MATHÉMATIQUES DE L'INFORMATIQUE

L'épreuve traite du problème de l'empilement de "paquets" dans des "boîtes" de capacité donnée c. Ce problème d'optimisation combinatoire est aussi connu sous le nom de *Bin Packing*. Il est en général NP-complet mais diverses heuristiques permettent de trouver des solutions approchées satisfaisantes pour de nombreuses applications.

DÉFINITIONS: Dans la suite du problème, N est un entier ≥ 1 (représentant le nombre de paquets), c est un réel strictement positif représentant la capacité d'une boîte. Une *entrée* est un couple $(X; c)$:

$$(X; c) = \big((X_1, X_2, \ldots, X_N);\ c\big)$$

où les X_i (représentant la taille des paquets) appartiennent à l'intervalle réel fermé $[0, c]$.

Une application f de l'intervalle entier $[1..N]$ dans un intervalle entier $[1..M]$ sera appelée un *empilement* de l'entrée $(X; c) = \big((X_1, X_2, \ldots, X_N);\ c\big)$ si sont satisfaites les deux conditions suivantes:

E1. La fonction f est surjective.

E2. Soit $\Omega_i = f^{-1}(\{i\})$; $\forall i \in [1..M]$:

$$\sum_{j \in \Omega_i} X_j \leq c.$$

La condition **E1** traduit le fait que toute boîte contient au moins un élément; la condition **E2** exprime que les éléments affectés à une boîte n'excèdent pas la capacité de boîte. L'entier M qui représente la cardinalité de l'image de f est appelé le *coût* de f et est noté $\|f\|$.

Etant donné une suite finie $X = (X_1, X_2, \ldots, X_N)$, on définit son *poids* $\pi(X)$ comme:

$$\pi(X) = \sum_{i=1}^{N} X_i.$$

Etant donné une entrée $E = (X; c)$, on définit le *coût optimal* associé $\omega(E)$ par:

$$\omega(E) = \min_f \{\|f\|\}$$

où f parcourt l'ensemble des empilements de E.

La notation $\lfloor u \rfloor$ représente le plus grand entier inférieur ou égal à u.

Dans tout le problème, on notera $E_N = [0, 1]^N$.

PARTIE I. GÉNÉRALITÉS.

Q1. Soit λ un réel strictement positif. Montrer que pour tout N et tout $X \in [0, c]^N$:

$$\omega\big((\lambda X_1, \lambda X_2, \ldots, \lambda X_N);\ \lambda c\big) = \omega\big((X_1, X_2, \ldots, X_N);\ c\big).$$

Montrer que pour tout N, et pour tout $X, Y \in [0, c]^N$:

$$\omega\big((X_1, X_2, \ldots, X_N, Y_1, Y_2, \ldots, Y_N);\ 2c\big) \leq \max\big\{\omega\big((X_1, X_2, \ldots, X_N);\ c\big),\ \omega\big((Y_1, Y_2, \ldots, Y_N);\ c\big)\big\}.$$

Montrer sous la condition $X_i \leq Y_i$ pour tout $i \in [1..N]$ que:

$$\omega\big((X_1, X_2, \ldots, X_N);\ c\big) \leq \omega\big((Y_1, Y_2, \ldots, Y_N);\ c\big).$$

En supposant que $X_i + Y_i \leq c$, déterminer le signe de la quantité:

$$\omega((X_1 + Y_1, X_2 + Y_2, \ldots, X_N + Y_N);\ c) - \omega((X_1, X_2, \ldots, X_N, Y_1, Y_2, \ldots, Y_N);\ c).$$

Q2. Soit $X = (X_1, X_2, \ldots, X_N)$ avec $X_i \in [0, c]$. Calculer $\omega(X; c)$ lorsque $\pi(X) \leq c$ et montrer que si $\pi(X) > c$:

$$\frac{\pi(X)}{c} \leq \omega(X; c) < 2\frac{\pi(X)}{c}.$$

Déterminer $\omega(X; 100)$ lorsque $X = (15, 30, 45, 50, 60, 75, 90)$.

Q3. Montrer que pour chaque N et c, et pour chaque entier M tel que $1 \leq M \leq N$, il existe une entrée $(X; c) = ((X_1, X_2, \ldots, X_N);\ c)$ telle que:

$$2\frac{\pi(X)}{c} - 1 = \omega(X; c) = M.$$

Q4. Il s'agit de montrer dans cette question qu'il n'est pas toujours avantageux de remplir complètement une boîte afin d'obtenir le placement optimal. On prendra ici $c = 1$. Construire pour tout $N \geq 6$ une suite $X = (X_1, X_2, \ldots, X_N)$ (avec $0 < X_i < 1$) telle que soient verifiées les deux conditions **C1**, **C2** suivantes:

C1. $X_1 + X_2 + X_3 = 1$.

C2. $\omega((X_1, X_2, \ldots, X_N); 1) < 1 + \omega(X_4, X_5, \ldots, X_N)$.

Construire une suite vérifiant de plus la condition:

C3. Pour tout ensemble d'indices $I \subseteq \{1, 2, \ldots, N\}$, $I \neq \{1, 2, 3\}$: on a:

$$\sum_{i \in I} X_i \neq 1.$$

PARTIE II. LE PLACEMENT SÉQUENTIEL.

L'algorithme de placement séquentiel ou *NextFit* est défini par la fonction Pascal suivante:

```
type vecteur = array[1..N] of real;
function NextFit(X : vecteur; c : real) : integer;
var boite : vecteur;
  i, M : integer;
begin
  for i :=1 to N do boite[i] :=0;
  M :=1;
  for i:=1 to N do
    begin
    if boite[M]+X[i] <= c then boite[M] := boite[M]+X[i]
      else begin M := M+1; boite[M] := X[i] end;
    f[i] := M
    end;
  NextFit := M
end;
```

Pour cette procédure, N est une constante entière et f est définie à l'extérieur de la procédure comme:

```
var f : array[1..N] of integer;
```

Q5. Quels sont la valeur de f et le résultat retourné lorsque la fonction NextFit est appliquée à l'exemple de la question Q2?

On posera pour simplifier les notations $\gamma_{NF}(X;c) = NextFit(X,c)$ (c'est-à-dire le résultat de l'appel de $NextFit(X,c)$). Montrer que l'exécution de *NextFit* sur un tableau de nombres X avec $X_i \leq c$, met dans les $\gamma_{NF}(X;c)$ premiers éléments du tableau f une suite d'entiers qui constitue un empilement de X de coût $\gamma_{NF}(X;c)$. Montrer que pour tout N et X:

$$\omega(X;c) \leq \gamma_{NF}(X;c) \leq 2\frac{\pi(X)}{c} + 1.$$

Q6. Montrer que pour tout N suffisamment grand, il existe une suite $X = (X_1, X_2, \ldots, X_N)$ avec $0 < X_i < 1$ telle que:

$$\frac{\gamma_{NF}(X;1)}{\omega(X;1)} > \frac{199}{100}.$$

Q7. On considère pour chaque $N \geq 1$ la suite particulière:

$$H^{<N>} = (\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots, \frac{1}{N+1}).$$

Montrer qu'il existe une constante $\delta > 0$ telle que:

$$\forall N \geq 1 : \quad \gamma_{NF}(H^{<N>};1) - \omega(H^{<N>};1) < \delta.$$

PARTIE III. LE PLACEMENT SÉQUENTIEL DÉCROISSANT.

Soit *TriDecroissant* une procédure de type:

```
procedure TriDecroissant(var X : vecteur);
```

qui trie les éléments de la suite X en ordre décroissant, de sorte qu'après exécution de la procédure, on ait:

$$X_1 \geq X_2 \geq X_3 \geq \cdots \geq X_N.$$

La procédure *NextFitDecroissant* est définie par:

```
function NextFitDecroissant(X : vecteur; c : real) : integer;
begin
  TriDecroissant(X);
  NextFitDecroissant:=NextFit(X,c)
end;
```

On note $\gamma_{NFD}(X;c)$ le résultat de *NextFitDecroissant* appliqué à l'entrée $(X;c)$. On fixe dans toute cette partie $c = 1$ et l'on se propose d'évaluer la valeur moyenne ou espérance de $\gamma_{NFD}(X;1)$ lorsque les composantes X_i sont des variables aléatoires indépendantes uniformément distribuées sur l'intervalle réel $[0,1]$.

RAPPELS: Soit g une fonction intégrable de E_N dans l'ensemble des réels. On appelle *espérance* de g et l'on note $\mathbf{E}[g]$ l'intégrale:

$$\mathbf{E}[g] = \int_0^1 \int_0^1 \cdots \int_0^1 g(X_1, X_2, \ldots, X_N)\, dX_1 dX_2 \cdots dX_N.$$

Soit $F \subset E_N$. On appelle *probabilité* de F, que l'on note $\mathbf{Pr}[F]$, l'espérance de la fonction g_F définie par:

$$g_F(X_1, X_2, \ldots, X_N) = \begin{cases} 1 & \text{si } (X_1, X_2, \ldots, X_N) \in F \\ 0 & \text{si } (X_1, X_2, \ldots, X_N) \notin F. \end{cases}$$

Q8. Calculer l'espérance de la fonction $\pi(X) = X_1 + X_2 + \cdots + X_N$.

Q9. Pour k entier, $1 \leq k < N$ et y tel que $0 \leq y \leq 1$, calculer les fonctions $\psi_k(y)$ et $\phi_k(y)$ définies par:

$$\psi_k(y) = \mathbf{Pr}\Big[\{X \in E_N \quad | \quad X_1 + X_2 + \cdots + X_k \leq y\}\Big];$$

$$\phi_k(y) = \mathbf{Pr}\Big[\{X \in E_N \quad | \quad X_1 + X_2 + \cdots + X_k \leq y \quad \text{et} \quad X_1 + X_2 + \cdots + X_k + X_{k+1} > 1\}\Big].$$

Soit $\bar{\alpha}_N$ l'espérance du nombre d'éléments X_i placés dans la première boîte, c'est-à-dire l'espérance de la cardinalité de $\Omega_1 = f^{-1}(\{1\})$, lors de l'application de l'algorithme *NextFit*. Déterminer $\lim \bar{\alpha}_N$ lorsque $N \to \infty$.

Soit, pour l'algorithme *NextFit*, $\bar{\beta}_N$ le remplissage moyen de la première boîte défini comme l'espérance de la quantité:

$$\sum_{j \in \Omega_1} X_j.$$

Calculer $\lim \bar{\beta}_N$ lorsque $N \to \infty$.

Q10. Pour r entier ≥ 2, on définit une partition de l'intervalle réel $[0,1]$ en r intervalles $\mathcal{I}_\ell$ définis comme suit:

- Pour ℓ tel que $1 \leq \ell < r$: $\mathcal{I}_\ell =]\frac{1}{\ell+1}, \frac{1}{\ell}]$.
- L'intervalle $\mathcal{I}_r$ est défini par $\mathcal{I}_r = [0, \frac{1}{r}]$.

Pour une suite $X \in E_N$ et un entier ℓ tel que $1 \leq \ell \leq r$, on définit la fonction $w_\ell(X)$ égale à la cardinalité de l'ensemble

$$\{i \quad | \quad 1 \leq i \leq N, \ X_i \in \mathcal{I}_\ell\}.$$

On pose:

$$V_r(X) = \sum_{\ell=1}^{r-1} \frac{w_\ell(X)}{\ell}; \quad W_r(X) = \sum_{\ell=1}^{r} \frac{w_\ell(X)}{\ell}.$$

Montrer que $\gamma_{NFD}(X;1)$ vérifie l'encadrement:

$$V_r(X) - r \leq \gamma_{NFD}(X;1) \leq W_r(X) + r.$$

Q11. Soit $\vec{n} = (n_1, n_2, \ldots, n_r)$ un vecteur à coordonnées entières. Montrer que la probabilité $p_{\vec{n}}$ définie par:

$$p_{\vec{n}} = \mathbf{Pr}\Big[\{X \in E_N \quad | \quad w_1(X) = n_1, \, w_2(X) = n_2, \, \ldots, \, w_r(X) = n_r\}\Big]$$

vaut lorsque $n_1 + n_2 + \cdots + n_r = N$, et $n_i \geq 0$:

$$p_{\vec{n}} = \frac{N!}{n_1! n_2! \cdots n_r!} \iota_1^{n_1} \iota_2^{n_2} \cdots \iota_r^{n_r}$$

où ι_ℓ représente la longueur de l'intervalle $\mathcal{I}_\ell$.

Q12. Calculer le polynôme Q en les indéterminées $t_1, t_2, \ldots, t_r$ défini par:

$$Q(t_1, t_2, \ldots, t_r) = \sum_{\vec{n}} p_{\vec{n}} \, t_1^{n_1} t_2^{n_2} \cdots t_r^{n_r},$$

où la sommation est étendue à tous les vecteurs $\vec{n} = (n_1, n_2, \ldots, n_r)$ entiers tels que $n_i \geq 0$ et $n_1 + n_2 + \cdots + n_r = N$. En déduire l'espérance des fonctions $V_r(X)$ et $W_r(X)$.

Q13. On définit $\bar{\gamma}_{NFD}(N)$ comme l'espérance de $\gamma_{NFD}(X;1)$ pour $X \in E_N$. En choisissant convenablement r en fonction de N, montrer l'existence de la limite:

$$\lim_{N\to\infty} \frac{\bar{\gamma}_{NFD}(N)}{N}$$

et évaluer cette limite. On rappelle que:

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}.$$

PARTIE IV. COUT DU PLACEMENT OPTIMAL.

L'objet de cette partie est de déterminer l'espérance du coût optimal lorsque $N \to \infty$. On prendra ici $c = 1$.

Soit $r = 2s+1$ un entier impair, avec $r \geq 3$. On définit une partition de l'intervalle $[0,1]$ en r intervalles $\mathcal{J}_\ell$ de la manière suivante:

- Pour ℓ tel que $1 \leq \ell < r$, $\mathcal{J}_\ell = [\frac{\ell-1}{r}, \frac{\ell}{r}[$.
- L'intervalle $\mathcal{J}_r$ est défini par: $\mathcal{J}_r = [\frac{r-1}{r}, 1]$.

On définit pour chaque ℓ et $X = (X_1, X_2, \ldots, X_N)$ l'ensemble:

$$\zeta_\ell(X) = \{i \mid 1 \leq i \leq N,\ X_i \in \mathcal{J}_\ell\}.$$

On note $Z_\ell(X)$ la cardinalité de l'ensemble $\zeta_\ell(X)$.

On considère une méthode de placement dite *méthode des placements complémentaires* et spécifiée comme suit:

A. Chaque élément de $\zeta_r(X)$ est placé dans une boîte séparée.

B. Pour j de 1 à s faire:

- Soit $\lambda_j = \min(Z_j(X), Z_{r-j}(X))$.
- Choisir deux sous-ensembles $\zeta'_j(X) \subseteq \zeta_j(X)$ et $\zeta'_{r-j}(X) \subseteq \zeta_{r-j}(X)$ de cardinalité λ_j. Chaque élément de $\zeta'_j(X)$ est apparié à un élément de ζ'_{r-j} et la paire est placée dans une boîte séparée.
- Les éléments restants de $\zeta_j(X)\backslash\zeta'_j(X) \cup \zeta_{r-j}(X)\backslash\zeta'_{r-j}(X)$ sont affectés chacun à une boîte différente.

Q14. Soit $\gamma_{PC}(X;1)$ le coût de l'empilement produit par un algorithme obéissant au schéma de placement complémentaire. Déterminer $\gamma_{PC}(X;1)$ lorsque $r = 5$ et

$$X = (\tfrac{15}{100}, \tfrac{30}{100}, \tfrac{45}{100}, \tfrac{50}{100}, \tfrac{60}{100}, \tfrac{75}{100}, \tfrac{90}{100}).$$

Etablir la majoration:

$$\gamma_{PC}(X;1) \leq \frac{N}{2} + \frac{r-1}{2}\Delta_r(X) + \frac{N}{2r},$$

où la fonction $\Delta_r(X)$, appelée écart de distribution, est définie comme:

$$\Delta_r(X) = \max_{1\leq \ell \leq r}\{\frac{N}{r} - Z_\ell(X)\}.$$

Q15. Pour $\vec{n} = (n_1, n_2, \ldots, n_r)$ un vecteur à coordonnées entières, déterminer les probabilités:

$$p'_{\vec{n}} = \mathbf{Pr}\Big[\{X \in E_N \mid Z_1(X) = n_1,\ Z_2(X) = n_2,\ \ldots,\ Z_r(X) = n_r\}\Big].$$

On définit pour tout b entier ≥ 0, la fonction:

$$e_b(z) = \sum_{j=0}^{b} \frac{z^j}{j!},$$

et l'on pose pour b entier avec $0 \leq b < \frac{N}{r}$:

$$q_{N,b} = \mathbf{Pr}\big[\{X \in E_N \quad | \quad \forall \ell : \ Z_\ell(X) > b\}\big].$$

Exprimer $q_{N,b}$ en fonction du coefficient de z^N dans le développement de Taylor de:

$$\left(e^z - e_b(z)\right)^r.$$

Q16. Soit Γ_ρ le cercle de rayon ρ centré à l'origine orienté positivement. Montrer que:

$$1 - q_{N,b} = \frac{N!}{2i\pi r^N} \int_{\Gamma_\rho} \left[e^{rz} - (e^z - e_b(z))^r\right] \frac{dz}{z^{N+1}}.$$

En prenant, $\rho = \frac{N}{r}$, établir la majoration:

$$1 - q_{N,b} \leq \frac{N! e^N}{N^N} \left[1 - \left(1 - e^{-N/r} e_b(\frac{N}{r})\right)^r\right].$$

Q17. Montrer que pour u réel avec $u > b \geq 0$, on a:

$$e_b(u) \leq (b+1)\frac{u^b}{b!} \qquad \text{et} \qquad e_b(u) < e^u.$$

Montrer qu'il existe deux constantes $C > 0$ et $K > 0$ telles que, lorsque $r = 2\lfloor \frac{N^{1/3}}{2} \rfloor + 1$ et $b = \lfloor N^{2/3} \rfloor - \lfloor N^{1/2} \rfloor$:

$$1 - q_{N,b} < K e^{-CN^{1/3}}.$$

On rappelle qu'on a par la formule de Stirling:

$$N! \sim N^N e^{-N} \sqrt{2\pi N} \qquad \text{lorsque } N \to \infty.$$

Q18. On définit $\bar{\omega}(N)$ comme égal à l'espérance de $\omega(X;1)$ pour $X \in E_N$. Montrer l'existence de la limite

$$\lim_{N \to \infty} \frac{\bar{\omega}(N)}{N},$$

et déterminer cette limite. Soit de même $\bar{\pi}(N)$ l'espérance de $\pi(X)$; évaluer

$$\lim_{N \to \infty} \frac{\bar{\omega}(N)}{\bar{\pi}(N)} \qquad \text{et} \qquad \lim_{N \to \infty} \frac{\bar{\gamma}_{NFD}(N)}{\bar{\omega}(N)}.$$

PARTIE V. *NP*-COMPLÉTUDE DU PROBLÈME GÉNÉRAL.

Dans cette partie, on part d'un problème classique dit de *recouvrement exact* ou *REC*. Une entrée de *REC* consiste en un couple $(t, \{S_j\}_{1 \leq j \leq J})$ où t est entier positif et $\{S_j\}_{1 \leq j \leq J}$ est une famille d'ensembles telle que chaque $S_j \subseteq [1..t]$. Le problème *REC* consiste à reconnaître s'il existe une sous famille $\{T_h\}_{1 \leq h \leq H}$ de la famille des $\{S_j\}$, $T_h = S_{j_h}$ avec $1 \leq j_1 < \ldots < j_H \leq J$, telle que:

R1. Les T_h sont deux à deux disjoints.

R2. L'union des T_h est égale à l'ensemble $\{1, 2, \ldots, t\}$.

On suppose dans la suite du problème que les entiers sont représentés de manière usuelle en binaire et que les ensembles et les suites sont donnés par la liste de leurs éléments. On admettra qu'avec ces conventions de représentation, le problème *REC* est *NP*-complet.

On rappelle que les opérations arithmétiques élémentaires sur les entiers ($+, -, \times$ et comparaisons) ont des coûts polynômiaux en la taille de leurs arguments. De même, pour les opérations élémentaires sur les listes (recherche d'un élément, concaténation) dont le coût est polynômial en la longueur *totale* des arguments.

Q19. Le problème de *Sac-à-Dos* (*KNAPSACK*) consiste, étant donné une suite d'entiers $\{a_i\}_{1 \leq i \leq t}$ et un entier b, à déterminer s'il existe un sous-ensemble des $\{a_i\}$ dont la somme soit exactement égale à b.

Construire une réduction calculable en temps polynomial (en fonction de la longueur de l'argument) du problème *REC* au problème *KNAPSACK*. Pour cela, on pourra prendre d supérieur au nombre J d'ensembles S_j de l'entrée de *REC*, et "coder" les ensembles S_j par des entiers en base d formés des chiffres 0 et 1.

Q20. Le problème dit de *partition* (*PART*) consiste, étant donné une suite $\{b_i\}_{1 \leq i \leq s}$ de nombres entiers à déterminer s'il existe un sous-ensemble $I \subset [1..s]$ tel que:

$$\sum_{h \in I} b_h = \sum_{h \notin I} b_h.$$

Construire une réduction polynomiale de *KNAPSACK* à *PART*.

Q21. On considère le problème d'empilement de paquets dans lequel les paquets sont à valeur entière et la capacité de boîte est un entier ≥ 1. Montrer que le problème de déterminer si une entrée $(X; c)$ est telle que:

$$\omega(X; c) = 2$$

est *NP*–complet.

Q22. Quelles conclusions algorithmiques générales sur l'empilement de paquets peut-on tirer de l'ensemble du problème lorsque la capacité de boîte c est égale à 1?