

LEAP Optimization version of Subset Sum
 (determine S^* of maximum $C(S^*) = \sum_{s \in S^*} s$
 among all subsets $S' \subseteq S : C(S') \leq t$)
EXHAUSTIVE SEARCH IN $\{C(S') \leq t : S' \subseteq S\}$

EXP-SS(S, t)

* Let $S = \{x_1, x_2, \dots, x_n\}$ *

$L_0 = \langle \rangle$

for $i \leftarrow 1$ to n do

$L_i \leftarrow \text{MERGE-VD}(L_{i-1}, L_{i-1} + x_i)$

return $\text{MAX}(L_n)$

CORRECTNESS: L_i contains all feasible costs
 of subsets in $S_i = \{x_1, \dots, x_i\}$

RUNNING TIME: In the worst case :

$$\begin{cases} |L_i| = 2 |L_{i-1}| \\ |L_0| = 1 \end{cases}$$

$$\Rightarrow |L_i| = 2^i$$

\Rightarrow The time of $\text{MERGE-VD}(L_i, L_{i-1} + x_i)$
 is $\Theta(2^i)$

$$\Rightarrow T_{\text{ESS}}(n) = \Theta(2^n) !$$

OBSERVATION: $2^{|S|}$ can be $\Theta(2^{\lfloor \log n \rfloor})$!

FROM EXHAUSTIVE SEARCH TO FPTAS:

Rather than just eliminating duplicates and zeroes $\geq t$, we also eliminate costs of subsets that are "well represented" by other costs.

Eliminating entries from the L_i^t 's reduces the size of the lists

APPROXIMATION IDEA: If two costs are sufficiently "close" we consider them as "duplicates", and keep only one of them (the smaller of the two to preserve feasibility).

REMARK: "Closeness" will depend on ϵ !

The smaller ϵ , the closer the costs must be so that the smaller cost can represent the larger

DEF For $S > 0$, a S -trimming of an ordered list L is a sublist L' of L such that:

$$\forall y \in L \exists z \in L' : \frac{y}{1+S} \leq z \leq y$$

In a S -trimming, element $z \in L'$ "is close" to element y . z is smaller than or equal to y but cannot be much smaller!

Observe that each element $z \in L'$ may represent itself (e.g. 0 always represents itself)

EXAMPLE $L = \langle 0, 10, 11, 12, 20, 21 \rangle$, $S = \frac{1}{10}$
 Then $L' = \langle 0, 10, 12, 20 \rangle$ is a S -trimming of L ,
 since $\frac{11}{10} \leq 10 \leq 11$ and $\frac{21}{1+1/10} \leq 20 \leq 21$

We can obtain the minimum size S -trimming greedily in linear time:

TRIM(L, S)

* Let $L = \langle y_0, y_1, \dots, y_{k-1} \rangle$ $\{ |L| = k \}$

GC $\leftarrow y_0$ {greedy choice}

$L' \leftarrow \langle GC \rangle$

for $i \leftarrow 1$ to $k-1$ do

if $(GC < y_i/(1+S))$ {GC not close enough to y_i }
 then

GC $\leftarrow y_i$

$L' \leftarrow \langle L', GC \rangle$

return L'

Correctness: $\forall y \in L - L' \exists z \in L' : y/(1+S) \leq z \leq y$

PROOF : let $y = y_i$ and let GC_i the greedy choice at iteration i . Clearly, $GC_i = y_j$ for some $j \leq i \Rightarrow y_j = GC_i \leq y = y_i$. Also, since $y = y_i \notin L'$ it must be $GC_i \geq y_i/(1+\delta)$. The proof follows setting $\varepsilon = GC_i \in L'$

EXERCISE Prove that $L' = \text{TRIM}(L, \delta)$ is a minimum trimming (standard analysis of a greedy algorithm). However, we will not use minimality directly but the following :

PROPERTY Let $L' = \langle z_0, z_1, \dots, z_{k-1} \rangle = \text{TRIM}(L, \delta)$. Then, for $1 \leq j < k-1$: $z_{j+1}/z_j > (1+\delta)$
(The property does not hold for $j=0$ since $z_0=0$)

PROOF z_{j+1} and z_j are two consecutive greedy choices. Thus it must be $z_j < z_{j+1}/(1+\delta) \Rightarrow z_{j+1}/z_j > (1+\delta)$

We are ready to transform EXR-SS($\langle S, \epsilon \rangle$) into a FPTAS:

(identical to

EXR-SS(S, ϵ)

but for trimming)

with $S = \underline{\Sigma}$

Z_M

APPROX-SS(S, ϵ, ε)

* Let $S = \{x_1, x_2, \dots, x_n\}$ *

$L'_0 = \langle \phi \rangle$

for $i \leftarrow 1$ to n do

$L'_i = \text{MERGE-VD}(L'_{i-1}, L'_{i-1} + x_i)$

$\Rightarrow L'_i = \text{TRIM}(L'_i, \varepsilon/(2n))$

return $\text{MAX}(L'_n)$

The algorithm is clearly correct, since it returns a feasible cost $z' = \text{MAX}(L_n)$
 $(L_n \subseteq L_m)$

Let $z^* = \text{cost}(S^*) = \sum_{s \in S^*} s$. We must prove that $\beta_\varepsilon = z^*/z' \leq 1 + \varepsilon$.

We have to study how the trimming error propagates.

LEMMA Let $S' \subseteq S_i = \{x_1, x_2, \dots, x_i\}$ and let $y = c(S') = \sum_{s \in S'} s$. For $0 \leq i \leq n$, let L'_i be the list computed by A_{SS} at the end of iteration i , for $0 \leq i \leq n$. Then

$$\exists z \in L'_i : \frac{y}{(1 + \frac{\varepsilon}{2n})^i} \leq z \leq y$$

The lemma states that the "closeness" of the representative z gets "looser" exponentially fast in i .

PROOF Induction on i . The base $i=1$ is trivial, since $L'_1 = \langle 0, x_1 \rangle = L_1$ (all possible costs in $S_1 = \{x_1\}$). Let $i > 1$ and assume that the HP holds for $i-1$. Two cases:

① $x_i \notin S' \subseteq \{x_1, \dots, x_{i-1}\}$. Then $S' \subseteq S_{i-1} = \{x_1, \dots, x_{i-1}\}$ thus HP holds and for $y = c(S') = \sum_{s \in S'} s$:

$$\exists z' \in L'_{i-1} : \frac{y}{(1 + \frac{\varepsilon}{2n})^{i-1}} \leq z' \leq y$$

Two cases:

1.a z' not eliminated in $\text{TRM}(L'_i, \varepsilon/2n)$

Then we have:

$$\frac{y}{(1+\varepsilon/2n)^{i_1}} \leq \frac{y}{(1+\varepsilon/2n)^{i_1}} \leq z' \leq y$$

\wedge
 L'_i

and the thesis follows setting $z = z'$

1.b z' is eliminated in $\text{TRM}(L'_i, \varepsilon/2n)$

Then $\exists z'' \in L'_i :$

$$\frac{y}{(1+\varepsilon/2n)^{i_1}} \leq \frac{y}{(1+\varepsilon/2n)^{i_1}} \cdot \frac{1}{(1+\varepsilon/2n)} \leq z' \leq z'' \leq z' \leq y$$

\wedge
 L'_i

and the thesis follows setting $z = z''$

② $x_i \in S' \subseteq S_i = \{x_1, \dots, x_i\}$. Let $y = c(S') = \sum_{S \in S'} S$

We have $S = S'' \cup \{x_i\}$, with $S'' \in S_{i-1}$,

and $y = c(S'') + x_i \Rightarrow c(S'') = y - x_i$

Then, by HP: $\exists z' \in L'_{i-1} :$

$$\frac{y-x_i}{(1+\varepsilon/2n)^{i_1}} \leq z' \leq y-x_i$$

Thus

$$\frac{y}{(1+\varepsilon/2n)^{i_1}} \leq \frac{y-x_i}{(1+\varepsilon/2n)^{i_1}} + x_i \leq z' + x_i \leq y$$

Observe that $z' + x_i \in L'_{i-1} + x_i$.

Again we have two subcases

2.a $z' + x_i \in L'_i \Rightarrow \frac{y}{(1+\varepsilon/2)^i} \leq z' + x_i \leq y$

(reasoning as in 1.a)

2.b $z' + x_i \notin L'_i \Rightarrow \exists z'' \in L_i : z'' \leq z' + x_i$

$$\frac{y}{(1+\varepsilon/2)^i} \leq z'' \leq y$$

(reasoning as in 1.b)

Let now $y^* = c(S^*) = \sum_{S \in S^*} s$, $S^* \subseteq S = S_n$

The revenue implies that $\exists z' \in L_n$:

$$\frac{y^*}{(1+\frac{\varepsilon}{2n})^n} \leq z' \leq y^*$$

Therefore: $\frac{y^*}{z'} \leq (1+\frac{\varepsilon}{2n})^n$

OBSERVATION We return $\text{MAX}(L_n) \geq z$

Thus $g(n) = \frac{c(S^*)}{\text{MAX}(L'_n)} \leq \frac{y^*}{z'} \leq (1+\frac{\varepsilon}{2n})^n$

It suffices to show that

$$(1+\frac{\varepsilon}{2n})^n \leq (1+\varepsilon)!$$

But:

$$\frac{d}{dn} (1+\frac{\varepsilon}{2n})^n > 0 \text{ and } \lim_{n \rightarrow \infty} (1+\frac{\varepsilon}{2n})^n = e^{\varepsilon/2}$$

Since $e^{\varepsilon/2} < 1+\varepsilon/2 + (\varepsilon/2)^2 < 1+\varepsilon/2 + \varepsilon/2 = 1+\varepsilon$

it follows that

$$g \leq \left(1 + \frac{\varepsilon}{2n}\right)^m < \lim_{n \rightarrow \infty} \left(1 + \frac{\varepsilon}{2n}\right)^m = e^{\varepsilon/2} < 1 + \varepsilon$$

\Rightarrow A-SS(s, t, ε) is a $(1+\varepsilon)$ -approximation!

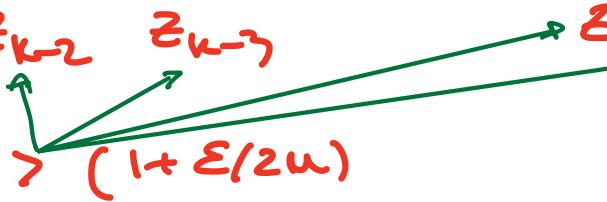
We are not finished! We need to bound the running time.

Each iteration $i \geq 1$ requires time $\Theta(|L'_i|)$. It suffices to bound $\sum_{i=1}^m |L'_i| = \sum_{i=0}^{m-1} |L_i|$.

Let $L'_i = \langle z_0 = 0, z_1, \dots, z_{k-1} \rangle$ ($|L'_i| = k$)

We prove that $\forall 1 \leq j < k-1 : z_{j+1}/z_j > \left(1 + \frac{\varepsilon}{2n}\right)$ (see PROPERTY). We can then write

$$\frac{z_{k-1}}{z_1} = \frac{z_{k-1}}{z_{k-2}} \cdot \frac{z_{k-2}}{z_{k-3}} \cdot \dots \cdot \frac{z_3}{z_2} \cdot \frac{z_2}{z_1}$$



telescoping product

thus

$$\frac{z_{k-1}}{z_1} > \left(1 + \frac{\varepsilon}{2n}\right)^{k-2}$$

Moreover

$$t \geq z_{k-1} > \left(1 + \frac{\varepsilon}{2n}\right)^{k-2} \cdot z_1 \geq \boxed{\left(1 + \frac{\varepsilon}{2n}\right)^{k-2}}$$

Let us "solve" in K :

$$(n-2) \ln\left(1 + \frac{\varepsilon}{2n}\right) < \ln t$$

$$\Rightarrow K < \frac{\ln t}{\ln\left(1 + \frac{\epsilon}{2m}\right)} + 2$$

$\|L'_i\|$

However, for $|x| < 1$: $\ln(1+x) \geq \frac{x}{1+x}$

$$\Rightarrow \|L'_i\| = K \leq \frac{\ln t}{\frac{\epsilon}{2m}\left(1 + \frac{\epsilon}{2m}\right)} + 2$$

$$\leq \frac{2m}{\epsilon} \left(1 + \frac{\epsilon}{2m}\right) \ln t + 2$$

$\leq 3t$

$$= O\left(\frac{n}{\epsilon} \log t\right)$$

Therefore

$$T_{ASS} = O\left(\sum_{i=0}^{n-1} \|L'_i\|\right) = O\left(\frac{n^2 \log t}{\epsilon}\right)$$

But $n = O(|\langle s, t \rangle|)$

and $\log t = O(|\langle s, t \rangle|)$

Thus $T_{ASS}(|\langle s, t \rangle|, \epsilon) = O\left(\frac{|\langle s, t \rangle|^3}{\epsilon}\right)$

which is polynomial both in $|\langle s, t \rangle|$
and $\frac{1}{\epsilon}$!

EXERCISE: Extend A-S to return the subset associated to $\text{MAX}(L^n)$

RANDOMIZED APPROXIMATION ALGORITHMS

- The use of randomization can help achieve simple and good approximations to hard optimization problems.

A RANDOMIZED ALGORITHM is an algorithm which may use the primitive $\text{RANDOM}(S)$, where S is any finite set.

$X = \text{RANDOM}(S) \in S$ is a random variable returning any element of S with probability $\frac{1}{|S|}$ (uniform distr. over S)

Different runs of RANDOM :

$X_1 = \text{RANDOM}(S_1)$; $Y = \text{RANDOM}(S_2)$ are probabilistically independent, in the sense that

$$\forall x \in S_1, \forall y \in S_2: \Pr((X_1 = x) \wedge (Y = y)) = \frac{1}{|S_1|} \cdot \frac{1}{|S_2|}$$

$X = \text{RANDOM}(S)$: extraction from set S with replacement

SPECIAL CASE

$X = \text{RANDOM}\{0, 1\}$ ($\Pr(X=0) = \Pr(X=1) = \frac{1}{2}$) is called a COIN FLIP

REMARK the use of RANDOM implies that:

1. The algorithm is not deterministic anymore.

⇒ returned value and running time are random variables

2. It may be the case that a randomized algorithm is not always correct

⇒ correctness is an event associated to a probability

Given an optimization problem Π a randomized approximation algorithm $A_\Pi^r(i)$ returns a solution $s \in S(i)$ which is a random variable (r.v.) and whose cost $c(s)$ is also r.v.

DEFINITION

The approximation ratio $\rho(i)$ of $A_{\text{PT}}^R(i) = S_A$ is defined as

$$\rho(i) = \max \left\{ \frac{E(\text{CCS}_A)}{c(S^*)}, \frac{c(S^*)}{E(\text{CGS}_A)} \right\}$$