Quelques problèmes additifs : bases, pseudo-puissances et ensembles k-libres

Victor Lambert

THÈSE

présentée à

L'ÉCOLE POLYTECHNIQUE

CENTRE DE MATHÉMATIQUES LAURENT SCHWARTZ

ÉCOLE DOCTORALE DE L'ÉCOLE POLYTECHNIQUE

PAR **Victor LAMBERT**

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : MATHEMATIQUES PURES

Quelques problèmes additifs : bases, pseudo-puissances et ensembles k-libres

---

Soutenue le 18 Juin 2015 après avis de :

L. HABSIEGER	Directeur de Recherche (CNRS, Université de Montréal)
F. HENNECART	Professeur (Université Jean Monnet, Saint-Étienne)

Devant la commission d'examen formée de :

E. BALANDRAUD	Maître de Conférences (Université P. et M. Curie, Paris)	**Examinateur**
J-M. DESHOUILLERS	Professeur émérite (ENSC, Bordeaux INP)	**Président du Jury**
L. HABSIEGER	Directeur de Recherche (CNRS, Université de Montréal)	**Rapporteur**
F. HENNECART	Professeur (Université Jean Monnet, Saint-Étienne)	**Rapporteur**
A. PLAGNE	Chercheur (École polytechnique)	**Directeur de thèse**
D. RENARD	Professeur associé (École polytechnique)	**Examinateur**
A. DE ROTON	Maître de Conférences (Université de Lorraine, Nancy)	**Examinatrice**

Remerciements

Je tiens tout d'abord à remercier mon directeur de thèse, Alain Plagne, pour m'avoir encadré et dirigé dans mes travaux de manière efficace, tout en me laissant une grande liberté. Au-delà de l'aspect mathématique, je lui suis extrêmement reconnaissant d'avoir compris mes ambitions professionnelles et de m'avoir toujours soutenu dans cette optique. Cela a été un réel plaisir de travailler et échanger avec lui durant ces trois années.

Je remercie aussi Laurent Habsieger et François Hennecart pour avoir accepté de rapporter ce manuscrit, et pour l'attention qu'ils ont portée à sa lecture. Merci également à Éric Balandraud, Anne de Roton, Jean-Marc Deshouillers et David Renard pour avoir accepté de faire partie de ce jury de thèse.

Un grand merci à mes collaborateurs mathématiques. Merci à Javier Cilleruelo de m'avoir fait confiance pour participer au premier projet d'article qui s'est concrétisé pour moi, mais aussi à Jean-Marc, Alain et Thai Hoang Le.

Et enfin, pour la partie mathématique, je tiens à remercier chaleureusement Jérémy le Borgne. Toujours présent pour répondre à mes questions, il a été un remarquable ami, d'une grande aide tout au long de ma thèse. Je lui dois énormément dans la finalisation de l'article correspondant au troisième chapitre de ce manuscrit. J'espère que nous parviendrons à élucider les questions mathématiques que nous avons laissées en suspens avec Éric.

Merci à Marine Amier et Pascale Fuseau, toujours sympathiques et d'une efficacité remarquable, qui m'ont prouvé que les taches administratives pouvaient finalement bien se passer. Merci aux différents doctorants (Thomas, Tatiana, Yakine, Jacek, Ivan, Rita, ...) d'avoir rendu mes passages au bureau des plus agréables. Merci à mon tuteur Thierry Paul de m'avoir parfaitement tuté.

Évidemment, j'ai adoré enseigner ici à l'X, et ce n'est d'ailleurs pas fini. Merci à Linda et Karine pour avoir rendu cette tâche aussi plaisante. Et surtout, un grand Merci à Pascale Harinck.

Je souhaite remercier ici ceux qui m'ont initié aux mathématiques, Denis Choimet et Bernard Luron, mes deux professeurs de CPGE. C'est certainement grâce à eux que j'ai trouvé ma vocation.

Merci à mes différents relecteurs de manuscrit ou d'articles en anglais, qui ont

essayé tant bien que mal de trouver toutes les erreurs : Chachou, Colette, Bibi, Pierrick, ainsi que mes parents. Merci également à Popo de m'avoir fourni un ordinateur en mode sauvetage de dernière minute.

À mes parents, je vous suis éternellement reconnaissant de m'avoir toujours soutenu et de m'avoir donné le goût des maths depuis mon plus jeune âge, allant jusqu'à me faire croire que j'étais l'inventeur de la racine cubique (cela n'a jamais été publié).

Je tiens finalement à remercier l'ensemble de mes amis, sans qui ma vie durant ces trois années n'aurait pas été la même. Merci aux Nantais : Hugo, Stan, Bibi, Bedam, la Mitch, Marjo, MH, Dalya, Camille, Gilles et le nouveau venu Armand. Merci à Fannie d'être toujours autant présente pour moi. Merci au club des 140 carreaux : Adrien, Julien, Jozy. Et merci à Renan de se dévouer toujours aussi gentiment. Merci à Laulau d'avoir si bien colloqué. Merci à Laurent d'être toujours marrant, de vouloir me faire voler et de laisser Drogba marquer. Merci à Simon de si bien liker. Et plus généralement, merci à Alain, Alex, Alice, Anne V, Anne P, Antoine, Cécile, Chachou, Charlotte, Clémence, Coko, Coline, Elio, Fanny, Florent, François G, François E, au mec de Manon, à Guillaume, Huitrounette, Jéjé, Judith, Laurent, Ludo, Maël, Manon, Manu, Manue, Marie, Marty, Morgane, Nathanaël, Nico, Pierre, Oli35 et Aenarion JD, Raf, Ronan, Tibo G, Tibo V et Xavier.

Et enfin, les mots me manquent pour dire à quel point cette thèse doit à Pauline. Tout a été plus facile avec toi à mes côtés et je ne te remercierai jamais assez d'illuminer ma vie.

Quelques problèmes additifs : bases, pseudo-puissances et ensembles k-libres

Victor LAMBERT
Thèse sous la direction d'Alain PLAGNE

18 Juin 2015

Table des matières

Introduction

Cette thèse est consacrée à l'étude de problèmes de théorie additive des nombres. Elle est composée de trois chapitres disjoints, correspondant à des articles coécrits (ou écrit pour le dernier) durant ces trois années de doctorat. Les deux premiers ont pour sujet commun les *bases additives*, auxquelles nous allons nous intéresser maintenant.

Étant donnés un ensemble $\mathcal{A}$ muni d'une structure additive, et A un sous-ensemble de $\mathcal{A}$, on note

$$hA = \{a_1 + \cdots + a_h, a_i \in A,\ 1 \leqslant i \leqslant h\},$$

à ne pas confondre avec $h \cdot A = \{ha, a \in A\}$. De plus, on utilise la notation $A \sim B$ si A et B sont deux ensembles qui ne diffèrent que d'un nombre fini d'éléments. On dit que A est une base additive de $\mathcal{A}$ s'il existe un entier naturel h tel que $hA \sim \mathcal{A}$. On peut également trouver dans la littérature l'appellation *base asymptotique* pour désigner le même objet. Dans le cas où un tel entier h existe, on peut définir l'ordre de A en tant que base additive

$$\mathrm{ord}^*(A) = \min\{h \text{ tel que } hA \sim \mathcal{A}\}.$$

On peut affaiblir la condition sur A en demandant que tout élément de $\mathcal{A}$, sauf un nombre fini, s'écrive comme somme d'*au plus* h éléments de A, c'est à dire

$$A \cup 2A \cup \cdots \cup hA \sim \mathcal{A}.$$

On dit dans ce cas que A est une *base faible*, et on peut définir à nouveau l'ordre faible de A que l'on note $\mathrm{ord}(A)$. Ces deux définitions coïncident lorsque 0 appartient à A. Mais dans le cas contraire, est-ce qu'une base faible est nécessairement une base ? Et dans ce cas, que dire de $\mathrm{ord}^*(A)$ par rapport à $\mathrm{ord}(A)$? Ces questions seront centrales dans le Chapitre 1.

Ces problèmes ont été largement étudiés dans le cadre des entiers naturels. Nous essaierons d'apporter un maximum de réponses lorsque $\mathcal{A}$ est un groupe abélien infini quelconque G. Nous nous intéresserons aux trois fonctions E, X et S, définies ultérieurement sur $\mathbb{N}$ puis étendues à G, qui contiennent de nombreuses propriétés des bases additives.

Dans le second chapitre, on s'intéressera aux pseudo puissances s-ièmes ($s \geq 2$), une version probabiliste des puissances s-ièmes d'entiers. La question de savoir si l'ensemble des puissances s-ièmes forme une base additive de $\mathbb{N}$ et la détermination de son ordre portent le nom de problème de Waring. On crée une suite aléatoire d'entiers de même répartition (presque sûrement) que les puissances s-ièmes. Des travaux précédents montrent qu'une telle suite A est presque sûrement une base d'ordre $s+1$, et on a donc en particulier $sA \not\sim \mathbb{N}$. Lorsqu'on observe la preuve de ce résultat, on comprend d'une certaine manière que pour passer de sA à $\mathbb{N}$ privé d'un nombre fini d'éléments, ajouter un $s+1$-ième élément de A est trop fort. C'est pourquoi on étudie dans ce Chapitre 2 les *compléments additifs* de A, c'est-à-dire les ensembles B vérifiant

$$sA + B \sim \mathbb{N},$$

presque sûrement. L'objectif est de déterminer à quel point un complément additif peut-être "petit".

Le troisième chapitre traite, pour sa part, des ensembles k-libres, à savoir les ensembles A tels que

$$A \cap k \cdot A = \emptyset.$$

Il est assez naturel de se demander quelle est la taille maximale d'un tel ensemble. La réponse est parfaitement connue pour $\mathbb{N}$, et nous en donnerons d'ailleurs une nouvelle preuve. L'intérêt est que nous pourrons adapter celle-ci pour déterminer la taille minimale d'un ensemble k-libre qui est maximum au sens de l'inclusion dans $\mathbb{N}$. Nous étudierons ensuite le problème initial dans $\mathbb{Z}/n\mathbb{Z}$, pour lequel la contrainte modulaire joue un rôle prépondérant. Les méthodes employées seront très différentes, selon la relation arithmétique entre k et n. En particulier, nous démontrerons un résultat sur des arbres combinatoires, dans l'étude du cas général.

Bien que nécessairement inspiré des articles, chaque chapitre revisite assez largement leurs contenus. Certaines preuves sont plus détaillées et nous donnons plus d'éléments en introduction de chaque partie afin de faciliter la lecture du manuscrit.

Sans mention contraire, les résultats énoncés dans cette thèse sont dus aux auteurs des articles correspondant à chacun des chapitres.

Chapitre 1

Sur les bases additives dans les groupes abéliens

Ce chapitre reprend, à peu de choses près et en français, le texte d'un article écrit en collaboration avec Thái Hoàng Lê et Alain Plagne.

1.1 Introduction

Dans toute la suite, $(G,+)$ est un semi-groupe commutatif. De plus, on rappelle que si A est un sous-ensemble de G et h un entier, hA est l'ensemble formé des sommes de h éléments de A (non nécessairement distincts). Pour deux sous-ensembles A et B de G, on note $A \sim B$ si leur différence symétrique est finie.

On va s'intéresser à la notion de *base additive*. Bien qu'on en ait déjà parlé dans l'introduction générale du manuscrit, nous avons besoin ici de définir différents types de bases additives.

Définition 1.1. On dit que l'ensemble $A \subset G$ est
- *une base (asymptotique) faible* d'ordre au plus h si $A \cup \cdots \cup hA \sim G$.
- *une base faible parfaite* d'ordre au plus h si $A \cup \cdots \cup hA = G$.
- *une base (asymptotique)* d'ordre au plus h si $hA \sim G$.
- *une base parfaite* d'ordre au plus h si $hA = G$.

Si h est le plus petit entier tel que $hA \sim G$, on dit que A est une base d'ordre h et on note $\mathrm{ord}^*_G(A) = h$. Si un tel h n'existe pas, on pose $\mathrm{ord}^*_G(A) = \infty$. On définit évidemment l'ordre de la même façon pour les autres notions, et pour une base faible d'ordre h, on note $\mathrm{ord}_G(A) = h$.

On peut facilement relier ces notions grâce à la remarque suivante. Dans le cas où G contient 0, A est une base faible si et seulement si $A \cup \{0\}$ est une base, et on a $\mathrm{ord}_G(A) = \mathrm{ord}^*_G(A \cup \{0\})$. Bien entendu, les bases (faibles ou non) ne présentent

un intérêt que lorsque G est infini. En revanche, les bases parfaites (faibles ou non) ont un sens dans un cadre fini également.

Historiquement, les bases additives n'ont été étudiées que dans les cas $G = \mathbb{N}$ et $G = \mathbb{Z}$ (cf. [11] pour les entiers relatifs). En particulier, la question de savoir comment se comportait une base lorsqu'on lui enlevait un élément a donné lieu à de nombreux résultats, sur les fonctions E, X et S que nous introduirons dans les trois sous-sections qui suivent. Nous y donnerons certains résultats connus dans $\mathbb{N}$, et expliquerons comment ces fonctions se comportent dans un groupe G infini. Pour un panorama plus exhaustif de ce thème de recherche, dans le cas des entiers naturels, on conseille la lecture de [13] ou [5].

Mais avant de chercher des propriétés sur les bases additives dans un groupe quelconque, il est naturel de se poser la question d'existence de bases d'ordre h, pour $h \geq 1$. Le cas $h = 1$ est trivial. En effet, il suffit de considérer G tout entier ou G privé d'un nombre fini d'éléments selon ce qu'on veut obtenir. Dans le Théorème 1.1 ci-dessous, on montre un résultat plus fort que la simple existence d'une base d'ordre h. Il existe en fait une *base minimale* A d'ordre h, c'est-à-dire que pour tout $a \in A$, $A \setminus \{a\}$ n'est plus une base d'ordre h (cela peut être une base d'ordre plus élevé). Dit autrement, chaque élément de A est nécessaire à ce que A soit une base d'ordre h.

Théorème 1.1. *soient G un groupe abélien infini et h un entier, $h \geq 2$. Alors G admet une base parfaite minimale d'ordre h.*

La démonstration de ce théorème sera l'objet de la section 1.3.

1.1.1 Élements exceptionnels

La première question qu'on se pose est de savoir si lorsqu'on enlève un élément a de A une base additive, $A \setminus \{a\}$ est toujours une base. Considérons un premier exemple simple pour se familiariser avec ces problèmes. On considère dans $\mathbb{N}$ la base additive suivante, d'ordre 2 (celle-ci est d'ailleurs parfaite) :

$$A = \{1\} \cup 2\mathbb{N}.$$

$A \setminus \{1\}$ n'est plus une base, puisque cet ensemble n'engendre plus que les nombres pairs. Si on enlève un autre élément a, $A \setminus \{a\}$ reste une base ; elle n'est plus parfaite d'ordre 2, mais ce n'est pas ce qui nous intéresse ici.

Définition 1.2. Soit $a \in A$, où A est une base de G. On dit que a est un élément *exceptionnel* si $A \setminus \{a\}$ n'est plus une base de G. Dans le cas contraire, a est un élément *régulier*. On note A^* l'ensemble des éléments réguliers de A.

Dans l'exemple précédent, A admet un seul élément exceptionnel. Grekos [6] a montré que le nombre d'éléments exceptionnels de A peut être majoré par $h-1$, ce qui donne un sens à la définition suivante.

$$E(h) = \max_{hA\sim\mathbb{N}} |A \setminus A^*|. \tag{1.1}$$

Plagne, dans [15], a obtenu l'équivalent suivant, lorsque $h \to \infty$:

$$E(h) \sim 2\sqrt{\frac{h}{\log h}}. \tag{1.2}$$

Pour un groupe G quelconque, il n'est a priori pas clair que la fonction E_G est bien définie, c'est-à-dire que le nombre d'éléments exceptionnels peut être majoré en fonction de h seulement. Le théorème suivant répond affirmativement à ce problème, et montre de plus que la borne obtenue est la meilleure possible.

Théorème 1.2. *(i) Pour tout G groupe abélien infini et pour tout $h \geq 1$,*

$$E_G(h) \leq h-1,$$

(ii) il existe un groupe G infini qui vérifie $E_G(h) = h-1$ pour tout $h \geq 1$,

(iii) pour tout $h \geq 1$, il existe un groupe G infini (qui dépend de h) pour lequel $E_G(h) = 0$.

Remarque. En fait, $E_G(1) = 0$ pour tout G, donc seul le cas $h \geq 2$ nous intéressera réellement dans les preuves.

Comme on va le voir plus tard, les points (ii) et (iii) viendront de l'étude du cas $G = \mathbb{F}_p[t]$ l'anneau des polynômes sur le corps $\mathbb{F}_p$, pour lequel on a

$$E_{\mathbb{F}_p[t]}(h) = \left[\frac{h-1}{p-1}\right].$$

Le Théorème 1.2 sera démontré dans la section 1.4.

1.1.2 La fonction d'Erdős et Graham

On se demande maintenant, lorsque a est un élément régulier de A base d'ordre h, comment se comporte l'ordre de $A \setminus \{a\}$. Erdős et Graham [3] ont étudié cette question dans $\mathbb{N}$ et montré qu'on peut majorer l'ordre de $A \setminus \{a\}$ en fonction de h seulement, ce qui donne un sens à la fonction

$$X(h) = \max_{hA\sim\mathbb{N}} \max_{a\in A^*} \mathrm{ord}^*(A \setminus \{a\}). \tag{1.3}$$

À l'heure actuelle, les meilleures minoration et majoration de $X(h)$ sont dues à Plagne [14], qui a amélioré celles obtenues par Stöhr [16], Grekos [6] et Nash [12] notamment. On a

$$\left\lceil \frac{h(h+4)}{3} \right\rceil \leq X(h) \leq \frac{h(h+1)}{2} + \left\lceil \frac{h-1}{3} \right\rceil. \tag{1.4}$$

Erdős et Graham [4] ont conjecturé qu'il existe une constante α telle que $X(h) \sim \alpha h^2$ quand $h \to \infty$, mais ce problème est toujours ouvert. Les inégalités (1.4) conduisent à $X(1) = 1, X(2) = 4, X(3) = 7$, mais la valeur de $X(4)$ reste inconnue.

On s'intéresse ici à ce problème dans le cas d'un groupe G quelconque, et on définit de même

$$X_G(h) = \max_{hA \sim G} \max_{a \in A^*} \mathrm{ord}_G^*(A \setminus \{a\}). \tag{1.5}$$

Dans [3], Erdős et Graham utilisent une version différente de la fonction X, à savoir

$$\begin{aligned} x_G(h) &= \max\{\mathrm{ord}_G^*(A) : \cup_{i=1}^h iA \sim G \text{ et } \mathrm{ord}_G^*(A) < \infty\} \\ &= \max\{\mathrm{ord}_G^*(A) : \cup_{i=1}^h iA \sim G \text{ et } \langle A - A \rangle = G\} \end{aligned} \tag{1.6}$$

avec $G = \mathbb{N}$ dans leur cas, et où $\langle B \rangle$ désigne le sous-groupe engendré par un ensemble B dans G. Dans la mesure où il nous sera préférable de travailler avec l'une ou l'autre de ces deux fonctions selon les cas, voyons dès à présent pourquoi elles sont égales (lorsqu'elles sont bien définies).

Lemme 1.1. *Pour tout groupe infini G, $X_G = x_G$.*

Démonstration. Soient $h \geq 1$ et A une base d'ordre au plus h de G. Soit $a \in A$ un élément régulier de A, alors $B := A - a$ est également une base d'ordre au plus h et contient 0. Ainsi, $B \setminus \{0\}$ est une base faible. De plus,

$$\mathrm{ord}_G^*(B \setminus \{0\}) = \mathrm{ord}_G^*(A \setminus \{a\}),$$

ce qui implique $X_G(h) \leq x_G(h)$.

Pour l'autre sens, d'après les définitions de X_G et x_G, on a clairement $h \leq X_G(h)$ et $h \leq x_G(h)$. Si $x_G(h) = h$, alors

$$X_G(h) = h = x_G(h),$$

puisque $X_G(h) \leq x_G(h)$. Ainsi, on peut supposer $x_G(h) > h$ (remarquons que c'est en fait toujours le cas, dès que $h \geq 2$ d'après le Théorème 1.1). Soit B une base faible d'ordre au plus h de G satisfaisant

$$h < \mathrm{ord}_G^*(B) < \infty.$$

Alors $0 \notin B$ (sinon, $\mathrm{ord}^*_G(B) = \mathrm{ord}_G(B) \leq h$). Posons $A = B \cup \{0\}$. A est donc une base d'ordre au plus h et 0 est un élément régulier de A car $A \setminus \{0\} = B$. En outre,

$$\mathrm{ord}^*_G(A \setminus \{0\}) = \mathrm{ord}^*_G(B),$$

ce qui donne bien $x_G(h) \leq X_G(h)$. □

L'étude de X_G est nettement moins concluante que celle de E_G. En effet, on ne sait même pas si $X_G(h)$ est fini pour tout G. Cependant, nous sommes capables de répondre à ce problème et donner des bornes pour $X_G(h)$ pour une large catégorie de groupes.

Avant d'énoncer nos résultats, nous avons besoin de rappeler la définition de la fonction arithmétique Ω :

$$\Omega(n) = \alpha_1 + \cdots + \alpha_k, \tag{1.7}$$

si $n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ est la factorisation en produit de nombres premiers distincts de n.

Rappelons de plus que nous désignons par $m \cdot A$ le sous-ensemble de G

$$m \cdot A = \{ma : a \in A\}.$$

En adaptant l'idée d'Erdős et Graham, on obtient le théorème suivant.

Théorème 1.3. *Soit G un groupe abélien infini tel que pour tout entier $1 \leq m \leq h$, le quotient $G/m \cdot G$ est fini, alors*

$$X_G(h) \leq h^2 + h \cdot \max_{1 \leq m \leq h} \Omega\left(|G/m \cdot G|\right) + h - 1. \tag{1.8}$$

Parmi les groupes qui satisfont l'hypothèse du Théorème 1.3, on retrouve notamment les groupes abéliens de type fini, les groupes divisibles (i.e les groupes G tels que $m \cdot G = G$ pour tout $m \in \mathbb{Z}^+$, ce qui inclut $\mathbb{R}$ et $\mathbb{Q}$) et $\mathbb{Z}_p$ (les entiers p-adiques).

Pour ce qui est de la minoration, on démontre que l'inégalité (1.4) se généralise aux groupes admettant $\mathbb{Z}$ comme quotient.

Théorème 1.4. *Soit G un groupe abélien infini. Supposons qu'il existe un sous-groupe H de G tel que $G/H \cong \mathbb{Z}$ (isomorphisme algébrique), alors pour tout entier $h \geq 1$, on a*

$$X_G(h) \geq \left\lceil \frac{h(h+4)}{3} \right\rceil.$$

Cela pourrait laisser penser que $X_G(h)$ a une croissance quadratique. Nous montrons que ce n'est en fait pas le cas, en exhibant une classe de groupes pour lesquels $X_G(h)$ a une croissance linéaire.

Théorème 1.5. *Soient p un nombre premier et G un groupe abélien infini qui a la propriété que tout élément non nul de G est d'ordre p.*

(i) Pour tout entier $h \geq p$, on a

$$X_G(h) \leq ph + p - 1.$$

(ii) Pour tout entier $h \geq 3(p-1)/2$, on a

$$X_G(h) \geq 2h - 3p + 3.$$

En particulier, si $p = 2$, $X_G(h) \sim 2h$ quand $h \to \infty$.

Bien que nous ne soyons pas capables de dire en général si $X_G(h)$ est fini, on peut le confirmer et même en donner un encadrement lorsque $h = 2$ ou $h = 3$.

Théorème 1.6. *Pour tout groupe abélien infini G, on a*

(i) $3 \leq X_G(2) \leq 5$.

(ii) $4 \leq X_G(3) \leq 17$.

La section 1.5 sera consacrée aux démonstrations de ces différents résultats.

1.1.3 La fonction de Grekos

Dans les exemples de constructions de bases donnant de bonnes minorations pour $X(h)$, on s'aperçoit que bien souvent, seul un nombre fini d'éléments a de A^* vérifient $\mathrm{ord}^*(A \setminus \{a\}) = X(h)$. C'est l'origine de l'introduction de la fonction S définie par la formule suivante

$$S(h) = \max_{hA \sim \mathbb{N}} \limsup_{a \in A^*} \mathrm{ord}^*(A \setminus \{a\}).$$

Cette fonction, due à Grekos, désigne, en d'autres mots, la valeur minimale de s telle que pour tout A satisfaisant $hA \sim \mathbb{N}$, il n'y a qu'un nombre fini d'éléments $a \in A$ vérifiant

$$\mathrm{ord}^*(A \setminus \{a\}) > s.$$

Il a été conjecturé par Grekos que l'ordre de grandeur de S est plus petit que celui de X, ce qui a été confirmé par Cassaigne et Plagne [3] qui ont prouvé que

$$h + 1 \leq S(h) \leq 2h \tag{1.9}$$

pour tout $h \geq 2$ (évidemment, $S(1) = 1$). On sait aussi que $S(2) = 3$, mais la valeur de $S(3)$ reste inconnue. Déterminer s'il existe une constante β telle que $S(h) \sim \beta h$ quand $h \to \infty$ est un problème ouvert qui paraît extrêmement difficile.

De manière analogue, on définit $S_G(h)$ dans un groupe abélien G infini comme étant la valeur minimale de s telle que pour tout A satisfaisant $hA \sim G$, le nombre de $a \in A$ vérifiant

$$\mathrm{ord}^*(A \setminus \{a\}) > s$$

est fini. À nouveau, il n'est a priori pas clair que S_G est bien définie. Déjà, d'après la définition, on a

$$S_G(h) \leq X_G(h).$$

On montre dans le théorème suivant non seulement que $S_G(h)$ est bien défini pour tout groupe abélien infini, et pour tout $h \geq 1$, mais aussi que les bornes dans (1.9) restent valables. On généralise en fait les arguments de [2]. Nous avons cependant besoin de la notion de *moyennabilité* sur les groupes, ce qui rend la preuve non élémentaire.

Théorème 1.7. *Pour tout groupe abélien infini G, on a $h + 1 \leq S_G(h) \leq 2h$.*

Au contraire du Théorème 1.2, on ne sait pas si ces bornes sont les meilleures possibles. Cependant, on peut démontrer que dans le cas $h = 2$, la borne inférieure est la bonne.

Théorème 1.8. *Pour tout groupe abélien infini G, on a $S_G(2) = 3$.*

Remarque. Le théorème précédent veut exactement dire que si on considère une base A d'ordre 2 de G, il y a un nombre fini d'éléments a qui vérifient $\mathrm{ord}^*_G(A \setminus \{a\}) > 3$. En fait, à travers la preuve, on verra qu'il ne peut exister qu'*un seul* tel élément a.

Nous introduirons la notion de *moyennabilité* et prouverons ces deux résultats dans la section 1.6.

Avant de passer aux preuves des résultats énoncés dans cette introduction, nous allons démontrer quelques résultats préliminaires.

1.2 Résultats préliminaires

1.2.1 Quelques observations

On commence par donner quelques lemmes, dont certains s'apparentent parfois à des remarques tellement ils sont immédiats. On les donnera d'ailleurs quelquefois sans preuve.

Lemme 1.2. *Soient G un groupe abélien infini et $A \subset G$. Si A est une base (respectivement une base parfaite) de G et $b \in G$, alors $A - b = \{a - b : a \in A\}$ est une base (respectivement une base parfaite) de même ordre.*

Démonstration. Cela découle immédiatement du fait que pour tout entier h, $h(A-b) = hA - hb$. □

On a en fait déjà utilisé ce résultat dans la preuve du Lemme 1.1.

Si H est un sous-groupe de G, pour tout $x \in G$, on note $\overline{x}$ la classe de x dans G/H. Les trois prochains lemmes vont nous permettre d'introduire les systèmes de représentants associés à un quotient G/H, utiles à la construction de bases pour G.

Lemme 1.3. *soient G un groupe abélien et H un sous-groupe de G. Considérons $\Lambda \subset G$ un système de représentants de G/H dans G, c'est-à-dire, pour tout $x \in G$, il y a exactement un élément $\lambda \in \Lambda$ tel que $x+H = \lambda+H$. Alors, pour tout $x \in G$, il y a une unique façon d'écrire*

$$x = \lambda + g$$

avec $\lambda \in \Lambda, g \in H$. En particulier, si $H \neq G$ et $H \neq \{0\}$ alors $\Lambda \cup H$ est une base parfaite d'ordre 2 de G.

Nous aurons besoin d'un système de représentants particulier que nous introduisons maintenant.

Lemme 1.4. *Soient G un groupe abélien et H un sous-groupe de G. Alors, il existe un système de représentants Λ de G/H dans G tel que $0 \in \Lambda$ et $\Lambda = -\Lambda$.*

Démonstration. On choisit un représentant dans chaque classe de G suivant H. Bien entendu, on commence par choisir 0 comme représentant de la classe H. Ensuite, si λ est le représentant d'une classe B, alors on choisit $-\lambda$ comme représentant de la classe $-B$. □

On observe ensuite que les bases parfaites d'un quotient peuvent facilement s'étendre au groupe tout entier, ce qui n'est pas le cas pour les bases non parfaites.

Lemme 1.5. *Soient G un groupe abélien infini et H un sous-groupe de G. Soient également $A \subset G/H$,*

$$B = \{x \in G : \overline{x} \in A\}$$

et h un entier supérieur ou égal à 1. On a alors :

(i) $hA = G/H$ si et seulement si $hB = G$,

(ii) $\bigcup_{i=1}^{h} iA = G/H$ si et seulement si $\bigcup_{i=1}^{h} iB = G$.

Le lemme suivant précise que toute base est parfaite, quitte à augmenter l'ordre de 1.

Lemme 1.6. *Soient G un groupe abélien infini et $A \subset G$.*

(i) Si $hA \sim G$, *alors* $(h+1)A = G$,

(ii) Si $\bigcup_{i=1}^{h} iA \sim G$, *alors* $\bigcup_{i=2}^{h+1} iA = G$.

Démonstration. Supposons $hA \sim G$, et soit x un élément quelconque de G. Comme $x - A$ est infini, il a nécessairement une intersection non vide avec hA. Ainsi, $x \in (h+1)A$.

Supposons à présent $\bigcup_{i=1}^{h} iA \sim G$, et soit x quelconque dans G. Comme $x - A$ est infini, il a nécessairement une intersection non vide avec rA pour $1 \leq r \leq h$. Ainsi,

$$x \in (r+1)A \subset \bigcup_{i=2}^{h+1} iA.$$

□

Pour établir des bornes de X_G, on aura besoin du résultat suivant, qui dit que si deux sous-ensembles de la forme nA ont une intersection non vide, alors on peut trouver une suite arbitrairement longue de sous-ensembles de A dont l'intersection est non vide.

Lemme 1.7. *Soient* $A \subset G$ *et* m, n *des entiers naturels non nuls. Si*

$$c \in nA \cap (n+m)A$$

alors pour tout entier naturel k, *on a*

$$kc \in knA \cap (kn+m)A \cap \cdots \cap (kn+km)A.$$

1.2.2 Caractérisation des bases à la façon d'Erdős et Graham

Dans [3], Erdős et Graham ont prouvé qu'une base faible A de $\mathbb{N}$ est une base si et seulement si

$$\text{pgcd}(A - A) = 1 \tag{1.10}$$

où $A - A = \{a_1 - a_2 : a_1, a_2 \in A\}$. Cela donne immédiatement un critère de régularité pour un élément d'une base de $\mathbb{N}$. En effet, si A est une base de $\mathbb{N}$, alors $a \in A$ est régulier si et seulement si

$$\text{pgcd}(A \setminus \{a\} - A \setminus \{a\}) = 1. \tag{1.11}$$

Nous allons à présent généraliser ces caractérisations dans le cas d'un groupe arbitraire. Ce n'est plus le pgcd$(A - A)$ qui intervient désormais (cela n'a pas de sens dans G quelconque), mais $\langle A - A \rangle$, le sous-groupe engendré par $A - A$ dans G.

Lemme 1.8. *Soient G un groupe abélien infini et A une base faible de G. Alors A est une base si et seulement si $\langle A - A\rangle = G$.*

Démonstration. Supposons que A soit une base faible d'ordre au plus h, c'est-à-dire,

$$G \sim \bigcup_{i=1}^{h} iA.$$

Posons $H = \langle A - A\rangle$. L'image de a dans G/H est la même pour tout $a \in A$. Ainsi, pour tout s, l'image de sA dans G/H est un singleton. Cela signifie que A ne peut être une base que si $G = H$.

Réciproquement, supposons $G = H$. On commence par montrer qu'il existe n tel que

$$nA \cap (n+1)A \neq \emptyset.$$

En effet, soit a un élément de A, on peut d'après l'hypothèse écrire a comme une combinaison linéaire d'éléments de $A - A$

$$a = \sum_{k=1}^{t} \alpha_k (a_k - b_k)$$

avec $a_k, b_k \in A$ et $\alpha_k \in \mathbb{N}^*$ pour tout k.

Ainsi, l'élément

$$c = a + \sum_{k=1}^{t} \alpha_k b_k = \sum_{k=1}^{t} \alpha_k a_k$$

est à la fois dans nA et $(n+1)A$, en posant $n = \sum_{k=1}^{t} \alpha_k$. D'après le Lemme 1.7,

$$(h-1)c \in \bigcap_{i=0}^{h-1} ((h-1)n + i)A.$$

Or, pour tout $x \in G$, sauf un nombre fini, on a

$$x - (h-1)c \in \bigcup_{i=1}^{h} iA.$$

Il vient alors que pour tout $x \in G$, sauf un nombre fini,

$$x = x - (h-1)c + (h-1)c \in ((h-1)n + h)A.$$

Ainsi, A est une base d'ordre au plus $(h-1)n + h$. □

On en vient maintenant à la condition de régularité d'un élément a.

Lemme 1.9. *Soient G un groupe abélien infini et A une base de G. Alors $a \in A$ est régulier si et seulement si $\langle A \setminus \{a\} - A \setminus \{a\} \rangle = G$.*

Démonstration. On aimerait appliquer le Lemme 1.8, mais on ne sait pas si $A \setminus \{a\}$ est une base faible. Cependant, remarquons que $B := A - a$ est également une base (Lemme 1.2), qui de plus contient 0. Ainsi, $B \setminus \{0\}$ est une base faible. D'où

$$\begin{aligned} A \setminus \{a\} \text{ est une base} &\iff B \setminus \{0\} \text{ est une base} \\ &\iff \langle B \setminus \{0\} - B \setminus \{0\} \rangle = G \\ &\iff \langle A \setminus \{a\} - A \setminus \{a\} \rangle = G. \end{aligned}$$

□

1.3 Existence de bases additives

Dans le cas de $\mathbb{N}$, il est connu depuis Härtter [9] que $\mathbb{N}$ admet des bases minimales de tout ordre (sa preuve n'est pas constructive). Un exemple concret de base minimale d'ordre h de $\mathbb{N}$ est donné par

$$A = \left\{ \sum_{f \in \mathcal{F}} 2^f : \mathcal{F} \text{ est un ensemble fini de } \mathbb{N} \text{ inclus dans une classe modulo } h \right\}.$$

On pourra lire [10] pour trouver des propriétés supplémentaires de cette base.

Montrons d'abord que dès qu'il existe une représentation spéciale des éléments de G similaire à celle de $\mathbb{N}$ que l'on vient juste de donner, il existe des bases minimales parfaites de tout ordre.

Proposition 1.1. *Soit G un groupe abélien infini. Supposons qu'il existe une suite infinie de sous-ensembles $(\Lambda_i)_{i=0}^{\infty}$ de G satisfaisant les propriétés suivantes :*

(i) $0 \in \Lambda_i$, *pour tout* $i \in \mathbb{N}$,

(ii) $-\Lambda_i = \Lambda_i$, *pour tout* $i \in \mathbb{N}$,

(iii) Tout élément $x \in G$ admet une unique *représentation de la forme*

$$x = \lambda_0(x) + \lambda_1(x) + \cdots$$

où $\lambda_i(x) \in \Lambda_i$ pour tout i, et $\lambda_i(x) \neq 0$ pour un nombre fini d'indices i. En d'autres termes, G est égal à la "somme directe" $\oplus_{i=0}^{\infty} \Lambda_i$. [1]

Alors, pour tout entier $h \geq 2$, G admet une base parfaite minimale d'ordre h.

1. On ne peut en fait pas parler de somme directe ici, les Λ_i étant seulement des ensembles, pas des groupes.

Démonstration. Pour $x \in G$, l'ensemble $\{i \in \mathbb{N} : \lambda_i(x) \neq 0\}$ est appelé le *support* de x.

Clairement, si x et y ont des supports disjoints,

$$\lambda_i(x+y) = \lambda_i(x) + \lambda_i(y).$$

Considérons $\mathbb{N} = N_1 \cup \cdots \cup N_h$ une partition de $\mathbb{N}$ en h ensembles infinis (disjoints). Notons A_j l'ensemble des $x \in G$ à support dans N_j, et posons finalement

$$A = \cup_{j=1}^{h} A_j.$$

Par définition, $0 \in A$. De plus, tout élément $x \in G$ peut s'écrire de manière *unique* comme

$$x = a_1 + \cdots + a_h \tag{1.12}$$

avec $a_j \in A_j$ pour tout $j = 1, \ldots, h$. Lorsque $a_1, \ldots, a_h \neq 0$, x ne peut être écrit comme une somme de moins de h éléments de A. Cela montre que A est une base parfaite d'ordre h. Cependant, A n'est pas minimale. On va en fait montrer que $B := A \setminus \{0\}$ est une base parfaite minimale d'ordre h.

Tout d'abord, prouvons que $hB = G$. Dans l'expression (1.12), certains (peut-être même tous) des a_j peuvent valoir 0. Étant donné $a \in A_j$, on peut l'écrire

$$a = (a + \lambda) + (-\lambda)$$

où λ est n'importe quel élément de $\Lambda_k \setminus \{0\}$ et $k \in N_j$ est un élément hors du support de a. Remarquons que d'après l'hypothèse, $-\lambda \in \Lambda_k$ également. Ainsi, tout élément de A_j, qu'il soit nul ou non, peut s'écrire comme somme de deux éléments non nuls de A_j. On peut donc faire croître le nombre d'éléments non nuls dans (1.12) jusqu'à h, ce qui montre bien $hB = G$.

Il nous reste à voir que B est minimale. Soit donc a un élément de B. Sans perte de généralité, on peut supposer que $a \in A_1 \setminus \{0\}$. Considérons un élément $x \in G$ de la forme

$$x = a + a_2 + \cdots + a_h$$

avec $a_j \in A_j \setminus \{0\}$ pour tout $j = 2, \ldots, h$. Comme les N_j forment une partition de $\mathbb{N}$, il y a bien une unique façon d'écrire x comme somme de h éléments de B, et a apparaît dans cette expression. Ainsi, x ne peut pas s'écrire comme somme de h éléments de $B \setminus \{a\}$. Comme il y a une infinité de tels x, on a $\mathrm{ord}^*(A \setminus \{a\}) \geq h+1$, et B est bien minimale. □

Venons-en à présent à la preuve du Théorème 1.1.

Preuve du Théorème 1.1. Il nous reste à construire une suite $(\Lambda_i)_{i=0}^{\infty}$ d'ensembles satisfaisant les hypothèses de la Proposition 1.1. Pour cela, on distingue deux cas.

Premier cas : G admet un élément d'ordre infini. On peut alors supposer que $\mathbb{Z} < G$. Soit alors Λ_0 un système de représentants $G/\mathbb{Z}$ dans G. Par le Lemme 1.3, tout élément $x \in G$ peut s'écrire d'une unique façon

$$x = n + \lambda_0$$

avec $\lambda_0 \in \Lambda_0$ et $n \in \mathbb{Z}$. De plus, on peut, d'après le Lemme 1.4, choisir Λ_0 tel que $0 \in \Lambda_0$ et $\Lambda_0 = -\Lambda_0$. Remarquons que tout entier relatif n s'écrit d'une *unique* manière comme

$$n = \sum_{i=0}^{k} a_i 3^i$$

où $a_i \in \{0, 1, -1\}$ pour tout i (cette représentation est connue dans la littérature sous le nom de *système ternaire balancé*). Posons $\Lambda_i = \{0, 3^{i-1}, -3^{i-1}\}$ pour $i \geq 1$. Ainsi, tout élément $x \in G$ admet une unique représentation de la forme

$$x = \lambda_0(x) + \lambda_1(x) + \cdots \tag{1.13}$$

avec $\lambda_i(x) \in \Lambda_i$ pour tout i, et $\lambda_i(x) \neq 0$ pour un nombre fini d'indices i.
Deuxième cas : Tout élément de G est d'ordre fini.

Soit $g_1 \in G$. D'après l'hypothèse, $G_1 := \langle g_1 \rangle$ est fini. On considère alors $g_2 \in G \setminus G_1$, et on pose $G_2 := \langle g_1, g_2 \rangle$. On a donc $G_1 \lneq G_2$ et G_2 est fini. On construit ainsi une chaîne infinie strictement croissante de sous-groupes de G

$$G_1 \lneq G_2 \lneq \cdots .$$

Pour tout entier $i \geq 2$, considérons $\Lambda_i \ni 0$ un système de représentants de G_i/G_{i-1} dans G_i vérifiant de plus $\Lambda_i = -\Lambda_i$. D'après le Lemme 1.3, on peut écrire tout $x \in G_i$ de manière unique, sous la forme

$$x = \lambda + g$$

avec $\lambda \in \Lambda_i$ et $g \in G_{i-1}$. Posons de plus $\Lambda_1 = G_1$. Ainsi, tout $x \in \cup_{i=1}^{\infty} G_i$ peut s'écrire d'une unique façon comme

$$x = \lambda_1(x) + \lambda_2(x) + \cdots$$

avec $\lambda_i(x) \in \Lambda_i$ pour tout $i = 1, 2, \ldots$, et seul un nombre fini de $\lambda_i(x)$ est non nul (en effet, si $x \in G_k$, alors $\lambda_i(x) = 0$ pour tout $i \geq k + 1$).

Finalement, définissons $\Lambda_0 \ni 0$ un système de représentants de $G/\cup_{i=1}^{\infty} G_i$ dans G, tel que $\Lambda_0 = -\Lambda_0$. Alors, tout x de G admet une unique décomposition de la forme

$$x = \lambda_0(x) + \lambda_1(x) + \lambda_2(x) + \cdots$$

où $\lambda_i(x) \in \Lambda_i$ pour tout $i = 0, 1, 2, \ldots$, et où il n'y a qu'un nombre fini de $\lambda_i(x)$ non nuls. Puisque $\Lambda_1 = G_1$, on a bien $\Lambda_i = -\Lambda_i$ pour tout i. Le théorème est alors démontré. □

1.4 La fonction E

Dans cette section, on s'intéresse aux nombres d'éléments exceptionnels d'une base additive dans un groupe abélien G. Rappelons simplement la définition de la fonction E_G :

$$E_G(h) = \max_{hA \sim G} |A \setminus A^*|$$

où A^* est l'ensemble des éléments réguliers de A.

Preuve du Théorème 1.2 (i). On doit donc montrer que si $hA \sim G$, alors A ne peut pas avoir plus de $h-1$ éléments exceptionnels.

D'après le Lemme 1.9, si $a \in A$ est un élément exceptionnel, alors $\langle A - A \rangle = G$ tandis que $\langle A \setminus \{a\} - A \setminus \{a\} \rangle \neq G$. S'il existe $a' \in A \setminus \{a\}$ tel que

$$a - a' \in \langle A \setminus \{a\} - A \setminus \{a\} \rangle,$$

alors pour tout $a'' \in A \setminus \{a\}$,

$$a - a'' = a - a' + a' - a'' \in \langle A \setminus \{a\} - A \setminus \{a\} \rangle,$$

et finalement

$$\langle A \setminus \{a\} - A \setminus \{a\} \rangle \supset \langle A - A \rangle = G.$$

Ainsi, $a - a'$ n'est pas dans $\langle A \setminus \{a\} - A \setminus \{a\} \rangle$ pour tout $a' \in A \setminus \{a\}$.

Supposons, par l'absurde, qu'il existe au moins h éléments exceptionnels $a_1, \ldots, a_h$ dans A. G étant infini, A l'est également. Considérons a_0 un élément de $A \setminus \{a_1, \ldots, a_h\}$. Comme $hA \sim G$, il existe $a \in A \setminus \{a_0, a_1, \ldots, a_h\}$ tel que

$$a_0 + a_1 + a_2 + \cdots + a_h - a$$

soit égal à la somme $b_1 + \cdots + b_h$ de h éléments de A. En conséquence,

$$\sum_{i=0}^{h} (a_i - a) = \sum_{i=1}^{h} (b_i - a).$$

Certains b_i peuvent être égaux à certains a_i. On est face à deux possibilités :

Premier cas : $\{a_1, a_2, \ldots, a_h\} \neq \{b_1, b_2, \ldots, b_h\}$. Cela signifie qu'une fois qu'on a enlevé de chaque côté les termes identiques, il reste du côté gauche de l'égalité au moins un a_i avec $i \neq 0$. Mais cela implique que $a_i - a \in \langle A \setminus \{a_i\} - A \setminus \{a_i\} \rangle$, ce qui fournit une contradiction.
Deuxième cas : $\{a_1, a_2, \ldots, a_h\} = \{b_1, b_2, \ldots, b_h\}$. Cela nous conduit à $a_0 = a$, à nouveau une contradiction. □

Remarque. Dans $\mathbb{N}$, le fait que toute base admet un nombre fini d'éléments exceptionnels provient immédiatement du critère d'Erdős et Graham (1.10) (cf. [13, Théorème 1]). Cependant, nous n'avons pas pu généraliser cette preuve à un groupe général, dans la mesure où elle utilisait une propriété bien spécifique à $\mathbb{Z}$, à savoir que toute suite strictement croissante de sous-groupes est finie.

Les assertions du Théorème 1.2 (ii) et (iii) sont des conséquences immédiates de la proposition suivante.

Proposition 1.2. *Soit $G = \mathbb{F}_p[t]$ l'anneau des polynômes sur le corps $\mathbb{F}_p$. Pour tout entier $h \geq 2$, on a*

$$E_G(h) = \left[\frac{h-1}{p-1}\right].$$

En particulier, si $p = 2$, $E_G(h) = h-1$ pour tout $h \geq 2$. Et si $p > h$, $E_G(h) = 0$. On ne peut donc pas obtenir une minoration universelle non triviale de E_G.

Démonstration. Commençons par montrer que

$$E_G(h) \leq \left[\frac{h-1}{p-1}\right].$$

On va avoir pour cela recours à des arguments similaires à ceux de la preuve du Théorème 1.2 (i).

Supposons que $A \subset \mathbb{F}_p[t]$ est une base d'ordre h, et que $a_1, \ldots, a_k$ sont tous les éléments exceptionnels de A. Supposons de plus, par l'absurde, que $k(p-1) \geq h$. Alors, il existe des entiers $0 \leq \alpha_1, \ldots, \alpha_k \leq p-1$ tels que

$$\alpha_1 + \cdots + \alpha_k = h.$$

Considérons a_0 un élément de $A \setminus \{a_1, \ldots, a_k\}$. Étant donné que $hA \sim G$ et que A est infini, il existe $a \in A \setminus \{a_0, a_1, \ldots, a_k\}$ tel que

$$\sum_{i=1}^{k} \alpha_i a_i + a_0 - a$$

peut s'écrire comme la somme $\sum_{j=1}^{h} b_i$ de h éléments de A. Ainsi,

$$\sum_{i=1}^{k} \alpha_i (a_i - a) + (a_0 - a) = \sum_{j=1}^{h} (b_i - a).$$

Comme $a_0 - a \neq 0$, les multi-ensembles $\{a_1(\alpha_1 \text{ fois}), \ldots, a_k(\alpha_k \text{ fois})\}$ et $\{b_1, \ldots, b_h\}$ sont distincts. Une nouvelle fois, lorsqu'on simplifie des deux côtés de l'égalité par les termes identiques, on constate qu'il existe $1 \leq i \leq k$, et $0 < \beta \leq \alpha_i$ tels que

$\beta(a_i - a)$ appartient à $\langle A \setminus \{a_i\} - A \setminus \{a_i\}\rangle$. $\mathbb{F}_p$ étant un corps, cela implique que $a_i - a$ appartient à ce sous-espace, ce qui contredit le Lemme 1.9 car a_i est exceptionnel.

Finalement, $h - 1 \geq (p-1)k$, d'où $k \leq [(h-1)/(p-1)]$.

Pour montrer qu'il y a bien égalité, on va s'intéresser à un exemple assez simple. Posons

$$k = \left[\frac{h-1}{p-1}\right].$$

En écrivant la division euclidienne de h par $p-1$, il vient $h = k(p-1) + r + 1$ où $0 \leq r < p-1$.

Définissons alors

$$A = \left\{1, t, \cdots, t^{k-1}\right\} \bigcup t^k \cdot \mathbb{F}_p \bigcup \cdots \bigcup t^{k+r-1} \cdot \mathbb{F}_p \bigcup t^{k+r} \cdot \mathbb{F}_p[t].$$

(Les ensembles $t^k \cdot \mathbb{F}_p, \ldots, t^{k+r-1} \cdot \mathbb{F}_p$ n'apparaissent pas si $r = 0$.)

Déjà, A est une base d'ordre $k(p-1) + r + 1 = h$. En effet, il est facile de voir que tout élément de $\mathbb{F}_p[t]$ peut s'écrire comme somme de $k(p-1) + r + 1$ éléments de A (remarquons que $0 \in A$). De plus, pour tout $P(t) \in \mathbb{F}[t] \setminus \{0\}$, l'élément

$$\sum_{i=0}^{k-1}(p-1)t^i + \sum_{i=k}^{k+r-1} t^i + P(t)t^{k+r}$$

ne peut pas s'exprimer comme somme d'au plus $h-1$ éléments de A.

Grâce au Lemme 1.9, on voit aisément que les éléments exceptionnels de A sont exactement

$$\left\{t^i : i = 0, \ldots, k-1\right\},$$

ce qui donne bien k éléments exceptionnels. □

1.5 La fonction X

Dans cette section, on étudie des bornes pour la fonction X_G. Rappelons qu'on pourra utiliser selon le contexte les deux définitions de la fonction, à savoir (1.5) et (1.6), qui sont équivalentes, d'après le Lemme 1.1.

1.5.1 Bornes générales

Afin de prouver le Théorème 1.3, nous aurons besoin du lemme suivant, faisant intervenir la fonction Ω définie en (1.7).

Remarque. On appelle *longueur* d'un groupe la taille $l(G)$ de sa plus longue chaîne de sous-groupes (i.e suite strictement croissante de sous-groupes). On obtient en fait ce maximum lorsque les quotients successifs G_{i+1}/G_i sont des groupes simples. Dans le cas de $\mathbb{Z}/n\mathbb{Z}$, si $n = \prod p_i^{\alpha_i}$, il est clair que $l(G) = \sum \alpha_i$. De plus, si G et G' sont de longueurs finies, $l(G \times G') = l(G) + l(G')$. Ainsi, pour un groupe abélien fini

$$G = \mathbb{Z}/n_1\mathbb{Z} \times \mathbb{Z}/n_2\mathbb{Z} \times \cdots \times \mathbb{Z}/n_k\mathbb{Z}$$

de cardinal $n = \prod n_j = \prod p_i^{\beta_i}$, on a $l(G) = \sum \beta_i = \Omega(|G|)$. C'est dans ce rôle que Ω va intervenir dans la suite.

Lemme 1.10. *Soient G un groupe abélien fini qui est de plus m-torsion (c'est-à-dire, $mx = 0$ pour tout $x \in G$) et $A \subset G$ tel que $\langle A - A \rangle = G$. Alors, pour tout entier $s \geq \Omega(|G|)$, on a $smA = G$.*

Démonstration. Puisque $\langle A - A \rangle = G$, on peut choisir des éléments $a_1, a_2, \ldots$ dans A de telle façon que pour tout entier k, si $\langle A_k - A_k \rangle \neq G$, alors $\langle A_k - A_k \rangle \lneq \langle A_{k+1} - A_{k+1} \rangle$, où

$$A_k = \{a_1, \ldots, a_k\}.$$

D'après la remarque précédant l'énoncé du lemme, il existe $t \leq \Omega(|G|)$ tel que $\langle A_t - A_t \rangle = G$. Ainsi, tout élément $x \in G$ admet une représentation de la forme

$$x = \sum_{i,j=1}^{t} \alpha_{i,j}(a_i - a_j)$$

où $\alpha_{i,j} \in \mathbb{Z}$. Comme G est m-torsion, en réarrangeant les termes du membre de droite, on peut écrire

$$x = \sum_{i}^{t} \beta_i a_i$$

avec $0 \leq \beta_i < m$ pour tout $i = 1, \ldots, t$. On sait en plus que $\sum_{i=1}^{t} \beta_i$ est un multiple de m (car $\sum_{i=1}^{t} \beta_i \equiv \sum_{i,j=1}^{t} (\alpha_{i,j} - \alpha_{i,j}) \equiv 0 \pmod{m}$). Comme $0 \in mA$, on peut ajouter autant de zéros que l'on souhaite, ce qui assure que $x \in tmA \subset smA$, ce qu'on voulait obtenir. □

Preuve du Théorème 1.3. On utilise ici la définition (1.6). Soit A une base faible d'ordre au plus h de G satisfaisant $\langle A - A \rangle = G$. Posons

$$s = \max_{1 \leq m \leq h} \Omega\left(|G/m \cdot G|\right).$$

Comme $(h+1)A \subset G \sim \bigcup_{i=1}^{h} iA$, il existe un entier n tel que $1 \leq n \leq h$ et $nA \cap (h+1)A \neq \emptyset$.

Posons $m = h + 1 - n$ et considérons $c \in nA \cap (n+m)A$. D'après le Lemme 1.7, on a

$$(h-1)c \in \bigcap_{i=0}^{h-1} ((h-1)n + im)A.$$

Or, $G \setminus \left(\bigcup_{i=1}^{h} iA\right)$ est fini, d'où

$$m \cdot G \setminus \left(\bigcup_{i=1}^{h} miA\right)$$

est également fini. On en déduit alors que

$$(h-1)c + m \cdot G \setminus ((h-1)n + hm)A \tag{1.14}$$

est fini.

Par ailleurs, le groupe $G/m \cdot G$ est fini par hypothèse et clairement m-torsion. On a également $\langle \overline{A} - \overline{A} \rangle = G/m \cdot G$, où $\overline{A}$ désigne l'image de A par la projection $G \to G/m \cdot G$. D'après le Lemme 1.10, on sait que

$$sm\overline{A} = G/m \cdot G.$$

Autrement dit, il existe un système de représentants $\{x_1, \ldots, x_k\}$ de $G/m \cdot G$ dans G tel que

$$x_j \in smA \tag{1.15}$$

pour tout $j = 1, \ldots, k$.

Mais pour tout $x \in G$, il existe $1 \leq j \leq k$ tel que $x - (h-1)c - x_j \in m \cdot G$, ce qui conduit d'après (1.14) à ce que pour tout $x \in G$, sauf un nombre fini, on a

$$x - x_j \in ((h-1)n + hm)A. \tag{1.16}$$

En écrivant maintenant

$$x = x - x_j + x_j$$

et en utilisant (1.15) et (1.16), il vient

$$G \sim (sm + (h-1)n + hm)A.$$

Ainsi, A est une base d'ordre au plus

$$sm + (h-1)n + hm = sm + (h-1)(m+n) + m \leq h^2 + sh + h - 1,$$

car $m + n = h + 1$. □

Remarque. L'hypothèse du Théorème 1.3 est notamment satisfaite si $G/m \cdot G$ est fini pour tout m. Les groupes divisibles (i.e. tels que $m \cdot G = G$, pour tout $m \geq 1$), parmi lesquels $\mathbb{R}$ et $\mathbb{Q}$, vérifient évidemment cette propriété. On peut aisément voir que les groupes abéliens de type fini satisfont également cette propriété. C'est aussi le cas du groupe $\mathbb{Z}_p$ des entiers p-adiques, puisque $\mathbb{Z}_p/m \cdot \mathbb{Z}_p \cong \mathbb{Z}/p^l\mathbb{Z}$, où p^l est la plus grande puissance de p qui divise m. Les groupes *juste infinis* (tous leurs quotients non triviaux sont finis) forment une autre classe de groupes pour lesquels le précédent théorème s'applique. Remarquons que $\mathbb{Z}_p$ n'est pas juste infini, puisque $\mathbb{Z}_p/\mathbb{Z}$ est infini.

On s'intéresse maintenant à la minoration, à travers la preuve du Théorème 1.4. En fait, on va retrouver la minoration connue pour les entiers (cf. (1.4)), simplement parce que la base qui donne cette borne dans $\mathbb{N}$ est une base parfaite.

Preuve du Théorème 1.4. Soit

$$g = \left[\frac{h(h+4)}{3}\right] + 1$$

et $k = g - 1$. D'après le Théorème 20 de [14], il existe un ensemble $A \subset \mathbb{Z}/g\mathbb{Z}$ de deux éléments tel que :

(i) $A \cup 2A \cup \cdots \cup hA = \mathbb{Z}/g\mathbb{Z}$,

(ii) $(k-1)A \neq \mathbb{Z}/g\mathbb{Z}$,

(iii) $kA = \mathbb{Z}/g\mathbb{Z}$.

Comme $\mathbb{Z}$ est un quotient de G par hypothèse, $\mathbb{Z}/g\mathbb{Z}$ est également un quotient de G. C'est-à-dire qu'il existe un sous-groupe K de G tel que $G/K \cong \mathbb{Z}/g\mathbb{Z}$.

Soit alors $B = \{x \in G : \overline{x} \in A\}$, où $\overline{x}$ désigne la classe de x dans G/K. D'après le Lemme 1.5, B vérifie

(i) $B \cup 2B \cup \cdots \cup hB = G$,

(ii) $(k-1)B \neq G$,

(iii) $kB = G$.

Autrement dit, $\mathrm{ord}^*_G(B) = k$. Mais selon la définition (1.6), c'est exactement dire que

$$X_G(h) \geq k = \left[\frac{h(h+4)}{3}\right].$$

□

1.5.2 Le cas des groupes p-torsion

Dans cette section, on suppose que $px = 0$ pour tout $x \in G$, avec p un nombre premier. Dans ce cas de torsion, l'inclusion

$$nA \subset (n+p)A$$

est vraie pour tout n. Cette simple observation va nous permettre d'améliorer considérablement la majoration de $X_G(h)$.

Preuve du Théorème 1.5 (i). Ici, nous utilisons à nouveau la définition (1.6) de X_G. On considère A une base faible d'ordre au plus h et on suppose $\mathrm{ord}^*_G(A) = k$. Comme $nA \subset (n+p)A$ pour tout n, $\cup_{i=h-p+1}^{h} A \sim G$ $(h \geq p)$. Le Lemme 1.6 implique que

$$\bigcup_{i=h-p+2}^{h+1} iA = G. \tag{1.17}$$

De plus, on a clairement

$$\bigcup_{i=h-p+3}^{h+2} iA = G. \tag{1.18}$$

Distinguons alors deux cas :

Premier cas : $(h+2)A \cap nA = \emptyset$, pour tout $h-p+3 \leq n \leq h+1$. Alors, d'après (1.17) et (1.18), et puisque $(h-p+2)A \subset (h+2)A$, il vient

$$(h-p+2)A = (h+2)A$$

En ajoutant pA des deux côtés, on obtient par récurrence

$$(h-p+2)A = (h+2+lp)A$$

pour tout $l \geq 0$. Si l est suffisamment grand, $h+2+lp \geq k$ et $(h+2+lp)A = G$. C'est pourquoi $(h-p+2)A = G$ et $k \leq h-p+2 \leq h$.
Deuxième cas : $nA \cap (h+2)A \neq \emptyset$ pour un certain $h-p+3 \leq n \leq h+1$. Posons dans ce cas $m = h+2-n$, et remarquons que m est premier avec p (car compris entre 1 et $p-1$). On procède comme au début de la preuve du Théorème 1.3. Si $c \in nA \cap (n+m)A$, alors d'après le Lemme 1.7, on a

$$(p-1)c \in \bigcap_{i=0}^{p-1} ((p-1)n + im)A. \tag{1.19}$$

Comme $\mathrm{pgcd}(m,p) = 1$, $\{im\}_{i=0}^{p-1}$ parcourt exactement toutes les classes modulo p. Si on considère $0 \leq j < p$ le représentant de im modulo p, alors $im \equiv j \pmod p$ et $im \geq j$, ce qui conduit à

$$(h-p+1+im)A \supset (h-p+1+j)A.$$

On en déduit que

$$\bigcup_{i=h-p+1}^{h} iA \subset \bigcup_{i=0}^{p-1}(h-p+1+im)A.$$

Ainsi,

$$G \sim \bigcup_{i=0}^{p-1}(h-p+1+im)A.$$

Pour tout $x \in G$, sauf un nombre fini, on a

$$x-(p-1)c \in \bigcup_{i=0}^{p-1}(h-p+1+im)A. \tag{1.20}$$

En combinant (1.20) et (1.19) on voit que pour tout $x \in G$, sauf un nombre fini,

$$\begin{aligned} x \in & ((h-p+1)+(p-1)m+(p-1)n)\,A \\ & = (h-p+1+(p-1)(h+2))A = (hp+p-1)A. \end{aligned}$$

Cela montre bien $\mathrm{ord}^*_G(A) \leq hp+p-1$. □

Afin d'obtenir une minoration de $X_G(h)$, on utilise la même idée que pour le Théorème 1.4, en exhibant une base parfaite dans un quotient de G. Remarquons que G est un espace vectoriel sur $\mathbb{F}_p$ et qu'en conséquence, tout quotient fini de G est isomorphe à $\mathbb{F}_p^d$, pour un certain d. Cette observation nous conduit à nous intéresser de plus près au cas de $\mathbb{F}_p^d$. Le prochain lemme permet de bien comprendre les bases faibles parfaites de cardinal d dans $\mathbb{F}_p^d$.

Lemme 1.11. *Soit $A = \{e_1, \ldots, e_d\} \subset \mathbb{F}_p^d$. A est une base faible parfaite de $\mathbb{F}_p^d$ si et seulement si les vecteurs $e_1, \ldots, e_d$ sont linéairement indépendants. Lorsque cette condition est satisfaite, tout élément de $\mathbb{F}_p^d$ peut s'exprimer comme somme de $\leq (p-1)d$ éléments de A, et on ne peut pas faire mieux que $(p-1)d$.*

Démonstration. Tout d'abord, il est évident que $iA \in \langle A \rangle$ pour tout i. Si A est une base faible parfaite, on a nécessairement $\langle A \rangle = \mathbb{F}_p^d$, ce qui implique que les vecteurs $e_1, \ldots, e_d$ sont linéairement indépendants. Supposons maintenant que les vecteurs $e_1, \ldots, e_d$ sont linéairement indépendants. Pour tous $0 \leq \alpha_1, \ldots, \alpha_d \leq p-1$, l'élément $\sum_{i=1}^d \alpha_i e_i$ est la somme de $\sum_{i=1}^d \alpha_i \leq (p-1)d$ éléments de A. En outre, $\sum_{i=1}^d (p-1)e_i$ ne peut pas s'écrire comme somme de moins de $(p-1)d$ éléments de A. □

Cela nous amène à une caractérisation des bases parfaites de cardinal $d+1$ dans $\mathbb{F}_p^d$.

Lemme 1.12. *Soit $A = \{e_1, \cdots, e_d, \alpha_1 e_1 + \cdots + \alpha_d e_d\} \subset \mathbb{F}_p^d$, avec les vecteurs $e_1, \ldots, e_d$ linéairement indépendants. Alors, A est une base parfaite de $\mathbb{F}_p^d$ si et seulement si*

$$\sum_{i=1}^{d} \alpha_i \not\equiv 1 \bmod p.$$

De plus, si cette condition est satisfaite, $d(p-1)A = \mathbb{F}_p^d$ et $(d(p-1)-1)A \neq \mathbb{F}_p^d$. Autrement dit, A est une base parfaite d'ordre $d(p-1)$.

Démonstration. Tout d'abord, A est une base parfaite si et seulement si $A - a$ est une base parfaite (Lemme 1.2). Remarquons tout de même que cette propriété n'est pas vraie pour une base vectorielle. On a :

$$A - e_1 = \{0, e_2 - e_1, \cdots, e_d - e_1, (\alpha_1 - 1)e_1 + \cdots + \alpha_d e_d\}.$$

Or, $(A - e_1)$ est une base parfaite si et seulement si $(A - e_1) \setminus \{0\}$ est une base faible parfaite. Et d'après le Lemme 1.11, il nous suffit de vérifier dans quels cas $(A - e_1) \setminus \{0\}$ est une famille de d vecteurs indépendants. On peut alors calculer le déterminant de cette famille :

$$\begin{aligned}
&\det(\{e_2 - e_1, \cdots, e_d - e_1, (\alpha_1 - 1)e_1 + \cdots + \alpha_d e_d\}) \\
&\qquad = \begin{vmatrix}
-1 & -1 & \cdots & \cdots & -1 & -1 & \alpha_1 - 1 \\
1 & 0 & \cdots & \cdots & 0 & 0 & \alpha_2 \\
0 & 1 & 0 & \cdots & 0 & 0 & \alpha_3 \\
0 & 0 & 1 & \ddots & \cdots & \vdots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\
\vdots & \vdots & & \ddots & 1 & 0 & \vdots \\
0 & 0 & \cdots & \cdots & 0 & 1 & \alpha_d
\end{vmatrix} \\
&\qquad = \sum_{i=1}^{d} \alpha_i - 1,
\end{aligned}$$

ce qui montre la première partie du lemme. L'ordre de A vient directement de la deuxième partie du Lemme 1.11. □

Nous sommes à présent en mesure de construire une base parfaite de $\mathbb{F}_p^d$ qui pourra jouer un rôle similaire à celui de la base A dans la preuve du Théorème 1.4.

Lemme 1.13. *Soient $e_1, \cdots, e_d$ des vecteurs linéairement indépendants de $\mathbb{F}_p^d$. Supposons que $d \not\equiv 1 \bmod p$. Alors, l'ensemble*

$$A = \{e_1, \cdots, e_d, e_1 + \cdots + e_d\}$$

satisfait les propriétés suivantes :

(i) tout élément de $\mathbb{F}_p^d$ peut s'écrire comme somme d'au plus $(d+1)(p-1)/2$ éléments de A,

(ii) $(d(p-1)-1)A \neq \mathbb{F}_p^d$,

(iii) $d(p-1)A = \mathbb{F}_p^d$.

Démonstration. Les deux dernières assertions découlent directement du Lemme 1.12 et de l'hypothèse $d \not\equiv 1 \bmod p$.

Pour démontrer la première assertion, posons $a = \sum_{i=1}^{d} e_i$. Considérons un élément quelconque $x = x_1e_1 + \cdots + x_de_d \in \mathbb{F}_p^d$, et définissons

$$\alpha_i = |\{x_j \equiv i \bmod p\}|\,.$$

Pour tout $0 \leq i \leq p-1$, on peut écrire

$$x = ia + \sum_{j=1}^{d}(x_j - i)e_j = ia + \sum_{j=1}^{d} y_j e_i$$

avec $0 \leq y_j \leq p-1$. Dans cette décomposition de x, on utilise $i + \alpha_{i+1} + 2\alpha_{i+2} + \cdots + (p-1)\alpha_{i+p-1}$ éléments de A. Ainsi, x peut s'écrire à l'aide de

$$\min_i \{i + \alpha_{i+1} + 2\alpha_{i+2} + \cdots + (p-1)\alpha_{i+p-1}\}$$

éléments de A.

Étant donné que

$$\begin{aligned}\sum_{i=0}^{p-1}(i + \alpha_{i+1} + 2\alpha_{i+2} + \cdots + (p-1)\alpha_{i+p-1}) &= \left(\sum_{i=0}^{p-1}\alpha_i\right)\frac{p(p-1)}{2} + \frac{p(p-1)}{2} \\ &= (d+1)\,\frac{p(p-1)}{2}\end{aligned}$$

le minimum sur les i est au plus $(d+1)(p-1)/2$, ce qui est bien le résultat souhaité. □

Remarque. Quand on démontre la première assertion, on utilise simplement le fait que $\min_i$ est plus petit que la moyenne sur tous les i. On pourrait potentiellement améliorer la borne $(d+1)(p-1)/2$ en travaillant plus précisément. Il nous paraît possible d'obtenir $d(p-1)/2$ à la place, mais ce ne serait pas d'un intérêt considérable.

Preuve du Théorème 1.5 (ii). Si $[2h/(p-1)-1] \not\equiv 1 \bmod p$, posons

$$d = \left[\frac{2h}{p-1} - 1\right].$$

Sinon, on choisit

$$d = \left[\frac{2h}{p-1} - 2\right].$$

La condition $h \geq 3(p-1)/2$ nous assure $d \geq 1$.

On considère $A \subset \mathbb{F}_p^d$ l'ensemble décrit dans le Lemme 1.13. D'après le travail effectué précédemment, et comme $(d+1)(p-1)/2 \leq h$ pour les deux choix de d, on a

(i) $A \cup 2A \cup \cdots \cup hA = \mathbb{F}_p^d$,

(ii) $(d(p-1)-1)A \neq \mathbb{F}_p^d$,

(iii) $d(p-1)A = \mathbb{F}_p^d$.

Or, il existe un sous-groupe K de G tel que $G/K \cong \mathbb{F}_p^d$. Si on définit $B = \{x \in G : \overline{x} \in A\}$, où $\overline{x}$ désigne la classe de x dans G/K, le Lemme 1.5 implique

(i) $B \cup 2B \cup \cdots \cup hB = G$,

(ii) $(d(p-1)-1)B \neq G$,

(iii) $d(p-1)B = G$.

On a donc bien

$$X_G(h) \geq \mathrm{ord}_G^*(B) = d(p-1) \geq (p-1)\left(\frac{2h}{p-1} - 3\right) = 2h - 3p + 3.$$

□

Dans la preuve, on obtient une meilleure borne dans le cas $d = [2h/(p-1) - 1] \not\equiv 1 \bmod p$, à savoir $X_G(h) \geq 2h - 2p + 2$. D'autre part, si $p = 2$, $d = 2h - 2$. Les Theorèmes 1.5 (i) et 1.5 (ii) donnent alors l'encadrement

$$2h - 2 \leq X_G(h) \leq 2h + 1$$

pour G un groupe 2-torsion. Il pourrait être intéressant de déterminer la valeur exacte de $X_{\mathbb{F}_2[t]}(h)$ par exemple.

1.5.3 $X_G(2)$ et $X_G(3)$

Dans cette section, on démontre le Théorème 1.6. Tout d'abord, les minorations $X_G(2) \geq 3$ et $X_G(3) \geq 4$ sont des conséquences immédiates du Théorème 1.1. Pour les majorations, on utilise encore une fois la définition (1.6) de X_G.

Preuve du Théorème 1.6(i). Supposons que

$$A \cup 2A \sim G$$

et $\mathrm{ord}^*_G(A) = k$ est fini. En adaptant la preuve du Lemme 1.6, pour tout $l \geq 2$, on a

$$lA \cup (l+1)A = G.$$

Premier cas : Il existe c dans $2A \cap 3A$. Dans ce cas, pour tout $x \in G$, sauf un nombre fini, $x - c \in A \cup 2A$. Ainsi, pour tout $x \in G$, sauf un nombre fini, $x = x - c + c \in 4A$, et $4A \sim G$.
Deuxième cas : $2A \cap 3A = \emptyset$. S'il existe $c \in 3A \cap 4A$, alors par le même procédé, on obtient $5A \sim G$. Supposons alors que $3A \cap 4A = \emptyset$. Comme $2A \cup 3A = 3A \cup 4A = G$, cela implique $2A = 4A$. Mais alors, $2A = 2mA$ pour tout $m \geq 1$. Si $2m > k$, on a $2mA = G$ et donc $2A = G$.

Dans tous les cas, $\mathrm{ord}^*_G(A) \leq 5$. □

Preuve du Théorème 1.6(ii). Supposons que

$$A \cup 2A \cup 3A \sim G$$

et $\mathrm{ord}^*_G(A) = k$ est fini. D'après le Lemme 1.6, pour tout $l \geq 2$, on a

$$lA \cup (l+1)A \cup (l+2)A = G.$$

Remarquons que si $iA \cap (i+1)A \neq \emptyset$, alors $2iA \cap (2i+1)A \cap (2i+2)A \neq \emptyset$, ce qui donne $(2i+3)A \sim G$. Ainsi, on peut supposer que

$$iA \cap (i+1)A = \emptyset \text{ pour } 1 \leq i \leq 7.$$

Autrement,

$$\mathrm{ord}^*(A) \leq 2i + 3 \leq 17.$$

On distingue cette fois-ci trois cas, ce qui est suffisant car $0 \in 2A \cup 3A \cup 4A$.

Premier cas : $0 \in 2A$. Dans ce cas, $4A \supset 2A$ et $5A \supset 3A$. Il vient alors $3A \cup 4A = G = 4A \cup 5A$. Par hypothèse, ce sont des partitions de G, ce qui conduit à $3A = 5A$. Mais alors, $3A = (2m+3)A$ pour tout $m \geq 1$. En conséquence, $3A = G$ et $\mathrm{ord}^*_G(A) \leq 3$.

Deuxième cas : $0 \in 3A$. On a donc $5A \supset 2A$ et $6A \supset 3A$. Comme $5A \cap 6A = \emptyset$, $5A \cap 3A = \emptyset$. Ainsi, $3A \cup 4A \cup 5A = G$ est une partition de G. D'autre part, comme $2A \cup 3A \cup 4A = G$, on en déduit que $2A = 5A$ Comme dans le cas précédent, on obtient $\mathrm{ord}^*_G(A) \leq 2$.

Troisième cas : $0 \in 4A$. Cette fois-ci, $6A \supset 2A$, $7A \supset 3A$, $8A \supset 4A$, et $2A \cup 3A \cup 4A = G = 6A \cup 7A \cup 8A = G$. Comme $7A$ est disjoint de $6A$ et $8A$, on en déduit $3A = 7A$. Et cela donne $\mathrm{ord}^*_G(A) \leq 3$. □

1.6 La fonction S

La clé pour généraliser les arguments de Cassaigne et Plagne [2] est la notion de *moyennabilité* d'un groupe.

1.6.1 Moyennabilité

On se place ici dans le cadre d'un groupe G abélien infini qu'on munit de la topologie discrète.

Remarque. En fait, on définit la moyennabilité plus généralement sur les groupes localement compacts, munis de la mesure de Haar μ (unique à translation près). Pour un tel groupe G, on note $L^\infty(G,\mu)$ l'espace de Banach des fonctions mesurables essentiellement bornées de G dans $\mathbb{C}$. S'intéresser ici seulement aux groupes discrets simplifiera notre propos, dans la mesure où on a dans ce cas $L^\infty(G,\mu) = l^\infty(G)$ l'ensemble des fonctions bornées de G dans $\mathbb{C}$.

Parmi les nombreuses définitions de la moyennabilité, on travaille avec celle faisant intervenir les *moyennes invariantes*.

Définition 1.3. Une moyenne invariante à droite sur G (avec G un groupe discret, non nécessairement abélien) est une fonctionnelle linéaire $\Lambda : l^\infty(G) \to \mathbb{R}$ vérifiant :

1. Λ est positive : si $f \geq 0$ sur G, alors $\Lambda(f) \geq 0$,
2. Λ est de norme 1 : $\Lambda(1_G) = 1$ où 1_G est la fonction caractéristique de G,
3. Λ est invariante à droite : $\Lambda(\tau_g f) = \Lambda(f)$ pour tout $f \in l^\infty(G)$ et $g \in G$, avec τ_g la translation à droite ($\tau_g(f(x)) = f(xg)$).

Dans le cas d'un groupe abélien, il ne sera pas nécessaire de préciser "à droite".

Définition 1.4. Le groupe G est dit *moyennable* s'il existe une moyenne invariante à droite sur G.

Commençons par énoncer le Théorème G.2.1 de [1] qu'on utilisera fondamentalement par la suite.

Théorème 1.9 (Markov,Kakutani)**.** *Tout groupe topologique* ***abélien*** *est moyennable.*

Même s'il s'agit du seul résultat utile pour nous, rappelons maintenant d'autres résultats bien connus à ce sujet. Pour un aperçu plus complet, on conseille fortement de lire [1, Appendice G].

La propriété de moyennabilité se transmet aux sous-groupes fermés et aux quotients d'un groupe. Or, il est assez facile de voir que le groupe libre à deux

éléments F_2 n'est pas moyennable. Pour cela, on peut se baser sur le graphe de Cayley du groupe, on s'aperçoit qu'une moyenne ne peut pas charger les points, puis, en utilisant l'invariance de la moyenne et les symétries du graphe, on peut montrer que la moyenne est nécessairement la fonction nulle. Ainsi, tout groupe qui contient F_2 n'est pas moyennable. Von Neumann s'est d'ailleurs demandé si c'était le seul cas d'obstruction à la moyennabilité. Mais on sait maintenant construire des groupes non moyennables qui ne contiennent pas F_2.

Remarque. Dans le cas des groupes localement compacts, dans la définition de la moyenne, on remplace $l^\infty(G)$ par $L^\infty(G, \mu)$. La preuve qui va suivre peut être adaptée à ces groupes lorsqu'ils sont abéliens, si on change la définition de l'ordre. En effet, il ne faut plus demander à ce que $hA \sim G$, mais plutôt $\mu(G \setminus hA) = 0$ dans ce cas.

1.6.2 Preuve du Théorème 1.7

Démonstration. La minoration $h + 1 \leq S_G(h)$ est une conséquence immédiate du Théorème 1.1. Il reste à démontrer la majoration. G est un groupe abélien, donc d'après le Théorème 1.9, on peut se munir de Λ une moyenne invariante sur G. G étant infini, on s'aperçoit aisément que $\Lambda(1_I) = 0$ pour tout sous-ensemble fini $I \subset G$, où 1_I est la fonction caractéristique de I (il suffit de le constater pour un singleton). En conséquence, pour tout B sous-ensemble G tel que $B \sim G$, par linéarité de Λ,

$$\Lambda(1_B) = \Lambda(1_G - 1_{G\setminus B}) = \Lambda(1_G) - \Lambda(1_{G\setminus B}) = 1. \tag{1.21}$$

Considérons A une base d'ordre h de G. Sans perte de généralité, on peut supposer que $0 \in A$.

Pour chaque élément $a \in A$, on définit f_a une fonction sur G par

$$f_a(x) = \begin{cases} 1, & \text{si } x \in hA \setminus h\,(A \setminus \{a\}) \\ 0, & \text{autrement.} \end{cases}$$

Autrement dit, $f_a(x) = 1$ si et seulement si a est essentiel dans toute représentation de x comme somme de h éléments de A.

Comme il est fait dans la preuve de [2], on observe deux choses. Premièrement, pour tout $x \in G$ et tout sous-ensemble fini $I \subset A$, on a $\sum_{a\in I} f_a(x) \leq h$. En effet, si $x \notin hA$, alors clairement, $f_a(x) = 0$ pour tout $a \in A$. Supposons que $x \in hA$, et fixons une représentation

$$x = a_1 + \cdots + a_h$$

avec $a_i \in A$. Alors $f_a(x)$ peut valoir 1 seulement si a est l'un des a_i, et il y a au plus h tels éléments.

Mais alors, $h1_G - \sum_{a\in I} f_a \geq 0$. En utilisant la positivité et la linéarité de Λ, on a finalement prouvé le fait suivant.

Fait 1. Pour tout sous-ensemble fini $I \subset A$, on a

$$\sum_{a\in I} \Lambda(f_a) \leq h.$$

La deuxième chose que l'on veut montrer est :

Fait 2. Si $a \in A$ est tel que $\Lambda(f_a) < 1/h$, alors il existe $x \in G$ tel que

$$x + ia \in h(A \setminus \{a\})$$

pour tout $i = 0, 1, \ldots, h-1$.

En effet, comme Λ est invariante par translation et linéaire,

$$1 > h\Lambda(f_a) = \sum_{i=0}^{h-1} \Lambda\left(\tau_{ia} f_a\right) = \Lambda\left(\sum_{i=0}^{h-1} \tau_{ia} f_a\right).$$

Or, en posant

$$B = \left\{ x \in G \text{ tel que } \sum_{i=0}^{h-1} \tau_{ia} f_a(x) \geq 1 \right\},$$

si $B \sim G$, on a

$$\sum_{i=0}^{h-1} \tau_{ia} f_a \geq 1_B,$$

ce qui donne d'après (1.21)

$$\Lambda\left(\sum_{i=0}^{h-1} \tau_{ia} f_a\right) \geq \Lambda(1_B) = 1.$$

Ainsi, $B \not\sim G$, donc B^c, le complémentaire de B, est infini. Or, tout élément x de B^c vérifie

$$1 > \sum_{i=0}^{h-1} \tau_{ia} f_a(x) = \sum_{i=0}^{h-1} f_a(x + ia).$$

En conséquence, quelque soit $x \in B^c$, $f_a(x + ia) = 0$ pour tout $i = 0, 1, \ldots, h-1$.

Mais comme $hA \sim G$, il existe $x \in B^c$ tel que $x + ia \in hA$ pour tout $i = 0, 1, \ldots, h-1$. Et ce x vérifie donc

$$x + ia \in h(A \setminus \{a\})$$

pour tout $i = 0, 1, \ldots, h-1$, ce qui est le résultat souhaité.

D'après le Fait 1, on sait que pour tout $a \in A$, sauf un nombre fini, on a $a \neq 0$ et $\Lambda(f_a) < 1/h$ (on ne peut en effet pas avoir plus de h^2 élements a tels que $\Lambda(f_a) \geq 1/h$). Pour un tel a, considérons x vérifiant $x + ia \in h(A \setminus \{a\})$ pour tout $i = 0, 1, \ldots, h-1$, dont l'existence est donnée par le Fait 2. Ainsi, pour tout $y \in G$, sauf un nombre fini, on a $y - x \in hA$ et $y - x \neq ha$. Mais alors, si on écrit

$$y - x = a_1 + \cdots + a_h,$$

où les a_j sont dans A, il y a au plus $h-1$ a_j égaux à a. En notant i ce nombre $(0 \leq i \leq h-1)$, on a

$$y - x - ia \in (h-i)(A \setminus \{a\}).$$

Cela implique que

$$y = (y - x - ia) + (x + ia) \in (2h - i)(A \setminus \{a\}) \subset 2h(A \setminus \{a\})$$

car $A \setminus \{a\}$ contient 0. On a bien démontré que $A \setminus \{a\}$ est une base d'ordre au plus $2h$. □

1.6.3 Preuve du Théorème 1.8

Démonstration. On a déjà $S_G(2) \geq 3$. Il reste à montrer que $S_G(2) \leq 3$. Soit donc A une base d'ordre 2 de G. On dit que $b \in A$ est *mauvais* si $\mathrm{ord}^*(A \setminus \{b\}) \geq 4$ et *bon* autrement. L'objectif est de montrer que A n'admet qu'un nombre fini de mauvais éléments.

En considérant $A - c$ à la place de A, avec c un élément de A, on peut supposer que $0 \in A$.

Commençons par examiner les propriétés d'un mauvais élément $b \in A, b \neq 0$. Notons $A_b = A \setminus \{b\}$. A étant une base d'ordre 2, on a

$$G \sim 2A \sim 2A_b \cup (A_b + b)\,. \tag{1.22}$$

Soit a un élément quelconque de A_b. Alors,

$$G \sim 2A + a \sim (2A_b + a) \cup (A_b + b + a) \subset 3A_b \cup (A_b + b + a)\,. \tag{1.23}$$

De plus, comme $0 \in A_b$, on déduit de (1.22) l'inclusion

$$G \sim 2A_b \cup (A_b + b) \subset 3A_b \cup (A_b + b). \tag{1.24}$$

D'après (1.23) et (1.24), les ensembles $(A_b + b + a)$ et $(A_b + b)$ contiennent tous deux $G \setminus 3A_b$, à un nombre fini d'éléments près. Or, b étant mauvais, $G \setminus 3A_b$ est infini, ce qui implique que $(A_b + b + a) \cap (A_b + b)$ est infini. Autrement dit, on a démontré la chose suivante :

Propriété 1 : Pour tout $a \in A_b$, $(A_b + a) \cap A_b$ est infini.

Ensuite on prouve :

Propriété 2 : $(A_b + b) \cap A_b = \emptyset$.

En effet, supposons, par l'absurde, qu'il existe $a_1, a_2 \in A_b$ tels que $b + a_1 = a_2$. Pour tout $x \in A$, sauf un nombre fini, on a $x - a_1 \in 2A \setminus \{2b\}$. Si $x - a_1 \in 2A_b$, alors $x \in 3A_b$. Sinon, $x - a_1 \in A_b + b$, mais alors $x \in A_b + a_2 \subset 2A_b \subset 3A_b$. Ainsi $3A_b \sim G$, ce qui fournit une contradiction.

Supposons à présent qu'il existe un autre mauvais élément $b' \in A, b' \neq 0$. Par la Propriété 1, on sait que $(A_b + b') \cap A_b$ est infini. En conséquence, $(A \setminus \{b, b'\} + b') \cap (A \setminus \{b, b'\})$ est infini. Mais cela contredit la Propriété 2 (où on remplace b par b'). □

En fait, la preuve montre qu'il y a au plus un mauvais élément dans A. En effet, on a montré que pour tout $c \in A$, il y a au plus un mauvais élément dans A qui soit différent de c. En appliquant cette observation à un bon élément c (dont on connaît maintenant l'existence), cela implique qu'il y a bien au plus un mauvais élément dans A.

Bibliographie

[1] B. Bekka, P. de la Harpe, A. Valette, *Kazhdan's Property (T)*, New Mathematical Monographs, 11. Cambridge University Press, Cambridge, 2008.

[2] J. Cassaigne, A. Plagne, *Grekos' S function has a linear growth*, Proceedings of the American Mathematical Society 132 (2004), 2833–2840.

[3] P. Erdős, R. L. Graham, *On bases with an exact order*, Acta Arith. 37 (1980) 201–207.

[4] P. Erdős, R. L. Graham, *Old and new problems and results in combinatorial number theory*, Monogr. Enseign. Math. 28, 1980.

[5] G. Grekos, *Extremal problems about asymptotic bases : a survey*, Combinatorial number theory, 237–242, de Gruyter, Berlin, 2007.

[6] G. Grekos, *Sur l'ordre d'une base additive*, Séminaire de Théorie des Nombres de Bordeaux (Talence, 1987–1988), Exp. No. 31, 13 pp.

[7] G. Grekos, *Extremal problems about additive bases*, Acta Math. Inform. Univ. Ostraviensis 6 (1998), no. 1, 87–92.

[8] G. Grekos, *Minimal additive bases and related problems*, Number theory days, 1980 (Exeter, 1980), 300–305, London Math. Soc. Lecture Note Ser., 56, Cambridge Univ. Press, Cambridge, 1982.

[9] E. Härtter, *Ein Beitrag zur Theorie der Minimalbasen*, J. reine angew. Math. 196 (1956), 170–204.

[10] M. B. Nathanson, *Minimal bases and powers of 2*, Acta Arith. 49 (1988), 525–532.

[11] M. B. Nathanson, *Unique representation bases for the integers*, Acta Arith. 108(1) (2003), 1–8.

[12] J. C. M. Nash, *Some applications of a theorem of M. Kneser*, J. Number Theory 44 (1993), 1–8.

[13] A. Plagne, *Problemas combinatorios sobre bases aditivas*, Gaceta de la Real Sociedad Matemática Española 9 (2006), 191–201.

[14] A. Plagne, *A propos de la fonction X d'Erdős et Graham*, Annales de l'Institut Fourier 54 (2004), no. 6, 1717–1767.

[15] A. Plagne, *Sur le nombre d'éléments exceptionnels d'une base additive*, Journal für die Reine und Angewandte Mathematik 616 (2008), 47–65.

[16] A. Stöhr, *Gelöste und ungelöste Fragen über Basen der natürlichen Zahlenreihe I, II*, J. reine angew. Math. 194 (1955), 40–65 and 111–140.

Chapitre 2

Propriétés additives des pseudo puissances s-ièmes

Ce chapitre reprend, à peu de choses près et en français, le texte d'un article écrit en collaboration avec Javier Cilleruelo, Jean-Marc Deshouillers et Alain Plagne.

2.1 Introduction

Considérons s un entier supérieur ou égal à 2. Nous allons nous intéresser aux pseudo puissances s-ièmes, ce qui correspond à une suite aléatoire d'entiers qui a la même répartition que les vraies puissances s-ièmes. C'est pourquoi il semble assez naturel, avant de s'intéresser aux travaux que nous avons effectués, de parler du problème de Waring, ce qui est en fait la version non probabiliste de ce qui suivra.

2.1.1 Problème de Waring

En 1770, Waring s'est demandé si, pour tout entier naturel k, il existe un entier naturel m tel que tout entier soit la somme d'au plus m puissances k-ièmes d'entiers positifs. Une réponse affirmative fut apportée par Hilbert en 1909, à l'aide d'une méthode essentiellement algébrique. Naturellement, on a ensuite cherché à déterminer le plus petit entier m vérifiant la condition de Waring, que l'on note $g(k)$. Il s'agit, compte tenu des définitions données dans le Chapitre 1, de l'ordre des puissances k-ièmes en tant que *base parfaite* (ici, comme 0 est une puissance k-ième, il n'y a pas de distinction entre *base* et *base faible*). Plusieurs travaux ont permis de connaître certaines valeurs de g : $g(2) = 4$ (Lagrange, 1770), $g(3) = 9$ (Kepner et Wieferich, 1909), $g(4) = 19$ (Balasubramanian, Deshouillers et Dress,

1986), $g(5) = 37$ (Chen, 1964), $g(6) = 73$ (Pillai, 1940)...

Si on pose

$$\alpha_k = 2^k \left[\left(\frac{3}{2} \right)^k \right] - 1,$$

comme $\alpha_k \leq 3^k$, on sait qu'il ne peut s'écrire comme somme de puissances k-ièmes qu'à l'aide de 1^k et 2^k. Son écriture la plus "économique" est alors

$$\alpha_k = \left(\left[\left(\frac{3}{2} \right)^k \right] - 1 \right) 2^k + \left(2^k - 1 \right) 1^k.$$

Cela implique

$$g(k) \geq 2^k + \left[\left(\frac{3}{2} \right)^k \right] - 2.$$

L'égalité est une conjecture, vérifiée pour les $k \leq 471600000$ (cf. [11]).

Une autre quantité, mise en évidence par Hardy et Littlewood, intéresse les mathématiciens. Il s'agit du plus petit entier m tel que tout entier "suffisamment grand" soit somme de m puissances k-ièmes d'entiers positifs, que l'on note $G(k)$. En d'autres mots, c'est l'ordre des puissances k-ièmes en tant que base (asymptotique). Les seuls valeurs exactes de G connues sont $G(2) = 4$ (Lagrange) et $G(4) = 16$ (Davenport, 1939). Pour les autres valeurs de k, des encadrements existent tout de même. On sait à l'heure actuelle (Théorème 12.5 dans [13]) qu'il existe une constante C telle que

$$G(k) \leq k \left(\log k + \log \log k + C \right).$$

Par des arguments combinatoires, on peut voir que $G(k) \geq k+1$, et il est conjecturé que $G(k) = k + 1$ en l'absence de restrictions modulaires.

2.1.2 Modèle probabiliste et résultats connus

Le modèle probabiliste auquel nous allons maintenant nous intéresser a été proposé par Erdős et Rényi dans [8], en 1960, pour créer des suites A asymptotiquement proches des puissances s-ièmes. Plus précisément, on considère un espace de probabilité $(\mathcal{U}, \mathcal{T}, \mathbb{P})$ et une suite de variables aléatoires indépendantes de Bernoulli $(\xi_n)_{n \in \mathbb{N}}$ à valeurs dans $\{0, 1\}$ et tels que

$$\mathbb{P}(\xi_n = 1) = \frac{1}{s} n^{-1+1/s} \quad \text{et} \quad \mathbb{P}(\xi_n = 0) = 1 - \frac{1}{s} n^{-1+1/s}.$$

À chaque $u \in \mathcal{U}$, on associe une suite d'entiers $A = A_u$ telle que $n \in A_u$ si et seulement si $\xi_n(u) = 1$. En d'autres mots, les évènements $\{n \in A\}$ sont indépendants et la probabilité que n appartienne à A est égale à $\mathbb{P}(n \in A) = n^{-1+1/s}/s$.

La fonction de comptage de cette suite aléatoire A satisfait presque sûrement la relation asymptotique $|A \cap [\![1, x]\!]| \sim x^{1/s}$ lorsque x tend vers l'infini (cf. [8] ou [12]), ce qui explique la terminologie *pseudo puissances s-ièmes*. Notons que Deshouillers, Hennecart et Landreau ont donné dans [5] un raffinement de ce modèle plus adapté à l'étude des sommes de carrés.

Comme on a pu le voir pour le problème de Waring, on cherche à déterminer pour quels k on a $kA \sim \mathbb{N}$. Ici, en termes probabilistes, on veut savoir pour quels k tout entier s'écrit presque sûrement comme somme de k élements de A. Dans cette optique, Goguel [9] a prouvé en 1975 que, presque sûrement, l'ensemble somme

$$sA = \{a_1 + \cdots + a_s \text{ with } \ a_i \in A\}$$

a pour densité $1 - e^{-\lambda_s}$ où

$$\lambda_s = \frac{\Gamma^s(1/s)}{s^s \ s!}.$$

Notons que cette quantité, faisant intervenir la fonction Γ d'Euler, apparaîtra fréquemment tout au long de l'étude. Ainsi, on sait que, presque sûrement, A n'est pas une base d'ordre s. Il a d'ailleurs été montré récemment, par Cilleruelo et Deshouillers (cf. [4]) que la suite $(b_n)_{n\in\mathbb{N}}$ des éléments ordonnés de sA a, presque sûrement, une infinité de "trous" de taille logarithmique. Ils ont en fait démontré le résultat

$$\limsup_{n\to+\infty} \frac{b_{n+1} - b_n}{\log b_n} = \frac{1}{\lambda_s}. \tag{2.1}$$

Ainsi, les sommes de s éléments de A ne sont pas suffisantes pour engendrer les entiers. Deshouillers et Iosifescu, en étudiant la probabilité pour un entier de ne pas être somme de $s+1$ puissances s-ièmes, ont prouvé dans [6] que, presque sûrement, une suite de pseudo puissances s-ièmes est une base d'ordre $s+1$.

Dès lors qu'on sait que s n'est pas suffisant alors que $s+1$ l'est, on peut se demander à quel point on a besoin d'un $(s+1)$-ième élément. Dans la suite, nous appellerons *complément additif* de sA un ensemble B tel que tout entier s'écrit, presque sûrement, comme somme de s éléments de A et d'un élément de B. Il s'agit du principal objet de notre étude.

2.1.3 Nouveaux résultats

Le premier résultat montre que le $(s+1)$-ième élément peut être pris "petit" parmi les pseudo puissances s-ièmes.

Théorème 2.1. *Soient $s \geq 2$ un entier et $c > (\lambda_s(1 - 2\lambda_s))^{-1}$. Presque sûrement, une suite de pseudo puissances s-ièmes A a la propriété suivante : tout entier suffisamment grand n peut s'écrire sous la forme*

$$n = a_1 + \cdots + a_{s+1}, \qquad \text{avec } a_i \in A, \qquad a_{s+1} < (c \log n)^s.$$

Nous avons certaines raisons de penser que la conclusion n'est plus vraie si $c < \lambda_s^{-1}$, ce dont nous discuterons à la fin de la section 2.3. Remarquons que $\lambda_s < 1/2$ pour $s \geq 2$, et on a donc bien $c > 0$.

Si l'on considère maintenant une suite fixée B (donc non aléatoire), le théorème suivant donne une condition suffisante sur la fonction de comptage $B(n) = |B \cap [\![1, n]\!]|$ pour qu'elle soit un complément additif de sA.

Théorème 2.2. *Soit $s \geq 2$ un entier. Soit B une suite d'entiers fixée vérifiant*

$$\liminf_{n\to\infty} \frac{B(n)}{\log n} > \lambda_s^{-1}.$$

Alors, une suite de pseudo puissances s-ièmes A a, presque sûrement, la propriété suivante : tout entier suffisamment grand n peut s'écrire sous la forme

$$n = a_1 + \cdots + a_s + b, \qquad \text{avec les } a_i \text{ distincts dans } A \text{ et où } b \in B.$$

On démontre ensuite que le Théorème 2.2 est fort au sens où la constante λ_s^{-1} intervenant dans le résultat est la plus petite possible.

Théorème 2.3. *Soit $s \geq 2$ un entier. Soit B une suite d'entiers fixée vérifiant*

$$\liminf_{n\to\infty} \frac{B(n)}{\log n} < \lambda_s^{-1}.$$

Alors, une suite de pseudo puissances s-ièmes A a, presque sûrement, la propriété suivante : il existe une infinité d'entiers n qui ne peuvent pas s'écrire sous la forme

$$n = a_1 + \cdots + a_s + b, \qquad \text{avec les } a_i \text{ distincts dans } A \text{ et où } b \in B.$$

Les Théorèmes 2.2 and 2.3 font apparaître un phénomène de seuil, qui nous amène naturellement à nous demander comment se comportent les suites B vérifiant

$$\liminf_{n\to\infty} \frac{B(n)}{\log n} = \lambda_s^{-1}. \tag{2.2}$$

Nous verrons dans la dernière section, que dans ce cas limite, il existe des suites B vérifiant (2.2) et la conclusion du Théorème 2.2, tandis que d'autres satisfont (2.2) et la conclusion du Théorème 2.3.

Avant d'aborder les preuves des trois théorèmes (sections 2.3, 2.4 et 2.5), nous allons avoir besoin de différents lemmes, qui feront l'objet de la section 2.2, dans laquelle nous rappellerons également quelques résultats de probabilités bien connus.

2.2 Lemmes préparatoires

Nous utiliserons la notation standard $\ll$ due à Vinogradov pour dire "plus petit à une constante multiplicative près". Ces constantes, dans notre travail, dépendront toujours et seulement du paramètre $s \geq 2$. Nous ne rappellerons pas cette dépendance lorsque nous utiliserons cette notation.

Le premier lemme, calculatoire et technique, apparaît déjà dans [4], si ce n'est pour le troisième point où l'on démontre ici un résultat plus général.

Lemme 2.1. *Soient s et t deux entiers tels que $s \geq 2$ et $1 \leq t \leq s-1$. On a alors*

(i) pour $z \geq 1$,

$$\sum_{\substack{1\leq x_1,\dots,x_t\\ x_1+\cdots+x_t=z}} (x_1\cdots x_t)^{-1+1/s} \ll z^{-1+t/s},$$

(ii) pour $z \geq 2$,

$$\sum_{\substack{1\leq x_1,\dots,x_t\\ x_1+\cdots+x_t<z}} (x_1\cdots x_t)^{-1+1/s}\Big(z-(x_1+\cdots+x_t)\Big)^{-2t/s} \ll z^{-1/s}\log z,$$

(iii) si g est une fonction positive satisfaisant $g(z) = o(z)$ lorsque z tend vers l'infini, alors

$$\lim_{\substack{z\to+\infty\\ z\in\mathbb{N}}} \sum_{\substack{g(z)\leq x_s<\cdots<x_1\\ x_1+\cdots+x_s=z}} (x_1\cdots x_s)^{-1+1/s} = s^s\lambda_s.$$

Démonstration. Comme dit précédemment, les points (i) et (ii) proviennent du Lemme 1 de [4], en prenant $a_1 = \cdots = a_s = 1$. Nous préférons tout de même en donner les preuves ici, puisque c'est le type d'inégalités qu'on aura à manipuler fréquemment par la suite.

(i) On peut supposer que x_t est plus grand que z/t. Étant donné que $-1+1/s < 0$, on a

$$\begin{aligned}
\sum_{\substack{1\leq x_1,\cdots,x_t\\ x_1+\cdots+x_t=z}} (x_1\cdots x_t)^{-1+1/s} &\leq \left(\frac{z}{t}\right)^{-1+1/s} \sum_{x_1,\cdots,x_{t-1}<z} (x_1\cdots x_{t-1})^{-1+1/s}\\
&\ll z^{-1+1/s}\left(\sum_{x<z} x^{-1+1/s}\right)^{t-1}\\
&\ll z^{-1+1/s}\left(z^{1/s}\right)^{t-1} \text{ (comparaison somme-intégrale)}\\
&\ll z^{-1+t/s}.
\end{aligned}$$

(ii) Pour démontrer cette inégalité, nous utiliserons aussi la "comparaison somme-intégrale", ce que nous indiquons par (*) dans ce qui suit :

$$\begin{aligned}
& \sum_{\substack{1\le x_1,\dots,x_t\\ x_1+\cdots+x_t<z}} (x_1\cdots x_t)^{-1+1/s}\Big(z-(x_1+\cdots+x_t)\Big)^{-2t/s}\\
= \quad & \sum_{m<z}(z-m)^{-2t/s}\sum_{\substack{1\le x_1,\cdots,x_t\\ x_1+\cdots+x_t=m}}(x_1\cdots x_t)^{-1+1/s}\\
(\text{par (i)}) \quad \ll \quad & \sum_{m<z}(z-m)^{-2t/s}m^{-1+t/s}.
\end{aligned}$$

En séparant la somme de droite en deux, on a finalement

$$\begin{aligned}
& \sum_{\substack{1\le x_1,\dots,x_t\\ x_1+\cdots+x_t<z}} (x_1\cdots x_t)^{-1+1/s}\Big(z-(x_1+\cdots+x_t)\Big)^{-2t/s}\\
\ll \quad & \sum_{m\le z/2}(z-m)^{-2t/s}m^{-1+t/s}+\sum_{z/2<m<z}(z-m)^{-2t/s}m^{-1+t/s}\\
(*) \quad \ll \quad & z^{-2t/s}z^{t/s}+z^{-1+t/s}\sum_{z/2<m<z}(z-m)^{-2t/s}\\
(*) \quad \ll \quad & z^{-t/s}+z^{-1+t/s}(1+\log z+z^{1-2t/s})\\
\ll \quad & z^{-t/s}+z^{-1+t/s}\log z\\
\ll \quad & z^{-1/s}\log z.
\end{aligned}$$

Remarquons que la présence du logarithme n'a de sens que pour $s=2$ et $t=1$.

(iii) Quant au troisième point, on le retrouve également dans [4] pour le cas particulier $g(z)=1$. Pour étendre cela à une fonction plus générale, il est suffisant de montrer que

$$\sum_{\substack{1\le x_s\le g(z)\\ 1\le x_{s-1}<\cdots<x_1\\ x_1+\cdots+x_s=z}}(x_1\cdots x_s)^{-1+1/s}=o(1).$$

Pour voir cela, utilisons (i) avec $t = s-1$ et majorons de la façon suivante :

$$\begin{aligned}
\sum_{\substack{1\le x_s\le g(z)\\ 1\le x_{s-1}<\cdots<x_1\\ x_1+\cdots+x_s=z}} (x_1\cdots x_s)^{-1+1/s} &\le \sum_{1\le x_s<g(z)} x_s^{-1+1/s} \sum_{\substack{1\le x_{s-1}<\cdots<x_1\\ x_1+\cdots+x_{s-1}=z-x_s}} (x_1\cdots x_{s-1})^{-1+1/s}\\
&\ll \sum_{1\le x_s<g(z)} x_s^{-1+1/s}(z-x_s)^{-1/s}\\
&\ll (z-g(z))^{-1/s} \sum_{1\le x_s<g(z)} x_s^{-1+1/s}\\
&\ll \left(\frac{g(z)}{z}\right)^{1/s}\\
&= o(1),
\end{aligned}$$

ce qui est le résultat souhaité.

□

Rappelons maintenant quelques résultats connus de probabilités. Pour fixer les notations, X désigne une variable aléatoire sur un espace probabilisé fixé, tandis que nous noterons classiquement $\mathbb{P}, \mathbb{E}$ et $\mathbb{V}$ la probabilité, l'espérance et la variance. Commençons par l'inégalité de Bienaymé-Tchebychev, que nous énonçons dans une forme adaptée à notre propos :

$$\mathbb{P}\left(X < \frac{\mathbb{E}[X]}{2}\right) \le \frac{4\mathbb{V}(X)}{\mathbb{E}[X]^2}. \tag{2.3}$$

Nous utiliserons également de manière cruciale le lemme de Borel-Cantelli.

Théorème 2.4 (Lemme de Borel-Cantelli)**.** *Soit $(F_i)_{i\in\mathbb{N}}$ une suite d'évènements. Si $\sum_{i=1}^{+\infty} \mathbb{P}(F_i) < +\infty$ alors, avec probabilité 1, seul un nombre fini d'évènements F_i a lieu.*

Pour finir, nous aurons besoin de résultats dus à Janson [10] (voir aussi [2]), donnant des inégalités de corrélation. Nous le présentons ici comme il est fait dans le théorème 8.1.1 de [1].

On utilise la notation suivante : si Ω est un ensemble dont les éléments sont eux-mêmes des ensembles, alors pour ω, ω' dans Ω, la notation $\omega \sim \omega'$ signifie que $\omega \neq \omega'$ et $\omega \cap \omega' \neq \emptyset$. De plus, E^c désigne l'événement complémentaire de E.

Théorème 2.5 (Inégalités de Janson)**.** *Soit $(E_\omega)_{\omega\in\Omega}$ une famille finie d'événements indexés par les éléments de Ω. Supposons de plus que $\mathbb{P}(E_\omega) \le 1/2$ pour tout $\omega \in \Omega$. Alors, la quantité $\mathbb{P}\Big(\bigcap_{\omega\in\Omega} E_\omega^c\Big)$ satisfait*

(i) pour la borne inférieure

$$\mathbb{P}\Big(\bigcap_{\omega\in\Omega} E_\omega^c\Big) \geq \prod_{\omega\in\Omega} \mathbb{P}(E_\omega^c)$$

et

(ii) pour la borne supérieure

$$\mathbb{P}\Big(\bigcap_{\omega\in\Omega} E_\omega^c\Big) \leq \Big(\prod_{\omega\in\Omega} \mathbb{P}(E_\omega^c)\Big) \exp\Big(2 \sum_{\substack{\omega,\omega'\in\Omega \\ \omega\sim\omega'}} \mathbb{P}(E_\omega \cap E_{\omega'})\Big).$$

2.3 Preuve du Théorème 2.1

On rappelle que $\lambda_s < 1/2$ quand $s \geq 2$ et on choisit un réel $c > (\lambda_s(1-2\lambda_s))^{-1}$.

On représente les ensembles de $s+1$ éléments distincts par $\omega = \{x_1, \ldots, x_{s+1}\}$ avec

$$x_{s+1} < \cdots < x_1.$$

De plus, on note $\sigma(\omega) = x_1 + \cdots + x_{s+1}$ et, pour chaque n, on introduit

$$\Omega_n = \{\omega \text{ tel que } \sigma(\omega) = n,\ x_{s+1} < (c\log n)^s \text{ et } (c\log n)^s < x_s\}.$$

Si on désigne par E_ω l'événement $\omega \subset A$ et par $\mathbb{I}$ la fonction indicatrice d'un événement, la fonction

$$r(n, A) = \sum_{\omega\in\Omega_n} \mathbb{I}(E_\omega)$$

compte le nombre de représentations de n sous la forme $n = x_1 + \cdots + x_{s+1}$, où

$$x_i \in A, \quad (c\log n)^s < x_s < \cdots < x_1, \quad \text{et} \quad x_{s+1} < (c\log n)^s.$$

L'objectif est de montrer que, presque sûrement, cette fonction n'est nulle que pour un nombre fini d'éléments. Ainsi, si on démontre que la série des $\mathbb{P}(r(n,A) = 0)$ converge, le lemme de Borel-Cantelli donnera le résultat. Comme pour tout ω dans Ω_n, $\mathbb{P}\Big(E_\omega\Big) \leq 1/2$ (même plus petit que $1/s^{s+1}$), on va pouvoir appliquer la deuxième inégalité de Janson :

$$\mathbb{P}(r(n,A) = 0) = \mathbb{P}\Big(\bigcap_{\omega\in\Omega_n} E_\omega^c\Big) \leq \prod_{\omega\in\Omega_n} \mathbb{P}\Big(E_\omega^c\Big) \times \exp\left(2\Delta_n\right),$$

avec

$$\Delta_n = \sum_{\substack{\omega,\omega'\in\Omega_n \\ \omega\sim\omega'}} \mathbb{P}(E_\omega \cap E_{\omega'}). \tag{2.4}$$

Commençons par étudier le produit.

Lemme 2.2. *Lorsque n tend vers l'infini, on a*

$$\prod_{\omega\in\Omega_n} \mathbb{P}\Big(E_\omega^c\Big) = \exp\Big(-(1+o(1))c\lambda_s \log n\Big).$$

Démonstration. Si on somme les probabilités des E_ω, quand ω parcourt Ω_n, on obtient

$$\sum_{\omega\in\Omega_n} \mathbb{P}(E_\omega) = \frac{1}{s^{s+1}} \sum_{1\le x_{s+1}<(c\log n)^s} x_{s+1}^{1/s-1} \sum_{\substack{(c\log n)^s<x_s<\cdots<x_1\\ x_1+\cdots+x_s=n-x_{s+1}}} (x_1\dots x_s)^{1/s-1}.$$

Pour chaque $x_{s+1} < (c\log n)^s$, on peut appliquer le Lemme 2.1 (iii) avec $z = n - x_{s+1} \sim n$, ce qui donne

$$\sum_{\omega\in\Omega_n} \mathbb{P}(E_\omega) = (1+o(1))\frac{\lambda_s}{s} \sum_{1\le x_{s+1}<(c\log n)^s} x_{s+1}^{1/s-1} = (1+o(1))c\lambda_s\log n,$$

et le résultat vient alors de la relation

$$\prod_{\omega\in\Omega_n} \mathbb{P}\Big(E_\omega^c\Big) = \exp\left(\sum_{\omega\in\Omega_n} \log(1-(\mathbb{P}(E_\omega))\right) = \exp\left(-(1+o(1))\sum_{\omega\in\Omega_n}\mathbb{P}(E_\omega)\right).$$

□

On s'intéresse maintenant au terme de corrélation Δ_n défini par (2.4).

Lemme 2.3. *Lorsque n tend vers l'infini, on a*

$$\Delta_n \le (1+o(1))c\lambda_s^2\log n.$$

Démonstration. Afin de décomposer la somme qui définit Δ_n, on introduit

$$\Delta_n(k) = \sum_{\substack{\omega,\omega'\in\Omega_n\\ \omega\sim\omega'\in\Omega_n\\ x_{s+1}=y_{s+1}\\ |\omega\cap\omega'|=k}} \mathbb{P}(E_\omega\cap E_{\omega'}) \quad \text{et} \quad \Delta_n'(k) = \sum_{\substack{\omega,\omega'\in\Omega_n\\ \omega\sim\omega'\in\Omega_n\\ x_{s+1}\neq y_{s+1}\\ |\omega\cap\omega'|=k}} \mathbb{P}(E_\omega\cap E_{\omega'})$$

de sorte que

$$\Delta_n = \sum_{k=1}^{s-1}\Delta_n(k) + \sum_{k=1}^{s-1}\Delta_n'(k).$$

Étudions chacun de ces termes séparément, en commençant par $\Delta_n(1)$ qui fournit la contribution principale, comme on va le voir.

(i) On a

$$\begin{aligned}\Delta_n(1) &= \frac{1}{s^{2s+1}} \sum_{\substack{(c\log n)^s<x_s<\cdots<x_1\\(c\log n)^s<y_s<\cdots<y_1\\x_{s+1}<(c\log n)^s\\x_1+\cdots+x_s=y_1+\cdots+y_s=n-x_{s+1}\\x_i\neq y_j \text{ pour tous indices } i \text{ et } j}} (x_1\cdots x_{s+1}y_1\cdots y_s)^{-1+1/s}\\ &\leq \frac{1}{s^{2s+1}} \sum_{1\leq x_{s+1}<(c\log n)^s} x_{s+1}^{-1+1/s} \left(\sum_{\substack{1\leq x_s<\cdots<x_1\\x_1+\cdots+x_s=n-x_{s+1}}} (x_1\cdots x_s)^{-1+1/s}\right)^2.\end{aligned}$$

Pour chaque $x_{s+1} < (c\log n)^s$, appliquons le Lemme 2.1 (iii) avec $z = n-x_{s+1} \sim n$, qui conduit à

$$\begin{aligned}\Delta_n(1) &\leq (1+o(1))\frac{1}{s^{2s+1}}(s^s\lambda_s)^2 \sum_{1\leq x_{s+1}<(c\log n)^s} x_{s+1}^{-1+1/s}\\ &\leq (1+o(1))c\lambda_s^2\log n\end{aligned}$$

quand n tend vers l'infini.

(ii) Pour $2 \leq k \leq s-1$, on a

$$\begin{aligned}\Delta_n(k) &= \frac{1}{s^{2s+2-k}} \sum_{\substack{K,K'\subset\{1,\ldots,s\}\\|K|=|K'|=k-1}} \sum_{\substack{(c\log n)^s<x_s<\cdots<x_1\\(c\log n)^s<y_s<\cdots<y_1\\1\leq x_{s+1}<(c\log n)^s\\\sum_{i\notin K}x_i=\sum_{i\notin K'}y_i=n-\left(\sum_{i\in K}x_i\right)-x_{s+1}\\x_i\neq y_j \text{ pour tous indices } i\notin K \text{ et } j\notin K'\\\{x_i \text{ for } i\in K\}=\{y_i \text{ for } i\in K'\}}} \left(\left(\prod_{i=1}^{s+1}x_i\right)\left(\prod_{i\notin K'}y_i\right)\right)^{-1+1/s}\\ &\ll \sum_{\substack{(c\log n)^s<x_s<\cdots<x_1\\(c\log n)^s<y_s<\cdots<y_k\\1\leq x_{s+1}<(c\log n)^s\\x_k+\cdots+x_s=y_k+\cdots+y_s=n-(x_1+\cdots+x_{k-1}+x_{s+1})}} (x_1\cdots x_{s+1}y_k\cdots y_s)^{-1+1/s},\end{aligned}$$

après regroupement des termes similaires. Ainsi,

$$\begin{aligned}\Delta_n(k) \ll \sum_{1\leq x_{s+1}<(c\log n)^s} x_{s+1}^{-1+1/s} &\sum_{\substack{(c\log n)^s<x_1,\ldots,x_{k-1}\\x_1+\cdots+x_{k-1}<n-x_{s+1}}} (x_1\cdots x_{k-1})^{-1+1/s}\\ &\times\left(\sum_{\substack{(c\log n)^s<x_k,\ldots,x_s\\x_k+\cdots+x_s=n-x_1-\cdots-x_{k-1}-x_{s+1}}} (x_k\cdots x_s)^{-1+1/s}\right)^2.\end{aligned}$$

On utilise tout d'abord le Lemme 2.1 (i) avec $z = n - x_1 - \cdots - x_{k-1} - x_{s+1} \geq 1$ pour majorer le dernier terme. Cela donne

$$\Delta_n(k) \ll \sum_{1\leq x_{s+1}<(c\log n)^s} x_{s+1}^{-1+1/s} \sum_{\substack{1\leq x_1,\dots,x_{k-1}\\ x_1+\cdots+x_{k-1}<n-x_{s+1}}} (x_1\cdots x_{k-1})^{-1+1/s} \times\Big(n - x_{s+1} - (x_1+\cdots+x_{k-1})\Big)^{-2(k-1)/s}$$

et en appliquant maintenant le Lemme 2.1 (ii) avec $z = n - x_{s+1} \geq 2$, on obtient

$$\begin{aligned}\Delta_n(k) &\ll \sum_{1\leq x_{s+1}<(c\log n)^s} x_{s+1}^{-1+1/s}(n-x_{s+1})^{-1/s}\log(n-x_{s+1})\\ &\ll n^{-1/s}\log n \sum_{1\leq x_{s+1}<(c\log n)^s} x_{s+1}^{-1+1/s}\\ &\ll n^{-1/s}\log^2 n.\end{aligned}$$

(iii) Finalement, pour $1 \leq k \leq s-1$, en utilisant une décomposition similaire, on a

$$\Delta'_n(k) \ll \sum_{\substack{1\leq x_s<\cdots<x_1\\ 1\leq y_s<\cdots<y_{k+1}\\ 1\leq x_{s+1},y_{k+1}<(c\log n)^s\\ x_{k+1}+\cdots+x_{s+1}=y_{k+1}+\cdots+y_{s+1}=n-(x_1+\cdots+x_k)}} (x_1\cdots x_{s+1}y_{k+1}\cdots y_{s+1})^{-1+1/s}.$$

D'où

$$\Delta'_n(k) \ll \sum_{\substack{1\leq x_1,\dots,x_k\\ x_1+\cdots+x_k<n}} (x_1\cdots x_k)^{-1+1/s} S(n;x_1,\dots,x_k)^2$$

avec

$$S(n;x_1,\dots,x_k) = \sum_{\substack{1\leq x_{k+1},\dots,x_s\\ 1\leq x_{s+1}<(c\log n)^s\\ x_{k+1}+\cdots+x_{s+1}=n-(x_1+\cdots+x_k)}} (x_{k+1}\cdots x_{s+1})^{-1+1/s}.$$

Il reste à étudier cette somme, pour laquelle on distingue deux cas :

(a) Tout d'abord, si $x_1+\cdots+x_k < n - 2(c\log n)^s$ alors

$$S(n;x_1,\dots,x_k) = \sum_{1\leq x_{s+1}<(c\log n)^s} x_{s+1}^{-1+1/s} \sum_{\substack{1\leq x_{k+1},\dots,x_s\\ x_{k+1}+\cdots+x_s=n-x_{s+1}-(x_1+\cdots+x_k)}} (x_{k+1}\cdots x_s)^{-1+1/s}$$

ce qu'on peut, grâce au Lemme 2.1 (i) appliqué avec $z = n-x_{s+1}-(x_1+\cdots+x_k) \geq 1$, majorer comme suit :

$$\begin{aligned}S(n;x_1,\dots,x_k) &\ll \sum_{1\leq x_{s+1}<(c\log n)^s} x_{s+1}^{-1+1/s}(n-(x_1+\cdots+x_k)-x_{s+1})^{-k/s}\\ &\ll (n-(x_1+\cdots+x_k))^{-k/s}\log n.\end{aligned}$$

(b) Ensuite, dans le cas $n-2(c\log n)^s \le x_1+\cdots+x_k < n$, on applique le Lemme 2.1 (i) avec $z = n-(x_1+\cdots+x_k) \ge 1$. D'où

$$\begin{aligned} S(n;x_1,\dots,x_k) &\le \sum_{\substack{1\le x_{k+1},\dots,x_{s+1}\\ x_{k+1}+\cdots+x_{s+1}=n-(x_1+\cdots+x_k)}} (x_{k+1}\cdots x_{s+1})^{-1+1/s} \\ &\ll (n-(x_1+\cdots+x_k))^{(1-k)/s} \\ &\ll 1. \end{aligned}$$

En décomposant la somme initiale et grâce aux majorations obtenues en (a) et (b), on a

$$\begin{aligned} \Delta_n'(k) &\ll \sum_{\substack{1\le x_1,\dots,x_k\\ x_1+\cdots+x_k<n-2(c\log n)^s}} (x_1\cdots x_k)^{-1+1/s} S(n;x_1,\dots,x_k)^2 \\ &\qquad + \sum_{\substack{1\le x_1,\dots,x_k\\ n-2(c\log n)^s\le x_1+\cdots+x_k<n}} (x_1\cdots x_k)^{-1+1/s} S(n;x_1,\dots,x_k)^2 \\ &\ll \log^2 n \sum_{\substack{1\le x_1,\dots,x_k\\ x_1+\cdots+x_k<n-2(c\log n)^s}} (x_1\cdots x_k)^{-1+1/s} (n-(x_1+\cdots+x_k))^{-2k/s} \\ &\qquad + \sum_{n-2(c\log n)^s\le r<n} \sum_{\substack{1\le x_1,\dots,x_k\\ x_1+\cdots+x_k=r}} (x_1\cdots x_k)^{-1+1/s}. \end{aligned}$$

On applique alors le Lemme 2.1 (ii) avec $t=k$ et $z=n$ au premier terme et le Lemme 2.1 (i) avec $t=k$ et $z=r$ pour chaque terme de la deuxième somme, ce qui conduit à

$$\begin{aligned} \Delta_n'(k) &\ll \log^2 n \sum_{\substack{1\le x_1,\dots,x_k\\ x_1+\cdots+x_k<n}} (x_1\cdots x_k)^{-1+1/s} (n-(x_1+\cdots+x_k))^{-2k/s} \\ &\qquad + \sum_{n-2(c\log n)^s\le r<n} r^{-1+k/s} \\ &\ll n^{-1/s}\log^3 n + n^{-1+k/s}(\log n)^s \\ &\ll n^{-1/s}\log^{s+1} n. \end{aligned}$$

La conclusion du lemme découle des bornes obtenues en (i), (ii) et (iii). □

En regroupant les résultats des Lemmes 2.2 et 2.3, il vient

$$\mathbb{P}(r(n,A)=0) \le \exp\left(-(1+o(1))c\lambda_s(1-2\lambda_s)\log n\right),$$

qui est le terme général d'une série convergente dès que $c\lambda_s(1-2\lambda_s) > 1$, ce qui conclut la preuve du Théorème 2.1. Comme il est dit dans [6], le facteur 2

apparaissant dans l'inégalité de Janson peut être amélioré à toute constante plus grande que 1 ; cependant, le terme de corrélation reste du même ordre de grandeur que le terme principal.

Que peut-on dire à propos d'un résultat inverse ? La première inégalité de corrélation de Janson conduit à

$$\mathbb{P}(r(n,A)=0) \geq \exp\left(-(1+o(1))c\lambda_s \log n\right), \tag{2.5}$$

qui est le terme général d'une série divergente dès que $c\lambda_s < 1$. Tout d'abord, ce qu'on a obtenu dans la preuve du Théorème 2.1 est en fait plus fort que ce qui était annoncé. En effet, $r(n,A)$ compte en fait des représentations spécifiques de n (les x_i sont deux à deux disjoints et un seul d'entre eux est plus petit que $(c\log n)^s$). Ainsi, on a démontré que, presque sûrement, on peut représenter tout entier n sous cette forme particulière. Nier cela n'est donc pas équivalent à nier la conclusion du theorème. Cependant, il n'est en fait pas difficile d'obtenir une borne du type (2.5) en prenant en compte toutes les représentations. En fait, on aimerait appliquer un théorème "inverse" de Borel-Cantelli, ce qui nécessite une indépendance entre les événements $\{r(n,A)=0\}$; malheureusement, la condition de [7] n'est pas vérifiée dans notre cas.

2.4 Preuve du Théorème 2.2

La stratégie que l'on va employer pour démontrer le Théorème 2.2 est globalement la même que pour la preuve précédente (application de la deuxième inégalité de Janson et du Lemme de Borel-Cantelli). Toutefois, cela nécessite une mise en place différente.

D'après l'hypothèse du théorème, il existe une constante

$$c > \lambda_s^{-1}$$

telle que la fonction de comptage de la suite B (fixée) vérifie

$$B(n) \geq c(1+o(1))\log n. \tag{2.6}$$

Pour tout entier n, on définit $m = m(n)$ le plus petit entier positif tel que

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

D'après (2.6) et la définition de m, on a

$$m \leq n^{(1+o(1))\frac{1+\lambda_s^{-1}/c}{2}} = o(n),$$

ce qui sera utile dans la preuve.

On représente les ensembles de s éléments distincts par $\omega = \{x_1, \ldots, x_s\}$ avec $x_1 > \cdots > x_s$. On note à nouveau $\sigma(\omega) = x_1 + \cdots + x_s$ et pour chaque n, on introduit

$$\Omega_n = \{\omega \text{ tel que } \sigma(\omega) = n - b \text{ pour un certain } b \in B,\ b < m\},$$

Si on note E_ω l'événement $\omega \subset A$, alors l'événement "n ne peut s'écrire sous la forme $n = a_1 + \cdots + a_s + b$ avec $a_1 > \cdots > a_s,\ a_i \in A,\ b \in B,\ b < m$", qu'on note F_n, peut être interprété comme

$$F_n = \bigcap_{\omega \in \Omega_n} E_\omega^c.$$

Comme dans la section précédente, nous avons besoin des deux lemmes suivants pour traiter le membre de droite de la deuxième inégalité de Janson.

Lemme 2.4. *On a*

$$\sum_{\omega \in \Omega_n} \mathbb{P}(E_\omega) = (1 + o(1)) \frac{c\lambda_s + 1}{2} \log n.$$

Démonstration. En effet, le Lemme 2.1 (iii) donne

$$\begin{aligned} \sum_{\omega \in \Omega_n} \mathbb{P}(E_\omega) &= \sum_{b<m} \frac{1}{s^s} \sum_{\substack{1 \le x_1 < \cdots < x_s \\ x_1 + \cdots + x_s = n-b}} (x_1 \cdots x_s)^{-1+1/s} \\ &= (1 + o(1)) B(m) \lambda_s \\ &= (1 + o(1)) \frac{c\lambda_s + 1}{2} \log n. \end{aligned}$$

□

Lemme 2.5. *On a*

$$\sum_{\substack{\omega \sim \omega' \\ \omega, \omega' \in \Omega_n}} \mathbb{P}(E_\omega \cap E_{\omega'}) \ll n^{-1/s} (\log n)^3.$$

Démonstration. En effet, on peut écrire

$$\sum_{\substack{\omega \sim \omega' \\ \omega, \omega' \in \Omega_n}} \mathbb{P}(E_\omega \cap E_{\omega'}) = \sum_{\substack{b \le b' < m \\ b, b' \in B}} \sum_{k=1}^{s-1} \Delta_n(k; b, b')$$

où

$$\Delta_n(k;b,b') = \sum_{\substack{\omega,\omega'\in\Omega_n \\ \sigma(\omega)=n-b,\ \sigma(\omega')=n-b' \\ |\omega\cap\omega'|=k}} \mathbb{P}(E_\omega \cap E_{\omega'}).$$

Or,

$$\Delta_n(k;b,b') \ll \sum_{\substack{1\le x_1<\cdots<x_k \\ x_1+\cdots+x_k<n-b'}} (x_1\cdots x_k)^{-1+1/s} \left(\sum_{\substack{x_{k+1},\ldots,x_s \\ x_{k+1}+\cdots+x_s=n-b-(x_1+\cdots+x_k)}} (x_{k+1}\cdots x_s)^{-1+1/s} \right) \times \left(\sum_{\substack{y_{k+1},\ldots,y_s \\ y_{k+1}+\cdots+y_s=n-b'-(x_1+\cdots+x_k)}} (y_{k+1}\cdots y_s)^{-1+1/s} \right).$$

D'après le Lemme 2.1 (i), pour $\zeta = b$ ou b', on obtient

$$\begin{aligned} \sum_{\substack{1\le x_{k+1},\ldots,x_s \\ x_{k+1}+\cdots+x_s=n-\zeta-(x_1+\cdots+x_k)}} (x_{k+1}\cdots x_s)^{-1+1/s} &\ll (n-\zeta-(x_1+\cdots+x_k))^{-k/s} \\ &\ll (n-b'-(x_1+\cdots+x_k))^{-k/s}. \end{aligned}$$

Ainsi, en appliquant cette majoration, puis le Lemme 2.1 (ii), il vient

$$\begin{aligned} \Delta_n(k;b,b') &\ll \sum_{\substack{1\le x_1<\cdots<x_k \\ x_1+\cdots+x_k<n-b'}} (x_1\cdots x_k)^{-1+1/s}(n-b'-(x_1+\cdots+x_k))^{-2k/s} \\ &\ll (n-b')^{-1/s}\log(n-b'). \end{aligned}$$

Finalement, en sommant toutes les contributions, on a

$$\sum_{\substack{\omega\sim\omega' \\ \omega,\omega'\in\Omega_n}} \mathbb{P}(E_\omega\cap E_{\omega'}) \ll \sum_{\substack{b\le b'<m \\ b,b'\in B}} (n-b')^{-1/s}\log(n-b') \ll B(m)^2 n^{-1/s}\log n.$$

et $B(m) \ll \log n$ permet de conclure la preuve du lemme. □

Venons-en maintenant à la démonstration du théorème. Par la deuxième inégalité de Janson,

$$\mathbb{P}(F_n) \le \prod_{\omega\in\Omega_n}(1-\mathbb{P}(E_\omega)) \exp\left(2 \sum_{\substack{\omega\sim\omega' \\ \omega,\omega'\in\Omega_n}} \mathbb{P}(E_\omega\cap E_{\omega'}) \right)$$

ce qui donne, en utilisant $\log(1-x) < -x$ (vrai pour $x > 0$),

$$\log \mathbb{P}(F_n) \leq -\sum_{\omega \in \Omega_n} \mathbb{P}(E_\omega) + 2 \sum_{\substack{\omega \sim \omega' \\ \omega, \omega' \in \Omega_n}} \mathbb{P}(E_\omega \cap E_{\omega'}). \tag{2.7}$$

À l'aide des Lemmes 2.4 et 2.5, (2.7) devient

$$\log \mathbb{P}(F_n) \leq -(1+o(1)) \frac{c\lambda_s + 1}{2} \log n,$$

d'où

$$\mathbb{P}(F_n) \leq n^{-(1+o(1))\frac{c\lambda_s+1}{2}}.$$

Comme $c > \lambda_s^{-1}$, $(c\lambda_s + 1)/2 > 1$ et la somme $\sum_n \mathbb{P}(F_n)$ est finie. Le lemme de Borel-Cantelli implique que, presque sûrement, seul un nombre fini d'événements F_n peut arriver.

2.5 Preuve du Théorème 2.3

La stratégie de la preuve n'est cette fois plus du tout la même. On utilise les mêmes notations que dans la section précédente, si ce n'est que maintenant

$$\Omega_n = \{\omega \text{ tel que } \sigma(\omega) = n - b \text{ pour un certain } b \in B\}.$$

On définit l'événement F_n : "n ne peut s'écrire sous la forme $n = x_1 + \cdots + x_s + b$ avec $x_1, \ldots, x_s \in A$, $x_s < \cdots < x_1$ et $b \in B$." En d'autres mots,

$$F_n = \bigcap_{\omega \in \Omega_n} E_\omega^c.$$

L'hypothèse du Théorème 2.3 est équivalente à

$$\liminf_{n \to +\infty} \frac{B(n)}{\log n} = c$$

pour $c < \lambda_s^{-1}$. Cela implique l'existence d'une suite $(N_i)_{i \in \mathbb{N}}$ d'entiers tels que

$$B(N_i) = c(1+o(1)) \log N_i. \tag{2.8}$$

Dans la suite, si N est un entier, nous dirons qu'un entier n est *bon (pour N)* si $N/2 \leq n \leq N$ et

$$|n - b| > (\log N)^{4s}$$

pour tout $b \in B$. Dans le cas contraire, nous dirons que n est *mauvais (pour N)*.

On considère la variable aléatoire ($\mathbb{I}$ est la fonction indicatrice d'un événement)

$$X_N = \sum_{\substack{N/2 \le n \le N \\ n \text{ est bon}}} \mathbb{I}(F_n).$$

On note

$$\mu_N = \mathbb{E}(X_N) \quad \text{et} \quad \sigma_N^2 = \mathbb{V}(X_N).$$

L'objectif est de montrer que

$$\lim_{i \to +\infty} \mu_{N_i} = +\infty \tag{2.9}$$

et

$$\sigma_{N_i}^2 \ll \frac{\mu_{N_i}^2}{\log N_i}. \tag{2.10}$$

Si (2.10) est vraie, en utilisant l'inégalité de Bienaymé-Tchebychev (2.3), on obtient

$$\mathbb{P}\left(X_{N_i} < \frac{\mu_{N_i}}{2}\right) < \frac{4\sigma_{N_i}^2}{\mu_{N_i}^2} \ll \frac{1}{\log N_i}. \tag{2.11}$$

Le Théorème 2.3 découle alors immédiatement de (2.11) et (2.9).

Il nous reste finalement à démontrer ces deux points. Soit donc N un élément de la suite $(N_i)_{i \in \mathbb{N}}$.

2.5.1 Estimation de μ_N

Proposition 2.1. *On a*

$$\mu_N \ge N^{(1-c\lambda_s)(1+o(1))}.$$

Démonstration. Par définition,

$$\mu_N = \sum_{\substack{N/2 \le n \le N \\ n \text{ bon}}} \mathbb{P}(F_n). \tag{2.12}$$

Soit n un entier bon pour N. La première inégalité de Janson (Théorème 2.5 (i)) donne

$$\mathbb{P}(F_n) \ge \prod_{\omega \in \Omega_n} \left(1 - \mathbb{P}(E_\omega)\right).$$

Puisque $\log(1-x) = -x + O(x^2)$, on obtient

$$\log(\mathbb{P}(F_n)) \ge -\sum_{\omega \in \Omega_n} \mathbb{P}(E_\omega) + O\left(\sum_{\omega \in \Omega_n} \mathbb{P}(E_\omega)^2\right). \tag{2.13}$$

D'une part, comme n est bon,

$$\begin{aligned}\sum_{\omega\in\Omega_n}\mathbb{P}(E_\omega) &= \frac{1}{s^s}\sum_{\substack{b<n-(\log N)^{4s}\\ b\in B}}\sum_{\substack{1\le x_s<\cdots<x_1\\ x_1+\cdots+x_s=n-b}}(x_1\cdots x_s)^{-1+1/s} \qquad (2.14)\\ &= \frac{1}{s^s}\sum_{\substack{b<n-(\log N)^{4s}\\ b\in B}} s^s\lambda_s(1+o(1))\\ &\le \lambda_s(1+o(1))B(N)\\ &\le c\lambda_s(1+o(1))\log N.\end{aligned}$$

D'autre part,

$$\begin{aligned}\sum_{\omega\in\Omega_n}\mathbb{P}(E_\omega)^2 &= \sum_{\substack{b<n-(\log N)^{4s}\\ b\in B}}\left(\frac{1}{s^s}\right)^2\sum_{\substack{1\le x_s<\cdots<x_1\\ x_1+\cdots+x_s=n-b}}(x_1\cdots x_s)^{-2+2/s}\\ &\ll \sum_{\substack{b<n-(\log N)^{4s}\\ b\in B}}(n-b)^{-2+2/s}\sum_{\substack{1\le x_s<\cdots<x_2\\ x_2+\cdots+x_s\le n-b}}(x_2\cdots x_s)^{-2+2/s}\end{aligned}$$

en remarquant que $x_1\ge(n-b)/s$ dans chaque terme de la somme. En se servant du fait que n est bon, on poursuit la majoration comme suit :

$$\begin{aligned}\sum_{\omega\in\Omega_n}\mathbb{P}(E_\omega)^2 &\ll \sum_{\substack{b<n-(\log N)^{4s}\\ b\in B}}(n-b)^{-2+2/s}\left(\sum_{x=1}^{n-b}x^{-2+2/s}\right)^{s-1}\\ &\ll \sum_{\substack{b<n\\ b\in B}}(\log^{4s}N)^{-2+2/s}\left(\sum_{x=1}^{n-b}x^{-1}\right)^{s-1}\\ &\ll \sum_{\substack{b<n\\ b\in B}}(\log N)^{-8s+8}(\log N)^{s-1}\\ &\ll (\log N)^{-7s+7}B(N)\\ &\ll (\log N)^{-6}.\end{aligned}$$

Ainsi, (2.13) et (2.14) implique que

$$\mathbb{P}(F_n)\ge N^{-c\lambda_s(1+o(1))} \qquad (2.15)$$

lorsque n est bon.

Majorons maintenant le nombre M_n de mauvais n.

$$\begin{aligned} M_n &= |\{N/2 \le n \le N : \ n \text{ mauvais}\}| \\ &= |\{N/2 \le n \le N : \ |n-b| < (\log N)^{4s} \text{ pour un certain } b \in B\}| \\ &\le \sum_{b<N} 2(\log N)^{4s} \\ &\ll (\log N)^{4s+1}. \end{aligned}$$

Finalement, en combinant cela avec les équations (2.12) et (2.15), il vient

$$\begin{aligned} \mu_N &= \sum_{\substack{N/2\le n\le N \\ n \text{ bon}}} N^{-c\lambda_s(1+o(1))} \\ &\ge \sum_{N/2\le n\le N} N^{-c\lambda_s(1+o(1))} - \sum_{\substack{N/2\le n\le N \\ n \text{ mauvais}}} N^{-c\lambda_s(1+o(1))} \\ &\ge N^{(1+o(1))(1-c\lambda_s)} - O\left((\log N)^{4s+1}\right) \\ &\ge N^{(1+o(1))(1-c\lambda_s)} \end{aligned}$$

car $1 - c\lambda_s > 0$. □

Étudions maintenant la variance, pour démontrer (2.10).

2.5.2 Estimation de σ_N^2

Rappelons qu'étant donné un ensemble B, son *ensemble de différence* $B - B$ est défini par

$$B - B = \{b - b' \text{ avec } b, b' \in B\}.$$

Pour commencer,

$$\sigma_N^2 = 2 \sum_{\substack{N/2\le n<m\le N \\ n,m \text{ bons}}} \Big(\mathbb{P}(F_n \cap F_m) - \mathbb{P}(F_n)\mathbb{P}(F_m)\Big) + \sum_{\substack{N/2\le n\le N \\ n \text{ bon}}} \Big(\mathbb{P}(F_n) - \mathbb{P}^2(F_n)\Big).$$

Le deuxième terme ne pose pas de soucis, puisqu'il suffit de le majorer par μ_N. Les prochains lemmes vont nous permettre de mieux appréhender la première somme.

Lemme 2.6. *Soit $B_N = \{b \le N \text{ avec } b \in B\}$. Soient n et m deux entiers positifs tels que $n < m \le N$ et $m - n \notin B_N - B_N$, alors*

$$\mathbb{P}(F_n \cap F_m) \le \mathbb{P}(F_n)\mathbb{P}(F_m) \exp\left(2 \sum_{\substack{\omega,\omega' \in \Omega_n \cup \Omega_m \\ \omega \sim \omega'}} \mathbb{P}(E_\omega \cap E_{\omega'})\right).$$

Démonstration. On remarque que

$$F_n \cap F_m = \bigcap_{\omega \in \Omega_n \cup \Omega_m} E_\omega^c$$

et que la condition $m - n \notin B_N - B_N$ implique que $\Omega_n \cap \Omega_m = \emptyset$. La deuxième inégalité de Janson appliquée à $\Omega = \Omega_n \cup \Omega_m$ donne

$$\begin{aligned}
\mathbb{P}(F_n \cap F_m) &\leq \prod_{\omega \in \Omega_n \cup \Omega_m} \mathbb{P}(E_\omega^c) \exp\left(2 \sum_{\substack{\omega, \omega' \in \Omega_n \cup \Omega_m \\ \omega \sim \omega'}} \mathbb{P}(E_\omega \cap E_{\omega'}) \right) \\
&= \prod_{\omega \in \Omega_n} \mathbb{P}(E_\omega^c) \prod_{\omega \in \Omega_m} \mathbb{P}(E_\omega^c) \exp\left(2 \sum_{\substack{\omega, \omega' \in \Omega_n \cup \Omega_m \\ \omega \sim \omega'}} \mathbb{P}(E_\omega \cap E_{\omega'}) \right) \\
&\leq \mathbb{P}(F_n)\mathbb{P}(F_m) \exp\left(2 \sum_{\substack{\omega, \omega' \in \Omega_n \cup \Omega_m \\ \omega \sim \omega'}} \mathbb{P}(E_\omega \cap E_{\omega'}) \right)
\end{aligned}$$

où la dernière majoration provient de la première inégalité de Janson appliquée à la fois à Ω_n et Ω_m. Le lemme est alors prouvé. □

On s'intéresse maintenant à l'argument de l'exponentielle.

Lemme 2.7. *Soient N, n, m des entiers. Si n et m sont bons pour N, alors*

$$\sum_{\substack{\omega \in \Omega_n, \omega' \in \Omega_m \\ \omega \sim \omega'}} \mathbb{P}(E_\omega \cap E_{\omega'}) \ll \frac{1}{\log N}.$$

Démonstration. On peut écrire

$$\sum_{\substack{\omega \in \Omega_n, \omega' \in \Omega_m \\ \omega \sim \omega'}} \mathbb{P}(E_\omega \cap E_{\omega'}) = \sum_{\substack{1 \leq b < n \\ 1 \leq b' < m \\ b, b' \in B}} \sum_{k=1}^{s-1} \Delta_{n,m}(k; b, b')$$

où, pour $k \geq 1$,

$$\Delta_{n,m}(k; b, b') = \sum_{\substack{\omega \in \Omega_n, \omega' \in \Omega_m \\ \sigma(\omega) = n-b,\ \sigma(\omega') = m-b' \\ |\omega \cap \omega'| = k}} P(E_\omega \cap E_{\omega'}).$$

Supposons que $n-b \leq m-b'$. Alors,

$$\Delta_{n,m}(k;b,b') \ll \sum_{\substack{1\leq x_1,\dots,x_k \\ x_1+\cdots+x_k<n-b}} (x_1\cdots x_k)^{-1+\frac{1}{s}} \left(\sum_{\substack{1\leq x_{k+1},\dots,x_s \\ x_{k+1}+\cdots+x_s=n-b-(x_1+\cdots+x_k)}} (x_{k+1}\cdots x_s)^{-1+\frac{1}{s}} \right) \times \left(\sum_{\substack{1\leq y_{k+1},\dots,y_s \\ y_{k+1}+\cdots+y_s=m-b'-(x_1+\cdots+x_k)}} (y_{k+1}\cdots y_s)^{-1+1/s} \right).$$

Le Lemme 2.1 (i) appliqué deux fois montre que

$$\begin{aligned}
\Delta_{n,m}(k;b,b') &\ll \sum_{\substack{1\leq x_1,\dots,x_k \\ x_1+\cdots+x_k<n-b}} (x_1\cdots x_k)^{-1+\frac{1}{s}} (n-b-(x_1+\cdots+x_k))^{-\frac{k}{s}} \\
&\qquad \times (m-b'-(x_1+\cdots+x_k))^{-\frac{k}{s}} \\
&\ll \sum_{\substack{1\leq x_1,\dots,x_k \\ x_1+\cdots+x_k<n-b}} (x_1\cdots x_k)^{-1+1/s} (n-b-(x_1+\cdots+x_k))^{-2k/s} \\
&\ll (n-b)^{-1/s}\log(n-b) \\
&\ll \frac{1}{\log^3 N}.
\end{aligned}$$

car $(\log N)^{4s} \leq n-b \leq N$.

Si $m-b' < n-b$ on procède de la même façon. Ainsi,

$$\begin{aligned}
\sum_{\substack{\omega\in\Omega_n,\omega'\in\Omega_m \\ \omega\sim\omega'}} \mathbb{P}(E_\omega \cap E_{\omega'}) &\ll \sum_{\substack{1\leq b<n \\ b\in B}} \sum_{\substack{1\leq b'<m \\ b'\in B}} \frac{1}{\log^3 N} \\
&\ll \frac{(B(N))^2}{\log^3 N} \\
&\ll \frac{1}{\log N},
\end{aligned}$$

d'où le résultat. □

Corollaire 2.1. *Soient N, n et m des entiers. Si n et m sont bons pour N et $m-n \notin B_N - B_N$ alors*

$$\mathbb{P}(F_n \cap F_m) - \mathbb{P}(F_n)\mathbb{P}(F_m) \ll \frac{1}{\log N}\mathbb{P}(F_n)\mathbb{P}(F_m).$$

Démonstration. D'après le Lemme 2.6,

$$\mathbb{P}(F_n \cap F_m) - \mathbb{P}(F_n)\mathbb{P}(F_m) \leq \mathbb{P}(F_n)\mathbb{P}(F_m) \left(\exp\left(2 \sum_{\substack{\omega,\omega' \in \Omega_n \cup \Omega_m \\ \omega \sim \omega'}} \mathbb{P}(E_\omega \cap E_{\omega'}) \right) - 1 \right).$$

Or,

$$\begin{aligned} \sum_{\substack{\omega,\omega' \in \Omega_n \cup \Omega_m \\ \omega \sim \omega'}} \mathbb{P}(E_\omega \cap E_{\omega'}) &= \sum_{\substack{\omega,\omega' \in \Omega_n \\ \omega \sim \omega'}} \mathbb{P}(E_\omega \cap E_{\omega'}) + \sum_{\substack{\omega,\omega' \in \Omega_m \\ \omega \sim \omega'}} \mathbb{P}(E_\omega \cap E_{\omega'}) \\ &\quad + \sum_{\substack{\omega \in \Omega_n, \omega' \in \Omega_m \\ \omega \sim \omega'}} \mathbb{P}(E_\omega \cap E_{\omega'}). \end{aligned}$$

On termine la preuve en appliquant le Lemme 2.7 aux trois sommes et en utilisant l'équivalent $e^x - 1 \sim x$ quand x tend vers 0. □

Proposition 2.2. *On a finalement*

$$\sigma_N^2 \ll \frac{\mu_N^2}{\log N}.$$

Démonstration. Pour commencer,

$$\sigma_N^2 = 2 \sum_{\substack{N/2 \leq n < m \leq N \\ n,m \text{ bons}}} \Big(\mathbb{P}(F_n \cap F_m) - \mathbb{P}(F_n)\mathbb{P}(F_m) \Big) + \sum_{\substack{N/2 \leq n \leq N \\ n \text{ bon}}} \Big(\mathbb{P}(F_n) - \mathbb{P}^2(F_n) \Big).$$

On décompose cette expression de la façon suivante

$$\sigma_N^2 = 2\Sigma_1 + 2\Sigma_2 + \Sigma_3,$$

où

$$\begin{aligned} \Sigma_1 &= \sum_{\substack{N/2 < n < m \leq N \\ n-m \notin B_N - B_N \\ n,m \text{ bons}}} \Big(\mathbb{P}(F_n \cap F_m) - \mathbb{P}(F_n)\mathbb{P}(F_m) \Big), \\ \Sigma_2 &= \sum_{\substack{N/2 \leq n < m \leq N \\ n-m \in B_N - B_N \\ n,m \text{ bons}}} \Big(\mathbb{P}(F_n \cap F_m) - \mathbb{P}(F_n)\mathbb{P}(F_m) \Big), \\ \Sigma_3 &= \sum_{\substack{N/2 \leq n \leq N \\ n \text{ bon}}} \Big(\mathbb{P}(F_n) - \mathbb{P}^2(F_n) \Big). \end{aligned}$$

Il est clair que

$$\Sigma_3 \leq \mu_N.$$

Pour borner Σ_2, on utilise la majoration triviale

$$\mathbb{P}(F_n \cap F_m) - \mathbb{P}(F_n)\mathbb{P}(F_m) \leq \mathbb{P}(F_m),$$

ce qui conduit à

$$\begin{aligned}
\Sigma_2 &\leq \sum_{\substack{N/2\leq m<N \\ m \text{ bon}}} \mathbb{P}(F_m)|\{N/2 \leq n \leq N \text{ tel que } n \in B_N - B_N + m\}| \\
&\leq |B_N - B_N| \sum_{\substack{N/2\leq m<N \\ m \text{ bon}}} \mathbb{P}(F_m) \\
&\ll |B_N|^2 \sum_{\substack{N/2\leq m<N \\ m \text{ bon}}} \mathbb{P}(F_m) \\
&\ll \log^2 N \ \mu_N.
\end{aligned}$$

Finalement, par le Corollaire 2.1, on obtient

$$\Sigma_1 \ll \frac{1}{\log N} \sum_{\substack{N/2<n<m\leq N \\ n-m\notin B_N-B_N \\ n,m \text{ bons}}} \mathbb{P}(F_n)\mathbb{P}(F_m) \leq \frac{1}{\log N}\Big(\sum_{\substack{N/2\leq n\leq N \\ n \text{ bon}}} \mathbb{P}(F_n)\Big)^2 = \frac{\mu_N^2}{\log N}.$$

Lorsqu'on ajoute les trois contributions de Σ_1, Σ_2 et Σ_3, on a

$$\sigma_N^2 \ll \frac{\mu_N^2}{\log N} + \log^2 N \ \mu_N + \mu_N \ll \mu_N^2 \left(\frac{1}{\log N} + \frac{\log^2 N}{\mu_N}\right). \tag{2.16}$$

Posons alors

$$\varepsilon = \frac{1 - c\lambda_s}{2} > 0,$$

la Proposition 2.1 permet de minorer μ_N

$$\mu_N \geq N^{2\varepsilon + o(1)} \gg \log^3 N. \tag{2.17}$$

On obtient le résultat en injectant (2.17) dans (2.16). □

2.6 Le cas limite

Voyons maintenant ce qu'il se passe au niveau du seuil, c'est-à-dire pour les suites B satisfaisant

$$\liminf_{n\to\infty} \frac{B(n)}{\log n} = \lambda_s^{-1}.$$

Dans cette section, nous allons exhiber de telles suites vérifiant la conclusion du Théorème 2.2, et d'autres vérifiant la conclusion du Théorème 2.3. Étant donné une suite u à valeurs dans $[0,1]$, notons qu'on peut facilement, à l'aide de la méthode gloutonne, construire une suite B de fonction de comptage $B(n) = u_n$.

Nous allons avoir besoin du raffinement du Lemme 2.1 (iii),

$$\sum_{\substack{1\leq x_s<\cdots<x_1\\ x_1+\cdots+x_s=n}} (x_1\cdots x_s)^{-1+1/s} = s^s\lambda_s + O(n^{-1/(s+1)}) \tag{2.18}$$

dont voici une idée de preuve. Posant $g(n) = n^{1/(s+1)}$, on sépare la somme sur x_s au niveau de $g(n)$. Pour la somme où $x_s \geq g(n)$, on reconnaît une somme de Riemann pour l'intégrale $\int\cdots\int (t_1\ldots t_s)^{-1+1/s} dt_1\ldots dt_s$ sur la zone de l'hyperplan $t_1+\cdots+t_s = 1$ délimitée par $g(n)/n < t_s < \cdots < t_1 \leq 1$; l'erreur dans l'approximation de l'intégrale par la somme de Riemann est un $O(1/g(n))$. L'erreur dans la troncature de la somme (cf. la preuve du Lemme 2.1 (iii)) est un $O((g(n)/n)^{1/s})$, il en est donc de même de l'erreur de la troncature de l'intégrale. On obtient donc globalement une erreur en $O(n^{-1/(s+1)})$, ce qui est suffisant pour ce qu'on veut montrer. En fait, en regardant de plus près ce qu'il se passe autour de 0 et en intégrant l'erreur, on peut obtenir un terme d'erreur en $O(n^{-1/s})$.

Considérons maintenant par exemple la suite B définie par la fonction de comptage

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

On peut suivre la preuve du Théorème 2.2 (nous devons cependant faire attention, et poser $m = n/2$ dans ce cas). L'équation (2.18) conduit à

$$\begin{aligned}\sum_{\omega\in\Omega_n}\mathbb{P}(E_\omega) &= \frac{1}{s^s}\sum_{b<n/2}\sum_{\substack{1\leq x_s<\cdots<x_1\\ x_1+\cdots+x_s=n}} (x_1\cdots x_s)^{-1+1/s}\\ &= B(n/2)\left(\lambda_s + O(n^{-1/(s+1)})\right)\\ &= \log n + 2\log\log n + O(1).\end{aligned}$$

En suivant le même raisonnement que dans la section 2.4, on obtient

$$\mathbb{P}(F_n) \leq e^{-(\log n + 2\log\log n + O(1))} \ll \frac{1}{n\log^2 n}.$$

Ainsi, $\sum_n \mathbb{P}(F_n) < \infty$ et on peut appliquer le Lemme de Borel Cantelli pour conclure que la suite B est, presque sûrement, un complément additif de sA.

Pour le cas contraire, considérons par exemple une suite B définie par la fonction de comptage

$$B(n) = \left\lfloor \lambda_s^{-1}\log n - t(n) \right\rfloor,$$

où $t(n)$ est une fonction croissante vérifiant $t(n) = o(\log n)$. On peut suivre la preuve du Théorème 2.3, à la différence que maintenant, l'exposant $2\epsilon + o(1)$ de (2.17) est $2\epsilon_N \sim \lambda_s t(N)/\log N$. Ainsi, nous pouvons choisir pour $t(n)$ toute fonction telle que $\mu_N \gg N^{\epsilon_N} \gg \log^3 N$. Par exemple,

$$t(N) = 4\lambda_s^{-1} \frac{\log\log N}{\log N}$$

vérifie bien cela. La suite B satisfait donc bien la conclusion du Théorème 2.3.

Bibliographie

[1] N. Alon and Spencer, *The probabilistic method*, 3rd edition, Wiley-Interscience, 2008.

[2] R. B. Boppona and J. H. Spencer, *A useful elementary correlation inequality*, J. Comb. Th. Ser. A 50 (1989), 305–307.

[3] J. Cilleruelo, *Sidon bases*, preprint

[4] J. Cilleruelo and J-M Deshouillers, *Gaps in sumsets of s pseudo s-th power sequences*, preprint.

[5] J-M Deshouillers, F. Hennecart and B. Landreau, *Sums of powers : an arithmetic refinement to the probabilistic model of Erdős and Rényi*, Acta Arith. 85 (1998), 13-33.

[6] J.-M. Deshouillers and M. Iosifescu, *Sommes de s+1 pseudo-puissances s-ièmes*, Rev. Roumaine Math. Pures Appl. 45 (2000), no. 3, 427–435 (2001).

[7] P. Erdős and A. Rényi, *On Cantor's series with convergent* $\sum 1/q_n$, Ann. Univ. Sci. Budapest. Eötvös. sect. math. 2 (1959), 93–109.

[8] P. Erdős and A. Rényi, *Additive properties of random sequences of positive integers*, Acta Arith. 6 (1960), 83–110.

[9] J. H. Goguel, *Über Summen von zufälligen Folgen natürlischen Zahlen*, J. Reine Angew. Math. 278/279 (1975), 63–77.

[10] S. Janson, T. Łuczak and A. Ruciński, *An exponential bound for the probability of nonexistence of a specified subgraph in a random graph*, Random Graphs 87 (Poznań, 1987), Wiley, 73–87.

[11] M. Kubina and M. C. Wunderlich, *Extending Waring's conjecture to 471,600,000*, Math. Comp., vol. 55, 1990, p. 815-820.

[12] B. Landreau, *Étude probabiliste des sommes de s puissances s-ièmes*, Compositio Math. 99 (1995), 1–31.

[13] R.C.Vaughan, *The Hardy-Littlewood method*, Cambridge University Press, n°125 (1997).

Chapitre 3

Ensembles k-libres

Ce chapitre reprend, à peu de choses près et en français, le texte d'un article [4] accepté pour publication dans The Electronic Journal of Combinatorics.

3.1 Introduction

Dans ce chapitre, nous nous intéressons aux ensembles k-libres, que nous pouvons définir dans un monoïde quelconque, même si nous les étudierons ici uniquement dans le cas des entiers naturels et des anneaux $\mathbb{Z}/n\mathbb{Z}$. Tout au long de l'étude, k et n seront des entiers naturels différents de 0. Lorsque nous dirons qu'un ensemble est optimal, cela signifiera qu'il est maximal au sens de la taille (cardinal), à ne pas confondre avec maximal au sens de l'inclusion.

Définition 3.1. Un ensemble A d'un monoïde M est dit k-libre si et seulement si $kx \neq y$ pour tout x, y dans A.

Un ensemble 1-libre étant nécessairement vide, on considèrera maintenant $k \geq 2$. Au-delà de leur propre intérêt, ces ensembles apparaissent naturellement dans l'étude des ensembles k-Sidon ("k-fold Sidon set" en anglais).

Définition 3.2. $A \subset \mathbb{Z}$ est un un ensemble k-Sidon s'il n'admet que les solutions triviales aux équations de la forme $c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 = 0$ où $0 \leqslant |c_i| \leqslant k$, avec les c_i dans $\mathbb{Z}$ et $c_1 + c_2 + c_3 + c_4 = 0$.

Dans cette définition, quitte à réordonner les c_i, on considère que les solutions sont "triviales" dans les cas suivants :

(i) $\{x, x, x, x\}$ est toujours une solution triviale,

(ii) si $c_1 = c_2 = -c_3 = -c_4$, $\{x, y, y, x\}$ est une solution triviale,

(iii) si $c_1 = -c_3$ et $c_2 = -c_4$, $\{x, y, x, y\}$ est une solution triviale.

Ces ensembles ont été introduits par Lazebnik et Verstraëte (voir [3]) dans un travail sur le nombre de Turán généralisé. Remarquons pour commencer qu'un ensemble 1-Sidon est un ensemble de Sidon au sens usuel ($x_1+x_2 = x_3+x_4$ n'admet que les solutions triviales). Si on note $D^*(A) = \{a_1 - a_2, a_1 \neq a_2 \in A\}$, l'ensemble des différences de A privé de 0, un 2-Sidon A est un ensemble de Sidon pour lequel $D^*(A)$ est 2-libre (ou sans double). Plus généralement, si A est un ensemble k-Sidon, $D^*(A)$ est k'-libre, pour tout $1 < k' \leq k$. Bien que cette propriété ne soit pas suffisante en générale pour que A soit un ensemble k-Sidon (pour $k = 3$ par exemple, l'équation $3x_1 = 2x_3 + x_4$ n'admet également que des solutions triviales), c'est en utilisant seulement celle-ci que Cilleruelo et Timons ont prouvé dans [2] que pour tout entier $k \geqslant 1$, un ensemble k-Sidon $A \subset [\![0, n]\!]$ a au plus $(n/k)^{1/2} + O((nk)^{1/4})$ éléments.

On sait seulement que le terme principal $(n/k)^{1/2}$ est optimal pour $k = 1$. En effet, les ensembles de Sidon ont été très largement étudiés (cf. [5]), et on connaît en particulier trois constructions d'ensembles de Sidon maximaux au sens de la taille dans $\mathbb{Z}/n\mathbb{Z}$ pour certains n. Bose et Chowla ont prouvé dans [1] l'existence d'un ensemble de Sidon de cardinal $q + 1$ dans $\mathbb{Z}/\left(q^2 + q + 1\right)\mathbb{Z}$ (ensembles de Singer, voir aussi [7]) et de cardinal q dans $\mathbb{Z}/\left(q^2 - 1\right)\mathbb{Z}$ (ensembles de Bose) où q est une puissance d'un nombre premier. La troisième construction optimale connue a été donnée par Ruzsa dans [6] pour $\mathbb{Z}/\left(p^2 - p\right)\mathbb{Z}$ lorsque p est premier. Pour $k = 2$, si $n = 2^{2^t+1} + 2^t + 1$ avec t un entier positif, on peut extraire (voir [3]) d'un ensemble de Singer un ensemble 2-Sidon dans $\mathbb{Z}/n\mathbb{Z}$ de taille

$$|A| \geqslant \frac{n^{1/2}}{2} - 3.$$

Pour $k \geqslant 3$, on ne sait même pas s'il existe une constante $c_k > 0$ telle que pour tout entier $n \geqslant 1$, il existe un ensemble k-Sidon $A \subset [\![0, n]\!]$ vérifiant $|A| \geqslant c_k n^{1/2}$.

C'est donc en s'intéressant à ces ensembles que nous avons été amenés à étudier les ensembles k-libres. On parlera dans la section 3.2 du cas des entiers naturels, où des résultats étaient déjà connus. Mais pour ces différents problèmes où l'on cherche à optimiser la taille d'un ensemble sous des contraintes arithmétiques, il est important et utile d'étudier le cas des ensembles modulaires $\mathbb{Z}/n\mathbb{Z}$. C'est dans cette optique que l'on va s'intéresser aux ensembles k-libres optimaux dans $\mathbb{Z}/n\mathbb{Z}$, ce dont nous parlerons dans la section 3.3.

3.2 Les ensembles k-libres dans $\mathbb{N}$

Nous appellerons désormais ensemble sans double les ensembles 2-libres. On note

$$r_k(n) = \max\left\{|A|, A \subset [\![1, n]\!], A \text{ ensemble } k\text{-libre}\right\}.$$

Nous verrons à la fin de la section 3.2.1 que rien ne change si on regarde les ensembles k-libres optimaux dans $\mathbb{N}$ tout entier, c'est-à-dire que l'on considère des ensembles infinis cette fois-ci. En effet, les ensembles k-libres optimaux qu'on va construire dans $[\![1, n]\!]$ s'étendent naturellement aux entiers, et la densité maximale obtenue sera donc préservée.

Le premier résultat a été obtenu par Wang en 1989 dans [9] où il a traité le cas des ensembles sans double :

Théorème 3.1 (Wang)**.** *Si n s'écrit $n = a_q 4^q + a_{q-1}4^{q-1} + \cdots + a_1 4 + a_0$ en base 4, alors*

$$r_2(n) = \frac{2n}{3} + \frac{1}{3}\sum_{k=0}^{q} a_k - d$$

où d est le nombre de a_i égaux à 2 ou 3. Cela donne en particulier le résultat asymptotique

$$r_2(n) = \frac{2n}{3} + O(\log_4 n).$$

Wakeham et Wood, dans [8], se sont intéressés à une généralisation des ensembles k-libres, les ensembles $\{a, b\}$-multiplicatifs ($ax \neq by$ pour tous $x, y \in A$), pour lesquels ils ont démontré le résultat suivant :

Théorème 3.2 (Wakeham, Wood)**.** *Soient $b > a \geq 1$ et $g = \mathrm{pgcd}(a, b)$, alors un ensemble $\{a, b\}$-multiplicatif optimal dans $[\![1, n]\!]$ a une densité $\frac{b}{b+g}$.*

Remarque. Le cas des ensembles sans double correspondant à $a = 1$, $b = 2$ et $g = 1$, on retrouve bien le résultat du Théorème 3.1. De plus, le cas des ensembles k-libres est entièrement traité par le théorème précédent ($a = 1$ et $b = k$), et on trouve dans ce cas une densité $\frac{k}{k+1}$ pour un ensemble k-libre optimal. On constate que c'est dans le cas des ensembles sans double ($k = 2$) que les ensembles optimaux sont les moins denses.

Dans la section qui suit, nous allons donner une nouvelle preuve de l'égalité

$$r_k(n) = \frac{k}{k+1}n + O(\log_k^2(n)).$$

Cela va nous permettre d'obtenir une vision adéquate pour traiter ensuite le cas des ensembles modulaires, mais aussi pour donner la taille minimale d'un ensemble k-libre maximal au sens de l'inclusion dans la section 3.2.2.

En effet, on définit

$$\tilde{R}_k(n) = \min\left\{|A|, A \subset [\![1, n]\!] \text{ ensemble } k\text{-libre maximal au sens de l'inclusion}\right\}.$$

Et nous démontrerons le résultat suivant :

Théorème 3.3.

$$\tilde{R}_k(n) = \frac{k^2}{k^2+k+1}n + O(\log_k^2(n)).$$

3.2.1 Ensembles k-libres optimaux

On définit $\mathcal{O}^k(x) := \{k^j x, j \in \mathbb{N}\}$, ce qu'on appellera l'orbite de x (sous-entendu par la multiplication par k). Afin d'étudier les ensembles k-libres dans $[\![1,n]\!]$, on va partitionner les entiers entre 1 et n comme suit :

$$[\![1,n]\!] = \bigsqcup_{i \not\equiv 0 \pmod{k}} \left(\mathcal{O}^k(i) \bigcap [\![1,n]\!]\right).$$

L'intérêt de cette partition est que les orbites que l'on considère ici sont indépendantes pour notre problème au sens où si x appartient à $\mathcal{O}^k(i)$, kx aussi. Ainsi, on se ramène à voir un ensemble k-libre comme un ensemble qui ne possède pas deux éléments consécutifs au sein de chacune de ces orbites.

Maintenant, pour une orbite donnée $\mathcal{O}^k(i)$, un ensemble k-libre optimal contient $\left\lceil \left|\mathcal{O}^k(i)\right| /2 \right\rceil$ éléments. Pour voir cela, il suffit de considérer la parité du cardinal de l'orbite. En sommant ces quantités sur chacune des orbites, nous allons obtenir $r_k(n)$. Cette méthode nous permet en outre de connaître tous les ensembles optimaux possibles.

Si on note $A_i :=]\!]\frac{n}{k^{i+1}}, \frac{n}{k^i}]\!]$, on a

$$[\![1,n]\!] = \bigcup_{i=0}^{d} A_i$$

où $d = [\log_k(n)]$. De plus,

$$|A_i| = \frac{n}{k^i} - \frac{n}{k^{i+1}} + \alpha(i)$$

avec $|\alpha(i)| \leqslant 1$. Et le nombre de $j \not\equiv 0 \pmod{k}$ dans A_i est

$$\left(1 - \frac{1}{k}\right)\left(\frac{n}{k^i} - \frac{n}{k^{i+1}} + \alpha(i)\right) + \epsilon(i)$$

avec $|\epsilon(i)| \leqslant 1$. Chaque élément de A_i a une orbite de taille $i+1$, ce qui nous permet de calculer $r_k(n)$:

$$\begin{aligned} r_k(n) &= \sum_{i=0}^{d} \left\lceil \frac{i+1}{2} \right\rceil \left(\left(1 - \frac{1}{k}\right)\left(\frac{n}{k^i} - \frac{n}{k^{i+1}} + \alpha(i)\right) + \epsilon(i)\right) \\ &= \sum_{i=0}^{d} \left\lceil \frac{i+1}{2} \right\rceil \left(1 - \frac{1}{k}\right)\left(\frac{n}{k^i} - \frac{n}{k^{i+1}}\right) + O(\log_k^2(n)). \end{aligned}$$

En regroupant par deux les termes pour lesquels la partie entière est la même, afin de faire apparaître un comportement télescopique, on obtient :

$$\begin{aligned} r_k(n) &= \left(1-\frac{1}{k}\right)\sum_{i=0}^{\lfloor \frac{d}{2} \rfloor}(i+1)\left(\frac{n}{k^{2i}}-\frac{n}{k^{2i+1}}+\frac{n}{k^{2i+1}}-\frac{n}{k^{2i+2}}\right) \\ &\quad + \beta(n) + O(\log_k^2(n)) \\ &= \left(1-\frac{1}{k}\right)\sum_{i=0}^{\lfloor \frac{d}{2} \rfloor}(i+1)\left(\frac{n}{k^{2i}}-\frac{n}{k^{2i+2}}\right) + \beta(n) + O(\log_k^2(n)) \end{aligned}$$

avec

$$|\beta(n)| \leqslant \left(1-\frac{1}{k}\right)\left(\frac{d}{2}+1\right)\left(\frac{n}{k^d}-\frac{n}{k^{d+1}}\right) = O(\log_k(n)).$$

Finalement,

$$\begin{aligned} r_k(n) &= \left(1-\frac{1}{k}\right)\sum_{i=0}^{\lfloor \frac{d}{2} \rfloor}\frac{n}{k^{2i}} + O(\log_k^2(n)) \\ &= \frac{k}{k+1}n + O(\log_k^2(n)). \end{aligned}$$

Remarque. Si on note $\mathcal{I}_k$ l'ensemble des entiers i qui ne sont pas des multiples de k, l'ensemble

$$B = \bigcup_{l=0}^{+\infty} k^{2l}\mathcal{I}_k$$

où l'union est disjointe, est un ensemble k-libre dans $\mathbb{N}$. En effet, on a cette fois une partition de $\mathbb{N}\setminus\{0\}$ (0 n'est jamais dans un ensemble k-libre) en orbites infinies

$$\mathbb{N}\setminus\{0\} = \bigcup_{i\not\equiv 0 \pmod{k}} \mathcal{O}^k(i)$$

dans lesquelles on prend un élément sur deux, en commençant par le premier. Il s'agit clairement du prolongement des ensembles qu'on a pu considérer auparavant. B est donc un ensemble k-libre optimal dans $\mathbb{N}$, de densité $k/(k+1)$.

3.2.2 Ensembles k-libres maximaux au sens de l'inclusion

En suivant le même schéma, nous allons calculer la taille minimale d'un ensemble k-libre maximal au sens de l'inclusion. En effet, considérant toujours la partition en orbites

$$[\![1,n]\!] = \bigsqcup_{i\not\equiv 0 \pmod{k}} \left(\mathcal{O}^k(i)\bigcap [\![1,n]\!]\right),$$

pour qu'un ensemble B soit maximal au sens de l'inclusion des ensembles k-libres, cela signifie que sur chacune de ces orbites, B ne contient pas deux éléments consécutifs, il n'existe pas trois éléments consécutifs qui ne soient pas dans B (sinon, on pourrait rajouter par exemple le deuxième de ces trois éléments), l'un des deux premiers éléments de l'orbite est dans B et l'un des deux derniers éléments de l'orbite est dans B. Réciproquement, il est clair que si un ensemble vérifie ces quatre propriétés, il est k-libre, et si on lui ajoute un élément, il ne l'est plus.

Essayer de minimiser la taille d'un tel ensemble nous conduit donc immédiatement au problème combinatoire suivant :

Dans $[\![1, j]\!]$, quelle est la taille minimale d'un ensemble E vérifiant les propriétés suivantes $(\mathcal{P})$

- $1 \in E$ ou $2 \in E$
- $j - 1 \in E$ ou $j \in E$
- $i \in E \Rightarrow (i-1) \notin E$ et $(i+1) \notin E$
- $\forall i \in [\![2, j-1]\!], \{i-1, i, i+1\} \cap E \neq \emptyset$

Notons $h(j)$ la taille minimale d'un ensemble vérifiant $(\mathcal{P})$.

Lemme 3.1.

$$h(j) = \left\lceil \frac{j}{3} \right\rceil$$

Démonstration. Déjà, lorsque $j = 3l$, si on prend $B = \{2, 5, \cdots, 2 + 3(l-1)\}$, B vérifie bien $(\mathcal{P})$ et est de taille $l = j/3$. Etant donné qu'on doit prendre au moins un élément parmi $\{3i+1, 3i+2, 3i+3\}, \forall i \in [\![0, l-1]\!]$, $h(j) \geqslant l$. Finalement, on a bien $h(3l) = l$

Si $j = 3l - 1$, on partitionne de la façon suivante

$$[\![1, 3l-1]\!] = \{1, 2\} \bigcup \left(\bigcup_{i \in [\![1, l-1]\!]} \{3i, 3i+1, 3i+2\} \right)$$

Comme on doit avoir au moins un élément dans chacun de ces ensembles, on a $h(3l-1) \geqslant l$. Or $B = \{2, 5, \cdots, 2 + 3(l-1)\}$ fonctionne à nouveau. Donc $h(3l-1) = l$.

Si $j = 3l - 2$, on partitionne de la façon suivante

$$[\![1, 3l-2]\!] = \{1, 2\} \bigcup \left(\bigcup_{i \in [\![1, l-2]\!]} \{3i, 3i+1, 3i+2\} \right) \bigcup \{3l-3, 3l-2\}$$

Comme on doit avoir au moins un élément dans chacun de ces ensembles, on a $h(3l-2) \geqslant l$; Or $B = \{1, 4, \cdots, 1 + 3(l-1)\}$ fonctionne. Donc $h(3l-2) = l$.

□

Nous pouvons désormais prouver le Théorème 3.3 en effectuant des calculs similaires à ceux utilisés pour $r_k(n)$.

Démonstration. Avec $d = [\log_k(n)]$, on aboutit comme précédemment à :

$$\begin{aligned}\tilde{R_k}(n) &= \sum_{i=0}^{d} \left\lceil \frac{i+1}{3} \right\rceil \left(\left(1 - \frac{1}{k}\right) \left(\frac{n}{k^i} - \frac{n}{k^{i+1}} + \alpha(i) \right) + \epsilon(i) \right) \\ &= \sum_{i=0}^{d} \left\lceil \frac{i+1}{3} \right\rceil \left(1 - \frac{1}{k}\right) \left(\frac{n}{k^i} - \frac{n}{k^{i+1}} \right) + O(\log_k^2(n)).\end{aligned}$$

Et en regroupant cette fois-ci les termes par trois, on obtient :

$$\begin{aligned}\tilde{R_k}(n) &= \left(1 - \frac{1}{k}\right) \sum_{i=0}^{\lfloor \frac{d}{3} \rfloor} (i+1) \left(\frac{n}{k^{3i}} - \frac{n}{k^{3i+1}} + \frac{n}{k^{3i+1}} - \frac{n}{k^{3i+2}} + \frac{n}{k^{3i+2}} - \frac{n}{k^{3i+3}} \right) \\ &\quad + \beta(n) + O(\log_k^2(n)) \\ &= \left(1 - \frac{1}{k}\right) \sum_{i=0}^{\lfloor \frac{d}{3} \rfloor} (i+1) \left(\frac{n}{k^{3i}} - \frac{n}{k^{3i+3}} \right) + \beta(n) + O(\log_k^2(n))\end{aligned}$$

avec

$$|\beta(n)| \leqslant \left(1 - \frac{1}{k}\right) \times 2 \left(\frac{d}{3} + 1 \right) \left(\frac{n}{k^{d-1}} - \frac{n}{k^{d+1}} \right) = O(\log_k(n)).$$

Finalement,

$$\begin{aligned}\tilde{R_k}(n) &= \left(1 - \frac{1}{k}\right) \sum_{i=0}^{\lfloor \frac{d}{3} \rfloor} \frac{n}{k^{3i}} + O(\log_k^2(n)) \\ &= \frac{k^2}{k^2 + k + 1} n + O(\log_k^2(n)).\end{aligned}$$

□

3.3 Les ensembles k-libres modulaires

Dans cette partie, nous étudions les ensembles k-libres dans $\mathbb{Z}/n\mathbb{Z}$. Remarquons tout d'abord que cela englobe le cas des ensembles $\{a, b\}$-multiplicatifs dans $\mathbb{Z}/n\mathbb{Z}$ lorsque $\text{pgcd}(a, n) = 1$, puisqu'il s'agit alors d'un ensemble ba^{-1}-libre.

On définit

$$R_k(n) = \max \{ |A|, A \text{ est un ensemble } k\text{-libre dans } \mathbb{Z}/n\mathbb{Z} \}$$

et nous allons voir comment obtenir cette quantité récursivement en n (Théorèmes 3.4, 3.5, 3.6 et 3.7). En outre, comme dans le cas des entiers naturels, les preuves seront constructives.

L'étude de ces ensembles avec la contrainte modulaire dépend fortement des propriétés arithmétiques entre n et k, c'est pourquoi nous présenterons les résultats en quatre théorèmes. Nous commencerons par traiter le cas où k et n sont premiers entre eux, ce qu'on peut d'ailleurs considérer comme le cas le plus important. En effet, comme le précise Cilleruelo et Timmons dans [2], quand on définit un ensemble k-Sidon dans $\mathbb{Z}/n\mathbb{Z}$, on doit ajouter la condition "n est premier avec tous les entiers entre 1 et k". Sinon, on pourrait avoir $c_i(a_1 - a_2) = 0$ avec $a_1 \neq a_2$ pour $|c_i| \leqslant k$, ce qui donnerait une solution non triviale à $c_i(x_1 - x_2) + x_3 - x_4 = 0$ par exemple.

Avant d'énoncer les premiers résultats, introduisons quelques notations. Pour des entiers k et d premiers entre eux, on note $\ell_k(d)$ l'ordre multiplicatif de k dans $(\mathbb{Z}/d\mathbb{Z})^*$. On désignera par I la fonction indicatrice des nombres impairs et par φ la fonction indicatrice d'Euler. On est désormais en mesure de donner le premier théorème.

Théorème 3.4. *Si* $\mathrm{pgcd}(n, k) = 1$,

$$R_k(n) = \frac{n-1}{2} - \sum_{d|n, d\neq 1} \frac{\varphi(d) I(\ell_k(d))}{2\ell_k(d)}.$$

Remarque. En ce qui concerne la borne supérieure pour la taille d'un ensemble 2-Sidon, ce sont les "petits" $R_2(n)$ qui nous intéressent. Dans le cas où $n = 2^m - 1$ est un nombre premier de Mersenne, ce qui implique m premier, alors $\ell_2(n) = m$, et

$$R_2(n) = \frac{n-1}{2} - \frac{n-1}{2\log_2(n+1)}.$$

Ainsi, pour un ensemble 2-Sidon A dans $\mathbb{Z}/n\mathbb{Z}$, comme $D^*(A)$ est un ensemble 2-libre, on a

$$2\binom{|A|}{2} \leq R_2(n),$$

ce qui donne

$$|A| \leqslant \sqrt{\frac{n-1}{2} - \frac{n-1}{2\log_2(n+1)} + \frac{1}{4}} + \frac{1}{2}.$$

En outre, on prouvera dans la section 3.3.1.2 qu'à k fixé, le terme d'erreur est un $o(n)$. Ainsi, $R_k(n) = (n-1)/2 - o(n)$.

Lorsque k divise n, le problème est plus facile et on a les deux résultats suivants :

Théorème 3.5. *Si m n'est pas divisible par k, alors*

$$R_k(km) = (k-1)m.$$

Si k^2 divise n, on obtient une formule récursive :

Théorème 3.6. *Considérons les entiers k, m, et n. On a alors :*

$$R_k(k^2 m) = R_k(m) + \left(k^2 - k\right) m.$$

Remarquons que les trois théorèmes précédents permettent de calculer précisément $R_k(n)$ lorsque k est premier. De plus, rappelons que la densité maximale d'un ensemble k-libre dans $[\![1, n]\!]$ est $k/(k+1)$. Dans le cas modulaire, en appliquant le Théorème 3.6, on obtient

$$R_k(k^{2m}) = \frac{k}{k+1}\left(k^{2m} - 1\right)$$

ce qui conduit à la proposition suivante.

Proposition 3.1. *Soit k un entier, $k \geqslant 1$, on a*

$$\limsup_n \frac{R_k(n)}{n} = \frac{k}{k+1}.$$

Maintenant, pour illustrer les deux théorèmes précédents, appliquons les à un exemple, et calculons $R_{15}(826875)$:

$$\begin{aligned} R_{15}(826875) &= R_{3.5}(3^3.5^4.7^2) \\ &= R_{3.5}(3.5^2.7^2) + (15^2 - 15).3.5^2.7^2 \\ &= (15-1)5.7^2 + (15^2 - 15).3.5^2.7^2 \\ &= 775180. \end{aligned}$$

Nous reviendrons sur cet exemple dans la section 3.3.2.3.

Pour le cas général, nous ne pouvons obtenir une formule satisfaisante, mais nous donnerons un algorithme dans la section 2.4 qui permet de calculer $R_k(n)$.

Théorème 3.7. *Il existe un algorithme qui donne la taille maximale d'un ensemble k-libre dans $\mathbb{Z}/n\mathbb{Z}$ et une méthode pour en construire un en $O((\log(n))^2)$ opérations.*

Cet algorithme et la complexité associée présupposent que l'on connaisse les factorisations en nombre premiers de k et n, qui sont malheureusement difficiles à obtenir en général. Cependant, nous verrons dans la section 3.3.2.3 comment appliquer cet algorithme pour calculer R_k et obtenir une formule explicite pour de nouvelles catégories de k et n. Ce sera l'objet du Théorème 3.8.

3.3.1 Les trois premiers théorèmes

3.3.1.1 Lemmes préparatoires

Introduisons tout d'abord quelques notations utiles pour la suite. Rappelons qu'on note $\mathcal{O}^k(x) := \{k^j x, j \in \mathbb{N}\}$ l'orbite de x (pour la multiplication par k), $\ell_k(d)$ l'ordre multiplicatif de k dans $(\mathbb{Z}/d\mathbb{Z})^*$ et $k \cdot A := \{ka, a \in A\}$, à ne pas confondre avec kA. On définit de plus, pour m diviseur de n, A_m un sous-ensemble de $\mathbb{Z}/n\mathbb{Z}$ par

$$A_m = \{x, \operatorname{pgcd}(x,n) = m\} = \left\{x = mu, \operatorname{pgcd}\left(u, \frac{n}{m}\right) = 1\right\}$$

dont le cardinal vérifie $|A_m| = \varphi(n/m)$.

La partition de $\mathbb{Z}/n\mathbb{Z}$ en l'union des A_m, indexée par les diviseurs de n, nous sera très utile pour l'étude des ensembles k-libres. Il est donc naturel de se demander comment celle-ci se comporte vis-à-vis de la multiplication par k. Le premier lemme permet de décrire $k \cdot A_m$.

Lemme 3.2. *Si la décomposition en facteurs premiers de n s'écrit*

$$n = \prod_{i=1}^{r} p_i^{n_i} \prod_{i=r+1}^{s} p_i^{n_i} \text{ et } k = u \prod_{i=1}^{r} p_i^{k_i}$$

où $\operatorname{pgcd}(u, p_i) = 1, \forall i \in [\![1, s]\!]$, *alors, tout m diviseur de n s'écrit sous la forme*

$$m = \prod_{i=1}^{r} p_i^{m_i} \prod_{i=r+1}^{s} p_i^{m_i}$$

avec $m_i \leqslant n_i, \forall i \in [\![1, s]\!]$, *et nous avons*

$$k \cdot A_m = A_{m'} \text{ où } m' = m \prod_{i=1}^{r} p_i^{\min(k_i, n_i - m_i)}.$$

Démonstration. Soit $x \in A_m$, alors $x = mv$ avec $\operatorname{pgcd}(v, n/m) = 1$. Ainsi, on a

$$\begin{aligned}
\operatorname{pgcd}(kx, n) &= m \operatorname{pgcd}(kv, \frac{n}{m}) \\
&= m \operatorname{pgcd}(k, \frac{n}{m}) \\
&= m \operatorname{pgcd}(\operatorname{pgcd}(k, n), \frac{n}{m}) \\
&= m \operatorname{pgcd}(\prod_{i=1}^{r} p_i^{k_i}, \prod_{i=1}^{r} p_i^{n_i - m_i} \prod_{i=r+1}^{s} p_i^{n_i - m_i}).
\end{aligned}$$

On obtient ainsi $k \cdot A_m \subset A_{m'}$.

Réciproquement, on sait maintenant qu'il existe $y \in A_{m'}$ tel que $y = kx$ et $x \in A_m$ (puisque $A_m \neq \emptyset$). Mais pour tout z dans $A_{m'}$, il existe w vérifiant $\text{pgcd}(w, n) = 1$ et $z = wy$. Il est clair que xw appartient à A_m et $z = kxw$, ce qui conclut la preuve. □

Maintenant, nous déterminons dans le lemme suivant la taille de l'orbite d'un élément dans un cas qui nous intéressera par la suite.

Lemme 3.3. *Soient m un diviseur de n, et k un entier tel que* $\text{pgcd}(k, n/m) = 1$ *et $x \in A_m$. Alors*

$$\left|\mathcal{O}^k(x)\right| = \ell_k\left(\frac{n}{m}\right).$$

Démonstration. Comme $x \in A_m$, si on note $< x >$ le sous-groupe engendré par x, on a

$$< x > \cong \mathbb{Z}/\left(\frac{n}{m}\right)\mathbb{Z}.$$

Mais puisque k est inversible dans ce sous-groupe

$$\mathcal{O}^k(x) \cong \mathcal{O}^k(1) =< k > \subset \mathbb{Z}/\left(\frac{n}{m}\right)\mathbb{Z}$$

et la taille de $< k >$ dans ce sous-groupe est exactement $\ell_k(n/m)$.

□

3.3.1.2 Preuves des Théorèmes 3.4, 3.5 et 3.6

Commençons par le Théorème 3.4, le cas $\text{pgcd}(n, k) = 1$.

Démonstration. Rappelons que I désigne la fonction indicatrice des nombres impairs.

Considérons la partition

$$(\mathbb{Z}/n\mathbb{Z}) \setminus \{0\} = \bigsqcup_{m|n, m<n} A_m.$$

Dans le cas où n est premier, cette partition est triviale.

D'après le Lemme 3.2, $\mathcal{O}^k(x) \subset A_m$, pour tout x dans A_m, puisque les k_i sont nuls dans ce cas précis. De plus, par le Lemme 3.3, si $x \in A_m$, on a

$$\left|\mathcal{O}^k(x)\right| = \ell_k\left(\frac{n}{m}\right).$$

Ainsi, on peut partitionner A_m en $\varphi(n/m)/\ell_k(n/m)$ orbites distinctes de longueur $\ell_k(n/m)$. Et ces orbites sont naturellement indépendantes du point de vue de la multiplication par k. Maintenant, au sein de chacune de ces orbites, pour obtenir

un ensemble k-libre optimal, nous devons prendre le plus d'éléments possibles sans qu'il y ait d'éléments consécutifs. Le raisonnement est similaire à ce qu'on a vu dans le cas des entiers naturels, mais ici, les orbites sont cycliques, c'est pourquoi lorsque la longueur l d'une orbite est paire, on peut prendre jusqu'à $l/2$ éléments, tandis que lorsque l est impaire, il y en a au plus $(l-1)/2$. Cela nous conduit à la formule

$$\begin{aligned} R_k(n) &= \sum_{d|n, d\neq 1} \frac{\varphi(d)}{\ell_k(d)} \left(\frac{\ell_k(d) - I(\ell_k(d))}{2} \right) \\ &= \frac{n-1}{2} - \sum_{d|n, d\neq 1} \frac{\varphi(d) I(\ell_k(d))}{2\ell_k(d)}. \end{aligned}$$

□

Analysons maintenant le terme d'erreur. À k fixé, on a l'asymptotique

$$R_k(n) = (n-1)/2 - o(n).$$

En effet, pour tout $\varepsilon > 0$, il existe d_0 tel que $\log_k d_0 \geqslant 1/\varepsilon$ et il existe n vérifiant $d_0^2/6 \leqslant \varepsilon n/2$. Ainsi,

$$\begin{aligned} \sum_{d|n, d\neq 1} \frac{\varphi(d) I(\ell_k(d))}{2\ell_k(d)} &= \sum_{d|n, d\neq 1, d\leqslant d_0} \frac{\varphi(d) I(\ell_k(d))}{2\ell_k(d)} + \sum_{d|n, d\neq 1, d> d_0} \frac{\varphi(d) I(\ell_k(d))}{2\ell_k(d)} \\ &\leqslant \sum_{d|n, d\neq 1, d\leqslant d_0} \frac{\varphi(d)}{6} + \sum_{d|n, d\neq 1, d> d_0} \frac{\varphi(d)}{2 log_k d} \\ &\leqslant \frac{d_0^2}{6} + \frac{\varepsilon n}{2} \\ &\leqslant \varepsilon n. \end{aligned}$$

Considérons maintenant le cas $n = k^2 m$, pour lequel on va utiliser la partition suivante de $\mathbb{Z}/n\mathbb{Z}$:

Lemme 3.4. *Pour $n = k^2 m$, on a*

$$\mathbb{Z}/n\mathbb{Z} = \left(k^2\mathbb{Z}/n\mathbb{Z}\right) \bigsqcup \left(\bigcup_{h\not\equiv 0 \pmod{k}} \{h, kh\} \right)$$

où la première union est disjointe.

Démonstration. En effet, si $x \not\equiv 0 \pmod{k^2}$ et $x \equiv 0 \pmod{k}$, alors $x = kh$ avec $h \not\equiv 0 \pmod{k}$. Ainsi, nous retrouvons bien tous les éléments dans cette union. De plus, si on a $h \not\equiv 0 \pmod{k}$, alors $kh \not\equiv 0 \pmod{k^2}$, ce qui montre que cette union est bien disjointe. □

Voyons maintenant en quoi c'est une bonne répartition des éléments pour notre problème, à travers la preuve du Théorème 3.6 :

Démonstration. Remarquons deux choses :
- $k^2\mathbb{Z}/n\mathbb{Z}$ est stable pour la multiplication par k.
- Si $h \not\equiv 0 \pmod{k}$, on ne peut pas écrire $h = ku$ dans $k^2\mathbb{Z}/n\mathbb{Z}$.

Soit alors A un ensemble k-libre dans $\mathbb{Z}/n\mathbb{Z}$. Premièrement, pour chaque $h \not\equiv 0 \pmod{k}$, au plus un élément de $\{h, kh\}$ est dans A. En outre, par la première remarque, $A \cap k^2\mathbb{Z}/n\mathbb{Z}$ est aussi un ensemble k-libre, ce qu'on peut voir comme un ensemble k-libre dans $\mathbb{Z}/m\mathbb{Z}$. Cela conduit à

$$R_k(k^2m) \leqslant R_k(m) + |\{h \not\equiv 0 \pmod{k}\}| = R_k(m) + \left(k^2 - k\right) m.$$

Construisons désormais un ensemble k-libre optimal. D'après la deuxième remarque, on peut prendre tout $h \not\equiv 0 \pmod{k}$ dans A, et on sait alors que $kh \notin k^2\mathbb{Z}/n\mathbb{Z}$, alors on peut prendre $R_k(m)$ éléments de $k^2\mathbb{Z}/n\mathbb{Z}$ dans A. Ainsi, on a

$$R_k(k^2m) = R_k(m) + \left(k^2 - k\right) m$$

ce qui conclut la preuve.

□

Finalement, venons-en au cas $n = km$ avec $m \not\equiv 0 \pmod{k}$. Dans ce cas, on a :

Lemme 3.5. *Si $n = km$ avec $m \not\equiv 0 \pmod{k}$,*

$$\mathbb{Z}/n\mathbb{Z} = \bigcup_{h \not\equiv 0 \pmod{k}} \{h, kh\}.$$

Démonstration. Si $x \equiv 0 \pmod{k}$, il existe u tel que $x = ku$. Si $u \not\equiv 0 \pmod{k}$, x est de la forme souhaitée. Sinon, $u \equiv 0 \pmod{k}$, alors il existe v, $u = kv$ et nous avons $x = x + n = x + km = k^2v + km$. Mais $m \not\equiv 0 \pmod{k}$ par hypothèse, alors on peut écrire $m = lk + a$ avec $a \not\equiv 0 \pmod{k}$, ce qui donne

$$x + km = k(kv + lk + a).$$

Comme $h = kv + lk + a \not\equiv 0 \pmod{k}$, on est parvenu à écrire $x = x + km = kh$ avec $h \not\equiv 0 \pmod{k}$, ce qui conclut le lemme. □

On peut maintenant facilement prouver le Théorème 3.5.

Démonstration. Si A est un ensemble k-libre, pour chaque $h \not\equiv 0 \pmod{k}$, au plus un élément de $\{h, kh\}$ est dans A, et donc $|A| \leqslant (k-1)m$. Si $h \not\equiv 0 \pmod{k}$, on ne peut pas écrire $h = ku$ dans $\mathbb{Z}/n\mathbb{Z}$ étant donné que $n = km$. Ainsi $\{h \not\equiv 0 \pmod{k}\}$ est un ensemble k-libre et on a bien

$$R_k(km) = (k-1)m.$$

□

3.3.2 Dans le cas général

La situation dans le cas général est nettement plus complexe. En effet, contrairement à ce qui se passe dans le cadre du Théorème 3.4, l'orbite d'un élement n'est pas nécessairement incluse dans un A_m. Il va falloir gérer le fait qu'on passe d'un A_m à un autre en multipliant les éléments par k. Pour cela, notre stratégie sera de créer un graphe qui aura pour sommets les diviseurs de n et qui seront reliés si $k \cdot A_m = A_{m'}$ lorsque $m \neq m'$. Il nous faudra cependant traiter à part les m tels que $k \cdot A_m = A_m$. Il s'agira en fait des racines de notre graphe (une fois qu'on l'aura interprété comme une forêt). Une fois cela fait, on aura envie de prendre certains A_m sans qu'aucun ne soit relié, et de façon à maximiser le cardinal de notre ensemble k-libre. C'est pourquoi nous allons avoir besoin d'un résultat sur les arbres enracinés dont les sommets sont pondérés. C'est l'objet de la section suivante.

3.3.2.1 Un algorithme sur les arbres enracinés

Soit T un arbre enraciné dont l'ensemble des noeuds est $V = \{v_i\}_{i \in I}$ où I est un ensemble fini et E est l'ensemble des arêtes. On associe une valeur $\alpha_i \geqslant 0$ à chaque noeud v_i et on note l_i son niveau (autrement appelé hauteur ou profondeur). Rappelons que le niveau d'un noeud est égal au nombre de noeuds à partir de la racine (en comptant la racine) pour aller jusqu'au noeud. On peut le définir de manière équivalente comme étant 1+ le nombre minimal d'arêtes entre le noeud et la racine. Supposons que T a la propriété suivante :

$$\text{Si } v_i \text{ est le parent de } v_j, \text{ alors } \alpha_i < \alpha_j. \tag{3.1}$$

En d'autres mots, α croît strictement sur chaque branche. On remarque que cette condition implique que si v_i n'est pas la racine de T, $\alpha_i > 0$. Nous recherchons un sous-ensemble A de I satisfaisant :

$$\forall (i,j) \in A^2, (v_i, v_j) \notin E \text{ et } \alpha_i \neq 0 \tag{3.2}$$

qui maximise la quantité

$$\Lambda_A = \sum_{i \in A} \alpha_i.$$

Notons l la profondeur maximum de T et construisons un ensemble B de la façon suivante :
Initialisation : $B = \{v_i | l_i = l\}$.
Ensuite, pour k allant de 1 à $l-1$: pour tout i tel que $l_i = l - k$, on ajoute v_i à B si et seulement $\alpha_i \neq 0$ et v_i n'a pas d'enfant dans B.

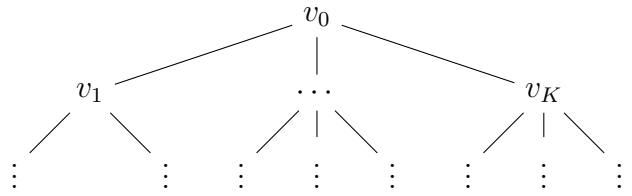
Il est clair que B satisfait (3.2). En fait, on va voir que B est l'ensemble recherché.

Lemme 3.6. *B est l'unique sous-ensemble de I maximisant Λ_A parmi les ensembles A satisfaisant (3.2).*

Démonstration. Procédons par récurrence sur la taille de I. Si $|I| = 1$, il n'y a rien à faire.

Soit n un entier et supposons que le lemme est vrai pour k plus petit que n. Supposons $|I| = n + 1$. Soient B l'ensemble provenant de l'algorithme expliqué ci-dessus, appliqué à T, et C un sous-ensemble de I maximisant Λ_A parmi les ensembles A satisfaisant (3.2). On veut montrer que $B = C$.

Notons v_0 la racine de T, et v_i, pour $i \in [\![1, K]\!]$, les enfants de v_0. On définit également, pour tout i dans $[\![1, K]\!]$, T_i le sous-arbre (enraciné) de T qui a pour racine v_i, $B_i = B \cap T_i$ et $C_i = C \cap T_i$. On pourra éventuellement mieux visualiser ces ensembles en s'appuyant sur le schéma ci-dessous, où T_1 est par exemple le sous-arbre (à gauche) issu de v_1.



Ce qui est intéressant dans cet algorithme, c'est que si on l'applique à T_i, on obtient l'ensemble B_i. Cela vient du fait que l'algorithme a pour point de départ les éléments de profondeur maximale dans l'arbre. Ainsi, d'après l'hypothèse de récurrence, on a pour tout i, $\Lambda_{C_i} \leqslant \Lambda_{B_i}$ avec égalité si et seulement si $B_i = C_i$.

Si $v_0 \in B$ et $v_0 \in C$, on a

$$\Lambda_C - \alpha_0 = \sum_{i=1}^{K} \Lambda_{C_i} \leqslant \sum_{i=1}^{K} \Lambda_{B_i} = \Lambda_B - \alpha_0.$$

Ainsi, par définition de C, il s'agit d'une égalité, et on a finalement $B = C$.

Si $v_0 \notin B$ et $v_0 \notin C$, on a

$$\Lambda_C = \sum_{i=1}^{K} \Lambda_{C_i} \leqslant \sum_{i=1}^{K} \Lambda_{B_i} = \Lambda_B$$

ce qui assure $B = C$ pour la même raison.

Si $v_0 \in B$ et $v_0 \notin C$, comme $\alpha_0 > 0$ (autrement, v_0 n'est pas dans B d'après l'algorithme), on a

$$\Lambda_C = \sum_{i=1}^{K} \Lambda_{C_i} \leqslant \sum_{i=1}^{K} \Lambda_{B_i} = \Lambda_B - \alpha_0 < \Lambda_B$$

ce qui conduit à une contradiction.

Si $v_0 \notin B$ et $v_0 \in C$, $\alpha_0 > 0$ (comme C satisfait (3.2)) et cela signifie qu'il existe i_0 dans $[\![1, K]\!]$ tel que $v_{i_0} \in B$. Considérons alors la branche issue de v_0 qui contient v_{i_0}. Si $K > 1$, elle contient strictement moins de $n+1$ noeuds et on peut alors appliquer l'hypothèse de récurrence pour obtenir $\Lambda_{C_{i_0}} + \alpha_0 < \Lambda_{B_{i_0}}$. Ainsi,

$$\Lambda_C = \sum_{i \neq i_0} \Lambda_{C_i} + \Lambda_{C_{i_0}} + \alpha_0 < \sum_{i \neq i_0} \Lambda_{B_i} + \Lambda_{B_{i_0}} = \Lambda_B$$

et on a une contradiction. Si $K = 1$, notons v_1 l'unique enfant de v_0. On a $v_1 \in B$ et $v_1 \notin C$. En considérant maintenant n'importe quel sous-arbre enraciné en les enfants de v_1, on a (avec des notations évidentes) :

$$\Lambda_C = \sum \Lambda_{C'_i} + \alpha_0 < \sum \Lambda_{B'_i} + \alpha_1 = \Lambda_B$$

puisque $\alpha_1 > \alpha_0$. Cela fournit à nouveau une contradiction.

Finalement, dans chacun de ces cas, on obtient $B = C$. Le lemme est alors prouvé.

□

Nous allons maintenant voir comment on se ramène à un arbre de ce type pour notre problème.

3.3.2.2 L'algorithmique du cas général

L'objectif est de définir une forêt (une union disjointe d'arbres enracinés) satisfaisant (3.1) et telle que l'algorithme du paragraphe précédent conduise à un ensemble k-libre optimal.

Définissons le graphe $G = (V, E)$ par l'ensemble de ses sommets $V = \{m\}_{m|n}$ et de ses arêtes E :

$$(m, m') \in E \text{ si et seulement si } m < m' \text{ et } k \cdot A_m = A_{m'}. \tag{3.3}$$

Essayons de bien comprendre ce graphe avant d'assigner des valeurs aux différents sommets qui nous permettront de résoudre le problème. On écrit

$$n = \prod_{i=1}^{r} p_i^{n_i} \prod_{i=r+1}^{s} p_i^{n_i} \text{ et } k = u \prod_{i=1}^{r} p_i^{k_i}$$

avec $\text{pgcd}(u, p_i) = 1, \forall i \in [\![1, s]\!]$ et $k_i > 0$ pour tout i. Notons $\mathcal{M}$ l'ensemble des diviseurs de n de la forme

$$m = \prod_{i=1}^{s} p_i^{m_i}$$

avec $m_i \leqslant n_i, \forall i \in [\![1, s]\!]$, et tel qu'il existe $i_0 \leqslant r$ vérifiant $m_{i_0} < \min(k_{i_0}, n_{i_0})$. La proposition suivante donne la structure de G.

Proposition 3.2. *G est une union disjointe d'arbres enracinés. En outre :*

(i) Une composante connexe de G est entièrement déterminée par le choix de $\{d_i\}_{i=r+1}^{s}$ avec $d_i \leqslant n_i$.

(ii) Les feuilles sont exactement les éléments de $\mathcal{M}$.

(iii) La racine de l'arbre défini par $\{d_i\}_{i=r+1}^{s}$ est

$$m = \prod_{i=1}^{r} p_i^{n_i} \prod_{i=r+1}^{s} p_i^{d_i}.$$

(iv) La profondeur de m est $j_m + 1$ où

$$j_m = \min \{j | jk_i \geqslant n_i - m_i, \forall i \in [\![1, r]\!]\}.$$

Démonstration. On définit

$$k^j * m = m \prod_{i=1}^{r} p_i^{\min(jk_i, n_i - m_i)}.$$

D'après le Lemme 3.2, $A_{k*m} = k \cdot A_m$, donc si (m, m') est une arête, on a $m_i = m'_i$ pour tout i dans $[\![r+1, s]\!]$. Ainsi, s'il existe un chemin entre deux sommets, ils ont les mêmes $\{d_i\}_{i=r+1}^{s}$. La réciproque, pour le premier point, sera démontrée à la fin de cette preuve.

Dans le lemme suivant, on montre qu'un sommet est ou bien dans $\mathcal{M}$ ou bien est un ancêtre d'un élément de $\mathcal{M}$.

Lemme 3.7. *Soit $m' = \prod_{i=1}^{s} p_i^{m'_i}$ un diviseur de n qui n'est pas dans $\mathcal{M}$, alors il existe $t > 0$ et m dans $\mathcal{M}$ tel que $m' = k^t * m$.*

Démonstration. Soit t defini par

$$t = \min\left\{j | \exists i_0 \leqslant r, m'_{i_0} - jk_{i_0} < \min(k_{i_0}, n_{i_0})\right\}$$

et introduisons $\alpha_i = \max(0, m'_i - tk_i)$ et

$$m = \prod_{i=1}^{r} p_i^{\alpha_i} \prod_{i=r+1}^{s} p_i^{m'_i}$$

qui appartient à $\mathcal{M}$ par définition de t. Remarquons que $t > 0$ puisque $m' \notin \mathcal{M}$. On a alors

$$\begin{aligned} k^t * m &= m \prod_{i=1}^{r} p_i^{\min(tk_i, n_i - m_i)} \\ &= \prod_{i=1}^{r} p_i^{\alpha_i + \min(tk_i, n_i - \alpha_i)} \prod_{i=r+1}^{s} p_i^{m'_i}. \end{aligned}$$

Nous devons étudier trois cas :

- $\alpha_i = 0$ et $k_i < n_i$: $m'_i \geqslant k_i$ car $m' \notin M$, alors $m'_i = tk_i$ par définition de t et on obtient dans ce cas $\alpha_i + \min(tk_i, n_i - \alpha_i) = m'_i$.
- $\alpha_i = 0$ et $n_i \leqslant k_i$: $m'_i = n_i$ car $m' \notin M$ et on a $\alpha_i + \min(tk_i, n_i - \alpha_i) = n_i = m'_i$.
- Sinon, $n_i - \alpha_i = n_i - m'_i + tk_i \geqslant tk_i$, donc $\alpha_i + \min(tk_i, n_i - \alpha_i) = m'_i$.

On obtient $k^t * m = m'$, ce qui était attendu. □

Réciproquement, si $m \in M$ et $t > 0$, $k^t * m \notin M$, donc les éléments de $\mathcal{M}$ n'ont pas d'enfant. De plus, si on regarde

$$m = \prod_{i=1}^{r} p_i^{n_i} \prod_{i=r+1}^{s} p_i^{d_i}$$

il est clair que $k * m = m$, donc m n'a pas de parent, et il s'agit bien d'un racine. Finalement, pour montrer le dernier point et la réciproque du premier point, si m' a les mêmes $\{d_i\}_{i=r+1}^{s}$, on a $k^{j_m} * m' = m$ et $k^{j_m - 1} * m' \neq m$ par définition de j_m.

On a bien obtenu les quatre conclusions de la proposition.

□

Voyons maintenant quelles valeurs donner aux sommets (ou noeuds). On a construit ce graphe de telle sorte qu'un noeud différent de la racine est envoyé sur son parent via la multiplication par k (si on identifie le noeud m à l'ensemble A_m). Il nous faut alors regarder ce qu'il se passe pour les racines du graphe, c'est-à-dire les m satisfaisant $k \cdot A_m = A_m$. Le lemme suivant donne la taille maximale d'un ensemble k-libre inclus dans A_m où m est une racine du graphe.

Lemme 3.8. *Si m est une racine du graphe G (ce qui est équivalent à dire* $\mathrm{pgcd}(k, n/m) = 1$*), la taille maximale d'un ensemble k-libre inclus dans A_m est*

$$R_k(A_m) = \frac{\varphi(n/m)}{\ell_k(n/m)} \left(\frac{\ell_k(n/m) - I(\ell_k(n/m))}{2} \right).$$

Démonstration. On a un isomorphisme :

$$A_m \cong A'_1 = \left\{ x \in \mathbb{Z}/(n/m)\mathbb{Z}, \mathrm{pgcd}(x, \frac{n}{m}) = 1 \right\}.$$

Comme on est dans le cas $\mathrm{pgcd}(k, n/m) = 1$, si on procède comme dans la preuve du Théorème 3.4, on a alors immédiatement le résultat.

□

Ainsi, on va assigner les valeurs aux noeuds m comme suit :

$$\alpha_m = \begin{cases} R_k(A_m) \text{ si } m \text{ est une racine} \\ \varphi\left(\dfrac{n}{m}\right) \text{ sinon.} \end{cases}$$

Remarquons que G vérifie la propriété (3.1), que nous rappelons ici :

$$\text{Si } v_i \text{ est le parent de } v_j, \text{ alors } \alpha_i < \alpha_j.$$

En appliquant l'algorithme de la section 3.3.2.1, un obtient un ensemble de noeuds de G, que l'on va noter B. Afin de construire un ensemble k-libre, on va prendre l'union des A_m pour m dans B qui n'est pas une racine de G, et pour les racines m qui sont dans B on ajoute K_m un ensemble k-libre optimal inclus dans A_m. Plus précisément, on définit

$$\overline{B} := \left(\bigsqcup_{\substack{m \in B \\ \mathrm{pgcd}(k,n/m) \neq 1}} A_m \right) \bigsqcup \left(\bigsqcup_{\substack{m \in B \\ \mathrm{pgcd}(k,n/m) = 1}} K_m \right)$$

qui est clairement un ensemble k-libre car B vérifie (3.2) et par définition de K_m.

Voyons maintenant pourquoi cet ensemble est celui que l'on recherche.

Proposition 3.3. *$\overline{B}$ est un ensemble k-libre optimal dans $\mathbb{Z}/n\mathbb{Z}$ et a un cardinal*

$$\sum_{\substack{m \in B \\ \mathrm{pgcd}(k,n/m) \neq 1}} \varphi\left(\frac{n}{m}\right) + \sum_{\substack{m \in B \\ \mathrm{pgcd}(k,n/m) = 1}} R_k(A_m).$$

Démonstration. Supposons que C est un ensemble k-libre optimal de $\mathbb{Z}/n\mathbb{Z}$ vérifiant $|C| > \left|\overline{B}\right|$. Soient alors x un élément de $C\backslash\overline{B}$ que l'on prend de profondeur t maximale parmi de tels éléments, m le diviseur de n tel que $x \in A_m$ et T_i la composante connexe de G (ou l'arbre enraciné) contenant m.

Premier cas : $t = 1$ et $m \notin B$. Ainsi, m est une racine qui n'est pas dans B, ce qui signifie qu'il existe m' un enfant de m dans B (autrement $\alpha_m = R_k(A_m) = 0$ et C ne serait pas un ensemble k-libre). Or, l'ensemble $k^{-1}(x) = \{y \in A_{m'} | y = kx\}$ n'a aucun élément dans C mais a un cardinal

$$\left|k^{-1}(x)\right| = \frac{\varphi(n/m')}{\varphi(n/m)} > 1.$$

Ainsi, en remplaçant, dans C, $\{x\}$ par $k^{-1}(x)$, on obtient toujours un ensemble k-libre. En effet, comme t est la profondeur maximale d'un élément de $C\backslash\overline{B}$, aucun élément de C n'est dans un A_{m_0} avec m_0 un enfant de m' (puisque m' est dans B). L'ensemble k-libre que l'on obtient ainsi est de taille strictement plus grande que C, ce qui fournit une contradiction.

Deuxième cas : $t > 1$. Par construction de $\overline{B}$, m n'appartient pas à B et on peut procéder exactement comme dans le cas précédent.

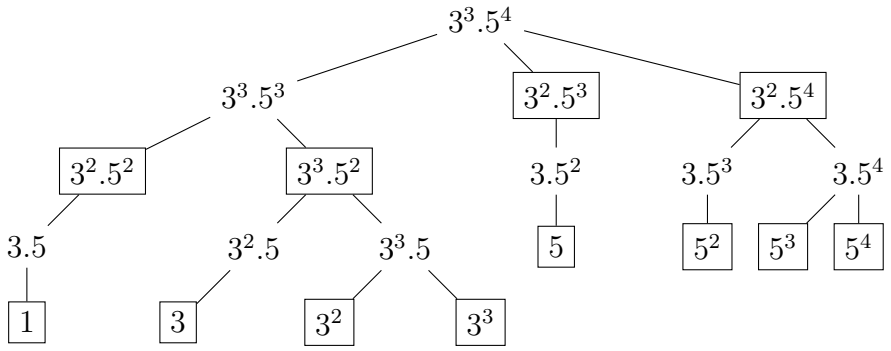
Ces deux cas menant à une contradiction, tout élément x de $C\backslash\overline{B}$ satisfait $t = 1$ et m appartient à B. Ainsi, m est une racine et on peut remplacer $C \cap A_m$ par K_m pour toutes les racines, ce qui nous donne $|C| \leqslant \left|\overline{B}\right|$. Le calcul de la taille de $\overline{B}$ est pour sa part immédiat.

□

Ainsi, pour obtenir $R_k(n)$, dans le cas où les factorisations en nombres premiers de k et n sont connues, on doit construire G ($O(\log(n))$ opérations), lui appliquer l'algorithme ($O(\log(n))$ opérations), calculer α_m pour les m dans B ($O((\log(n))^2)$ opérations puisque nous connaissons la factorisation de m) et finalement ajouter ces différentes contributions.

3.3.2.3 Illustrations de l'algorithme

Commençons par illustrer cette méthode sur l'exemple que nous avons déjà considéré au début de la partie 3.3, $n = 3^3.5^4.7^2 = 826875$ et $k = 3.5 = 15$. Dans ce cas, nous obtenons une forêt à trois arbres, de racines $3^3.5^4$, $3^3.5^4.7$ a $3^3.5^4.7^2$. On représente le premier ci-dessous. Pour obtenir le deuxième, on doit multiplier chaque noeud par 7, et pour le troisième, par 7^2. En appliquant l'algorithme, on obtient un ensemble de noeuds que l'on a encadré :



Pour obtenir la taille d'un ensemble 15-libre optimal dans $\mathbb{Z}/826875\mathbb{Z}$, nous devons alors sommer les $\varphi(n/m)$ pour les m choisis par l'algorithme dans chaque arbre. Et nous retrouvons le résultat $R_{15}(826875) = 775180$ que nous avions précédemment déduit des Théorèmes 3.5 et 3.6.

Cette façon de calculer $R_k(n)$ ne donne pas une formule générale, c'est pourquoi nous étudions dans le théorème suivant trois nouveaux cas particuliers.

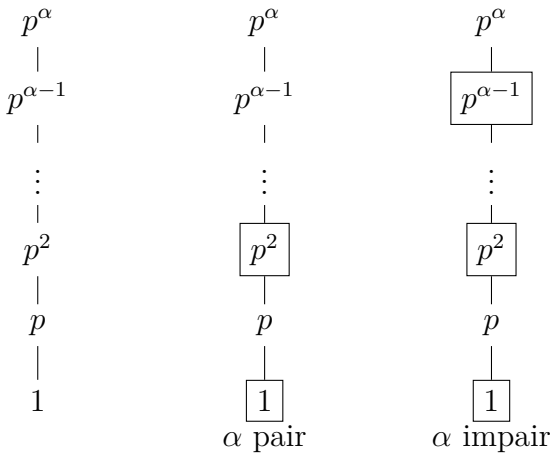
Théorème 3.8. *Soient p et q des nombres premiers, α, β et u des entiers.*

1. *Si* $\mathrm{pgcd}(u,p) = 1$,
$$R_{up}(p^\alpha) = \sum_{i=0}^{\lfloor \frac{\alpha-1}{2} \rfloor} \varphi(p^{\alpha-2i}).$$

2. *Si* $\mathrm{pgcd}(u,p) = 1$,
$$R_{up^2}(p^\alpha) = \sum_{i=0}^{\lfloor \frac{\alpha-1}{4} \rfloor} \left(\varphi(p^{\alpha-4i}) + \varphi(p^{\alpha-4i-1})\right).$$

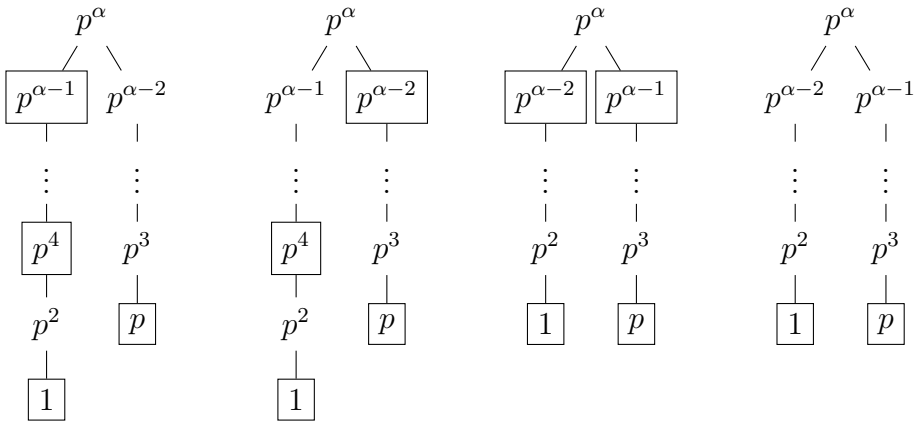
3. *Si* $\mathrm{pgcd}(u,p) = \mathrm{pgcd}(u,q) = 1$,
$$R_{up}(p^\alpha q^\beta) = \sum_{j=0}^{\beta} \sum_{i=0}^{\lfloor \frac{\alpha-1}{2} \rfloor} \varphi(p^{\alpha-2i} q^{\beta-j})$$
$$R_{up^2}(p^\alpha q^\beta) = \sum_{j=0}^{\beta} \sum_{i=0}^{\lfloor \frac{\alpha-1}{4} \rfloor} \left(\varphi(p^{\alpha-4i} q^{\beta-j}) + \varphi(p^{\alpha-4i-1} q^{\beta-j})\right).$$

Démonstration. 1. Le premier graphe qui suit est celui qu'on obtient dans ce cas particulier ($n = p^\alpha$, $k = up$ avec $\mathrm{pgcd}(u,p) = 1$), et où on a là aussi encadré l'ensemble des noeuds issus de l'algorithme. Le deuxième graphe correspond à α pair et le troisième à α impair :



Comme $A_{p^\alpha} = \{0\}$, $R_{up}(p^\alpha) = 0$, c'est pourquoi p^α n'est jamais considéré par l'algorithme. En appliquant la Proposition 3.3, on obtient le résultat.

2. On donne ci-dessous le graphe issu de l'algorithme ($n = p^\alpha$, $k = up^2$ avec $\text{pgcd}(u, p) = 1$), qui dépend de la valeur de α modulo 4 (on utilise aussi $R_{up^2}(p^\alpha) = 0$) :



Et on calcule $R_k(n)$ à nouveau grâce à la Proposition 3.3.

3. Si $k = up$ et $n = p^\alpha q^\beta$ où $\text{pgcd}(u, p) = gcd(u, q) = 1$, on obtient une forêt de $\beta + 1$ arbres $(T_j)_{j=0\cdots\beta}$ où T_j est représenté ci-dessous :

$$
\begin{array}{c}
p^\alpha q^j \\
| \\
p^{\alpha-1} q^j \\
| \\
\vdots \\
| \\
p^2 q^j \\
| \\
p q^j \\
| \\
q^j
\end{array}
$$

On obtient alors la taille d'un ensemble k-libre optimal sur T_j (sous-entendu inclus dans l'union des A_m, où m parcourt les noeuds de T_j) :

$$\sum_{i=0}^{\lfloor \frac{\alpha-1}{2} \rfloor} \varphi(p^{\alpha-2i} q^{\beta-j}).$$

Le résultat provient ensuite de l'addition des contributions de chaque T_j.
Pour $k = up^2$, il s'agit d'une conséquence du deuxième cas.

□

De la même manière, on pourrait évidemment aller plus loin et étudier les cas $k = up^3$ ou $n = p^\alpha q^\beta r^\gamma$ par exemple, mais cela donnerait des formules de moins en moins agréables.

Bibliographie

[1] R. C. Bose, S. Chowla, *Theorems in the additive theory of numbers* , Commentarii Mathematici Helvetici, 37, 141-147, 1962/1963.

[2] J. Cilleruelo and C. Timmons, *k-fold Sidon sets*, Electronic Journal of Combinatorics, 4, 2014, 9pp.

[3] F. Lazebnik, J. Verstraëte, *On hypergraphs of girth five*, Electronic Journal of Combinatorics, 10, 2003, #R25, 9pp.

[4] V. Lambert, *On modular k-free sets*, Electronic Journal of Combinatorics, 22, 2015, 18pp.

[5] K. O'Bryant, *A complete annotated bibliography of work related to Sidon sequences*, Electronic Journal of Combinatorics, DS 11, 2004, 39pp.

[6] I. Ruzsa, *Solving a linear equation in a set of integers I*, Acta Arithmetica, 65, 259-282, 1993.

[7] J. Singer, *A theorem in finite projective geometry and some applications to number theory*, Transactions of the American Mathematical Society 43, 377-385, 1938.

[8] D. Wakeham and D.R.Wood, *On Multiplicative Sidon Sets*, Integers 13, Paper No. A26, 2013.

[9] E.T.H. Wang, *On Double-Free Sets of Integers*, Ars Combinatoria, 28, 97-100, 1989.