

Complexité d'algorithmes basés sur des opérations sur les arbres

S. Dulucq
S. Gire

LaBRI, Université Bordeaux I

avec le soutien du P.R.C. Mathématiques et Informatique

E-mail : dulucq@geocub.greco-prog.fr

Abstract

We consider operations on trees like paths reversals and standard path compression used in a recent mutual exclusion algorithm [8] or the algorithm for maintaining disjoint sets under union [11]. We give exact values for the worst-case cost of a sequence of these operations performed on an arbitrary initial tree. To obtain these bounds, we apply the potential function method of amortized analysis introduced by Tarjan [10].

La structure d'arbre est l'une des plus importantes et des plus spécifiques de l'informatique. Certains algorithmes, utilisant cette structure pour la gestion dynamique ou logique d'informations, sont basés sur des transformations agissant sur les arbres.

Citons par exemple l'algorithme de gestion de partitions [1, 4, 11] qui consiste à maintenir une collection d'ensembles disjoints (ou parts) où les seules opérations autorisées sont l'union de deux parts et la recherche de la part contenant un élément donné. Cet algorithme est utilisé pour la gestion des "Common" et "Equivalence" en Fortran [5], la recherche d'un arbre recouvrant de poids minimal d'un graphe [3], le problème de l'équivalence de deux automates d'états finis [1], ...

Le but des transformations réalisées sur ces structures de données arborescentes, lors de chaque accès à un élément, est de diminuer le coût des opérations ultérieures.

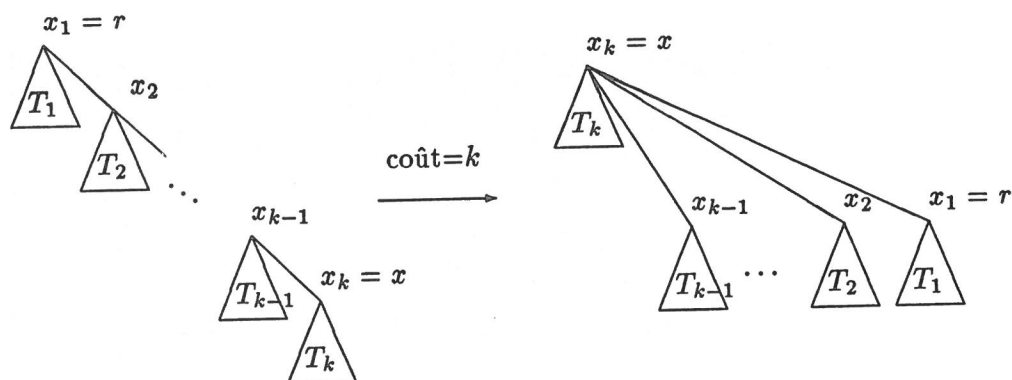


Figure 1: Inversion de type 0 du chemin $x \rightarrow r$.

Dans [9, 10], Sleator et Tarjan montrent l'intérêt des structures arborescentes modifiables suivant de telles transformations, par rapport aux structures de données auxquelles sont associées des contraintes, comme par exemple les conditions d'équilibre dans les arbres AVL.

Les transformations les plus fréquemment utilisées sont la compression standard, l'éclatement, le partage et l'inversion de chemin [11], cette dernière transformation se trouvant à la base d'un nouvel algorithme d'exclusion mutuelle [8].

Parmi ces diverses transformations, nous nous intéressons ici plus particulièrement à la compression standard et aux divers types d'inversions de chemin dans les arbres.

L'inversion de chemin de type 0 est à la base d'un algorithme d'exclusion mutuelle dans un réseau réparti, algorithme récemment proposé par Naïmi et Tréhel. Cet algorithme a pour but d'assurer l'accès exclusif de n processus à une ressource, ces processus communiquant par échange de messages. Il consiste à maintenir une organisation logique des processus sous forme arborescente, la demande d'accès à la ressource se traduisant par une inversion (de type 0) du chemin dans l'arbre menant du processus demandeur au processus situé à la racine (figure 1).

Le coût d'une telle inversion de chemin $x \rightarrow r$ dans un arbre est la longueur de ce chemin en nombre de sommets (k sur la figure 1) et correspond au nombre de messages envoyés par les processus (chaque processus du chemin prévient son "père" de la demande d'accès du processus x à la ressource).

Récemment, Arnold, Delest et Dulucq [2] ont montré que le coût moyen M_n (pour n processus) d'une telle transformation est exactement H_{n-1} le

$(n - 1)$ ième nombre harmonique.

Nous nous intéressons ici au coût au pire de cet algorithme, et donc de l'inversion de chemin dans les arbres à n sommets. Lorsque l'arbre est filiforme (une seule feuille f), l'inversion du chemin $f \rightarrow \text{racine}$ est de coût n et l'arbre obtenu est le "peigne" ($n - 1$ feuilles); toute nouvelle inversion de chemin sur cet arbre serait alors de coût égal à 2. Cette constatation nous amène à non pas considérer la complexité au pire d'une seule inversion (qui vaut donc n), mais la notion de complexité amortie introduite par Tarjan [10]. Cette complexité correspond à maximiser le coût moyen d'une suite d'inversions.

Soient donc, pour les arbres à n sommets,

$$\text{Ch}(n, m) = \max_w \{ \text{coût moyen}(w) : w \text{ suite de } m \text{ inversions} \},$$

$$\text{Cy}(n) = \max_{\bar{w}} \{ \text{coût moyen}(\bar{w}) : \bar{w} \text{ cycle d'inversions} \},$$

où un cycle \bar{w} correspond à une suite d'inversions telles que le premier et le dernier arbre de cette suite soient identiques.

Ginat, Sleator et Tarjan [6] ont récemment montré que

$$\text{Ch}(n, m) \leq 1 + \log n + \frac{n \log n}{2m} ,$$

$$\text{Cy}(n) \leq 1 + \log n .$$

(Nous notons $\log n$ le logarithme à base 2 de n).

Dans une première partie, nous décrivons la méthode introduite par Tarjan [10] et employée dans [6]. Le point essentiel de cette méthode consiste en la définition d'une fonction sur la structure d'arbre, appelée *potentiel*.

Le choix d'un meilleur potentiel que celui considéré dans [6] nous permet de montrer dans la seconde partie que

$$\text{Ch}(n, m) \leq 1 + 2f(n) + \frac{nf(n)}{m} ,$$

où

$$f(n) = \frac{1}{2} \left(\lfloor \log n \rfloor + \frac{n - 2^{\lfloor \log n \rfloor}}{2^{\lfloor \log n \rfloor}} \right) .$$

De plus, en mettant en évidence des cycles construits à partir d'une classe particulière d'arbres, les arbres binômiaux, nous montrons que

$$\text{Cy}(n) = 1 + \lfloor \log n \rfloor + \frac{n - 2^{\lfloor \log n \rfloor}}{2^{\lfloor \log n \rfloor}} ,$$

la longueur des cycles correspondants étant

$$\frac{2^{\lfloor \log n \rfloor}}{\text{pgcd}(2^{\lfloor \log n \rfloor}, n - 2^{\lfloor \log n \rfloor})} .$$

Ce résultat, venant compléter celui obtenu par Arnold, Delest et Dulucq [2], montre que l'algorithme d'exclusion mutuelle de Naïmi et Tréhel [8] a une complexité en moyenne et au pire (complexité amortie) logarithmique, ce qui en fait le meilleur algorithme connu à ce jour pour réaliser l'exclusion mutuelle.

Nous nous intéressons également dans ce travail au coût au pire de deux autres transformations sur les arbres intervenant dans la gestion de partitions : la compression standard de chemin (figure 2) et l'inversion de chemin de type p où $1 \leq p < n - 1$ dans les arbres à n sommets.

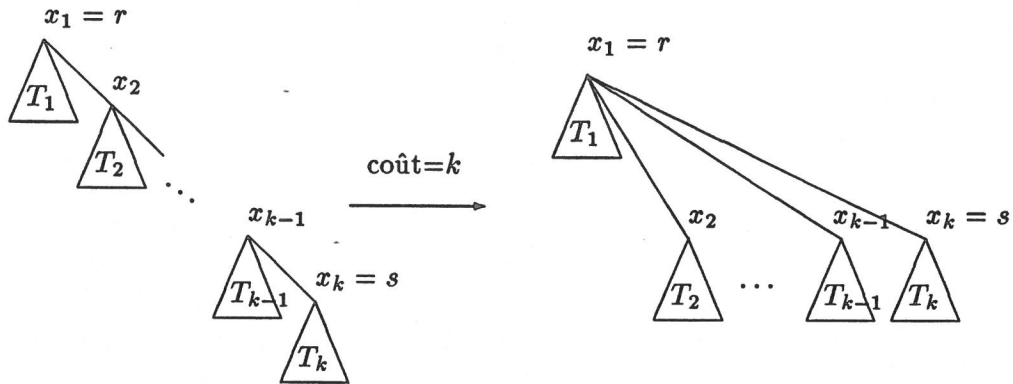


Figure 2: Compression standard du chemin $r \rightarrow s$.

Dans les paragraphes 3 et 4, en utilisant la même technique que pour l'inversion de chemin (de type 0), nous montrons que dans le cas de l'inversion de chemin de type 1 (figure 5 du paragraphe 3)

$$\text{Cy}(n) = 2 + 2f(n - 1) ,$$

et que pour l'inversion de type p où $1 < p < n - 1$

$$\text{Ch}(n, m) = \begin{cases} 3 + \frac{n-3}{m} & \text{si } m \leq n-p, \\ 2 + \frac{2n-p-3}{m} & \text{sinon.} \end{cases}$$

Enfin, pour terminer, nous montrons que dans le cas de la compression standard

$$\text{Ch}(n, m) = 2 + \frac{n-2}{m}.$$

1 Complexité amortie

Cette notion a été introduite par Tarjan ([10]). Nous rappelons ici les principaux résultats la concernant, qui seront utilisés par la suite.

Soit $w = \varphi_1, \dots, \varphi_m$ une suite de m transformations sur un arbre T_0 ayant n sommets. On appelle coût moyen de la suite w la quantité

$$\text{coût moyen}(w) = \frac{1}{m} \sum_{i=1}^m c(\varphi_i)$$

où $c(\varphi_i)$ est le coût de la transformation φ_i .

Lorsque l'arbre obtenu à l'issue de ces transformations est égal à l'arbre initial T_0 , on nommera *cycle* cette suite de transformations.

Définition 1

$$\text{Ch}(n, m) = \max_w \{ \text{coût moyen}(w) : w \text{ suite de } m \text{ transformations} \}$$

$$\text{Cy}(n) = \max_{\bar{w}} \{ \text{coût moyen}(\bar{w}) : \bar{w} \text{ cycle de transformations} \}$$

A tout arbre T , on associe un nombre réel $\phi(T)$, appelé *potentiel* de T .

Le *coût amorti* d'une transformation φ sur un arbre T est alors défini par

$$a(\varphi) = c(\varphi) + \phi(\varphi(T)) - \phi(T),$$

où $c(\varphi)$ est le coût de la transformation φ et $\varphi(T)$ est l'arbre obtenu à l'issue de la transformation.

Lemme 2 *Le coût moyen d'une suite de m transformations est la quantité*

$$\frac{1}{m} \left(\sum_{i=1}^m a(\varphi_i) + \phi_0 - \phi_m \right)$$

où $a(\varphi_i)$ correspond au coût amorti de la i ème transformation φ_i et ϕ_0 (resp. ϕ_m) représente le potentiel de l'arbre initial (resp. final).

Corollaire 3 Le coût moyen d'un cycle de m transformations est

$$\frac{1}{m} \sum_{i=1}^m c(\varphi_i) = \frac{1}{m} \sum_{i=1}^m a(\varphi_i) \quad .$$

2 L'inversion de chemin

Etant donné un arbre T de racine r et un sommet x de T , l'arbre obtenu par inversion (de type 0) du chemin joignant r à x (noté $r \rightarrow x$) est l'arbre T' où la nouvelle racine est x et tous les autres sommets du chemin ont x pour père. Le coût d'une telle transformation est égal à la longueur (en nombre de sommets) du chemin $r \rightarrow x$ (k sur la figure 1).

Théorème 4 Le coût maximal $Ch(n, m)$ d'une suite de m inversions de chemin sur un arbre à n sommets vérifie

$$Ch(n, m) \leq 1 + \left(\frac{n}{2m} + 1\right)(\lfloor \log n \rfloor + \frac{n - 2^{\lfloor \log n \rfloor}}{2^{\lfloor \log n \rfloor}})$$

Théorème 5 Pour tout n , il existe un cycle d'inversions sur un arbre à n sommets de coût moyen

$$1 + \lfloor \log n \rfloor + \frac{n - 2^{\lfloor \log n \rfloor}}{2^{\lfloor \log n \rfloor}}$$

et de longueur

$$\frac{1}{\text{pgcd}(2^{\lfloor \log n \rfloor}, n - 2^{\lfloor \log n \rfloor})} \quad .$$

Corollaire 6 Le coût maximal $Cy(n)$ d'un cycle d'inversions sur un arbre à n sommets est

$$Cy(n) = 1 + \lfloor \log n \rfloor + \frac{n - 2^{\lfloor \log n \rfloor}}{2^{\lfloor \log n \rfloor}} \quad .$$

Notations et définitions.

Pour tout entier positif p , nous notons $f(p) = \frac{1}{2}(\lfloor \log p \rfloor + \frac{p - 2^{\lfloor \log p \rfloor}}{2^{\lfloor \log p \rfloor}})$.

Pour tout sommet x d'un arbre T , nous définissons $s_T(x)$ comme étant le nombre de descendants de x , x compris.

Nous considérons le potentiel d'un arbre T défini par

$$\phi(T) = \sum_{x \in T} f(s_T(x)) \quad .$$

Lemme 7 *Le coût amorti d'une inversion de chemin φ sur un arbre à n sommets vérifie $a(\varphi) \leq 1 + 2f(n)$.*

Preuve.

Soit l'inversion φ du chemin $x_1 \rightarrow x_k$ sur l'arbre à n sommets T (figure 1).

Soit $T' = \varphi(T)$. Nous avons

- $\forall x \neq x_i \quad i \in \{1, \dots, k\} \quad s_T(x) = s_{T'}(x)$,
- $s_{T'}(x_k) = s_T(x_1) = n$,
- $\forall i = 1, \dots, k-1 \quad s_{T'}(x_i) = s_T(x_i) - s_T(x_{i+1})$.

Le coût amorti de l'inversion s'exprime alors sous la forme

$$a(\varphi) = 1 + \sum_{i=1}^{k-1} (1 + f(s_T(x_i) - s_T(x_{i+1})) - f(s_T(x_{i+1}))) .$$

La fonction f vérifiant la propriété $1 + f(p - q) \leq 2f(p) - f(q)$ pour tout entiers $p > q \geq 1$. (Une démonstration en est donnée dans [7]), il vient

$$a(\varphi) \leq 1 + 2 \sum_{i=1}^{k-1} (f(s_T(x_i)) - f(s_T(x_{i+1}))) = 1 + 2(f(s_T(x_1)) - f(s_T(x_k))) ,$$

et donc $a(\varphi) \leq 1 + 2f(n)$. \square

Preuve du théorème 4.

Etant donné un arbre T à n sommets, son potentiel vérifie $f(n) \leq \phi(T) \leq nf(n)$. Ainsi, $\phi_0 - \phi_m \leq (n-1)f(n)$. Des lemmes 2 et 7, nous déduisons

immédiatement $Ch(n, m) \leq 1 + 2f(n) + \frac{nf(n)}{m}$. \square

Étapes de la preuve du théorème 5.

Nous ne donnons ici que les grandes étapes de cette preuve (pour plus de détails, voir [7]).

Le cycle de coût maximal $Cy(n)$ est construit à partir d'une classe d'arbres appelés *arbres binômiaux* (figure 3), l'arbre binomial B_k étant de taille 2^k et de hauteur $k + 1$.

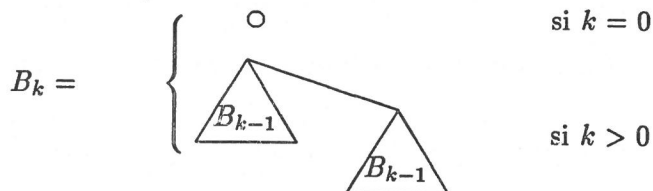


Figure 3: Arbre binomial B_k .

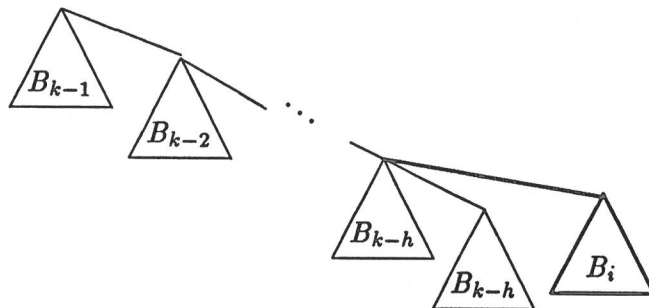


Figure 4: $B_k[B_i, h]$

La preuve du théorème est basé sur la décomposition du nombre de sommets n en base 2.

- 1er cas : $n = 2^k$

Une inversion du chemin *racine* $\rightarrow s$ où s est le sommet de hauteur $k + 1$ sur B_k (feuille la plus basse) est de coût $k + 1$ et redonne l'arbre B_k . Le cycle de longueur 1 ainsi obtenu est de coût $k + 1 = 1 + 2f(n)$.

- 2ième cas : $n = 2^k + 2^i$, $0 \leq i < k$.

Nous notons $B_k[B_i, h]$, $0 \leq i < k$, $1 \leq h \leq k - i + 1$, l'arbre composé de l'arbre binomial B_i greffé à hauteur h sur la branche principale (la plus longue) de l'arbre binomial B_k (voir figure 4). Nous notons $B_k[B_i, (h)]$ l'arbre composé de B_i greffé à hauteur h sur B_k , non nécessairement sur la branche principale.

Pour tout h , $1 \leq h < k - i + 1$, si φ est l'inversion du chemin *racine* $\rightarrow s$ où s est le sommet de hauteur $k + 1$ sur B_k , nous avons

- . $\varphi^{2^{h-1}}(B_k[B_i, h]) = B_k[B_i, h + 1]$,
- . $\forall j < 2^{h-1}$ $\varphi^j(B_k[B_i, h]) = B_k[B_i, (h')]$ avec $h' < h + 1$.

Si ψ est l'inversion du chemin *racine* $\rightarrow s$ où s est le sommet de hauteur $i + 1$ de B_i (ou $k + 2$ dans $B_k[B_i, k - i + 1]$),

- . $\psi(B_k[B_i, k - i + 1]) = B_k[B_i, 1]$ et le coût de cette inversion vaut $k + 2$.

Considérons l'arbre $T_0 = B_k[B_i, k - i]$. En effectuant

$$m = 2^{k-i-1} + 1 + \sum_{j=1}^{k-i-1} 2^{k-j} = 2^{k-i} = \frac{2^{\lfloor \log n \rfloor}}{\text{pgcd}(2^{\lfloor \log n \rfloor}, n - 2^{\lfloor \log n \rfloor})}$$

inversions φ ou ψ sur T_0 , on obtient à nouveau l'arbre T_0 . Le coût moyen de ce cycle est alors

$$\frac{1}{m}((k + 2) + (k + 1)(m - 1)) = k + 1 + \frac{1}{2^{k-j}} = 1 + 2f(n).$$

- 3ème cas : $n = 2^k + 2^{i_1} + \dots + 2^{i_p}$ avec $0 \leq i_p < i_{p-1} < \dots < i_1 < k$, $0 \leq p \leq k$.

Les notations et propriétés précédentes se généralisent pour ce cas.

Considérons T_0 l'arbre à n sommets composé de l'arbre binomial B_k sur la branche principale duquel sont greffés les arbres binomiaux B_i aux hauteurs $k - i$ pour $i \in \{i_1, \dots, i_p\}$. En effectuant

$$m = \text{ppcm}(2^{k-i_1}, 2^{k-i_2}, \dots, 2^{k-i_p}) = 2^{k-i_p} = \frac{2^{\lfloor \log n \rfloor}}{\text{pgcd}(2^{\lfloor \log n \rfloor}, n - 2^{\lfloor \log n \rfloor})}$$

inversions φ ou ψ sur T_0 , on obtient à nouveau T_0 . Le coût total de ces inversions est

$$(k+2) \left(\sum_{i \in \{i_1, \dots, i_p\}} 2^{i-i_p} \right) \quad (\text{pour les inversions du type } \varphi)$$

$$+ (k+1) (2^{k-i_p} - \sum_{i \in \{i_1, \dots, i_p\}} 2^{i-i_p}) \quad (\text{pour les inversions du type } \psi)$$

ce qui donne un coût moyen pour ce cycle d'inversions de $1 + 2f(n)$. \square

Preuve du corollaire 6.

Ce résultat est une conséquence immédiate des théorème 5, lemme 7 et corollaire 3.

3 L'inversion de chemin de type 1

Etant donné un arbre T de racine r et un sommet s de T , l'arbre obtenu par inversion de type 1 du chemin $r \rightarrow s$ est l'arbre T' où le sommet s a r pour père et les sommets du chemin autres que s et r ont s pour père. Ceci correspond à une inversion de type 0 sur le sous arbre de T de racine x_2 (figure 5). Le coût d'une telle transformation est égal à la longueur du

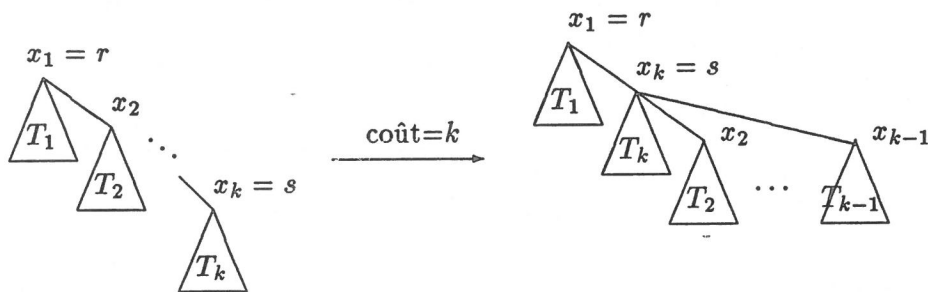


Figure 5: Inversion de type 1 du chemin $r \rightarrow s$.

chemin (k sur la figure 5).

Théorème 8 *Le coût maximal $Ch(n, m)$ d'une suite de m inversions de chemin de type 1 sur un arbre à n sommets vérifie*

$$Ch(n, m) \leq 2 + \lfloor \log(n-1) \rfloor + \frac{(n-1) - 2^{\lfloor \log(n-1) \rfloor}}{2^{\lfloor \log(n-1) \rfloor}} + \frac{n}{2m} \left(\lfloor \log n \rfloor + \frac{n - 2^{\lfloor \log n \rfloor}}{2^{\lfloor \log n \rfloor}} \right)$$

Théorème 9 *Pour tout n , il existe un cycle d'inversions de type 1 sur un arbre à n sommets de coût moyen*

$$\text{et de longueur } \frac{2 + \lfloor \log(n-1) \rfloor + \frac{(n-1) - 2^{\lfloor \log(n-1) \rfloor}}{2^{\lfloor \log(n-1) \rfloor}}}{\text{pgcd}(2^{\lfloor \log(n-1) \rfloor}, (n-1) - 2^{\lfloor \log(n-1) \rfloor})} .$$

Corollaire 10 *Le coût maximal $Cy(n)$ d'un cycle d'inversion de type 1 sur un arbre à n sommets est $Cy(n) = 2 + \lfloor \log(n-1) \rfloor + \frac{(n-1) - 2^{\lfloor \log(n-1) \rfloor}}{2^{\lfloor \log(n-1) \rfloor}}$.*

La démonstration de ces résultats se fait de manière analogue au cas de l'inversion de type 0. Ici aussi, nous prenons comme *potentiel* d'un arbre T la quantité $\phi(T) = \sum_{x \in T} f(s_T(x))$.

Lemme 11 *Le coût amorti d'une inversion de chemin φ de type 1 sur un arbre à n sommets vérifie $a(\varphi) \leq 2 + 2f(n-1)$*

Preuve.

En procédant comme dans le paragraphe précédent (lemme 7), le coût amorti de l'inversion de type 1 d'un chemin $x_1 \rightarrow x_k$ de l'arbre T (figure 5) devient

$$\begin{aligned} a(\varphi) &= 2 + \sum_{i=2}^{k-1} (1 + f(s(x_i) - s(x_{i+1})) - f(s(x_{i+1}))) \\ &\leq 2 + 2 \sum_{i=2}^{k-1} (f(s(x_i)) - f(s(x_{i+1}))) = 2 + 2(f(s(x_2)) - f(s(x_k))) , \end{aligned}$$

d'où $a(\varphi) \leq 2 + 2f(n-1)$. \square

Preuve du théorème 8.

La preuve est identique au cas de l'inversion de type 0.

Le coût maximal vérifie $Ch(n, m) \leq 2 + 2f(n-1) + \frac{nf(n)}{m}$

Preuve du théorème 9.

Soient $T'_0, T'_1, \dots, T'_{m-1}, T'_m = T'_0$ les arbres à $n-1$ sommets qui forment le cycle d'inversions de type 0 de coût maximal $Cy(n-1) = 1 + 2f(n-1)$, où

$$m = \frac{2^{\lfloor \log(n-1) \rfloor}}{\text{pgcd}(2^{\lfloor \log(n-1) \rfloor}, (n-1) - 2^{\lfloor \log(n-1) \rfloor})} .$$

Considérons l'arbre T_0 à n sommets constitué d'une racine r à laquelle est accroché l'arbre T'_0 . En effectuant les m inversions de type 1 sur T_0 (correspondant aux m inversions de type 0 sur les arbres T'_i), on obtient un cycle de longueur m formé des arbres $T_0, \dots, T_{m-1}, T_m = T_0$ où chaque arbre T_i ($i = 1, \dots, m$) est constitué de la racine r à laquelle est accroché l'arbre T'_i . Le coût moyen de ce cycle est alors $1 + (1 + 2f(n-1)) = 2 + 2f(n-1)$ \square

Preuve du corollaire 10.

Ce résultat est une conséquence immédiate des théorème 9, lemme 11 et corollaire 3.

4 L'inversion de chemin de type p

Etant donné un arbre T de racine r et un sommet s de T , l'arbre obtenu par inversion de type p , $1 < p < n-1$, du chemin $r \rightarrow s$ est l'arbre T' où les $p-1$ premiers sommets (non compris r) du chemin ainsi que le sommet s ont r pour père et les autres sommets du chemin ont s pour père (figure 6). Le coût d'une telle transformation est égal au nombre de sommets du chemin $r \rightarrow s$ (k sur la figure 6).

Théorème 12 *Le coût maximal $Ch(n, m)$ d'une suite de m inversions de chemin de type p , $1 < p < n-1$, sur un arbre à n sommets est*

$$Ch(n, m) = \begin{cases} 3 + \frac{n-3}{m} & \text{si } m \leq n-p , \\ 2 + \frac{2n-p-3}{m} & \text{sinon .} \end{cases}$$

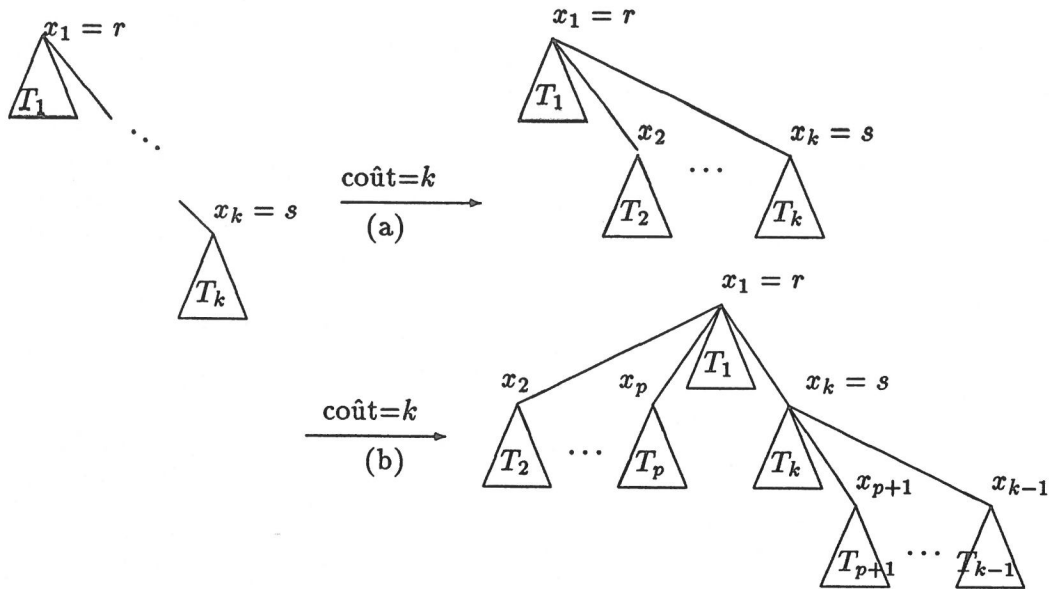


Figure 6: Inversion de type p ($p \geq 1$) du chemin $r \rightarrow s$ dans les cas $k \leq p$ (a) et $k > p$ (b).

Notation. Notons $N^i(T)$ le nombre de sommets à hauteur i dans l'arbre T .

Remarque 13 Soit T' l'arbre obtenu par inversion de type p sur un sommet s de T .

- (i) Si s est à hauteur 1 ($s = \text{racine}$) ou 2, alors $T' = T$.
- (ii) Si s est à hauteur k , $2 < k \leq p$, alors $N^2(T') = N^2(T) + k - 2$.
- (iii) Si s est à hauteur k , $p < k$, alors $N^2(T') = N^2(T) + p - 1$.

Dans le cas de l'inversion de type p , nous définissons le potentiel d'un arbre T comme étant la quantité $\phi(T) = -N^2(T) - N^3(T)$.

Lemme 14 Le coût amorti d'une inversion φ de type p vérifie $a(\varphi) \leq 3$.

Preuve.

Soit l'inversion de type p du chemin $x_1 \rightarrow x_k$ sur l'arbre T . En utilisant les notations de la figure 6, deux cas se présentent suivant la longueur du chemin.

- $k \leq p$

Le coût amorti de la transformation est

$$\begin{aligned} a(\varphi) &= k + (-N^2(T_1) - (k-1) - N^3(T_1) - \sum_{i=2}^k N^2(T_i)) \\ &\quad - (-N^2(T_1) - 1 - N^3(T_1) - N^2(T_2) - 1) \\ &= 3 - \sum_{i=3}^k N^2(T_i) \leq 3 \end{aligned}$$

- $k > p$

Le coût amorti de la transformation est

$$\begin{aligned} a(\varphi) &= k + (-N^2(T_1) - p - N^3(T_1) - \sum_{i=2}^p N^2(T_i) - N^2(T_k) - (k-p-1)) \\ &\quad - (-N^2(T_1) - 1 - N^3(T_1) - N^2(T_2) - 1) \\ &= 3 - (\sum_{i=3}^p N^2(T_i) + N^2(T_k)) \leq 3 \quad \square \end{aligned}$$

Lemme 15 *Le coût maximal d'une suite de m inversions de type p sur un arbre à n sommets vérifie* $Ch(n, m) \leq 3 + \frac{n-3}{m}$.

Preuve.

Pour tout arbre T à n ($n \geq 2$) sommets, le nombre de sommets à hauteur 2 ou 3 est compris entre 2 et $n-1$. D'où $-(n-1) \leq \phi(T) \leq -2$, et ainsi $\phi_0 - \phi_m \leq n-3$.

Des lemmes 2 et 14, nous déduisons alors que $Ch(n, m) \leq 3 + \frac{n-3}{m}$ \square

Lemme 16 *Pour tout n et $m \leq n-p$, il existe une suite de m inversions de type p sur un arbre à n sommets de coût moyen* $3 + \frac{n-3}{m}$.

Considérons T_0 l'arbre filiforme à n sommets (une seule feuille f). En effectuant $n-p$ inversions de type p sur T_0 (1 inversion de coût n et $n-p-1$ de coût 3), on obtient le "peigne" ($n-1$ feuilles), arbre invariant par toute inversion de coût 2 (figure 7). Ainsi, si l'on effectue m inversions, $m \leq n-p$, le coût moyen du chemin obtenu est $\frac{1}{m}(n + 3(m-1)) = 3 + \frac{n-3}{m}$ \square

Si le nombre m d'inversions effectuées vérifie $m > n-p$, ce chemin a un coût moyen égal à $\frac{1}{m}(n + 3(n-p-1) + 2(m-n-p)) = 2 + \frac{2n-p-3}{m}$.

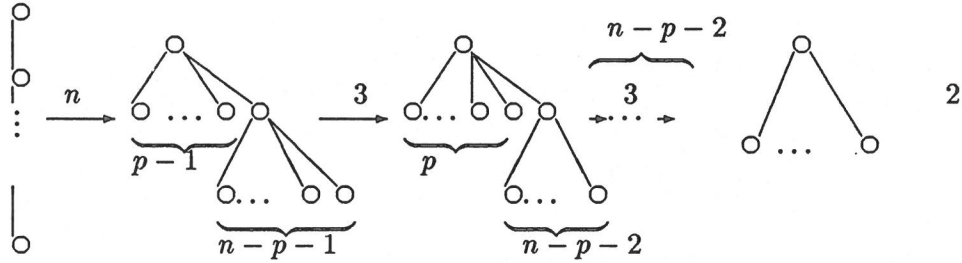


Figure 7: Chemin de coût maximal (Les nombres indiquent les coûts de chaque inversion)

Preuve du théorème 12.

Si $m \leq n - p$

des lemmes 15 et 16, nous déduisons $Ch(n, m) = 3 + \frac{n-3}{m}$.

Si $m > n - p$

montrons par l'absurde que $Ch(n, m) \leq 2 + \frac{2n-p-3}{m}$.

Considérons une suite de m inversions de type p où $m = n - p + q$ avec $q > 0$, d'arbre initial (resp. final) T_0 (resp. T_m) et de coût total $C > 2m + 2n - p - 3$. Supposons que cette suite se compose de q_1 inversions de coût 2, $0 \leq q_1 \leq m$.

Une inversion de coût 2 sur un arbre T redonnant le même arbre T , on peut extraire de cette suite de m inversions une suite de $m' = m - q_1$ inversions $\varphi_1, \dots, \varphi_{m'}$ telles que $c(\varphi_i) > 2$ pour $i = 1, \dots, m'$. Le coût total de cette suite est $\sum_{i=1}^{m'} c(\varphi_i) = C - 2q_1$.

Si $q_1 \geq q$

Nous avons $\sum_{i=1}^{m'} c(\varphi_i) > 3m' + n - 3$, ce qui est impossible d'après le lemme 15.

Si $q_1 < q$

Deux cas se présentent

- 1er cas. Parmi les m' inversions, au moins une est de coût $> p$.

D'après les remarques 13(ii) et (iii), nous avons

$$N^2(T_m) - N^2(T_0) \geq (p-1) + (m'-1) > n-2, \text{ ce qui est impossible.}$$

- 2ième cas. Pour tout $i = 1, \dots, m'$ $2 < c(\varphi_i) < p$.

D'après la remarque 13(ii),

$$N^2(T_m) - N^2(T_0) = \sum_{i=1}^{m'} (c(\varphi_i) - 2) = C - 2m > n - 2,$$

ce qui est impossible.

Ainsi, $Ch(n, m) \leq 2 + \frac{2n-p-3}{m}$.

Le chemin de la figure 7 a un coût moyen égal à $2 + \frac{2n - p - 3}{m}$.

Ainsi, $Ch(n, m) = 2 + \frac{2n - p - 3}{m}$ \square

5 La compression standard

Etant donné un arbre T de racine r et un sommet s de T , l'arbre obtenu par compression standard du chemin $r \rightarrow s$ est l'arbre T' où tous les sommets du chemin ont r pour père. Le coût de cette transformation est égal à la longueur du chemin (k sur la figure 2).

La compression standard sur un arbre à n sommets peut être considérée comme une inversion de type $p \geq n - 1$.

Théorème 17 *Le coût maximal $Ch(n, m)$ d'une suite de m compressions standards sur un arbre à n sommets est $Ch(n, m) = 2 + \frac{n - 2}{m}$.*

Dans ce cas, nous considérons le potentiel d'un arbre T défini par la quantité $\phi(T) = -N^2(T)$.

Lemme 18 *Le coût amorti d'une compression standard φ est $a(\varphi) = 2$*

Preuve du lemme.

Soit φ la compression standard du chemin $x_1 \rightarrow x_k$ sur l'arbre T (figure 2).

Le coût amorti de cette compression est

$$a(\varphi) = k + (-N^2(T_1) - (k - 1)) - (-N^2(T_1) - 1) = 2 \quad \square$$

Preuve du théorème 17.

Etant donné un arbre T à n sommets, son arité est comprise entre 1 et $n - 1$.

Ainsi, $-(n - 1) \leq \phi(T) \leq -1$, et $\phi_0 - \phi_m \leq n - 2$. Des lemmes 2 et 18, nous

déduisons alors que $Ch(n, m) \leq 2 + \frac{n - 2}{m}$.

Cette borne est atteinte. En effet, considérons l'arbre filiforme à n sommets. Si l'on effectue une compression standard du chemin *racine* \rightarrow *feuille* (de coût n), on obtient le "peigne", et toute nouvelle compression standard redonne le peigne (coût=2). Le coût total d'une telle suite de m compressions

standards est $n + 2(m - 1)$, ce qui donne un coût moyen de $2 + \frac{n - 2}{m}$.
Ainsi, $\text{Ch}(n, m) = 2 + \frac{n - 2}{m} \quad \square$

Références

- [1] A.V. Aho, J.E. Hopcroft, J.D. Ullman. "The design and analysis of computer algorithms", Addison-Wesley Reading, Mass. (1974).
- [2] A. Arnold, M.P. Delest, S. Dulucq. Complexité moyenne de l'algorithme d'exclusion mutuelle de Naïmi et Tréhel, Actes du 57ième congrès de l'ACFAS, UQAM, Montréal, 15-19 mai 1989, S. Brlek éd.
- [3] D. Cheriton, R.E. Tarjan. Finding minimum spanning trees, SIAM J. Comput. 5, 4(1976), 724-734.
- [4] M.J. Fisher. Efficiency of equivalence algorithms, "Complexity of computer computations", R.E. Miller, J.W. Thatcher, Eds Plenum Press, New York (1972), 153-168.
- [5] M.J. Fisher, B.A. Galler. An improved equivalence algorithm, Comm. ACM 7, 5(1964), 301-303.
- [6] D. Ginat, D.D. Sleator, R.E. Tarjan. A tight amortized bound for path reversal, Information Processing Letters, 31(1989), 3-5.
- [7] S. Gire. Transformations combinatoires sur les arbres, Mémoire de DEA, Université de Bordeaux I (1990).
- [8] M. Naïmi, M. Tréhel. Un algorithme distribué d'exclusion mutuelle en $\log n$, T.S.I. 6, 2(1987), 141-150.
- [9] D.D. Sleator, R.E. Tarjan. Self adjusting binary search trees, J. ACM 32, 3(1985), 652-686.
- [10] R.E. Tarjan. Amortized computational complexity, SIAM J. Algebraic Discrete Methods 6(1985), 306-318.
- [11] R.E. Tarjan, J. Van Leeuwen. Worst case analysis of set union algorithms, J. ACM 31, 2(1984), 245-281.