Université Paris-Sud

Mémoire de Magistère

Lucile DEVIN

24 octobre 2014

Table des matières

Remerciements

Je remercie Florent Jouve pour sa confiance et sa disponibilité, Frédéric Le Roux pour m'avoir encouragée à me mettre à Latex, Nathalie Carrierre pour son sourire et son accueil chaleureux.

Un mois après le début de mon magistère, mon père s'est fait licencier abusivement, ma mère s'est retrouvée victime collatérale, je les remercie de leur soutien malgré tous leurs problèmes (et de m'avoir malheureusement convaincu que je faisais le bon choix en m'éloignant du monde de l'entreprise).

Je remercie aussi la FMJH — et particulièrement Dominique Hulin pour avoir proposé mon dossier — pour le soutien financier lors de ma deuxième année de magistère.

Merci à Ataùlfo, Clément et aux joueuses de tarot, grâce à qui ces années de magistère ont été lumineuses et colorées. Merci à Alfonso pour la géométrie algébrique, et la relecture de ce mémoire. Enfin et toujours, je remercie Valentin.

Le magistère de Mathématiques à Orsay

Première année (2010-2011)

J'ai commencé ma formation dans le magistère d'Orsay en Licence Mathématiques Fondamentales et Appliquées. Au premier semestre j'ai suivi comme tout le monde le tronc commun de connaissances de base, permettant d'assurer aux étudiants venant d'horizons différents le même niveau nécessaire à la poursuite du cursus. Plus portée vers les choses abstraites, j'ai choisi au second semestre de suivre le cours *graphes-algèbre linéaire*, et le cours de *programmation-algorithmique*, (ce qui ne m'empêche pas d'utiliser toujours l'algorithme de pire complexité lorsque je programme). J'ai suivi avec beaucoup de plaisir le cours d'introduction à la topologie générale qui donnait l'impression de commencer à approcher les vraies mathématiques.

Cette année a aussi donné lieu à un premier travail de recherche, j'ai étudié le troisième problème de Hilbert avec Frédéric Le Roux. Ce sujet m'a attirée car je m'étais questionnée sur le paradoxe de Banach-Tarski pour mon sujet de TIPE (en classe préparatoire aux concours des grandes écoles), et j'espérais mieux comprendre les tenants et aboutissants de ce mystère. La question de Hilbert est de savoir si en découpant (sympathiquement) un polyèdre, on peut réarranger les morceaux et obtenir n'importe quel polyèdre de même volume. Dehn a apporté une réponse négative en introduisant un invariant stable par découpage sympathique : il trouve deux polyèdres de même volume mais d'invariants différent. Mon mémoire en annexe tente d'expliquer son cheminement.

Pour finir cette année, et appliquer les mathématiques, j'ai effectué un stage de cinq semaines dans un laboratoire de cardio-biologie dans l'université de Chatenay-Malabry. Il s'est trouvé qu'il y a assez peu de mathématiques à faire dans ce domaine car on utilise plutôt des logiciels bien pratiques qui font les mathématiques à notre place, j'ai donc surtout fait du traitement de données (comme vous pouvez le lire dans mon rapport en annexe). J'ai cependant découvert la vie de laboratoire, et mis les pieds dans la recherche actuelle même si ce n'était pas dans mon sujet de prédilection.

Deuxième année (2011-2012)

En première année de master, il fallait faire des choix. Je me suis tournée vers l'*algèbre*, *arithmétique* et *géométrie*, tout en gardant un peu d'analyse — notamment le cours de *distributions* —pour ne pas risquer de passer à côté d'un sujet utile pour l'agrégation. Les cours proposés dans le cadre du magistère — *Introduction à la Théorie Spectrale et à l'analyse Harmonique* puis *Introduction aux Systèmes Dynamiques*

— ainsi que les diverses options auxquelles j'ai pu assister — *Logique* et *Histoire des Mathématiques* — ont permis d'enrichir ma culture mathématiques pour faire un choix de deuxième année réfléchi.

Pour le travail encadré de recherche, Michel Raynaud proposait de s'intéresser au théorème de la progression arithmétique de Dirichlet. Clément Sarrazin étant également attiré par le sujet, nous avons travaillé ensemble à la compréhension de ce théorème. Le but était de lire le chapitre correspondant dans le *Cours d'arithmétique* de Serre [9], et de l'expliquer (à deux) en moins d'une heure et demie au tableau (cela n'a pas donné lieu à la rédaction d'un mémoire). Pour rappel, le théorème de Dirichlet assure qu'il y a une infinité de premiers dans chaque classe de congruence inversible modulo un entier fixé m. Le théorème prouvé dans le livre de Serre prouve mieux que ça : il y a "autant" de premiers dans chaque classe inversible modulo m. La preuve du théorème demande des connaissances sur les séries de Dirichlet, et donc de l'analyse complexe, mais aussi sur les caractères des groupes cycliques, enfin sur les fonction L mêlant les caractères et l'analyse complexe. Ce travail était très enrichissant et m'a permis de découvrir la théorie analytique des nombres, domaine de mon sujet de thèse.

A la fin de cette année, j'ai réussi le concours d'entrée en troisième année de l'ENS de Cachan. J'y ai alors préparé l'agrégation que j'ai obtenue en juillet 2013.

Troisième année (2013-2014)

Une fois mon agrégation assurée, je suis revenue comme prévu à la faculté d'Orsay, pour suivre la seconde année de master mention Analyse Arithmétique et Géométrie. Après de longues hésitations dans le choix des cours au premier semestre, j'ai suivi celui de *Géométrie Algébrique*, et celui de *Théorie des Nombres*.

J'ai été très enthousiasmée par la théorie des nombres et j'ai commencé mon mémoire assez tôt encadrée par Florent Jouve. Je me suis intéressée à la méthode de Stepanov pour estimer le nombre de points rationnels de courbes algébriques sur un corps fini (en simplifiant, le nombre de solutions dans un corps fini d'une équation à deux inconnues). Il se trouve que des théorèmes donnant les estimations de Stepanov avaient déjà été trouvés par Weil ou Hasse avant lui, mais la méthode de Stepanov est bien plus élémentaire que les preuves établies précédemment. En effet, elle n'utilise que des propriétés des polynômes sur les corps (lien entre degré et nombre de racines), et de l'algèbre linéaire (théorème du rang). De plus il existe des cas où la méthode de Stepanov donne un meilleur résultat que les estimations de Weil, c'est le cas notamment dans la partie 3 de mon mémoire, en annexe.

Au second semestre, sur les conseils de Florent Jouve, j'ai suivi les cours de *Marches aléatoires sur les groupes*, *Introduction à la théorie analytique des nombres* et *Cribles en théorie analytique des nombres*. Ces trois cours sont directement liés à mon sujet de thèse et me sont donc très profitables.

En juillet j'ai participé à l'école d'été de Théorie analytique des nombres organisée par l'Institut des Hautes Études Scientifiques. Ces deux semaines ont accéléré ma formation dans le domaine de la théorie analytique des nombres et me permettent de commencer ma thèse avec une assez large vision du domaine.

Présentation d'un domaine de recherche : La Théorie Analytique des Nombres

Théorie analytique des nombres

Les nombres premiers fascinent les hommes depuis qu'ils comptent et partagent. Depuis l'antiquité grecque, on se pose des questions sur les nombres entiers, sur les nombres premiers. Maintenant qu'on sait qu'il y a une infinité des nombres premiers, on peut se poser des questions sur leur répartition.

L'arithmétique est l'une des branches des mathématiques à avoir trouvé très tardivement des applications, et les applications qu'on en a reposent surtout sur le fait que l'on ne sait pas répondre à des questions. Ainsi la cryptologie reposant sur les nombres premiers a connu une grande avancée avec l'arrivée des technologies informatiques. La sécurité dépend du fait que l'on ne sait pas factoriser de grands nombres (c'est le cas pour la méthode RSA, on peut la voir expliquée dans [3, Chap. 1, Problème 1]).

Les questions de théorie des nombres peuvent être aisées à poser, mais beaucoup moins à résoudre. Ainsi en est-il de la conjecture de Goldbach : *tout nombre pair (> 2) peut s'écrire comme la somme de deux nombres premiers*, c'est un sujet dont on peut discuter à table, avec des enfants dès la sixième, mais il est difficile d'expliquer pourquoi on ne sait pas encore démontrer si c'est vrai.

Une autre conjecture résistante est celle des nombres premiers jumeaux : *il existe une infinité de premiers p tels que $p+2$ soit aussi premier.*

Le but de ma thèse n'est bien sûr pas de résoudre ces problèmes centenaires, mais j'espère obtenir les outils pour comprendre les avancées récentes sur le sujet. En effet on a le résultat suivant.

Théorème 1 (2013 Helfgott, [5])**.** *Tout entier impair (≥ 7) peut s'écrire comme la somme de trois nombres premiers.*

Ce théorème est une amélioration du résultat suivant que l'on peut trouver dans [13, Chap.3, th.3.4].

Théorème 2 (1937 Vinogradov)**.** *Il existe $N_0 \in \mathbb{N}$ tel que tout entier impair supérieur à N_0 peut s'écrire comme la somme de trois nombres premiers.*

Vinogradov donnait un N_0 non calculable, puis a été trouvé une borne de l'ordre de 10^{50000}, trop grande pour vérifier techniquement tous les impairs inférieurs. Le travail de Helfgott a été entre autres de ramener cette borne à une quantité suffisamment petite pour que les vérifications des impairs inférieurs puissent se faire par ordinateur.

Sur la conjecture des premiers jumeaux, on a aussi un résultat récent.

Soit p_n le n-ème nombre premier, on pose $H_1 = \liminf_{n\to\infty}(p_{n+1} - p_n)$. La conjecture des premiers jumeaux revient à dire $H_1 = 2$.

Théorème 3 (2014 Maynard, Zhang, Polymath8 et 8b)**.** *On a* $H_1 \leq 246$.

Ce résultat est le fruit de plusieurs améliorations du résultat de Zhang de 2013 montrant $H_1 \leq 7.10^7$, et du résultat de Maynard de la même année, qui a montré avec une autre méthode $H_1 \leq 600$. Il a été expliqué par Tao (instigateur des projets Polymath8 et 8b) lors de l'école d'été de 2014 [12].

On s'intéresse aux propriétés des nombres entiers et des structures liées aux entiers. Dans ma thèse, le centre d'intérêt est $\mathbb{Z}[X_1, \ldots, X_n]$ l'ensemble des polynômes à coefficients entiers. Mon sujet s'inscrit dans le domaine de la géométrie arithmétique, dont le but est de définir et d'étudier les lieux d'annulations d'ensembles de polynômes. Ce domaine a réellement pris de l'importance au vingtième siècle avec Deligne, Grothendieck et leurs collaborateurs qui ont fait le lien entre la géométrie et l'arithmétique, notamment en prouvant un analogue de l'hypothèse de Riemann pour les variétés algébriques sur les corps finis, pour cela la référence classique est [1].

Propriétés arithmétiques du nombre de $\mathbb{F}_q$-points des variétés algébriques

Une façon élémentaire de comprendre les variétés algébriques (définies du $\mathbb{Z}$) est de les voir comme des lieux d'annulation de systèmes de polynômes à coefficients entiers. On appellera un tel ensemble de zéros une variété affine (définie sur $\mathbb{Z}$). Alors une variété algébrique sera un "recollement" de plusieurs variétés affines.

On définit aussi les variétés projectives qui sont les lieux d'annulations de systèmes de polynômes homogènes, on les voit plongées dans un espace projectif.

Soit $(f_i(X_1, \ldots, X_n) \in \mathbb{Z}[X_1, \ldots, X_n])_{i\in I}$ une famille de polynômes à coefficients entiers, ils définissent une variété algébrique $V = \mathrm{Spec}(\mathbb{Z}[X_1, \ldots, X_n]/(f_i)_i)$ sur $\mathbb{Z}$.

Pour p un nombre premier, on peut regarder le système "modulo p", les équations sont alors dans le corps $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. On définit $V(\mathbb{F}_p)$ comme l'ensemble des solutions du système d'équations $f_i(x_1, \ldots, x_n) = 0$ dans le corps $\mathbb{F}_p$, on l'appelle ensemble des $\mathbb{F}_p$-points de V.

On peut aussi chercher des solutions dans une extension $\mathbb{F}_q$ avec q une puissance de p, on étudie alors $V(\mathbb{F}_q)$, ensemble des $\mathbb{F}_q$-points de V.

Ces ensembles sont finis puisque les corps dans lesquels on travaille sont finis, on peut donc s'intéresser à leur cardinal. Dans la suite, si X est un ensemble fini, on notera $|X|$ son cardinal.

Les sommes d'exponentielles indexées par $\mathbb{F}_q$ peuvent souvent être reliées au nombre de $\mathbb{F}_q$-points d'une variété algébrique. On souhaite pouvoir mieux estimer ces nombres afin d'avoir une meilleure approximation des sommes d'exponentielles.

De plus on peut généraliser le concept de sommes exponentielles en les indexant sur l'ensemble des points rationnels d'une variété algébrique. L'estimation fine d'une somme de ce type est l'un des points clés de la preuve récente de Zhang sur les écarts entre nombres premiers.

Dans un article de 2001, Fouvry et Katz [2, th.8.1] ont montré que l'on peut améliorer l'estimation de certaines sommes exponentielles indexées sur les points rationnels d'une variété algébrique. Cette amélioration dépend de la congruence modulo p du nombre de points $\mathbb{F}_q$-rationnels (q une puissance de p).

C'est donc à ce genre de propriétés arithmétiques que l'on veux s'intéresser.

On aimerait montrer que "de façon générique", $|V(\mathbb{F}_q)| \neq 0 \ [\text{mod } p]$.

Une diversité d'outils

Le principe de la théorie analytique des nombres est d'attaquer les problèmes arithmétiques grâce à des outils de l'analyse. Elle se distingue donc par une grande variété d'outils, allant de l'analyse complexe aux probabilités, sans tout de même négliger l'algèbre linéaire et la géométrie algébrique. En voici un avant-goût.

Analyse complexe

En théorie analytique des nombres, on doit souvent traiter avec des fonctions de la variable complexe, l'exemple le plus connu est celui de la fonction zêta de Riemann, $\zeta : s \mapsto \sum_{n\geq 1} n^{-s}$ pour $\text{Re}(s) > 1$.

On peut prolonger cette fonction de façon méromorphe au plan complexe, et s'intéresser à ses zéros (voir [9, Chap.VI, prop.10] pour un prolongement jusqu'à $\text{Re}(s) > 0$, et par exemple [8, Chap.II, prop.III.1] pour le plan complexe entier). L'hypothèse de Riemann conjecture que tous les zéros non triviaux de la fonction ζ sont sur la droite $\text{Re}(s) = \frac{1}{2}$.

Cette hypothèse demeure improuvée, mais on sait dire que les zéros sont tous "proches" de cette droite et cela suffit à prouver le résultat suivant (que l'on trouve par exemple dans [4, annexe C]), on note $\mathbb{P}$ l'ensemble des nombres premiers.

Théorème 4 (des nombres premiers). *Soit $\pi(x) = |\{p \in \mathbb{P}, p \leq x\}|$, on a*

$$\pi(x) \sim_{x\to\infty} \frac{x}{\log(x)}.$$

Ainsi l'analyse complexe nous donne une première information sur la répartition des nombres premiers. Le lien entre ζ et les nombres premiers se fait grâce à la formule d'Euler :

$$\zeta(s) = \prod_{p\in\mathbb{P}} \frac{1}{1-\frac{1}{p^s}}.$$

On peut ensuite généraliser la fonction ζ de Riemann, en voici deux exemples.

Les fonctions L de Dirichlet : Soit $q > 1$ un entier, on définit un caractère de Dirichlet modulo q comme une fonction $\chi : \mathbb{Z} \to \mathbb{C}$ vérifiant pour tout $m, n \in \mathbb{Z}$:

— $\chi(m + q) = \chi(m)$

— $\chi(m) = 0 \Leftrightarrow (q, m) > 1$

— $\chi(mn) = \chi(m)\chi(n)$.

Alors la fonction L associée à χ est $s \mapsto L(s, \chi) = \sum_{n\geq 1} \chi(n)n^{-s}$. Ces fonctions sont bien définies sur le demi-plan complexe $\text{Re}(s) > 1$.

Pour $q > 1$ entier, on a un caractère constant (dit trivial) $m \mapsto 1$ si $(q, m) = 1$ et 0 sinon. Alors

$$L(s, \chi) = \zeta(s) \times \left(\prod_{p \in \mathbb{P}, p|q} \frac{1}{1 - \frac{1}{p^s}} \right)^{-1},$$

c'est "presque" la fonction ζ.

Si χ n'est pas le caractère trivial, alors la fonction $L(., \chi)$ se prolonge holomorphiquement au demi-plan $\mathrm{Re}(s) > 0$.

L'étude de telles fonctions sert par exemple à montrer le résultat suivant dont on peut trouver une preuve dans [9, Chap.VI, th.2].

Théorème 5 (de la progression arithmétique)**.** *Soit $m > 1$ entier, $a \in (\mathbb{Z}/m\mathbb{Z})^*$, alors*

$$\{p \in \mathbb{P}, p \equiv a[\mathrm{mod}\ m]\}$$

est infini. Et même, il y a "autant" de premiers dans chaque classe inversible modulo m.

Les fonctions Z de variétés algébriques : Soit V une variété algébrique sur $\mathbb{F}_q$ (les équations sont définies sur $\mathbb{F}_q$), avec q une puissance d'un nombre premier. On s'intéresse en général au nombre de $\mathbb{F}_{q^n}$-points de V. Une bonne façon d'appréhender une suite de nombres est de regarder la série génératrice, mais il se trouve que l'on obtient de meilleurs résultats en regardant la fonction dont la dérivée logarithmique est la série génératrice. On définit donc

$$Z(V, T) = \exp\left(\sum_{n \geq 1} \frac{|V(\mathbb{F}_{q^n})|}{n} T^n \right).$$

Cette fonction ressemble à celle de Riemann si l'on remplace la variable T par q^{-s} avec s variable complexe. Il se trouve que cette fonction zêta est beaucoup mieux connue. On trouve par exemple dans [11, Chap.V, th.2.2] le résultat suivant.

Théorème 6 (Conjectures de Weil)**.** *Soit V une variété projective lisse de dimension n sur $\mathbb{F}_q$, alors*

Rationalité : $Z(V, T) \in \mathbb{Q}(T)$,

Équation fonctionnelle : Il existe un entier ϵ tel que

$$Z(V, \frac{1}{q^n T}) = \pm q^{n\epsilon/2} T^\epsilon Z(V, T),$$

Hypothèse de Riemann : On a une factorisation

$$Z(V, T) = \frac{P_1(T) \dots P_{2n-1}(T)}{P_2(T) \dots P_{2n}(T)}$$

où chaque P_i appartient à $\mathbb{Z}[T]$. De plus, $P_0(T) = 1 - T$, $P_{2n}(T) = 1 - q^n T$, et pour tout $1 \leq i \leq 2n - 1$, on a une factorisation dans $\mathbb{C}$:

$$P_i(T) = \prod_j (1 - \alpha_{ij} T)$$

avec $|\alpha_{ij}| = q^{i/2}$.

L'hypothèse de Riemann démontrée ici nous donne dans le cas des courbes ($n = 1$) que les zéros de la fonction $s \mapsto Z(V, q^{-s})$ sont tous sur la droite $\mathrm{Re}(s) = \frac{1}{2}$.

On peut alors déduire une formule pour le nombre de points rationnels des variétés projectives lisses :

$$|V(\mathbb{F}_q)| = q^n + \sum_{i=1}^{2n-1} \sum_j \alpha_{ij} + 1.$$

Géométrie algébrique

En fait cette formule découle d'un théorème plus compliqué utilisant la cohomologie ℓ-adique.

Si V est une variété algébrique sur $\mathbb{F}_q$, alors on a un $\mathbb{F}_q$-endomorphisme naturel $\mathrm{Frob}_q : V \to V$ appelé endomorphisme de Frobenius, qui vaut l'identité sur les points $\mathbb{F}_q$-rationnels de V. Dans le cas d'une variété affine, définie par des équations polynômiales $f_i(x_1, \ldots, x_n) = 0$ à coefficients dans $\mathbb{F}_q$, l'endomorphisme de Frobenius est :

$$\mathrm{Frob}_q : (x_1, \ldots, x_n) \mapsto (x_1^q, \ldots, x_n^q).$$

On définit ensuite les groupes de cohomologie ℓ-adique, pour ℓ un premier fixé différent de p ($q = p^e$), $H^i(V, \ell) = H^i_c(V \otimes_{\mathbb{F}_q} \overline{\mathbb{F}_q}, \mathbb{Q}_\ell)$, le c en indice indiquant la cohomologie à support compact, $\overline{\mathbb{F}_q}$ est une clôture algébrique fixée de $\mathbb{F}_q$ et $\mathbb{Q}_\ell$ est la complétion ℓ-adique[1] de $\mathbb{Q}$. Ce sont des $\mathbb{Q}_\ell$-espaces vectoriels de dimension finie, et pour $i > 2\dim(V)$, $H^i(V, \ell) = 0$.

Frob_q agissant sur V, il agit sur $H^i(V, \ell)$ pour tout i, on note $\mathrm{Frob}_q \mid H^i(V, \ell)$ l'endomorphisme correspondant à cette action. On a (par exemple dans [10, Th. 4.2]) le résultat suivant.

Théorème 7 (Grothendieck-Lefschetz). *Pour V une variété algébrique sur $\mathbb{F}_q$, ℓ un nombre premier ne divisant pas q,*

$$|V(\mathbb{F}_q)| = \sum_{i=0}^{2d} \mathrm{tr}(\mathrm{Frob}_q \mid H^i(V, \ell)).$$

Ce théorème lie la géométrie de la variété V via ses groupes de cohomologie, à son nombre de points rationnels. Des connaissances sur les dimensions de ces espaces vont nous donner une estimation du nombre de points. On a notamment pour les courbes le résultat suivant détaillé dans [6, chap.4].

Théorème 8 (Hasse-Weil). *Soit C une courbe projective lisse sur $\mathbb{F}_q$, alors*

$$||C(\mathbb{F}_q)| - (q + 1)| \leq 2g\sqrt{q}$$

où g est le genre de la courbe ($g = \dim H^1(C, \ell)$).

Sommes d'exponentielles

On s'intéresse dans ces formules à une somme de racines de l'unité ($\sum e^{i\theta}$). On a donc envie de savoir bien estimer des "sommes d'exponentielles", c'est le nom général que l'on donne à des sommes ou intégrales d'exponentielles ou de caractères.

1. le corps local $\mathbb{Q}_\ell$ est la complétion de $\mathbb{Q}$ pour la valeur absolue issue de la valuation ℓ-adique : $\frac{a}{b} \mapsto v_\ell(a) - v_\ell(b)$, où pour un entier $a = \ell^e m$, m premier à ℓ, $v_\ell(a) = e$

Méthode du cercle : Notons $e : t \mapsto \exp(2i\pi t)$.

L'idée de la méthode du cercle de Hardy et Littlewood est de se servir la relation d'orthogonalité

$$\begin{aligned} \int_0^1 e(ax)dx &= 1 \text{ si } a = 0 \\ &= 0 \text{ si } a \neq 0, \end{aligned} \tag{1}$$

ou pour $q \geq 1$

$$\begin{aligned} \frac{1}{q}\sum_{x=0}^{q} e(\frac{ax}{q}) &= 1 \text{ si } a \equiv 0 \text{ [mod } q] \\ &= 0 \text{ sinon.} \end{aligned}$$

Alors on peut écrire un nombre de solutions d'équation comme une somme d'exponentielles. On en a un exemple dans [13, Chap.1] pour résoudre le problème de Waring. On se pose la question de savoir combien de termes sont nécessaires pour écrire tout nombre comme somme de carrés, de cubes, de bicarrés... Posons $R(n,k,s)$ le nombre de façons d'écrire n comme somme de s puissances k-ièmes. On a

$$R(n,k,s) = \int_0^1 \left(\sum_{m=1}^{P} e(m^k x)\right)^s e(-nx)\ dx$$

pour P assez grand.

En effet

$$\begin{aligned} & \int_0^1 \left(\sum_{m=1}^{P} e(m^k x)\right)^s e(-nx)\ dx \\ = & \sum_{m_1=1}^{P} \dots \sum_{m_s=1}^{P} \int_0^1 e((m_1^k + \dots + m_s^k - n)x)\ dx \\ = & R(n,k,s) \end{aligned}$$

d'après la formule (1).

Il reste alors à estimer une intégrale. Cela peut se faire en utilisant la technique des arcs mineurs/arcs majeurs (voir [13, 2.1]). On coupe l'intégrale entre nombres proches d'un rationnel à petit dénominateur qui vont donner le terme principal de l'intégrale, et ceux qui sont plus loin qui donnent un terme de reste.

Et dans l'autre sens, cette méthode permet de faire le lien entre une variété algébrique et une somme d'exponentielles. On peut s'en servir aussi pour étudier des sommes d'exponentielles que l'on lie à des variétés algébriques dont on sait estimer le nombre de points.

On pourra en trouver plusieurs exemples dans mon mémoire de M2 [annexe C]. En effet, la méthode de Stepanov, grâce à des considérations d'algèbre linéaire, nous donne une estimation du nombre de points rationnels de courbes. On en déduit des estimations de sommes exponentielles, notamment les sommes de Kloosterman ($K(a,b;p) = \sum_{x=1}^{p-1} e(\frac{ax+bx^{-1}}{p})$) et les sommes de Heilbronn ($H_p(a) = \sum_{x=0}^{p-1} e(\frac{ax^p}{p^2})$).

Grand Crible

Enfin, le grand crible, comme présenté par Kowalski dans son livre [7], permet de démontrer des propriétés génériques. Les théorèmes récents de Zhang et Maynard reposent tous deux sur des idées de ce type. C'est en utilisant une méthode de grand crible que l'on espère pouvoir donner un résultat sur les variétés "génériques".

On étudie des objets "globaux" (disons en caractéristique nulle, ou dans un "grand" groupe) pour lesquels on veut voir si une propriété est "génériquement" vérifiée. On regarde alors la propriété "locale" (modulo ℓ, ou plus généralement modulo un sous-groupe d'indice fini) associée.

Une situation de crible est la donnée d'un triplet $(Y, \Lambda, (\rho_\ell)_{\ell\in\Lambda})$, où Y est un ensemble (souvent un groupe), Λ est un ensemble d'indices (souvent sous-ensemble des nombres premiers) et pour tout indice ℓ, $\rho_\ell : Y \to Y_\ell$ est une projection vers un ensemble fini. Pour $\mathcal{L}$ partie finie de Λ, on s'intéresse alors à des ensembles du type

$$\{y \in Y, y \notin \Omega\} \subset \{y \in Y, \forall \ell \in \mathcal{L}, \rho_\ell(y) \notin \Omega_\ell\},$$

où $\Omega_\ell = \rho_\ell(\Omega) \subset Y_\ell$ pour tout ℓ. S'il existe $C \geq 0$ telle que pour tout $\ell \in \Lambda$ on a $\frac{|\Omega_\ell|}{|Y_\ell|} > C$ alors on parle de grand crible.

On veut mesurer la taille de ces ensembles pour pouvoir dire que "beaucoup" de y sont dans Ω. On associe à la situation de crible un espace probabilisé (X, μ) et une application $F : X \to Y$ vérifiant pour tout ℓ dans Λ, $\rho(\ell) \circ F$ est une variable aléatoire (pour la mesure de comptage sur l'ensemble fini Y_ℓ). L'idée est de montrer que ces ensembles sont "exponentiellement petits" en fonction de $|\mathcal{L}|$, et donc pour un $x \in X$ "générique", pour tout $\ell \in \Lambda, \rho_\ell(F(x)) \in \Omega_\ell$.

Par exemple, on peut s'intéresser au nombre de points de courbes génériques sur $\mathbb{F}_p$. Définissons une famille de courbes à un paramètre, pour $f \in \mathbb{F}_p[X]$ fixé qui n'est pas un carré dans $\overline{\mathbb{F}_p}[X]$, on considère la famille $C_t : y^2 = f(x)(x-t)$ dans le plan. Alors le théorème de Hasse-Weil (pour les courbes affines) nous assure que $|C_t(\mathbb{F}_p)| = p + a_t$ avec $|a_t| \leq 2g\sqrt{p}$. Ainsi si p est assez grand, $|C_t(\mathbb{F}_p)| \equiv 0 \pmod p$ si et seulement si $|C_t(\mathbb{F}_p)| = p$. On peut alors étudier l'ensemble

$$\{t \in \mathbb{F}_p, a_t = 0\} \subset \{t \in \mathbb{F}_p, \forall \ell \leq L, a_t \equiv 0 \pmod \ell\}.$$

Ici on a pris $X = \mathbb{F}_q$ muni de la probabilité uniforme, $F : t \mapsto a_t \in \mathbb{Z}$, Λ est l'ensemble des premiers différents de p, $Y_\ell = \mathbb{Z}/\ell\mathbb{Z}$ et $\Omega_\ell = \mathbb{Z}/\ell\mathbb{Z} - \{0\}$.

On va obtenir des majorations dans le cas de grand crible grâce au résultat suivant [7, prop.2.3].

Théorème 9 (Inégalité de grand crible)**.** *Pour une situation de crible $(Y, \Lambda, (\rho_\ell)_{\ell\in\Lambda})$ et (X, μ, F), on a*

$$\mu(\{x \in X, \forall \ell \in \mathcal{L}, \rho_\ell(F(x)) \notin \Omega_\ell\}) \leq \Delta H^{-1}$$

où Δ est la "constante de grand crible" qui ne dépend que des ensembles X, Y, $\mathcal{L}$ et $(Y_\ell)_{\ell\in\mathcal{L}}$, et H est liée à la densité des Ω_ℓ pour $\ell \in \mathcal{L}$.

La constante de grand crible Δ est vue comme la norme d'un opérateur dans un espace vectoriel de dimension finie, son estimation se ramène souvent à un problème de sommes d'exponentielles. Tandis que H se déduit souvent d'un calcul de combinatoire dans Y_ℓ.

Marches aléatoires. Trouver un bon ensemble (X, μ) n'est pas toujours évident, par exemple si on veut étudier les variétés sur $\mathbb{Z}$, on préférerait pouvoir choisir $X = \mathbb{Z}$ qui n'a pas de mesure finie uniforme. L'idée est alors de considérer une marche aléatoire sur l'ensemble.

On définit une marche aléatoire sur $\mathbb{Z}$ par

$$\left\{ \begin{array}{rcl} X_0 & = & 0 \\ X_{k+1} & = & \xi_k + X_k \end{array} \right. ,$$

où pour tout k, $\xi_k \in \{-1, 1\}$ avec probabilité uniforme.

Cela nous fournit une famille de mesures de probabilité sur $\mathbb{Z}$ par $\mu_k(n) = P(X_k = n)$. On s'intéresse à la probabilité que X_k vérifie la propriété donnée :

$$P(F(X_k) \notin \Omega) \leq P(\forall l \leq L, \rho_l(F(X_k)) \notin \Omega_l),$$

on se retrouve ainsi dans une situation de crible. On voudra ensuite faire tendre k vers l'infini.

On aimerait un résultat pour "presque toute" variété, et pour "presque tout" premier. On a donc dans l'idée de construire un "double crible" imbriqué, ce qui n'a pas encore été fait et semble assez compliqué (pour utiliser trois ans de thèse).

Une première piste

Une façon de parler de "beaucoup" de premiers est donnée par la notion de densité. Soit A un sous-ensemble de l'ensemble $\mathbb{P}$ des nombres premiers, on définit

$$\text{dens-sup}(A) = \limsup_{x \to \infty} \frac{|\{p \in A, p \leq x\}|}{|\{p \in \mathbb{P}, p \leq x\}|}$$

et

$$\text{dens-inf}(A) = \liminf_{x \to \infty} \frac{|\{p \in A, p \leq x\}|}{|\{p \in \mathbb{P}, p \leq x\}|}.$$

Si ces deux quantités sont égales, on dit que l'ensemble A admet une densité (naturelle), leur valeur est appelée *densité* de A.

Il est clair qu'un ensemble qui admet une densité(-inférieure) non nulle est infini.

Dans un livre récent, Serre montre en utilisant (entre autres) la cohomologie l-adique le résultat suivant [10, th.6.3].

Théorème 10. *Soit V une variété algébrique, pour $m \geq 2$ entier, il existe sous-ensemble fini S_m de l'ensemble $\mathbb{P}$ des nombres premiers, tel que pour tout entier a,*

$$\{p \notin S_m, |V(\mathbb{F}_p) \equiv a \ [\text{mod } m]\}$$

est soit vide soit de densité strictement positive.

On obtient donc des renseignements arithmétiques sur le nombre de points rationnels des variétés et on pense pouvoir en déduire des résultats modulo p.

On obtient notamment un résultat pour les courbes en utilisant une version faible de la borne de Weil.

Corollaire 1. *Si C est une courbe algébrique vérifiant pour tout premier p de bonne réduction, $0 < |C(\mathbb{F}_p)| < 2p$, alors*

$$\{p \notin S_C, C(\mathbb{F}_p) \neq 0 \ [\text{mod } p]\}$$

est soit vide soit de densité inférieure strictement positive.

Ce résultat est un petit pas mais n'est bien sûr pas suffisant, on aimerait pouvoir dire que l'ensemble est de densité 1. Il faudrait aussi généraliser aux variétés de dimensions supérieures.

Enfin on voit qu'il suffit de trouver un premier pour en avoir une infinité, encore faut-il savoir comment trouver ce premier premier. On pourra commencer par chercher algorithmiquement sur des surfaces au hasard si l'on trouve facilement un premier qui ne divise pas le nombre de points rationnels associés.

Bibliographie

[1] P. Deligne. *Cohomologie étale*. Lect. Notes in Math. 569. Springer Verlag, 1977.

[2] E. Fouvry and N. Katz. A general stratification theorem for exponential sums, and applications. *Journal für die reine und angewandte Mathematik*, (540) :115–166, 2001.

[3] X. Gourdon. *Les maths en tête — Algèbre*. Ellipses, 2009.

[4] X. Gourdon. *Les maths en tête — Analyse*. Ellipses, 2009.

[5] H. A. Helfgott. Major arcs for Goldbach problem. Preprint disponible sur arXiv :1305.2897v4.

[6] B. Klinger. Etale cohomology and the Weil conjectures. notes de cours disponibles sur `http://webusers.imj-prg.fr/~bruno.klingler/cours/Weil.pdf`.

[7] E. Kowalski. *The large sieve and its applications*. Cambridge University Press, 2008.

[8] H. Queffélec and C. Zuily. *Analyse pour l'agrégation*. Dunod, 2007.

[9] J.-P. Serre. *Cours d'arithmétique*. PUF, 1970.

[10] J.-P. Serre. *Lectures on $N_X(p)$*. CRC Press, 2012.

[11] J. H. Silverman. *The Arithmetic of Elliptic Curves*. Graduate Texts in Mathematics 106. Springer-Verlag, 1986.

[12] T. Tao. Bounded gap between primes. In *École d'été de théorie analytique des nombres*, 2014. Disponible sur `https://www.youtube.com/watch?v=HUiSkxCGhCk&list=PLx5f8IelFRgEXleMSOdwidL2P5GTlL0uY`.

[13] R.C. Vaughan. *The Hardy-Littlewood method*. Cambridge University Press, 1981.

Annexe A

Travail de Recherche de première année :
Le troisième problème de Hilbert

Le troisième problème de Hilbert

Comment calculer le volume d'une pyramide ? Tout le monde connait la formule "$\frac{base \times hauteur}{3}$", mais sa démonstration n'est pas triviale, et nécessite un passage à la limite. Est-ce juste que l'on n'a pas encore les moyens d'éviter le passage à la limite ou est-ce vraiment impossible ?

Bolyai et Gerwien ont montré que l'aire d'un polygone peut toujours se calculer de façon simple, c'est-à-dire sans processus infini. En effet, on peut découper le polygone en un nombre fini de plus petits polygones et les réassembler pour obtenir un carré. Et l'aire du carré est connue et calculable sans passage à la limite. Le troisième problème que posa Hilbert, lors de sa fameuse intervention au Congrès International de 1900 peut se traduire ainsi : Peut-on calculer le volume de la pyramide en se ramenant au cube par découpages ?

En 1896, Bricard avait déjà trouvé que le tétraèdre régulier (pyramide dont toutes les faces sont des triangles équilatéraux) ne peut pas devenir un cube par découpages et recollements. Il montra ainsi que le calcul du volume de cette pyramide ne peut pas se faire simplement. Mais sa preuve se révèla lacunaire. Dehn trouva en 1902 une preuve exacte à l'hypothèse de Bricard, mais sa rédaction n'était pas très accessible. La preuve fut beaucoup simplifiée et améliorée pendant les années qui suivirent. Nous allons ici exposer une preuve de ce résultat.

1 Définitions

Définition 1 (polyèdre)**.**
Un polyèdre *est l'enveloppe convexe d'un nombre fini de points non-coplanaires de l'espace* $\mathbb{R}^3$

Ainsi le cube et le tétradèdre sont deux polyèdres.

Définition 2 (congruence)**.**
Deux polyèdres sont dits congruents *s'il existe une isométrie qui fait passer de l'un à l'autre.*

Définition 3 (équidécomposabilité)**.**
Deux polyèdres A et B sont dits équidécomposables en polyèdres *(ou équidécomposables) s'il existe $n \in \mathbb{N}^*$ tel que $A = A_1 \cup \cdots \cup A_n$ et $B = B_1 \cup \cdots \cup B_n$, où :*

- *les A_i sont des polyèdres deux à deux d'intérieur disjoints ;*
- *les B_i sont des polyèdres deux à deux d'intérieur disjoints ;*
- *pour tout i dans $\{1, \cdots, n\}$, A_i est congruent à B_i.*

Deux polyèdres congruents ont le même volume par définition de l'isométrie, donc deux polyèdres équidécomposables ont le même volume. La question est de savoir si deux polyèdres de même volume sont équidécomposables. Nous allons

montrer que ce n'est pas toujours le cas : le tétraèdre régulier et le cube ne sont pas équidécomposables.

Définition 4 (angle dièdre)**.**
*Pour e une arète d'un polyèdre, l'*angle dièdre *lui correspondant est l'angle entre les demi-droites définies par l'intersection d'un plan perpendiculaire à e avec les deux faces adjacentes à cette arète. Un angle dièdre est défini modulo π (c'est un élément de $\mathbb{R}/\pi\mathbb{Z}$).*

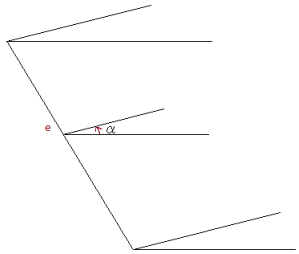

FIGURE 1 – angle dièdre.

2 L'invariant de Dehn

Pour M ensemble fini d'éléments de $\mathbb{R}$, on notera $V(M)$ le $\mathbb{Q}$ espace vectoriel engendré par M :

$$V(M) = \{ \sum_{finies} q_i x_i, x_i \in M, q_i \in \mathbb{Q}\}.$$

Définition 5 (L'invariant de Dehn)**.**
Soient
- *A un polyèdre ;*
- *$l_1, \cdots, l_k$ les longueurs de ses arètes ;*
- *$\alpha_1, \cdots, \alpha_k$ les angles dièdres correspondants ;*
- *$M = \{\alpha_1, \cdots, \alpha_k, \pi\}$ l'ensemble contenant π et les angles dièdres de A ;*
- *f une forme linéaire sur $V(M)$, telle que $f(\pi) = 0$.*

Alors l'invariant de Dehn de A relativement à f *est :*

$$\mathrm{D}_f(A) = l_1 f(\alpha_1) + \cdots + l_k f(\alpha_k).$$

Le choix de f telle que $f(\pi) = 0$ permet de bien définir f sur un ensemble d'angles dièdres.

Théorème 1 (Dehn-Hadwiger)**.**
Soient A et B deux polyèdres et M un ensemble contenant π et les angles dièdres de A et B.
S'il existe f une forme linéaire sur $V(M)$ telle que $f(\pi) = 0$ vérifiant $\mathrm{D}_f(A) \neq \mathrm{D}_f(B)$ alors A et B ne sont pas équidécomposables en polyèdres.

L'idée principale vient du fait que l'invariant de Dehn est préservé par la décomposition en polyèdres. Commençons par étendre la forme linéaire à un décomposition en polyèdres.

Sous-Lemme 1.
Si $\bar{M}$ contient M et f est une forme linéaire sur $V(M)$, on peut l'étendre en une forme linéaire sur $V(\bar{M})$.

Démonstration.
Comme $\bar{M}$ contient M, $V(M)$ est un sous-espace vectoriel de $V(\bar{M})$. Et on connait l'image de f sur la base de $V(M)$, il nous suffit alors de compléter la base de $V(M)$ en une base de $V(\bar{M})$ et de choisir de façon arbitraire l'image par f des vecteurs ajoutés. On obtient ainsi le prolongement souhaité. □

Lemme 1. *Soit P un polyèdre et $P = P_1 \cup \cdots \cup P_n$ une décomposition de P en polyèdres. On note M l'ensemble contenant π et les angles dièdres de P, et $\bar{M}$ l'ensemble contenant M et tous les angles dièdres des polyèdres de la décomposition.*

Soit f une forme linéaire sur $V(M)$, que l'on étend sur $V(\bar{M})$ en utilisant le sous-lemme 1. On peut alors calculer les invariants de Dehn de P et de tous les P_i relativement à f, et :

$$\mathrm{D}_f(P) = \mathrm{D}_f(P_1) + \cdots + \mathrm{D}_f(P_n).$$

Démonstration.
Dans la décomposition de P en polyèdres, on peut considérer tous les "bouts d'arète" e, qui sont limités par les sommets de polyèdres ou les intersections d'arètes de polyèdres. Il y a un nombre fini de bouts d'arètes car chaque polyèdre a un nombre fini d'arètes et de sommets.

Chaque "bout d'arète" e contenu dans une arète de P, contribue à la somme $\sum_{boutsd'arètesdeP} l(e) \times f(\alpha(e))$ avec la valeur $l(e) \times f(\alpha(e))$ où l donne la longueur d'un segment et $\alpha(e)$ donne l'angle dièdre associé à l'arète contenant le "bout d'arète" e.

Un "bout d'arète" peut être commun à plusieurs P_i, notons alors $\alpha_i(e)$ l'angle dièdre associé à e dans P_i (qui vaut zéro si e n'est pas un bout d'arète de P_i). En utilisant linéarité de f, la contribution du bout d'arète e dans $\sum_{i=1}^{n} \sum_{boutsd'arète} l(e) \times f(\alpha_i(e))$ sera

$$l(e) \times (f(\alpha_1(e)) + \cdots + f(\alpha_n(e))) = l(e) \times (f(\alpha_1(e) + \cdots + \alpha_n(e))).$$

Et on calcule que :

– si e est contenu dans une arête de P (cas a), la somme des angles dièdres qui lui sont associés vaut $\alpha(e)$, sa contribution sera alors $l(e) \times f(\alpha(e))$;
– si e est contenu dans l'intérieur d'une face de P (cas b), la somme des angles dièdres qui lui sont associés vaut π, sa contribution sera alors $l(e) \times f(\pi) = 0$;
– si e est contenu dans l'intérieur de P (cas c et d), la somme des angles dièdres qui lui sont associés vaut π si le bout d'arête est contenu dans la face d'un polyèdre et 2π sinon. Sa contribution sera donc $l(e) \times f(2\pi) = 0$ ou $l(e) \times f(\pi) = 0$.

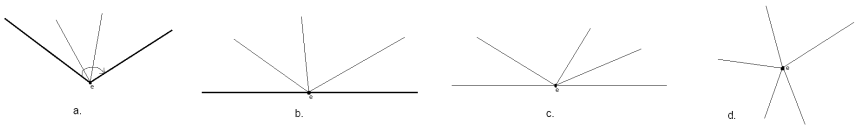

FIGURE 2 – Différents cas pour la localisation du bout d'arête e, vus dans le plan perpendiculaire à e.

Cela permet d'écrire :

$$\sum_{bouts d'arête de P} l(e) \times f(\alpha(e)) = \sum_{i=1}^{n} \sum_{bouts d'arête} l(e) \times f(\alpha_i(e)).$$

Or on considère tous les "bouts d'arête" donc en sommant les longueurs des "bouts d'arête" contenus dans une unique arête de P, on obtient exactement la longueur de cette arête et ces "bouts d'arête" ont le même angle dièdre associé :

$$\sum_{e \subset a} l(e) \times f(\alpha(e)) = \sum_{e \subset a} l(e) \times f(\alpha(a)) = l(a) \times f(\alpha(a)).$$

D'où, en sommant sur les arêtes de P :

$$\mathrm{D}_f(P) = \sum_{bouts d'arêtes de P} l(e) \times f(\alpha(e)).$$

Et de la même façon, pour $i \in 1, \cdots, n$, par construction des bouts d'arête, toute arête de P_i est constitué d'un nombre entier non nul de bouts d'arête, on obtient :

$$\mathrm{D}_f(P_i) = \sum_{bouts d'arêtes} l(e) \times f(\alpha_i(e)).$$

Finalement,

$$\mathrm{D}_f(P) = \mathrm{D}_f(P_1) + \cdots + \mathrm{D}_f(P_n).$$

□

Nous sommes désormais en mesure de prouver le théorème 1 de Dehn-Hadwiger.

Démonstration.
Soit M l'ensemble contenant π et les angles dièdres des polyèdres A et B. Soit f une forme linéaire définie sur $V(M)$, telle que $f(\pi) = 0$ et $\mathrm{D}_f(A) \neq \mathrm{D}_f(B)$. Raisonnons par l'absurde, en supposant que les polyèdres A et B sont équidécomposables en polyèdres.

Alors, on peut écrire $A = A_1 \cup \cdots \cup A_n$ et $B = B_1 \cup \cdots \cup B_n$ avec, pour tout i A_i congruent à B_i.

Soit $\bar{M}$ l'ensemble composé de π et de tous les angles dièdres qui apparaissent dans la décomposition ($\bar{M}$ est fini). En étendant f à $V(\bar{M})$ grace au sous-lemme 1 on peut écrire pour tout i, $\mathrm{D}_f(A_i) = \mathrm{D}_f(B_i)$ car les deux polyèdres sont congruents.

On a donc en sommant pour i de 1 à n et en utilisant le lemme 1, $\mathrm{D}_f(A) = \mathrm{D}_f(A_1) + \cdots + \mathrm{D}_f(A_n) = \mathrm{D}_f(B_1) + \cdots + \mathrm{D}_f(B_n) = \mathrm{D}_f(B)$ ce qui contredit l'hypothèse. Donc A et B ne sont pas équidécomposables en polyèdres. □

3 L'invariant de Dehn dans le bon espace

On peut voir l'invariant de Dehn comme un élément du produit tensoriel :

$$\mathbb{R} \otimes_{\mathbb{Q}} (\mathbb{R}/\pi\mathbb{Q}) = \{\sum_{i=1}^{n} l_i \otimes \alpha_i, n \in \mathbb{N}, l_i \in \mathbb{R}, \alpha_i \in \mathbb{R}/\pi\mathbb{Q}\}.$$

En effet, la forme linéaire f s'annule en π, on peut donc la factoriser par $\pi\mathbb{Q}$ -le sous-espace vectoriel de $\mathbb{R}$ vu comme $\mathbb{Q}$-espace vectoriel engendré par π- et ainsi considérer les angles dièdres comme des éléments de $\mathbb{R}/\pi\mathbb{Q}$. Alors que les longueurs des arètes ou des bouts d'arètes sont des nombres réels. Et les propriétés du produit tensoriel traduisent bien les propriétés de l'invariant de Dehn.

Pour $l \in \mathbb{R}; \alpha, \beta \in \mathbb{R}/\pi\mathbb{Q}$, l'égalité $l \otimes \alpha + l \otimes \beta = l \otimes (\alpha + \beta)$ traduit le fait que l'on peut sommer les angles dièdres correspondants au même bout d'arète sur les différents polyèdres.

Pour $k, l \in \mathbb{R}; \alpha \in \mathbb{R}/\pi\mathbb{Q}$, l'égalité $k \otimes \alpha + l \otimes \alpha = (k + l) \otimes \alpha$ traduit la sommation des bouts d'arètes qui composent une arète du polyèdre, pour obtenir la contribution de cette arète.

Cela permet donc de comprendre plus précisément l'invariance par décomposition en polyèdres. Nous pouvons alors donner une nouvelle preuve du théorème de Dehn-Hadwiger, en utilisant l'invariant de Dehn comme un élément du produit tensoriel.

Définition 6 (L'invariant de Dehn, nouvelle définition)**.**
*Soient A un polyèdre ; $l_1, \cdots, l_k$ les longueurs de ses arètes et $\alpha_1, \cdots, \alpha_k$ les angles dièdres correspondants vus comme éléments de $\mathbb{R}/\pi\mathbb{Q}$. L'*invariant de Dehn de A *est alors :*

$$\mathrm{D}(A) = \sum_{i=1}^{k} l_i \otimes \alpha_i.$$

Lemme 2.
Soit P un polyèdre et $P = P_1 \cup \cdots \cup P_n$ une décomposition de P en polyèdres. Alors,

$$\mathrm{D}(P) = \mathrm{D}(P_1) + \cdots + \mathrm{D}(P_n).$$

Démonstration.
Soit P un polyèdre, écrivons $P = P_1 \cup \cdots \cup P_n$ une décomposition de P en polyèdres. Chaque morceau d'arète e de P contribue à $\mathrm{D}(P)$ avec la valeur $l(e) \otimes \alpha(e)$ où l donne la longueur d'un segment.

Dans la décomposition de P en polyèdres, on peut considérer tous les "bouts d'arètes" e, définis de la même façon que précedemment. On note de la même façon $\alpha_i(e)$ l'angle dièdre associé à e dans P_i (qui vaut zéro si e n'est pas un bout d'arète de P_i). Sa contribution dans $D(P_1) + \cdots + \mathrm{D}(P_n)$ sera

$$l(e) \otimes \alpha_1(e) + \cdots + l(e) \otimes \alpha_n(e) = l(e) \otimes (\alpha_1(e) + \cdots + \alpha_n(e))$$

par définition de la somme d'éléments du produit tensoriel. Et de la même façon,

- si e est contenu dans une arète de P, la somme des angles dièdres qui lui sont associés vaut $\alpha(e)$, sa contribution sera alors $l(e) \otimes \alpha(e)$;
- si e est contenu dans l'intérieur d'une face de P, la somme des angles dièdres qui lui sont associés vaut $\pi = 0[\pi]$, sa contribution sera alors $l(e) \otimes 0 = 0.(l(e) \otimes 1) = 0_{\mathbb{R} \otimes_{\mathbb{Q}} (\mathbb{R}/\pi\mathbb{Q})}$;
- si e est contenu dans l'intérieur de P, la somme des angles dièdres qui lui sont associés vaut $\pi = 0[\pi]$ si le bout d'arète est contenu dans la face d'un polyèdre et $2\pi = 0[\pi]$ sinon. Sa contribution sera donc $l(e) \otimes 0 = 0$.

Finalement, chaque "bout d'arète" contribue de la même façon dans les deux termes : on a donc bien l'égalité

$$\mathrm{D}(P) = D(P_1) + \cdots + \mathrm{D}(P_n).$$

□

Cela permet de démontrer de nouveau le théorème exactement de la même façon. Nous pouvons observer que le produit tensoriel est le bon espace pour comprendre l'invariant de Dehn : la preuve est plus claire et ne s'appuie que sur les propriétés de cet espace.

4 Un résultat utile

Lemme 3.
Si $\cos(\psi) = \frac{1}{3}$ alors $\frac{\psi}{\pi}$ n'est pas rationnel.

Commençons à prouver par récurrence le

Sous-Lemme 2.
Pour $n \in \mathbb{N}$, $\cos(nt)$ est un polynôme à coefficients entiers en $\cos(t)$ de degré n, et de coefficient dominant 2^{n-1} si $n \geq 1$.

Démonstration.
L'initialisation est simple puisque $\cos(0 \times t) = 1 = \cos(t)^0$ et $\cos(1 \times t) = 2^0 \cos(t)$.

Supposons que l'affirmation soit vraie pour un certain $n \in \mathbb{N}, n \geq 1$ et pour $n-1$. En sommant les formules trigonométriques :

$$\cos((n+1)t) = \cos(nt)\cos(t) - \sin(nt)\sin(t)$$

$$\cos((n-1)t) = \cos(nt)\cos(t) + \sin(nt)\sin(t)$$

on obtient

$$\cos((n+1)t) = 2\cos(t)\cos(nt) - \cos((n-1)t)$$

$\cos((n+1)t)$ est donc un polynôme à coefficients entiers en $\cos(t)$ de degré $n+1$ et de coefficient dominant $2 \times 2^{n-1} = 2^n$. On obtient ainsi ce qui avait été annoncé. □

On notera T_n le polynôme vérifiant $T_n(\cos(t)) = \cos(nt)$.

Sous-Lemme 3.
Soit P un polynôme à coefficients entiers, de coefficient dominant $a \in \mathbb{N}^*$. Soient $q \in \mathbb{N}^*$ et $p \in \mathbb{Z}$ tels que p et q sont premiers entre eux. Si $\frac{p}{q}$ est une racine de P, alors q divise a.

Démonstration.
Écrivons $P[X] = aX^n - \sum_{i=0}^{n-1} a_i X^i$. On a alors :

$$a(\frac{p}{q})^n - \sum_{i=0}^{n-1} a_i (\frac{p}{q})^i = 0.$$

En multipliant par q^n on obtient :

$$ap^n = \sum_{i=0}^{n-1} a_i p^i q^{n-i}.$$

q divise le terme de droite donc il divise le terme de gauche. Comme il est premier avec p il l'est aussi avec p^n, donc q divise a. □

Nous pouvons maintenant prouver le lemme 3 :

Démonstration.
Raisonnons par l'absurde en supposant que $\frac{\psi}{\pi}$ est dans $\mathbb{Q}$. Il existe donc $p \in \mathbb{Z}, q \in \mathbb{N}^*$ avec p et q premiers entre eux tels que $\frac{\psi}{\pi} = \frac{p}{q}$. Alors

$$T_q(\frac{1}{3}) = T_q(\cos(\psi)) = T_q(\cos(\frac{p\pi}{q})) = \cos(p\pi) = \pm 1$$

c'est-à-dire que $\frac{1}{3}$ est racine du polynôme à coefficients entiers $T_q \mp 1$ de coefficient dominant 2^{q-1}. Mais comme 3 ne divise pas 2^{q-1}, on obtient une contradiction. Donc $\frac{\psi}{\pi}$ n'est pas rationel. □

5 Le cube et le tétraèdre régulier ne sont pas équidécomposables

Démonstration.
Commençons par calculer l'angle dièdre entre les faces du tétraèdre régulier. Nommons abcd le tétraèdre régulier de coté 1.

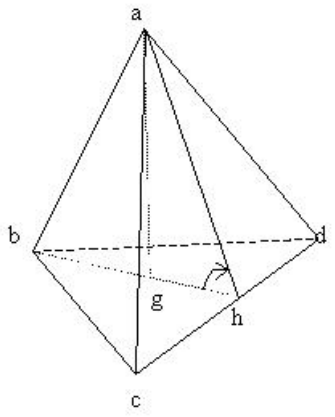

FIGURE 3 – Le tétraèdre régulier.

La hauteur issue du sommet a coupe la face (bcd) en son centre de gravité g, (bg) est perpendiculaire à (cd) et la coupe en h (milieu de (cd)). L'angle dièdre du tétraèdre a donc pour cosinus $\frac{gh}{ah} = \frac{1}{3}$.

Nommons c la longueur du coté du cube de même volume que le tétraèdre. On a

$$\mathrm{D}(cube) = \sum_{k=1}^{12} c \otimes \frac{\pi}{2} = \frac{12}{2} \cdot c \otimes \pi = 6 \cdot c \otimes 0 = 0$$

et

$$\mathrm{D}_f(tétraèdre) = \sum_{k=1}^{6} 1 \otimes \psi = 6 \otimes \psi.$$

Or nous avons vu grâce au lemme 3 que $\psi = arcos(\frac{1}{3})$ n'est pas un multiple rationnel de π. Donc $\mathrm{D}_f(tétraèdre)$ est un élément non-nul de $\mathbb{R} \otimes_{\mathbb{Q}} (\mathbb{R}/\pi\mathbb{Q})$ donc différent de $\mathrm{D}_f(cube)$.

D'après le théorème de Dehn-Hadwiger ceci implique que le cube et le tétraèdre ne sont pas équidécomposables en polyèdres. □

Conclusion

Nous avons ainsi montré que le volume du tétraèdre régulier ne peut pas se calculer par découpages et recollements en se ramenant au cube. Nous n'avons donc pas trouvé de moyen d'éviter le processus infini pour le calcul de certains volumes. Cela nous incite à penser que le volume du polyèdre est une notion beaucoup plus compliquée que celle de l'aire d'un polygone.

Cette intuition, pointée du doigt par Hilbert en 1900, sera renforcée une vingtaine d'années plus tard par les résultats de Hausdorff puis de Banach et Tarski. Ces mathématiciens ont en effet démontré (en utilisant l'axiome du choix) que la boule unité de l'espace $\mathbb{R}^3$ est équidécomposable à deux fois elle même (mais avec des morceaux qui ne sont plus des polyèdres). Un tel résultat permettrait de montrer si la notion de volume était aussi simple que le volume de la boule est égal au volume de deux fois cette boule, d'où le nom de paradoxe de Banach et Tarski. Ils montrent aussi qu'une telle manipulation ne peux pas s'effectuer dans le plan où l'aire peut être définie de façon plus simple.

Historique : Calcul des aires et volumes

III^{ème} siècle avant J.C. Euclide calcule le volume de la pyramide par un (joli) processus infini.

1830 Bolyai et Gerwien montrent que deux polygones de même aire sont toujours équidécomposables en polygones.

1896 Bricard donne une preuve fausse de la non-équidécomposabilité du cube et du tétraèdre régulier.

1900 Hilbert expose ses 23 problèmes du siècle, le troisième demande de trouver deux polyèdres qui ne sont pas équidécomposables en polyèdres.

1902 Dehn répond au troisième problème de Hilbert : le cube et le tétraèdre régulier ne sont pas équidécomposables en polyèdres.

1914 Hausdorff montre que la sphère de l'espace à laquelle on retire un ensemble dénombrable de points est équidécomposable à deux fois elle-même.

1924 Banach et Tarski reprennent et rendent plus remarquable le résultat de Hausdorff : la boule unité est équidécomposable à deux fois elle même.

Bibliographie

- *A New Approach to Hilbert's third problem*, David Benko, The Mathematical Association of America,Octobre 2007, p.665-676.
- *Raisonnements divins*, Chapitre 8, M.Aigner et G.M.Ziegler
- *Hilbert's Third Problem*, Vladimir G. Boltianski, V. H. Winston & Sons, 1978.
- *Algèbre*, chapitre 16, S.Lang, Dunod, 2004.
- *Aires et volumes : découpage et recollement*, Daniel Perrin, Images des mathématiques, 2010.

Annexe B

Rapport de stage de première année : Application des mathématiques en Cardiobiologie

Rapport de stage
Modélisation du second messager au laboratoire de l'INSERM - Signalisation et Physiopathologie Cardiaque - U769

Lucile Devin

juin 2011

1 Mes attentes concernant le stage

N'ayant pas de proches dans le monde de la recherche, avant cette première expérience en laboratoire, je n'avais pas beaucoup d'idée de ce à quoi ressemblent le travail de chercheur et les mathématiques appliquées. Je ne savais donc pas vraiment à quoi m'attendre et j'avais peur de ne pas être tout à fait au niveau, de devoir créer un programme de résolution d'équations différentielles à partir de pas grand chose (ce que j'ai un peu commencé à faire la première semaine), alors que je ne suis pas très à l'aise en programmation, et je craignais aussi de ne pas comprendre ce que je faisais au niveau du sens biologique.

2 L'organisation du laboratoire

L'unité 769 de l'INSERM, est répartie en trois équipes d'une quinzaine de chercheurs sur les étages 2, 4 et 5 de la tour D4 de la faculté de pharmacie de Chatenay-Malabry. Le directeur du laboratoire, Rodolphe Fischmeister est responsable des trois équipes, il est souvent en réunion, soit avec des chercheurs du laboratoire, soit à l'extérieur pour d'autres projets, ou en conférence. Chaque équipe est dirigée par un directeur de recherche, il y a ensuite les chargés de recherche comme mon maitre de stage, Grégoire Vandecasteele (mais il a passé le concours de directeur de recherche pendant mon stage), des maitres de conférence, enseignants-chercheurs, médecins, post-doc et doctorants comme Zeineb Haj Slimane avec qui je travaille pendant mon stage. Jean-Luc Mazet était le mathématicien du laboratoire, il a déjà fait un modèle de la cellule cardiaque pour comprendre le rôle des phosphodiestérases en 2008, il revient au laboratoire pour m'aider à comprendre la modélisation, il participe beaucoup à la conception du modèle.

Les horaires de travail sont assez libres, les chercheurs arrivent pour la plupart entre 9h et 9h30, certains préfèrent arriver dès l'ouverture du laboratoire à 7h, on m'a dit que ceux du deuxième étage arrivent plutôt autour de 10h. Cela peut aussi dépendre du travail que l'on projette d'effectuer, certaines expériences durent longtemps, il faut mieux arriver tôt les jours où l'on veut faire de telles expériences. Les horaires de départ varient de la même manière, et toujours selon le travail et les réunions, on sort rarement avant 17h30.

Les bureaux sont répartis entre les salles d'expériences, ce sont des bureaux pour deux ou trois personnes. Jean-Luc et moi étions installés dans le bureau de Grégoire qu'il partage avec un autre chercheur. Les biologistes travaillent dans leur bureau environ un jour sur deux, lorsqu'il faut analyser les données fournies par les expériences de la veille avec l'ordinateur, ils en ont beaucoup plus besoin lors de la rédaction d'article.

Il y a une réunion du laboratoire tous les lundis matin à 9h30, et une réunion d'équipe à peu près tout les quinze jours. Elles ont lieu dans une grande salle au cinquième étage qui sert aussi de kitchenette pour ceux qui mangent au laboratoire, pour les réceptions et pour boire un café.

La réunion du laboratoire commence par un bilan financier, puis parfois une concertation sur le matériel à acheter, des remarques administratives, les nouvelles comme « Patrick est passé hors-classe ». Ensuite un ou deux membres du laboratoire doivent faire un exposé, par exemple résumer une conférence à laquelle les autres n'ont pas assisté, raconter un article en cours d'écriture, expliquer les recherches qu'il fait. Parfois des chercheurs d'autres laboratoires viennent faire une conférence, cela remplace alors les exposés, lorsque l'intervenant est étranger la conférence se fait en anglais, il faut non seulement connaitre des bases sur le sujet exposé mais aussi maitriser la langue si on veut comprendre quelque chose.

Pendant la réunion d'équipe, chacun présente les résultats qu'il a obtenus, et surtout les problèmes qu'il a rencontré pour pouvoir les résoudre ensemble. Cette réunion se fait en plus petit comité que la réunion du laboratoire, cela permet à chacun d'exposer l'avancement de ses travaux, et l'on peut mieux s'étendre sur chaque cas.

Pendant mon stage nous avons mis en place des réunions de modélisation environ une fois par semaine. Nous nous retrouvions avec Rodolphe, Grégoire, Jean-Luc et Zeineb, pour discuter des résultats de Zeineb, voir où j'en étais dans le modèle, savoir si tout se passait bien. Je pouvais poser des questions pour avoir des précisions sur certaines réactions que je devais modéliser et que je ne comprenais pas, ou demander des précisions sur des résultats. Bien sûr je pouvais aussi poser des questions hors de ces réunions à Grégoire, Jean-Luc ou Zeineb.

3 Petit nécessaire sur le fonctionnement de la cellule cardiaque

Grégoire a commencé à m'expliquer dès mon arrivée le fonctionnement de la cellule cardiaque pour que je puisse assimiler ce qui allait se dire en réunion du laboratoire juste après. Voici à peu près ce qui sert à comprendre ce sur quoi j'ai travaillé durant mon stage, pour les non biologistes qui comme moi ont arrété les cours de « sciences de la vie et de la terre » en terminale.

La cellule cardiaque a une particularité importante, elle doit se contracter de façon régulière. Pour cela l'acyclase (AC) fixé sur la membrane plasmique fabrique à partir d'« énergie » (ATP) de l'AMP cyclique qui va elle même se fixer sur une protéine kinase AMPc-dépendante (PKA). La PKA est composée d'une sous-unité régulatrice et d'une sous-unité catalytique. Lorsque deux AMP cyclique sont fixées sur la sous-unité régulatrice, la sous-unité catalytique se détache et va entre-autre activer les canaux calciques (le calcium est responable de la contraction), et diffuser dans le noyau où il peut agir sur l'expression des gènes. Les phosphodiestérases (PDE, la spécialité du laboratoire) inhibent l'AMP cyclique, la cellule peut ainsi se décontracter.

Lorsque l'on a peur ou que l'on produit un effort, le cœur bat plus rapidement et plus fort. Le système nerveux envoie de la noradrénaline, une neuro-hormone

proche de l'adrénaline, elle sert de messager pour activer les battements du coeur. L'isoprénaline (Iso), dont on se sert plutôt en laboratoire a le même effet, elle se fixe sur un récepteur β-adrenergique (Rb1). Celui-ci dissocie la protéine Gs dont un morceau va se fixer à l'acyclase pour augmenter son action de production d'AMP cyclique. l'AMP cyclique sert alors de second messager, dans la cellule, pour augmenter les battements cardiaques.

Le sujet de thèse de Zeineb concerne l'action de la sous-unité catalytique de la PKA dans le noyau de cellules cardiques de rat adulte. Elle se sert pour situer ces sous-unités de protéines AKAR. Elles sont fabriquées de telle sorte qu'elles changent de conformation lorsqu'elles sont phosphorylées (AKAR-P) par la sous-unité catalytique, et grâce à de la fluorescence, on peut mesurer le quotient de protéines phosphorylées sur les protéines non-activées. La phosphatase (PP1) déphosphoryle AKAR-P. On peut aussi adresser ces protéines dans le cytoplasme, AKAR-nes, ou dans le noyau, AKAR-nls, de la cellule, et ainsi différencier les deux lieux et mesurer indépendamment les fluorescences.

4 Carnet de bord

Semaine 1 : Lecture d'articles, prise en main de Python

J'ai commencé la première semaine par lire beaucoup d'articles consacrés au sujet de recherche et à la modélisation. J'avais accès grâce à internet à un très grand nombre de publications (en anglais bien sûr) dans des journaux scientifiques reconnus, et il fallait sélectionner ceux qui pouvaient être liés à mon sujet, Grégoire m'a donné les premiers et indispensables à lire, et il m'aidait par la suite à trouver les bons mots-clés pour trouver les articles intéressants. Il fallait surtout que je trouve des équations utilisés par d'autres pour modéliser les réactions qui nous intéressaient et trouver les bonnes constantes.

L'idée de départ était de pouvoir rentrer toutes ces équations dans un logiciel de programmation (Python) et de le faire résoudre pour obtenir des courbes de concentration des espèces en fonction du temps. Je me suis donc en même temps initiée à ce langage.

Pendant la première réunion de modélisation, le premier jour, Zeineb nous a présenté ses résultats : dans le cas normal, elle observe une action de la sous-unité catalytique de la PKA importante dans le cytoplasme et très faible dans le noyau, et lorsqu'elle surrexprime l'inhibiteur de la phosphatase 1, l'action de la sous-unité catalytique de la PKA ressemble à celle de l'état normal dans le cytoplasme, par contre dans le noyau, elle observe une action beaucoup plus importante que dans l'état normal. Nous avons dégagées trois hypothèses pouvant chacune expliquer un tel phénomène.

- La phosphatase 1 facilite la sortie de la sous-unité catalytique hors du noyau, donc lorsqu'elle est inhibée la sous-unité catalytique s'accumule dans le noyau puisqu'elle en sort moins.
- Un complexe contenant une protéine PKA est adressé sur la membrane nucléaire, et par des réactions que je ne comprends pas vraiment, l'inhibiteur

facilite l'entrée de la sous-unité catalytique de la PKA de ce complexe dans le noyau, Jean-Luc propose de commencer par modéliser cette hypothèse en rajoutant un compartiment péri-nucléaire où l'entrée de sous-unité catalytique est facilitée.
- L'inhibiteur peut entrer dans le noyau, il y inhibe alors la phosphatase 1 qui avait une action contraire à celle de la sous-unité catalytique, ce qui explique l'augmentation d'action de la sous-unité catalytique.

Je prépare une présentation au propre des schémas des trois hypothèses que nous avions dégagées, comme compte rendu de la réunion. Il faut ensuite en déduire quelles équations sont à prendre ou ne pas prendre en compte selon l'hypothèse.

Semaine 2 : Découverte de Virtual Cell

Mais au détour d'un article, je lis qu'un auteur ne fournit pas ses équations mais donne un logiciel en ligne, Virtual Cell, qui lui a permis de faire son modèle. Lors de la réunion modélisation qui suit, je présente mes schémas, explique que je ne maitrise pas encore le logiciel de programmation Python, et mentionne Virtual Cell. Il est décidé que je vais désormais faire les modèles en utilisant ce logiciel que nous téléchargeons en ligne.

Je commence par étudier le fonctionnement du logiciel en me basant sur des modèles d'autres chercheurs. Je tatonne un peu avec le logiciel d'abord, en essayant de rajouter des espèces et des réactions aux modèles existants, et je finis par lire la notice d'utilisation en ligne (qui ne m'est pas d'un réel secours). Grégoire est très intéressé et participe bien à mes recherches sur le fonctionnement du logiciel. Puis je reprends le modèle de l'article et je dois rajouter les éléments dont on veut tenir compte, notamment un noyau dans la cellule. Pour cela, il faut comprendre toutes les réacions du modèle existant pour ne pas garder des erreurs ni en rajouter. Jean-Luc vérifie surtout que les équations respectent les principes fondamentaux de la thermodynamique, et il semblerait que non. Nous remarquons en effet qu'il n'y a pas toujours conservation de la matière lors des réactions modélisées, et les résultats sont instables. J'ai plutôt du mal à trouver les problèmes dans les équations qui causent ce disfonctionnement car le logiciel Virtual Cell ne permet pas de voir vraiment la façon dont il résout les équations, Jean-Luc modifie quelques réactions pour stabiliser les résultats.

Semaines 3-4 : Modélisation et erreurs de calcul, nouvelles données

En me basant sur deux modèles, je parviens à modéliser deux des trois hypothèses, pour la troisième il me faudrait inventer une équation de réaction, et je ne sais pas vraiment la forme qu'elle pourrait avoir. Mais mes modèles ne sont pas encore vraiment réalistes : il y a beaucoup de données auxquelles j'attribue des valeurs fantaisistes car je ne les trouve pas dans les articles publiés. J'ai donc besoin de me concerter avec les autres pour avoir leur avis, certaines

données resteront surement sans valeur fixe, il faudra les faire varier pour coller aux résultats expérimentaux.

En attendant la prochaine réunion, je relis quelques publications, et j'en cherche de nouvelles en espèrant trouver de quoi améliorer mes données fantaisistes. Grégoire me donne un article que je n'avais pas encore étudié, il traite de la modélisation de cellule cardiaque de chiens et est très complet (une centaine de pages d'annexe pour donner les équations et les constantes). Mis à part certaines concentrations initiales qui varient entre le chien et le rat, les équations semblent plus cohérentes que celles que nous avons étudiées précedemment. Je reconstruis le modèle de Virtual Cell en utilisant ces nouvelles équations, mais je me rends compte qu'une des équations, celle correspondant à la production d'AMP cyclique, n'est pas théorique mais expérimentale : elle est basée sur l'équation de Hill. Jean-Luc modifie les constantes pour approximer la courbe obtenue avec l'équation de Hill par une équation plus simple, type Michaëlis-Menten, que le logiciel Virtual Cell peut prendre en compte plus simplement.

Semaine 5 : Calcul du volume du noyau, conclusion du stage

Une des données que nous ne connaissons pas est le volume des noyaux de la cellule, il y a deux noyaux mais on a décidé de faire un modèle avec un seul noyau pour commencer simplement. Patrick a pris une série de photos grâce au microscope confocal en faisant varier la hauteur de façon régulière sur une cellule cardiaque dans laquelle le marqueur AKAR-nls est activé (donc les noyaux ressortent plus lumineux). Le logiciel d'analyse d'image ImageJ nous permet en modifiant le contraste de faire ressortir les noyaux. On peut ensuite calculer l'aire des flaques blanches sur chaque image, en sommant sur toutes les photos et en multipliant par la hauteur de variation (constante et assez petite) on obtient une bonne approximation du volume de chaque noyau.

Nous avons fait beaucoup de modifications depuis le premier modèle, je résume dans un tableau toutes les équations essayées avec leur provenance et les constantes correspondantes afin de clarifier ce que j'ai fait pour le présenter en réunion.

Construire un tel modèle n'est pas si simple et demande beaucoup de temps, au bout des cinq semaines mon modèle n'est toujours pas utilisable, j'explique à Zeineb le fonctionnement du logiciel, la façon de trouver et comprendre les équations dans les publications. Ainsi elle pourra continuer ce que je n'ai pas fini si ça l'intéresse pour sa thèse. Le vendredi après-midi nous avons une dernière réunion de modélisation, je fais un bilan de mes cinq semaines de stage et on parle de ce qu'il reste à faire, à améliorer, dans le modèle, puis des expériences à effectuer pour donner une idée des contantes inconnues et comment les calculer à partir de ces données expérimentales, ils auront beaucoup de travail encore après mon départ.

5 La modélisation

La compartimentation

Nous avons décidé de commencer avec un modèle en compartiments, plus simple qu'un modèle géométrique où il faut prendre en compte la forme de la cellule et la diffusion des éléments dans les milieux. Chaque compartiment est considéré comme un point et tous les éléments dans un compartiment peuvent réagir entre eux selon des réaction données. Nous désignons trois ou quatre compartiments selon l'hypothèse, le premier correspond à ce qu'il se passe au niveau de la membrane, le second au cytoplasme et le troisième au noyau, dans l'hypothèse 2, on rajoute un compartiment périnucléaire entre le cytoplasme et le noyau. Il peut y avoir des échanges entre compartiments séparés par ce que l'on modélise comme une membrane, c'est ainsi que l'on modélise par exemple la diffusion de la sous-unité catalytique du compartiment cytoplasmique au compartiment nucléaire.

Modélisation grâce au logiciel Virtual Cell

Le logiciel Virtual Cell est conçu pour les biologistes qui veulent construire un modèle réaliste de cellule. Il offre une approche assez simple, on se contente de positionner dans la cellule des compartiments, puis des éléments dans les compartiments (en posant des points verts à l'endroit souhaité) et on les relie par des réactions. Il faut ensuite donner des valeurs initales pour les concentrations de chaque élément, en général on choisit des conditions initiales nulles pour la plupart des produits qui vont être créés par les réactions, bien que ce ne soit pas très réaliste. Le logiciel se charge ensuite de traduire le schéma créé en système d'équation et de le résoudre dans des conditions données. Il rend un tableau ou un graphe indiquant les variations de concentration de chaque espèce en fonction du temps. Avec conditions initiales nulles, on peut observer, si tout se déroule bien, un temps à partir duquel toutes les concentrations sont équilibrées et non nulles. On choisit alors ce temps comme état initial de la cellule et on applique des changements dans la cellule à partir de cet état d'équilibre. On peut ensuite comparer ces résultats théoriques aux résultats expérimentaux.

Ce logiciel peut être très pratique pour ceux qui n'ont pas envie d'écrire des formules mathématiques compliquées mais on ne peut pas vraiment comprendre comment il fonctionne pour résoudre les équations. Il est donc peu aisé de régler les problèmes, nous avons souvent eu des erreurs dues à l'instabilité de la résolution sans pouvoir vraiment trouver quelle était la cause.

Modèle pour les réactions

Pour modéliser les réactions il faut entrer dans l'interface les éléments qui vont réagir, ceux qui sont produits et les enzymes qui catalysent la réaction, puis les relier entre eux par une réaction. Il faut ensuite donner la vitesse de réaction, qui dépend en général des concentrations de tous les éléments mis en

jeu et de certains paramètres dont il faut aussi donner les valeurs. On se base presque uniquement sur les publications antérieures pour trouver les vitesses de réactions, il y en a trois formes principales.

La loi des masses donne une vitesse sous la forme $Kf \times [réactifs] - Kr \times [produits]$. Elle permet de modéliser la plupart des réactions en les décomposant en réactions élémentaires. On peut aussi l'utiliser pour traduire ce qui est donné comme un équilibre instantanné dans les publication, en donnant une constante de formation (Kf) suffisamment grande.

L'équation de Michaëlis-Menten correspond à une vitesse de la forme $\frac{V_{max}}{k+[S]}$. Elle traduit le plus souvent une réaction enzymatique, le substrat (S) se lie à l'enzyme (V_{max} est souvent proportionnel à la concentration d'enzyme) qui le transforme en produit. Elle traduit en une seule équation les réactions correspondant à la liaison du substrat sur l'enzyme, à l'action de l'enzyme et à la séparation du produit de l'enzyme, ces trois réactions pouvant aussi être modélisées séparément par une loi des masses.

L'équation de Hill permet de modéliser beaucoup de courbes expérimentales, la vitesse,correspondante est alors de la forme $\frac{V_\infty}{1+(\frac{k}{x})^n} + V_0$ où x est la concentration d'un élément,V_∞ la vitesse maximale de cette réaction, V_0 la vitesse initiale, k la concentration à laquelle la moitié de la vitesse est atteinte, n est un nombre positif, dans le cas d'une réaction avec catalyseur, il indique le nombre de sites de l'enzyme auxquels le substrat peut se lier. On essaie d'éviter une telle vitesse de réaction car la puissance pas toujours entière est génante pour les conditions initiales nulles.

6 L'intérêt du modèle en biologie

Le but de mon stage était de faire un modèle de cellule cardiaque pour tenter de comprendre les causes des changements de la cinétique de diffusion de la sous-unité catalytique de la PKA dans le noyau. Trois hypothèses ont été dégagées et je devais modéliser chacune de ces hypothèses pour voir laquelle pouvait correspondre le mieux aux observations, ou au moins quelles hypothèses pouvaient être écartées. Un modèle permet de tester une hypothèse, en effet on construit un modèle en se basant sur les articles publiés (et reconnus) avec une hypothèse d'action précise et si ce que l'on observe est différent de ce que le modèle prédit, cela peut indiquer que l'hypothèse est erronée.

L'intérêt d'un modèle est aussi de faire découvrir de nouvelles réactions, d'indiquer des zones où il faut chercher quelque chose que l'on ne sait pas encore décrire. Le but du modèle est de sembler le plus réaliste possible, en prenant en compte toutes les réactions connues qui peuvent être liées au point qui nous intéresse. Et si un modèle qui semble coller au plus près de la réalité prédit un résultat en contradiction avec des observations, cela peut être qu'une réaction non-modélisée car inconnue est à considérer, et donc à mettre à jour par les chercheurs. Le modèle mathématique est donc un réel objet de recherche en biologie.

J'ai remarqué que malgré l'intérêt que cela peut avoir, peu de biologistes font de la modélisation, ou du moins qui en parlent dans leurs publications, à la fin de mon stage j'avais trouvé seulement quatre articles donnant des équations pour modéliser l'action du second messager AMP cyclique dans la cellule cardiaque dont deux du même auteur. Il est vrai que construire un modèle prend du temps que l'on ne peut plus alors passer à observer des expériences au microscope, et les biologistes ne sont pas forcéments friands de programmation. Mais le logiciel Virtual Cell m'a l'air de vouloir mettre cela à la portée du novice, sauf qu'il contient encore quelques erreurs. C'est pour cela que les laboratoires de biologie ont aussi besoin de mathématiciens ou informaticiens.

Les mathématiciens qui font des modèles ne sont plus vraiment reconnus par la communauté mathématique comme de réels mathématiciens, et pas non plus comme des biologistes par leurs collègues chercheurs, et pourtant le décloisonnement et l'échange entre les disciplines ne peut être qu'un avantage pour la recherche.

7 Bilan du stage

Pendant ce stage les chercheurs ont très bien su s'adapter à mon niveau, acceptant de réexpliquer lentement, et même lors des présentations en réunion certains expliquaient ce qu'il faisaient en partant de la base pour que je puisse comprendre même si ce n'était pas mon sujet de travail. Et Jean-Luc m'a beaucoup aidée pour la partie programmation. Lorsque nous travaillions sur Python, un logiciel nouveau pour tout les deux, il me laissait découvrir la base toute seule, mais ensuite ses connaissances sur la programmation en général (et surtout sur Fortran) m'ont été bien utiles. Puis dans les semaines suivantes, où nous étions passés à Virtual Cell, il a passé beaucoup de temps à étudier ce qui ne tournait pas rond dans les modèles que nous reprenions, et ensuite à m'expliquer ce qu'il fallait changer et pourquoi.

J'ai donc toujours été bien entourée de gens prêts à m'expliquer leurs sujets de recherche et aussi intéressés par ce que je faisais (les doctorants voulaient que je fasse un exposé pendant une réunion d'équipe pour qu'ils puissent mieux comprendre ce sur quoi je travaillais). J'ai beaucoup aimé l'ambiance du laboratoire, chacun est passioné par son sujet et explique aux autres l'intéret de ce qu'il fait et à quoi cela peut mener, parfois quelqu'un a déjà fouillé ce sujet alors on peut s'entraider, se donner des publications à consulter. Je retire de ce stage une première expérience de travail avec des chercheurs. Cela m'a permis d'avoir un premier contact avec la programmation en Python, je me fais désormais une bonne idée de ce à quoi ressemble un modèle en biologie cellulaire, et surtout je peux briller en société en parlant de la cellule cardiaque.

Avant ce stage je n'avais pas vraiment d'idée de ce à quoi ressemble le métier de chercheur et cette expérience m'a permis de le découvrir et m'a donné envie de cesser de rejetter l'idée de faire de la recherche mon métier. Cela me semble très enrichissant de pouvoir rencontrer autant de gens passionnés et passionants, de réfléchir aux moyens de justifier ou infirmer les hypothèses que l'on a, trouver

des idées neuves.

Remerciements

à Rodolphe Fischmeister,
Grégoire Vandecasteele,
Jean-Luc Mazet,
Zeineb Haj Slimane,
Françoise Boussac,
Patrick Lechène,
Phillipe Mateo,
Aziz Guellich, Fabien Hubert, Hind Mehel

Annexe C

Mémoire de fin d'études : Applications de la méthode de Stepanov

Université Paris-Sud

23 juin 2014
Mémoire de M2

Applications de la méthode de Stepanov

Lucile DEVIN

Encadrant : Florent JOUVE

Introduction

La méthode de Stepanov est une application d'un résultat classique d'algèbre : sur un corps, un polynôme de degré d a au plus d racines. Ainsi, si un polynôme non nul de degré fixé s'annule sur un ensemble, alors le cardinal de l'ensemble est inférieur au degré du polynôme. Partant de cela, on va avec des considérations d'algèbre linéaire, pouvoir majorer des nombres de points rationnels de courbes algébriques sur des corps finis. De façon plus précise on montrera qu'il existe un polynôme non nul ayant des zéros d'ordre assez grand en chacun des points de l'ensemble qui nous intéresse. Cela se fera en montrant qu'une application linéaire n'est pas injective pour certaines conditions sur le degré du polynôme et l'ordre des zéros. Un bon choix de ces paramètres nous donnera alors une majoration du cardinal de l'ensemble.

Les conjectures de Weil pour les courbes algébriques sur les corps finis, donnent en général la meilleure estimation du nombre de points rationnels, mais de façon beaucoup moins élémentaire que ce que l'on va voir ici. On arrive cependant avec la méthode de Stepanov à des résultats dignes de ceux obtenus par la méthode de Weil. On peut aussi traiter des cas où la méthode géométrique est insuffisante.

Nous allons illustrer cette méthode par quatre exemples. Tout d'abord, la méthode de Stepanov donne de façon élémentaire une bonne estimation du nombre de points rationnels des courbes hyperelliptiques. Les sommes exponentielles sont des objets récurrents de la théorie analytique des nombres, dans le cas où elles sont indexées par les points $\mathbb{F}_q$-rationnels d'un ouvert d'une courbe projective lisse sur $\mathbb{F}_q$, la méthode de Stepanov va aussi permettre de majorer des modules de ces sommes. Nous verrons notamment une majoration pour les sommes de Kloosterman, appliquée directement de l'estimation du nombre de points rationnels des courbes hyperelliptiques (et que l'on peut trouver dans [4]). Nous verrons ensuite une majoration pour les sommes de Heilbronn, d'après [3], où le résultat n'est cependant pas optimal. Pour finir, la méthode de Stepanov s'applique aux courbes algébriques projectives non-singulière générales comme exposé dans [1], et on a une bonne estimation à condition que le cardinal du corps soit un carré parfait. La démonstration n'est plus si élémentaire puisqu'elle nécessite le théorème de Riemann-Roch.

Dans toute la suite, p sera un nombre premier, q une puissance de p cardinal d'un corps fini $\mathbb{F}_q$.

1 Méthode de Stepanov pour les courbes hyperelliptiques

Soit q une puissance d'un premier p. On fixe $\overline{\mathbb{F}_q}/\mathbb{F}_q$, une clôture algébrique de $\mathbb{F}_q$. On considère une courbe algébrique $\mathcal{C}_f/\mathbb{F}_q$ définie par

$$\mathcal{C}_f(\overline{\mathbb{F}_q}) = \{(x,y) \in \overline{\mathbb{F}}_q^2, y^2 = f(x)\} \tag{1}$$

où f est un polynôme à coefficients dans $\mathbb{F}_q$ qui n'est pas un carré dans $\overline{\mathbb{F}_q}[X]$. La méthode de Stepanov va nous donner une bonne approximation de $|\mathcal{C}_f(\mathbb{F}_q)| = N$, le nombre de points $\mathbb{F}_q$ rationnels de $\mathcal{C}_f$.

Théorème 1. *Si $m = \deg(f) \geq 3$ et $q > 4m^2$, alors*

$$|N - q| < 8m\sqrt{q}$$

On voit ici la force de la méthode de Stepanov, en effet cette majoration est du même ordre que celle obtenue grâce aux conjectures de Weil. On va cependant prouver ce théorème de façon plus élémentaire.

On remarque que si $p = 2$, $y \mapsto y^2$ est un automorphisme de $\mathbb{F}_q$ donc $N = q$. On suppose désormais que $p \neq 2$.

Distinguons deux cas dans les points rationnels

— $y = 0$ on s'intéresse aux zéros de f, on pose $N_0 = |\{(x,0) \in \mathbb{F}_q^2, f(x) = 0\}|$

— $y \neq 0$ alors (x, y) est un point de $\mathcal{C}_f(\mathbb{F}_q)$ seulement si $f(x)$ est un carré dans $\mathbb{F}_q$ c'est-à-dire que $f(x)^{\frac{q-1}{2}} = 1$ et alors $(x, -y)$ est aussi un point de $\mathcal{C}_f(\mathbb{F}_q)$.

On pose alors $g = f^{\frac{q-1}{2}}$ et $N_1 = |\{x \in \mathbb{F}_q, g(x) = 1\}|$.

Ainsi $N = N_0 + 2N_1$.

En généralisant un peu, posons pour a dans $\mathbb{F}_q$, $S_a = \{x \in \mathbb{F}_q, f(x) = 0 \text{ ou } g(x) = a\}$. On cherche à déterminer $|S_1|$.

L'idée de Stepanov est de trouver un polynôme dont on sait majorer le degré et qui a des zéros d'ordre fixé en chacun des points de S_a.

1.1 Dérivation de Hasse

Usuellement pour évaluer l'ordre des zéros d'un polynôme on le dérive, mais dans les corps finis, la dérivée d'un polynôme peut s'annuler sans qu'il soit constant. On utilise donc une nouvelle notion de dérivation : la dérivation de Hasse sur $K[X]$ où K est un corps quelconque. Pour tout n, k entiers, on pose

$$E^k X^n = \binom{n}{k} X^{n-k}$$

que l'on étend par linéarité sur $K[X]$.

Lemme 1. *Pour $f, g \in K[X]$, k entier on a*

$$E^k(fg) = \sum_{j=0}^{k} (E^j f)(E^{k-j} g).$$

Corollaire 1. *Pour $a \in K$, n, k entiers, on a*

$$E^k (X - a)^n = \binom{n}{k} (X - a)^{n-k}.$$

Pour f, $g \in K[X]$, n entier, il existe $h \in K[X]$ avec $\deg h \leq \deg f + k \deg g - k$ *et*

$$E^k(fg^n) = hg^{n-k}.$$

Lemme 2. *Soit $f \in K[X]$, $a \in K$, l un entier, si $\forall k < l$, $E^k f(a) = 0$ alors f a un zéro en a d'ordre supérieur ou égal à l (c'est-à-dire que $(X - a)^l \mid f$).*

Lemme 3. *Pour $K = \mathbb{F}_q$, $h \in K[X, Y]$, $r = h(X, X^q)$, k entier, on a*

$$E^k r = (E_X^k h)(X, X^q).$$

où E_X est la dérivation de Hasse par rapport à la première variable.

On peut trouver des preuves de ces résultats dans le livre [4, p.283-284].

1.2 Construction du polynôme

Proposition 1. *Si $q > 8m$, l entier tel que $m < l \leq \frac{q}{8}$, $a \in \mathbb{F}_q$. Alors il existe $r \in \mathbb{F}_q[X]$ avec $\deg r < \frac{q-1}{2}l + 2ml(l-1) + mq$, admettant un zéro d'ordre supérieur ou égal à l en tout point de S_a.*

On cherche r sous la forme $r = f^l \sum_{j=0}^{J}(r_j + s_j g)X^{jq}$, avec $\deg r_j, \deg s_j \leq \frac{q-1}{2} - m$, Alors

$$\begin{aligned}\deg r &\leq l\deg f + \deg g + \frac{q-1}{2} - m + Jq \\ &\leq m(\frac{q}{8} + \frac{q-1}{2} - 1) + \frac{q-1}{2} + Jq \\ &\leq (m+J)q\end{aligned}$$

car $m \geq 3$.

On voudra pouvoir s'assurer que r n'est pas nul.

Lemme 4. *r est le polynôme nul si et seulement si pour tout j, r_j et s_j sont nuls.*

Démonstration. f n'est pas nul, donc quitte à changer X en $X + a$ pour un certain $a \in \mathbb{F}_q$, on peut supposer que $f(0) \neq 0$.

Supposons par l'absurde que r est nul mais pas tous les r_j, s_j, soit k le plus petit indice tel que $r_k \neq 0$ ou $s_k \neq 0$.

Alors $f^l X^{kq}$ divise r et,

$$\begin{aligned}\frac{r}{f^l X^{kq}} &= \sum_{j=k}^{J}(r_j + s_j g)X^{(j-k)q} \\ &= h_0 + h_1 g\end{aligned}$$

On a $h_0 + h_1 g = 0$ donc

$$\begin{aligned}h_0^2 &= h_1^2 g^2 \\ h_0^2 f &= h_1^2 f^q\end{aligned}$$

Or $f \in \mathbb{F}_q[X]$ donc $f(X)^q = f(X^q) \equiv f(0)[\mathrm{mod} X^q]$. Ainsi

$$r_k^2 f \equiv s_k^2 f(0)[\mathrm{mod} X^q].$$

Or ces polynômes sont de degrés inférieurs à q :
$\deg r_k^2 f \leq q - 1 - 2m + m < q$ et $\deg s_k^2 f(0) \leq q - 1 - 2m < q$. Donc

$$r_k^2 f = s_k^2 f(0).$$

$f(0)$ étant un carré dans $\overline{\mathbb{F}_q}$, cela contredit l'hypothèse que f n'est pas un carré dans $\overline{\mathbb{F}_q}[X]$. □

Lemme 5 (Dérivation de Hasse de r)**.** *Pour tout $k \leq l$, pour tout $j \leq J$, il existe $r_j^{(k)}$, $s_j^{(k)}$, de degré inférieur ou égal à $\frac{q-1}{2} - m + k(m-1)$ tels que*

$$E^k r = f^{k-l} \sum_{j=0}^{J}(r_j^{(k)} + s_j^{(k)} g)X^{jq}$$

Démonstration. Le polynôme r peut s'écrire $r = h(X, X^q)$ où $h(X,Y) = f(X)^l \sum_{j=0}^{J}(r_j(X)+s_j(X)g(X))Y^j$. Donc par le lemme 3, pour tout $k \leq l$,

$$E^k r(X) = (E_X^k h)(X, X^q) = \sum_{j=0}^{J}(E^k(f^l(X)r_j(X)) + E^k(s_j(X)f^{l+\frac{q-1}{2}}(X)))X^{qj}.$$

Par le corollaire 1, pour tout $j \leq J$, il existe $r_j^{(k)}$, $s_j^{(k)}$, de degré inférieur ou égal à $\frac{q-1}{2} - m + k(m-1)$ tels que

$$E^k r = \sum_{j=0}^{J}(r_j^{(k)} f^{l-k} + s_j^{(k)} f^{\frac{q-1}{2}+l-k}).$$

□

On veut que r ait des zéros d'ordre supérieur ou égal à l aux points de S_a. D'après le lemme 2, il nous suffit d'avoir $E^k r(x) = 0$ pour tout $k < l$ et tout $x \in S_a$.

Si $x \in S_a$, on a

— Soit $f(x) = 0$, donc x est un zéro de r d'ordre au moins l.

— Soit $f(x) \neq 0$, alors $g(x) = a$. Donc

$$\begin{aligned} E^k r(x) &= f^{k-l}(x)\sum_{j=0}^{J}(r_j^{(k)}(x) + as_j^{(k)}(x))x^{jq} \\ &= f^{k-l}(x)\sum_{j=0}^{J}(r_j^{(k)}(x) + as_j^{(k)}(x))x^{j} \end{aligned}$$

Posons pour $k \leq l$, $\sigma^{(k)} = \sum_{j=0}^{J}(r_j^{(k)}(x) + as_j^{(k)}(x))x^j$. Il nous suffit donc d'annuler les $\sigma^{(k)}$ pour $k < l$. Les coefficients des polynômes $\sigma^{(k)}$ sont des combinaisons linéaires en les coefficients des r_j, s_j. On cherche donc à résoudre un système d'équations linéaires homogènes. Il y a autant d'équations que de coefficients des polynômes $\sigma^{(k)}$. Or $\deg \sigma_k \leq J + \frac{q-1}{2} - m + k(m-1)$, donc il y a au plus $E := l(J + \frac{q-1}{2} - m) + \frac{l(l-1)}{2}(m-1))$ équations. Les inconnues du système sont les coefficients des r_j, s_j, il y en a $I := (q-1-2m)(J+1)$.

Pour un choix de J assez grand on va avoir $E < I$, c'est-à-dire plus d'inconnues que d'équations. $J = \frac{l}{q}(\frac{q-1}{2} + 2m(l-1))$ convient et donne la borne annoncée sur $\deg r$. On aura donc une solution non nulle, et donc un polynôme r non nul avec zéros d'ordre au moins l en tous les points de S_a.

1.3 Preuve du théorème 1

On se place dans les hypothèses du théorème, soit $q > 4m^2$, alors $q > 8m$ car $m > 2$. Pour $a \in \mathbb{F}_q$ quelconque, on applique la proposition 1. Pour $m < l \leq \frac{q}{8}$, l entier, il existe un polynôme non nul r avec des zéros d'ordre au moins l aux points de S_a.

Alors $l|S_a| \leq \deg r \leq \frac{q-1}{2}l + 2ml(l-1) + mq$. Donc

$$|S_a| \leq \frac{q-1}{2} + 2m(l-1) + \frac{mq}{l}.$$

Prenons $l = 1 + \lfloor \frac{\sqrt{q}}{2} \rfloor$ (on a bien $m < l \leq \frac{q}{8}$).

Alors

$$|S_a| \leq \frac{q-1}{2} + m\sqrt{q} + 2m\sqrt{q} < \frac{q-1}{2} + 4m\sqrt{q}.$$

En choisissant $a = 1$ on déduit

$$N_0 + N_1 = |S_1| < \frac{q-1}{2} + 4m\sqrt{q}.$$

Donc

$$N = N_0 + 2N_1 \leq 2(N_0 + N_1) < q + 8m\sqrt{q}. \tag{2}$$

De plus en remarquant que pour tout $x \in \mathbb{F}_q$, $f(x)(g(x)-1)(g(x)+1) = 0$. On déduit en posant $N_{-1} = |\{x \in \mathbb{F}_q, g(x) = -1\}|$ que $N_0 + N_1 + N_{-1} = q$ et

$$N_0 + N_{-1} = |S_{-1}| < \frac{q}{2} + 4m\sqrt{q}.$$

Donc $N_1 = q - (N_0 + N_{-1}) > \frac{q}{2} - 4m\sqrt{q}$.

Ainsi

$$N = N_0 + 2N_1 \geq 2N_1 > q - 8m\sqrt{q}. \tag{3}$$

On déduit de (2) et (3) le théorème 1.

2 Borne pour les sommes de Kloosterman

Soient ψ un caractère additif sur $\mathbb{F}_q$, $a, b \in \mathbb{F}_q$ on définit la somme de Kloosterman :

$$K(\psi; a, b) = \sum_{x \in \mathbb{F}_q^*} \psi(ax + bx^{-1}).$$

On remarque que $K(\psi; a, b) = \sum_{x \in \mathbb{F}_q^*} \psi(ax)\psi(bx^{-1})$. Or pour ψ caractère non trivial, les $\psi_a : x \mapsto \psi(ax)$ décrivent l'ensemble des caractères de $\mathbb{F}_q$ lorsque a décrit $\mathbb{F}_q$. On pourrait donc définir les sommes de Kloosterman de façon équivalente par $K(\psi, \phi) = \sum_{x \in \mathbb{F}_q^*} \psi(x)\phi(x)$, pour ψ, ϕ caractères additifs de $\mathbb{F}_q$.

On définit aussi les sommes compagnes

$$K_n(\psi; a, b) = \sum_{x \in \mathbb{F}_{q^n}^*} \psi(a \operatorname{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(x) + b \operatorname{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(x^{-1})),$$

et la fonction zêta associée :

$$Z((\psi; a, b), T) = \exp\left(\sum_{n \geq 1} \frac{K_n(\psi; a, b)}{n} T^n\right) \in \mathbb{C}[[T]].$$

2.1 Rationnalité de la fonction Zêta

Théorème 2 (Carlitz). *Si ψ est non trivial et a, b tous deux non nuls, alors*

$$Z((\psi; a, b), T) = 1 + K(\psi; a, b)T + qT^2.$$

On peut alors écrire $Z((\psi;a,b),T) = (1-\alpha T)(1-\beta T)$, $K(\psi;a,b) = -(\alpha+\beta)$.

Démonstration. Soit ψ un caractère additif de $\mathbb{F}_q$ non trivial et a, b dans $\mathbb{F}_q$ tous deux non nuls. Soit $G \subset \mathbb{F}_q(X)^*$ le sous-groupe des quotients de polynômes unitaires qui n'ont ni zéro ni pôle en zéro. On peut définir un caractère multiplicatif de G en donnant uniquement sa valeur sur les polynômes unitaires. Soit $\eta : X^d - a_1X^{d-1} + \ldots + (-1)^{d-1}a_{d-1}X + (-1)^d a_d \mapsto \psi_a(a_1)\psi_b(a_{d-1}/a_d)$.

Lemme 6. *η définit un caractère multiplicatif de l'ensemble des polynômes unitaires. Il s'étend à G en posant $\eta(\frac{1}{P}) = \frac{1}{\eta(P)}$.*

Démonstration. En effet, si $P = X^d - a_1X^{d-1} + \ldots + (-1)^{d-1}a_{d-1}X + (-1)^d a_d$, $Q = X^e - b_1X^{e-1} + \ldots + (-1)^{e-1}b_{e-1}X + (-1)^e b_e$, alors

$$PQ = X^{d+e} - (a_1+b_1)X^{d+e-1} + \ldots + (-1)^{d+e-1}(a_{d-1}b_e + a_d b - e - 1)X + (-1)^{d+e}a_d b_e.$$

Donc, $\eta(PQ) = \psi_a(a_1+b_1)\psi_b(\frac{a_{d-1}}{a_d} + \frac{b_{e-1}}{b_e}) = \eta(P)\eta(Q)$. □

On peut ensuite étendre η en une fonction complètement multiplicative sur $\mathbb{F}_q[X]$ en posant $\eta(P) = 0$ si $P(0) = 0$.

On définit alors la fonction L associée à η :

$$L(s,\eta) := \sum_{P \in \mathbb{F}_q[X]} \eta(P) q^{-\deg(P)s} = \prod_{P \in \mathbb{F}_q[X]\text{irred}} (1 - \eta(P) q^{-s\deg(P)})$$

pour $s \in \mathbb{C}$.

Lemme 7. *La fonction L associée à η vérifie, pour tout s dans $\mathbb{C}$,*

$$L(s,\eta) = 1 + K(\psi;a,b)q^{-s} + q^{1-2s}.$$

Démonstration. En réarrangeant les termes selon le degré, on a

$$L(s,\eta) = \sum_{d \geq 0} \sum_{\substack{P \in \mathbb{F}_q[X] \\ \deg(P) = d}} \eta(P) q^{-ds}.$$

On étudie alors $\eta(P)$ selon le degré de P unitaire avec $P(0) \neq 0$.

— Si $\deg(P) = 0$ alors $P = 1$ donc $\eta(P) = 1$.

— Si $\deg(P) = 1$, $P = X + \alpha$, $\alpha \in \mathbb{F}_q^*$ donc $\eta(P) = \psi(a\alpha + b\alpha^{-1})$.
Alors

$$\sum_{\substack{P \in \mathbb{F}_q[X] \\ \deg(P) = 1}} \eta(P) = \sum_{\alpha \in \mathbb{F}_q^*} \psi(a\alpha + b\alpha^{-1}) = K(\psi;a,b).$$

— Si $\deg(P) = 2$, $P = X^2 + \alpha X + \beta$, $\alpha \in \mathbb{F}_q$, $\beta \in \mathbb{F}_q^*$ donc $\eta(P) = \psi(a\alpha + b\alpha\beta^{-1})$.
Alors

$$\begin{aligned}
\sum_{\substack{P \in \mathbb{F}_q[X] \\ \deg(P) = 2}} \eta(P) &= \sum_{\alpha \in \mathbb{F}_q} \sum_{\beta \in \mathbb{F}_q^*} \psi(a\alpha + b\alpha\beta^{-1}) \\
&= \sum_{\beta \in \mathbb{F}_q^*} \psi(0) + \sum_{\alpha \in \mathbb{F}_q^*} \sum_{\beta \in \mathbb{F}_q^*} \psi_a(\alpha)\psi_b(\alpha\beta^{-1}) \\
&= q - 1 + (-1)(-1) = q
\end{aligned}$$

— Si $\deg(P) \geq 3$, $P = X^d - a_1 X^{d-1} + \ldots + (-1)^{d-1} a_{d-1} X + (-1)^d a_d$.

$$\sum_{\substack{P \in \mathbb{F}_q[X] \\ \deg(P)=d}} \eta(P) = \sum_{a_1 \in \mathbb{F}_q} \ldots \sum_{a_{d-1} \in \mathbb{F}_q} \sum_{a_d \in \mathbb{F}_q^*} \psi_a(a_1)\psi_b(a_{d-1}/a_d) = 0$$

Ainsi

$$L(s,\eta) = 1 + K(\psi; a, b) q^{-s} + q q^{-2s}.$$

□

Lemme 8. *On a pour tout s où les deux fonctions sont définies, $Z((\psi; a, b), q^{-s}) = L(s, \eta)$.*

Démonstration. Montrons que les dérivées logarithmiques des deux fonctions de s sont égales.

$$\frac{Z'((\psi; a, b), q^{-s})}{Z((\psi; a, b), q^{-s})} = -\log(q) \sum_{n=0}^{\infty} K_n(\psi; a, b) q^{-sn}.$$

On prend la dérivée logarithmique de $L(s, \eta) = \prod_{P \in \mathbb{F}_q[X]\text{irred}} (1 - \eta(P) q^{-s \deg(P)})$.

$$\begin{aligned}
\frac{L'(s,\eta)}{L(s,\eta)} &= -\sum_{P \in \mathbb{F}_q[X]\text{irred}} (1 - \eta(P) q^{-s\deg(P)})^{-1} \eta(P) q^{-s\deg(P)} \log(q) \deg(P) \\
&= -\log(q) \sum_{d \geq 0} \sum_{P\text{irred } \deg(P)=d} d \sum_{k=1}^{\infty} (\eta(P) q^{-sd})^k \\
&= -\log(q) \sum_{n=0}^{\infty} \left(\sum_{d|n} d \sum_{P\text{irred } \deg(P)=d} \eta(P)^{n/d} \right) q^{-sn}.
\end{aligned}$$

Comme $\lim_{s \to +\infty} L(s, \eta) = 1 = \lim_{s \to +\infty} Z((\psi; a, b), q^{-s})$, il nous suffit de montrer que

$$\sum_{d|n} d \sum_{P\text{irred } \deg(P)=d} \eta(P)^{n/d} = K_n(\psi; a, b)$$

pour tout n.

Soit n un entier et d un diviseur de n. Soit P un polynôme unitaire irréductible de degré d, non nul en 0, $P = X^d - a_1 X^{d-1} + \ldots + (-1)^{d-1} a_{d-1} X + (-1)^d a_d$, $\eta(P) = \psi_a(a_1)\psi_b(a_{d-1}/a_d)$. Soient $x_1, \ldots, x_d$ les racines de P dans $\mathbb{F}_{q^n}$, alors pour $i \in \{1, \ldots, d\}$, $\mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(x_i) = \frac{n}{d} a_1$.

De plus le polynôme $(-1)^d a_d^{-1} X^d P(\frac{1}{X}) = X^d - \frac{a_{d-1}}{a_d} X^{d-1} + \ldots$ a pour racines les x_i^{-1}. Donc pour tout i, $\mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(x_i^{-1}) = \frac{n}{d} \frac{a_{d-1}}{a_d}$.

Ainsi

$$\eta(P)^{n/d} = \psi_a(\mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(x_i)) \psi_b(\mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(x_i^{-1})).$$

En sommant sur l'ensemble des racines de P on obtient

$$d\eta(P)^{n/d} = \sum_{i=1}^{d} \psi_a(\mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(x_i)) \psi_b(\mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(x_i^{-1})).$$

Finalement on somme sur l'ensemble des polynômes unitaires irréductibles non nuls en 0 de degré divisant n :

$$\sum_{d|n} d \sum_{P \text{irred} \deg(P)=d} \eta(P)^{n/d} = \sum_{x \in \mathbb{F}_{q^n}^*} \psi_a(\mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(x))\psi_b(\mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(x^{-1})) = K_n(\psi; a, b).$$

Ce qui conclut. □

Ces deux derniers lemmes prouvent le théorème 2. □

2.2 Majoration des sommes de Kloosterman

Théorème 3. *Si ψ est non trivial, a, b tous deux non nuls, et $p \neq 2$, alors $|\alpha| = |\beta| = \sqrt{q}$ donc $|K(\psi; a, b)| \leq 2\sqrt{q}$.*

De nouveau la méthode de Stepanov nous donne l'ordre attendu par les conjectures de Weil.
Fixons $a, b \in \mathbb{F}_q^*$ et posons $g = aX + bX^{-1} \in \mathbb{F}_q(X)$.

Lemme 9. *Pour $x \in \mathbb{F}_{q^n}$, on a*

$$\sum_{\psi \in \widehat{\mathbb{F}_q}} \psi(\mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(x)) = |\{y \in \mathbb{F}_{q^n} \mid y^q - y = x\}|$$

Démonstration. Soit $y \in \mathbb{F}_{q^n}$, soit $P(X) \in \mathbb{F}_q[X]$ son polynôme minimal, alors $P(X)^q = P(X^q)$ donc y^q est conjugué à y. On a donc pour tout $y \in \mathbb{F}_{q^n}$, $\mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(y^q - y) = \mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(y^q) - \mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(y) = 0$.

— Si $\mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(x) \neq 0$, l'équation $y^q - y = x$ n'a pas de solution, donc $|\{y \in \mathbb{F}_{q^n} \mid y^q - y = x\}| = 0$.
De plus par orthogonalité des caractères, $\sum_{\psi \in \widehat{\mathbb{F}_q}} \psi(\mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(x)) = 0$.

— Si $\mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(x) = 0$, l'équation $y^q - y = x$ a au plus q solutions (racines d'un polynôme de degré q). Si y est une solution de cette équation, on a exactement q solutions données par les $y + a$ avec $a \in \mathbb{F}_q$.
On a en fait une bijection entre les classes à gauche de $\mathbb{F}_{q^n}$ relativement à $\mathbb{F}_q$ et le noyau de la trace donnée par $y + \mathbb{F}_q \mapsto y^q - y$. Donc $|\{y \in \mathbb{F}_{q^n} \mid y^q - y = x\}| = q$.
Or $\sum_{\psi \in \widehat{\mathbb{F}_q}} \psi(0) = |\widehat{\mathbb{F}_q}| = q$.

□

On en déduit pour $n \geq 1$,

$$\begin{aligned}
\sum_{\psi} K_n(\psi; a, b) &= \sum_{\psi} \sum_{x \in \mathbb{F}_{q^n}^*} \psi(\mathrm{tr}_{\mathbb{F}_{q^n}/\mathbb{F}_q}(g(x))) \\
&= \sum_{x \in \mathbb{F}_{q^n}^*} |\{y \in \mathbb{F}_{q^n} \mid y^q - y = g(x)\}| \\
&= |\{(x, y) \in \mathbb{F}_{q^n}^* \times \mathbb{F}_{q^n} \mid y^q - y = g(x)\}|.
\end{aligned}$$

Regardons le cas où ψ est le caractère trivial.

$$K_n(1, 1) = \sum_{x \in \mathbb{F}_{q^n}^*} 1 = q^n - 1$$

Pour ψ non trivial, on a $Z((\psi;a,b),T) = (1-\alpha_\psi T)(1-\beta_\psi T)$. Donc $K_n(\psi;a,b) = -(\alpha_\psi^n + \beta_\psi^n)$. Ainsi

$$|\{(x,y) \in \mathbb{F}_{q^n}^* \times \mathbb{F}_{q^n} \mid y^q - y = g(x)\}| = q^n - 1 - \sum_{\psi \neq \psi_0} \alpha_\psi^n + \beta_\psi^n.$$

On est donc ramené à étudier le nombre de points $\mathbb{F}_{q^n}$-rationnels de la courbe $y^q - y = g(x)$. C'est-à-dire la courbe $ax^2 - (y^q - y)x + b = 0$ (car $b \neq 0$ donc $x \neq 0$). On nomme N_n le nombre de points $\mathbb{F}_{q^n}$-rationnels de cette courbe.

Comme on a supposé $p \neq 2$, les solutions de l'équation sont les $(\frac{(y^q-y)+v}{2a}, y)$, où v est solution de $v^2 = (y^q-y)^2 - 4ab$. Donc N_n est égal au nombre de points $\mathbb{F}_{q^n}$-rationnels de

$$D_{a,b} : v^2 = (y^q - y)^2 - 4ab.$$

Comme $(y^q-y)^2 - 4ab$ n'est pas un carré dans $\overline{\mathbb{F}_q}[y]$ (car $4ab \neq 0$), $D_{a,b}$ est une courbe algébrique du type (1). On applique le théorème 1, avec $m = 2q$. Donc si $q^n > 16q$,

$$|N_n - q^n| < 16q^{1+\frac{n}{2}}.$$

Ainsi, pour $n \geq 4$,

$$\frac{1}{q}\left|1 + \sum_{\psi \neq \psi_0} \alpha_\psi^n + \beta_\psi^n\right| < 16q^{n/2}.$$

Établissons un lemme général qui nous resservira plus tard.

Lemme 10. *Soient $\omega_1, \ldots, \omega_k \in \mathbb{C}$. On suppose qu'il existe A et B constantes positives, et un $n_0 \in \mathbb{N}$, tels que $\forall n \geq n_0$, $|\sum_{i=1}^k \omega_i^n| \leq AB^n$.*

Alors, $\forall 1 \leq i \leq k, |\omega_i| \leq B$.

Démonstration. Considérons

$$\sum_{n=0}^{\infty}\left(\sum_{i=1}^{k}\omega_i^n\right)T^n = \sum_{i=1}^{k}\frac{1}{1-T\omega_i},$$

c'est une série convergente sur le disque $|T| < \frac{1}{B}$. En particulier elle n'admet pas de pôle à l'intérieur de ce disque. Donc pour tout i, $|\omega_i^{-1}| \geq \frac{1}{B}$, d'où

$$|\omega_i| \leq B.$$

□

Ceci appliqué à la famille des α_ψ, β_ψ avec $A = 16q$ et $B = \sqrt{q}$ donne pour tout ψ caractère non trivial $|\alpha_\psi|, |\beta_\psi| \leq \sqrt{q}$.

Comme de plus $\alpha_\psi \beta_\psi = q$, on a l'égalité

$$|\alpha_\psi| = |\beta_\psi| = \sqrt{q},$$

donc le théorème annoncé.

3 Borne pour les sommes de Heilbronn

D'après [2] et [3].
On utilise ici la fonction 1-périodique $x \mapsto e(x) = \exp(2i\pi x)$.
Pour p premier fixé, a entier non divisible par p, on définit

$$H_p(a) = \sum_{n=0}^{p-1} e\left(\frac{an^p}{p^2}\right),$$

appelé somme de Heilbronn. Dans cette partie, on s'attache à trouver un majorant pour $H_p(a)$.
Dans leur article [3], Heath-Brown et Konyagin ont montré :

Théorème 4. *On a $H_p(a) \ll p^{7/8}$ avec une constante implicite absolue.*

On s'attendrait en fait à un majorant en $\ll p^{3/4}$ mais il n'y a pas encore de preuve d'une telle majoration.

3.1 Réduction du problème au comptage de points rationnels d'une courbe

Définissons $H_0(a) = H_{0,p}(a) = \sum_{n=1}^{p-1} e\left(\frac{an^p}{p^2}\right) = H_p(a) - 1$, alors un majorant de $H_0(a)$ donnera un majorant de $H_p(a)$. On va en fait montrer que :

Théorème 5. $\sum_{r=0}^{p-1} |H_0(a+rp)|^4 \ll p^{7/2}$ *avec une constante implicite absolue.*

Cela donne bien le résultat car en effet :

$$|H_0(a)|^4 \leq \sum_{r=0}^{p-1} |H_0(a+rp)|^4$$

et $|H_0(a)|^4 \ll p^{7/2}$ implique $|H_0(a)| \ll p^{7/8}$.
Soient

$$f(X) = X + \frac{X^2}{2} + \ldots + \frac{X^{p-1}}{p-1} \in \mathbb{F}_p[X]$$

et

$$\mathcal{B} = \left\{(x_1, x_2) \in (\mathbb{F}_p - \{0,1\})^2, f(x_1) = f(x_2)\right\}.$$

Lemme 11. *On a $\sum_{r=0}^{p-1} |H_0(a+rp)|^4 \leq p^3 + p^2 \# \mathcal{B}$*

Démonstration. Pour $m \in \mathbb{F}_p^*$, on a $H_0(a) = H_0(am^p)$, donc :

$$(p-1)\sum_{r=0}^{p-1} |H_0(a+rp)|^4 = \sum_{r=0}^{p-1}\sum_{m=1}^{p-1} |H_0((a+rp)m^p)|^4.$$

Or, $(r,m) \mapsto (a+rp)m^p$ est injective de $\mathbb{F}_p \times \mathbb{F}_p^*$ dans $\mathbb{Z}/p^2\mathbb{Z}$. Donc,

$$(p-1)\sum_{r=0}^{p-1} |H_0(a+rp)|^4 \leq \sum_{n=0}^{p^2-1} |H_0(n)|^4.$$

Mais $|H_0(n)|^4 = H_0(n)^2\overline{H_0(n)}^2 = \sum_{m_1=1}^{p-1}\dots\sum_{m_4=1}^{p-1} e\left(\frac{n(m_1^p+m_2^p-m_3^p-m_4^p)}{p^2}\right)$.
D'où

$$\begin{aligned}(p-1)\sum_{r=0}^{p-1}|H_0(a+rp)|^4 &\leq \sum_{m_1=1}^{p-1}\dots\sum_{m_4=1}^{p-1}\sum_{n=0}^{p^2-1} e\left(\frac{n(m_1^p+m_2^p-m_3^p-m_4^p)}{p^2}\right)\\ &= p^2\#\left\{(m_1,m_2,m_3,m_4)\in\mathbb{F}_p^{*4}, m_1^p+m_2^p=m_3^p+m_4^p \pmod{p^2}\right\}.\end{aligned}$$

On cherche donc le cardinal de cet ensemble. Si

$$m_1^p+m_2^p=m_3^p+m_4^p \pmod{p^2},$$

alors en particulier

$$m_1^p+m_2^p=m_3^p+m_4^p \pmod{p}$$

donc

$$m_1-m_3=m_4-m_2 \pmod{p}.$$

Posons $b=m_1-m_3$,

Si $b=0$, alors $m_1=m_3$ et $m_4=m_2$, on a $p-1$ choix pour m_1 et de même pour m_4, ce qui donne $(p-1)^2$ éléments.

Sinon, $b\in\mathbb{F}_p^*$ (ce qui fait $p-1$ choix pour b), on pose $m_1=\nu_1 b$, $m_4=\nu_2 b$, alors $m_3=(\nu_1-1)b$, $m_2=(\nu_2-1)b$ (donc $\nu_1,\nu_2\in\mathbb{F}_p-\{0,1\}$).
$m_1^p+m_2^p=m_3^p+m_4^p \pmod{p^2}$ se traduit par :

$$(\nu_1^p-(\nu_1-1)^p)b^p=(\nu_2^p-(\nu_2-1)^p)b^p \pmod{p^2}$$
$$(\nu_1^p-(\nu_1-1)^p)=(\nu_2^p-(\nu_2-1)^p) \pmod{p^2} \quad (4)$$

car b est inversible modulo p donc b^p l'est aussi, et il est donc premier à p^2.

Lemme 12. *Pour $x\in\mathbb{F}_p$, $x^p-(x-1)^p=1-pf(x) \pmod{p^2}$.*

Démonstration.

$$x^p-(x-1)^p=-\sum_{k=1}^{p}\binom{p}{k}(-1)^{p-k}x^k \pmod{p^2}.$$

Or, pour $k\neq 0,p$

$$\begin{aligned}\binom{p}{k} &= \frac{p(p-1)\dots(p-k+1)}{k(k-1)\dots 2.1} \pmod{p^2}\\ &= \frac{p(-1)(-2)\dots(-k+1)}{k(k-1)\dots 2.1} \pmod{p^2}\\ &= (-1)^{k-1}\frac{p}{k} \pmod{p^2}.\end{aligned}$$

Donc

$$\begin{aligned}x^p-(x-1)^p &= p\sum_{k=1}^{p-1}(-1)^p\frac{x^k}{k}+(-1)^{p+1} \pmod{p^2}\\ &= 1-pf(x) \pmod{p^2}\end{aligned}$$

car p est impair. □

On peut donc remplacer la condition (4) par :

$$1 - pf(\nu_1) = 1 - pf(\nu_2) \text{ [mod } p^2]$$

d'où

$$f(\nu_1) = f(\nu_2) \text{ [mod } p]$$

et $\nu_1, \nu_2 \in \mathbb{F}_p - \{0,1\}$. Ainsi,

$$\#\left\{(m_1,m_2,m_3,m_4) \in \mathbb{F}_p^{*4}, m_1^p + m_2^p = m_3^p + m_4^p \text{ [mod } p^2]\right\} = (p-1)^2 + (p-1)\#\mathcal{B}.$$

D'où

$$\begin{aligned}(p-1)\sum_{r=0}^{p-1}|H_0(a+rp)|^4 &\leq p^2((p-1)^2 + (p-1)\#\mathcal{B}\\ \sum_{r=0}^{p-1}|H_0(a+rp)|^4 &\leq p^3 + p^2\#\mathcal{B}\end{aligned}$$

□

3.2 Méthode de Stepanov

Le but est maintenant d'estimer $\#\mathcal{B} = \#\left\{(x_1,x_2) \in (\mathbb{F}_p - \{0,1\})^2, f(x_1) = f(x_2)\right\}$.

Théorème 6. *On a $\#\mathcal{B} \ll p^{3/2}$ avec une constante implicite absolue.*

Cela permettra de démontrer le théorème 5, en effet on aura alors $p^3 + p^2\#\mathcal{B} \ll p^{7/2}$, donc d'après le lemme 11, $\sum_{r=0}^{p-1}|H_0(a+rp)|^4 \ll p^{7/2}$.

Comme $\mathcal{B}$ est une courbe, on pourrait s'attendre à une majoration du type $\ll p$, ce qui donnerait le majorant $\ll p^{3/4}$ pour les sommes de Heilbronn, mais un tel majorant n'a pas encore été trouvé.

Pour $u \in \mathbb{F}_p$, on pose $\mathcal{F}(u) = \{x \in \mathbb{F}_p - \{0,1\}, f(x) = u\}$. Alors

$$\begin{aligned}\#\mathcal{B} &= \sum_{u\in\mathbb{F}_p} \#\{x_1 \in \mathbb{F}_p - \{0,1\}, f(x_1) = u\}\,\#\{x_2 \in \mathbb{F}_p - \{0,1\}, f(x_2) = u\}\\ &= \sum_{u\in\mathbb{F}_p} (\#\mathcal{F}(u))^2 \qquad (5)\end{aligned}$$

et

$$\sum_{u\in\mathbb{F}_p} \#\mathcal{F}(u) = p. \qquad (6)$$

On veut estimer $\#\mathcal{F}(u)$, pour cela on estime le cardinal d'une réunion de tels ensembles : Pour $\mathcal{U} \subset \mathbb{F}_p$ une partie de cardinal T, on pose $\mathcal{G}_\mathcal{U} = \bigcup_{u\in\mathcal{U}} \mathcal{F}(u)$.

Lemme 13. *Si $\#\mathcal{U} = T \geq 1$, alors $\#\mathcal{G}_\mathcal{U} \ll (pT)^{2/3}$ avec une constante implicite absolue.*

Démonstration. Utilisant la méthode de Stepanov

On commence par remarquer que la borne triviale sur $\#\mathcal{G}_{\mathcal{U}}$ est p, donc pour $T \geq p^{1/2}$ le résultat est automatique. On suppose donc désormais $T \ll p^{1/2}$.

Pour appliquer la méthode de Stepanov, on cherche un polynôme $\Phi(X,Y,Z) \in \mathbb{F}_p[X,Y,Z]$ vérifiant $\deg_X \Phi < A$, $\deg_Y \Phi < B$ et $\deg_Z \Phi < C$, tel que le polynôme $\Psi(X) = \Phi(X, f(X), X^p)$ soit non nul et ait des zéros d'ordre supérieur à D en tout point de $\mathcal{G}_{\mathcal{U}}$ (les constantes A, B, C, D seront fixées plus tard selon les conditions que l'on trouvera pour l'existence de tels polynômes). En effet cela nous donnera

$$D\#\mathcal{G}_{\mathcal{U}} \leq \deg \Psi \leq A + p(B+C).$$

Pour que Ψ ait un zéro d'ordre au moins D au point $x \in \mathbb{F}_p$, il faut que

$$\forall n < D, \frac{d^n \Psi(X)}{dX^n}|_{X=x} = 0.$$

Comme $0, 1 \notin \mathcal{G}_{\mathcal{U}}$, c'est équivalent à

$$\forall n < D,\ (X(1-X))^n \frac{d^n \Psi(X)}{dX^n}|_{X=x} = 0.$$

Pour chaque a, b, c,

$$\begin{aligned}
(X(1-X))^n \quad & \frac{d^n(X^a f(X)^b X^{pc})}{dX^n} \\
= \quad & X^{pc}(X(1-X))^n \frac{d^n(X^a f(X)^b)}{dX^n} \\
= \quad & X^{pc} \sum_{k=0}^{n} \sum_{k+k_1+\ldots+k_b=n} C(n,k,k_1,\ldots,k_b)(X(1-X))^k \frac{d^k X^a}{dX^k} \prod_{i=1}^{b} (X(1-X))^{k_i} \frac{d^{k_i} f(X)}{dX^{k_i}}
\end{aligned}$$

Or

— $(X(1-X))^k \frac{d^k X^a}{dX^k}$ est nul si $k > a$ et est un polynôme de degré $a+k$ sinon (puisque la dérivée k-ème d'un polynôme de degré a est de degré $a-k$ si $k \leq a$, sinon elle est nulle).

— Pour tout $k \geq 1$, il existe $q_k(X), h_k(X) \in \mathbb{F}_p[X]$ avec $\deg(q_k) \leq k+1$, $\deg(h_k) \leq k-1$ tels que

$$(X(1-X))^k \frac{d^k f(X)}{dX^k} = q_k(X) + h_k(X)(X^p - X)$$

(par récurrence, pour les détails on peux voir la partie 3 de [2]).

Ainsi,

$$\begin{aligned}
(X(1-X))^n \quad & \frac{d^n(X^a f(X)^b X^{pc})}{dX^n} \\
= \quad & X^c \sum_{k=0}^{\min(a,n)} \sum_{k+k_1+\ldots+k_b=n} C(n,k,k_1,\ldots,k_b) r_k(X) q_{k_1}(X) \ldots q_{k_l}(X) f(X)^{b-l} \ [\mathrm{mod}\ X^p - X]
\end{aligned}$$

où $l \in \{0, \ldots, \min(b,n)\}$

$$(X(1-X))^n \frac{d^n(X^a f(X)^b X^{pc})}{dX^n} = \sum_{l=0}^{\min(b,n)} P_l(X;a,b,c,n) f(X)^{b-l} \ [\mathrm{mod}\ X^p - X]$$

avec $\deg(P_l(X;a,b,c,n)) \leq c+a+k+k_1+1+\ldots+k_l+1 = c+a+n+l \leq a+2n+c$.

Pour $u \in \mathbb{F}_p$, on pose $P(X;a,b,c,n,u) = \sum_{l=0}^{\min(b,n)} P_l(X;a,b,c,n)u^{b-l}$ de telle sorte que, pour tout $x \in \mathcal{F}(u)$,

$$(X(1-X))^n \frac{d^n(X^a f(X)^b X^{pc})}{dX^n}|_{X=x} = P(x;a,b,c,n,u)$$

car $x^p - x = 0$.

Revenons à $\Phi(X,Y,Z) = \sum_{a,b,c} \lambda_{a,b,c} X^a Y^b Z^c$ et posons $P_{n,u}(X) = \sum_{a,b,c} \lambda_{a,b,c} P(X;a,b,c,n,u)$, de sorte que $\deg(P_{n,u}) \leq A+2n+C$ et

$$(X(1-X))^n \frac{d^n \Psi(X)}{dX^n}|_{X=x} = P_{n,u}(x)$$

pour tout $x \in \mathcal{F}(u)$.

On veut choisir les $\lambda_{a,b,c}$ tels que les $P_{n,u}$ soient identiquement nuls pour tout $n < D$ et tout $u \in \mathcal{U}$. Les coefficients des $P_{n,u}$ sont des formes linéaires en les coefficients de Φ, comme nous voulons annuler $D \times T$ polynômes de degré inférieur à $A+2D+C$ strictement, cela revient à annuler $DT(A+2D+C)$ formes linéaires sur un espace de dimension ABC. Pour qu'il existe un polynôme Φ non nul qui convienne, il suffit donc que :

$$DT(A+2D+C) < ABC.$$

Lemme 14. *Soit $F(X,Y) \in \mathbb{F}_p[X,Y]$ non nul avec $m = \deg_X(F) \leq A$, $n = \deg_y(F) \leq B$, alors si $AB \leq p$, $X^p \nmid F(X, f(X))$.*

On commence par montrer :

Lemme 15. *Soient m, n entiers fixés, inférieurs stricts à p, alors pour $a \leq m$, $1 \leq b \leq n$ on a*

$$(1-X)^{m+1} \frac{d^{m+1}(X^a f(X)^b)}{dX^{m+1}} \equiv G(X, f(X); a, b) \ [\mathrm{mod} X^{p-1-m}],$$

avec $\deg_X G(X,Y;a,b) \leq a$, $\deg_Y G(X,Y;a,b) \leq b-1$ et tel que le coefficient devant $X^a Y^{b-1}$ de $G(X,Y;a,b)$ est non nul.

Démonstration. On a $(1-X)^k \frac{d^k X^a}{dX^k}$ est nul si $k > a$ et est un polynôme de degré a sinon.

De plus :

$$\begin{aligned}
(1-X)^l \frac{d^l f(X)}{dX^l} &= (1-X)^l \frac{d^{l-1}}{dX^{l-1}}(1+X+\ldots+X^{p-2}) \\
&= (1-X)^l \frac{d^{l-1}}{dX^{l-1}} \left(\frac{1}{1-X} + X^{p-1} P_1 \right) \\
&= (1-X)^l \left(\frac{(l-1)!}{(1-X)^{l-1}} + X^{p-l} P_2 \right) \\
&= (l-1)! \ [\mathrm{mod} X^{p-l}].
\end{aligned}$$

Par la formule de Leibniz on a alors :

$$(1-X)^{m+1} \frac{d^{m+1}(X^a f(X)^b}{dX^{m+1}} = \sum_{k=0}^{a} \sum_{k+l_1+\ldots+l_b=m+1} C^{m+1}_{k,l_1,\ldots,l_b} g_k(X) \prod_{i=1}^{I_{k,l}} (l_i-1)! f(X)^{b-I_{k,l}} \ [\mathrm{mod} X^{p-1-m}]$$

où $I_{k,l}$ est le nombre de l_i non nuls (on peut réarranger les l_i par ordre décroissant en changeant la constante $C^{m+1}_{k,l_1,\dots,l_b}$).

Ainsi $\deg_X G(X,Y;a,b) \leq a$. Et $I_{k,l} \geq 1$ pour tout k,l car $k \leq a < m+1$ donc il y a toujours au moins un l_i non nul, donc $\deg_Y G(X,Y;a,b) \leq b-1$.

On s'intéresse alors au coefficient devant $X^a f(X)^{b-1}$. Pour $k \in \{0,\dots,a\}$ fixé, le coefficient devant X^a de $(1-X)^k \frac{d^k X^a}{dX^k}$ est $(-1)^k \frac{a!}{(a-k)!}$. Le coefficient devant $f(X)^{b-1}$ de $(1-X)^{m+1-k}\frac{d^{m+1-k}f(X)^b}{dX^{m+1-k}}$ est $\sum_{i=1}^{b}(m+1-l_i)!$ avec $k+l_i = m+1$, donc $b(m-k)!$.

Donc le coefficient devant $X^a f(X)^{b-1}$ est :

$$\begin{aligned}\sum_{k=0}^{a}\binom{m+1}{k}(-1)^k\frac{a!}{(a-k)!}b(m-k)! &= b.a!(m-a)!\sum_{k=0}^{a}(-1)^k\binom{m+1}{k}\binom{m-k}{m-a}\\ &= b.a!(m-a)!(-1)^a \qquad (7)\\ &\neq 0\end{aligned}$$

à condition que $a, b, m-a$ soient tous inférieurs strictement à p. Cela est vérifié par l'hypothèse de départ. L'égalité (7) n'est pas tout à fait évidente, elle est montrée dans la partie 3 de [2] en remarquant que la somme que l'on cherche à évaluer est au signe près le coefficient devant X^{m-a} de $\frac{1-X^{m+1}}{1-X}$. □

On peut maintenant démontrer le lemme 14.

Démonstration. On prouve par récurrence sur $n = \deg_Y(F)$ la proposition : si $F(X,Y) \in \mathbb{F}_p[X,Y]$ non nul avec $m = \deg_X(F)$, $n = \deg_y(F)$ et $(m+1)(n+1) \leq p$, alors $X^{(m+1)(n+1)} \nmid F(X, f(X))$.

Le cas $n = 0$ donne un polynôme $F(X)$ de degré m non nul donc X^{m+1} ne divise pas $F(X)$.

Soit n non nul et supposons le résultat vrai au rang $n-1$. Supposons que l'on ait $X^{(m+1)(n+1)} \mid F(X, f(X))$ alors

$$X^{(m+1)n} \mid (1-X)^{m+1}\frac{d^{m+1}F(X,f(X))}{dX^{m+1}} = G(X,f(X)) \ [\mathrm{mod} X^{p-1-m}]$$

par linéarité sur le résultat du lemme précédent, où $G(X,Y) \in \mathbb{F}_p[X,Y]$ vérifie $m' = \deg_X(G(X,Y)) \leq m$ et $\deg_Y(G(X,Y)) = n-1$. Comme $(m+1)n \leq p-1-m$, $X^{(m+1)n} \mid G(X,f(X))$, en particulier $X^{(m'+1)n} \mid G(X,f(X))$ or G est un polynôme non nul, ce qui contredit l'hypothèse de récurrence. Ainsi $X^{(m+1)(n+1)} \nmid F(X,f(X))$, ce qui conclut. □

Ainsi si Φ est non nul et que $AB \leq p$, il existe c_0 minimal tel que $\Phi(X,Y,Z) = \sum_{c=c_0}^{C} F_c(X,Y)Z^c$ avec $F_{c_0}(X,Y) \neq 0$, $\deg_X(F_{c_0}) \leq A$, $\deg_y(F_{c_0}) \leq B$.
Alors $X^p \nmid F_{c_0}(X,f(X))$, donc en particulier $F_{c_0}(X,f(X))$ est non nul, et $\Psi(X) \neq 0$.

On peut alors choisir $A = \lfloor p^{2/3}T^{-1/3} \rfloor$, $B = C = \lfloor p^{1/3}T^{1/3} \rfloor$ et $D = \left\lfloor \frac{p^{2/3}T^{-1/3}}{16} \right\rfloor$.
On a bien $AB \leq p^{2/3}T^{-1/3}p^{1/3}T^{1/3} = p$ et

$$DT(A+2D+C) \leq \frac{9}{128}p^{4/3}T^{1/3} + \frac{pT}{16},$$

$$ABC = p^{4/3}T^{1/3} + O(p)$$

Donc

$$DT(A+2D+C) < ABC$$

pour p assez grand.

Finalement

$$\begin{aligned} D\#\mathcal{G}_U &\leq A + p(B+C) \\ p^{2/3}T^{-1/3}\#\mathcal{G}_U &\ll p^{2/3}T^{-1/3} + p.p^{1/3}T^{1/3} \\ \#\mathcal{G}_U &\ll p^{2/3}T^{2/3} \end{aligned}$$

d'où la majoration annoncée, de plus la constante implicite est absolue. Ceci achève la preuve du Lemme 13. □

3.3 Preuve du Théorème 4

On peut maintenant prouver le théorème. On numérote les éléments de $\mathbb{F}_p$ pour avoir

$$\#\mathcal{F}(u_1) \geq \#\mathcal{F}(u_2) \geq \dots$$

Et on prend $\mathcal{U} = \{u_1, \dots, u_T\}$. Alors,

$$T\#\mathcal{F}(u_T) \leq \#\mathcal{G}_U \ll p^{2/3}T^{2/3},$$

donc

$$\#\mathcal{F}(u_T) \ll p^{2/3}T^{-1/3}.$$

Ainsi

$$\begin{aligned} \sum_{\frac{N}{2}<T\leq N} (\#\mathcal{F}(u_T))^2 &\ll \sum_{\frac{N}{2}<T\leq N} (p^{2/3}T^{-1/3})^2 \\ &\leq p^{4/3}\frac{N}{2}\left(\frac{N}{2}\right)^{-2/3} \\ &\ll p^{4/3}N^{1/3}. \end{aligned}$$

On a aussi en utilisant (6)

$$\begin{aligned} \sum_{\frac{N}{2}<T\leq N} (\#\mathcal{F}(u_T))^2 &\leq \sup_{\frac{N}{2}<T\leq N} (\#\mathcal{F}(u_T)) \sum_{\frac{N}{2}<T\leq N} \#\mathcal{F}(u_T) \\ &\ll p^{2/3}N^{-1/3}p \\ &\ll p^{5/3}N^{-1/3}. \end{aligned}$$

En utilisant ces deux majorations, d'après (5), on trouve

$$\begin{aligned} \#\mathcal{B} &= \sum_{1\leq 2^k<p^{1/2}} \sum_{2^{k-1}<T\leq 2^k} (\#\mathcal{F}(u_T))^2 + \sum_{p^{1/2}\leq 2^k<2p} \sum_{2^{k-1}<T\leq 2^k} (\#\mathcal{F}(u_T))^2 \\ &\ll \sum_{1\leq 2^k<p^{1/2}} p^{4/3}2^{k/3} + \sum_{p^{1/2}\leq 2^k<2p} p^{5/3}2^{-k/3} \\ &\ll p^{4/3}p^{1/6} + p^{5/3}p^{-1/6} \\ &\ll p^{3/2}. \end{aligned}$$

Les constantes implicites restent absolues. Ce qui conclut d'après le lemme 11 et le théorème 5.

3.4 Commentaires sur la majoration obtenue

On pourrait s'attendre plutôt à une majoration $\#\mathcal{B} \ll p$, en effet $\mathcal{B}$ est une courbe dans le plan $\mathbb{F}_p^2$, intuitivement, une courbe a autant de points qu'une droite. Avec une telle majoration, on trouverait dans le lemme 11, $\sum_{r=0}^{p-1} |H_0(a+rp)|^4 \ll p^3$. Donc

$$|H_0(a)| \ll p^{3/4}.$$

Les sommes de Heilbronn ont justement été construites pour que la méthode géométrique ne fonctionne pas bien, et il semblerait pour le moment que la méthode de Stepanov soit celle qui donne les meilleurs résultats.

En effet, les points $\mathbb{F}_p$-rationnels de $\mathcal{B}$ sont tous singuliers. En effet en utilisant le critère jacobien, on trouve $\frac{d(f(x)-f(y))}{dx} = \frac{x^p-x}{x(x-1)} = 0$ pour $x \in \mathbb{F}_p - \{0,1\}$. On ne peut donc pas appliquer les résultats de la partie suivante dédiée aux courbes non-singulières.

En étudiant la courbe $\mathcal{B}$, on a pu remarquer au moins pour les petits premiers, qu'elle est réunion de trois composantes irréductibles. En effet

$$f(x) - f(y) = (x-y)(x+y-1)h(x,y)$$

où h est un polynôme de degré $p-3$, irréductible pour toutes les valeurs de $p < 100$. Ainsi $\mathcal{B}$ est réunion des points $\mathbb{F}_q$-rationnels de deux droites de cardinal p et de la courbe définie par l'annulation de h. Soit $\mathcal{H} = \{(x,y) \in (\mathbb{F}_p - \{0,1\})^2 \mid h(x,y) = 0\}$, l'ensemble des points $\mathbb{F}_q$-rationnels de cette courbe.

Nous avons calculé grâce au logiciel `sage` le cardinal de la courbe $\mathcal{H}$ pour les premiers plus petits que 100 et mis en comparaison avec la valeur trouvée par la méthode de Stepanov $p^{3/2}$.

p	5	7	11	13	17	19	23	29	31	37	41	43	47	53	59	61	67
$\#\mathcal{H}$	4	8	20	20	40	32	40	100	68	80	76	104	88	124	328	176	164
$p^{3/2}$	11	18	36	46	70	82	110	156	172	225	262	281	322	385	453	476	548

p	71	73	79	83	89	97
$\#\mathcal{H}$	208	200	308	268	244	224
$p^{3/2}$	598	623	702	756	839	955

Il semblerait que la borne $p^{3/2}$ est trop élevée. On cherche alors une façon de d'améliorer la majoration.

Le résultat de Lang-Weil s'applique aux variétés algébrique géométriquement irréductibles sur un corps fini qui se plongent dans un espace projectif de dimension finie. En supposant que h est irréductible sur une clôture algébrique de $\mathbb{F}_p$, on peut supposer que le théorème de Lang-Weil s'applique. En effet, $\mathcal{H}$ est une courbe plane donc elle se plonge dans l'espace projectif de dimension deux. Le théorème donne donc

$$|\#\mathcal{H} - p| \leq (d-1)(d-2)p^{1/2} + A(d)$$

où d est le degré de $\mathcal{H}$ c'est-à-dire $\deg(h) = p-3$. L'estimation n'est donc pas intéressante puisque $(d-1)(d-2)p^{1/2}$ est de l'ordre de $p^{5/2}$. On a même une estimation moins bonne que celle obtenue grâce à la méthode de Stepanov.

4 Borne pour le nombre de points rationnels pour les courbes sur les corps finis

D'après [1].

Soit $q = p^\alpha$, avec α pair, $k = \mathbb{F}_q$ un corps fini. On considère $\mathcal{C}/k$ une courbe projective non-singulière de genre g, $\nu_r(\mathcal{C})$ le nombre de points $\mathbb{F}_{q^r}$-rationnels de $\mathcal{C}$. On définit :

$$Z(\mathcal{C}, T) = \exp\left(\sum_{r=1}^{\infty} \frac{\nu_r(\mathcal{C})}{r} T^r\right) \in \mathbb{C}[[T]]$$

Théorème 7. *On a :*

$$Z(\mathcal{C}, T) = \frac{P(T)}{(1-T)(1-qT)}$$

où $P(T) = \prod_{i=1}^{2g}(1-\omega_i T) \in \mathbb{Z}[T]$, *vérifie une équation fonctionnelle équivalente à pour tout* i, $\omega_i \omega_{2g-i} = q$.

Ainsi $\nu_r(\mathcal{C}) = q^r + 1 + \sum_{i=1}^{2g} \omega_i^r$.

La méthode de Stepanov permet de montrer le

Théorème 8. *Si* $q > (g+1)^4$, *alors*

$$|\nu_r(\mathcal{C}) - q^r - 1| \leq 2gq^{r/2}$$

On a donc une majoration du même ordre que celle obtenue par la conjecture de Weil, à condition que q soit un carré parfait.

Il suffit de prouver le résultat pour $r = 1$. On va regarder $\mathcal{C}$ comme définie sur $\overline{k}$ (clôture algébrique de k), alors on a l'automorphisme de Frobenius $\varphi : \mathcal{C} \to \mathcal{C}$, et $\nu_r(\mathcal{C})$ est le cardinal de l'ensemble des points fixes de φ.

4.1 Théorème de Riemann-Roch

Pour x un point de $\mathcal{C}/\bar{k}$, x est un point régulier donc $\bar{k}[\mathcal{C}]_x$ est un anneau de valuation discrète. Nommons M_x son idéal maximal. On peut alors définir une valuation sur $\bar{k}[\mathcal{C}]_x$,

$$f \mapsto \mathrm{ord}_x(f) = \max\{d \in \mathbb{N} \mid f \in M_x^d\}$$

si f est non nul et $\mathrm{ord}_x(0) = \infty$. Cette valuation s'étend à $\bar{k}(\mathcal{C})$ en posant $\mathrm{ord}_x(\frac{f}{g}) = \mathrm{ord}_x(f) - \mathrm{ord}_x(g)$.

Lemme 16. *Si* $t \in \bar{k}(\mathcal{C})$ *est une uniformisante en un point* x *de* $\mathcal{C}$, *alors* $\bar{k}(\mathcal{C})$ *est une extension finie séparable de* $\bar{k}(t)$.

On peut voir les éléments de $k(\mathcal{C})$ comme des fonctions rationnelles $\mathcal{C} \to \mathbb{P}^1$. Si $\mathrm{ord}_x(f) > 0$, on dit que f à un zéro en x, si $\mathrm{ord}_x(f) < 0$, f a un pôle en x. On peut à chaque $f \in k(\mathcal{C})^*$ associer la somme formelle $\sum_{x \in \mathcal{C}} \mathrm{ord}_x(f)[x] =: \mathrm{div}(f)$.

On peut généraliser cela. On appelle diviseur de $\mathcal{C}$ (ou de $k(\mathcal{C})$) toute somme formelle finie $D = \sum_{x \in \mathcal{C}} n_x[x]$, où les n_x sont des entiers presque tous nuls. L'ensemble des diviseurs de $k(\mathcal{C})$ forme un groupe abélien noté $Div(k(\mathcal{C}))$.

On a alors un morphisme de groupes

$$\begin{aligned} \deg : Div(K) &\to \mathbb{Z} \\ \sum_{x\in\mathcal{C}} n_x[x] &\mapsto \sum_{x\in\mathcal{C}} n_x. \end{aligned}$$

Lemme 17. *Pour $f \in k(\mathcal{C})^*$, $\deg(\mathrm{div}(f)) = 0$. C'est-à-dire qu'une fonction rationnelle a autant de zéros que de pôles.*

Le groupe $Div(K)$ est muni d'un ordre partiel défini par $\sum_{x\in\mathcal{C}} n_x[x] \leq \sum_{x\in\mathcal{C}} n'_x[x]$ si $n_x \leq n'_x$ pour tout x. On définit alors pour $D \in Div(K)$, $\mathcal{L}(D) = \{0\} \cup \{f \in K^*, \mathrm{div}(f) \geq D\}$.

Théorème 9. *Pour D diviseur de K, $\mathcal{L}(D)$ est un espace vectoriel de dimension finie sur $\mathbb{F}_q$.*

On note alors $\ell(D)$ sa dimension.

Théorème 10 (Riemann-Roch)**.** *Il existe un diviseur canonique $\mathcal{K}$ tel que pour tout $D \in Div(K)$,*

$$\ell(D) - \ell(\mathcal{K} - D) = \deg D - g + 1$$

où $g = \frac{\deg \mathcal{K}}{2} + 1$

On trouve tous ces résultats dans la partie II de [5].
g est appelé le genre de la courbe $\mathcal{C}$.

4.2 Majoration de $\nu_1(\mathcal{C})$

4.2.1 Application du théorème de Riemann-Roch

Théorème 11. *Si $q > (g+1)^4$, alors*

$$\nu_1(\mathcal{C}) - q - 1 \leq (2g+1)q^{1/2}.$$

Si φ n'a pas de point fixe, le résultat est clair, on suppose donc qu'il existe un point fixe x_0.
On définit alors $R_m = \left\{f \in \overline{k}(\mathcal{C}), (f) \geq -m[x_0]\right\} (= \mathcal{L}(m[x_0])$ dans les notations du théorème de Riemann-Roch). On déduit du théorème 10 :

Lemme 18. *Pour m entier, on a*

$$\begin{aligned} \dim R_m &\leq m+1 && (8)\\ \dim R_m &\geq m+1-g \text{ avec égalité si } m > 2g-2 && (9)\\ \dim R_{m+1} &\leq \dim R_m + 1 && (10)\\ R_m \circ \varphi &\subseteq R_{mq} && (11) \end{aligned}$$

et tout élément de $R_m \circ \varphi$ est une puissance q-ème vérifiant $(f \circ \varphi) = q\varphi^{-1}((f))$.

Démonstration. En effet, si $f \in R_m$, il existe $d \leq m$

$$f(x) = \frac{\prod_{i=1}^d (x - x_i)}{(x - x_0)^d}$$

$$f \circ \varphi(x) = \frac{\prod_{i=1}^d (x^q - x_i)}{(x^q - x_0)^d} = \frac{\prod_{i=1}^d (x^q - y_i^q)}{(x^q - x_0^q)^d} = \left(\frac{\prod_{i=1}^d (x - y_i)}{(x - x_0)^d}\right)^q.$$

car φ est un automorphisme, il existe y_i tel que $x_i = \varphi(y_i)$. □

On définit aussi $R_l(p^\mu) = \left\{f^{p^\mu}, f \in R_l\right\} \subset R_{lp^\mu}$. Et pour $A \subset R_m$, $B \subset R_n$, on définit AB le sous-espace vectoriel de R_{m+n} engendré par les sommes de produits d'éléments de A par des éléments de B.

Lemme 19. *Si $lp^\mu < q$,*

$$\begin{array}{rcl} R_l(p^\mu) \otimes R_m \circ \varphi & \to & R_l(p^\mu)(R_m \circ \varphi) \\ f \otimes g & \mapsto & fg \end{array}$$

est un isomorphisme.

Démonstration. La surjectivité est claire. Pour $m \geq 1$ il existe $s_1, \ldots, s_r$ base de R_m telle que pour tout i dans $\{1, \ldots, r-1\}$, $\operatorname{ord}_{x_0} s_i < \operatorname{ord}_{x_0} s_{i+1}$ (par récurrence en utilisant (10)).

On montre que pour $\sigma_1, \ldots, \sigma_r \in R_m$, si $\sum_{i=1}^r \sigma_i^{p^\mu}(s_i \circ \varphi) = 0$ alors tous les σ_i sont nuls.

En effet soit ρ le plus petit indice tel que $\sigma_\rho \neq 0$. Alors

$$\begin{array}{rcl} \sigma_\rho^{p^\mu}.s_\rho \circ \varphi & = & \displaystyle\sum_{i=\rho+1}^{r} \sigma_i^{p^\mu}.s_i \circ \varphi \\ p^\mu \operatorname{ord}_{x_0} \sigma_\rho + q.\operatorname{ord}_{x_0} s_\rho & \geq & \displaystyle\min_{i \in \{\rho+1,\ldots,r\}} (p^\mu \operatorname{ord}_{x_0} \sigma_i + q.\operatorname{ord}_{x_0} s_i) \\ & \geq & -lp^\mu + q.\operatorname{ord}_{x_0} s_{\rho+1} \\ p^\mu \operatorname{ord}_{x_0} \sigma_\rho & \geq & -lp^\mu + q.\underbrace{(\operatorname{ord}_{x_0} s_{\rho+1} - \operatorname{ord}_{x_0} s_\rho)}_{\geq 1} > 0 \end{array}$$

donc $\operatorname{ord}_{x_0} \sigma_\rho > 0$, σ_ρ a un zéro en x_0 or il n'a pas de pôle en dehors de x_0 donc d'après le lemme 17, $\sigma_\rho \equiv 0$ ce qui mène à la contradiction. □

Corollaire 2. $\dim R_l \times \dim R_m = \dim R_l(p^\mu)(R_m \circ \varphi)$.

4.2.2 Méthode de Stepanov

On suppose toujours $lp^\mu < q$, on définit

$$\begin{array}{rcl} \delta : R_l(p^\mu)(R_m \circ \varphi) & \to & R_l(p^\mu)R_m \hookrightarrow R_{lp^\mu+m} \\ \sum \sigma_i^{p^\mu} f_i \circ \varphi & \mapsto & \sum \sigma_i^{p^\mu} f_i \end{array}$$

induit par l'isomorphisme du lemme et le morphisme linéaire $R_m \circ \varphi \to R_m$. Par le théorème du rang, on a :

$$\begin{array}{rcl} \dim \ker \delta & = & \dim R_l(p^\mu)(R_m \circ \varphi) - \dim Im\delta \\ & \geq & \dim R_l \dim R_m - \dim R_{lp^\mu+m} \\ & \geq & (l+1-g)(m+1-g) - (lp^\mu + m + 1 - g) \end{array}$$

D'après 9 car si $l, m \geq g$ alors $lp^\mu + m > 2g - 2$.

Lemme 20. *Si $\ker \delta \neq \{0\}$, alors les éléments non nuls de $\ker \delta$ s'annulent en tout point fixe de φ avec un ordre supérieur ou égal à p^μ, sauf en x_0.*

Démonstration. En effet, soit $f \in \ker\delta - \{0\}$ et $x \neq x_0$ un point fixe de φ, alors

$$\begin{aligned} f(x) &= \sum(\sigma_i^{p^\mu}(s_i \circ \varphi))(x) \\ &= \sum \sigma_i^{p^\mu}(x) s_i(x) \\ &= \delta(f)(x) = 0 \end{aligned}$$

Or les éléments de $R_l(p^\mu)$ sont des puissances p^μ-èmes, et les éléments de $R_m \circ \varphi$ sont des puissances q-èmes (comme $p^\mu < q$, p^μ divise q, et ce sont aussi des puissances p^μ-èmes). Donc f est une puissance p^μ-ème, donc ses zéros sont d'ordre p^μ. □

La fonction $f \in R_l(p^\mu)R_m \circ \varphi \subset R_{lp^\mu+qm}$ a au plus un pôle en x_0, d'ordre au plus $lp^\mu + mq$. Et comme f a autant de zéros que de pôles (lemme 17) on déduit :

$$p^\mu(\nu_1(\mathcal{C}) - 1) \leq lp^\mu + qm$$

D'où

$$\nu_1(\mathcal{C}) \leq l + \frac{q}{p^\mu}m + 1.$$

La fonction f est notre "polynôme" de la méthode de Stepanov, au lieu de majorer son degré, on sait majorer son nombre de pôles.

On a une telle majoration si $l, m \geq g$, $lp^\mu < q$ et $(l+1-g)(m+1-g) > lp^\mu + m + 1 - g$. On choisit $\mu = \frac{\alpha}{2}$, $m = p^\mu + 2g$, $l = \lfloor \frac{g}{g+1}p^\mu \rfloor + g + 1$. Alors on a bien $l, m \geq g$,

$$\begin{aligned} lp^\mu &= \lfloor \frac{g}{g+1}p^\mu \rfloor p^\mu + (g+1)p^\mu \\ &\leq \frac{g}{g+1}p^{2\mu} + (g+1)p^\mu = \frac{gq + (g+1)^2 p^\mu}{g+1} \\ &< \frac{gq+q}{g+1} = q \end{aligned}$$

et

$$\begin{aligned} (l+1-g)(m+1-g) - (lp^\mu + m + 1 - g) &= (\lfloor \frac{g}{g+1}p^\mu \rfloor + 2)(p^\mu + g + 1) - (\lfloor \frac{g}{g+1}p^\mu \rfloor p^\mu + (g+1)p^\mu + p^\mu + g + 1) \\ &= \lfloor \frac{g}{g+1}p^\mu \rfloor (g+1) + 2(g+1) - gp^\mu - g - 1 \\ &> (\frac{g}{g+1}p^\mu - 1)(g+1) - gp^\mu + g + 1 = 0. \end{aligned}$$

Alors

$$\begin{aligned} \nu_1(\mathcal{C}) &\leq l + \frac{q}{p^\mu}m + 1 \\ &\leq q + (2g+1)q^{1/2} + 1. \end{aligned}$$

Ce qui démontre le théorème 11.

Il faut remarquer ici qu'on a utilisé le fait que q est un carré parfait pour avoir l'ordre de la majoration le meilleur possible. En effet, si α était impair, on choisirait $\alpha = 2\mu + 1$ pour avoir les mêmes conditions que précédemment, mais alors on obtiendrait $\nu_1(\mathcal{C}) \leq q + (2g+1)(qp)^{1/2} + 1$.

4.3 Preuve du théorème 8

Théorème 12. *On a l'estimation*

$$\nu_1(\mathcal{C}) = q + O(q^{1/2})$$

Démonstration. D'après le lemme 16, $\bar{k}(\mathcal{C})$ corps des fonctions de $\mathcal{C}$ contient un sous corps $\bar{k}(t)$ purement transcendant sur $\bar{k}$.

Soit N la clôture galoisienne de $\bar{k}(\mathcal{C})/\bar{k}(t)$, alors N est le corps des fonctions d'une courbe $\mathcal{C}'$.

$$\begin{array}{rcl} N & \rightsquigarrow & \mathcal{C}' \\ \uparrow & & \downarrow \\ \bar{k}(\mathcal{C}) & \rightsquigarrow & \mathcal{C} \\ \uparrow & & \downarrow \\ \bar{k}(t) & \rightsquigarrow & \mathbb{P}^1 \end{array}$$

Soit g' le genre de $\mathcal{C}$, G le groupe de Galois de l'extension $\bar{k}(\mathcal{C}')/\bar{k}(t)$, H sous-groupe de G correspondant à $\bar{k}(\mathcal{C}')/\bar{k}(\mathcal{C})$. On voit l'action de G sur $\bar{k}(\mathcal{C}')$ comme une action sur $\mathcal{C}'$.

Soit $x \in \mathbb{P}^1_{\bar{k}}$, non ramifié, rationnel sur k, y un point de $\mathcal{C}'$ au dessus de x. Alors $\varphi(y) = \eta \cdot y$ pour un certain η dans G. Posons

$$\nu_1(\mathcal{C}', \eta) = \#\{y \in \mathcal{C}' \mid \exists x \in \mathbb{P}^1_{\bar{k}}, x \text{ rationnel sur } k \text{ non ramifié dans } \mathcal{C}'/\mathbb{P}^1_{\bar{k}} \text{ avec } \varphi(y) = \eta \cdot y\}$$

Alors de la même façon que dans la preuve du théorème 8, mais en regardant cette fois

$$\begin{array}{rcl} \delta_\eta : R_l(p^\mu)(R_m \circ \varphi) & \to & R_l(p^\mu)(R_m \circ \eta) \\ \sum \sigma_i^{p^\mu} f_i \circ \varphi & \mapsto & \sigma_i^{p^\mu} f_i \circ \eta, \end{array}$$

on déduit

$$\nu_1(\mathcal{C}', \eta) \leq q + (2g' + 1)q^{1/2} + 1.$$

On a aussi

$$\begin{array}{rcl} \displaystyle\sum_{\eta \in G} \nu_1(\mathcal{C}', \eta) & = & \#\{y \in \mathcal{C}' \mid \exists x \in \mathbb{P}^1_{\bar{k}} \text{ rationnel sur } k \text{ et non ramifié } \} \\ & = & |G|\nu_1(\mathbb{P}^1_{\bar{k}}) + O(1) = |G|(q+1) + O(1), \end{array}$$

où le $O(1)$ vient des points de $\mathbb{P}^1$ qui sont k-rationnels mais ramifiés. On en déduit que $\nu_1(\mathcal{C}', \eta) = q + O(q^{1/2})$. De plus

$$\sum_{\eta \in H} \nu_1(\mathcal{C}', \eta) = \nu_1(\mathcal{C}) + O(1)$$

ainsi

$$\nu_1(\mathcal{C}) = q + O(q^{1/2}).$$

□

On a pour tout entier r, $\nu_r(\mathcal{C}) = q^r - \sum_{i=1}^{2g} \omega_i^r = q^r + O(q^{r/2})$. Donc $\sum_{i=1}^{2g} \omega_i^r = O(q^{r/2})$, où la constante dans le O est indépendante de q. C'est-à-dire qu'il existe $A > 0$ tel que $|\sum_{i=1}^{2g} \omega_i^r| \leq A q^{r/2}$.

Alors, d'après le lemme 10, pour tout i

$$|\omega_i| \leq q^{1/2}.$$

Comme l'équation fonctionnelle de $Z(\mathcal{C}, T)$ donne $\omega_i \omega_{2g-i} = q$, on a pour tout i, $|\omega_i| = q^{1/2}$. Ainsi

$$|\nu_r(\mathcal{C}) - q^r - 1| = |\sum_{i=1}^{2g} \omega_i^r| \leq 2gq^{r/2}.$$

Ce qui conclut.

Références

[1] E. Bombieri. Counting points on curves over finite fields. In *Séminaire N. Bourbaki*, number exp. n430, 1972-1973.

[2] D. R. Heath-Brown. An estimate for Heilbronn's exponential sum. In *Analytic number theory : Proceedings of a conference in honor of Heini Halberstam*, pages 451–463, 1996.

[3] D. R. Heath-Brown and S. Konyagin. New bounds for Gauss sums derived from kth powers, and for Heilbronn's exponential sum. *Quart. J. Math.*, (51) :221–235, 2000.

[4] H. Iwaniec and E. Kowalski. *Analytic Number Theory*. Number 53. AMS, 2004.

[5] J. H. Silverman. *The Arithmetic of Elliptic Curves*. Number 106. Springer-Verlag Graduate Texts in Mathematics, 1986.