

Test de propriétés et BDD

Alpha DIALLO, Lamine KEITA, Sacha MEMMI

Avril 2020



Encadrant : Antoine Genitrini

Université Pierre et Marie Curie

Paris, France

Pstl

Master 1, Semestre 2

Table des matières

1	Abstract	3
2	Introduction	5
3	BDD et ROBDD	9
4	(RO)BDD : propriétés	11
5	Comptage	14
6	Combinatoire sur les squelette	16
7	Unranking	20
8	Quickcheck	25
9	Conclusion	27
10	Appendix	28
10.1	Appendix 1	28

1 Abstract

Les fonction booléennes sont très largement utilisées en science. Elles sont aussi largement utilisées parce que simple à manipuler et à mettre en place leur utilité est également très large. On les utilise par exemple en électronique pour travailler avec des circuits logique.

Les structures avec lesquelles nous travaillons dans ce rapport sont ce que l'on appelle des robdd, reduced ordered binary decision diagram. Elles représentent les fonctions booléennes en tant que graphes (dirigés acycliques). Elles sont l'une des méthode préférée pour représenter les fonctions booléenne en particulier en informatique. Elles sont préférées car facile à implémenter, quelques ligne de codes dans n'importe quel langage sont suffisantes pour les représenter, mais également parce qu'elles sont simple à manipuler.

En particulier nous cherchons à énumérer le nombre de robdd unique représentant une fonction ayant un nombre définie de variable selon leurs tailles. Ce problème a intéressé le domaine scientifique depuis que ces structures ont vu le jour, il présente de nombreux intérêts dans tous les domaines où l'on travail avec des robdd.

Puis en plus de rappeler la méthode standard de génération des robdd, nous donnons un nouveau moyen d'obtenir ces structures. Nous obtenons par cette nouvelle méthode un générateur uniforme de robdd que nous pouvons insérer dans des outils tels que Quickcheck, outil que nous verrons dans ce rapport. Nous utiliserons l'outil Quickcheck sur de petits exemples.

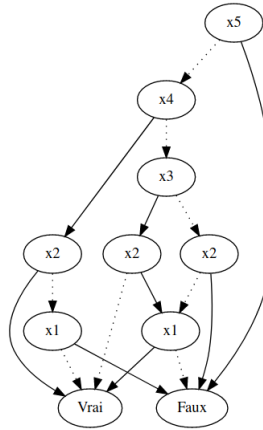


FIGURE 1 – Robdd généré avec la méthode du unranking

Avec notre méthode nous avons pu générer le robdd de taille 10 sur 5 variables de la figure 2.

La méthode de génération de robdd présenter dans ce rapport s'appuie sur celle présenter dans [4], en plus d'implémenter des algorithmes introduits dans ce papier nous en fournissons certains non décrits dans celui-ci [le(s)quel(s) ?]. Y en a-t-il vraiment ? (3) XXX].

2 Introduction

L'idée derrière les arbres de décision binaires a été introduite par Shannon en 1938 dans [8]. C'est en 1959 que Lee introduit l'idée de bdd dans [6].

Les diagrammes de fonction binaires sont principalement des outils pour les ingénieurs (et autre professions travaillant dans le hardware en informatique) qui manipulent des circuit électroniques, cas où l'utilisation de tables de vérité se révèle efficace. Dans ce cas là diviser leurs circuits complexes en sous circuits simple comme le permettent les diagrammes de décision binaires (concept que l'on appelle "diviser pour régner") est très efficace.

Plus particulièrement les circuits électroniques sont devenus si compliqués que les robdd sont à peine suffisant pour les traiter comme l'a noté Knuth il y a de ça plus de 10 ans [5].

Un bdd (binary decision diagram) est un diagramme de décision binaire pouvant représenter une fonction booléenne, l'utilité des bdd ne se limite a représenter des fonctions booléenne mais dans ce rapport nous nous limiterons à celle ci. Par exemple le bdd de la figure 2 peut être utilisé pour représenter la fonction booléenne : $x_3(x_2 + \bar{x}_3) + x_3(\bar{x}_2 + x_1)$, ainsi que la table de vérité ci dessous :

x_3	x_2	x_1	R
⊥	⊥	⊥	⊥
⊥	⊥	⊤	⊤
⊥	⊤	⊥	⊤
⊥	⊤	⊤	⊤
⊤	⊥	⊥	⊥
⊥	⊤	⊤	⊤
⊤	⊤	⊥	⊤
⊤	⊤	⊤	⊤

Un robdd (reduced ordered binary decision diagram) est comme son nom l'indique la forme réduite et ordonné d'un bdd. Dans la littérature le terme bdd est presque toujours utilisé pour évoquer un robdd, nous différencierons tout d'abord entre ces deux termes pour mettre l'accent sur l'aspect réduction. Les bdd que nous considérons dans ce rapport sont toujours ordonnés, quand on parle de bdd nous faisons en fait référence à un obdd (ordered binary decision diagrams).

Les robdd n'ont été introduit qu'en 1986 [2]. Assez tard dans l'histoire de l'informatique. Les robdd ont été une telle révolution que [2] est pendant des années resté le rapport le plus cité dans la littérature scientifique comme l'a noté Knuth :

"His (Randal E. Bryant's) introduction to the subject [IEEE Trans. C-35 (1986),677-691] became for many years the most cited papers in all of computer science"

De nombreuses autres variantes de robdd ont également été créées depuis [9].

Le principe d'un robdd est simple, on cherche à conserver de l'espace mémoire en se débarrassant des structures redondantes, la mémoire étant une préoccupation importante en informatique. Pour réduire un bdd, nous supprimons les sous structures redondantes, nous supprimons également les noeuds non essentiels. Nous représentons les bdd en tant qu'arbre plat et les robdd en tant que dag (directed acyclic diagrams) [7]. Plus généralement les robdd sont des sous types de bdd et tous les types de bdd peuvent être représentés par un dag.

“A Boolean function is a function in mathematics and logic whose arguments, as well as the function itself, assume values from a two-element set (usually $\{0,1\}$).”

- Wikipedia

Il existe de nombreux moyens pour représenter des fonctions booléennes. Ceux préférés en science de l'informatique sont les diagrammes, pas seulement grâce à la manière dont ces diagrammes peuvent être réduits, mais aussi parce qu'ils peuvent être conservés en mémoire en tant que multiples noeuds qui ne sont pas forcément adjacents en mémoire les uns aux autres. Ceci rend leur conservation plus simple et efficace.

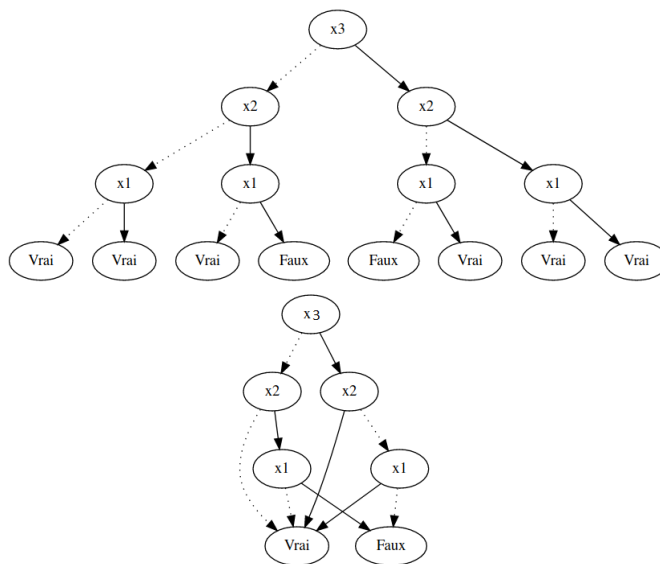


FIGURE 2 – Arbre de decision binaire en haut et dag représentant sa forme réduite en bas

Un exemple d'arbre de décision binaire et un dag représentant sa forme réduite sont donnés dans la figure 2. Nous suivons ici le consensus dans la littérature

et représentons les liens vers les fils haut par une ligne solide, et ceux vers les fils bas par une ligne en pointillé. Le fils haut d'un noeud est le noeud successeur de celui marqué par un arc 1 (la valeur de transition vers ce noeud dans la fonction de transition est 1), et le fils bas est le noeud successeur marqué par un arc 0 (cette fois ci la valeur dans la fonction de transition est 0).

En plus de l'utilité évidente des fonctions booléennes on peut aussi noter que les bdd sont devenus aussi populaire parce que comme Knuth l'a noté dans [5] les bdd simples sont relativement facile à fusionner pour obtenir des bdd plus complexes.

Dans ce rapport notre but est de présenter un nouveau procédé permettant d'énumérer des robdd sans passer par le processus de compression.

Le processus de compression est relativement simple avec une complexité polynomiale, mais la méthode standard de génération des bdd d'un nombre donné de variable k utilisé jusqu'ici était impraticable pour $k > 4$ [3]. Pour construire les bdd ayant k variable dans la méthode standard nous construisons tout d'abord toutes les fonctions booléennes ayant k variables, puis leurs bdd, puis compressons ces bdd afin d'obtenir leurs robdd et enfin nous supprimons les robdd apparaissant plus d'une fois. Ceci avait une complexité doublement exponentielle de $O(2^{2^k})$.

Nous définissons la taille d'un robdd par le nombre de noeuds internes (non terminaux) qu'il contient.

Pour énumérer ces structures dans la méthode que nous introduisons ici nous utilisons de nombreuses propriétés des robdd. Nous utilisons ensuite des méthodes dites de unranking et ranking basées sur le modèle standard introduit dans [1] en définissant un ordre total sur les robdd et en décomposant chacun par une constitution de sous robdd.

Le processus de ranking consiste à attribuer à chaque objet d'un ensemble un identifiant unique, celui du unranking correspond au processus inverse, retrouver depuis son identifiant un objet au sein d'un ensemble.

Nous obtenons grâce à cette méthode un générateur uniforme de robdd.

Avec la méthode standard utilisée dans [3] les robdd de taille proche de la taille maximale étaient créés avec une grande probabilité, mais dans celle ci chaque robdd a un identifiant unique dans l'ensemble. On peut également modifier ce générateur pour ne créer que des robdd ayant certaine classes de propriétés.

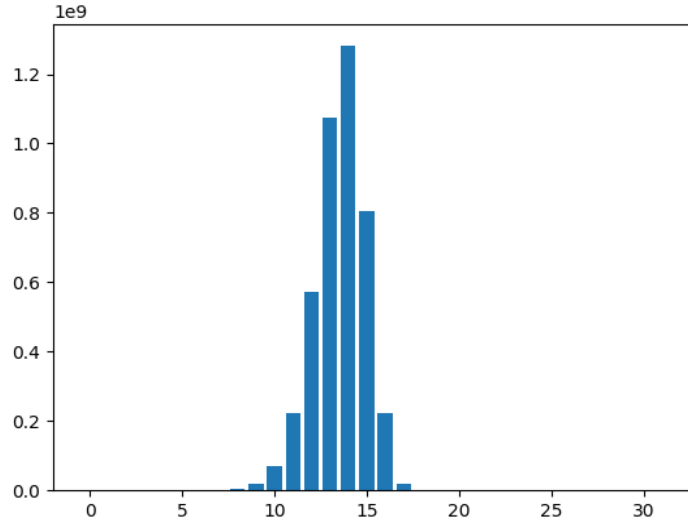


FIGURE 3 – Distribution des taille de bdd ayant 5 variable

Avec notre méthode nous arrivons à comptabiliser les bdd ayant 5 variables, avec une implantation en python et un matériel limité, en 14 secondes.

Nos résultats sont similaires à ce qui était estimé dans [3] pour 5 variables, comme on peut le voir sur la figure 3 (on rappelle qu'ici on ne prend pas en compte les feuille dans le calcul de la taille d'un bdd).

Dans ce rapport nous considérerons que nous ne rencontrons jamais le robdd réduit au noeud unique \top ou \perp .

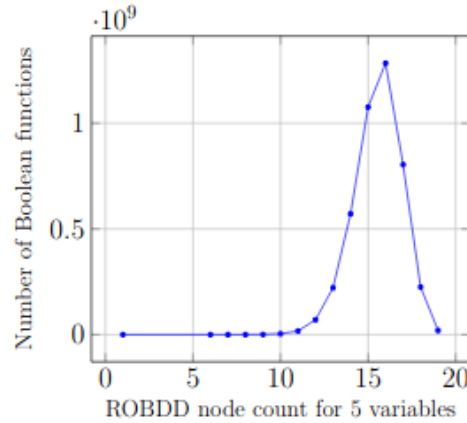


FIGURE 4 – Distribution des taille de bdd ayant 5 variable selon [3]

3 BDD et ROBDD

Dans cette section nous cherchons à définir les structures basiques avec lesquelles nous travaillons, les bdd et les robdd.

Un bdd est composé de noeuds ayant un label. Chaque noeud non terminal possède deux enfants, ou fils, un bas et un haut liés à leurs parents par des liens notés respectivement 0 et 1. Chaque noeud a un index qui détermine son ordre dans le bdd. Les noeuds sont traités dans l'ordre décroissants de leurs indexes.

La manière dont nous ordonnons les noeuds est importante comme noté dans [3] et trouver l'ordre le plus efficace est un problème NP (nondeterministic polynomial time).

La racine a un index k qui est l'index maximal de la bdd (ou robdd), et les noeuds terminaux ont des index nuls. Tous les noeuds internes (quand ils existent) ont des indexes allant de k à 0, non inclus. Pour un bdd l'index de la racine est le nombre de variables de la fonction booléenne qu'il représente, pas forcément pour un robdd.

Un bdd associe a chaque affectation de variables une valeur de retour, pour connaître cette valeur nous allons tout d'abord à la racine, puis nous suivons un chemin en se fiant à la valeur de chaque variable données en entrée selon l'index du noeud courant. Si la variable d'index du noeud courant est 0 nous nous rendons à l'enfant bas du noeud courant, l'enfant haut si la variable a pour valeur 1. Nous finissons par arriver à un noeud terminal, que dans le reste de ce rapport nous appellerons feuille, de valeur \top (true) ou \perp (false) qui est la valeur de retour de la bdd pour les variables considérées.

Pour le bdd de la figure 2 si l'on donnait les variables 101 en tant qu'entrée elle retournerait \top (true).

La méthode classique pour obtenir un robdd est de réduire un arbre de décision binaire. Définissons maintenant comment on effectue cette réduction.

Pour effectuer cette réduction nous suivons tout d'abord deux règles, que nous notons R et M.

Considérons Δ un dag représentant une fonction booléenne f . Considérons également α et β deux noeuds distincts dans Δ .

Règle M : Si α et β sont des racine de sous graphes isomorphes alors fusionner β et α .

Règle R : Si les deux fils de β sont isomorphes à α , alors faire pointer tous les liens entrant de β vers α et supprimer β .

Nous appliquons récursivement ses deux règles à tous les noeuds de Δ .

Deux ordres de parcours peuvent être considérés, $post_o$ et pre_o .

Dans le $post_o$ nous visitons tous d'abord le fils bas du noeud courant, puis son fils haut et enfin le noeud lui-même. Dans le pre_o nous visitons tous d'abord le noeud courant, puis son fils bas, et enfin son fils haut.

Nous choisissons ici de considérer l'ordre de traversé $post_o$ pour effectuer chacune de nos compressions, puisque c'est celui préféré en science de l'informatique. Choisir un ordre en particulier est important pour s'assurer que des bdd équivalents soient compressés en tant que robdd identique une fois l'ordre des variables établie.

Nous pouvons désormais définir formellement le processus de compression.

Compression : Considérons Δ le bdd de la fonction f . Quand le noeud λ est visité, λ étant l'enfant d'un noeud μ . Si un sous-arbre Λ , celui qui a pour racine λ , a déjà été vu lors de la traversé de Δ , sous arbre dont la racine est ω , alors Λ est supprimé de Δ et le noeud μ obtient un pointeur vers ω , remplaçant le lien allant précédemment de μ à λ . Une fois que la traversé de Δ achevée, le dag résultant est le robdd de f .

Voir Appendix 10.1 pour l'algorithme de compression.

Définissons maintenant formellement un robdd, cette définition ainsi que les notations que l'on utilise se basent sur celles définies dans [4].

Considérons un robdd $\Delta = (Q, I, r, \delta)$. Q est l'ensemble de noeuds que l'on obtient par une traversé en $post_o$ de Δ . r est la racine de Δ , I est la fonction qui associe pour tout noeud $\lambda \in Q$ son index et δ est la fonction de transition complète qui associe à chaque noeuds ses fils.

- $Q = ([feuille1], [feuille2], \dots, r)$ ($[feuille1]$ étant la première feuille que l'on rencontre lorsque l'on parcourt Δ en $post_o$ et $[feuille2]$ la seconde)
- $I = Q \rightarrow \{0, \dots, k\}$ où k est l'index de r
- $\delta = (Q \setminus \{\top, \perp\}) * \{0, 1\} \rightarrow Q$

On note que Q a une taille forcément supérieure ou égale à 3.

On peut ainsi représenter le dag de la figure 2 par un robdd $\Delta = (Q, I, r, \delta)$ ayant les propriétés suivantes :

- $Q = (\perp, \top, x1, x1, x2, x2, x3)$

- $I = \{\perp, \top, x1, x1, x2, x2, x3\} \rightarrow \{0, 0, 1, 1, 2, 2, 3\}$
- $\delta = \{x1 * 0 \rightarrow \perp, x1 * 1 \rightarrow \top, x1 * 0 \rightarrow \top, x1 * 1 \rightarrow x2 * 0 \rightarrow x1, x2 * 1 \rightarrow \top, x2 * 0 \rightarrow \top, x2 * 1 \rightarrow x1, x3 * 0 \rightarrow x2, x3 * 1 \rightarrow x2\} Q$
- $r = x3$

Dans le reste de ce rapport nous utiliserons le terme bdd pour faire référence aux robdd.

4 (RO)BDD : propriétés

Dans cette section nous définissons des propriétés que l'on assigne aux bdd. Pour ce faire nous mettons en place un ensemble de notation et d'opérations. Toutes les notations utilisées dans cette section seront utilisées dans le reste de ce rapport. Les opérations que nous définissons pour les listes sont également applicables aux ensembles. On se contente ici de redéfinir les notations définies dans [4].

Pour deux noeuds α et β d'un dag Φ nous écrivons $\alpha \Leftarrow_{post} \beta$ si le noeud α est visité avant le noeud β en $post_o$ lors de la traversée de Φ , et $\alpha \Leftarrow_{pre} \beta$ si le noeud α est visité avant le noeud β en pre_o lors de la traversée de Φ .

Pour deux listes l et l' de tailles respectives m et n , où $l = (p0, p1, \dots, pm)$ et $l' = (p'0, p'1, \dots, p'n)$ nous écrivons $L = l + l'$ l'opération telle que si $m < n$ avec $L = (P0, P1, \dots, Pn)$ alors nous avons pour $i \leq m$ $Pi = pi + p'i$ et pour $m < i \leq n$ $Pi = p'i$.

Pour une liste l nous notons $|l|$ le nombre d'élément dans l , et $\|l\|$ la somme de tous ses éléments.

Profil d'un ensemble : Le profil $P = \text{profil}(S)$ d'un ensemble S est simplement une liste du nombre d'élément de S ayant un certain index, on note $P = (p0, p1, \dots, pj, \dots, pk)$ où pj est le nombre d'éléments dans l'ensemble S d'index j et k l'index maximal de S .

Le processus d'extension de cette définition de profil à des graphes est évident.

Squelette d'un bdd : Considérons un bdd $\Delta = (Q, I, r, \delta)$ enraciné en r . Le squelette $T = (Q', I, r, \delta')$ de Δ est un dag tel que $Q' = Q / \{\top, \perp\}$ où chaque noeud $\lambda \in Q'$ a été obtenu par la traversée en $post_o$ de Δ en faisant omission des deux feuilles. Les liens du squelette sont décrit en utilisant une fonction de transition partielle $\delta' : Q' \times \{0, 1\} \rightarrow Q' \cup \{\text{nil}\}$ où nil est un symbole spécial désignant une transition non définie.

Décrire simplement un squelette est un bdd auquel on retire les deux feuilles et les pointeurs.

Dans le reste de ce rapport nous nous référerons aux valeurs non définies de la fonction de transition partielle en tant que nil transition.

Nous notons n le nombre de noeud d'un dag Δ d'un bdd, nous savons que son squelette S a le même nombre de noeud que Δ à l'exception des deux feuilles. Nous pouvons ainsi noter que S a $(n-2)$ noeuds. Nous savons également que tous les noeuds non terminaux de Δ ont deux enfants, ainsi le nombre de transition dans Δ est $(n-2)*2$, ce qui est le nombre total de transition dans S (à la fois les transition nil et non nil). Nous savons également que chaque noeud $\lambda \in S$ a nécessairement un degré entrant de 1 à l'exception de la racine ainsi le nombre restant de transition non nil dans S est $(n-3)$. Nous pouvons maintenant conclure que S a un nombre de nil transition égal à $(n-1)$.

Sous arbre : Un sous arbre $TA(\lambda)$ est un arbre enraciné en λ appartenant à un autre arbre.

Pool : Considérons un dag D , la pool $PT(\lambda)$ d'un noeud $\lambda \in D$ est :

$$PT(\lambda) = \{\tau' \in D \mid \tau' \Leftarrow_{pre} \lambda \text{ et } I(\tau') < I(\lambda)\} \cup \{\top, \perp\} .$$

Pool profil : Le pool profil de λ est comme son nom l'indique le profil de sa pool que l'on note : $profil(PT(\lambda))$.

Level set : Considérons un dag D , le level set $\lambda \in D$ est :

$$ST(\lambda) = \{\theta' \in D \mid \theta' \Leftarrow_{pre} \lambda \text{ et } I(\theta') = I(\lambda)\} .$$

Level rank : Considérons $ST(\lambda)$ le level set du noeud λ , nous écrivons $RS(\lambda)$ son sibling rank tel que :

$$RS(\lambda) = |ST(\lambda)|$$

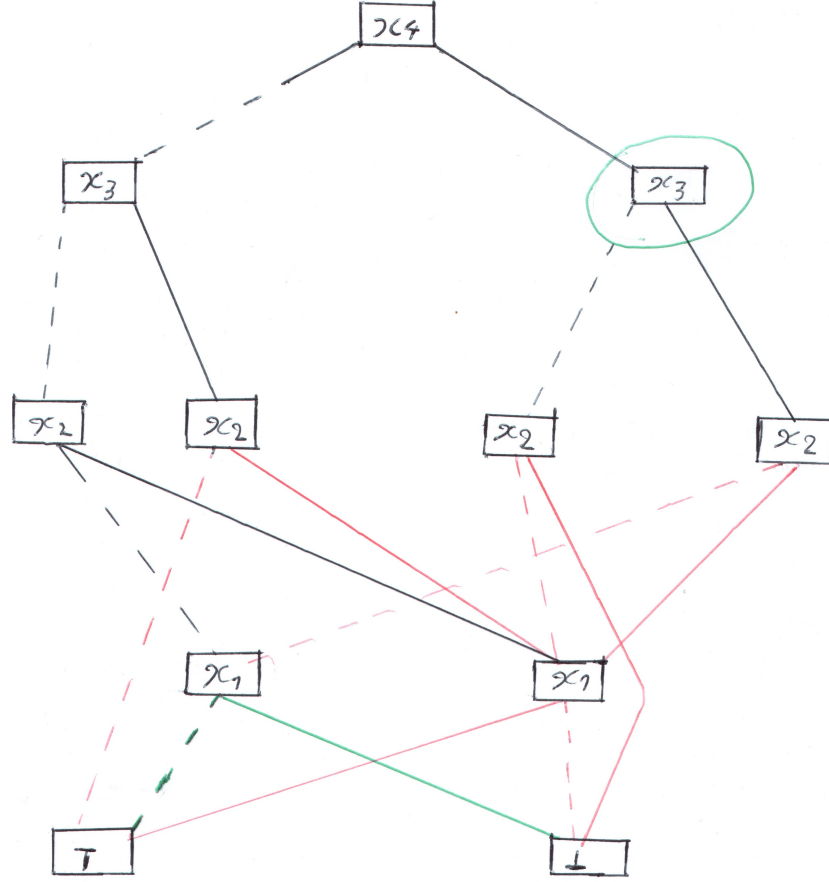


FIGURE 5 – Dag

On considère dans la figure 3 un dag. Les pointeurs sont représentés en rouge. Le squelette du robdd de ce dag est le graphe réduit aux transitions en noirs. La pool du noeud entouré en vert est : $\{x_2, x_2, x_1, x_1, \top, \perp\}$. Le profil de ce bdd est : $\{2, 2, 4, 2, 1\}$.

Nous savons de la manière dont nous définissons les bdd que pour être valide un bdd doit avoir les propriétés suivantes :

FACT 1.1 : *De la règle r nous faisons la conclusion que deux noeuds de même index ne peuvent avoir les mêmes descendants, autrement ils auraient été fusionnés.*

FACT 1.2 : *A partir de la règle m nous savons qu'un noeud ne peut avoir deux enfants identiques, autrement ce noeud aurait été supprimé.*

5 Comptage

La méthode classique permettant d'énumérer les bdd ayant un nombre de variable k et une taille n est la suivante :

1. Énumérer toutes les 2^{2^k} fonctions booléennes et dresser leurs arbres de décision binaires
2. Appliquer le processus de compressions pour obtenir leurs bdd
3. Filtrer les bdd de taille n
4. Supprimer les bdd réapparaissant plus d'une fois afin de ne conserver qu'une seule de leurs occurrences

Nous proposons ici une méthode récursive permettant d'effectuer cette énumération sans passer ni par le processus de compression ni celui de filtration. Ces deux processus sont hautement coûteux avec une complexité doublement exponentielle $O(2^{2^k})$.

Proposition 1 : Considérons $T = (Q', I, r, \delta')$ un squelette avec un ensemble de noeuds Q' , racine r et fonction de transition partiel $\delta' : Q' \times \{0, 1\} \rightarrow Q' \cup \{\text{nil}\}$. La fonction de transition entière $\delta : Q' \times \{0, 1\} \rightarrow Q' \cup \{\top, \perp\}$ est la fonction de transition d'un bdd avec un squelette T si et seulement si pour tout noeuds $\lambda \in Q'$, noté $\lambda 0 = \delta(\lambda, 0)$ et $\lambda 1 = \delta(\lambda, 1)$, la paire $(\lambda 0, \lambda 1)$ satisfait :

1. Si $\delta'(\lambda, 0) \neq \text{nil}$ et $\delta'(\lambda, 1) \neq \text{nil}$ alors

$$\lambda \sigma = \delta'(\lambda, \sigma) \text{ pour } \sigma \in \{0, 1\}.$$

2. Si $\delta'(\lambda, 0) = \delta'(\lambda, 1) = \text{nil}$, alors

$$\lambda \sigma \in \text{PT}(\lambda) \text{ pour } \sigma \in \{0, 1\} \text{ et } \lambda 0 \neq \lambda 1, \text{ et il n'y a aucun noeud } \lambda' \neq \lambda \text{ avec } I(\lambda) = I(\lambda') \text{ tel que } \delta(\lambda', \cdot) = \delta(\lambda, \cdot).$$

3. Si $\delta'(\lambda, 0) = \text{nil}$ et $\delta'(\lambda, 1) \neq \text{nil}$, alors

$$\lambda 0 \in \text{PT}(\lambda) \text{ et } \lambda 1 = \delta'(\lambda, 1).$$

4. Si $\delta'(\lambda, 0) \neq \text{nil}$ et $\delta'(\lambda, 1) = \text{nil}$, alors

$$\lambda 0 = \delta'(\lambda, 0) \text{ et } \lambda 1 \in \text{PT}(\lambda) \cup \text{TA}(\lambda 0) \setminus \lambda 0.$$

Preuve. Puisque $\delta(\cdot, \cdot)$ doit étendre $\delta'(\cdot, \cdot)$, le cas 1. est triviale puisque nous devons étendre la fonction de transition uniquement où l'on a $\delta(\lambda, \sigma) = \text{nil}$. Dans le cas 2., nous devons choisir pour $(\lambda 0, \lambda 1)$ deux noeuds dans $PT(\lambda)$. De plus $\lambda 0 \neq \lambda 1$ comme déclaré dans FACT 1.2 et un autre noeud avec le même index que λ ne peut pas avoir les mêmes enfants $(\lambda 0, \lambda 1)$ comme déclaré dans FACT 1.1. Dans le cas 3., l'enfant bas doit être choisie dans la pool de λ puisque l'on doit préserver le squelette. Dans le cas 4., l'enfant haut de λ est aussi choisi dans la pool de λ ou dans $TA(\lambda 0)$ (et doit être différent de $\lambda 0$).

Nous définissons maintenant un procédé permettant de calculer le nombre de bdd correspondant à un squelette donné, nous appellerons le nombre de bdd correspondant à un squelette S le poids de S .

Poids d'un noeud : Considérons un squelette $T = (Q', I, \delta', r)$, le poids $wt(\lambda)$ d'un noeud $\lambda \in Q'$ est le nombre de transition possible pour compléter la fonction de transition partielle $\delta'(\lambda, \cdot)$ et former un bdd valide de squelette T .

Proposition 2 (Poids d'un noeud) : Considérons $T = (Q', I, \delta', r)$ un squelette, le poids $wt(\lambda)$ d'un noeud $\lambda \in T$ est :

$$wt(\lambda) = \begin{cases} 1 & \text{si } \delta'(\lambda, 0) \neq \text{nil et } \delta'(\lambda, 1) \neq \text{nil} \\ pts(\lambda)(pts(\lambda) - 1) - ST(\lambda) & \text{si } \delta'(\lambda, 0) = \delta'(\lambda, 1) = \text{nil} \\ pts(\lambda) + pts0(\lambda) - 1 & \text{si } \delta'(\lambda, 0) = \text{nil et } \delta'(\lambda, 1) \neq \text{nil} \\ pts(\lambda) & \text{si } \delta'(\lambda, 0) \neq \text{nil et } \delta'(\lambda, 1) = \text{nil} \end{cases} \quad (1)$$

Avec $pts(\lambda) = ||profil(PT(\lambda))||$ et $pts0 = ||profil(TA(\delta'(\lambda, 0)))||$.

Preuve. A partir de la proposition 1 nous énumérons tous les liens possibles (liant à un noeud) nous pouvons utiliser pour compléter la fonction de transition partielle. Dans le premier cas nous ne pouvons compléter que avec les liens déjà existant (comme déclaré dans 1) de proposition 1) le cas est déjà définie). Dans le second cas nous énumérons tous les liens possible de 2) de la proposition 1. Dans le troisième et quatrième cas nous faisons la même chose pour respectivement 3) et 4).

Le poids d'un sous arbre est le poids cumulé de ses noeuds.

Poids d'un squelette : Considérons un squelette $T = (Q', I, \delta', r)$, le poids $W(T)$ de T est le poids cumulé de ses noeuds.

$$W(T) = \prod wt(\lambda) \mid \lambda \in Q'$$

Dans le cas où le poids d'un noeud $\lambda \in Q'$ est nul ou négative nous savons que le squelette T est invalide parce que T n'est pas le squelette du moindre bdd. Ceci arrive uniquement lorsque λ a ses deux enfants égaux à nil (puisque l'on a nécessairement $||profil(PT(\lambda))|| \geq 2$).

Cette formule peut être aisément décrite de manière récursive, via une traversée en $post_o$ où toutes les informations nécessaires pour calculer le poids du noeud ont déjà été traités.

Nous notons T_{nk} l'ensemble des squelette de taille n et de nombre de variable k , et N_{nk} le nombre de bdd de taille n et de nombre de variable k . Pour déterminer N_{nk} tout ce que nous avons à faire est déterminer la totalité des squelette de taille n et de nombre de nombre de variable k et ensuite déterminer leurs poids, nous faisons ensuite l'addition de leurs poids.

$$N_{nk} = \sum W(T) \mid T \in T_{nk}$$

6 Combinatoire sur les squelette

Ce que nous voulions dans la section précédente était d'énumérer tout les bdd de taille n et de nombre de variable k . Pour ce faire nous devons déterminer les squelette de taille n et de nombre de variable k . Ceci n'est un processus aisé à informatiser.

Nous définissons ici un processus récursive où nous construisons les squelette par chacun de leurs noeuds et des que l'on a déterminé que ses squelette ne sont pas valides on les ignore.

Nous définissons un squelette Tt en tant que tuple tel que :

$$Tt = T' \cup N_i \cup T'' \mid nil$$

où N_i est un noeud d'index i , ceci signifie que le squelette est soit réduit à nil soit à une racine et deux squelette. Nous designerons dans la suite les squelette T' et T'' qui composent Tt par le terme de sous squelette.

Pour considerer un squelette enraciné en r en particulier nous prendrons en compte trois de ses paramètres. Sa pool $pT(T)$, le level rank de sa racine $ST(r)$ et sa taille que nous designerons par $\|T\|$.

Ces paramètre sont important parce qu'ils sont ce dont on a besoin pour déterminer le poids des squelette que nous considérons.

Nous utiliserons la notation $T_{m,p,s}$ pour considérer le squelette enraciné en r de taille m ($m = \|T\|$), de pool profil p ($p = \text{profil}(pT(r))$) et de level rank s ($s = ST(r)$) représenté par un tuple Tt tel décrit plus haut.

Proposition 3 : Pour $T_{m,p,s}$ nous considérons des tailles pour chaque sous squelette $T' \cup T'' \in T$ de respectivement i et $m - 1 - i$, avec $0 \leq i \leq m - 1$. Pour chacune des tailles de sous squelette considérés nous considérons chacune des valeurs d'index que l'on peut assigner à leurs racines.

Nous avons ainsi un algorithme où nous considérons toute les tailles possible ainsi que les indexes des sous-squelette, le level rank de la racine étant déjà donné par le pool profil du squelette courant (ainsi que par le profil de T' dans le cas du level rank de T''). k correspond dans la suite à la taille de p .

1. Si $m > 0$
 - (a) $T' \in \bigcup_{k_0 \in K} T_{i, p[:k_0-1], p_{k_0-1}}$
 - (b) $T'' \in \bigcup_{k_0 \in K} T_{j, p1[:k_0-1], p1_{k_0-1}}$
 - (c) $W(T) = WTP(N_k, T', T'') * W(T') * W(T'')$
2. Si $m = 0$
 - (a) $T = \text{nil}$
 - (b) $W(T) = 1$

Avec $j = m-1-i$, on note $p[:i] = (p_0, p_1 \dots p_i)$ pour designer dans un ensemble p que l'on ne considère que les éléments ayant un index plus petit ou égal à i et p_k lorsqu'on désigne l'élément d'index k . On a également $p1 = p + \text{profil}(\text{TA}(T'))$ et $K = \{0, 1, \dots, k\}$. Enfin on note $W(T)$ le poids du squelette T et $WTP(N, T', T'')$ le poids du noeud N qui a pour fils bas le squelette T' et pour fils haut le squelette T'' .

Avec l'algorithme nous ne considérons pour chaque récursion que les squelette de même valeurs m , p et s . En s'inspirant de ce procédé on peut écrire un algorithme $\text{COUNT}(m, p, s)$ prenant en argument une taille de squelette m , un profil p et un level rank s et retournant un dictionnaire où chaque clé est le profil d'un squelette et la valeur la somme des poids des squelette ayant ce profil.

Quelques exemples d'exécution de la fonction COUNT :

- $\text{COUNT}(3, (2, 0, 0), 0)$ nous retourne $\{(0, 1, 1, 1) : 56, (0, 2, 0, 1) : 2, (0, 0, 2, 1) : 2\}$
- $\text{COUNT}(2, (2, 0), 0)$ nous retourne $\{(0, 1, 1) : 8\}$
- $\text{COUNT}(1, (2), 1)$ nous retourne $\{(0, 1) : 1\}$
- $\text{COUNT}(1, (2), 0)$ nous retourne $\{(0, 1) : 2\}$

Dans le premier cas le dictionnaire retourné nous indique qu'il existe 56 squelette unique de taille 3 ayant pour profil $(0, 1, 1, 1)$, 2 ayant pour profil $(0, 2, 0, 1)$ et encore 2 ayant pour profil $(0, 0, 2, 1)$.

S'appuyant sur l'algorithme COUNT on définit $N(n, k)$ qui renvoie le nombre de bdd de taille n et d'index k :

$$N(n, k) = \sum_{(t, w) \in \text{COUNT}(n-2, (2, 0, \dots, 0), 0)} w$$

$e^{(k)}$ est une liste de $k+1$ composants tous nuls sauf le dernier qui est égal à 1 tel que $e^{(k)} = (0, 0, \dots, 1)$.

La complexité de l'algorithme COUNT est exponentiel de l'ordre de $O(2^{3k^2/2+k})$.

L'algorithme COUNT s'appuie sur la méthode de comptage et de génération des squelettes que l'on a décrit plus haut. Cet algorithme vérifie tout d'abord si le noeud courant est vide ($m=0$) si tel est le cas on sait qu'on est sur une nil-transition. Si le noeud courant n'est pas vide ($m>0$) on génère la totalité

des sous arbre gauche et droit et déterminons récursivement leurs poids. Nous stockons leurs poids dans un dictionnaire selon leurs profil.

On décrit l'algorithme COUNT en pseudo code ci-dessous.

```

1: function COUNT(m,p,s)
2:    $d \leftarrow \{\}$ 
3:    $k \leftarrow \text{Len}(p)-1$  ▷ Len(p) : renvoie la taille de la liste p
4:   if m=0 then
5:      $S \leftarrow (\sum_{j=0}^{k-1} p_j)(\sum_{j=0}^{k-1} p_j - 1)$ 
6:     if  $S > 0$  then
7:        $d \leftarrow \{e^{(k)} : S\}$ 
8:     end if
9:   else
10:    for  $i \leftarrow 0$  to m-1 do
11:       $d_0 \leftarrow \{\}$ 
12:      if i=0 then
13:         $d_0 \leftarrow \{() : \sum_{i=0}^{k-1} p_i\}$ 
14:      else
15:        for  $k_0 \leftarrow 1$  to k-1 do
16:           $d_0 \leftarrow d_0 \cup \text{COUNT}(i, p[: k_0 - 1], p_{k_0})$ 
17:        end for
18:      end if
19:      for  $(l, w_0) \leftarrow d_0$  do
20:         $d_1 \leftarrow \{\}$ 
21:         $p' \leftarrow p + l$ 
22:        if n-1-i=0 then
23:           $d_1 \leftarrow \{() : -1 + \sum_{i=0}^{k-1} p'_i\}$ 
24:        else
25:          for  $k_1 \leftarrow 1$  to k-1 do
26:             $d_1 \leftarrow d_1 \cup \text{COUNT}(n-1-i, p'[: k_1 - 1], p'_{k_1})$ 
27:          end for
28:        end if
29:        for  $(r, w_1) \leftarrow d_1$  do
30:           $w \leftarrow w_1 * w_0$ 
31:           $t \leftarrow l + r + e^k$ 
32:          if  $t \in d$  then
33:             $d[t] \leftarrow d[t] + w$ 
34:          else
35:             $d \leftarrow d \cup \{t : w\}$ 
36:          end if
37:        end for
38:      end for
39:    end for
40:  end if
41:  return d
42: end function

```

7 Unranking

Notre approche pour le unranking s'appuie sur la méthode introduite dans [1] où l'on décompose un arbre binaire en ses sous arbre gauche et droits et en attribuant à chaque décomposition un ordre total pour un bdd de taille et d'index fixé.

En s'inspirant de la Proposition 3 on définit une méthode récursive b permettant de déterminer le nombre d'arbres binaires ayant une certaine taille n .

$$b_n = \begin{cases} 1 & \text{si } n = 0 \\ \sum_{k=0}^{n-1} b_k * b_{n-1-k} & \text{sinon} \end{cases} \quad (2)$$

En s'appuyant sur cette méthode on peut définir une décomposition sur un arbre de taille fixé n . Cette méthode s'appuie sur le fait que b_n est définie en une somme de produit, où chaque élément du produit peut être vu comme la représentation d'un arbre ayant un sous arbre gauche et un sous arbre droit de taille fixe. Par exemple $B_7 = 429$, et on peut le décomposer en :

$$B_7 = B_0 * B_6 + B_1 * B_5 + B_2 * B_4 + B_3 * B_3 + B_4 * B_2 + B_5 * B_1 + B_6 * B_0$$

$$429 = 1 * 132 + 1 * 42 + 2 * 14 + 5 * 5 + 14 * 2 + 42 * 1 + 132 * 1$$

Où $B_2 * B_4$ représente un arbre composé d'un sous arbre droit de taille 4 et d'un sous arbre gauche de taille 2.

Ainsi, si je cherche l'arbre n°250 de taille 7, j'obtiens : $250 = 132 + 42 + 28 + 25 + 23$. On trouve donc que l'arbre 250 est composé d'un enfant gauche de taille 4 et d'un enfant droit de taille 2. Et plus particulièrement, parmi les 28 arbres ayant cette décomposition 4x2, c'est l'arbre numéro 23 qui nous intéresse.

Pour ensuite retrouver le numéro des sous arbre gauche et droit on procède comme ceci :

- $23 // B_2$ pour le numéro du sous arbre gauche
- $23 \% B_2$ pour le numéro du sous arbre droit

En s'appuyant récursivement sur cette méthode on obtient l'arbre de la figure 6.

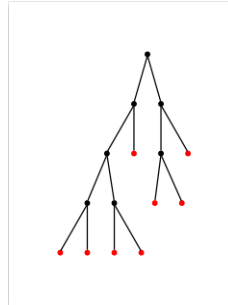


FIGURE 6 – Arbre de taille 7 et de rang 250 généré

En s'inspirant de cette methode on definie une fonction $\text{DECOMPOSE}(r', n', s, p')$ permettant de generer un bdd de taille n' , ayant un profil p' et un rang r' . s est une variable contenant le squelette du bdd a generer.

[XXX donner explication methode decompose]

```

1: function DECOMPOSE(rank, n, squelette, p)
2:   p_target  $\leftarrow$  list(p)
3:   r  $\leftarrow$  rank
4:   s  $\leftarrow$  squelette.get_profil()
5:   k  $\leftarrow$  len(p_target) - 1
6:   q  $\leftarrow$  s[k]
7:   q[0]  $\leftarrow$  2
8:   i  $\leftarrow$  n - 1
9:   while i  $\geq$  0 do
10:    if i = 0 then
11:      d0  $\leftarrow$  () :sum(q, 1, k - 1) + 2
12:    else
13:      d0  $\leftarrow$  {}
14:      for k0  $\leftarrow$  range(1, k) do
15:        d0.update(count(i, q[k0], q[k0]))
16:      end for
17:    end if
18:    for (l, w0) in d0.items() do
19:      q_prim  $\leftarrow$  sum_profil(q, l)
20:      if n - 1 - i = 0 then
21:        d1  $\leftarrow$  {() :sum(q_prim, 1, k - 1) + 1}
22:      else
23:        d1  $\leftarrow$  {}
24:        for k1 in range(1, k) do
25:          d1.update(count(n-1-i, q_prim[k1], q_prim[k1]))
26:        end for
27:      end if
28:      for (h, w1) in d1.items() do
29:        t  $\leftarrow$  sum_profil(sum_profil(e_k(k), list(l)), list(h))
30:        w  $\leftarrow$  w0 * w1
31:        if t = p_target then
32:          if w > 0 then
33:            r  $\leftarrow$  r - w
34:          end if
35:          if r < 0 then
36:            r  $\leftarrow$  r + w
37:            r0  $\leftarrow$  r % w0
38:            r1  $\leftarrow$  r // w0
39:            return (i, r0, l, r1, h)
40:          end if
41:        end if
42:      end for
43:    end for
    i  $\leftarrow$  i - 1

```

On décrit ensuite une fonction $\text{GENERATE}(r,n,s,p)$ permettant de générer un squelette de taille n , de profil p et de rang r . La variable s est une variable globale qui contiendra en fin d'exécution le bdd généré. Cette fonction prend en argument un squelette s contenant les noeuds initialement dans la pool (ici $et\perp$) et crée au fur et à mesure le bdd souhaité en faisant appel à la fonction decompose et en mettant à jour le squelette.

```

1: function  $\text{GENERATE}(rank, n, \text{squelette}, p\_target)$ 
2:    $k \leftarrow \text{len}(p\_target) - 1$ 
3:   if  $n = 0$  then
4:      $node \leftarrow \text{squelette.unrank\_singleton}(rank)$ 
5:     return  $node$ 
6:   end if
7:   if  $n = 1$  then  $lo, hi \leftarrow \text{squelette.unrank\_pair}(rank, k)$ 
8:   else
9:      $i, r0, l, r1, h \leftarrow \text{DECOMPOSE}(rank, n, \text{squelette}, p\_target)$ 
10:     $lo \leftarrow \text{GENERATE}(r0, i, \text{squelette}, l)$ 
11:    if  $n - 1 - i = 0$  then  $hi \leftarrow \text{squelette.unrank\_singleton}(r1)$ 
12:      if  $hi = lo$  then  $hi \leftarrow \text{GENERATE}(r1 - 1, n - 1 - i, \text{squelette}, h)$ 
13:    end if
14:    else
15:       $hi \leftarrow \text{GENERATE}(r1, n - 1 - i, \text{squelette}, h)$ 
16:    end if
17:  end if
18:   $node \leftarrow \text{squelette.add\_node}(rank, k, lo, hi)$ 
19:  return  $node$ 

```

Dans la fonction GENERATE on fait en sorte d'identifier les noeuds que l'on peut assigner aux nil-transitions du noeud courant. Pour un noeud v ayant deux nil-transitions on vérifie donc avant de lui assigner en tant que haut fils un noeud n et fils bas un noeud m que $I(m) < I(v)$ et $I(n) < I(v)$, on vérifie également qu'il n'existe pas $v' \in \text{squelette}$ tel que $v \neq v'$, où v a pour fils bas m' et pour fils haut n' , que $m = m'$ et $n = n'$.

On fait ensuite de même lorsque le noeud courant a une seule nil-transition en tant que fils bas ou en tant que fils haut tout en respectant les règles que l'on a déjà décrit pour la génération des bdd.

Pour cela on utilise ces fonctions :

- $\text{unrank_singleton}(r)$: cette fonction retourne un noeud compatible avec le noeud courant de rang r .
- $\text{unrank_pair}(rank, k)$: Cette fonction retourne une paire de noeud dont l'index est strictement inférieur à k , on omet également les paires de noeuds déjà assignés en tant que descendant d'un noeud précédent de même index.
- $\text{add_node}(lo, hi, k)$: Cette fonction crée le noeud d'index k et ayant pour descendants bas et haut lo et hi .

Enfin on décrit une fonction $\text{GEN_BDD}(r,n,k)$ permettant de générer un bdd de taille n , d'index k et de rang r en utilisant la fonction GENERATE décrite précédemment.

```

1: function GEN_BDD(rank, n, k)
2:   squelette  $\leftarrow$  squelette(k)
3:   r  $\leftarrow$  rank
4:   p  $\leftarrow$  [0] * k
5:   p[0]  $\leftarrow$  2
6:   d  $\leftarrow$  count(n - 2, p, 0)
7:   for (t, w) in d.items() do
8:     r  $\leftarrow$  r - w
9:     if r < 0 then :
10:      r  $\leftarrow$  r + w
11:      generate(r, n - 2, squelette, t)
12:     return squelette
13:

```

squelette est ici une classe d'objet permettant de représenter un squelette.

La méthode du unranking nous permet de mettre au point un générateur aléatoire uniforme de bdd.

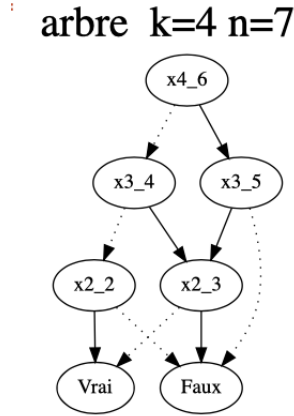


FIGURE 7 – Graphe(A) rang = 3104

L'ensemble dans lequel on a choisi le squelette du graphe(A) de la figure 7 est : $\{(0, 2, 1, 1, 1) : 584, (0, 1, 2, 1, 1) : 1256, (0, 2, 2, 0, 1) : 74, (0, 1, 1, 2, 1) : 1112, (0, 2, 0, 2, 1) : 74, (0, 0, 2, 2, 1) : 74\}$.

Avec un rang=3104, on déduit le profile du squelette du graphe qui est (0, 0, 2, 2, 1). Puisqu'en faisant la somme $584+1256+74+1112+74 = 3100$, on remarque qu'on a pas atteint 3104 donc le seule choix reste le dernier.

Avec un rang=3104 et le profile de squelette généré on se retrouve avec $x1=0$; $x2=2$; $x3=2$; $x4 = 1$ et le graphe de la figure 7.

8 Quickcheck

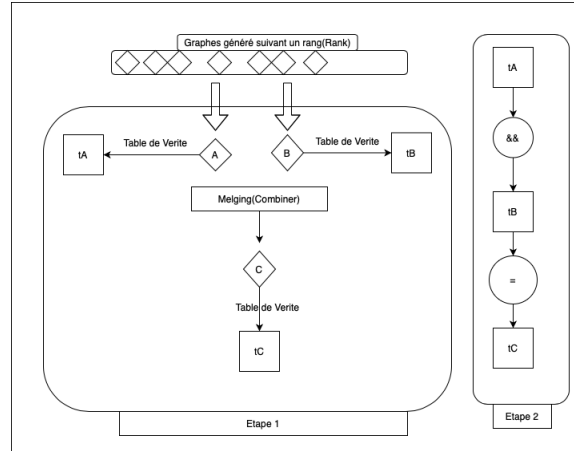


FIGURE 8 – Résumé des tests

arbre $k=4$ $n=7$

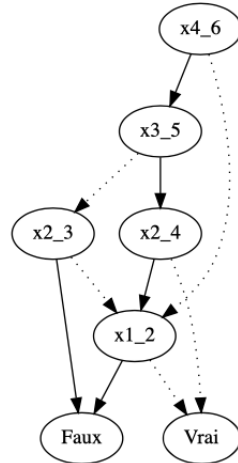


FIGURE 9 – Graphe(B) rang(rank) = 1577

Dans nos tests, il s'agit d'utiliser une bibliothèque quickcheck pour tester les propriétés de notre générateur. Pour cela on a due recourir à une stratégie avec

la bibliothèque Hypothesis en langage Python. Pour commencer nous générons un certains nombres de graphes (G) à partir d'une suites de rang uniformément généré par Hypothesis. Pour chaque couple (A,B) de l'ensemble G, nous produisons leur combinaison C puis nous générons la table de vérité de chacun des trois (Etape 1). Enfin nous testons la formule de l'étape 2.

La fonction melding, combine deux BDD(graphes). Le résultat est un graphe qui une conjonction des deux BDD combinés. on peut voir cela avec le graphe(C) qui la combinaison des graphes A et B.

Avec cette méthode nous avons validé qu'une partie, il faut savoir que d'autre types de tests sont possibles. Comme par exemples des tests comparatifs avec d'autres bibliothèques.

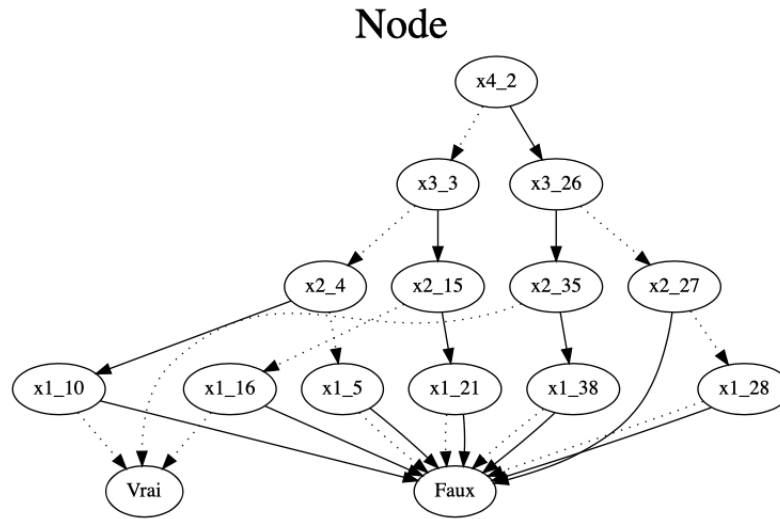


FIGURE 10 – Graphe(C) combinaison de A et B

Pour l'exemple de ces trois graphes on peut observer les tables de vérités pour une validation avec ces choix de rangs sur la figure 11.

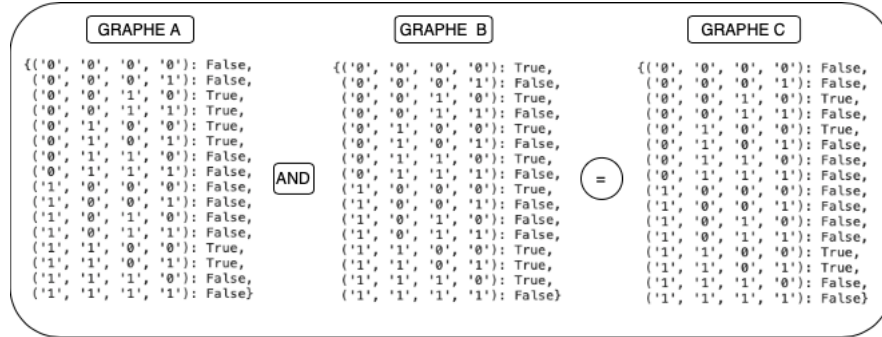


FIGURE 11 – A et B \Leftrightarrow C

9 Conclusion

Il ne fait aucun doute que la methode presenté dans ce rapport offre un gain en temps et en espace gigantesque pour la resolution du probleme présenté, la generation de bdd. Nous avons vu que même avec un materiel limité nous pouvions generer un grand nombre de bdd en un temps records, ce qui n'était pas possible avec la methode standard.

Cette generation en plus de sa simplicité presente d'autre gain tel que la possibilité de l'adapter selon nos besoins. Nous pouvons ainsi l'adapter pour pouvoir generer d'autre types d'arbre de decisions ainsi que pourquoi pas d'arbre. Mais comme nous l'avons vu ce gain ne se limite pas a la seule generaiton de bdd, cette methode nous offre une toute nouvelle comprehension et moyen d'étude de ces structures que l'on nomme bdd.

On note également que l'on pourrait encore améliorer notre méthode, par exemple en faisant en sorte de ne considérer que les bdd valides, ou en améliorant le processus de comptage.

10 Appendix

10.1 Appendix 1

On présente ici l'algorithme de compression en pseudo code

Require:

Le noeud initial est celui de la racine de l'arbre à compresser

table est une valeur globale stockant un ensemble de sous arbre et initialement vide

count est un compteur globale initialisé à 0

```
1: function COMPRESSION(n)
2:   if is_terminal(n) then      ▷ is_terminal(n) : Vérifie si n est une feuille
3:     if Label(n)=False then      ▷ Label(n) : Renvoie le label de n
4:       return 0
5:     else
6:       Return1
7:     end if
8:   else
9:     lo ← Left(n)                ▷ Left(n) : Renvoie le sous arbre gauche de n
10:    hi ← Right(n)
11:    lo ← Compression(lo)
12:    hi ← Compression(hi)
13:    if lo = hi then return lo
14:    else
15:      triple ← (INDEX(n), lo, hi)    ▷ Index(n) : Renvoie l'index de n
16:      value_found ← FIND(triple, table)  ▷ Find(n) : Renvoie l'arbre
17:      if value_found = None then
18:        uid ← count
19:        INSERT((triple, uid), table)  ▷ INSERT((triple, uid), table) :
20:        Insert triple dans la table avec l'identifiant uid.
21:        count ← count + 1
22:        return uid
23:      elsereturn value_found
24:    end if
25:  end if
26: end function=0
```

On peut faire remarquer que cette algorithme a une complexité polynomiale $O(n)$ où n est le nombre de noeud de l'arbre à compresser.

Références

- [1] Herbert S. Wilf ALBERT NIJENHUIS. “Combinatorial algorithms : An update”. In : *SIAM* (1989).
- [2] R.E BRYANT. “Graph-based algorithms for boolean function manipulation”. In : *IEEE Trans. Computers* 35.8 (1986), p. 677-691. DOI : <https://doi.org/10.1109/TC.1986.1676819>.
- [3] Didier Verna JIM NEWTON. “A theoretical and numerical analysis of the worst-case size of reduced ordered binary decision diagrams”. In : *ACM Transactions on Computational Logic* 20.6 (2019), p. 1-36. DOI : <https://doi.org/10.1145/3274279>.
- [4] Antoine Genitrini JULIEN CLÉMENT. “Binary Decision Diagrams : from Tree Compaction to Sampling”. In : (2020).
- [5] Donald E. KNUTH. *The Art of Computer Programming*. T. 4A. Addison-Wesley, 2011. Chap. Combinatorial Algorithms. ISBN : 9780201038040.
- [6] C. Y. LEE. “Representation of Switching Circuits by Binary-Decision Programs”. In : *Bell System Technical Journal* 38 (1959), p. 985-999. DOI : <https://doi.org/10.1002/j.1538-7305.1959.tb01585.x>.
- [7] P. Sipala P. FLAJOLET et J.-M. STEYAERT. “Analytic variations on the common subexpression problem [In Automata, languages and programming (Coventry, 1990)]”. In : *Lecture Notes in Comput. Sci.* 443 (1990), p. 220-234. DOI : <https://doi.org/10.1007/BFb0032034>.
- [8] Claude SHANNON. *Trans. Amer. Inst. Electrical Engineers*. T. 57. 1938, p. 713-723.
- [9] Ingo WEGENER. *Branching Programs and Binary Decision Diagrams : Theory and Applications*. Sous la dir. de SIAM. Monographs on Discrete Mathematics and Applications (Book 4). Society for Industrial et Applied Mathematics, 2000. ISBN : 9780898714586. DOI : <https://doi.org/10.1137/1.9780898719789>.