Sorbonne Université

3I003 : Algorithmique

Projet :

Confitures

Ahmed Boukerram
Angelo Ortiz

Licence d'Informatique
Année 2018/2019

Table des matières

Table des figures

Liste des tableaux

Introduction

Sur un forum de cuisine, on pourrait trouver une question portant sur le nombre minimal de pots nécessaires pour la préservation d'une certaine quantité de confiture préparée. Il est important de remplir au complet chaque bocal afin de prolonger au maximum le temps de conservation du produit. Mais la capacité des bocaux n'est pas la même pour tous. C'est pourquoi on veut proposer des algorithmes résolvant cette question : on se fixe pour objectif d'utiliser le moins possible de bocaux.

Pour définir ce problème, on dispose des données suivantes. Il y a S décigrammes de confiture que l'on doit verser dans des bocaux vides. On compte aussi plusieurs bocaux de diverses capacités que l'on range en k classes : chaque classe de bocal correspond à une capacité distincte. On appellera *système de capacités* l'ensemble des capacités dont on dispose. On note ces k capacités par un tableau V de taille k numéroté par ordre croissant :

- $V[1] < V[2] < \cdots < V[k]$; et
- chaque capacité $V[i]$ est exprimée par la quantité en décigrammes que l'on peut mettre dans un bocal.

On se fixe aussi certaines contraintes qui garantissent l'existence d'une solution au problème :

- la quantité S est un nombre entier de décigrammes ;
- la capacité minimal d'un bocal est 1 décigramme, i.e. $V[1] = 1$; et
- on dispose d'une très grande quantité (supposée ainsi illimitée) de bocaux de chacune des capacités.

En effet, en tenant compte des trois hypothèses, on remarque que l'on peut toujours verser la totalité de confiture dans des pots de 1 décigramme.

Notre objectif est de remplir le moins possible de bocaux et qu'ils soient tous remplis exactement à sa capacité maximale. Ainsi, étant donnés :

— un système de k capacités $V[i] \in \mathbb{N}$, $i \in \{1, \ldots, k\}$, avec $V[1] = 1$,

— et une quantité totale $S \in \mathbb{N}$ de confiture,

le but est de déterminer le nombre minimum de bocaux tels que la somme de leurs capacités est égale à S. On cherche donc à retourner un couple (n, A), où n est le nombre de bocaux utilisés et A est un tableau de taille k tel que $A[i]$ représente le nombre de bocaux de capacité $V[i]$ à remplir au complet. On a ainsi que $n = \sum_{i=1}^{k} A[i]$.

Les objectifs de ce projet sont l'analyse théorique et la mise en œuvre de trois algorithmes résolvant le problème décrit ci-dessus. Dans un premier temps, on formalise le problème et propose plusieurs algorithmes dont on fera l'analyse de complexité. Dans un deuxième temps, on implémente ces algorithmes au moyen du langage de programmation Python afin de vérifier expérimentalement les complexités trouvées précédemment.

Première partie
Partie théorique

1 Algorithme I : Recherche exhaustive

Question 1

Soit la propriété suivante $P_1(s), s \in \mathbb{Z}$: `RechercheExhaustive`(k, V, s) se termine et renvoie le nombre minimal de bocaux à remplir pour une quantité s de confiture avec un système de k capacités décrit dans les cases du tableau $V[1\,..\,k]$. Montrons-la par récurrence forte sur s.

Preuve

Cas négatif : Pour $s < 0$, la quantité de confiture est négative. Dans ce cas, on ne peut pas remplir les bocaux et on dit que le nombre de bocaux est infini. Par ailleurs, dans le corps de l'algorithme, on rentre dans le bloc `then` du premier branchement conditionnel et renvoie bien $+\infty$. De plus, on n'a effectué que des instructions élémentaires en nombre fini. De ces faits, `RechercheExhaustive` est valide et se termine dans le cas négatif.

Base : Pour $s = 0$, la quantité de confiture est nulle. Aussi, on rentre dans le bloc `else` du premier branchement conditionnel, puis dans le bloc `then` du deuxième branchement conditionnel. Donc, on renvoie 0, ce qui est correct puisque aucun bocal ne doit être rempli.

De la même manière que dans le cas où $s < 0$, on montre ici que l'algorithme se termine car la suite d'instructions effectuées ne comporte que des instructions élémentaires. De ces faits, $P_1(0)$ est vraie.

Induction : Supposons $P_1(s)$ vraie pour tout $s < s_0$, pour un $s_0 > 0$ fixé. Montrons $P_1(s_0)$ vraie.

Comme la quantité de confiture s_0 est positive, on a besoin d'*au moins* un bocal où la verser. On remarque qu'il n'y a pas de contrainte sur ce premier bocal. Pour s'assurer de bien calculer le nombre minimum de bocaux à remplir, on essaie donc toutes les capacités possibles pour le premier bocal, puis on résout le problème récursif induit par chacune d'entre elles. De cette manière il suffit de choisir la configuration optimale parmi les k possibles.

Par ailleurs, en ce qui concerne l'algorithme `RechercheExhaustive`, on rentre dans le bloc `else` du premier branchement conditionnel, puis dans le bloc `else` du deuxième branchement conditionnel, puisque $s_0 > 0$. On sait que la capacité minimum est 1 dg. On en déduit que le nombre *maximum* de bocaux à utiliser est s_0.

On cherche à améliorer la solution de départ. Pour ce faire, on teste chaque capacité $V[i], i \in \{1, \ldots, k\}$, pour le premier bocal à remplir. Il reste ainsi $s_0 - V[i]$ décigrammes de confiture à verser. D'après les conditions du problème, $V[i] \geq 1$, d'où $s_0 - V[i] < s_0$. Par hyporhèse de récurrence, `RechercheExhaustive`(k, V, s),

où $s = s_0 - V[i]$, se termine et est valide. On rajoute 1 à cette solution en raison du premier choix. On teste ensuite si cette solution améliore la solution courante et, le cas échéant, on la met à jour. On répète ce procédé pour chacune des capacités du tableau $V[1\,..\,k]$.

Comme mentionné précédemment, à la fin de l'examen des k capacités pour le premier bocal, on se retrouve avec la solution optimale. Étant donné que l'appel `RechercheExhaustive`(k, V, s_0) renvoie cette solution optimale, il est valide. Par ailleurs, par hypothèses de récurrence, les appels récursifs se terminent. On effectue k tours de boucle du bloc `for` qui contient ces appels récursifs et quelques instructions élémentaires de plus. On en déduit que `RechercheExhaustive`(k, V, s_0) se termine.

Conclusion :

$$\left.\begin{array}{l} \forall s < 0, P_1(s) \text{ vraie} \\ P_1(0) \text{ vraie} \\ \forall s_0 > 0, [(\forall s < s_0, P_1(s)) \Rightarrow P_1(s_0)] \end{array}\right\} \begin{array}{l} \forall s \in \mathbb{Z}, \texttt{RechercheExhaustive}(k, V, s) \\ \text{se termine et est valide.} \end{array}$$

Cas particulier

D'après le résultat précédant, pour $s = S \in \mathbb{N}$ (la quantité de confiture dans le cadre de ce projet), on a que l'algorithme est valide et se termine lorsqu'il est appelé par l'appel initial `RechercheExhaustive`(k, V, S).

Question 2

Soient $b(s)$ et $c(s)$ les suites définies par

$$b(s) = \begin{cases} 0 & \text{si } s = 0 \\ 2 & \text{si } s = 1 \\ 2 \cdot b(s-2) + 2 & \text{si } s \geq 2 \end{cases} \text{ et } c(s) = \begin{cases} 0 & \text{si } s = 0 \\ 2 \cdot c(s-1) + 2 & \text{si } s \geq 1 \end{cases}$$

Supposons que les seules capacités de bocal disponibles soient 1 dg et 2 dg. Soit $a(s)$ le nombre d'appels récursifs effectués par `RechercheExhaustive`$(2, [1, 2], s)$.

a) La suite $a(s)$ est définie par

$$a(s) = \begin{cases} 0 & \text{si } s = 0 \\ 2 & \text{si } s = 1 \\ a(s-2) + a(s-1) + 2 & \text{si } s \geq 2 \end{cases}$$

b) Soit la propriéte suivante $P_2(s), s \geq 0 : b(s) \leq a(s) \leq c(s)$. Montrons-la par récurrence d'ordre 2 sur s.

Preuve

Base : Pour $s = 0$, on a $b(0) = a(0) = c(0) = 0$. Donc, $P_2(0)$ est vraie.
Pour $s = 1$, on a $b(1) = a(1) = c(1) = 2$. Donc, $P_2(1)$ est vraie.

Induction : Supposons $P_2(s-2)$ et $P_2(s-1)$ vraies pour un $s \geq 2$ fixé. Montrons $P_2(s)$ vraie.
D'après les hypothèses de récurrence, on a :

$$b(s-2) \leq a(s-2) \leq c(s-2) \tag{1}$$
$$b(s-1) \leq a(s-1) \leq c(s-1) \tag{2}$$

Par définition de ces suites récursives, on a :

$$b(s) = 2 \cdot b(s-2) + 2 \tag{3}$$
$$a(s) = a(s-2) + a(s-1) + 2 \tag{4}$$
$$c(s) = 2 \cdot c(s-1) + 2 \tag{5}$$

Étant donné que les suites $b(s)$ et $c(s)$ sont croissantes (c.f. annexe A), les inégalités (1) et (2) deviennent respectivement

$$b(s-2) \leq a(s-2) \leq c(s-1)$$
$$b(s-2) \leq a(s-1) \leq c(s-1)$$

Il en résulte que

$$2 \cdot b(s-2) + 2 \leq a(s-2) + a(s-1) + 2 \leq 2 \cdot c(s-1) + 2 \tag{6}$$

En remplaçant les expressions (3), (4) et (5) en (6), on obtient les inégalités cherchées : $b(s) \leq a(s) \leq c(s)$.

Conclusion :

$$\left.\begin{array}{l} P_2(0) \text{ et } P_2(1) \text{ vraies} \\ \forall s \geq 2, [(P_2(s-2) \text{ et } P_2(s-1)) \Rightarrow P_2(s)] \end{array}\right\} \forall s \in \mathbb{N}, b(s) \leq a(s) \leq c(s)$$

c) Soit la propriéte suivante $P_3(s)$, $s \geq 0$: $c(s) = 2^{s+1} - 2$. Montrons-la par récurrence faible sur s.

Preuve

Base : Pour $s = 0$, on a $c(0) = 0$. De plus, $2^{0+1} - 2 = 2^1 - 2 = 0$. Donc, $P_3(0)$ est vraie.

Induction : Supposons $P_3(s)$ vraie pour un $s \geq 0$ fixé. Montrons $P_3(s+1)$ vraie.
On a $s + 1 \geq 1$.

Par définition,

$$c(s+1) = 2 \cdot c(s) + 2 \tag{7}$$

Or, par hypothèse de récurrence,

$$c(s) = 2^{s+1} - 2 \tag{8}$$

En remplaçant (8) en (7), on obtient

$$c(s+1) = 2 \cdot (2^{s+1} - 2) + 2 = 2^{s+2} - 2$$

D'où, $P_3(s+1)$ est vraie.

Conclusion :

$$\left.\begin{array}{l} P_3(0) \text{ vraie} \\ \forall s \geq 0, [P_3(s) \Rightarrow P_3(s+1)] \end{array}\right\} \ \forall s \in \mathbb{N}, c(s) = 2^{s+1} - 2$$

d) Soit la propriéte suivante $P_4(s)$, $s \geq 0$: $b(s) = c(\lceil \frac{s}{2} \rceil)$. Montrons-la par récurrence d'ordre 2 sur s.

Preuve

Base : Pour $s = 0$, on a $b(0) = 0 = c(\lceil \frac{0}{2} \rceil)$. Donc, $P_4(0)$ est vraie.
Pour $s = 1$, on a $b(1) = 2 = 2 \cdot 0 + 2 = c(\lceil \frac{1}{2} \rceil)$. Donc, $P_4(1)$ est vraie.

Induction : Supposons $P_4(s-2)$ et $P_4(s-1)$ vraies pour un $s \geq 2$ fixé. Montrons $P_4(s)$ vraie.
On a $s \geq 2$ et $\lceil \frac{s}{2} \rceil \geq 1$.
Par définition,

$$b(s) = 2 \cdot b(s-2) + 2 \tag{9}$$

$$c\left(\left\lceil \frac{s}{2} \right\rceil\right) = 2 \cdot c\left(\left\lceil \frac{s}{2} \right\rceil - 1\right) + 2 \tag{10}$$

Or, par hypothèse de récurrence,

$$b(s-2) = c\left(\left\lceil \frac{s-2}{2} \right\rceil\right) \tag{11}$$

En remplaçant (11) en (9), on obtient

$$b(s) = 2 \cdot c\left(\left\lceil \frac{s}{2} - 1 \right\rceil\right) + 2 \tag{12}$$

De plus, on sait que

$$\forall x \in \mathbb{R}, \forall p \in \mathbb{Z}, \lceil x + p \rceil = \lceil x \rceil + p$$

En particulier, pour (12) on a

$$b(s) = 2 \cdot c\left(\left\lceil \frac{s}{2} \right\rceil - 1\right) + 2 \tag{13}$$

Par transitivité entre (10) et (13), on obtient bien $b(s) = c(\lceil \frac{s}{2} \rceil)$.
De ce fait, $P_4(s)$ est vraie.

Conclusion :

$$\left.\begin{array}{l} P_4(0) \text{ et } P_4(1) \text{ vraies} \\ \forall s \geq 2, [(P_4(s-2) \text{ et } P_4(s-1)) \Rightarrow P_4(s)] \end{array}\right\} \ \forall s \in \mathbb{N}, b(s) = c(\lceil \tfrac{s}{2} \rceil)$$

e) On déduit des deux questions précédentes que

$$b(s) = 2^{\lceil \frac{s}{2} \rceil + 1} - 2$$

D'après la question b), on a $b(s) \leq a(s) \leq c(s)$.
De ces faits,

$$2^{\lceil \frac{s}{2} \rceil + 1} - 2 \leq a(s) \leq 2^{s+1} - 2$$

Pour le calcul de la complexité temporelle de l'algorithme, considérons seulement les appels récursifs effectués par `RechercheExhaustive`$(2, [1, 2], s)$. La complexité pour ce système de deux capacités est donnée par $a(S)$. La complexité est ainsi en $\Omega\left(\sqrt{2}^S\right)$ et $\mathcal{O}(2^S)$.

N.B. Pour la complexité de cet algorithme dans le cas général, regardez l'annexe B.

2 Algorithme II : Programmation dynamique

Question 3

On note $m(S)$ le nombre minimum de bocaux pour S décigrammes de confiture et un tableau de capacités V. On définit une famille de problèmes intermédiaires de la façon suivante : étant donnés un entier s et un entier $i \in \{1, \ldots, k\}$, on note $m(s, i)$ le nombre minimum de bocaux nécessaires pour une quantité totale s en ne choisissant des bocaux que dans le système de capacités $V[1], V[2], \ldots, V[i]$.

On pose :

$$\begin{aligned} m(0, i) &= 0 && \forall i \in \{1, \ldots, k\} \\ m(s, 0) &= +\infty && \forall s \geq 1 \\ m(s, i) &= +\infty && \forall i \in \{1, \ldots, k\}\ \forall s < 0 \end{aligned}$$

a) On a $m(S) = m(S, k)$.

b) Soit la relation de récurrence suivante pour tout $i \in \{1, \ldots, k\}$

$$m(s, i) = \begin{cases} 0 & \text{si } s = 0 \\ \min\{m(s, i-1), m(s - V[i], i) + 1\} & \text{sinon} \end{cases}$$

Montrons-la par récurrence.

Preuve

Soit s la capacité de confiture à traiter. Deux cas sont possibles :

i) $\underline{s = 0 :}$ Dans ce cas, il n'y a pas de confiture à préserver. On n'utilise donc aucun bocal, i.e. on a $m(0, i) = 0$. ■

ii) $\underline{s \neq 0 :}$ Puisque $m(s, i) = +\infty$ pour toute quantité $s < 0$, la relation de récurrence ne concerne dans ce cas que les quantités $s > 0$. De même, on a $m(s, 0) = +\infty$ pour toute quantité $s \geq 1$. Donc, on ne considère que les $i > 0$.

Étant donné le système de capacités $V[1], V[2], \ldots, V[i]$, deux choix sont possibles : soit on utilise des bocaux de capacité $V[i]$, soit on ne s'en sert plus.

D'un côté, le premier choix implique forcément l'utilisation d'*au moins* un bocal de ladite capacité. Ceci se traduit mathématiquement par $m(s - V[i], i) + 1$.

D'un autre côté, le deuxième choix rejete les bocaux de ladite capacité. Donc, le résultat est donné par $m(s, i - 1)$.

Comme on cherche à minimiser le nombre de bocaux à utiliser, $m(s, i)$ doit valoir le minimum entre ces deux résultats. ■

Question 4

Soit M un tableau doublement indicé tel que la case $M[s, i]$, où $s \in \{0, \ldots, S\}$ et $i \in \{0, \ldots, k\}$, contient la valeur $m(s, i)$.

a) Soit le couple $(s, i) \in \{1, \ldots, S\} \times \{1, \ldots, k\}$. D'après la relation de récurrence prouvée ci-dessus, pour calculer $m(s, i)$, on a besoin des valeurs stockées dans les cases $(s, i-1)$ et $(s-V[i], i)$, si $s-V[i] \geq 0$. Autrement dit, la case à gauche et toutes les cases au-dessus de (s, i) doivent être traitées préalablement au calcul de la valeur de cette dernière.

Cela implique que le remplissage du tableau doit commencer à la case $(0, 0)$ et aller jusqu'à la case (S, k) en suivant un parcours matriciel. Ce parcours peut donc se faire soit par lignes, soit par colonnes.

N.B. Dans le cadre du projet, on a sélectionné le parcours séquentiel des lignes.

b) On déduit des deux questions précédentes un algorithme `AlgoProgDyn` qui pour un système de k capacités $V[1], V[2], \ldots, V[k]$ et une quantité S de confiture retourne $m(S)$. En voici le pseudo-code :

Algorithm 1 AlgoProgDyn

Require: $k \geq 0$ **and** $S \geq 0$

```
1:  function ConfitureForwards(k : integer, V : k-integer array, S : integer)
2:      M : (S + 1) × (k + 1) integer matrix filled with +∞
3:      for s = 0 to S do
4:          for i = 0 to k do
5:              M[s][i] ← GetMsi(V, M, s, i)
6:      return M[S][k]                                  ▷ the value of m(S)

7:  function GetMsi(V : k-integer array, M : (S + 1) × (k + 1) integer matrix,
    s : integer, i : integer)
8:      if s = 0 then
9:          return 0
10:     else if s < 0 then
11:         return +∞
12:     else if i = 0 then
13:         return +∞
14:     else if M[s][i] < +∞ then                       ▷ M[s][i] already calculated
15:         return M[s][i]
16:     else
17:         return min(GetMsi(V, M, s, i − 1), GetMsi(V, M, s − V[i], i) + 1)
```

c) Pour l'analyse de la complexité temporelle on utilise l'accès à une case de la matrice M comme instruction élémentaire. Ainsi, on remarque que pour chacune des $(S + 1) \cdot (k + 1)$ itérations de la boucle interne on accède au plus 5 fois à des éléments du tableau : l'une pour l'écriture de la case et les quatre restantes pour les lectures des cases dont on cherche le minimum des

valeurs. Par ailleurs, la création et l'initialisation de la matrice M se font en $\Theta(S \cdot k)$. Donc, la complexité temporelle de l'algorithme est en $\Theta(S \cdot k)$.

Quant à la complexité spatiale, on se sert de deux structures de données gourmandes, à savoir le tableau V de k cases d'entiers et la matrice M de $(S+1) \cdot (k+1)$ cases d'entiers. En prenant comme taille de base celle d'un entier, on réalise que la complexité spatiale de l'algorithme est aussi en $\Theta(S \cdot k)$.

Question 5

On désire à présent permettre à l'algorithme de retourner un tableau A indiquant le nombre de bocaux pris pour chaque type de capacités.

a) Dans un premier temps, on décide de placer dans chaque case du tableau $M[s, i]$, un tableau A indiquant les bocaux pris pour la capacité totale s et des capacités de bocaux prises parmi $V[1], \ldots, V[i]$.

Pour ce faire, on effectue quelques modifications à l'algorithme `AlgoProgDyn`. Tout d'abord, la matrice M devient une matrice de tableaux d'entiers d'ordre $(S+1) \times (k+1)$. De plus, la fonction GetMsi est modifiée de sorte à utiliser le tableau de bocaux utilisés pour chaque case de la matrice lorsque nécessaire. Finalement, on remplace l'appel à GetMsi par un appel à GetA à la ligne 5 de l'algorithme initial. Cette nouvelle fonction reprend la logique de la fonction GetMsi initiale, sauf que l'on renvoie le tableau A correspondant au lieu du nombre total de bocaux utilisés.

Voici le pseudo-code des modifications effectuées :

```
1: function GetMsi(V : k-integer array, M : (S + 1) × (k + 1) integer array matrix, s : integer, i : integer)
2:     A : integer array
3:     if s = 0 then
4:         return 0
5:     else if s < 0 then
6:         return +∞
7:     else if i = 0 then
8:         return +∞
9:     else
10:        A ← M[s][i]
11:        return Σ_{j=1}^{i} A[j]

12: function GetA(V : k-integer array, M : (S+1)×(k+1) integer array matrix, s : integer, i : integer)
13:    A : integer array filled with i zeros
14:    if s > 0 and i > 0 then
15:        left ← GetMsi(V, M, s, i − 1)
16:        up ← GetMsi(V, M, s − V[i], i) + 1
17:        if left < up then
18:            A[1 . . i − 1] ← M[s][i − 1]
19:        else
20:            A[1 . . i] ← M[s − V[i]][i]
21:            A[i] ← A[i] + 1
22:    return A
```

En termes de mémoire, chaque case $M[s][i]$, où $i > 0$ contient maintenant un tableau à i cases d'entier. Ainsi, on a $\frac{k\cdot(k+1)}{2}$ cases d'entier dans chaque ligne, i.e. il y a $(S+1)\cdot\frac{k\cdot(k+1)}{2}$ cases d'entier dans toute la matrice M. De ce fait, la complexité spatiale de cet algorithme est en $\Theta(S\cdot k^2)$.

b) Dans un deuxième temps, on souhaite éviter de copier dans chaque case du tableau M, les tableaux A des bocaux utilisés. Pour ce faire, on conserve l'algorithme initial et on reconstitue *a posteriori* le tableau A de la solution optimale. Voici le pseudo-code du deuxième algorithme :

Algorithm 2 AlgoProgDynRet

Require: $k \geq 0$ **and** $S \geq 0$

```
1:  function ConfitureBackwards(k : integer, V : k-integer array, S : integer,
    M : (S + 1) × (k + 1) integer matrix)
2:      A : integer array filled with k zeros
3:      s, i : integers
4:      s ← S
5:      i ← k
6:      while s ≠ 0 do
7:          if i > 0 and M[s][i] = M[s][i − 1] then    ▷ no more V[i]-capacity jars
8:              i ← i − 1
9:          else                                        ▷ M[s][i] = M[s − V[i]][i] + 1
10:             A[i] ← A[i] + 1
11:             s ← s − V[i]
12:     return A
```

Force est de constater que l'algorithme suit *dans le pire cas* un chemin de déplacements unitaires sans detours de la case (S, k) à la case $(0, 0)$. C'est pourquoi on effectue *au plus* $S + k$ tours de boucle. On remarque aussi que le corps de la boucle comporte uniquement des instructions élémentaires. La complexité temporelle de la boucle de l'algorithme-retour est ainsi en $\mathcal{O}(S + k)$.

Par ailleurs, l'initialisation du tableau A est en $\Theta(k)$. Ce tableau A est la seule addition en termes de mémoire par rapport à l'algorithme de base et il occupe autant de cases mémoire que le tableau V. De ces faits, les compléxités temporelle et spatiale de l'algorithme-retour sont respectivement en $\mathcal{O}(S+k)$ et $\Theta(S \cdot k)$.

Question 6

On constate que, pour l'algorithme global, ce sont les complexités de l'algorithme initial qui l'emportent, d'où celles du premier algorithme sont retenues, à savoir des complexités en $\Theta(S \cdot k)$.

Cependant, il est important de remarquer que cette complexité dépend de la valeur de S et non pas de sa taille de codage, en l'occurrence $\lfloor \log_2(S) \rfloor + 1$. De ce fait, on dit que cet algorithme est pseudo-polynomial : il est exponentiel en la longueur de ses paramètres, mais polynomial en leur valeur numérique.

3 Algorithme III : Cas particulier et algorithme glouton

Question 7

Voici le pseudo-code de l'algorithme `AlgoGlouton` :

Algorithm 3 AlgoGlouton

Require: $k \geq 1$ **and** $S \geq 0$

```
1:  function MAIN(k : integer, V : k-integer array, S : integer)
2:      A : integer array filled with k zeros
3:      s, i, n : integers
4:      s ← S
5:      i ← k
6:      n ← 0
7:      while s ≠ 0 do
8:          A[i] ← s ÷ V[i]
9:          s ← s mod V[i]
10:         n ← n + A[i]
11:         i ← i − 1
12:     return (n, A)
```

L'algorithme `AlgoGlouton` consiste en un boucle comportant deux divisions euclidiennes, une addition et une soustraction. Pour le calcul de la complexité temporelle, considérons seulement las divisions euclidiennes. La complexité de la j-ième itération est donnée par

$$T(j) = \begin{cases} (\lfloor \log_2(\max(S, V[k])) \rfloor + 1)^2 & \text{si } j = 1 \\ (\lfloor \log_2(V[k-j+1]) \rfloor + 1)^2 & \text{sinon} \end{cases}$$

Dans le pire des cas, on effectue k itérations, d'où la complexité temporelle est donnée par

$$\begin{aligned} \sum_{j=1}^{k} T[j] &= (\lfloor \log_2(\max(S, V[k])) \rfloor + 1)^2 + \sum_{j=2}^{k} (\lfloor \log_2(V[k-j+1]) \rfloor + 1)^2 \\ &= (\lfloor \log_2(\max(S, V[k])) \rfloor + 1)^2 + \sum_{l=1}^{k-1} (\lfloor \log_2(V[l]) \rfloor + 1)^2 \\ &\leq \left[\lfloor \log_2(\max(S, V[k])) \rfloor + 1 + \sum_{l=1}^{k-1} (\lfloor \log_2(V[l]) \rfloor + 1) \right]^2 \\ &\leq \left[\log_2(S) + \sum_{l=1}^{k} (\log_2(V[l]) + 1) \right]^2 \end{aligned}$$

Étant donné que les bocaux sont de capacité finie, on peut bien assumer qu'il existe une constante positive A qui est une borne supérieure de toutes les tailles de codage des capacités du système V.

Il en résulte que

$$\sum_{j=1}^{k} T[j] \leq [\log_2(S) + A \cdot k]^2$$

De ce fait, la complexité temporelle de l'algorithme `AlgoGlouton` dans le pire des cas est en $\mathcal{O}((\log(S)+k)^2)$. On remarque que cet algorithme est bien polynomial en la longueur de ses entrées.

Question 8

On dit qu'un système de capacités est **glouton-compatible** si, pour ce système de capacités, l'algorithme `AlgoGlouton` produit la solution optimale quelle que soit soit la quantité totale S. Montrons qu'il existe des systèmes de capacités qui ne sont pas glouton-compatibles.

Preuve

Il suffit de donner un exemple de système de capacités V tel qu'il y a au moins une quantité totale S pour laquelle l'algorithme `AlgoGlouton` ne renvoie pas la solution optimale.

Soit le système de capacités $V = [1, 4, 5]$. Soit $S = 28$ la quantité totale de confiture. L'algorithme `AlgoGlouton` produit la solution $(8, [3, 0, 5])$, alors que la solution optimale est $(6, [0, 2, 4])$, puisque 5 bocaux ne suffisent pas. Ceci est donc un contre-exemple cherché. □

Question 9

Montrons que, pour tout entier $d \geq 2$, le système de k capacités **Expo** est glouton-compatible.

Soit $g = (g_1, \ldots, g_k)$ la solution produite par l'algorithme `AlgoGlouton` pour le système **Expo**. Supposons qu'il existe une solution optimale $o = (o_1, \ldots, o_k)$ différente de g.

a) Montrons qu'il existe un plus grand indice j pris dans $\{1, \ldots, k\}$ tel que $o_j < g_j$.

   ### Preuve

   D'après les conditions du problème, on sait que les solutions g et o sont différentes. Donc, il existe *au moins* un indice $j \in \{1, \ldots, k\}$ tel que $o_j < g_j$, car sinon la solution g n'utiliserait toute la quantité de confiture. Par ailleurs, on sait que k est fini, d'où il existe un plus grand indice $j \in \{1, \ldots, k\}$ remplissant ladite condition. □

b) Montrons que $\sum_{i=1}^{j} V[i] \cdot g_i = \sum_{i=1}^{j} V[i] \cdot o_i$

Preuve

D'aprés la question précédente, on a $o_l \geq g_l$ pour tout l dans $\{j+1, \ldots, k\}$. De plus, puisque au cours de l'algorithme glouton on choisit le nombre maximum possible de bocaux de capacité $V[l]$, on a $o_l \leq g_l$. On en déduit que $o_l = g_l$. Donc, on a

$$\sum_{i=1}^{k} V[i] \cdot g_i = \sum_{i=1}^{k} V[i] \cdot o_i \tag{14}$$

$$\sum_{i=j+1}^{k} V[i] \cdot g_i = \sum_{i=j+1}^{k} V[i] \cdot o_i \tag{15}$$

En soustrayant (15) de (14), on obtient $\sum_{i=1}^{j} V[i] \cdot g_i = \sum_{i=1}^{j} V[i] \cdot o_i$ □

c) Montrons que $\sum_{i=1}^{j-1} V[i] \cdot o_i \geq V[j]$

Preuve

On sait que

$$\sum_{i=1}^{j-1} V[i] \cdot g_i \geq 0$$

En utilisant le résultat de la question précédente, il en résulte que

$$V[j] \cdot g_j \leq \sum_{i=1}^{j} V[i] \cdot o_i$$

D'où,

$$V[j] \cdot (g_j - o_j) \leq \sum_{i=1}^{j-1} V[i] \cdot o_i$$

Or $g_j - o_j \geq 1$ d'après la question a) et le fait que ce sont des entiers naturels.

Donc, $V[j] \leq \sum_{i=1}^{j-1} V[i] \cdot o_i$ □

Cette propriété précise l'une des manières d'utiliser le moins possible de bocaux pour remplir une quantité de confiture donnée. En effet, elle consiste à remplir au complet tous les bocaux des plus grandes capacités possibles jusqu'à l'arrivée à une certaine capacité. Pour cette dernière, on préfère ne pas remplir tous les bocaux possibles, mais de se servir plutôt de ceux de capacité plus petite.

d) Supposons que $o_i \leq d-1$ pour tout $i \in \{1, \ldots, j-1\}$. D'après la question précédente, on a

$$\begin{aligned} V[j] &\leq \sum_{i=1}^{j-1} V[i] \cdot o_i \\ &\leq \sum_{i=1}^{j-1} V[i] \cdot (d-1) \\ &\leq (d-1) \cdot \sum_{i=1}^{j-1} 2^{i-1} \\ &\leq (d-1) \cdot \frac{d^{j-1}-1}{d-1} \\ d^{j-1} &\leq d^{j-1} - 1 \qquad \text{car } d-1 \geq 1 \end{aligned}$$

On arrive donc à une contradiction, i.e. il existe un indice $l \in \{1, \ldots, j-1\}$ tel que $o_l \geq d$. □

e) Posons la suite o' telle que $o'_l = o_l - d, o'_{l+1} = o_{l+1} + 1$ et $o'_i = o_i$ pour tout $i \neq l$ et $i \neq l+1$. Montrons que o' est une solution du problème.

Preuve

D'après les conditions du problème, on a

$$V[l] \cdot d = V[l+1]$$

D'où,

$$0 = -V[l] \cdot d + V[l+1]$$

Donc,

$$V[l] \cdot o_l + V[l+1] \cdot o_{l+1} = V[l] \cdot (o_l - d) + V[l+1] \cdot (o_{l+1} + 1),$$

i.e.

$$V[l] \cdot o_l + V[l+1] \cdot o_{l+1} = V[l] \cdot o'_l + V[l+1] \cdot o'_{l+1}$$

Par ailleurs, on a $o_i = o'_i$ pour tout $i \in \{1, \ldots, k\} \setminus \{l, l+1\}$.
De ces faits,

$$\sum_{i=1}^{k} V[i] \cdot o_i = \sum_{i=1}^{k} V[i] \cdot o'_i,$$

i.e. o' est aussi une solution du problème. □

On remarque que la différence entre o' et o est que la première remplace d bocaux d'une certaine capacité dans la deuxième solution par un bocal de la capacité immédiatement supérieure. Autrement dit, la solution o' utilise $d-1$ bocaux de moins que la solution o.

De ce fait, la solution o n'est pas optimale, ce qui contredit l'hypothèse de départ. On en déduit que, pour le système **Expo**, toute solution o est non optimale ou elle est égale à g. Ceci équivaut à dire que, pour le système **Expo**, pour toute solution o, si elle est optimale alors est égale à g. On en déduit que toute solution optimale est égale à g. Or il existe une solution optimale à ce problème. Donc, l'algorithme `AlgoGlouton` renvoie l'unique solution optimale, à savoir g, pour le système de capacités **Expo**. □

Question 10

Montrons que tout système de capacités V avec $k = 2$ est glouton-compatible.

Preuve

On a montré dans la question précédente que le système **Expo** est glouton-compatible.

Soit $V = [1, p]$, où $p > 1$, un système de capacités quelconque. On remarque que ce système est un système **Expo** avec $k = 2$ et $d = p \geq 2$.

De ces faits, tout système de capacité V de longueur $k = 2$ est bien glouton-compatible. □

Question 11

Montrons que la complexité de l'algorithme `TestGloutonCompatible`(k, V) est en $\mathcal{O}(k^3)$, i.e. qu'il est polynomial.

Preuve

Le nombre d'itérations de la boucle externe est

$$\text{nbiter\_ext} = V[k-1] + V[k] - 1 - (V[3] + 2) + 1$$

Or

$$V[3] + 2 \geq 5$$
$$V[k-1] + V[k] - 1 \leq 2 \cdot V[k] - 2$$

De ces faits,

$$\text{nbiter\_ext} \leq 2 \cdot V[k] - 6,$$

i.e. le nombre d'itérations de la boucle externe est en $\Theta(V[k])$.

Par ailleurs, on effectue le même nombre d'itérations de la boucle interne pour chaque tour de la boucle externe, en l'occurrence k. De plus, chaque tour de la boucle interne comporte un premier test de vérité entre deux nombres obtenus en $\Theta(1)$, en l'occurrence $V[j]$ et S, et un deuxième test sur deux nombres obtenus chacun au moyen d'un appel à `AlgoGlouton`. D'après la question 7, sa complexitè est en $\mathcal{O}((\log(S) + k)^2)$. Or

$$S \leq 2 \cdot V[k] - 2,$$

i.e. $\log(S)$ est de l'ordre de A, à savoir la borne supérieure des tailles des capacités de bocaux. Donc, chaque tour de la boucle interne est fait en $\mathcal{O}(k^2)$.

De ces faits, la complexitè de l'algorithme `TestGloutonCompatible` est en $\mathcal{O}(k^3)$, car $V[k]$ est borné par 2^A. □

Deuxième partie

Mise en œuvre

On a implémenté les trois algorithmes en le langage de programmation Python. Dans les deux sections suivantes, on vérifie expérimentalement les complexités temporelles théoriques déterminées dans les sections précédentes et on essaie aussi de répondre à la question sur la mise en œuvre en pratique de l'algorithme glouton.

4 Analyse de complexité expérimentale

Pour cette section, on a considéré le système de capacités **Expo** pour des valeurs $d = 2, 3$ et 4. Pour chaque système, on a fait varier les valeurs des paramètres S et k. Bien évidemment, leur ordre de grandeur n'est pas le même pour les trois tests, comme vous pouvez vous en apercevoir ci-dessous.

Question 12

Pour comprendre les images données juste après, il faut préciser que les temps trouvés correspondent au temps CPU et non pas au temps d'exécution, et que l'on s'est arrêtés dès que le temps de calcul mésuré dépasse une minute.

D'après la première section, la complexité de l'algorithme `RechercheExhaustive` est exponentielle en la quantité de confiture. On remarque dans la figure 1 que le temps CPU de cet algorithme explose très rapidement. En effet, le temps CPU passe d'être très proche de 0 pour $S = 30$ à une valeur de quelques unités pour $S = 40$. C'est d'autant plus que pour $S = 50$ cela se trouve dans l'ordre des dizaines.

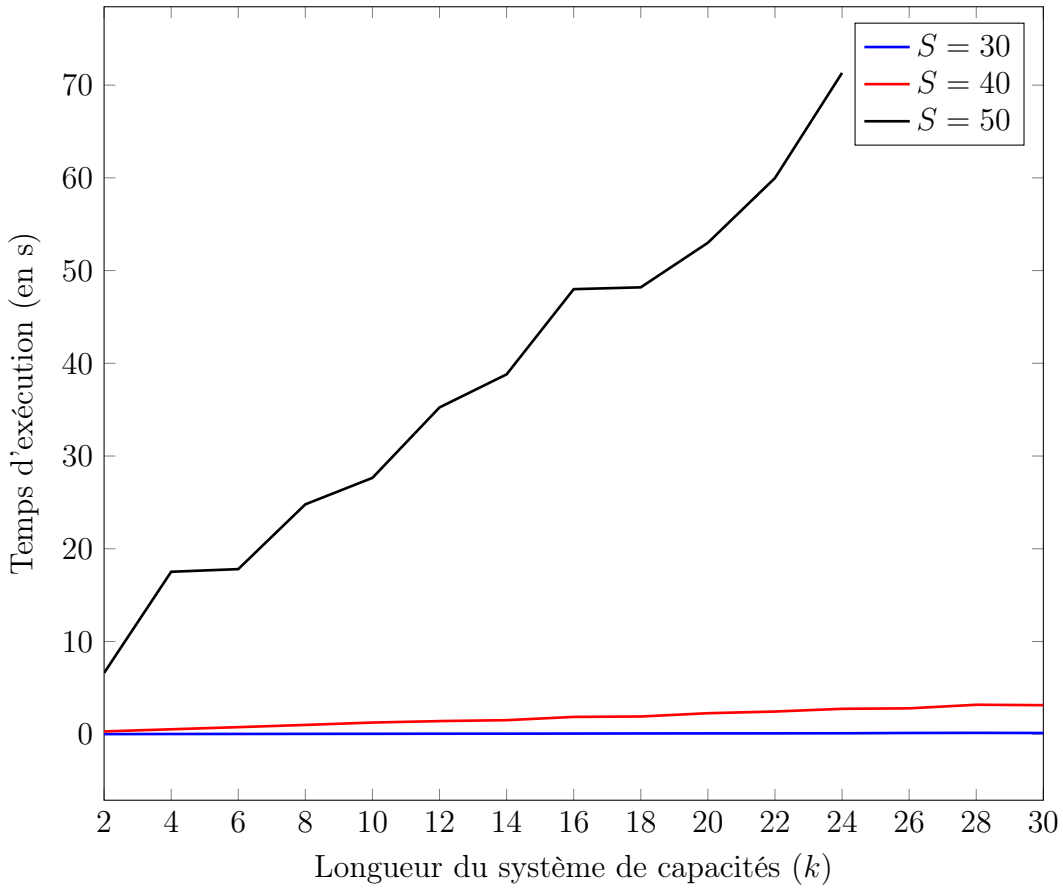

FIGURE 1 – Analyse expérimentale de la complexité de l'algorithme `RechercheExhaustive` pour un système **Expo**, avec $d = 4$, dont la taille est variable

Dans la deuxième section, on avait trouvé que la complexité de l'algorithme `AlgoProgDyn` dépend aussi bien de la quantité de confiture que de la taille du système de capacités. On a aussi en déduit qu'elle est pseudo-polynomiale. Dans la figure 2, on peut apprécier cette augmentation moins prononcée de l'algorithme en question. On le voit notamment si on trace une droite verticale. Le taux de variation est linéaire en la valeur numérique de S, ce qui montre bien le caractère pseudo-polynomial de l'algorithme par programmation dynamique.

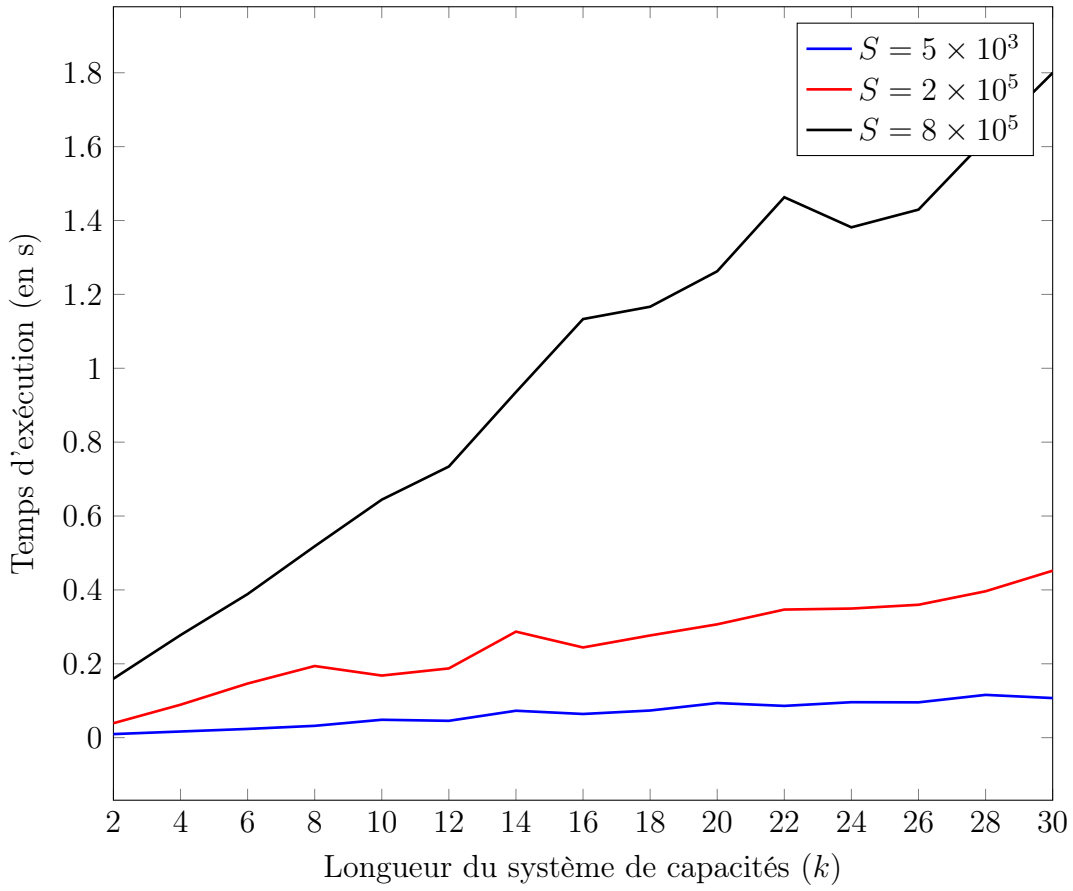

FIGURE 2 – Analyse expérimentale de la complexité de l'algorithme `AlgoProgDyn` pour un système **Expo**, avec $d = 3$, dont la taille est variable

Dans la section précédente, on a trouvé que la complexité de l'algorithme `AlgoGlouton` dépandait en grande partie de la taille du système de capacités, et, d'une façon moins importante, de la quantité de confiture à traiter. Dans le jeu de tests apparaissant dans la figure 3, on observe que l'algorithme est très faiblement susceptible aux variations dans S. Par ailleurs, pour une courbe donnée, on s'aperçoit que la variation n'est pas linéaire en la taille k, mais plutôt quadratique. Ceci est bien conforme à l'analyse théorique de la complexité de cet algorithme.

Il est important de remarquer que les courbes comportent à chaque fois des valeurs de S de plus en plus grandes, mais que le temps CPU diminue en même temps. Cela veut dire que, pour les systèmes de capacités **Expo**, on peut définir la relation d'ordre suivante selon l'efficacité croissante des algorithmes :

$$\texttt{RechercheExhaustive} \prec \texttt{AlgoProgDyn} \prec \texttt{AlgoGlouton}$$

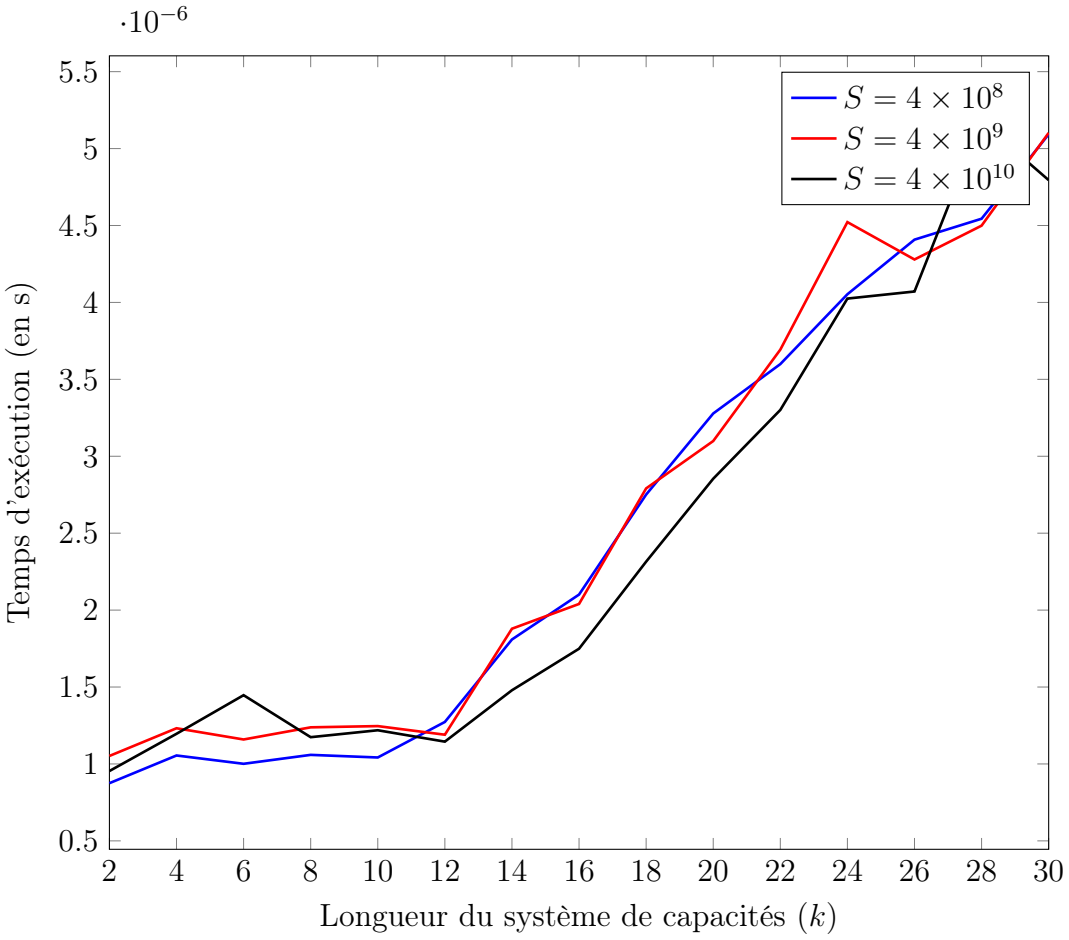

FIGURE 3 – Analyse expérimentale de la complexité de l'algorithme `AlgoGlouton` pour un système **Expo**, avec $d = 2$, dont la taille est variable

5 Utilisation de l'algorithme glouton

On s'intéresse maintenant à vérifier la faisabilité pratique de l'algorithme glouton. Deux points sont importants à ce sujet : la fréquence d'apparition des systèmes de capacités glouton-compatibles, et ses performances dans les cas contraire à l'égard de la solution optimale.

Pour ce faire, on a généré des systèmes de capacités de manière aléatoire. Les capacités ont été tirées aléatoirement dans $[\![2..100]\!]$. De plus, on a fait varier la taille k des systèmes de 3 à 30. Finalement, pour chacune de ces tailles, on a obtenu 20 systèmes distincts.

Question 13

Pour chacun des systèmes obtenus, on a testé s'il était ou pas glouton-compatible ; e cas échéant, on l'a pris en compte pour la détermination de la proportion de systèmes glouton-compatibles.

Cette proportion a été calculé de façon cumulée pour chacun des sous-intervalles engendrés. Vous retrouverez ces résultats dans la figure 4 ci-dessous.

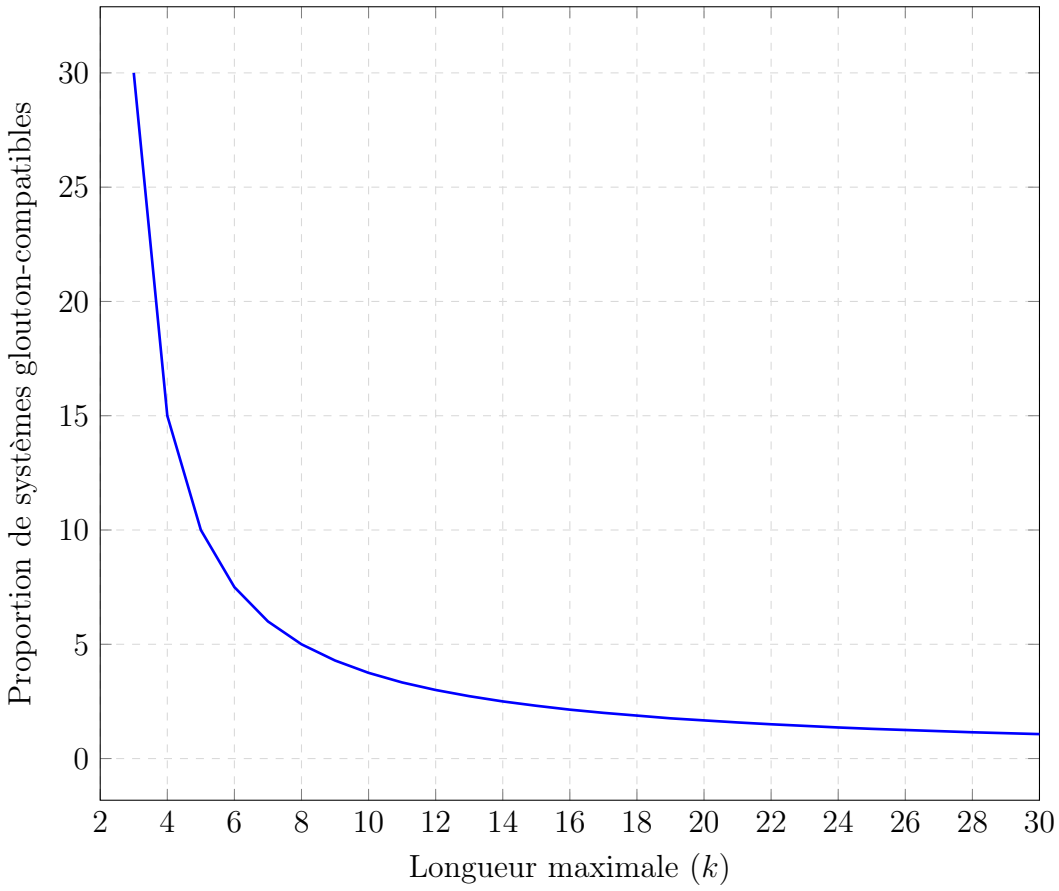

FIGURE 4 – Variation de la proportion de systèmes glouton-compatibles en fonction de la borne supérieure de l'intervalle de longueurs considéré

On déduit de cette courbe que les systèmes de capacités glouton-compatibles sont plus rares au fur et a mesure que l'on considère des tailles plus grandes. En effet, pour l'intervalle $[\![3..30]\!]$, on compte seulement 1,07 % de systèmes glouton-compatibles. Ceci veut dire qu'il y en a très peu qui sont non **Expo**, pour lesquels l'algorithme glouton est optimal et très efficace.

Question 14

En revanche, pour la plupart des systèmes générés, qui n'étaient pas glouton-compatibles, on a cherché à mesurer les performances de l'algorithme glouton. Ainsi, on a lancé les algorithmes `AlgoGlouton` (en question) et `AlgoProgDyn` (le meilleur) sur des diverses entrées et fait un résumé des résultats obtenus. Concrètement, on a considéré une quantité de confiture allant de 100 jusqu'à 1000 pour chaque système non glouton-compatible.

On a compilé nos résultats dans le tableau 1. On a mis en rouge le nombre de bocaux parce que c'est la caractéristique pour laquelle l'algorithme glouton est le moins performant ; le temps CPU étant son point fort, on l'a mis en bleu.

Comme mentionné précédemment, l'algorithme glouton est incontestablement le plus réactif. On peut s'en convaincre notamment avec la deuxième mesure du

Mesure	Écart			
	Moyen		Maximum	
	valeur	en %	valeur	en %
Nombre de bocaux	3.167	51.2	77	3850
Temps CPU (en s)	6.704×10^{-3}	1.716×10^{5}	4.218×10^{-2}	1.6×10^{6}

TABLE 1 – Statistiques sur les performances des algorithmes glouton et par programmation dynamique

tableau ci-dessus. De plus, il semblerait que l'algorithme glouton n'est pas du tout mauvais si on ne regardait que les écarts moyen et maximum du nombre de bocaux en valeur.

Cependant, il faut rappeler que la quantité maximale de confiture testée a été 1000, i.e. elle n'a pas été très élévée. De ce fait on justifie que les écarts en pourcentage montrent la faible efficacité de cet algorithme.

A Suites croissantes

On rappelle la définition des suites $b(s)$ et $c(s)$ comme suit :

$$b(s) = \begin{cases} 0 & \text{si } s = 0 \\ 2 & \text{si } s = 1 \\ 2 \cdot b(s-2) + 2 & \text{si } s \geq 2 \end{cases} \text{ et } c(s) = \begin{cases} 0 & \text{si } s = 0 \\ 2 \cdot c(s-1) + 2 & \text{si } s \geq 1 \end{cases}$$

Montrons que les deux suites sont croissantes.

Suite $b(s)$

Soit la propriété suivante $Q_1(s), s \geq 0 : b(s) \leq b(s+1)$. Montrons-la par récurrence d'ordre 2 sur s

Preuve

Base : On a $b(0) = 0 \leq 2 = b(1) \leq 2 \cdot 0 + 2 = b(2)$. Donc, la propriété est vérifiée aux rangs 0 et 1.

Induction : Supposons $Q_1(s-2)$ et $Q_1(s-1)$ vraies pour un $s \geq 2$ fixé. Montrons $Q_1(s)$ vraie.

Par hypothèse de récurrence,

$$b(s-2) \leq b(s-1)$$

D'où,

$$2 \cdot b(s-2) + 2 \leq 2 \cdot b(s-1) + 2 \tag{16}$$

Or, par définition,

$$b(s+1) = 2 \cdot b(s-1) + 2 \tag{17}$$

En remplaçant (17) en (16), on obtient $b(s) \leq b(s+1)$, i.e. la propriété est vérifiée au rang s.

Conclusion :

$$\left.\begin{array}{l} Q_1(0) \text{ et } Q_1(1) \text{ vraies} \\ \forall s \geq 2, [(Q_1(s-2) \text{ et } Q_1(s-1)) \Rightarrow Q_1(s)] \end{array}\right\} \forall s \in \mathbb{N}, b(s) \leq b(s+1)$$

On remarque que ceci correspond à la définition d'une suite croissante. □

Suite $c(s)$

Soit la propriété suivante $Q_2(s), s \geq 0 : c(s) \geq 0$. Montrons-la par récurrence faible sur s.

Preuve

Base : Pour $s = 0$, on a $c(0) = 0 \geq 0$. Donc, la propriété est vérifié aux rang 0.

Induction : Supposons $Q_2(s)$ vraie pour un $s \geq 0$ fixé. Montrons $Q_2(s+1)$ vraie.

Par définition,

$$c(s+1) = 2 \cdot c(s) + 2 \tag{18}$$

Or, par hypothèse de récurrence,

$$c(s) \geq 0$$

D'où,

$$2 \cdot c(s) + 2 \geq 2 \tag{19}$$

En remplaçant (18) en (19), il résulte que $c(s+1) \geq 2$. On en déduit par transitivité que la propriété est vérifiée au rang $s+1$.

Conclusion :

$$\left.\begin{array}{l} Q_2(0) \text{ vraie} \\ \forall s \geq 0, [Q_2(s) \Rightarrow Q_2(s+1)] \end{array}\right\} \ \forall s \in \mathbb{N}, c(s) \geq 0$$

Par ailleurs, on a

$$c(s) = 2 \cdot c(s-1) + 2, s \geq 1$$

D'où,

$$c(s) \geq 2 \cdot c(s-1) \tag{20}$$

Comme cette suite est positive, on sait que

$$2 \cdot c(s-1) \geq c(s-1) \tag{21}$$

Donc, par transitivité entre (20) et (21), $c(s) \geq c(s-1)$.
Autrement dit, pour tout $s \geq 0$, $c(s+1) \geq c(s)$. □

B Complexité de `RechercheExhaustive` dans le cas général

Pour le calcul de la complexité temporelle de l'algorithme `RechercheExhaustive` dans le cas général, considérons de nouveau seulement les appels récursifs effectués par `RechercheExhaustive`$(k, V[1, \ldots, k], s)$. La complexité pour ce système de capacités est donnée par une suite $d(s) \leq k \cdot 2^s$.

On utilise un raisonnement par récurrence forte. Pour fonder la base, on montre que

$$d(0) = 0$$
$$d(s) \leq k \cdot 2^{s-1}, \forall s \in \{1, \ldots, k+1\}$$

Puis, dans la phase d'induction, on retrouvera un terme linéaire en k qui est soustrait de l'expression exponentielle donnée ci-dessus.

La complexité est alors en $\mathcal{O}(k \cdot 2^s)$.