Projet ANR blanc BOOLE
Rapport intermédiaire à 30 mois

Mars 2012

1 F1: Circuits et formes normales booléennes.

Dans ce thème prennent place les travaux reliant diverses représentations des fonctions booléennes (par circuits, par polynômes) à la complexité de problèmes sur ces fonctions.

En collaboration avec, notamment, H. Vollmer (Leibniz Universität, Hanovre, Allemagne), N. Creignou a étudié la complexité de différents problèmes algorithmiques concernant les circuits booléens (tels l'équivalence de deux circuits ou l'énumération de toutes les solutions d'un circuit). Une classification complète de la complexité de ces problèmes selon le type de portes utilisées a été obtenue, ainsi que des algorithmes efficaces pour un certain nombre d'ensemble de types de portes; pour les autres ensembles de portes ces problèmes ont été prouvés au moins NP-difficiles.

H. Fournier, en collaboration avec G. Malod (Paris 7) et S. Mengel (Paderbron, Allemagne), s'est intéressé à la complexité de deux questions à propos des polynômes représentés par des circuits arithmétiques : tester si un monôme est présent, et compter les monômes. En étudiant ces questions sur diverses classes de circuits, on obtient des problèmes complets pour des sous-classes de la hiérarchie de comptage.

2 F2: Fonctions booléennes et cryptographie.

Il s'agit ici d'étudier les propriétés de fonctions booléennes, en particulier celles susceptibles d'applications cryptographiques telles que le fait d'être courbe, ou l'immunité algébrique. Les recherches sont de nature fondamentale, mais aussi plus pratiques, avec l'exploration de ces propriétés par la génération aléatoire de fonctions, soit directement, soit par l'intermédiaire de leurs représentations.

2.1 Immunité algébrique.

J. Clément, A. Genitrini et J.-M. Le Bars travaillent, avec A. Viola (Univ. de Montevideo, Uruguay), à l'application de la méthode des classes au problème de l'immunité algébrique d'une fonction booléenne. L'immunité algébrique est le plus petit degré algébrique d'un annulateur d'une fonction ou de son complémentaire. C'est un paramètre crucial : les fonctions booléennes doivent avoir une immunité algébrique élevée pour pouvoir être utilisées en cryptographie.

Leurs travaux ont porté dans un premier temps sur les espaces vectoriels des annulateurs de degré algébrique fixé d'une fonction booléenne, dont toutes les dimensions possibles jusqu'à 4 variables ont été déterminées expérimentalement. Les recherches actuelles portent sur les supports de ces espaces vectoriels.

2.2 Fonctions courbes.

Les intervenants sur ce sujet sont J. Clément, A. Genitrini, S. Mesnager, et J.-M. Le Bars, avec toujours A. Viola. Il s'agit ici d'énumérer et générer aléatoirement des fonctions courbes, qui

sont les fonctions de plus grande non-linéarité. Il est prévu que N. Carrasco (Uruguay), qui a déjà travaillé sur l'application de la méthode des classes à la génération aléatoire de fonctions booléennes satisfaisant certains critères, prenne en charge l'implémentation.

Par ailleurs, S. Mesnager a obtenu plusieurs résultats significatifs sur les fonctions courbes en utilisant le corps F_{2^n} au lieu de F_2^n (une fonction booléenne devient donc une fonction de F_{2^n} vers F_2), et commence depuis peu à étudier F_{2^n}, afin de mieux comprendre la correspondance entre ces deux représentations.

2.3 Énumération et génération aléatoire de diagrammes de décision binaires.

Les diagrammes de décision binaires (BDDs) ou leurs variantes sont souvent utilisés en pratique pour représenter une fonction booléenne. Pour mieux comprendre les aspects combinatoires des BDDs, et quels types de formules sont représentables avec des BDDs de taille fixée. J. Clément et J.-M. le Bars s'intéressent à l'énumération et à la génération aléatoire de ces objets.

2.4 Dualisation de fonctions booléennes monotones.

Le problème de la dualisation d'une fonction booléenne est le suivant: étant donnée une fonction booléenne monotone, écrite sous sa forme normale conjonctive, calculer la forme normale conjonctive de la fonction duale. Un problème ouvert est de savoir si dans le pire des cas la dualisation est un problème de la classe output-polynomiale (polynomiale en la taille de l'entrée plus celle de la sortie). L. Lhote, en collaboration avec F. Rioult (Greyc, Caen) a apporté un éclairage sur la complexité en moyenne du problème. Lorsque le nombre de variables est fixé mais pas le nombre de clauses, il est connu que le problème est très probablement output-polynomial. La question restait ouverte lorsque le nombre de variables et de clauses sont simultanément fixés, l'un dépendant éventuellement de l'autre. Sous certaines conditions, on peut montrer que le problème est output-polynomial en analysant un algorithme du domaine, et prouver au passage que la taille moyenne de la sortie est super-polynomiale.

3 F3: Satisfaisabilité.

Les travaux dans ce thème concernent d'une part les transitions de phase, approchées tant par l'analyse fine de problèmes spécifiques bien structurés que par une approche globale des CSP booléens, et d'autre part l'étude des performances d'un solveur de contraintes.

3.1 Formules quantifiées.

En collaboration avec U. Egly (T.U. Wien, Autriche), N. Creignou, H. Daudé et R. Rossignol avaient étudié la satisfaisabilité des formules booléennes quantifiées du type $\forall X \exists \varphi(X, Y)$, où φ est une conjonction de clauses contenant chacune une variable de X et deux variables de Y, et montré que la transition Sat/Insat est contrôlée par le rapport nombre de clauses sur nombre de variables existentielles. L'étude expérimentale qui a suivi a permis de préciser l'apparition du ratio critique associé à la transition, et d'obtenir la valeur limite de ce ratio.

3.2 Générateur de formules aléatoires.

N. Creignou a continué sa collaboration avec U. Egly et M. Seidl (Johannes Kepler University, Linz, Autriche). Leur travail a abouti à la construction d'un générateur de formules aléatoires (quantifiées ou non) non nécessairement en forme normale –cf. la page web http://big.tuwien.ac.at/staff/seidl/qbfgen/.

Ce générateur permet de créer des instances aléatoires en interprétant la spécification des formules souhaitées en XML. Ainsi grâce à un seul générateur on peut obtenir des instances aléatoires

de types variés sans aucun effort de programmation; de telles données sont utiles pour l'évaluation des solveurs dédiés à SAT et QSAT.

3.3 Solveur de contraintes.

J. du Boisberranger, D. Gardy et C. Truchet, en collaboration avec X. Lorca (Nantes), poursuivent leur travail d'exploration du comportement de la contrainte *AllDifferent*. Après avoir mis au point les outils permettant de juger de l'intérêt de considérer cette contrainte, il s'agit maintenant d'évaluer expérimentalement différentes stratégies de résolution de contraintes; ce travail se fait dans le cadre du solveur de contraintes *Choco/Galak* développé par l'équipe de Nantes.

4 F4: Logique quantitative.

Cette partie concerne l'étude de diverses lois de probabilité pouvant être définies sur les fonctions booléennes à partir de représentations arborescentes.

4.1 Influence des propriétés des opérateurs.

A. Genitrini et C. Mailler ont poursuivi leur étude des formules booléennes, lorsque les opérateurs sont commutatifs ou associatifs. En collaboration avec B. Gittenberger et V. Kraus (tous deux de la T.U. Wien, Autriche), ils ont montré que ces propriétés ne changent pas fondamentalement le type de la loi de probabilité sur les fonctions booléennes qui en découle: la probabilité d'une fonction f est toujours de la forme $\gamma(f)/n^{L(f)+1}$, avec $L(f)$ la complexité en arbre de la fonction f (bien que la constante $\gamma(f)$ dépende des propriétés des opérateurs). En particulier, de telles distributions ne présentent pas d'effet Shannon.

4.2 Influence de la notion de taille d'une expression.

La même équipe de chercheurs étudie actuellement un modèle avec une notion de taille différente: la taille d'un arbre est, non plus le nombre de ses feuilles, mais le nombre total de ses noeuds – afin d'être plus proche de la mémoire réellement utilisée pour représenter l'arbre. De manière surprenante, ce modèle (qui n'a de sens que lorsque les opérateurs sont associatifs) se comporte très différemment, avec en particulier un nombre important de feuilles au premier niveau, donc des arbres relativement "plats".

5 M1: Méthodologie: méthodes de combinatoire analytique.

Les sujets abordés ici sont de nature fondamentale; ils ont concerné le modèle des urnes de Polya (qui apparaît par exemple lors de l'analyse de certains types d'expressions arborescentes booléennes), les arbres enrichis, et des phénomènes d'allocations aléatoires.

5.1 Urnes de Polya.

Après avoir travaillé sur une classe d'urnes additives équilibrées à 2 paramètres, en utilisant une série génératrice algébrique et une méthode de cols multiples coalescents pour obtenir des résultats sur les distributions limites, B. Morcrette a établi des conditions nécessaires sur la structure des règles pour des urnes additives équilibrées afin de garantir l'algébricité de la fonction génératrice qui code le comportement de l'urne. De nouvelles classes d'urnes à série génératrice apparaissent; cela ouvre la porte à leur analyse asymptotique par le même genre de méthodes. Ce travail a été fait grâce aux méthodes de calcul formel, en collaboration avec A. Bostan, F. Chyzak et P. Dumas (Inria Rocquencourt).

B. Morcrette a également travaillé, en collaboration avec H. Mahmoud (Washington University, U.S.A.), sur des urnes équilibrées où les règles d'évolution sont dictées par des variables aléatoires.

Ils ont étendu la théorie combinatoire analytique introduite par P. Flajolet et al., afin d'obtenir un système différentiel permettant d'exprimer la fonction génératrice de ces urnes. Pour de nombreux modèles, ce système fournit des expressions exactes des distributions de probabilité.

Un dernier travail, en cours, se fait en collaboration avec P. Dumas et concerne les urnes non équilibrées, dont le comportement est régi par une équation aux dérivées partielles.

5.2 Arbres digitaux

B. Vallée, avec le doctorant Kanal Hun, qui a débuté sa thèse en octobre 2011, étudie les dst (digital search tree) sous le modèle d'une source générale, afin de comparer les principaux paramètres (dont la longueur de cheminement) avec les paramètres analogues des tries ou des patricia-tries. La première étape (obtention d'une expression algébrique des séries génératrices) est encourageante, et montre le rôle joué par une série génératrice (de type Dirichlet) de la source, analogue, mais non identique àigitt la série génératrice de Dirichlet qui apparaît pour les tries.

5.3 Arbres enrichis.

En collaboration avec O. Bodini et A. Jacquot (tous deux de l'université Paris 13) et B. Gittenberger, D. Gardy a poursuivi l'étude des arbres enrichis, dans le cas où le nombre d'arcs partant d'un noeud vers les feuilles est borné par une constante. La fonction génératrice obtenue ne rentrant pas dans le cadre classique de la combinatoire analytique, une étude asymptotique spécifique a dû être menée; elle a permis d'obtenir des informations sur l'asymptotique de ces objets. Une étude plus complète est en cours.

5.4 Allocations aléatoires.

J. du Boisberranger s'est intéressé au problème classique du *Collectionneur de Coupons* dans le cas particulier où les éléments à collectionner sont des mots d'un langage construit sur un alphabet dont les lettres ne sont pas équiprobables. Ce cas conduit à des distributions non uniformes sur les coupons, mais o ù cependant les mots ayant les mêmes lettres, et les mêmes répétitions de lettres, ont la même probabilité. Dans un tel cadre avec beaucoup de probabilités répétées, les hypothèses ayant permis d'obtenir des résultats antérieurs sur les coupons non uniformes ne s'appliquent plus. Une étude asymptotique a permis d'obtenir le temps d'attente de la collection complète, sous des hypothèses assez générales sur le langage; en particulier ceci a été appliqué à un algorithme de génération aléatoire d'ARN (collaboration avec Y. Ponty, LIX, Ecole Polytechnique).

6 M2: Méthodologie: méthodes probabilistes.

Dans cette partie se place l'étude d'un nouveau modèle de source pour les arbres digitaux. Nous y incluons aussi le travail d'écriture d'un livre qui relève à proprement parler des deux thèmes **M1** et **M2**.

6.1 Livre de synthèse.

B. Chauvin, J. Clément et D. Gardy ont entrepris l'écriture d'un livre sur les arbres utilisés en algorithmique. Il s'agit d'un travail de longue haleine, qui vise à présenter les travaux relatifs à l'évaluation des paramètres des structures arborescentes classiques en algorithmique, et à la complexité des différents algorithmes sur ces structures, de façon accessible à tout usager potentiel, venant aussi bien de l'informatique que des mathématiques, à partir d'un niveau Master et au-delà. Un but essentiel est de faire le pont, dans l'étude des arbres, entre l'approche probabiliste et celle de la combinatoire analytique.

6.2 Variable Length Markov Chain (VLMC).

Ces travaux sont menés par B. Chauvin et N. Pouyanne, en collaboration avec P. Cénac (Dijon) et F. Paccaut (Amiens). Après la mise en place du modèle probabiliste pour les VLMC en 2011, un second travail sur plusieurs cas de "peignes" comme arbre de contexte a permis d'aller plus loin sur ce modèle. Tout d'abord, les propriétés de mélange de la chaîne de caractères ainsi produite sont précisées, ce qui conduit à expliciter des modèles relativement simples, en tout cas faciles à implémenter et pour lesquels le mélange n'est pas exponentiel, comme cela est habituellement supposé dans une source probabiliste, mais seulement polynomial, ou même pas du tout mélangeant.

Une autre conséquence spectaculaire est la construction de tries suffixes à partir d'une telle source, qui sont inhabituels au sens où leur hauteur n'est pas en $\log n$ mais beaucoup plus grande (en puissance de n) et leur niveau de saturation non pas en $\log n$ mais beaucoup plus petit.