon} + 2 \leq \frac{3n \ln t}{\epsilon} + 2$$

(јер је $1+\epsilon/(2n) \leq 1 + 1/(2\cdot 1) = 3/2$). Овај израз је полином по n, $\ln t$ и по $1/\epsilon$. Пошто је величина улаза $O(n+\log t)$, тиме је завршен доказ теореме. ■