

UNIVERSITE D'ANTANANARIVO
Faculté des Sciences
Département de Mathématiques
et d'Informatique

UNIVERSITE D'ANTANANARIVO



Mémoire pour l'obtention du
Diplôme d'Etudes Approfondies de Mathématiques

Option : Mathématiques Appliquées
Spécialité : Informatique et Génie Mathématiques

APPROCHE ANALYTIQUE D'UN PROBLÈME
DIFFICILE D'OPTIMISATION : MAX-2-XORSAT

Soutenu par : **RASENDRAHASINA Solofoniaina Ndimbisoa Tinarivo Vonjy**
le 06 Octobre 2008

Devant la Commission d'Examen formée de

Président : Monsieur Arthur RANDRIANARIVONY
Professeur à l'Université d'Antananarivo
Rapporteur : Monsieur Vlady RAVELOMANANA
Maître de Conférence à l'Université de Paris XIII
Examineur : Monsieur Gérard RAZAFIMANANTSOA
Maître de Conférence à l'Université d'Antananarivo
Monsieur Marc RABIAZAMAHOLY
Maître de Conférence à l'Université d'Antananarivo
Monsieur Daniel RASOAMIARAMANANA
Maître de Conférence à l'Université d'Antananarivo

Approche Analytique d'un Problème Difficile d'Optimisation : MAX-2-XORSAT

RASENDRAHASINA Solofoniaina Ndimbisoa Tinarivo Vonjy

Table des matières

Remerciements	ii
Introduction	iii
1 Contextes généraux de MAX-2-XORSAT	1
1.1 Le contexte 2-XORSAT	1
1.2 Le Problème MAX-2-XORSAT	2
1.3 Méthodes et approches adoptées	4
2 Énumérations exactes et asymptotiques	6
2.1 Rappels sur les graphes connexes et 2-XORSAT	6
2.1.1 Les graphes connexes sans cycles de poids impairs	6
2.1.2 Des séries génératrices à la satisfaisabilité	8
2.2 Approche de MAX-2-XORSAT par les graphes	9
2.2.1 Le cas particulier $\ell = 1$	9
2.2.2 Les composantes multicycliques	12
2.3 Énumérations asymptotiques	14
3 Les instances aléatoires MAX-2-XORSAT	17
3.1 Clauses satisfaisables en phase sous critique	17
3.2 MAX-2-XORSAT voisin de la transition	21
Conclusion	23

Remerciements

Je souhaite adresser mes remerciements à Monsieur Vlady RAVELOMANANA du Laboratoire de l'Informatique de Paris Nord (LIPN) de l'Université de Paris 13. Il m'a permis de découvrir un thème de recherche passionnant : Approche Analytique d'un Problème difficile d'optimisation MAX-2-XORSAT. Je n'aurais sans doute jamais entrepris de tel travail sans ses conseils si pertinents et son encadrement exemplaire.

Je tiens également à remercier l'équipe des enseignants de l'Université de Paris 13 et de l'Université Libre de Bruxelles (ULB). Ils ont eu la bonne volonté de venir à Madagascar pour nous partager leur expérience et nous avons eu l'occasion de profiter de leur immense savoir.

La faculté des Sciences de l'Université d'Antananarivo, par le département de Mathématiques et Informatique, nous a fait découvrir cette formation. Que vous trouviez ici un témoignage de nos chaleureux remerciements.

J'aimerais remercier Monsieur Arthur RANDRIANARIVONY d'être président de mon jury. Il m'a consacré du temps malgré ses multiples obligations.

Monsieur Gérard RAZAFIMANANTSOA, Monsieur Marc RABIAZAMAHOLY et Monsieur Daniel RASOAMIARAMANANA ont immédiatement accepté d'être examinateurs de ce mémoire. Je les remercie pour leur disponibilité et leur gentillesse.

Un grand merci à ma famille pour le soutien et l'encouragement qu'elle a manifestés envers mes études.

Introduction

Les problèmes de satisfaction de contraintes ou d'optimisation combinatoire qui apparaissent fréquemment dans plusieurs domaines sont une préoccupation majeure de diverses communautés scientifiques : biologistes, informaticiens, physiciens et mathématiciens. Depuis une vingtaine d'années, les chercheurs (essentiellement informaticiens, physiciens et mathématiciens) ont pu établir que ces problèmes sont sujets à des phénomènes de transition de phase comme ceux rencontrés en physique. En physique statistique par exemple, une petite variation de certains paramètres de contrôle peut changer abruptement les comportements ou l'état d'un système arbitraire. En optimisation combinatoire, une petite variation de la densité de la formule peut changer la satisfaisabilité de celle-ci. Cette analogie est sans doute essentielle dans la compréhension des problèmes de satisfaction qui sont au coeur de la complexité informatique. MAX-2-XORSAT est un exemple typique de ces problèmes issus de l'informatique théorique et de la physique statistique. Nous nous intéressons à l'approche analytique de ce problème. Il est la version optimisation de 2-XORSAT, donc consiste à maximiser le nombre de clauses satisfaisables. Il illustre la difficulté de l'informatique à résoudre un problème en un temps « raisonnable », car à l'heure actuelle, il n'existe pas des algorithmes tournant en temps polynomial pour le résoudre.

Le premier chapitre est consacré aux contextes généraux du problème MAX-2-XORSAT. Nous allons commencer par décrire le contexte 2-XORSAT qui est le problème de décision correspondant à MAX-2-XORSAT. Nous allons brièvement rappeler les séries génératrices qui serviront de briques de base pour notre approche. Des résultats sur 2-XORSAT seront alors décrits en début du second chapitre, notamment les séries génératrices des configurations autorisées et l'étude de la satisfaisabilité d'une formule dans une *fenêtre critique*. Ensuite nous donnons des énumérations exactes de certains cas particuliers des configurations qui correspondent à MAX-2-XORSAT. Des énumérations asymptotiques sont aussi entamées dans cette partie. Enfin, le troisième chapitre sera consacré à l'étude des instances aléatoires de MAX-2-XORSAT. Dans le cas présent, nous nous limitons sur le nombre maximum de clauses satisfaisables en phase sous critique.

Chapitre 1

Contextes généraux de MAX-2-XORSAT

1.1 Le contexte 2-XORSAT

Pour les problèmes de satisfaction, deux questions principales peuvent se poser : existe-t-il une affectation de valeur aux variables réalisant toutes les contraintes, ou comment faire pour satisfaire un nombre maximum de contraintes ? Ces deux questions classent respectivement les problèmes de satisfaction en *problème de décision* et en *problème d'optimisation*. Le problème 2-XORSAT consistant à répondre *oui* s'il existe une solution à la formule donnée et *non* dans le cas contraire est donc un problème décisionnel. Il peut être formulé de la manière suivante. Nous avons une collection de n variables booléennes $x_1, x_2, \dots, x_n$ et une conjonction de m clauses de la forme $x_i \oplus x_j = \epsilon$ où $\epsilon \in \{0, 1\}$ avec $x_i \oplus x_j = 1$ si et seulement si $(x_i = 0 \text{ et } x_j = 1)$ OU $(x_i = 1 \text{ et } x_j = 0)$, et $x_i \oplus x_j = 0$ si et seulement si $x_i = x_j$. Notons par $\mathcal{P}(n, m)$ la classe de ces formules. Une clause $x_i \oplus x_j = \epsilon$ de l'*instance* ainsi formée est dite satisfaisable s'il existe une affectation (de vrai=1 ou faux=0) aux variables rendant l'équation vraie. Par la suite, une formule sur n variables et à m clauses de $\mathcal{P}(n, m)$ est dite satisfaisable si et seulement si toutes les clauses sont satisfaisables.

Exemple, soit le problème $\Pi_1 \in \mathcal{P}(4, 4)$ suivant

$$\Pi_1 = \begin{cases} x_1 \oplus x_2 = 1 \\ x_2 \oplus x_3 = 0 \\ x_3 \oplus x_4 = 1 \\ x_4 \oplus x_2 = 1 \end{cases}$$

Π_1 est satisfaisable car une solution serait $\{x_1 = 0, x_2 = 1, x_3 = 1, x_4 = 0\}$. La figure 1.1 représente le graphe correspondant.

Comme beaucoup de problèmes de satisfaction, 2-XORSAT se modélise en utilisant la théorie des graphes. Les formules peuvent être traduites sous forme de graphes où les sommets sont étiquetés par les variables x_i et les arêtes sont pondérées par 0 ou 1 représentant les relations entre deux sommets. Il est établi qu'une formule est satisfaisable si et seulement si il n'existe pas de cycles de poids impairs dans le graphe correspondant. Ce

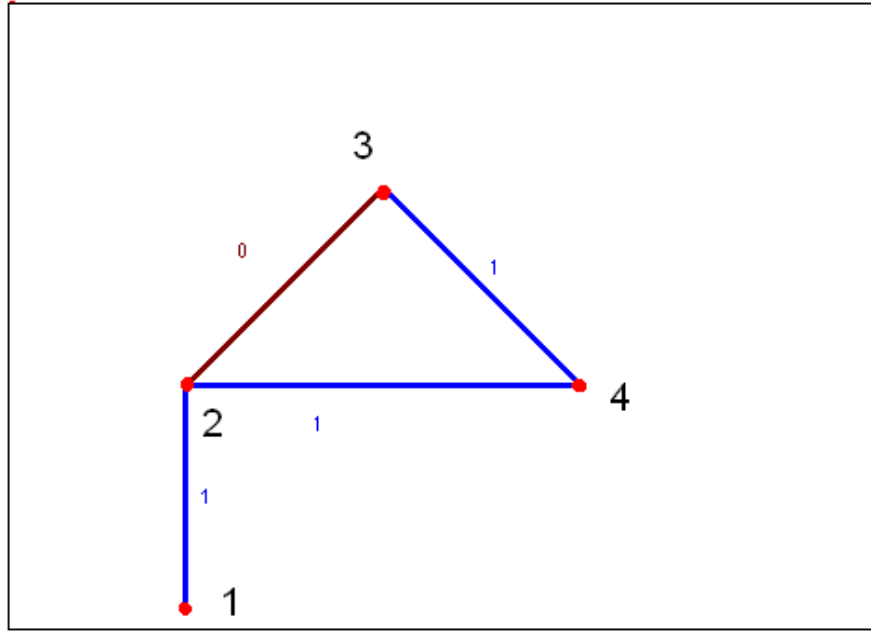


FIG. 1.1 – Une composante 2-XORSAT

problème est facile algorithmiquement car il se résout en temps polynomial. Par contre, s'il n'existe pas une solution au problème donné, MAX-2-XORSAT consiste alors à *maximiser le nombre de clauses satisfaisables* et ce problème est un problème difficile d'optimisation. Voyons de suite un exemple de formule non satisfaisable, $\Pi_2 \in \mathcal{P}(4, 4)$

$$\Pi_2 = \begin{cases} x_1 \oplus x_2 = 1 \\ x_2 \oplus x_3 = 0 \\ x_3 \oplus x_4 = 1 \\ x_4 \oplus x_2 = 0 \end{cases}$$

Π_2 ne possède pas de solutions. Dans cet exemple, le nombre maximum de clauses satisfaisables est 3. La figure 1.2 représente le graphe correspondant à Π_2 .

1.2 Le Problème MAX-2-XORSAT

Nous avons un modèle aléatoire d'un problème de satisfaction de contraintes booléennes. Reprenons l'instance que nous venons de formuler précédemment. Le nombre de clauses est m et le nombre de variables est n . Une clause $x_i \oplus x_j = \epsilon$ est tirée uniformément au hasard parmi toutes les clauses possibles. Comme le tirage est sans remise, chaque formule comportant m clauses apparaît avec une probabilité

$$\frac{1}{\binom{n(n-1)}{m}} \quad (1.1)$$

S'il n'existe pas une affectation de valeurs aux variables qui rendent toutes les clauses satisfaites, c'est à dire si le graphe traduisant la formule contient au moins un cycle de

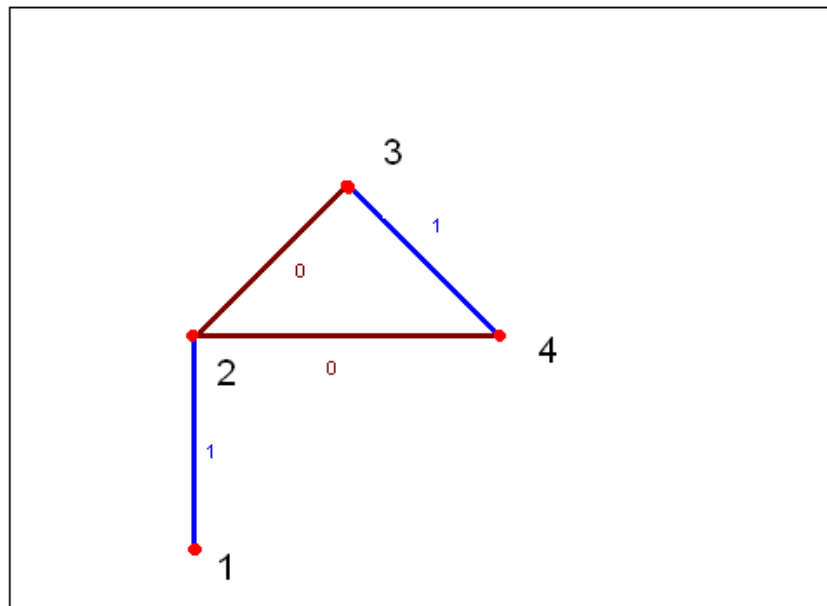


FIG. 1.2 – Une composante MAX-2-XORSAT

poids impair, que peut on dire à propos du nombre maximum de clauses satisfaisables ? Le problème *MAX-2-XORSAT* consiste à maximiser le nombre de clauses satisfaisables dans une formule. En d'autres termes, sur les 2^n affectations possibles des variables x_i , il s'agit de trouver celle(s) qui vont satisfaire le maximum de clauses dans une formule donnée.

Rappelons quelques principales motivations.

D'abord, l'une des principales motivations de ce domaine est la relation importante entre les phénomènes observés expérimentalement en physique statistique et les problèmes d'optimisation combinatoire que nous avons mentionnés au début. Le problème sert donc à valider certaines méthodes non encore rigoureuses en physique théorique notamment la méthode des "verres de spin".

Ensuite les problèmes de type *satisfaction* intéressent beaucoup de chercheurs. A ce propos, nous renvoyons à l'article de Talagrand [12] qui est une introduction au séminaire Bourbaki à l'intention des mathématiciens purs de ces objets décrits par des méthodes et des langages motivés par la physique théorique [6]. Ces problèmes sont célèbres non seulement par les difficultés intrinsèques qu'elles engendrent mais aussi par leur rôle central en théorie de complexité computationnelle.

Enfin les caractéristiques de 2-XORSAT nous conduisent naturellement à la version OPTIMISATION du problème : étant donné un système de m clauses sur n variables et un nombre positif R , on cherche s'il existe une affectation de valeurs aux variables qui rend R équations vraies. Il est démontré que pour $R = m$ et $m = cn$, une formule aléatoire de cn clauses sur n variables est satisfaisable avec une probabilité $p(n, cn)$ telle que

$$\lim_{n \rightarrow +\infty} p(n, cn) = \begin{cases} \exp(c/2) \cdot (1 - 2c)^{1/4} & \text{si } 0 \leq c \leq 1/2 \\ 0 & \text{ailleurs} \end{cases}$$

Ce résultat montre que la fonction $\lim_{n \rightarrow +\infty} p(n, cn)$ est une fonction décroissante de c . Notons que 2-XORSAT admet une transition de phase douce en $c = \frac{1}{2}$. Les résultats de 2-XORSAT [9] ont été obtenus en combinant la théorie des graphes (caractérisations par des graphes pondérés dans $\{0, 1\}$ sans cycles de poids impairs) et la combinatoire analytique.

C'est cette même approche que nous allons adopter pour aborder le problème. Nous définissons toute une famille de graphes avec cycles de poids impairs qui caractérisent le problème MAX-2-XORSAT, tels qu'en supprimant s arêtes on arrive à des configurations satisfaisables. Nous cherchons des séries génératrices associées à ces objets en utilisant le nombre minimal d'arêtes à enlever et l'excès du graphe comme paramètres. Ensuite, nous nous focalisons sur le nombre de clauses satisfaisables essentiellement en phase sous critique.

1.3 Méthodes et approches adoptées

Notre approche est basée sur les outils de la combinatoire analytique, en particulier sur l'utilisation des séries génératrices. Nous allons rappeler quelques définitions sur les séries génératrices et expliquer comment nous allons faire pour MAX-2-XORSAT.

Les séries génératrices en combinatoire

Nous disposons d'une classe d'objets $\mathcal{F}$, d'une fonction $|\cdot|$ de $\mathcal{F}$ à valeur dans $\mathbf{N}$ dite taille. Nous voulons des informations sur le nombre des objets de taille n , noté F_n . La série génératrice ordinaire et la série génératrice exponentielle associées à $\mathcal{F}$ sont définies respectivement de la manière suivante :

$$F(z) = \sum_{\alpha \in \mathcal{F}} z^{|\alpha|} = \sum_{n=0}^{+\infty} F_n z^n \quad \text{et} \quad F(z) = \sum_{\alpha \in \mathcal{F}} \frac{z^{|\alpha|}}{|\alpha|!} = \sum_{n=0}^{+\infty} F_n \frac{z^n}{n!}$$

Par exemple, si on considère l'ensemble des entiers naturels où la taille d'un entier est sa valeur, la série génératrice ordinaire est $F(z) = \sum_{n=0}^{+\infty} z^n = 1/(1-z)$ et la série génératrice exponentielle $F(z) = \sum_{n=0}^{+\infty} \frac{z^n}{n!} = \exp(z)$.

Dans ces exemples précis, les séries génératrices s'obtiennent directement. Mais lorsque les objets que l'on considère sont obtenus par une construction *simple* à partir d'autres classes d'objets combinatoires dont on connaît les séries génératrices, on peut utiliser la méthode dite la *méthode symbolique*¹. La série génératrice se déduit directement de la spécification de cette construction. Les tableaux 1.3 et 1.4 suivants récapitulent les correspondances classiques entre le type de la construction et la série génératrice associée.

Les séries génératrices ordinaires sont utilisées pour l'étude des objets non étiquetés et les séries génératrices exponentielles pour des objets étiquetés.

Dans ces tableaux, toutes les possibilités sont considérées, mais on peut se limiter à une suite, à un cycle ou à un ensemble d'objets de taille fixe k . Par exemple, si $\mathcal{F} = \text{SEQ}_k(\mathcal{A})$ alors $F(z) = A(z)^k$, si $\mathcal{F} = \text{SET}_k(\mathcal{A})$ alors $F(z) = A(z)^k/k!$. Ces séries génératrices sont dites monovariées. Mais si on associe également un certain nombre de paramètres qui dénombrent des caractéristiques supplémentaires des objets considérés, on dit que la série

¹Une description rigoureuse de la méthode symbolique peut être trouvée dans [2]

	<i>Construction</i>	<i>Série génératrice associée</i>
Union disjointe	$\mathcal{F} = \mathcal{A} \cup \mathcal{B}$	$F(z) = A(z) + B(z)$
Produit	$\mathcal{F} = \mathcal{A} \times \mathcal{B}$	$F(z) = A(z) \cdot B(z)$
Suite	$\mathcal{F} = \text{SEQ}(\mathcal{A})$	$F(z) = (1 - A(z))^{-1}$
Ensemble	$\mathcal{F} = \text{PSET}(\mathcal{A})$	$F(z) = \exp\left(\sum_{k=1}^{+\infty} \frac{(-1)^{k-1}}{k} B(z^k)\right)$

FIG. 1.3 – Séries génératrices ordinaires des constructions classiques

	<i>Construction</i>	<i>Série génératrice associée</i>
Union disjointe	$\mathcal{F} = \mathcal{A} \cup \mathcal{B}$	$F(z) = A(z) + B(z)$
Produit	$\mathcal{F} = \mathcal{A} \times \mathcal{B}$	$F(z) = A(z) \cdot B(z)$
Suite	$\mathcal{F} = \text{SEQ}(\mathcal{A})$	$F(z) = (1 - A(z))^{-1}$
Cycle	$\mathcal{F} = \text{CYC}(\mathcal{A})$	$F(z) = \log(1/1 - A(z))$
Ensemble	$\mathcal{F} = \text{SET}(\mathcal{A})$	$F(z) = \exp(A(z))$
Pointage	$\mathcal{F} = \Theta\mathcal{A}$	$F(z) = \frac{d}{dz}A(z)$
Substitution	$\mathcal{F} = \mathcal{A} \circ \mathcal{B}$	$F(z) = A(B(z))$

FIG. 1.4 – Séries génératrices exponentielles des constructions classiques

génératrice est multivariée. On définit la série génératrice exponentielle multivariée de la manière suivante :

$$F(w_1, w_2, \dots, w_\ell, z) = \sum_{\alpha \in \mathcal{F}} w_1^{p_1(\alpha)} w_2^{p_2(\alpha)} \dots w_\ell^{p_\ell(\alpha)} \frac{z^{|\alpha|}}{|\alpha|}$$

Par exemple, si $\mathcal{F}$ est la famille des graphes simples pondérés dans $\{0, 1\}$, z la variable des sommets, w la variable pour tracer les arêtes de poids 1 et u la variable pour tracer les arêtes de poids 0, la série génératrice exponentielle multivariée associée à $\mathcal{F}$ est

$$F(w, u, z) = \sum_{M \in \mathcal{F}} w^{m_1(M)} u^{m_0(M)} \frac{z^{|M|}}{|M|}$$

où $m_1(M)$ =le nombre d'arêtes de poids 1 du graphe M , $m_0(M)$ =le nombre d'arêtes de poids 0.

Dans la suite, on s'intéresse aux séries génératrices des graphes simples pondérés dans $\{0, 1\}$, en particulier les composantes connexes avec cycles de poids impairs. Dans le cas général, on se contente des énumérations asymptotiques de certains de ces graphes. Ces séries génératrices seront exploitées pour l'approche de MAX-2-XORSAT et les méthodes privilégiées sont celles issues de la combinatoire analytique.

Chapitre 2

Énumérations exactes et asymptotiques

Comme déjà mentionné au chapitre précédent, les problèmes 2-XORSAT et MAX-2-XORSAT sont étroitement liés au problème d'énumération des graphes pondérés dans $\{0, 1\}$ par l'existence ou non d'un (ou plusieurs) cycle(s) de poids impairs dans le graphe associé à la formule. Ce chapitre est consacré à l'énumération de ces objets paramétrés par le nombre d'excès et les cycles de poids impairs de la composante connexe considérée. On notera par $\mathcal{C}_{s,\ell}$ la famille des ℓ – *cycliques*¹ nécessitant la suppression de s arêtes pour ne contenir aucun cycle de poids impair et par $C_{s,\ell}$ la série génératrice exponentielle associée.

2.1 Rappels sur les graphes connexes et 2-XORSAT

Cette section comporte un survol sur les séries génératrices des bonnes configurations— les graphes sans cycles de poids impairs— et quelques résultats de 2-XORSAT.

2.1.1 Les graphes connexes sans cycles de poids impairs

Rappelons que la série génératrice exponentielle $t(z)$ des arbres de Cayley—arbres enracinés et étiquetés— satisfait l'équation suivante

$$t(z) = \sum_{n=1}^{+\infty} n^{n-1} \frac{z^n}{n!} = ze^{t(z)} \quad (2.1)$$

Comme un arbre à n sommets a exactement $n - 1$ arêtes et la pondération se fait dans un ensemble à deux éléments, un arbre de Cayley à n sommets fournit 2^{n-1} composantes acycliques, enracinées et pondérées dans $\{0, 1\}$, soit au total $2^{n-1}n^{n-1}$. Notons par $\mathcal{T}$ la famille de ces graphes et par $T(z)$ la série génératrice associée, alors on a

$$T(z) = \sum_{n=1}^{+\infty} (2n)^{n-1} \frac{z^n}{n!} = ze^{2T(z)} = \frac{1}{2} t(2z) \quad (2.2)$$

Remarquons que les éléments de $\mathcal{C}_{0,-1}$ sont obtenus en « déracinant » les éléments de $\mathcal{T}$. Par la suite, nous avons

$$C_{0,-1}(z) = \int_0^z \frac{T(z)}{z} dz = \int_0^z e^{2T(z)} dz$$

¹Un ℓ – *cyclique* est une composante connexe d'excès ℓ .

En effectuant le changement de variable $u = T(z)$, l'équation devient

$$\begin{aligned} C_{0,-1}(z) &= \int_0^{T(z)} (1 - 2u) du \\ &= T(z) - T(z)^2 \\ &= \sum_{n=1}^{\infty} 2^{n-1} n^{n-2} \frac{z^n}{n!} \end{aligned} \quad (2.3)$$

Les éléments de $\mathcal{C}_{0,0}$ sont les composantes unicycliques. Rappelons aussi que la série génératrice $c(z)$ des cycles est $c(z) = -\frac{1}{2} \log(1-z) - z/2 - z^2/4$. En observant qu'un cycle de taille n fournit 2^n cycles pondérés dans $\{0, 1\}$ et que à chaque cycle de poids impair on peut associer un cycle de poids pair, et réciproquement, nous avons $\tilde{C}_{0,0}(z) = \frac{1}{2} c(2z)$ où $\tilde{C}_{0,0}(z)$ est la série génératrice des cycles (lisses) de poids pairs. Par substitution, nous avons

$$C_{0,0}(z) = \tilde{C}_{0,0}(T(z)) = \frac{1}{4} \log \left(\frac{1}{1 - 2T(z)} \right) - \frac{1}{2} T(z) - \frac{1}{2} T(z)^2 \quad (2.4)$$

Notons par $\partial_z = z \frac{\partial}{\partial z}$ l'opérateur différentiel pour distinguer un sommet parmi les autres. En particulier nous avons de l'équation (2.2) avec $T \equiv T(z)$ que

$$\partial_z(T) = \frac{T}{1 - 2T} \quad \partial_z^2(T) = \frac{T}{(1 - 2T)^3} \quad (2.5)$$

Cela nous permet d'obtenir les séries génératrices $(C_{0,\ell})_{\ell \geq 1}$ qui satisfont l'équation différentielle définie de la manière suivante dans [9]

Théorème 1 Pour $\ell \geq 0$, $C_{0,\ell}(z)$ vérifie l'équation différentielle

$$\begin{aligned} (1 - 2T) \partial_z C_{0,\ell+1}(z) + (\ell + 1) C_{0,\ell+1}(z) &= \frac{\partial_z^2 - 3\partial_z - 2\ell}{2} C_{0,\ell}(z) \\ &+ \sum_{p=0}^{\ell} (\partial_z C_{0,p}(z)) (\partial_z C_{0,\ell-p}(z)) \end{aligned} \quad (2.6)$$

où T est donnée par l'équation (2.2)

Preuve Voir [9]. ■

Pour $\ell = 1$ et $\ell = 2$, on a respectivement

$$C_{0,1}(z) = \frac{2T^4(3 - T)}{3(1 - 2T)^3} \quad C_{0,2}(z) = \frac{T^4(1 + 28T - 46T^2 + 36T^3 - 8T^4)}{3(1 - 2T)^6}$$

Une décomposition en élément simple de $C_{0,1}(z)$ par rapport à T donne

$$C_{0,1}(z) = \frac{1}{12} T^2 - \frac{1}{8} T - \frac{1}{4} + \frac{5}{48} \frac{1}{(1 - 2T)^3} - \frac{19}{48} \frac{1}{(1 - 2T)^2} + \frac{13}{24} \frac{1}{(1 - 2T)}$$

On considère les composantes complexes² et on définit l'*excès total* d'un graphe comme suit : l'excès total d'un graphe est le nombre des arêtes plus le nombre des arbres moins

²Les composantes complexes sont les graphes connexes d'excès $\ell \geq 1$

le nombre des sommets. Notant par $E_{0,r}(z)$ la série génératrice exponentielle de tous les graphes complexes (connexes ou non) sans cycles de poids impairs et d'excès total positif r , $E_{0,0}(z) = 1$, $E_{0,1}(z) = C_{0,1}(z)$, nous avons la relation suivante [9]

$$E_{0,r}(z) = [x^r] \exp \left(\sum_{i=1}^{\infty} x^i C_{0,i}(z) \right) \quad (2.7)$$

2.1.2 Des séries génératrices à la satisfaisabilité

Les séries génératrices $C_{0,\ell}(z)$ et $E_{0,\ell}(z)$ vérifie les inégalités "à la Wright" ci-dessous — qui sont des encadrements d'une série quelconque par des séries génératrices simples [14, 11]. Des résultats précis sur la satisfaisabilité d'une formule aléatoire 2-XORSAT résultent de ces derniers, notamment dans la fenêtre critique $\frac{n}{2}(1 + \mu n^{-1/3})$ [9]. Il montre une bonne illustration sur les apports de la Combinatoire Analytique dans un problème de ce type. Nous disons $f(z) \preceq g(z)$ si et seulement si $[z^n]f(z) \leq [z^n]g(z)$ pour tout $n \geq 3$ entier, où $[z^n]f(z)$ représente le coefficient de z^n de la série génératrice $f(z)$.

Théorème 2 Soient $b_1 = \frac{5}{48}$, $c_1 = \frac{19}{48}$. Pour $\ell \geq 1$

$$\begin{aligned} 2(\ell+1)b_{\ell+1} &= 3\ell(\ell+1)b_{\ell} + 6 \sum_{p=1}^{\ell-1} p(\ell-p)b_p b_{\ell-p} \quad \text{et} \\ 3(3\ell+2)c_{\ell+1} &= 8(\ell+1)b_{\ell+1} + 3\ell b_{\ell} + (3\ell-1)(3\ell+2)c_{\ell} + 12 \sum_{p=1}^{\ell-1} p(3\ell-3p-1)b_p c_{\ell-p} \end{aligned} \quad (2.8)$$

Alors on a , pour $\ell \geq 1$

$$\frac{b_{\ell}}{(1-2T)^{3\ell}} - \frac{c_{\ell}}{(1-2T)^{3\ell-1}} \preceq C_{0,\ell}(z) \preceq \frac{b_{\ell}}{(1-2T)^{3\ell}} \quad (2.9)$$

Preuve Voir [9]. ■

Corollaire 1 Nous avons $E_{0,0}(z) = 1$ et $E_{0,1}(z) = C_{0,1}(z)$. Pour $\ell \geq 1$, $E_{0,\ell}(z)$ satisfait

$$\frac{e_{\ell}}{(1-2T)^{3\ell}} - \frac{f_{\ell}}{(1-2T)^{3\ell-1}} \preceq E_{0,\ell}(z) \preceq \frac{e_{\ell}}{(1-2T)^{3\ell}} \quad (2.10)$$

où les suites (b_{ℓ}) et (c_{ℓ}) sont définies par les équations (2.8); $f_1 = c_1 = \frac{19}{48}$ et e_{ℓ}, f_{ℓ} satisfont

$$e_{\ell} = \frac{1}{\ell} \sum_{k=1}^{\ell} k b_k e_{\ell-k} \quad (2.11)$$

$$f_{\ell} = c_{\ell} + \frac{1}{\ell} \sum_{k=1}^{\ell-1} (k c_k e_{\ell-k} + k b_k f_{\ell-k}), \quad (\ell \geq 2) \quad (2.12)$$

Preuve Voir [9]. ■

Théorème 3 *Soit une constante μ , réelle, fixée. La probabilité telle qu'une formule aléatoire 2-XORSAT de n variables et $m = \frac{n}{2}(1 + \mu n^{-1/3})$ clauses soit satisfaisable vérifie*

$$\lim_{n \rightarrow +\infty} n^{1/12} p(n, m) = \left(\sum_{r=0}^{\infty} \frac{\sqrt{2\pi} e^{1/4} e_r}{2^r} A(3r + 1/4, \mu) \right) \quad (2.13)$$

où la suite $(e_r)_{r \in \mathbf{N}}$ et le fonction A satisfont

$$\sum_{r=0}^{\infty} e_r x^r = \exp \left(\sum_{r=1}^{\infty} \frac{(6r)!}{2^{5r-1} 3^{2r} (3r)! (2r)!} x^r \right) \quad (2.14)$$

$$A(y, \mu) = \frac{e^{-\mu^3/6}}{3^{(1+y)/3}} \sum_{k \geq 0} \frac{\left(\frac{1}{2} 3^{2/3} \mu\right)^k}{k! \Gamma((y+1-2k)/3)} \quad (2.15)$$

Preuve Voir [9]. ■

2.2 Approche de MAX-2-XORSAT par les graphes

Tous ce qui suit est **nouveau** et constitue **l'essentiel de ce stage**. On s'intéresse aux graphes avec cycles de poids impairs. Il s'agit de donner une approche sur les calculs des séries génératrices $C_{s,\ell}(z)$ pour les objets de grandes tailles. Nous donnons une étude complète des bicycles ($\ell = 1$) avant d'introduire la méthode pour l'énumération asymptotique.

2.2.1 Le cas particulier $\ell = 1$

Rappelons que la série génératrice des unicycliques avec cycles de poids impairs (une arête est alors à supprimer) est donnée par l'équation (2.4) :

$$C_{1,0}(z) = \frac{1}{4} \log \left(\frac{1}{1-2T} \right) - \frac{T}{2} - \frac{T^2}{2} \quad (2.16)$$

Notons que pour avoir un bicycle dont une arête est à supprimer, soit trois chemins de poids différents³ sont joints en deux sommets par ses extrémités, soit deux cycles de poids différents sont reliés par un chemin. Si on note par $\mathcal{L}$ la famille de tous les chemins de longueur ≥ 2 et si on trace par w et u les arêtes de poids 1 et 0 respectivement, nous avons

$$\mathcal{L} + \mathcal{L} \cong \mathcal{Z} \times \text{SEQ}_{\geq 1}((w+u)\mathcal{Z})$$

Donc la série génératrice (multivariée) est

$$L(w, u, z) = \frac{1}{2} \frac{(u+w) z^2}{1 - (u+w) z} \quad (2.17)$$

³Nous disons que deux objets sont de poids différents si l'un est de poids pair et l'autre de poids impair

où L est la série génératrice associée à $\mathcal{L}$.

On a donc respectivement les séries génératrices des chemins de poids pairs P et des chemins de poids impairs I

$$P(w, u, z) = \frac{L(w, u, z) + L(-w, u, z)}{2} = \frac{1}{4} \frac{(u+w)z^2}{1-(u+w)z} + \frac{1}{4} \frac{(u-w)z^2}{1-(u-w)z} \quad (2.18)$$

$$I(w, u, z) = \frac{L(w, u, z) - L(-w, u, z)}{2} = \frac{1}{4} \frac{(u+w)z^2}{1-(u+w)z} - \frac{1}{4} \frac{(u-w)z^2}{1-(u-w)z} \quad (2.19)$$

Et en terme de série génératrice manovariée, (2.18) et (2.19) donnent

$$P(z) = P(1, 1, z) = \frac{1}{2} \frac{z^2}{1-2z} = I(1, 1, z) = I(z)$$

Les bicycliques dont une arête est à éliminer

Pour les bicycliques formés par trois chemins joints en deux sommets, nous avons les cas suivants :

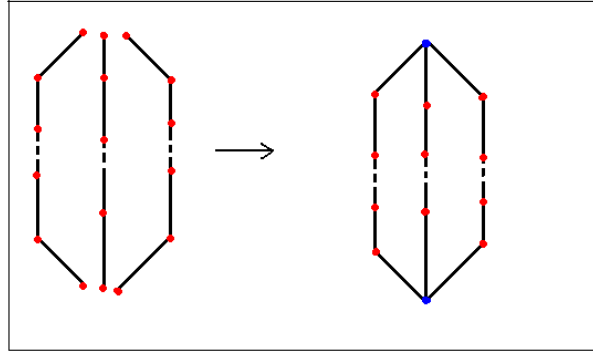


FIG. 2.1 – Trois chemins joints en deux sommets

Si tous les chemins sont de longueur ≥ 2 , la procédure de construction ci-dessus nous donne la spécification suivante, en désignant respectivement par $\mathcal{P}$ et $\mathcal{I}$ les classes des chemins de poids pairs et de poids impairs et en remarquant qu'il existe deux manières de placer un chemin

$$\text{SET}_2(2\mathcal{P}_{\geq 2}) \times 2\mathcal{I}_{\geq 2} + 2\mathcal{P}_{\geq 2} \times \text{SET}_2(2\mathcal{I}_{\geq 2}) \quad (2.20)$$

Si un chemin est de longueur 1 (les deux autres sont nécessairement de longueur au moins 2), on a

$$\begin{aligned} & \text{SET}_2(2\mathcal{P}_{\geq 2}) \times 2\mathcal{I}_{=1} + \text{SET}_2(2\mathcal{I}_{\geq 2}) \times 2\mathcal{P}_{=1} + \\ & (2\mathcal{P}_{\geq 2}) \times (2\mathcal{I}_{\geq 2}) \times (2\mathcal{I}_{=1}) + (2\mathcal{P}_{\geq 2}) \times (2\mathcal{I}_{\geq 2}) \times (2\mathcal{P}_{=1}) \end{aligned} \quad (2.21)$$

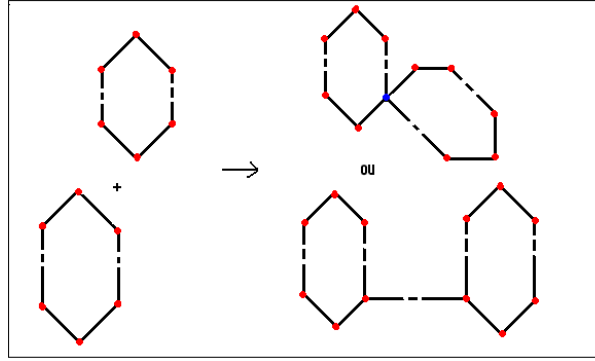


FIG. 2.2 – Deux cycles connectés par un chemin

Si le bicycle est construit à partir de deux cycles de poids différents reliés par un chemin quelconque, nous avons les spécifications suivantes en notant par $\mathcal{C}_p$ la classe des cycles de poids pairs et par $\mathcal{C}_i$ la classe des cycles de poids impairs

$$\Theta\mathcal{C}_p \times \mathcal{Z} \times \Theta\mathcal{C}_i \quad \text{si le chemin est de longueur 0} \quad (2.22)$$

$$\Theta\mathcal{C}_p \times \mathcal{L} \times \Theta\mathcal{C}_i \quad \text{si le chemin est de longueur } \geq 1 \quad (2.23)$$

Remarquons que dans les constructions (2.20) et (2.21), les deux sommets communs sont comptés trois fois et les configurations ainsi obtenues sont deux à deux symétriques ; dans les constructions (2.22) et (2.23) le(s) sommet(s) partagé(s) par les deux cycles avec le chemin du milieu est(sont) compté(s) deux fois . Alors, en enlevant ces répétitions, la série génératrice (monovariée) des bicyclics dont une arête est à enlever est

$$4 \frac{z^5}{(1-2z)^3} + 6 \frac{z^4}{(1-2z)^2} + 4 \frac{z^5}{(1-2z)^3}$$

Ce qui nous donne, par substitution, la série génératrice $C_{1,1}(z)$ des bicyclics nécessitant la suppression d'une arête pour éliminer les cycles de poids impairs

$$\begin{aligned} C_{1,1}(z) &= 2 \frac{T^4(3-2T)}{(1-2T)^3} \\ &= \frac{1}{2} T^2 - \frac{3}{8} + \frac{1}{4} \frac{1}{(1-2T)^3} - \frac{7}{8} \frac{1}{(1-2T)^2} + \frac{1}{(1-2T)} \end{aligned} \quad (2.24)$$

Les bicyclics dont deux arêtes sont à éliminer

Ces composantes sont formées de deux cycles de poids impairs joints par un chemin. Nous avons les constructions suivantes

$$\text{SET}_2(\Theta\mathcal{C}_i) \times \mathcal{Z} \quad \text{si le chemin est de longueur 0} \quad (2.25)$$

$$\text{SET}_2(\Theta\mathcal{C}_i) \times \mathcal{L} \quad \text{si le chemin est de longueur } \geq 1 \quad (2.26)$$

Comme dans les cas (2.22) et (2.23) des sommets sont partagés, alors par substitution nous avons

$$\begin{aligned} C_{2,1}(z) &= 2 \frac{T^5}{(1-2T)^3} \\ &= -\frac{1}{4}T^2 - \frac{3}{8}T - \frac{3}{8} + \frac{1}{16} \frac{1}{(1-2T)^3} - \frac{5}{16} \frac{1}{(1-2T)^2} + \frac{5}{8} \frac{1}{(1-2T)} \end{aligned} \quad (2.27)$$

2.2.2 Les composantes multicycliques

Pour $\ell \geq 2$ et $s \geq 1$, nous allons déterminer les séries génératrices $C'_{s,\ell}$ associées aux ℓ – cycliques ayant les s arêtes à éliminer pointées. Ce pointage est dû à la construction des ℓ – cycliques suivante.

La série $\frac{\partial_z^2 - \partial_z}{2} C_{0,\ell}$ compte un graphe connexe à ℓ excès sans cycles de poids impairs dont deux sommets quelconques sont distincts parmi les autres. Il n'existe qu'une et une seule manière de rajouter une arête w (resp. u) pour obtenir un (ou plusieurs) cycle(s) de poids impairs à partir de ce graphe :

- (i) soit on place cette arête w (resp. u) et on a des cycles de poids impairs
- (ii) soit on ne place pas cette arête w (resp. u)

Comme le nombre de la construction (i) est égal au nombre de la construction (ii), la série $\frac{\partial_z^2 - \partial_z}{4} C_{0,\ell}$ compte donc un graphe connexe à ℓ excès sans cycles de poids impairs dont deux sommets sont pointés et l'ajout d'une arête w (resp. u) entre ces deux sommets crée un ou plusieurs cycles de poids impairs. Donc nous avons :

$$C'_{1,\ell+1}(w, u, z) = (u + w) \frac{\partial_z^2 - \partial_z}{4} C_{0,\ell}(w, u, z) \quad (2.28)$$

En terme de série génératrice monovariée on a :

$$\begin{aligned} C'_{1,\ell+1}(z) &= (1 + 1) \frac{\partial_z^2 - \partial_z}{4} C_{0,\ell}(1, 1, z) \\ &= \frac{\partial_z^2 - \partial_z}{2} C_{0,\ell}(z) \end{aligned}$$

D'une manière réursive on a :

$$\begin{aligned}
C'_{1,\ell+1}(z) &= \frac{\partial_z^2 - \partial_z}{2} C'_{0,\ell}(z) \\
C'_{2,\ell+1}(z) &= \frac{\partial_z^2 - \partial_z}{2} C'_{1,\ell}(z) \\
&= \left(\frac{\partial_z^2 - \partial_z}{2} \right)^2 C'_{0,\ell}(z) \\
&\vdots \\
C'_{s+1,\ell+1}(z) &= \frac{\partial_z^2 - \partial_z}{2} C'_{s,\ell}(z) \\
&= \left(\frac{\partial_z^2 - \partial_z}{2} \right)^s C'_{1,\ell-s+1}(z) \\
&= \left(\frac{\partial_z^2 - \partial_z}{2} \right)^{s+1} C'_{0,\ell-s}(z)
\end{aligned}$$

D'où le théorème suivant :

Théorème 4 Nous avons pour $s = 0, \dots, \ell + 1$ et $\ell \geq 0$:

$$C'_{0,\ell}(z) = C_{0,\ell}(z) \quad \text{et} \quad C'_{s,\ell}(z) = \left(\frac{\partial_z^2 - \partial_z}{2} \right)^s C_{0,\ell-s}(z)$$

Proposition 1 Soit G un ℓ -cyclique à n sommets (sans poids sur les arêtes). Parmi les $2^{n+\ell}$ pondérations possibles, il y en a $\binom{\ell+1}{s} 2^{n-1}$ qui fournissent des ℓ -cycliques contenant exactement s cycles fondamentaux de poids impairs.

Preuve Soit Σ un arbre couvrant de G (Σ compte $n - 1$ arêtes). Chaque arête de $G \setminus \Sigma$ (au total $\ell + 1$) forme un cycle fondamental. Observons que la pondération de G est équivalente à la pondération d'un arbre couvrant de G et la pondération des $\ell + 1$ arêtes restantes déterminant les cycles fondamentaux. À une pondération de Σ , il existe $\binom{\ell+1}{s}$ manière de générer exactement s cycles fondamentaux de poids impairs. Comme il existe 2^{n-1} façons de pondérer Σ , un ℓ -cyclique fournit $\binom{\ell+1}{s} 2^{n-1}$ ℓ -cycliques ayant exactement s cycles fondamentaux de poids impairs. ■

Proposition 2 Si on note par $M_{s,\ell}(z)$ la série génératrice des ℓ -cycliques qui comportent exactement s cycles fondamentaux de poids impairs, on a

$$M_{s,\ell}(z) = \frac{\binom{\ell+1}{s}}{2} W_\ell(2z) \tag{2.29}$$

et

$$\frac{\beta_{s,\ell}}{(1-2T)^{3\ell}} - \frac{\alpha_{s,\ell}}{(1-2T)^{3\ell-1}} \preceq M_{s,\ell}(z) \preceq \frac{\beta_{s,\ell}}{(1-2T)^{3\ell}} \tag{2.30}$$

avec

$$\alpha_{s,\ell} = \binom{\ell+1}{s} \frac{c_\ell}{2} \quad \text{et} \quad \beta_{s,\ell} = \binom{\ell+1}{s} \frac{b_\ell}{2} \tag{2.31}$$

où $W_\ell(z)$ est la série génératrice des ℓ -cycliques sans poids sur les arêtes et les constantes c_ℓ et b_ℓ sont données par les équations (2.8).

Preuve C'est une conséquence directe de la proposition précédente et les inégalités de Wright. ■

2.3 Énumérations asymptotiques

Montrons d'abord que pour s et ℓ fixes, il existe $c_{s,\ell}$ telle que au voisinage de sa singularité, la série génératrice des ℓ – *cycliques* lisses $\tilde{C}_{s,\ell}(z)$ nécessitant la suppression de s arête pour éliminer les cycles de poids impairs vérifie

$$\tilde{C}_{s,\ell}(z) \sim \frac{c_{s,\ell}}{(1-2z)^{3\ell}} \quad (2.32)$$

Rappelons que il existe un polynome P_ℓ tel que la série génératrice $\tilde{W}_\ell(z)$ des ℓ – *cycliques* (lisses) est

$$\tilde{W}_\ell(z) = \frac{P_\ell(z)}{(1-z)^{3\ell}} \quad (2.33)$$

Vu la relation entre les composantes avec et sans poids sur les arêtes, il existe un polynome $Q_{s,\ell}(z)$ tel que

$$\tilde{C}_{s,\ell}(z) = \frac{Q_{s,\ell}(z)}{(1-2z)^{3\ell}} \quad (2.34)$$

Ce qui explique l'existence de la constante $c_{s,\ell}$; pour $z \rightarrow \frac{1}{2}$,

$$\tilde{C}_{s,\ell}(z) \sim \frac{c_{s,\ell}}{(1-2z)^{3\ell}}$$

De la même manière, si on considère la série génératrice $C_{s,\ell}^{(2)}(z)$ des ℓ – *cycliques* lisses avec isthme ou comme deux blocs articulés à un sommet (voir figure 2.3) qui nécessitent une suppression de s arêtes pour éliminer tous les cycles de poids impairs, il existe une constante $c_{s,\ell}^{(2)}$ telle qu'au voisinage de sa singularité, $C_{s,\ell}^{(2)}(z)$ vérifie

$$C_{s,\ell}^{(2)}(z) \sim \frac{c_{s,\ell}^{(2)}}{(1-2z)^{3\ell}} \quad (2.35)$$

Exemple :

Pour $s = 0$ et $\ell \geq 1$, de l'inégalité (2.9), nous avons

$$[z^n]C_{0,\ell}(z) \sim [z^n] \frac{c_{0,\ell}}{(1-2z)^{3\ell}}$$

avec $c_{0,\ell} = b_\ell$, en particulier $c_{0,1} = \frac{5}{48}$; si $s = 1$, $s = 2$ et $\ell = 1$, (2.24) et (2.27) donne $c_{1,1} = 1/4$ et $c_{2,1} = 1/16$.

Observons que pour $\ell = 1$, ces coefficients directeurs $c_{s,1}$ sont reliés par la manière de construire un bicyclic, car seuls les coefficients directeurs des séries génératrices dominantes de chaque objet mis en jeu déterminent les valeurs de $c_{s,\ell}$. Par exemple pour $s = 2$ et $\ell = 1$ la série génératrice des cycles impairs est dominée par $\frac{1}{4} \log(\frac{1}{1-2z})$. En

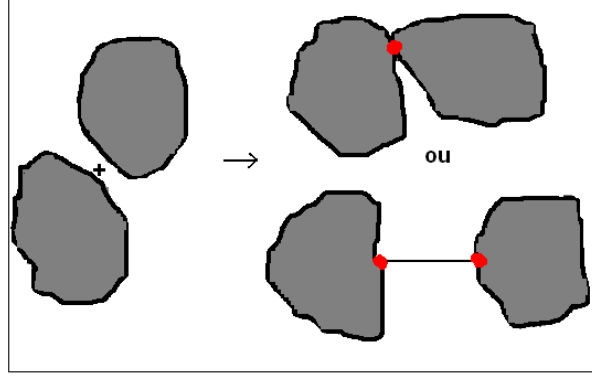


FIG. 2.3 – Deux blocs connectés par un chemin

pointant un sommet des deux cycles impairs à relier, la série génératrice d'un cycle impair pointé est dominée par $\frac{1}{4} \frac{1}{1-2z}$ et nous avons

$$c_{2,1} = 2 \times \underbrace{\frac{1}{2}}_{\text{facteur de symétrie}} \times \underbrace{\frac{1}{4}}_{\text{cycle pair pointé}} \times \underbrace{\frac{1}{4}}_{\text{cycle impair pointé}}$$

Une généralisation de la méthode serait donc souhaitable.

Proposition 3 Pour $s \geq 1$ et pour $\ell \geq 2$, on a

$$c_{s,\ell}^{(2)} = \frac{9}{2} \sum_{r=0}^s \sum_{p=1}^{\ell-2} p(\ell-1-p) c_{r,p} c_{s-r,\ell-1-p} + \frac{3}{2} \sum_{r=0}^1 (\ell-1) c_{r,0} c_{s-r,\ell-1} \quad (2.36)$$

avec $c_{1,0} = c_{0,0} = \frac{1}{4}$

Preuve Pour créer ce graphe, on part de deux blocs : p – *cyclique* dont r arêtes sont à éliminer plus un $(\ell-1-p)$ – *cyclique* dont $s-r$ arêtes sont à éliminer. Pour $z \rightarrow \frac{1}{2}$, les séries génératrices approchées des deux composantes multicycliques sont respectivement

$$\frac{c_{r,p}}{(1-2z)^{3p}} \quad \text{et} \quad \frac{c_{s-r,\ell-1-p}}{(1-2z)^{3(\ell-1-p)}}$$

En passant au pointage de chaque côté, on a

$$6pz \frac{c_{r,p}}{(1-2z)^{3p+1}} \quad 6(\ell-1-p)z \frac{c_{s-r,\ell-1-p}}{(1-2z)^{3(\ell-1-p)+1}}$$

En décomposant en élément simple et en ne récupérant que la série génératrice dominante, nous avons

$$6pz \frac{c_{r,p}}{(1-2z)^{3p+1}} \sim 3p \frac{c_{r,p}}{(1-2z)^{3p+1}} \quad (2.37)$$

$$6(\ell-1-p)z \frac{c_{s-r,\ell-1-p}}{(1-2z)^{3(\ell-1-p)+1}} \sim 3(\ell-1-p) \frac{c_{s-r,\ell-1-p}}{(1-2z)^{3(\ell-1-p)+1}} \quad (2.38)$$

Le produit cartésien donne avec le chemin du milieu

$$\frac{9p(\ell - 1 - p)c_{r,p}c_{s-r,\ell-1-p}}{(1 - 2z)^{3\ell}}$$

En sommant sur p et sur r , nous avons

$$\frac{9}{2} \sum_{r=0}^s \sum_{p=1}^{\ell-2} p(\ell - 1 - p)c_{r,p}c_{s-r,\ell-1-p}$$

ce qui explique le premier terme du second membre.

Si $p = 0$, les composantes concernées sont un cycle et un $(\ell - 1)$ -*cyclique*, et nous avons

$$\frac{3}{2} \sum_{r=0}^1 (\ell - 1)c_{r,0}c_{s-r,\ell-1}$$

qui donne le second terme du second membre. ■

Dans ce chapitre, nous avons apporté un éclairage nouveau sur le problème MAX-2-XORSAT. En effet, en donnant une méthode pour trouver les coefficients directeurs des séries génératrices (voir Proposition 3), nous pouvons aller plus en avant pour étudier la structure des solutions de ce problème difficile. C'est le but du prochain chapitre.

Chapitre 3

Les instances aléatoires MAX-2-XORSAT

Dans cette partie, on s'intéresse au nombre maximum de clauses satisfaisables en phase sous critique. Rappelons que $\mathcal{P}(n, m)$ désigne la classe des formules à m clauses sur n variables booléennes. Soit $\Pi(n, m) \in \mathcal{P}(n, m)$. Si $\Pi(n, m)$ n'est pas satisfaisable, que peut-on dire à propos du nombre maximum de clauses satisfaisables ? Pour ce faire, nous étudions le graphe associé à la formule $\Pi(n, m)$. Rappelons qu'une formule aléatoire n'est pas satisfaisable si et seulement si le graphe la représentant contient au moins un cycle de poids impairs. Donc, maximiser le nombre de clauses satisfaisables revient à minimiser le nombre de suppression d'arêtes pour éliminer tous les cycles de poids impairs. Notons par $X_{n,m}$ la variable aléatoire associée au nombre de suppression d'arêtes nécessaire pour éliminer tous les cycles de poids impairs, par $Z_{n,m}$ le nombre maximum de clauses satisfaisables. Donc, nous avons la relation $Z_{n,m} = m - X_{n,m}$.

3.1 Clauses satisfaisables en phase sous critique

Si $m < \frac{1}{2}n$, alors deux cas peuvent se présenter, ou bien $m/n \rightarrow 0$ ou bien il existe $c \in]0, \frac{1}{2}[$ tel que $\frac{m}{n} \rightarrow c$ pour $n \rightarrow \infty$.

Pour $m \ll n$,¹ le graphe ne contient que des arbres presque sûrement, c'est à dire avec une probabilité $1 - o(1)$ lorsque n et m tendent vers l'infini. Donc la formule est satisfaisable presque sûrement et $X_{n,m(n)} = 0$ et $Z_{n,m(n)} = m(n)$.

Pour $m = cn$ avec $0 < c < \frac{1}{2}$, nous avons le théorème suivant

Théorème 5 Pour R fixé, $n \rightarrow \infty$ et $m = cn$, $c \in]0, \frac{1}{2}[$

$$\mathbb{P}([X_{n,cn} = R]) = \frac{\lambda(c)^R}{R!} e^{-\lambda(c)-c^2/2} (1 + o(1)) \quad (3.1)$$

avec

$$\lambda(c) = \frac{1}{4} \log \left(\frac{1}{1-2c} \right) - \frac{c}{2} - \frac{c^2}{2}$$

Preuve On va d'abord déterminer la probabilité qu'une formule aléatoire $\Pi(n, m)$ ne contienne que

¹ $f \ll g$ si et seulement si $f(n)/g(n) \rightarrow 0$ quand $n \rightarrow \infty$

- (i) des arbres
- (ii) des cycles de poids pairs
- (iii) R cycles de poids impairs, R fixé

Notons par $F_R(z)$ la série génératrice exponentielle associée aux objets satisfaisant les conditions (i), (ii) et (iii) ci-dessus. Notons par $\kappa_{n,m}$ la variable aléatoire associée au nombre de cycles de poids impairs dans le graphe. Nous avons au total $m = cn$ clauses et n variables. Le graphe ne contient que des arbres et des cycles, alors le nombre de composantes acycliques est exactement $n - m$. En utilisant le tableau 1.4 et les équations (2.3), (2.4) et (2.16), on a

$$F_R(z) = \frac{(T - T^2)^{n-m}}{(n-m)!} \exp(C_{0,0}(z)) \frac{C_{1,0}(z)^R}{R!} \quad (3.2)$$

Calculons d'abord $\mathbb{P}([\kappa_{n,m} = R])$:

$$\mathbb{P}([\kappa_{n,m} = R]) = \frac{n!}{\binom{n(n-1)}{cn}} [z^n] F_R(z) \quad (3.3)$$

D'après la formule intégrale de Cauchy on a

$$[z^n] F_R(z) = \frac{1}{2\pi i} \oint \frac{(T - T^2)^{n-m}}{(n-m)!} \frac{e^{-T/2-T^2/2}}{(1-2T)^{1/4}} \frac{dz}{z^{n+1}} \quad (3.4)$$

En posant $u = 2T(z)$, on a $z = \frac{1}{2}ue^{-u}$ et $dz = \frac{1}{2}(1-u)e^{-u}$ et l'équation (3.3) devient :

$$(3.3) = \frac{1}{\binom{n(n-1)}{m}} \frac{n!}{(n-m)!} \frac{e^n 4^m}{2^n} \frac{1}{2\pi i} \oint g(u) \exp(nh(u)) \frac{du}{u} \quad (3.5)$$

avec

$$\begin{aligned} g(u) &= (1-3u)^{3/4} e^{-u/4-u^2/8} \frac{\left(\frac{1}{4} \log\left(\frac{1}{1-u}\right) - \frac{u}{4} - \frac{u^2}{8}\right)^R}{R!} \\ h(u) &= u - 1 - \log(u) + (1-c) \log(2u - u^2) \end{aligned}$$

Les points cols sont les solutions par rapport à u de l'équation $\frac{dh(u)}{du} = 0$, ce qui nous donne $u_0 = 1$ et $u_1 = 2c$. Remarquons que $h''(u_1) = \frac{(1-2c)}{4c(1-c)} > 0$. Posons $u = 2c \exp(i\theta)$. Si on développe $h(2c \exp(i\theta))$ d'ordre 3 et $g(2c \exp(i\theta))$ d'ordre 1 au voisinage de $\theta = 0$, nous avons

$$g(2c \exp(i\theta)) = g(2c) + O(\theta) = g(2c)(1 + O(\theta)) \quad (3.6)$$

$$\begin{aligned} nh(2c \exp(i\theta)) &= nh(2c) - n(2c)^2 h''(2c) \frac{\theta^2}{2} + O(n\theta^3) \\ &= nh(2c) - \beta(c) \frac{n\theta^2}{2} + O(n\theta^3) \end{aligned} \quad (3.7)$$

Il suffit de choisir θ_0 de telle manière que $n\theta_0^2 \rightarrow +\infty$ et $n\theta_0^3 \rightarrow 0$, ($\theta_0 = n^{-2/5}$ convient). L'intégration se fait de $-\pi$ à $+\pi$ en choisissant comme contour d'intégration l'ensemble

des z tel que $|z|=2c$. D'abord, la contribution de l'intégrale sur $I_1 = [-\pi, -\theta_0] \cup [\theta_0, \pi]$ est exponentiellement petite. En effet, la fonction définie par

$$\begin{aligned} f(\theta) &= \Re(h(e^{i\theta})) \\ &= 2c \cos \theta - 1 - c \log(2c) + (1-c) \log(4c + 4c^2 - 8c \cos \theta) \end{aligned}$$

est maximale en $\theta = 0$:

$$\begin{aligned} f'(\theta) = 0 &\Leftrightarrow 2c \sin(\theta) \left(\frac{1-c}{1+c^2-2c \cos(\theta)} - 1 \right) = 0 \\ &\Leftrightarrow \theta = 0 \quad (\text{sur } [-\pi, \pi]) \\ &\text{car } \left(\frac{1-c}{1+c^2-2c \cos(\theta)} > \frac{1}{1-c} > 1 \right) \end{aligned} \quad (3.8)$$

Donc le signe de la fonction dérivée f' dépend du signe de la fonction sinus. Alors $\forall \theta \in I_1$, $f(\theta) \leq f(\theta_0) = f(-\theta_0)$ (car la fonction f est paire). C'est à dire que

$$\begin{aligned} \int_{\theta_0}^{2\pi-\theta_0} g(2ce^{i\theta}) e^{nh(2ce^{i\theta})} d\theta &\leq \int_{\theta_0}^{2\pi-\theta_0} g(2ce^{i\theta}) d\theta (\exp(nf(\theta_0))) \\ &= K \exp(nf(\theta_0)) \end{aligned}$$

Or, $c < 1/2$ et $\cos x = 1 - x^2/2 + O(x^4)$ et $\log(1+x) = x + O(x^2)$ pour x voisin de 0, $K \exp(nf(\theta_0)) = \exp(O(-n))$ est exponentiellement faible sur I_1 . Donc la valeur de l'expression sous l'intégrale est concentrée sur $I_0 = [-\theta_0, +\theta_0]$.

$$\frac{1}{2\pi i} \int_{-\theta_0}^{\theta_0} g(u) \exp(nh(u)) \frac{du}{u} = \frac{1}{2\pi} \int_{-\theta_0}^{\theta_0} g(2ce^{i\theta}) \exp(nh(2ce^{i\theta})) d\theta. \quad (3.9)$$

Tenant compte des approximations ci-dessus, (3.9) devient

$$\begin{aligned} (3.9) &= \frac{1}{2\pi} \int_{-\theta_0}^{\theta_0} g(2c) \exp(nh(2c) - n\beta(c)\theta^2/2) d\theta (1 + O(n\theta_0))(1 + O(n\theta_0^3)) \\ &= \frac{1}{2\pi} \int_{-\theta_0}^{\theta_0} g(2c) \exp(nh(2c) - n\beta(c)\theta^2/2) d\theta (1 + o(1)) \end{aligned} \quad (3.10)$$

En utilisant la formule de Stirling, nous avons

$$\begin{aligned} \frac{1}{\binom{n(n-1)}{m}} \frac{n!}{(n-m)!} &\sim \sqrt{2\pi m} \frac{(c-c^2)^m}{(1-c)^n} \frac{e^{c+c/2-2m}}{(1-c)^{1/2}} \\ &= \sqrt{2\pi n} e^{c+c^2/2-2m} (1-c)^{m-n-1/2} c^{m+1/2} \end{aligned}$$

Les formes d'erreurs qui se cachent derrière le symbole $\sim$ proviennent de

$$\begin{aligned} (n(n-1)-m)^m &= n^m (n-1)^m \left(e^{-c^2} (1 + O(\frac{1}{n})) \right) \\ (n-1)^m &= n^m e^{-c} (1 + O(\frac{1}{n})) \\ (n(n-1)-m)^{n(n-1)} &= (n(n-1))^{n(n-1)} \left(e^{-m-c^2/2} (1 + O(\frac{1}{n})) \right) \\ N! &= \sqrt{2\pi N} (N/e)^N \left(1 + O(\frac{1}{N}) \right) \end{aligned}$$

Alors, on a

$$\frac{1}{\binom{n(n-1)}{m}} \frac{n!}{(n-m)!} = \sqrt{2\pi n} e^{c+c^2/2-2m} (1-c)^{m-n-1/2} c^{m+1/2} (1+o(1)) \quad (3.11)$$

En posant $x = \sqrt{n\beta(c)/2}\theta$ nous avons

$$\begin{aligned} \int_{-\theta_0}^{\theta_0} \exp(-n\beta(c)\theta^2/2) d\theta &= \int_{-k\theta_0}^{+k\theta_0} \exp(-x^2) \frac{dx}{\sqrt{n\beta(c)/2}} \\ &= \int_{-k'n^{1/10}}^{+k'n^{1/10}} \exp(-x^2) \frac{dx}{\sqrt{n\beta(c)/2}} \end{aligned}$$

et, en faisant $n \rightarrow +\infty$, nous avons

$$\frac{1}{2\pi} \int_{-\theta_0}^{\theta_0} \exp(-n\beta(c)\theta^2/2) d\theta = \frac{1}{\sqrt{2\pi n\beta(c)}} \quad (3.12)$$

Donc de (3.3), (3.5), (3.11) et (3.12) nous avons

$$\begin{aligned} \mathbb{P}([\kappa_{n,m} = R]) &= \exp(c/2) (1-2c)^{1/4} \frac{\left(\frac{1}{4} \log\left(\frac{1}{1-2c}\right) - \frac{c}{2} - \frac{c^2}{2}\right)^R}{R!} (1+o(1)) \\ &= \frac{\lambda(c)^R}{R!} e^{-\lambda(c)-c^2/2} (1+o(1)) \end{aligned} \quad (3.13)$$

où

$$\lambda(c) = \frac{1}{4} \log\left(\frac{1}{1-2c}\right) - \frac{c}{2} - \frac{c^2}{2}$$

Cherchons une relation entre $X_{n,m}$ et $\kappa_{n,m}$. $X_{n,m}$ compte le nombre total de suppression d'arêtes nécessaire (pour éliminer les cycles de poids impairs dans les composantes unicycliques et multicycliques). Or, la probabilité qu'un graphe aléatoire $\Pi(n, m = cn)$, avec $c < 1/2$, contienne une composante multicyclique est $O(n^{-1/2})$ [1], donc $X_{n,m} = \kappa_{n,m}$. Donc on a

$$\mathbb{P}([X_{n,cn} = R]) = \frac{\lambda(c)^R}{R!} e^{-\lambda(c)-c^2/2} (1+o(1)) \quad (3.14)$$

ce qui achève la démonstration du théorème. ■

Cela veut dire que pour un graphe aléatoire à n sommets et à cn arêtes, avec une probabilité $\mathbb{P}([X_{n,m} = R])$, IL FAUT ENLEVER R ARÊTES pour supprimer tous les cycles de poids impairs. En d'autre terme, nous avons le corollaire suivant vu la relation entre $X_{n,m}$ et $Z_{n,m}$,

Corollaire 2

$$\mathbb{P}([Z_{n,m} = R]) = \frac{\lambda(c)^{m-R}}{(m-R)!} e^{-\lambda(c)-c^2/2} (1+o(1)) \quad (3.15)$$

3.2 MAX-2-XORSAT voisin de la transition

Dans cette section, $m = \frac{n}{2}(1-y)$ où y dépend de n , est à valeur dans $\mathbf{R}_+$ avec $y \rightarrow 0$, mais $y^3 n \rightarrow \infty$ quand $n \rightarrow +\infty$. Et nous avons le nouveau théorème suivant

Théorème 6 *Si $n, m \rightarrow \infty$, $y \rightarrow 0$ et $y^3 n \rightarrow \infty$, alors pour un x fixé,*

$$\mathbb{P} \left(\left[\kappa_{n,m} + \frac{1}{4} \log(y) \leq x \sqrt{-\frac{1}{4} \log(y)} \right] \right) \rightarrow \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du \quad (3.16)$$

Preuve Notons par $\mathcal{G}(n, m)$ la famille des graphes qui ne contiennent pas des composantes complexes. Si $n, m \rightarrow \infty$, $y \rightarrow 0$ et $y^3 n \rightarrow \infty$, alors un graphe aléatoire à n sommets et à m arêtes appartient à $\mathcal{G}(n, m)$ presque sûrement [5], ce qui explique toujours qu'asymptotiquement $X_{n,m} = \kappa_{n,m}$.

On note par $\mathcal{K}_n$ la famille des graphes de taille n dont les composantes sont unicycliques et par $\mathcal{U}_n$ la classe des forêts de taille n . Et on définit

$$\begin{aligned} \mathcal{K}_{n,N} &= \{G \in \mathcal{K}_n / G \text{ contient exactement } N \text{ composantes connexes}\} \\ \mathcal{U}_{n,N} &= \{G \in \mathcal{U}_n / G \text{ contient exactement } N \text{ composantes connexes}\} \end{aligned}$$

On note par $k_n = |\mathcal{K}_n|$, $k_{n,N} = |\mathcal{K}_{n,N}|$ et $u_{n,N} = |\mathcal{U}_{n,N}|$.

Si on ignore la pondération sur les arêtes, le nombre $\bar{k}_n$ de graphes ne contenant que des composantes unicycliques, le nombre $\bar{k}_{n,N}$ de graphes ne contenant que N composantes unicycliques et le nombre $\bar{u}_{n,N}$ de forêts formées de N arbres à n sommets sont donnés respectivement par le Théorème 3, le Lemme 3 et le Théorème 6 de [5] :

pour $n \rightarrow \infty$

$$\bar{k}_n = \frac{n^{n-1/4} \sqrt{2\pi}}{e^{3/4} 2^{1/4} \Gamma(\frac{1}{4})} (1 + o(1)), \quad (3.17)$$

pour $\lambda_n = \frac{1}{4} \log(n)$, $u = (N - \lambda_n) / \sqrt{\lambda_n}$ avec $|u| \leq \sqrt[4]{\log(n)}$

$$\bar{k}_{n,N} = \frac{n^{n-1/4} \sqrt{2\pi}}{e^{3/4} 2^{1/4} \Gamma(\frac{1}{4})} \frac{\lambda_n^N}{N!} e^{-\lambda_n} \quad (3.18)$$

et pour $n_1 \sqrt{y/n} \rightarrow 0$

$$\bar{u}_{n-n_1, n-m} = \frac{(1-y^2)^{n-m} (n-n_1)!}{2^{n-m} (n-m)! (1-y)^n e^{-n(1-y)}} \sqrt{\frac{y}{2\pi n}} (1-y)^{n_1} e^{-n_1(1-y)} (1 + o(1)) \quad (3.19)$$

Comme un arbre compte une arête de moins que de sommets et il y a autant d'arêtes que de sommets dans un cycle, si on affecte un poids à chaque arête alors nous avons les égalités suivantes

$$k_n = 2^n \bar{k}_n, \quad k_{n,N} = 2^n \bar{k}_{n,N}, \quad u_{n-n_1, n-m} = 2^{m-n_1} \bar{u}_{n-n_1, n-m} \quad (3.20)$$

Par conséquent, le Théorème 4, le Théorème 7 de [5] restent encore valables. Alors, si on note ξ_n le nombre de cycles dans un graphe aléatoire de $\mathcal{K}_n$ et ζ_n le nombre de sommets dans les composantes cycliques d'un graphe aléatoire de $\mathcal{G}(n, m)$, nous avons

$$\mathbb{P}([\xi_n = N]) = \frac{|\mathcal{K}_{n,N}|}{|\mathcal{K}_n|} = \frac{\lambda_n^N}{N!} e^{-\lambda_n} (1 + o(1)) = \frac{2}{\sqrt{2\pi \log(n)}} e^{-u^2/2} (1 + o(1)) \quad (3.21)$$

et que la probabilité que le nombre de sommets dans les composantes unicycliques est

$$\mathbb{P}([\zeta_n = n_1]) = \frac{1}{\Gamma(\frac{1}{4})} t^{-3/4} e^{-t} \frac{y^2}{2} (1 + o(1)) \quad (3.22)$$

pour $t = y^2 n / 2$ élément d'un intervalle donné $[t_0, t_1] \subset \mathbf{R}_+$.

L'équation (3.21) nous explique que le nombre ξ_n de composantes connexes contenant un cycle dans un graphe aléatoire tend vers une loi limite loi de Poisson $\mathcal{P}(\frac{1}{4} \log(n))$ ou $\mathcal{N}(\frac{1}{4} \log(n), \frac{1}{4} \log(n))$. Comme le nombre de cycles de poids impairs est la moitié de tous les cycles, alors si Q_n désigne la variable aléatoire associée au nombre de composantes connexes qui contient un cycle de poids impairs est gaussienne $Q_n \rightsquigarrow \mathcal{N}(\frac{1}{8} \log(n), \frac{1}{8} \log(n))$. Donc pour $n_1 \rightarrow \infty$

$$Q_{n_1} \rightsquigarrow \mathcal{N}(\frac{1}{8} \log(n_1), \frac{1}{8} \log(n_1)) \quad (3.23)$$

Or

$$\mathbb{P}([\kappa_{n,m} + \frac{1}{4} \log(y) \leq x \sqrt{-\frac{1}{4} \log(y)}]) = \sum_{n_1=0}^{\infty} \mathbb{P}(v = n_1) \mathbb{P}([Q_{n_1} + \frac{1}{4} \log(y) \leq x \sqrt{-\frac{1}{4} \log(y)}]) \quad (3.24)$$

pour $t = y^2 n_1 / 2$ et $0 < t_0 < t_1 < \infty$. Si $y \rightarrow 0$ alors $\log(n_1) \sim -2 \log(y)$ car t appartient à un intervalle fermé. C'est à dire, $(Q_{n_1} + \frac{1}{4} \log(y)) / \sqrt{-\frac{1}{4} \log(y)}$ tend vers une loi $\mathcal{N}(0, 1)$ d'après (3.23). Donc

$$\mathbb{P}([\kappa_{n,m} + \frac{1}{4} \log(y) \leq x \sqrt{-\frac{1}{4} \log(y)}]) \rightarrow \Phi(x) \quad (3.25)$$

Le théorème s'en déduit car $X_{n,m} = \kappa_{n,m}$. ■

On peut donc dire que pour une formule aléatoire de $\mathcal{P}(n, m)$ avec $m = \frac{n}{2}(1 - y)$ où $y \rightarrow 0$ mais $y^3 n \rightarrow \infty$ quand $n \rightarrow +\infty$, par exemple $y = \log(n) n^{-1/3}$, la distribution du nombre de suppression d'arêtes nécessaire pour éliminer tous les cycles de poids impairs est centrée en $-\frac{1}{4} \log(y)$. On est passé de la loi de Poisson à la loi de Laplace-Gauss ; $X_{n,m}$ est donc asymptotiquement normale. L'évasement de cette distribution est $\sqrt{-\frac{1}{4} \log(y)}$. Donc en moyenne, le nombre maximum de clauses satisfaisables est $m + \frac{1}{4} \log(y)$.

Conclusion

Les résultats que nous avons obtenus se limitent en phase sous critique : la valeur de la densité de clauses par variables ne passe pas le point critique $\frac{1}{2}$. Cela nous a permis de donner des résultats précis sur l'étude du nombre maximum de clauses satisfaisables, car à ce stade, tous les cycles du graphe $G(n, m)$ représentant une formule $\Pi(n, m)$ de MAX-2-XORSAT à n variables et à m clauses sont élémentaires presque sûrement. En d'autre terme, pour R fixé, une suppression de R arêtes pour éliminer tous les cycles de poids signifient que $G(n, m)$ contient R cycles de poids impairs avec une probabilité tendant vers 1 pour $n \rightarrow \infty$.

Un prolongement naturel de ce travail sera donc une généralisation de ce que nous venons de commencer dans la section 3.2., c'est à dire dès que $m = \frac{n}{2} + \frac{yn^{2/3}}{2}$. À ce moment, deux cas seront à distinguer : dans un premier temps y sera un réel fixe et dans un second temps y varie avec la taille de la formule et que $y = y(n) \rightarrow \infty$ si $n \rightarrow \infty$. Pour ce faire, on fera appel aux coefficients directeurs $c_{s,\ell}$ de l'énumération asymptotique que nous avons initiée au chapitre 2 pour étudier les composantes complexes, car dans ces régimes le graphe contient un sous graphe complexe (connexe ou non) d'excès total positif ℓ' dont s' arêtes sont à éliminer pour rendre ce graphe à satisfaire la propriété « sans cycles de poids impairs ».

Bibliographie

- [1] Flajolet, P., Knuth, D.E. and Pittel B.(1989) : The first cycles in an evolving graph. *Discrete math*, 75 :167-215
- [2] Flajolet, P and Sedgewick, R.(2007) : Analytic Combinatorics. <http://algo.inria.fr/flajolet/Publications/books.html>
- [3] Farro, L. Khin, H. et Salmann, N.(2005) : Résolution pratique des problèmes NP-complets.
- [4] Hao Jin-Kao, Galinier, P. et Habib, M.(1999) : Métaheuristiques pour l'optimisation et l'affectation sous contraintes. *Revue d'Intelligence Artificielle*.
- [5] Kolchin, V.F (1986) : On the behavior on a random graph near a critical point. *Theory Probab.Appl*, 31 :439-452
- [6] Mézard, M., Parisi, G. et Virasiro, M. (1987) : Spin glass theory and beyond. *World Scientific*, Singapoure.
- [7] Monasson Rémi(2006) : Transitions de phases dans les problèmes d'optimisations. *Seminaire MIP, exposé*.
- [8] Semerjian, G.(2004) : Modèles dilués en physique statistique : Dynamiques hors d'équilibre et algorithme d'optimisation. *Thèse de Doctorat*.
- [9] Ravelomanana, V. and Daudé, H.(2007) : Random 2-XORSAT at the satisfiability threshold. *Lecture Notes in Computer Science*. Springer Verlag.
- [10] Ravelomanana, V(2006) : The Average Size of Giant Components between the Double-Jump.*Algorithmica*, 46 :529-555
- [11] Ravelomanana, V(2007) : Another proof of Wright's inequalities. *Information Processing Letters* 104 :12-23
- [12] Talagrand, M. (1999) : Verres de spin et optimisation combinatoire. *Séminaire Bourbaki*, 51ième année.
- [13] Wright, E., M.(1977) : The Number of Connected Sparsley Edged Graphs. *Journal of the Graph theory*, 1 :317-330.
- [14] Wright, E., M.(1980) : The Number of Connected Sparsley Edged Graphs. III. Asymptotic results. *Journal of the Graph theory*, 4 :393-407.

RASENDRAHASINA Solofoniaina
Ndimbisoa Tinarivo Vonjy
rasendrahasina@gmail.com

Approche analytique d'un problème difficile d'optimisation : MAX-2-XORSAT

Résumé. Nous présentons une approche du problème d'optimisation MAX-2-XORSAT basée sur la combinatoire énumérative et analytique. Dans ce mémoire, nous étudions les séries génératrices liées aux configurations optimales de MAX-2-XORSAT. En combinant ces outils avec ceux d'analyse complexe, nous quantifions alors le nombre maximum de clauses satisfaisables des instances aléatoires de MAX-2-XORSAT, dans la phase sous critique, lorsque ces instances sont peu denses.

Mots clés : graphes aléatoires, séries génératrices, combinatoire analytique, analyse complexe, Analyse en moyenne, transition de phase, SAT (satisfaction d'expressions booléennes), optimisation (MAX-SAT, MAX-CSP).

Analytical approach of hard optimisation problem : MAX-2-XORSAT

Abstract. We present an approach of the MAX-2-XORSAT optimization problem by means of enumerative and analytic combinatorics. In these words, we study the generating functions of optimal configurations of MAX-2-XORSAT. Combining these tools with complex analysis, we then quantify the maximum number of satisfiable clauses of random MAX-2-XORSAT formulas, in the subcritical case, whenever the random formulas are not too dense.

Keywords : random graphs, generating functions, analytic combinatorics, complex analysis, average-case analysis of algorithms, phase transition, SAT (boolean satisfiability problem), optimization (MAX-SAT, MAX-CSP).

ENCADREUR :
Vlady RAVELOMANANA
vlad@lipn.univ-paris13.fr