

Projet IPF: SUBSET-SUM-OPT

Romain PEREIRA

03/04/2018

Sommaire

| | | |
|----------|---|----------|
| 1 | Préambule | 2 |
| 2 | Spécificités techniques | 3 |
| 2.1 | Notations | 3 |
| 2.2 | Rappels/Complexités | 3 |
| 3 | Approche naïve | 4 |
| 3.1 | Question 1 : somme sur un ensemble | 4 |
| 3.2 | Question 2 : génération des parties d'un ensemble | 4 |
| 3.3 | Question 3 : résolution de SUBSET_SUM_OPT par force brute | 4 |
| 4 | Approche plus directe | 5 |
| 4.1 | Question 4 : sommes atteignables | 5 |
| 4.2 | Question 5 : résolution de SUBSET_SUM_OPT | 5 |
| 5 | Approche avec nettoyage | 6 |
| 5.1 | Question 6 ; fonction clean_up | 6 |
| 5.2 | Question 7 : résolution de SUBSET_SUM_OPT | 6 |
| 6 | Approche de type <i>Diviser pour régner</i> | 7 |
| 6.1 | Question 8 : is_feasible | 7 |
| 6.2 | Question 9 : best_feasible | 7 |
| 6.3 | Question 10 : résolution de SUBSET_SUM_OPT | 8 |
| 7 | Références | 9 |

1 Préambule

Ce projet est réalisé dans le cadre de mes études à l'ENSIIE. Rappel de l'énoncé:

SUBSET-SUM-OPT - Etant donnée un ensemble fini E d'entiers strictements positifs et un entier cible s , trouver l'entier $s' \leq s$ le plus grand possible, tel qu'il existe un sous-ensemble $E' \subseteq E$ vérifiant $\sum_{e \in E'} e = s'$.

Ce rapport présente (en pseudo-code), les algorithmes implémentés.

Le code OCaml est disponible dans le rendu.

Un Makefile est disponible pour compiler le projet et les tests:

- *make* : compile les fichiers sources avec les tests vers un exécutable **subset-sum-opt.out**
- *clean* : supprime les fichiers compilés temporaires (.mlo et .cmo)
- *fclean* : supprime tous les fichiers compilés (exécutable et temporaires)

Les tests sont unitaires et effectués dans l'ordre:

- Les fonctions du module *Listes*
- Les fonctions du module *Ensembles*
- Les fonctions du module *Approche_naïve*
- Les fonctions du module *Approche_plus_directe*
- Les fonctions du module *Approche_avec_nettoyage*
- Les fonctions du module *Approche_diviser_pour_regner*
- Les fonctions du module *Approche_diviser_pour_regner*
- Comparaisons des approches

2 Spécificités techniques

2.1 Notations

J'utiliserai les notations suivantes:

- Soit E est un ensemble, $Card(E)$ est le cardinal de E .
- Soit E est un ensemble, $P(E)$ est l'ensemble formé des parties de E

2.2 Rappels/Complexités

Soit E un ensemble tel que $Card(E) = n$, alors

$$Card(P(E)) = 2^n$$

$$\forall E' \in P(E), 0 \leq Card(E') \leq n$$

On considèrera les complexités pour les opérations suivantes:

- Soit (x, y) des scalaires, je considère que les 4 opérations élémentaires $(x + y, x - y, x * y, x / y)$ sont en $O(1)$
- Soit 2 ensembles X, Y , $X \cup Y$ est une opération en $O(Card(X) * Card(Y))$
- Soit $E \subset \mathbb{N}$, $\max E$ est une opération en $O(Card(E))$

3 Approche naïve

Cette approche consiste à déterminer tous les sous-ensembles E' de E , d'effectuer la somme sur tous les E' , et de renvoyer la somme la plus proche de s . Cette approche par 'force brute' est lourde.

3.1 Question 1 : somme sur un ensemble

Algorithm 1: Renvoie la somme des éléments de E

```
1: function SOMME( $E' \subset \mathbb{N}$ )
2:   if  $E' = \emptyset$  then
3:     return 0
4:   end if
5:   Soit  $x \in E'$ 
6:   return  $x + \text{Somme}(E' \setminus \{x\})$ 
7: end function
```

Complexité en $\boxed{O(n)}$, où $n = \text{Card}(E')$

3.2 Question 2 : génération des parties d'un ensemble

Algorithm 2: Renvoie l'ensemble des parties de E

```
1: function SOUS-ENSEMBLES( $E \subset \mathbb{N}$ )
2:   if  $E = \emptyset$  then
3:     return  $\{\emptyset\}$ 
4:   end if
5:   Soit  $x \in E$ 
6:   Soit  $P \leftarrow \text{Sous-ensembles}(E \setminus \{x\})$ 
7:   return  $P \cup \{E' \cup x \mid E' \in P\}$ 
8: end function
```

Complexité en $\boxed{O(2^n)}$, où $n = \text{Card}(E)$

3.3 Question 3 : résolution de SUBSET_SUM_OPT par force brute

Algorithm 3: Renvoie la réponse au problème SUBSET_SUM_OPT sur (E, s)

```
1: function SUBSET_SUM( $E \subset \mathbb{N}, s \in \mathbb{N}$ )
2:   Soit  $P \leftarrow \text{Sous-ensembles}(E)$ 
3:   return  $\max_{E' \in P} \{s' \mid s' = \text{Somme}(E') \text{ et } s' \leq s\}$ 
4: end function
```

Complexité en $\boxed{O(2^{n+1}n)}$, où $n = \text{Card}(E)$

4 Approche plus directe

Dans cette approche, plutot que de calculer l'ensemble des parties de E , on se propose de calculer l'ensemble des sommes atteignables en sommet sur les parties de E .

4.1 Question 4 : sommes atteignables

Algorithm 4: Renvoie l'ensemble des entiers s tels qu'il existe $E' \subseteq E$ vérifiant $\sum_{e \in E'} e = s$

```
1: function GET_ALL_SUMS( $E \subset \mathbb{N}$ )
2:   if  $E = \emptyset$  then
3:     return  $\{0\}$ 
4:   end if
5:   Soit  $x \in E$ 
6:    $S \leftarrow \text{get\_all\_sums}(E \setminus \{x\})$ 
7:   return  $S \cup \{x + s \mid s \in S\}$ 
8: end function
```

Si l'on suppose:

- $n = \text{Card}(E)$
- $m(n) = \text{Card}\{\text{sommes atteignables}\} = \{s \mid \exists E' \subseteq E \mid \sum_{e \in E'} e = s\} \leq 2^n$

Alors la complexité de cet algorithme est en $\boxed{O(m(n)^2 * n)}$

- n : nombre de récursion
- $m(n)^2$: l'union

4.2 Question 5 : résolution de SUBSET_SUM_OPT

Algorithm 5: Renvoie la réponse au problème SUBSET_SUM_OPT sur (E, s)

```
1: function SUBSET_SUM( $E \subset \mathbb{N}, s \in \mathbb{N}$ )
2:   Soit  $S \leftarrow \text{get\_all\_sums}(E)$ 
3:   return  $\max\{s' \in S \mid 0 \leq s' \leq s\}$ 
4: end function
```

La complexité de cet algorithme de résolution est donc en $\boxed{O(m(n)^2 * n)}$

5 Approche avec nettoyage

On peut réduire la complexité de l'approche précédente en réduisant ce que l'on a noté $m(n)$ (le nombre de sommes atteignables. Dans cette approche, on se propose d'ajouter un 'filtre' sur l'algorithme qui génère les sommes.

5.1 Question 6 ; fonction `clean_up`

Le filtre (l'algorithme '`clean_up`') est donnée dans l'énoncé. Cette fonction prends en paramètre:

- $E \subset \mathbb{N}$: l'ensemble a filtré
- $s \in \mathbb{N}$: un entier positif
- $\delta \in \mathbb{R}_+^*$: un réel positif (généralement $\ll 1$)

Cette fonction renvoie un nouvel ensemble $E' \subset E$, tel que:

- $\forall x \in E', x \leq s$
- Si l'on considère la suite croissante $(u_n)_{n \in \mathbb{N}}$ des éléments de E' , $\forall n \in \mathbb{N}, u_{n+1} \geq (1 + \delta)u_n$.

Autrement dit, tous les entiers strictement supérieurs à s sont supprimés, et si l'on considère 2 entiers consécutifs de l'ensemble trié x et y , tel que $x < y$, ils sont 'proches d'un rapport d'au moins $(1 + \delta)$, y est supprimé. Ce filtre supprime donc les éléments de l'ensemble qui sont 'proches' l'un de l'autre (au regard de δ)

5.2 Question 7 : résolution de SUBSET_SUM_OPT

L'algorithme de résolution est également fourni dans l'énoncé. Il est le même que celui de 'l'approche plus directe' 4, sauf que lors du calcul des sommes atteignables filtrés par la fonction '`clean_up`'.

Algorithm 6: Renvoie l'ensemble des entiers s tels qu'il existe $E' \subseteq E$ vérifiant $\sum_{e \in E'} e = s$, passant les tests du filtre

```
1: function GET_ALL_SUMS_2( $E \subset \mathbb{N}, \delta \in \mathbb{R}_+^*$ )
2:   if  $E = \emptyset$  then
3:     return  $\{0\}$ 
4:   end if
5:   Soit  $x \in E$ 
6:    $S \leftarrow \text{get\_all\_sums\_2}(E \setminus \{x\})$ 
7:   return clean_up( $S \cup \{x + s \mid s \in S\}$ )
8: end function
```

Algorithm 7: Renvoie la réponse au problème SUBSET_SUM_OPT sur (E, s)

```
1: function SUBSET_SUM( $E \subset \mathbb{N}, s \in \mathbb{N}$ )
2:   Soit  $S \leftarrow \text{get\_all\_sums\_2}(E, \delta)$ 
3:   return  $\max\{s' \in S \mid 0 \leq s' \leq s\}$ 
4: end function
```

La complexité de cet algorithme de résolution est donc en $\boxed{O(m'(n)^2 * n)}$, avec

$$m'(n) = \text{Card}\{\text{sommes atteignables filtrés}\} \leq m(n) \leq 2^n$$

. **Attention** cependant, pour δ trop grand, on perd l'optimalité du résultat.

6 Approche de type *Diviser pour régner*

6.1 Question 8 : `is_feasible`

On nous demande ici de créer une fonction qui sur la donnée d'un entier s , d'une liste $l1$ d'entier croissant, d'une liste $l2$ d'entier décroissant, renvoie *true* si s s'écrit comme la somme d'un élément de $l1$ et d'un élément de $l2$, et *false* sinon.

La fonction implémenté suit l'algorithme suivant, qui se sert au maximum du fait que les listes soit triés. Les 3 cas de retours possibles sont:

- **TOO_SMALL** : Soit $y \in l2$, $\forall x \in l1$, $x + y < s \Rightarrow \forall y' < y$, $x + y' < x + y < s$. En français, pour y fixé, la somme des $x + y$ avec x dans $l1$ est toujours inférieur à s . Pour un y plus petit, on ne risque donc pas d'atteindre une valeur plus proche de s .
- **REACHED** : on a réussi à atteindre s .
- **TOO_BIG** : On a parcouru toutes les valeurs de $l2$, pour toutes les valeurs de $l1$, sans avoir pu obtenir de couple $(x, y) \in l1 \times l2$ tel que $x + y \leq s$.

Algorithm 8: Renvoie **true** si $\exists (x, y) \in l1 \times l2 \mid x + y = s$, **faux** sinon

```
1: function IS_FEASIBLE( $s \in \mathbb{N}$ ,  $l1 \subset \mathbb{N}$ ,  $l2 \subset \mathbb{N}$ )
2:   for  $y \in l2$  (pris du plus grand au plus petit) do
3:     for  $x \in l1$  (pris du plus petit au plus grand) do
4:       if  $x + y = s$  then
5:         Renvoyer true (cas REACHED)
6:       else if  $x + y > s$  then
7:         Passer au  $y$  suivant (cas TOO_BIG)
8:       else (si  $x + y < s$ )
9:         Passer au  $x$  suivant (car  $x + y$  est déjà plus petit que 's')
10:      end if
11:    end for
12:    Renvoyer faux (cas TOO_SMALL)
13:  end for
14:  Renvoyer faux (on a eu que des cas TOO_BIG)
15: end function
```

6.2 Question 9 : `best_feasible`

Au regard de la Question 8 (6.1), on nous demande maintenant de créer une fonction qui sur la donnée d'un entier s , d'une liste $l1$ d'entier croissant, d'une liste $l2$ d'entier décroissant, renvoie le plus grand entier $s' \leq s$ somme d'un élément de $l1$ et d'un élément de $l2$

N.B: non explicitement dis dans le sujet, mais si un tel entier s' n'existe pas, -1 est renvoyé.

Algorithm 9: Renvoie $\max_{(x,y) \in l1 \times l2} \{x + y \mid x + y < s\}$

```
1: function BEST_FEASIBLE( $s \in \mathbb{N}$ ,  $l1 \subset \mathbb{N}$ ,  $l2 \subset \mathbb{N}$ )
2:    $s' \leftarrow -1$ 
3:   for  $y \in l2$  (pris du plus grand au plus petit) do
4:     for  $x \in l1$  (pris du plus petit au plus grand) do
5:       if  $x + y = s$  then
6:         Renvoyer  $s$  (cas REACHED)
7:       else if  $x + y > s$  then
8:         Passer au  $y$  suivant (cas TOO_BIG)
9:       else (si  $x + y < s$ )
10:        if  $x + y > s'$  then
11:           $s' \leftarrow x + y$ 
12:        end if
13:        Passer au  $x$  suivant
14:      end if
15:    end for
16:    Renvoyer  $s'$  (cas TOO_SMALL).
17:  end for
18:  Renvoyer  $s'$  (on a eu que des cas TOO_BIG)
19: end function
```

6.3 Question 10 : résolution de SUBSET_SUM_OPT

En se servant des fonctions 4.1, 6.2, en en ajoutant 2 fonctions subsidiaires (voir *Listes.ml*)

- **split** : sépare une liste en 2 sous liste de longueur égal (à 1 prêt)
- **reverse** : inverse les éléments d'une liste

on peut élaborer l'algorithme de résolution suivant, qui se base sur l'idée de 'diviser pour régner':

Algorithm 10: Renvoie la réponse au problème SUBSET_SUM_OPT sur (E, s)

```
1: function SUBSET_SUM( $E \subset \mathbb{N}$ ,  $s \in \mathbb{N}$ )
2:    $(l1, l2) \leftarrow \text{split}(l)$ 
3:    $(s1, s2) \leftarrow (\text{get\_all\_sums}(l1), \text{get\_all\_sums}(l2))$ 
4:   return  $\text{best\_feasible}(s, l1, \text{reverse}(l2))$ 
5: end function
```

7 Références

- [1] 'Writting Cache-Friendly code', Gerson Robboy, Portland State University
<http://web.cecs.pdx.edu/~jrb/cs201/lectures/cache.friendly.code.pdf>.