La Poste sous l'austérité

Épreuve pratique d'algorithmique et de programmation
Concours commun des Écoles normales supérieures

Durée de l'épreuve: 3 heures 30 minutes

Juin/Juillet 2017

ATTENTION !

N'oubliez en aucun cas de recopier votre u_0
à l'emplacement prévu sur votre fiche réponse

Important.

Il vous a été donné un numéro u_0 qui servira d'entrée à vos programmes. Les réponses attendues sont généralement courtes et doivent être données sur la fiche réponse fournie à la fin du sujet. À la fin du sujet, vous trouverez en fait deux fiches réponses. La première est un exemple des réponses attendues pour un $\widetilde{u_0}$ particulier (précisé sur cette même fiche et que nous notons avec un tilde pour éviter toute confusion!). Cette fiche est destinée à vous aider à vérifier le résultat de vos programmes en les testant avec $\widetilde{u_0}$ au lieu de u_0. Vous indiquerez vos réponses (correspondant à votre u_0) sur la seconde et vous la remettrez à l'examinateur à la fin de l'épreuve.

En ce qui concerne la partie orale de l'examen, lorsque la description d'un algorithme est demandée, vous devez présenter son fonctionnement de façon schématique, courte et précise. Vous ne devez en aucun cas recopier le code de vos procédures!
Quand on demande la complexité en temps ou en mémoire d'un algorithme en fonction d'un paramètre n, on demande l'ordre de grandeur en fonction du paramètre, par exemple: $O(n^2)$, $O(n \log n)$,...
Il est recommandé de commencer par lancer vos programmes sur de petites valeurs des paramètres et de ***tester vos programmes sur des petits exemples que vous aurez résolus préalablement à la main ou bien à l'aide de la fiche réponse type fournie en annexe***. Enfin, il est recommandé de lire l'intégralité du sujet avant de commencer afin d'effectuer les bons choix de structures de données dès le début.

Dans ce sujet, on s'intéresse à une ville qui n'a pas encore de bureau de poste. On voudrait ouvrir un ou plusieurs bureaux de poste tout en essayant de minimiser le coût mais aussi la distance de chaque maison au bureau le plus proche.
Dans la première section, on construit les instances du problème. On considère ensuite deux types de solutions à notre problème. La section 2 étudie les solutions exactes. La section 3 est indépéndante et considère le calcul de solutions approximatives.

1 Villes et bureaux de poste

Les maisons de notre ville sont chacune identifiées à un point. Par ailleurs, toutes les rues sont parallèles à l'axe des abscisses ou à l'axe des ordonnées. Ainsi, la distance entre deux points (x_1, y_1) et (x_2, y_2) est donnée par la norme 1 : on a $\|(x_1, y_1) - (x_2, y_2)\|_1 = |x_1 - x_2| + |y_1 - y_2|$.

Pour deux entiers naturels x, y, on rappelle que $x \bmod y$ est le reste de la division euclidienne de x par y. Pour un nombre réel x, on note par $\lfloor x \rfloor$ sa partie entière.

On considère les suites suivantes. Soit u_0 indiqué sur votre feuille (à reporter sur votre fiche réponse).
On définit

$$\begin{aligned}
&\forall t \geq 1, u_t = a \cdot u_{t-1} \bmod b,\\
&\forall t \geq 0, u_{t,0} = u_t,\\
&\forall t \geq 0, i \geq 1, u_{t,i} = c \cdot u_{t,i-1} \bmod d,\\
&\forall t \geq 0, i \geq 0, u_{t,i,0} = u_{t,i},\\
&\forall t, i \geq 0, j \geq 1, u_{t,i,j} = e \cdot u_{t,i,j-1} \bmod f,
\end{aligned}$$

On prendra $a = 263, b = 269, c = 271, d = 281, e = 283, f = 293$.

Question 1 *Calculer* ***a)*** u_2 ***b)*** u_{1000} ***c)*** $u_{1000,4}$ ***d)*** $u_{1000,4,2001}$.

Pour tout $t \geq 1$ et $n \geq 1$ donnés, on définit un ensemble de n maisons comme suit. On note par (x_i, y_i) les coordonnées de la i-ème maison, pour $1 \leq i \leq n$, où $x_i = u_{t,1,2i} - \lfloor f/2 \rfloor$, et $y_i = u_{t,1,2i+1} - \lfloor f/2 \rfloor$.

Question 2 *Donner le nombre de maisons qui sont à une distance inférieure ou égale à* 200 *(pour la norme* 1*) de l'origine* $(0, 0)$ *dans les ensembles de* n *maisons correspondant aux paramètres suivants.* ***a)*** $t = 1, n = 10$ ***b)*** $t = 2, n = 100$ ***c)*** $t = 3, n = 1000$

On voudrait ouvrir un certain nombre de bureaux de poste dans ces villes. On se donne m emplacements possibles, aux coordonnées (X_j, Y_j) pour $1 \leq j \leq m$, qui peuvent accueillir un bureau de poste, moyennant un coût c_j. Pour $t \geq 1$ et $m \geq 1$ donnés, on définit, pour tout $j \in [1, m]$,

$$\begin{aligned}
X_j &= u_{t,2,2j} - \lfloor f/2 \rfloor,\\
Y_j &= u_{t,2,2j+1} - \lfloor f/2 \rfloor,\\
c_j &= 10 \cdot u_{t,3,j}.
\end{aligned}$$

Ainsi, si on ouvre deux bureaux de poste, un à l'emplacement (X_1, Y_1), et un autre à l'emplacement (X_2, Y_2), cela nous coûte $c_1 + c_2$.
On note que certaines maisons (ou emplacements) peuvent avoir des coordonnées identiques. On les considérera comme des maisons (ou emplacements) distinctes.
Le rectangle englobant d'un ensemble de points est le plus petit rectangle fermé dont un côté est horizontal et qui contient tous les points de cet ensemble.

Question 3 *Calculer la surface du rectangle englobant des ensembles de m emplacements correspondant aux paramètres suivants.* ***a)*** $t = 1, m = 10$ ***b)*** $t = 2, m = 100$ ***c)*** $t = 3, m = 1000$

On note par $G(t, m, n)$ la ville avec n maisons définies plus haut et les m emplacements possibles pour les bureaux de poste et leur coût que l'on vient de définir.
Par exemple, pour $u_0 = 1$, la ville $G(1, 2, 4)$ est constitué de 4 maisons et de 2 emplacements avec $(x_1, y_1) = (-19, 49), (x_2, y_2) = (-45, 16), (x_3, y_3) = (-8, -61), (x_4, y_4) = (-117, -143)$, et $(X_1, Y_1) = (146, -136), c_1 = 1330$, $(X_2, Y_2) = (47, -25), c_2 = 1350$.

Notons $[i, j] = \{i, i + 1, \dots, j\}$. En plus du coût de l'ouverture des bureaux de poste, on voudrait considérer la distance que doit parcourir chaque habitant pour se rendre au bureau de poste le plus proche. Ainsi, étant donné $G(t, m, n)$, pour chaque maison $i \in [1, n]$ et chaque emplacement $j \in [1, m]$, notons par $d_{i,j}$ la distance entre (x_i, y_i) et (X_j, Y_j) (pour la norme 1).
Étant donnée une ville $G(t, m, n)$ qui a besoin d'au moins un bureau de poste, notre objectif est de sélectionner un sous-ensemble non-vide $I \subseteq [1, m]$ d'emplacements pour ouvrir des bureaux de poste, afin de minimiser le coût total défini comme suit :

$$\mathsf{coût}(G(t, m, n), I) = \sum_{j \in I} c_j + \sum_{i \in [1,n]} \min_{j \in I} d_{i,j}.$$

Ainsi, le coût total est obtenu en sommant le coût total d'ouverture des bureaux de poste aux emplacements I, et la distance de chaque maison au bureau le plus proche parmi I.

On propose d'étudier deux algorithmes pour ce problème dans les deux sous-sections suivantes. **Ces deux algorithmes sont indépendants**.

2 Solution exacte

On commence par la solution naïve : on voudrait énumérer tous les sous-ensembles $I \subseteq [1, m]$, et pour chaque I, évaluer son coût, et en choisir un qui minimise ce coût.
On définit le codage d'un ensemble $I \subseteq [1, m]$ comme une suite $\mathcal{B}_m(I) = b_{m-1} \dots b_1 b_0$ de m bits telle que $i \in I$ si et seulement si $b_{i-1} = 1$. On interprète la suite $\mathcal{B}_m(I)$ comme l'écriture en binaire d'un entier $B_m(I) \in \mathbb{N}$.

Question à développer pendant l'oral 1 *Décrire l'ensemble* $\{B_m(I) \mid I \subseteq [1, m]\}$.

Pour un entier naturel $x \in \mathbb{N}$ sur m bits, soit $I_m(x) \subseteq [1, m]$ le sous-ensemble tel que $B_m(I_m(x)) = x$.

Question 4 *On fixe* $m = 10$. *Pour chacune des valeurs de x suivantes, calculer* $\max I_m(x)$. *On notera* $\max \emptyset = -\infty$. ***a)*** $x = u_{10} \bmod 1024$ ***b)*** $x = u_{11} \bmod 1024$ ***c)*** $x = u_{12} \bmod 1024$

Question 5 *Calculer le coût optimal, c'est-à-dire* $\min_{I \subseteq [1,m]} \mathsf{coût}(G(t, m, n), I)$ *pour les instances suivantes.* ***a)*** $t = 1, m = 10, n = 20$ ***b)*** $t = 2, m = 11, n = 30$ ***c)*** $t = 3, m = 12, n = 100$ ***d)*** $t = 4, m = 15, n = 1000$

Question à développer pendant l'oral 2 *Quelle est la complexité de votre algorithme ? Donner une idée de l'ordre de grandeur de n, m pour lesquels on peut espérer la terminaison de cet algorithme en temps raisonnable ?*

Supposons qu'on ait retenu les emplacements $I \subseteq [1, m]$ avec $I \cap [k, m] = \emptyset$ pour un certain $k \in [1, m]$. On souhaite calculer rapidement une estimation du meilleur coût qu'on pourrait obtenir en étendant I avec un sous-ensemble de $[k, m]$. On propose le minorant suivant à cette quantité :

$$\mathsf{coût}_{\mathrm{est}}(G(t, m, n), I, k) = \sum_{j \in I} c_j + \sum_{i \in [1,n]} \min_{j \in I \cup [k,m]} d_{i,j}.$$

Question à développer pendant l'oral 3 *Montrer que pour toute ville $G(t, m, n)$, tout ensemble $I \subseteq [1, k-1]$ et J qui vérifie $I \subseteq J \subseteq I \cup [k, m]$, on a $\mathsf{coût}_{est}(G(t, m, n), I, k) \leq \mathsf{coût}(G(t, m, n), J)$. Décrire un algorithme pour calculer le coût optimal en se servant de la quantité $\mathsf{coût}_{est}(G(t, m, n), I, k)$.*

Question 6 *Calculer le coût optimal, c'est-à-dire $\min_{I \subseteq [1,m]} \mathsf{coût}(G(t, m, n), I)$ pour les instances suivantes.* ***a)*** $t = 1, m = 30, n = 100$ ***b)*** $t = 2, m = 32, n = 100$ ***c)*** $t = 3, m = 34, n = 100$ ***d)*** $t = 4, m = 36, n = 100$

3 Solution approximative

Pour un emplacement $j \in [1, m]$ et un ensemble $C' \subseteq [1, n]$ non-vide de maisons, on définit le coût unitaire de (j, C') comme suit :

$$\mathsf{coût}_{\mathrm{unit}}(j, C') = \frac{c_j + \sum_{i \in C'} d_{i,j}}{|C'|}.$$

L'idée du coût unitaire est de mesurer le coût d'ouvrir un bureau à l'emplacement j et de connecter les maisons de C' à ce bureau, rapporté au nombre de maisons concernées. Autrement dit, cette quantité nous dit combien cette opération nous coûte *par maison connectée*.

Définissons un ordre sur les sous-ensembles de $[1, n]$. Pour $A \subseteq [1, n]$ soit $\mathsf{mot}(A)$ le mot obtenu en concaténant les éléments de A, vus comme des lettres, dans l'ordre croissant. Par exemple, on a $\mathsf{mot}(\{3, 6, 1, 5\}) = 1 \cdot 3 \cdot 5 \cdot 6$. Soit $\leq_{\mathrm{lex}}$ l'ordre lexicographique sur les mots. Pour deux ensembles $A, B \subseteq [1, n]$, on définit $A \prec B$ si et seulement si

- soit $|A| > |B|$,
- soit $|A| = |B|$ et $\mathsf{mot}(A) \leq_{\mathrm{lex}} \mathsf{mot}(B)$.

On a, par exemple, $\{3, 6, 1, 5\} \prec \{2, 3, 4, 5\}$ et $\{2, 3, 4, 5, 6\} \prec \{3, 6, 1, 5\}$. Ainsi, étant donné un ensemble de sous-ensembles de $[1, n]$, le minimal pour l'ordre $\prec$ consiste à choisir, parmi ceux qui ont le cardinal le plus élevé, l'ensemble dont le mot est minimal pour l'ordre lexicographique.

On propose **l'algorithme d'approximation** suivant pour choisir un ensemble I. L'idée intuitive est la suivante. À chaque itération, on choisit un emplacement j et un ensemble de maisons C' tels que (j, C') a le coût unitaire minimal. On ouvre alors un bureau à l'emplacement j et on considère que toutes les maisons de C' sont connectées à ce bureau. On recommence jusqu'à ce que toutes les maisons soient connectées à un bureau de poste, ou que tous emplacements soient occupés.

Formellement, soit C l'ensemble des maisons qui restent à connecter à un bureau de poste, et B l'ensemble des emplacements libres. On a, initialement, $C := [1, n]$ et $B := [1, m]$. On répète l'opération suivante tant que $C \neq \emptyset$ et $B \neq \emptyset$:

1. Soit
$$\alpha = \min\{\mathsf{coût}_{\text{unit}}(j, C') \mid j \in B, \emptyset \neq C' \subseteq C\}.$$
   Soit $j_0 \in B$ le plus petit indice tel que $\alpha = \min_{C' \subseteq C} \mathsf{coût}_{\text{unit}}(j_0, C')$, et C_0 minimal pour l'ordre $\prec$ tel que $\alpha = \mathsf{coût}_{\text{unit}}(j_0, C_0)$.
2. On met à jour : $B := B \setminus \{j_0\}$, et $C := C \setminus C_0$.

L'algorithme renvoie alors $[1, m] \setminus B$ comme ensemble d'emplacements choisis.

Question à développer pendant l'oral 4 *Quel est le nombre de paires (j, C') qu'il faut considérer a priori dans chaque itération ?*

Lemme 1 *Soit (j, C') avec $j \in [1, m]$ et $C' \subseteq [1, n]$. S'il existe $i \in [1, n] \setminus C'$ et $k \in C'$ tels que $d_{i,j} < d_{k,j}$, alors il existe $C'' \subseteq [1, n]$ qui satisfait $\mathsf{coût}_{unit}(j, C'') < \mathsf{coût}_{unit}(j, C')$.*

Question à développer pendant l'oral 5 *Démontrer le lemme précédent. Décrire un algorithme de complexité polynomiale pour trouver la paire qui minimise le coût unitaire, et donner sa complexité.*

Question 7 *Donner la paire (j_0, C_0) calculée à la première itération de l'algorithme d'approximation sur les instances suivantes.* ***a)*** $t = 1, m = 5, n = 15$ ***b)*** $t = 2, m = 10, n = 20$ ***c)*** $t = 3, m = 50, n = 20$.

Question 8 *Pour les instances suivantes, appliquer l'algorithme décrit plus haut pour trouver les futurs bureaux de poste $I \subseteq [1, m]$. Donner le coût $\mathsf{coût}(G(t, m, n), I)$ où I est la solution calculée par l'algorithme d'approximation.* ***a)*** $t = 1, m = 5, n = 15$ ***b)*** $t = 2, m = 50, n = 200$ ***c)*** $t = 3, m = 500, n = 500$ ***d)*** $t = 4, m = 1000, n = 1000$.

Malheureusement, cet algorithme ne calcule pas toujours la solution optimale. On voudrait comprendre s'il donne néanmoins une solution approchée dans le sens suivant. On dit que $\eta \geq 1$ est un rapport d'approximation si pour toutes les villes G, en notant par $I_{\mathsf{alg}}(G)$ la solution calculée par l'algorithme d'approximation, et par $I_{\mathsf{opt}}(G)$ la solution optimale sur G, on a
$$\frac{\mathsf{coût}(G, I_{\mathsf{alg}}(G))}{\mathsf{coût}(G, I_{\mathsf{opt}}(G))} \leq \eta. \tag{1}$$

Autrement dit, η permet de borner uniformément l'erreur commise par l'algorithme d'approximation par rapport à la solution optimale.

On notera que si un tel rapport η existe, alors l'inégalité ci-dessus est vérifiée pour toute ville. Pour certaines villes particulières, il se peut bien sûr que l'algorithme d'approximation calcule une solution meilleure, voire optimale.

Pour comprendre la qualité des approximations calculées par notre algorithme, on considère la ville G suivante. Soient deux emplacements aux coordonnées $(0, 0)$, $(120, 0)$ avec les coûts d'ouverture de 100 et 200 respectivement. De plus, on a une maison en $(0, 10)$ et une autre en $(120, 10)$.

Question à développer pendant l'oral 6 *Calculer la solution optimale, et la solution calculée par l'algorithme d'approximation pour G. Que peut-on en déduire sur un rapport d'approximation η pour notre algorithme, si celui-ci existe ? Existe-t-il un rapport d'approximation pour cet algorithme ?*

Pour une instance donnée, notons la solution optimale calculée dans la section précédente par I_{opt}. Pour rappel, on a

$$\mathsf{coût}(G(t,m,n), I_{\mathsf{opt}}) = \min_{I \subseteq [1,m]} \mathsf{coût}(G(t,m,n), I).$$

Question 9 *Quel est le nombre d'instances où* $\mathsf{coût}(G(t,m,n), I_{\mathsf{alg}}) \leq 1.20 \cdot \mathsf{coût}(G(t,m,n), I_{\mathsf{opt}})$ *dans les ensembles d'instances suivants :* ***a)*** $t \in [1,10]$, $m = 10, n = 50$ ***b)*** $t \in [11,1000]$, $m = 10, n = 50$ ***c)*** $t \in [500,1000]$, $m = 10, n = 100$

Question à développer pendant l'oral 7 *Que peut-on faire pour diminuer le coût des solutions calculées par l'algorithme en pratique tout en gardant une complexité raisonnable ? Donner quelques idées informelles.*

3.1 Routes

On considère maintenant le cas où certaines maisons et emplacements ne sont pas reliés entre eux par une route directe mais seulement par une route indirecte qui passe par d'autres maisons ou emplacements. Autrement dit, le graphe dont les sommets sont les maisons et les emplacements n'est pas un graphe complet.

Considérons l'instance $G(t,m,n)$ du problème défini dans les sections précédentes. On définit le graphe $H(t,m,n)$ pondéré sur l'ensemble de sommets $V = \{\mathcal{M}_1, \ldots, \mathcal{M}_n, \mathcal{E}_{n+1}, \ldots, \mathcal{E}_{n+m}\}$ qui sont l'ensemble des maisons et des emplacements. On notera que certains sommets de ce graphe (maisons ou emplacements) peuvent avoir les mêmes coordonnées. Les arêtes E sont définies comme suit. Pour tout $i, j \in \{1, 2, \ldots, n\}$, on a

— Pour tout $1 \leq i < j \leq n$, $\{\mathcal{M}_i, \mathcal{M}_j\} \in E$ si et seulement si $u_{t,4,(ia+j \bmod 1001)} \leq \lfloor f/15 \rfloor$.

— Pour tout $1 \leq k < l \leq m$, $\{\mathcal{E}_k, \mathcal{E}_l\} \in E$ si et seulement si $u_{t,4,(kb+l \bmod 1001)} \leq \lfloor f/15 \rfloor$.

— Pour tout $i \in [1,n]$ et $k \in [1,m]$, $\{\mathcal{M}_i, \mathcal{E}_k\} \in E$ si et seulement si $u_{t,4,(ic+k \bmod 1001)} \leq \lfloor f/15 \rfloor$.

Comme dans les sections précédentes, le poids d'une arête est défini comme la distance pour la norme 1 entre les coordonnées de deux points (maisons ou emplacements).

Notons par $d'_{i,j}$ la longueur du plus court chemin de la maison i à l'emplacement j dans $H(t,m,n)$. En effet, la façon la moins coûteuse d'aller de la maison i à l'emplacement j est de suivre un tel plus court chemin, à défaut d'avoir une route directe (i,j).

Question 10 *Pour chaque instance* $H(t,m,n)$*, donner la quantité suivante :*

$$\min\{d'_{i,j} \mid \{\mathcal{M}_i, \mathcal{E}_j\} \notin E\},$$

où E *est l'ensemble des arêtes du graphe* $H(t,m,n)$*. (Le minimum d'un ensemble vide est* ∞*).* ***a)*** $t = 1, m = 5, n = 15$ ***b)*** $t = 2, m = 50, n = 200$ ***c)*** $t = 3, m = 500, n = 500$ ***d)*** $t = 4, m = 1000, n = 1000$.

Ainsi, pour calculer le coût d'ouvrir un bureau de poste à chaque emplacement d'un ensemble I pour $H(t,m,n)$, on définit

$$\mathsf{coût}(H(t,m,n), I) = \sum_{j \in I} c_j + \sum_{i \in [1,n]} \min_{j \in I} d'_{i,j}.$$

Question 11 *Pour les instances suivantes, appliquer l'algorithme d'approximation décrit plus haut. Donner le coût* coût$(H(t,m,n), I)$ *où* I *est la solution calculée par cet algorithme. Marquer* ∞ *s'il n'y a pas de solution.* ***a)*** $t = 1, m = 5, n = 15$ ***b)*** $t = 2, m = 50, n = 200$ ***c)*** $t = 3, m = 500, n = 500$ ***d)*** $t = 4, m = 750, n = 750$.

Question à développer pendant l'oral 8 *Décrire et justifier votre algorithme. Quelle est la complexité de cet algorithme ?*

Fiche réponse <u>type</u>: La Poste sous l'austérité

$\widetilde{u_0}$: 1

Question 1

a) 36

b) 226

c) 198

d) 170

Question 2

a) 5

b) 77

c) 836

Question 3

a) 60562

b) 82944

c) 80089

Question 4

a) 7

b) 5

c) 8

Question 5

a) 3775

b) 4595

c) 10231

d) 64593

Question 6

a) 9160

b) 8863

c) 9214

d) 9292

Question 7

a) 4, [1 2 7 9 10 15]

b) 5, [1 3 6 11 14 18]

c) 29, [8]

Question 8

a) 3729

b) 15278

c) 29572

d) 46193

Question 9

a) 3

b) 416

c) 184

Question 10

a) 261

b) 11

c) 11

d) 11

Question 11

a) ∞

b) 41700

c) 56699

d) 68334

Fiche réponse: La Poste sous l'austérité

Nom, prénom, u_0: ....

Question 1

a)

b)

c)

d)

Question 2

a)

b)

c)

Question 3

a)

b)

c)

Question 4

a)

b)

c)

Question 5

a)

b)

c)

d)

Question 6

a)

b)

c)

d)

Question 7

a)

b)

c)

Question 8

a)

b)

c)

d)

Question 9

a)

b)

c)

Question 10

a)

b)

c)

d)

Question 11

a)

b)

c)

d)