Contamination de cellules

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

Durée de l'épreuve: 3 heures 30 minutes

Juin/Juillet 2014

ATTENTION !

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

Important.

Sur votre table est indiqué 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 veut simuler la propagation de populations de bactéries. Notre univers (« boîte de Pétri ») sera un tableau bi-dimentionnel de cellules, de taille $n \times n$, sur lequel on placera k bactéries à des positions différentes (les foyers). Ces bactéries contaminerons toutes les cellules du tableau selon une règle simple : chaque cellule sera contaminée par la bactérie du foyer le plus proche.

1 Définitions

Une cellule est une paire $(x, y) \in \mathbb{N}^2$. Pour $n \in \mathbb{N}^*$, $C_n = [0, 1, \ldots n-1] \times [0, 1, \ldots n-1]$ sera l'ensemble des cellules de notre univers. Deux cellules (x, y) et (x', y') sont voisines si $|x - x'| + |y - y'| = 1$. À l'instant initial, k bactéries (identifiées par les entiers de 0 à $k-1$) sont placées sur des cellules de C_n. Deux bactéries différentes ne peuvent être placées sur la même cellule. Ces cellules seront les foyers ; formellement un foyer est une paire $((x, y), i)$, où i est le numéro de la bactérie et (x, y) une cellule sur laquelle la bactérie i est placée.
Chaque cellule d'un C_n va être contaminée par une bactérie d'un foyer. La règle de contamination est la suivante : chaque cellule (x, y) est contaminée par la bactérie du foyer le plus proche (selon une certaine distance choisie). Si plusieurs foyers sont à même distance d'une cellule, c'est la bactérie de numéro le plus petit qui contaminera la cellule parmi les foyers à distance minimum. (On ne considérera pas, dans le début du sujet, de notion de temps : on peut donc supposer que toutes les contaminations se font instantanément.)
La population de la bactérie i est l'ensemble des cellules contaminées par la bactérie i au cours du processus. L'ensemble des populations de bactéries formera donc une partition de C_n.
La Figure 1 donne un exemple de contamination (avec la distance de Manhattan, c'est à dire la distance associée à la norme 1). Les foyers sont représentés par des cercles, et les numéros correspondent aux numéros des bactéries.

2 Génération d'un ensemble de foyers

2.1 Générateur pseudo aléatoire

Soit $v_0 = u_0$, et pour tout $i \in \mathbb{N}$:

$$u_{i+1} = 16365 \cdot u_i \bmod 65521,$$

et

$$v_{i+1} = 16379 \cdot v_i \bmod 65519.$$

N'oubliez pas de reporter votre u_0 sur votre fiche réponse.

Question 1 *Calculez :*
a) (u_{100}, v_{100}) **b)** $(u_{1\,000}, v_{1\,000})$ **c)** $(u_{10\,000}, v_{10\,000})$.

Remarque : Calculer les termes de ces suites peut être coûteux, il peut donc être pertinent de les stocker plutôt que de les recalculer à chaque nouvelle utilisation.

9	9	9	9	9	9	9	9	9	9	9	1	1	1	1	1	1	1	1	1	1	1	1	1	5	5	5	5	5	5
9	9	9	9	9	9	9	9	9	9	9	1	1	1	1	1	1	1	1	1	1	1	1	1	5	5	5	5	5	5
9	9	9	9	9	9	9	9	9	9	9	9	1	1	1	1	1	1	1	1	1	1	1	1	5	5	5	5	5	5
9	9	9	9	9	9	9	9	9	9	9	2	2	2	2	2	2	2	2	1	1	1	1	1	5	5	5	5	5	5
4	9	9	9	9	9	9	9	9	9	2	2	2	2	2	2	2	2	2	2	1	1	1	1	5	5	5	5	5	5
4	4	9	9	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	5	5	5	5	5
4	4	4	4	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	5	5	5	5
4	4	4	4	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	5	5	5	5
4	4	4	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	5	5	5	5
4	4	3	3	3	3	3	3	3	3	3	3	0	2	2	2	2	2	2	2	2	2	2	2	2	2	5	5	5	5
4	3	3	3	3	3	3	3	3	3	3	3	0	0	2	2	2	2	2	2	2	2	2	2	2	2	5	5	5	5
4	3	3	3	3	3	3	3	3	3	3	0	0	0	0	2	2	2	2	2	2	2	2	2	2	2	5	5	5	5
4	3	3	3	3	3	3	3	3	3	0	0	0	0	0	0	2	2	2	2	2	2	2	2	2	2	5	5	7	7
4	3	3	3	3	3	3	3	3	0	0	0	0	0	0	0	0	2	2	2	2	2	2	2	2	7	7	7	7	7
4	3	3	3	3	3	3	3	0	0	0	0	0	0	0	0	0	0	2	2	2	2	2	2	7	7	7	7	7	7
4	3	3	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	7	7	7	7	7
4	3	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	7	7	7	7	7
4	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	7	7	7	7	7
4	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	7	7	7	7	7	7
4	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	7	7	7	7	7	7	7
4	8	8	8	8	8	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	0	0	0	0	0	0	0	0	0	0	0	0	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	6	6	0	0	0	0	0	0	0	0	0	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	6	6	6	0	0	0	0	0	0	0	7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	6	6	6	6	0	0	0	0	0	0	7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	6	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7

FIGURE 1 – Les populations de $\mathcal{P}_{30,10,1}$ (c'est à dire, $n = 30$, $k = 10$ et la distance est d_1). Les cercles représentent les foyers, et les lignes grises les frontières entre les populations. (L'origine est en bas à gauche.)

2.2 Génération de l'ensemble $\mathcal{F}_{n,k}$

On générera ici $\mathcal{F}_{n,k}$, qui sera notre ensemble de k foyers sur C_n.
Soit $w_{n,i} = (u_i \bmod n, v_i \bmod n)$. Notez que $w_{n,i}$ représente une cellule de C_n. On va placer nos k bactéries, dans l'ordre de 0 à $k-1$, sur la cellule $w_{n,i}$ où i est le plus petit entier naturel tel que $w_{n,i}$ ne soit pas déjà occupée par une autre bactérie. Formellement :

$$\mathcal{F}_{n,k} = \{(w_{n,\pi_n(0)}, 0), \ldots (w_{n,\pi_n(k-1)}, k-1)\}$$

où $\pi_n : \mathbb{N} \to \mathbb{N}$ tel que pour tout i, $\pi_n(i)$ soit le plus petit entier naturel tel que $w_{n,\pi_n(i)} \neq w_{n,\pi_n(j)}$ pour tout $0 \leq j < i$.

Question 2 *Calculez :*
a) $\pi_{20}(100)$ ***b)*** $\pi_{100}(1\,000)$ ***c)*** $\pi_{1\,000}(10\,000)$.

3 Calcul des cellules infectées

Soit d_r, $r \in \mathbb{N}^* \cup \{\infty\}$, la distance associée à la norme r. On travaillera avec les normes 1, 2 et ∞. On rappelle donc (en dimension 2) :

$$||(x,y)||_1 = |x| + |y|$$

$$||(x,y)||_2 = \sqrt{x^2+y^2}$$

$$||(x,y)||_\infty = \max(|x|,|y|).$$

Question 3 *Donnez la distance maximale entre 2 foyers de l'ensemble $\mathcal{F}_{30,10}$ pour la norme :*
a) d_1 **b)** d_2 **c)** d_∞.

Soit $\mathcal{P}_{n,k,r} = \{P_1, \ldots P_k\}$ l'ensemble des populations, où P_i est la population de la bactérie i, à la fin du processus pour la carte C_n, l'ensemble de foyers $\mathcal{F}_{n,k}$ et la norme r. La Figure 1 montre les populations à la fin du processus, dans le cas de la norme 1, pour $n = 30$ et $k = 10$. Implémentez un algorithme calculant l'ensemble des populations de bactéries à la fin du processus.

Question 4 *Donnez la taille maximum d'une population dans :*
a) $\mathcal{P}_{10,10,1}$ **b)** $\mathcal{P}_{100,100,2}$ **c)** $\mathcal{P}_{500,500,\infty}$.

Question à développer pendant l'oral : Quelle est la complexité de votre algorithme, au pire des cas, en fonction de n et k.

4 Cas de la norme 1 sur de grandes instances

À partir de maintenant, on ne considérera que la norme 1. On notera pour simplifier $\mathcal{P}_{n,k}$ à la place de $\mathcal{P}_{n,k,1}$. On veut un algorithme capable de calculer en un temps raisonnable les populations quand n et k sont grands, dans le cas de la distance d_1.
Pour simplifier la compréhension de l'algorithme, on considérera dans cette section une notion de temps.
Soit le processus suivant :
- À l'instant 0, les foyers sont infectés par la bactérie correspondante.
- À l'instant $t \in \mathbb{N}^*$, pour chaque bactérie i de 0 à $k-1$ (dans cet ordre), chaque cellule infectée à l'instant $t-1$ par la bactérie i infecte toutes ses cellules voisines non encore infectées.

Au bout d'un temps fini, toutes les cellules de C_n sont infectées.
Question à développer pendant l'oral : Expliquez pourquoi ce processus calcule les populations de bactéries dans le cas de la norme 1.

Implémentez un algorithme simulant le processus décrit précédemment.

Question 5 *Donnez la taille maximum d'une population pour les ensembles :*
a) $\mathcal{P}_{500,1\,000}$ **b)** $\mathcal{P}_{1\,000,10\,000}$ **c)** $\mathcal{P}_{1\,000,100\,000}$

Question à développer pendant l'oral : Donnez la complexité en espace et en temps de votre implémentation de l'algorithme, en fonction de n et k.
Question à développer pendant l'oral : Comment modifier l'algorithme pour calculer les populations avec la norme ∞ ? Et pour la norme 2 ?

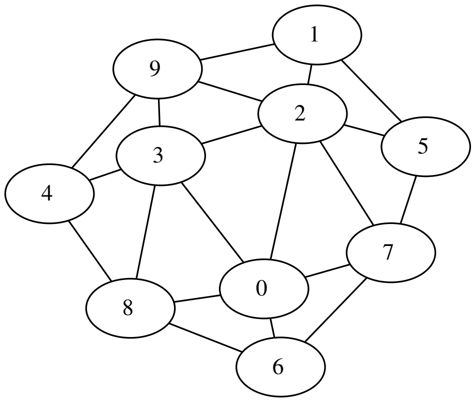

FIGURE 2 – Le graphe correspondant à $\mathcal{P}_{30,10}$.

5 Coloriage des populations

Dans cette section également, on ne considérera que la norme 1. On veut disposer d'un outil pour visualiser les populations de bactéries. On veut donc colorier chaque population avec une couleur telle que deux populations voisines ont deux couleurs différentes. On cherche également à ce que le nombre de couleurs que l'on doit utiliser ne soit pas trop grand.

On rappelle qu'un graphe (non orienté) est une paire (V, E), où V est l'ensemble des sommets, et E, l'ensemble des arêtes, est un ensemble de sous ensembles de taille 2 de V. Deux sommets u et v sont adjacents si $\{u, v\} \in E$. Le voisinage d'un sommet u, dénoté $N(u)$, est l'ensemble des sommets adjacents avec v. Le degré d'un sommet u est la taille du voisinage de u.
Deux populations A et B sont voisines si $A \neq B$ et s'il existe $(x, y) \in A$ et $(x', y') \in B$ tels que les cellules (x, y) et (x', y') sont voisines. Le graphe des populations associé à $\mathcal{P}_{n,k}$ est le graphe d'ensemble de sommets $\{1, 2, \dots k\}$, et tel que deux sommets i et i' sont adjacents si les populations P_i et $P_{i'}$ sont voisines (où $\mathcal{P}_{n,k} = (P_1, \dots P_k)$).
La Figure 2 représente par exemple le graphe des populations de $\mathcal{P}_{30,10}$.

5.1 Calcul du graphe des populations

Implémentez un algorithme calculant le graphe des populations de $\mathcal{P}_{n,k}$.

Question 6 *Donnez le nombre d'arêtes et le degré moyen des graphes des populations correspondants à :*
a) $\mathcal{P}_{10,30}$ ***b)*** $\mathcal{P}_{100,100}$ ***c)*** $\mathcal{P}_{500,1\,000}$
d) $\mathcal{P}_{100,10\,000}$ ***e)*** $\mathcal{P}_{2\,000,100\,000}$

Question à développer pendant l'oral : Quelles structures de données avez vous utilisées pour stocker votre graphe ? Donnez la complexité en espace et en temps de votre implémentation, en fonction de n et k.

5.2 Calcul d'une coloration valide

Une coloration d'un graphe $G = (V, E)$ est une fonction $c : V \to \mathbb{N}$, et une coloration est valide si pour toute arête $\{u, v\} \in E$, on a $c(u) \neq c(v)$.
Soit un graphe $G = (V, E)$ et un sommet $v \in V$. On notera par $G - v$ le graphe (V', E') où $V' = V \setminus \{v\}$ et $E' = \{\{x, y\} : \{x, y\} \in E$ et $v \notin \{x, y\}\}$ (c'est à dire $G - v$ est le graphe G privé du sommet v et des arêtes adjacentes v).
Implémentez l'algorithme suivant pour colorier les graphes (on suppose que l'ensemble des sommets du graphe est un ensemble d'entiers, comme c'est le cas pour les graphes de populations) :

Entrée : un graphe $G = (V, E)$.
Sortie : une coloration de G.

1. Choisissez v dans V de plus petit degré. (En cas d'égalité, choisissez le plus petit v parmi les sommets de plus petit degré.)
2. Calculez récursivement une coloration sur $G' = G - v$.
3. Pour chaque sommet u dans G', assignez à u dans G la même couleur que dans G'.
4. Assignez à v dans G la plus petit entier de $\mathbb{N}$ non assigné aux voisins de v dans G'.

Question à développer pendant l'oral : Expliquez pourquoi l'algorithme donne une coloration valide.

Question 7 *Donnez le nombre de sommets de chaque couleur de la coloration calculée par l'algorithme précédent, pour :*
a) $\mathcal{P}_{20,50}$ ***b)*** $\mathcal{P}_{100,100}$ ***c)*** $\mathcal{P}_{1\,000,1\,000}$
d) $\mathcal{P}_{2\,000,20\,000}$ ***e)*** $\mathcal{P}_{5\,000,500\,000}$ ***f)*** $\mathcal{P}_{20\,000,2\,000\,000}$

Attention, il s'agit de compter le nombre de populations de chaque couleur, et non pas le nombre de cellules de chaque couleur. Vous devrez peut être adapter vos algorithmes en tenant compte de la mémoire disponible pour faire passer les dernières instances.
Question à développer pendant l'oral : Donnez la complexité en espace et en temps de votre implantation de l'algorithme, en fonction du nombre de sommets du graphe. Quelles structures de données avez vous utilisées pour obtenir ce temps d'exécution ?
Question à développer pendant l'oral : Compte tenu des résultats obtenus aux questions 6 et 7, que pouvez vous conjecturer sur le nombre de couleurs utilisé par l'algorithme pour colorier le graphe ? Est-on loin d'une coloration optimale ? Expliquez pourquoi.

Fiche réponse type: Contamination de cellules

$\widetilde{u_0}$: 42

Question 1

a) $(49632, 41175)$

b) $(33168, 30138)$

c) $(41199, 46570)$

Question 2

a) 117

b) 1041

c) 10048

Question 3

a) 47

b) $\sqrt{1117} = 33.42$

c) 28

Question 4

a) 20

b) 246

c) 1752

Question 5

a) 773

b) 419

c) 48

Question 6

a) 66, 4.40

b) 265, 5.30

c) 2852, 5.70

d) 19800, 3.96

e) 288827, 5.78

Question 7

a) $(15, 14, 13, 8)$

b) $(29, 28, 27, 15, 1)$

c) $(288, 265, 237, 193, 17)$

d) $(5527, 5285, 4910, 3823, 455)$

e) $(138920, 133408, 123341, 94344, 9987)$

f) $(547122, 527386, 489736, 383917, 51839)$

Fiche réponse: Contamination de cellules

Nom, prénom, u_0: ....

Question 1

a)

b)

c)

Question 2

a)

b)

c)

Question 3

a)

b)

c)

Question 4

a)

b)

c)

Question 5

a)

b)

c)

Question 6

a)

b)

c)

d)

e)

Question 7

a)

b)

c)

d)

e)

f)