La ruée vers l'or

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

Durée de l'épreuve: 3 heures 30 minutes

Juin/Juillet 2010

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.

1 Introduction

Nous sommes en 1849, dans l'ouest américain, et de nombreux filons d'or viennent d'être découverts à Sutter's Mill, qui attirent une foule d'aventuriers cherchant à faire fortune. Afin d'éviter au maximum les rivalités et les affrontements, les autorités décident de contrôler l'attribution des terrains aux nouveaux venus[1]. Pour assurer un partage équitable des ressources, elles font procéder à des sondages pour évaluer la capacité aurifère du terrain et répartir les terres équitablement.

Par la suite, on modélise un terrain rectangulaire de taille $n \times m$ par une fonction T à deux variables, où $T(x, y)$ donne la quantité d'or estimée aux coordonnées (x, y), pour $1 \leqslant x \leqslant n$ et $1 \leqslant y \leqslant m$. On va chercher à découper ce terrain en p parcelles rectangulaires, de telle sorte que les parcelles soient disjointes et recouvrent la totalité du terrain. Une parcelle rectangulaire peut être définie par quatre coordonnées $(x_{\min}, x_{\max}, y_{\min}, y_{\max})$: la parcelle est alors l'ensemble des (x, y) tels que $x_{\min} \leqslant x \leqslant x_{\max}$ et $y_{\min} \leqslant y \leqslant y_{\max}$ La quantité (estimée) d'or dans une telle parcelle rectangulaire est appelée <u>valeur</u> de la parcelle, et est donnée par

$$\sum_{x_{\min} \leqslant x \leqslant x_{\max}} \ \sum_{y_{\min} \leqslant y \leqslant y_{\max}} T(x, y).$$

Pour obtenir une distribution équitable, on choisit de limiter la jalousie des prospecteurs en s'assurant qu'aucune parcelle n'a beaucoup plus d'or que les autres. Plus formellement, l'objectif est de trouver le découpage (ou partitionnement) du terrain en parcelles, de telle sorte que la valeur maximale des parcelles soit la plus petite possible. On appelle valeur d'un partitionnement la valeur maximale de toutes les parcelles obtenues.

2 Génération aléatoire de terrains

On note $M = 64\,007$. Considérons la suite d'entiers (u_k) définie pour $k \geqslant 0$ par :

$$u_k = \begin{cases} \textit{votre } \mathtt{u}_0 \ \textit{(à reporter sur votre fiche)} & \text{si } k = 0 \\ 15\,091 \times u_{k-1} \mod M & \text{si } k \geqslant 1 \end{cases}$$

Question 1 *Que valent :* ***a)*** u_{10} ***b)*** u_{100} ***c)*** $u_{1\,000}$

Définition 1 *Pour $n \in \mathbb{N}^*$, et $m \in \mathbb{N}^*$ on note $T_{n,m}$ le terrain de largeur n et de hauteur m, et dont la valeur pour l'abscisse x ($1 \leqslant x \leqslant n$) et l'ordonnée y ($1 \leqslant y \leqslant m$) est donnée par :*

$$T_{n,m}(x, y) = 1 + \left\lfloor 400 \times \frac{u_{x \times m + y}}{64007} \right\rfloor$$

Question 2 *Donner* ***a)*** $T_{10,10}(3, 5)$, ***b)*** $T_{15,20}(3, 5)$ *et* ***c)*** $T_{10,1}(5, 1)$.

[1] En réalité, la législation était quasi inexistante dans cette région mexicaine récemment occupée par l'armée américaine. La suite du sujet est donc purement fictive.

3 Partitionnement dans le cas mono-dimensionnel

> *NB : La taille des instances sur lesquelles les algorithmes doivent être testés est choisie de telle sorte qu'un algorithme de la complexité attendue donne un résultat en un temps raisonnable (au plus quelques dizaines de secondes).*

On s'intéresse dans cette partie au cas à une dimension, correspondant à une bande de terre. Pour simplifier les notations, on pose $L_n = T_{n,1}$ et $L_n(x) = T_{n,1}(x, 1)$. Dans cette partie, on étudiera les terrains L_{10}, L_{20}, L_{100}, $L_{1\,000}$, $L_{10\,000}$, et $L_{100\,000}$.
Pour une bande de terre L, on note $L[x_1; x_2]$ la parcelle du terrain L comprise entre x_1 et x_2, c'est-à-dire avec x tel que $x_1 \leqslant x \leqslant x_2$. La valeur d'une parcelle $L[x_1; x_2]$ est donnée par :

$$val(L[x_1; x_2]) = \sum_{x_1 \leqslant x \leqslant x_2} L(x)$$

Pour simplifier les algorithmes suivants, pour tout terrain L, on commence par calculer la valeur de toutes les parcelles de la forme $L[1; x]$, avec les sommes partielles suivantes :

$$S(x) = val(L[1; x]) = \sum_{i=1}^{x} L_n(i)$$

Question 3 *Donner les sommes partielles suivantes :* ***a)*** $S_{100}(50)$, ***b)*** $S_{1\,000}(500)$ *et* ***c)*** $S_{10\,000}(5000)$.

Pour les terrains étudiés, on s'assurera de pré-calculer et de stocker les valeurs des sommes partielles.

Question à développer pendant l'oral : Expliquer comment calculer la valeur d'une parcelle $L_n[x_1; x_2]$ à partir des $S_n(x)$ en temps constant.

Question 4 *On veut découper un terrain en deux parcelles. On rappelle que l'objectif est que la valeur maximale des parcelles soit minimale. Écrire un algorithme permettant de calculer la valeur optimale du partitionnement en deux parcelles pour les terrains suivants :*
a) L_{10}, ***b)*** L_{100}, ***c)*** $L_{1\,000}$, ***d)*** $L_{10\,000}$.

On commence par étudier une stratégie de partitionnement simple. On suppose que le nombre p de parcelles est une puissance de 2 ($p = 2^q$). On commence par découper le terrain en deux parties de telle sorte que la valeur maximale soit minimale. Dans le cas où deux découpages obtiendraient exactement la même valeur maximale, on choisit de donner la plus petite valeur à la parcelle de gauche. Ensuite, on continue en découpant de la sorte chacun des morceaux en $p/2$ parcelles.

Question 5 *Donner la valeur maximale d'une parcelle obtenue avec cette stratégie de partitionnement pour les terrains et nombres de parcelles suivants :*
a) L_{10}, $p = 4$, ***b)*** L_{100}, $p = 4$, ***c)*** $L_{1\,000}$, $p = 64$, ***d)*** $L_{10\,000}$, $p = 64$.

3.1 Recherche du partitionnement optimal

On cherche maintenant à calculer le partitionnement optimal d'un terrain en p parcelles, c'est-à-dire tel que la valeur maximale d'une parcelle est minimale. Pour un terrain donné L décrit à l'aide de ses sommes partielles S, on note V_x^p la valeur minimale d'un partitionnement du terrain $L[1;x]$ en p parcelles. On peut exprimer V_x^p de la façon suivante :

$$V_x^p = \begin{cases} S(x) & \text{si } p = 1 \\ \min_{p-1 \leqslant i \leqslant x-1} \max\left(V_i^{p-1}, val(L[i+1;x])\right) & \text{sinon} \end{cases} \tag{1}$$

Question à développer pendant l'oral : Justifier cette formule.

Question 6 *En déduire un algorithme pour calculer la valeur minimale du partitionnement mono-dimensionnel, et le tester sur les terrains et nombres de parcelles suivants :*
a) $L_{10},\ p = 4,$ ***b)*** $L_{100},\ p = 10,$ ***c)*** $L_{1000},\ p = 3.$

Question à développer pendant l'oral :
- Donner la complexité de l'algorithme précédent, en temps de calcul et en espace de stockage nécessaire. Peut-on réduire la taille de l'espace de stockage ?
- Comment obtenir les coordonnées des parcelles du partitionnement optimal à partir de sa valeur ?

On cherche à améliorer la complexité de l'algorithme de la question précédente, afin de le rendre linéaire en n. On s'appuie sur deux observations :
- Notons $i_p^{(x)}$ la valeur de i qui atteint le minimum dans l'équation 1, et qui définit donc la coordonnée gauche ($x_{\min}$) de la dernière parcelle du partitionnement correspondant à V_x^p. On remarque que $i_p^{(x+1)} \geqslant i_p^{(x)}$.
- Dans l'équation 1, le terme V_i^{p-1} est une fonction croissante en i, et le terme $val(L[i+1;x])$ est une fonction décroissante en i. Le minimum est atteint lorsque ces deux fonctions se croisent (juste avant ou juste après) : soit i^* le plus petit indice tel que $V_i^{p-1} \geqslant val(L[i+1;x])$, alors $i_p^{(x)} = i^*$ ou bien $i_p^{(x)} = i^* - 1$. Pour faire un choix déterministe, lorsque les deux solutions conduisent à un résultat identique, on choisira $i^* - 1$.

Question à développer pendant l'oral : Justifier la première observation.

Question 7 *À l'aide de ces observations, modifier l'algorithme précédent pour que sa complexité soit linéaire en n. Donner la valeur du partitionnement pour les terrains et nombres de parcelles suivants :*
a) $L_{1\,000},\ p = 20,$ ***b)*** $L_{10\,000},\ p = 20,$ ***c)*** $L_{10\,000},\ p = 5,$ ***d)*** $L_{100\,000},\ p = 5.$

3.2 Recherche améliorée du partitionnement optimal

On veut obtenir des algorithmes encore plus efficaces pour le cas mono-dimensionnel, permettant de gérer des tableaux de plus grande taille et des nombres de parcelles plus grands. Pour ceci, on commence par étudier la faisabilité d'un partitionnement avec un objectif donné.

Question 8 *Écrire un algorithme Test1D(L, p, B) efficace qui répond à la question suivante : peut-on ou non découper un terrain mono-dimensionnel L en p parcelles de telle sorte que la valeur de chaque parcelle soit inférieure ou égale à une borne B ? Le tester sur les valeurs suivantes :*
a) $L_{10\,000}, p = 80, B = 25\,050$,
b) $L_{100\,000}, p = 80, B = 250\,500$,
c) $L_{1\,000\,000}, p = 80, B = 2\,505\,000$.

Question à développer pendant l'oral : Pour un terrain L donné, on appelle B_1 la valeur minimale de la première parcelle telle que $Test1D(B_1) =$ vrai, et $L[0; x_1]$ la parcelle correspondante. Montrer que dans un partitionnement optimal, la première parcelle est soit $L[0; x_1]$, soit $L[0; x_1 - 1]$.

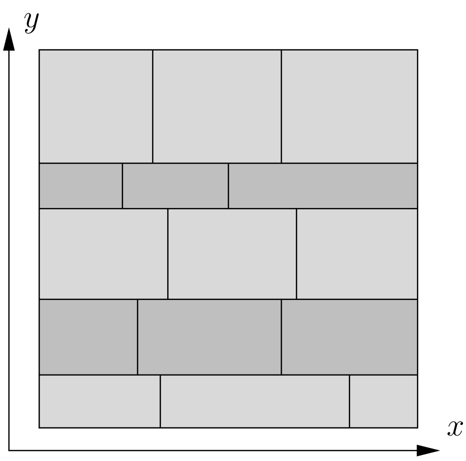
Question 9 *En s'appuyant sur l'observation précédente, et en la généralisant, donner un algorithme efficace pour calculer le partitionnement optimal utilisant l'algorithme Test1D. Le tester sur les terrains suivants, avec $p = 80$:* ***a)*** $L_{10\,000}$, ***b)*** $L_{100\,000}$ *et* ***c)*** $L_{1\,000\,000}$.

Question à développer pendant l'oral : Quelle est la complexité de cet algorithme ?

4 Partitionnement dans le cas bi-dimensionnel

On étudie maintenant le cas général des terrains à deux dimensions. Dans cette partie, on aura besoin d'un algorithme pour partitionner de façon optimale un terrain à une dimension. On pourra par exemple utiliser l'algorithme de la question 6 pour trouver la valeur optimale, suivi de celui de la question 8 pour construire le partitionnement. Dans cette partie, on utilisera les terrains $T_{10,15}$, $T_{15,20}$, et $T_{20,30}$.

On s'intéresse ici à une méthode de partitionnement à deux étapes, illustrée ci-contre : on découpe d'abord le terrain en p bandes horizontales de façon optimale, (de telle sorte que la valeur maximale d'une bande soit minimale), puis on applique un algorithme de partitionnement optimal mono-dimensionnel sur chacune des bandes pour les partager en q parcelles.



Partitionnement en
$p = 5$ bandes de $q = 3$ blocs.

Question 10 *Donner la valeur du partitionnement par bandes obtenu avec les terrains et valeurs de p et q suivants :*
a) $T_{10,15}$, $p = 3$, $q = 4$,
b) $T_{15,20}$, $p = 4$, $q = 6$,
c) $T_{20,30}$, $p = 5$, $q = 8$.

On cherche maintenant à obtenir le partitionnement optimal en p bandes de q parcelles. Pour un terrain donné, on s'intéresse à la bande horizontale contenant les éléments du terrain dont l'ordonnée y est telle que $y_1 \leqslant y \leqslant y_2$. On note $Part1D(y_1, y_2, k)$ la valeur du partitionnement mono-dimensionnel optimal de cette bande en k parcelles.

Question à développer pendant l'oral : Donner une expression récursive pour $B(y, p)$, le partitionnement optimal en p bandes de q parcelles des y premières lignes du terrain. On pourra utiliser la fonction *Part1D*.

Question 11 *En déduire un algorithme optimal pour calculer la valeur du partitionnement par bandes optimal pour les terrains et valeurs de p et q suivants :*
a) $T_{10,15}$, $p = 3$, $q = 4$,
b) $T_{15,20}$, $p = 4$, $q = 6$,
c) $T_{20,30}$, $p = 5$, $q = 8$.

Question à développer pendant l'oral :
- Donner la complexité de cet algorithme.
- Pour d'autres algorithmes de partitionnement bi-dimensionnel, on veut pouvoir calculer la valeur d'une parcelle rectangulaire en temps constant. Expliquer comment étendre le concept de sommes partielles vu à la partie précédente pour l'utiliser avec des parcelles rectangulaires.

$\diamondsuit$
$\diamondsuit \quad \diamondsuit$
$\diamondsuit$

Fiche réponse type: La ruée vers l'or

$\widetilde{u_0}$: 1

Question 1

a) 11035

b) 45405

c) 22387

Question 2

a) 220

b) 345

c) 182

Question 3

a) 9546

b) 101464

c) 1013638

Question 4

a) 974

b) 10139

c) 100480

d) 1010998

Question 5

a) 585

b) 5101

c) 3345

d) 31783

Question 6

a) 515

b) 2103

c) 67162

Question 7

a) 10146

b) 101197

c) 404450

d) 4004468

Question 8

a) non

b) oui

c) non

Question 9

a) 25394

b) 250400

c) 2506135

Question 10

a) 3065

b) 3386

c) 4079

Question 11

a) 3065

b) 3344

c) 3985

Fiche réponse: La ruée vers l'or

Nom, prénom, u_0: ....

Question 1

a)

b)

c)

Question 2

a)

b)

c)

Question 3

a)

b)

c)

Question 4

a)

b)

c)

d)

Question 5

a)

b)

c)

d)

Question 6

a)

b)

c)

Question 7

a)

b)

c)

d)

Question 8

a)

b)

c)

Question 9

a)

b)

c)

Question 10

a)

b)

c)

Question 11

a)

b)

c)