QUELQUES IDÉES SUR L'ALGORITHMIQUE DES ÉQUATIONS DIOPHANTIENNES

par

Guillaume Hanrot

1. Introduction

Ce mini-cours a pour objectif de s'intéresser, de façon très sommaire, aux aspects transcendants de l'algorithmique des équations diophantiennes.

Avant toute chose, un *caveat* s'impose : l'esprit du présent texte est celui d'un exposé semi-formel des idées de l'algorithmique des équations diophantiennes. On a cherché à mettre en avant les idées, et à expliquer les manipulations qui doivent être effectuées, plus qu'à énoncer une longue suite de théorèmes à constantes explicites. Le lecteur trouvera toutes les constantes nécessaires dans les différents articles auquel on renvoie dans le texte.

On appelle usuellement *équation diophantienne* une équation de la

$$P(x_1, \dots, x_n) = 0, \quad P \in \mathbb{Z}[X_1, \dots, X_n],$$

où les inconnues $x_1, \dots, x_n$ sont cherchées dans $\mathbb{Z}$. Il s'agit d'une situation très différente de la situation où les inconnues sont cherchées dans un corps algébriquement clos, la contrainte arithmétique limitant généralement de façon drastique le nombre de solutions. Dans ce texte, nous nous intéresserons également à des équations de type différent (équations diophantiennes exponentielles), où les inconnues peuvent intervenir comme exposants.

1.1. Problème de Hilbert et limites intrinsèques. — Il est difficile de commencer un texte traitant de l'algorithmique des équations diophantiennes sans mentionner le dixième problème de Hilbert. Au congrès de 1900, dans la liste de ses 23 problèmes, Hilbert pose la question de la conception d'une méthode générale qui, étant donné un polynôme P à n variables et à coefficients entiers, décide s'il existe des entiers $x_1, \ldots, x_n$ tels que $P(x_1, \ldots, x_n) = 0$. Une question hélas aussi ambitieuse que vaine : des travaux de logiciens culminant dans le résultat de Matjasevitch [**12**] ont montré que l'existence d'une solution entière est, en toute généralité, indécidable : la méthode générale que demande Hilbert n'existe pas.

Il faut toutefois nuancer les conséquences de ce résultat : si le problème, en toute généralité, est indécidable, rien n'empêche que pour des familles d'équations des algorithmes existent. Et nous verrons que c'est effectivement le cas.

En tout état de cause, un tour d'horizon rapide de l'existant limite rapidement les ambitions : on ne connaît pas de méthode, actuellement, qui permette en toute rigueur de décider si une courbe $y^2 = q(x)$, avec q de degré 4, admet ou non un point rationnel sur $\mathbb{Q}$...

1.2. Quelques grandes familles de méthodes. — On peut s'essayer à une taxonomie grossière des méthodes existantes.

La méthode la plus élémentaire (ne pas l'oublier, elle prouve l'absence de solutions dans bien des cas !) consiste en l'utilisation d'arguments de congruence. Ce type d'arguments permet, en exhibant un nombre premier p (ou un idéal premier $\mathfrak{p}$ d'un corps de nombres bien choisi) modulo lequel l'équation n'a aucune solution, de prouver que l'équation n'a aucune solution dans $\mathbb{Z}$. Plus généralement, quand une équation a peu de solutions modulo des premiers (typiquement, pour une équation de la forme $y^p = f(x)$, on s'attend à ce que les solutions soient peu nombreuses modulo $\ell \equiv 1 \bmod p$, car les puissances p-èmes modulo ℓ sont alors en proportion $1/p$), on peut utiliser le théorème chinois pour construire efficacement toutes les solutions potentielles inférieures à une borne donnée ; on atteint toutefois assez vite les limites de ce type de méthode.

Certaines équations sont justiciables de techniques purement arithmétiques (pgcd, fractions continues, algèbre linéaire), développées dans le cours de Karim Belabas ; nous examinerons sommairement dans une première partie le cas de l'équation de Pell.

Des méthodes plus sophistiquées, que nous explorerons dans ce texte, utilisent des arguments de transcendance. Nous nous concentrerons sur un aspect, dont l'algorithmique est bien établie, qui s'appuie sur la théorie des bornes inférieures pour les formes linéaires de logarithmes de nombres algébriques. Plus généralement, l'idée de ce type de méthodes est que les nombres algébriques jouissent de propriétés arithmétiques fortes (par exemple, le théorème de Roth affirme qu'ils sont mal approchés par des rationnels). Certains de ces résultats sont de nature effective, et fournissent dès lors des informations sur le nombre, ou – mieux – la taille des solutions potentielles de l'équation. Souvent, ces informations sont difficiles à exploiter (information de nombre peu exploitable, ou information de taille rendant impossible l'énumération), mais nous verrons comment, dans certains cas, une combinaison d'ingrédients algorithmiques et diophantiens permet de résoudre les problèmes.

Ce type de méthode sera d'abord illustré par des équations dont le traitement algébrique est élémentaire, de façon à se concentrer sur le processus algorithmique ; par la suite, ce texte culminera avec l'exposé de la méthode de Tzanakis et de Weger [**18**] pour la résolution de l'équation de Thue, dont le traitement algorithmique est quasiment le même, mais qui présuppose un travail préalable de réduction de l'équation à la forme souhaitée.

Enfin, les méthodes les plus avancées reposent sur l'utilisation de techniques issues de la géométrie arithmétique. La méthode la plus explorée dans cette direction est sans doute la méthode de Chabauty. À une courbe algébrique, la géométrie arithmétique permet d'associer une variété algébrique munie naturellement d'une structure de groupe – la jacobienne – ; le groupe des points sur $\mathbb{Q}$ est alors de type fini. Quand, de surcroît, le rang de la partie libre dudit groupe est plus petit que sa dimension, un argument dû à Chabauty prouve de façon effective la finitude du nombre de solutions. La restriction sur le rang du groupe n'est pas bénigne, mais diverses techniques permettent souvent de la contourner. Ces techniques sont sans doute les plus puissantes, et d'un point de vue algébrique et géométrique, les plus intrinsèques et élégantes. Elles nous emmèneraient toutefois trop loin, et nous ne les discuterons pas.

Le lecteur intéressé par une vision plus complète de la théorie et de l'algorithmique des équations diophantiennes est invité à consulter l'ouvrage de Smart [**15**], ou l'ouvrage d'Henri Cohen, en cours d'édition à l'heure où ces notes sont écrites.

2. Un aspect élémentaire

Cette partie se contente, essentiellement, de replacer dans une problématique générale (et donc, de renvoyer au texte de K. Belabas) un cas simple d'équations diophantiennes, le cas de l'équation de Pell-Fermat.

2.1. L'équation de Pell-Fermat. — On désigne sous ce nom l'étude de l'équation

$$x^2 - Dy^2 = 1,$$

ou plus généralement de l'équation $x^2 - Dy^2 = a$, où D est un nombre entier positif.

2.1.1. Cas où le second membre est 1. — On peut d'ores et déjà situer assez précisément les solutions de cette équation :

Proposition 1. — *Soit D un nombre entier, avec $D = fd^2$, f sans facteur carré. Si (x, y) est une solution de $x^2 - Dy^2 = 1$, alors $x - dy\sqrt{f}$ est une unité du corps $\mathbb{Q}(\sqrt{f})$. Inversement, à toute unité $l + m\sqrt{f}$ de norme 1 de ce corps avec $d|m$ on peut associer une solution de l'équation de Pell-Fermat.*

Démonstration. — L'équation de Pell-Fermat dit simplement que la norme de $x - dy\sqrt{f}$ vaut 1 ; or c'est un entier, c'est donc une unité. □

Notons que les unités de la forme indiquée constituent un sous-groupe de l'ensemble de toutes les unités ; étant donné $\varepsilon = \varepsilon_1 + \varepsilon_2\sqrt{f}$ une unité fondamentale (supposons pour simplifier que $f \equiv 1 \bmod 4$, de sorte que ε_1 et ε_2 sont dans $\mathbb{Z}$), il suffit de trouver le plus petit entier j tel que ε^j soit du type souhaité.

On peut voir les ε^j comme des éléments de l'anneau $\mathbb{Z}[X]/(X^2 - f)$; on déduit alors :

Proposition 2. — *Soit p un nombre premier ; les entiers k tels que $\varepsilon^k = \varepsilon_{1,k} + \varepsilon_{2,k}\sqrt{f}$ avec $p|\varepsilon_{2,k}$ sont les multiples de l'ordre de $\varepsilon \bmod p$ dans $(\mathbb{F}_p[X]/(X^2 - f))^*/\mathbb{F}_p^*$.*

En particulier, si f est un carré modulo p, cet ordre divise $p-1$, et si f est un non-carré modulo p, cet ordre divise $p+1$.

On conclut cette section en relevant que quand le membre de droite est 1, l'équation de Pell admet de façon systématique une infinité de solutions.

2.1.2. Membre droit quelconque. — Dans le cas général (membre de droite $\neq 1$), la stratégie consiste, dans un premier temps, à déterminer un ensemble complet de solutions de l'équation aux normes $N_{\mathbb{Q}(\sqrt{D})/\mathbb{Q}}(u) = a$, modulo l'action du groupe des unités. Si cet ensemble, disons E, est fini, on sait alors que pour tout u solution, il existe une unité η et $e \in E$ tel que $u = \eta e$; inversement, modulo l'étude du signe de la norme et les conditions de divisibilité comme dans la partie précédente, on obtient ainsi toutes les solutions. On est donc ramené à l'étude de l'équation aux normes.

2.1.3. Équation aux normes

Lemme 1. — *Soit K un corps de nombres ; il n'existe qu'un nombre fini d'entiers algébriques de norme donnée, modulo l'action du groupe des unités.*

Démonstration. — Il suffit *a fortiori* de prouver qu'il n'existe qu'un nombre fini d'idéaux de norme donnée ; en décomposant a, on peut de plus se limiter au cas où a est une puissance de nombre premier, $a = p^t$. Dans ce cas, tout idéal de norme a est parmi les $\mathfrak{p}_i^l$, $l \leqslant t$, $\mathfrak{p}_i$ un idéal au-dessus de p, donc dans un ensemble fini. $\square$

La preuve de ce lemme est essentiellement effective ; on peut même la rendre plus efficace en cherchant les idéaux de norme p^t sous la forme $\prod_{i=1}^{l} \mathfrak{p}_i^{j_i}$, ce qui conduit à une équation linéaire $\sum j_i f_i = t$, à laquelle on adjoint des équations linéaires pour traduire le fait que l'idéal est bien principal : si $\mathrm{Cl}(K)$ de la forme $\prod_{\ell=1}^{k} \mathbb{Z}/n_\ell\mathbb{Z}$, et que ϕ_ℓ désigne la projection d'un idéal sur une composante de $\mathrm{Cl}(K)$, on veut de surcroît que $\sum j_i \phi_\ell(\mathfrak{p}_i) \equiv 0 \bmod n_\ell$ pour tout ℓ. On est donc ramené... à résoudre un système d'équations diophantiennes linéaires, qui se résout par des techniques d'algèbre linéaire entière (mise sous forme normale de Smith, par exemple).

2.2. Le problème des bœufs d'Archimède. — Ce célèbre problème est généralement attribué à Archimède. Il a été découvert dans un manuscrit grec conservé dans une bibliothèque du nord de l'Allemagne en 1773. Le texte propose de compter les troupeaux du dieu du soleil, et en substance, il s'énonce de la façon suivante : il s'y trouve des taureaux et vaches de 4 couleurs différentes, blancs, noirs, tachetés et marrons. Pour les taureaux, le nombre de blancs est plus grand que le nombre des marrons de $1/2+1/3$ du nombre des noirs ; le nombre des noirs plus grand que les marrons de $1/4+1/5$ du nombre des tachetés ; le nombre des tachetés plus grand que le nombre des marrons de $1/6+1/7$ du nombre des blancs.

Pour les vaches, le nombre des blanches est $1/3+1/4$ du nombre total de têtes de bétail noires ; le nombre des noires, $1/4+1/5$ du nombre total de têtes de bétail tachetées ; le nombre de tachetées, $1/5+1/6$ du nombre total de têtes de bétail marrons ; le nombre des marrons, $1/6+1/7$ du nombre total de têtes de bétail blanches.

Ce problème se retranscrit simplement en le système de 7 équations à 8 inconnues suivant :

$$(1) \qquad \begin{cases} b = m + 5/6n \\ n = m + 9/20t \\ t = m + 13/42b \\ B = 7/12(n+N) \\ N = 9/20(t+T) \\ T = 11/30(m+M) \\ M = 13/42(b+B) \end{cases}$$

qui conduit à la solution, z étant un paramètre entier quelconque,

$$(2) \qquad \begin{cases} b = 10366482z, \\ n = 7460514z, \\ m = 4149387z, \\ t = 7358060z, \\ B = 7206360z, \\ N = 4893246z, \\ M = 5439213z, \\ T = 3515820z. \end{cases}$$

Le problème inclut toutefois une seconde partie : les taureaux blancs et noirs réunis peuvent être rangés en carré, et les taureaux marrons et tachetés en triangle. Ces deux contraintes conduisent alors à l'équation, avec x, y, z des inconnues entières :

$$x^2 = 17826996z, \quad y(y+1)/2 = 11507447z.$$

En éliminant z, on obtient $2471x^2 = 1914y(y+1)$. Cela impose en particulier que $1914|x$; posant $x = 1914X$, on a

$$y(y+1) - 4729494X^2 = 0,$$

soit encore, en posant $t = 2y + 1$,

$$t^2 - 4729494(2X)^2 = 1.$$

Pour que z soit bien entier, il nous faut finalement imposer que $4657|X$; noter que toute solution de $t^2 - 4729494u^2 = 1$ a toujours u pair (regarder l'équation modulo 8).

Notons ε l'unité fondamentale de $\mathbb{Q}(\sqrt{4729494})$. On cherche une puissance de ε de la forme $\varepsilon_1 + \varepsilon_2\sqrt{4729494}$ avec $2|\varepsilon_2$ et $4657|\varepsilon_2$. La discussion générale menée dans le paragraphe précédent montre alors que la plus petite solution est de la forme $\varepsilon^2, \varepsilon^{2329}$ ou ε^{4658}. C'est cette dernière solution qui est la bonne, et comme on trouve, par exemple avec GP, que $\log|\varepsilon| \approx 102$, la solution minimale au problème d'Archimède a donc un logarithme de l'ordre de $102 \cdot 2 \cdot 2329 = 475116$, et a en fait 206545 chiffres décimaux... Pour plus de détails, on pourra consulter [**19**].

3. Méthodes transcendantes

Dans cette partie, nous allons étudier les méthodes dites « transcendantes » pour les équations diophantiennes.

3.1. Approximation diophantienne des nombres algébriques

Il faut dire que l'étude de l'approximation diophantienne des nombres algébriques et transcendants et l'étude des équations diophantiennes ont souvent avancé de concert. L'objet général des aspects « transcendants » de la théorie des nombres est l'étude, dans le cas le plus simple, de l'approximation des nombres réels par des rationnels. La réponse la plus générale est fournie par le théorème de Dirichlet :

Théorème 1. — *Soit x un nombre réel, et Q un nombre entier strictement positif. Il existe un entier $q \leqslant Q$ tel que $d(qx, \mathbb{Z}) \leqslant 1/Q$.*

En particulier, il existe (p, q) tels que $|x - p/q| \leqslant 1/q^2$. On peut se demander, de façon plus générale, si 2 est l'exposant optimal dans cette proposition. Une question naturelle se pose : pour quel type de x l'exposant 2 est-il optimal ?

Rappelons la définition suivante :

Définition 1. — Un nombre complexe x est *algébrique* s'il existe un polynôme P non nul, à coefficients entiers tel que $P(x) = 0$.

Quand x est algébrique, il est alors relativement simple d'estimer à quel point x est approché par des rationnels p/q. En effet, si $P(x) = 0$ avec P irréductible sur $\mathbb{Q}$, on peut former $|P(p/q) - P(x)| = P(p/q)$ qui, en vertu de l'inégalité des accroissements finis sera, pour p/q assez proche de x, de l'ordre de $|x - p/q|$. Cependant, $P(p/q)$ est un rationnel non nul, de dénominateur $1/q^{\deg(P)}$. Il vient que

$$|x - p/q| \geqslant C(x)/q^{\deg(P)}.$$

En particulier, cet argument de Liouville, pour les nombres algébriques quadratiques ($\deg P = 2$) montre que le théorème de Dirichlet est optimal. Dans le cas général, il faut attendre le XX^e siècle pour obtenir les résultats suivants, dans l'ordre chronologique :

***Théorème 2* (Thue-Siegel-Dyson-Roth).** — *Soit x algébrique de degré d. Alors, pour tout $\varepsilon > 0$, il existe $C(x, \varepsilon)$ tel que pour tout p/q, on ait*

$$|x - p/q| > C(x, \varepsilon)/q^{f(d)+\varepsilon},$$

où

- $f(d) = d/2 + 1$ (*Thue,* [**17**]) ;
- $f(d) = 2\sqrt{d}$ (*Siegel,* [**14**]) ;
- $f(d) = \sqrt{2d}$ (*Dyson,* [**8**] – *et, simultanément, Gelfond* [**9**]) ;
- $f(d) = 2$ (*Roth,* [**13**]).

Le dernier résultat, dû à Roth, clôt quasiment le problème, puisque seules deux questions subsistent : préciser le ε du théorème, et (!) surtout, préciser la constante $C(x, \varepsilon)$, qui n'est effective dans aucun des résultats mentionnés. À ce titre, ces résultats, qui donnent un panorama très précis de l'approximation des algébriques par des rationnels, permettent aisément d'obtenir des résultats de finitude de nombre de solutions, mais nous sont inutiles d'un point de vue algorithmique... pour lequel une connaissance, même très grossière, de $C(x,\varepsilon)$, quitte à perdre sur l'exposant, est requise.

Avant de continuer, on peut donner une justification heuristique du résultat de Roth : presque tous les nombres réels ont des meilleures approximations à l'ordre 2 exactement. En effet, si X_f est l'ensemble des réels de $[0,1]$ admettant une suite d'approximations p_n/q_n avec $|x - p_n/q_n| \leqslant 1/f(q_n)$, avec $f(x) \to \infty$ quand $q \to \infty$, alors

$$X_f = \bigcap_{q_0=1}^{\infty} \bigcup_{q\geqslant q_0} \bigcup_{0\leqslant p\leqslant q} \left[\frac{p}{q} - \frac{1}{f(q)}, \frac{p}{q} + \frac{1}{f(q)}\right].$$

L'intersection étant décroissante et les ensembles de mesure finie, il vient

$$\begin{aligned}\mu(X_f) &= \lim_{q_0\to\infty} \mu\biggl(\bigcup_{q\geqslant q_0} \bigcup_{0\leqslant p\leqslant q} \left[\frac{p}{q} - \frac{1}{f(q)}, \frac{p}{q} + \frac{1}{f(q)}\right]\biggr)\\ &\leqslant \lim_{q_0\to\infty} \sum_{q\geqslant q_0} \frac{2(q+1)}{f(q)}.\end{aligned}$$

En particulier, si la série $\sum_{q\geqslant 1} q/f(q)$ converge, $\mu(X_f) = 0$. C'est le cas par exemple pour $f(x) = x^\alpha$ pour $\alpha > 2$.

3.2. Formes linéaires de logarithmes et résultats de Baker

Pour l'algorithmique des équations diophantiennes, le « salut » vient d'un point de vue assez différent, exploré par Gelfond et Schneider pour l'étude du 13-ème problème de Hilbert, puis généralisé par A. Baker dans les années 1960 dans une série d'articles [**1, 2**].

Le procédé consiste, *via* une étude algébrique de l'équation, à construire une *forme linéaire de logarithmes*

$$\Lambda(b_1,\ldots,b_n) = \left|\sum_{i=1}^{r} b_i \operatorname{Log} u_i\right| \tag{3}$$

où les quantités liées aux inconnues initiales sont les b_i, les autres quantités étant explicites et ne dépendant que de l'équation. Ici et dans la suite, Log est la détermination principale du logarithme complexe.

La quantité $\Lambda(b_1,\ldots,b_n)$ doit de plus avoir la propriété que pour toute solution de l'équation initiale, on doit pouvoir construire un n-uplet $(b_1,\ldots,b_n)$ tel que $\Lambda(b_1,\ldots,b_n)$ soit très petit, typiquement :

$$\Lambda(b_1,\ldots,b_n) \ll \exp(-C\cdot\max_i|b_i|). \tag{4}$$

La forme Λ (on parle de forme linéaire en logarithmes) est en général construite en prenant le logarithme d'une quantité très voisine de 1, et qui dépend donc de façon exponentielle des b_i. On va donc rencontrer naturellement ce type de méthodes lors de l'étude d'équations diophantiennes qui « cachent » une structure de groupe multiplicatif de type fini.

Intuitivement, une propriété telle que (4) a un caractère exceptionnel. L'équivalent n-dimensionnel du théorème de Dirichlet – conséquence aisée du principe des tiroirs – prédit que parmi les N^n valeurs de $|\sum_{i=1}^n b_i\alpha_i|$ pour $(b_i)\in[1,N]^n$ qui se trouvent dans $[0,N\sum_{i=1}^n|\alpha_i|]$, on peut en trouver deux distantes d'au plus $N^{1-n}\sum_{i=1}^n|\alpha_i|$, et donc une combinaison linéaire de cet ordre de grandeur avec $|b_i|\leqslant N$. Cette estimation est en fait essentiellement optimale : pour presque tout $(\alpha_1,\ldots,\alpha_n)$, on ne peut pas espérer obtenir un meilleur exposant.

Cet argument n'est bien entendu pas suffisant... car l'arithmétique de la situation est à prendre en compte. Mais dans le cas présent, nos nombres α_i sont bien particuliers ; ce sont des logarithmes de nombres algébriques. Dans ce cas, le résultat de Baker (qui lui a valu la médaille Fields en 1966), largement précisé, raffiné, et étendu depuis, nous fournit l'estimation dont nous avons besoin :

Théorème 3. — *Soit $u_1,\ldots,u_n$ des nombres algébriques. Il existe une constante $C(u_1,\ldots,u_n)$ telle que, pour tout n-uplet $(b_1,\ldots,b_n)$, si l'on pose $\Lambda(b_1,\ldots,b_n)=\sum b_i\operatorname{Log}u_i$, on a*

– soit $\Lambda(b_1,\ldots,b_n)=0$ (et les u_i sont multiplicativement dépendants) ;

– soit $\Lambda(b_1,\ldots,b_n) > \exp(-C(u_1,\ldots,u_n)\log\max|b_i|)$, où $C(u_1,\ldots,u_n)$ est une constante positive ne dépendant que des u_i.

On peut comparer ce résultat, dans le cas où les u_i sont des nombres entiers, à la méthode de Liouville. Dans ce cas, Λ est le logarithme d'un nombre $\prod u_i^{b_i}$; si ce nombre est différent de 1, sa distance à 1 est alors au moins égale à l'inverse de son dénominateur, qui est au plus $\prod |u_i|^{|b_i|} \leqslant (\prod |u_i|)^{\max |b_i|}$. En particulier, l'estimation qui est obtenue pour Λ est alors $\exp((\sum \log |u_i|) \max |b_i|)$, dans laquelle la dépendance en $B := \max |b_i|$ est bien moins bonne que précédemment. En revanche, la dépendance en les $\log |u_i|$ est, elle, meilleure que ce que l'on obtient : il va falloir, en général, remplacer la somme par un produit.

L'étude détaillée de cet exemple montre que si $u_i = p_i/q_i$ devient un rationnel, il faut remplacer $\log |u_i|$ par $\log \max(|p_i|, |q_i|)$. La généralisation de cette quantité pour un nombre algébrique est la *hauteur logarithmique absolue* :

Définition 2. — Soit α un nombre algébrique de degré d, $P = a_0 x^n + \sum_{d=0}^{n-1} a_{n-d} x^d \in \mathbb{Z}[X]$ son polynôme minimal, $\alpha_1, \ldots, \alpha_n$ ses racines. La hauteur logarithmique absolue de α est $1/n \cdot \log\left(|a_0| \prod_{i=1}^{n} \max(1, |\alpha_i|)\right)$.

Nous pouvons maintenant énoncer le théorème de Baker et Wüstholz [**4**] :

Théorème 4. — *Dans le théorème 3, on peut prendre*

$$C(u_1, \ldots, u_n) = 18(n+1)!\, n^{n+1} (32D)^{n+2} h(u_1) \cdots h(u_n),$$

où $D = [\mathbb{Q}(u_1, \ldots, u_n) : \mathbb{Q}]$ *et* $h(\cdot)$ *désigne la hauteur logarithmique absolue.*

Nous utiliserons cette version de la borne de Baker en raison de sa simplicité. De multiples (certaines meilleures, comme les récents travaux de Matveev) versions existent, en fonction des hypothèses plus ou moins fortes que l'on peut faire sur les u_i et sur n.

Certaines offrent davantage de liberté à l'« utilisateur » en s'exprimant en fonction de paramètres qui peuvent être optimisés selon la situation. Dans le cas présent, on verra que seul l'ordre de grandeur de C importe réellement. Une seule situation justifie alors réellement l'utilisation d'une expression nettement plus précise ; il s'agit du cas où l'on n'a que deux logarithmes, où l'on a le résultat suivant dû à Laurent, Mignotte et Nesterenko **[10]** :

Théorème 5. — *Soit* α_1, α_2 *deux nombres algébriques, et* $\Lambda = b_1 \log \alpha_1 - b_2 \log \alpha_2$. *On pose* $D = [\mathbb{Q}(\alpha_1, \alpha_2) : \mathbb{Q}]/[\mathbb{R}(\alpha_1, \alpha_2 : \mathbb{R})]$, *et on suppose donnés des réels* h_i *tels que*

$$h_i \geqslant \max(h(\alpha_i), \log \alpha_i / D, 1/D).$$

Alors, en posant $b = b_1/h_2 + b_2/h_1$, *on a*

$$\log |\Lambda| \geqslant -30.9 D^4 \max(\log b, 21/D, 1/2)^2 h_1 h_2.$$

La dépendance en $\log \max |b_i|$ est un peu moins bonne que précédemment, mais ceci est, pour notre usage, largement compensé par une constante nettement plus petite...

3.3. Équations exponentielles et conjecture ABC. — Plutôt que de donner un exposé abstrait et fastidieux du versant algorithmique du problème, étudions un cas où l'aspect « modélisation algébrique » est quasi-absent, et où l'on se limite à appliquer la borne de Baker, pour ensuite embrayer directement sur l'étude algorithmique.

Fixons $\{p_1, \ldots, p_r\}, \{q_1, \ldots, q_s\}$ des nombres premiers et un entier a. On s'intéresse aux solutions $b_1, \ldots, b_r, c_1, \ldots, c_s > 0$ de l'équation

$$\prod_{i=1}^{r} p_i^{b_i} - \prod_{i=1}^{s} q_i^{c_i} = a.$$

Ces solutions donnent, en règle général, de bons exemples de valeurs extrêmes pour la conjecture suivante, dite *abc*, et due à Masser et Oesterlé.

Conjecture 1. — *Pour tout* $\varepsilon > 0$, *il existe* $C(\varepsilon) \in \mathbb{R}$ *tel que pour tout triplet d'entiers naturels* (a, b, c) *tels que* $(a, b) = 1$ *et* $a + b = c$, *on a*

$$c \leqslant C(\varepsilon) N(abc)^{1+\varepsilon},$$

où $N(abc) = \prod_{p|abc} p$ *est le noyau sans facteur carré de abc.*

En d'autres termes, quand $a + b = c$, les trois entiers a, b, c ne sauraient avoir simultanément trop de facteurs multiples. Certains des exemples que l'on peut trouver par une stratégie voisine de celle décrite ici sont, typiquement, $1 + 2 \cdot 3^7 = 5^4 \cdot 7$, $11^2 + 3^2 \cdot 5^6 \cdot 7^3 = 2^{21} \cdot 23$ (tous deux dus à de Weger). Le « pire » (*i.e.*, qui maximise

$c/\log N(abc)$) exemple connu (trouvé par Reyssat) est donné par $2 + 3^{10} \cdot 109 = 23^5$.

3.3.1. Un cas vraiment simple. — On va commencer par le cas très particulier de l'équation $3^x - 2^y = 1$. Bien sûr, cette équation peut très facilement être résolue par des méthodes élémentaires (ce fut fait par L. Ben Gerson au milieu du XII$^{\text{e}}$ siècle : on peut typiquement remarquer que x est nécessairement pair, puis déduire la solution du fait que $3^{x/2} \pm 1$ sont tous deux des puissances de 2), mais elle nous permet d'illustrer notre propos de façon très simple.

De l'équation, on tire

$$\left|\frac{3^x}{2^y} - 1\right| \leqslant \frac{1}{2^y}.$$

En utilisant l'inégalité $|\log(1+x)| \leqslant |x|$, on voit que $|\log(3^x/2^y)| \leqslant \left|\frac{3^x}{2^y} - 1\right|$, d'où notre inégalité du type (4) :

$$|x \log 3 - y \log 2| \leqslant \frac{1}{2^y} = \frac{1}{2^{\max(|x|,|y|)}}. \tag{5}$$

Noter que pour parvenir à exprimer la borne supérieure en terme de $\max(|x|, |y|)$, il nous a fallu comparer les différents b_i (ici x et y) ; c'est le cas en général.

En outre, on déduit immédiatement de cette borne le lemme suivant, qui permet d'énumérer très rapidement toutes les solutions en-deçà d'une borne fixée M :

Lemme 2. — *Si (x, y) est une solution de $3^x - 2^y = 1$, avec $y \geqslant 2$, alors x/y est une réduite du développement en fraction continue de* $\log 2/\log 3$.

Démonstration. — En effet, on a alors $|x - y\log 2/\log 3| \leqslant \frac{1}{2^y \log 3} \leqslant 1/(2y)$. □

La borne de Laurent, Mignotte et Nesterenko fournit

$$|x \log 3 - y \log 2| \geqslant \exp(-34 \cdot \log\max(y/\log 3, x))$$

dès lors que le membre de gauche est non nul (l'inverse signifierait que $3^x = 2^y$...). Comme $y/\log 3 \approx x/\log 2 > x$, on voit que $34 \cdot (\log(|y|/\log 3))^2 \geqslant |y| \log 2$, d'où $y \leqslant 3095$.

Mais la théorie des fractions continues nous indique que la plus petite valeur de $|x\log 3-y\log 2|$ pour $|y|\leqslant 3095$ est fournie par la dernière réduite p/q du développement en fraction continue de $\log 2/\log 3$ vérifiant $q\leqslant 3095$. En particulier, cette dernière réduite étant 1054/665, on voit que $|x\log 3-y\log 2|\geqslant |665\log 3-1054\log 2|\geqslant 4\cdot 10^{-5}$; par suite, on trouve $2^{-y}\geqslant 4\cdot 10^{-5}$, soit $y\leqslant 14$...

Contrairement à ce que l'on pourrait croire, l'histoire ne s'arrête pas là, même si, dans le cas présent, l'estimation obtenue suffit. Cette nouvelle borne étant plus petite que l'ancienne, on peut raffiner le minorant de $|x\log 3-y\log 2|$ obtenu par les fractions continues, ce qui permet à nouveau de raffiner la borne sur x et y, qui permet à nouveau d'améliorer la borne sur $|x\log 3-y\log 2|$. Ce processus s'appelle l'étape de *réduction de la borne.* Il faut bien que tout cela ait une fin, mais en règle générale, on obtient de cette façon une borne très raisonnable pour $\max_i |b_i|$.

Dans le cas présent, une étape supplémentaire conduit à $y\leqslant 4$, suite à quoi la réduction bloque à $|y|\leqslant 3$.

On conclut enfin par énumération exhaustive que

Théorème 6. — *Les seuls couples* (x,y) *tels que* $3^x-2^y=1$ *sont* $(1,1),(2,3)$.

3.4. Le cas général. — De la même façon que précédemment, dans le cas général, on obtient l'inégalité

$$\left|\sum_{i=1}^{r} b_i\log p_i-\sum_{i=1}^{s} c_i\log q_i-\log a\right|\leqslant \frac{1}{\prod_{i=1}^{s} q_i^{c_i}}. \tag{6}$$

Avant de pouvoir appliquer la borne de Baker, il nous faut exprimer notre borne supérieure en termes non des seuls c_i, mais aussi des b_i ; en fait, les maxima des deux familles sont du même ordre de grandeur. En effet,

$$a+\prod_{i=1}^{s} q_i^{c_i}=\prod_{i=1}^{r} p_i^{b_i}, \tag{7}$$

et, donc

$$a+(\min_i q_i)^{\max_i c_i}\leqslant \left(\prod_{i=1}^{r} p_i\right)^{\max_i b_i},$$

ce qui permet de montrer (en échangeant les rôles de b_i et c_i) qu'il existe C_1 et C_2 ne dépendant que des p_i, q_i, a telles que

$$C_1 \max_i b_i \leqslant \max_i c_i \leqslant C_2 \max_i b_i.$$

Il s'ensuit, en particulier, que l'on peut trouver une constante C ne dépendant que des p_i, q_i, a, telle que

$$\left|\sum_{i=1}^{r} b_i \log p_i - \sum_{i=1}^{s} c_i \log q_i - \log a\right| \leqslant \exp(-C \max(|b_i|, |c_i|)).$$

Là encore, la borne de Baker fournit un réel C' tel que la quantité du membre de gauche soit minorée par

$$\exp(C' \log \max(|b_i|, |c_i|)).$$

In fine, on obtient donc une estimation de $\max(|b_i|, |c_i|)$, qui n'est souvent, hélas, pas utilisable telle quelle. C'est à ce stade que les arguments algorithmiques sont requis pour aller plus loin. Ils généralisent le rôle joué par les fractions continues pour l'équation $3^x - 2^y = 1$.

3.4.1. Réduction de la borne. — Nous sommes donc ramenés à une situation analogue au cas précédent, mais avec $r + s + 1 = n + 1$ logarithmes au lieu de 2 (ou à une situation inhomogène en $r + s = n$ logarithmes). Il nous faut donc remplacer les fractions continues par une « approximation diophantienne simultanée » en dimension $n + 1$. Historiquement, deux approches ont été utilisées à cette fin :

– Si l'on connaît une bonne approximation simultanée à n des $n+1$ termes, *i.e.*, $d(Qx_i, \mathbb{Z})$ petit pour tout $i \leqslant n$, alors on va pouvoir minorer Q fois la somme ; cette technique est connue sous le nom de lemme de Baker-Davenport ;

– Alternativement, la réduction des réseaux fournit une « presque plus petite » combinaison linéaire, et permet donc *a fortiori* de *minorer* les combinaisons linéaires des $\log p_i$ à coefficients bornés par B. C'est exactement ce dont on a besoin. Cette remarque est due à de Weger.

***Lemme 3* (Lemme de Baker-Davenport).** — *Soit $(b_i)_{1\leqslant i\leqslant n}$ des entiers, $(x_i)_{1\leqslant i\leqslant n+1}$ des réels, $B = \max_i |b_i|$. On suppose donné Q entier positif tel que $d(Qx_i, \mathbb{Z}) \leqslant \varepsilon$ pour tout $1 \leqslant i \leqslant n$. Alors $\left|\sum_{i=1}^{n+1} b_i x_i + x_{n+1}\right| \geqslant Q^{-1}(d(Qx_{n+1}, \mathbb{Z}) - nB\varepsilon)$.*

Démonstration. — On a

$$\begin{aligned}\Big|Q\Big(\sum_{i=1}^{n+1} b_i x_i + x_{n+1}\Big)\Big| &\geqslant d\Big(\sum_{i=1}^{n} Q b_i x_i + Q x_{n+1}, \mathbb{Z}\Big)\\ &\geqslant d(Qx_{n+1}, \mathbb{Z}) - \sum_{i=1}^{n} d(Qb_i x_i, \mathbb{Z})\\ &\geqslant d(Qx_{n+1}, \mathbb{Z}) - \max_{1\leqslant i\leqslant n} |b_i| \sum_{i=1}^{n} d(Qx_i, \mathbb{Z}),\\ &\geqslant d(Qx_{n+1}, \mathbb{Z}) - nB\varepsilon. \qquad \square\end{aligned}$$

Il suffit, dès lors, de choisir Q assez grand. Le principe des tiroirs suggère que l'on peut penser que $\varepsilon \approx Q^{-1/n}$, et *a priori*, si les x_i sont linéairement indépendants, on s'attend à ce que Qx_{n+1} soit « aléatoire » modulo 1. *A priori*, on va donc chercher Q légèrement plus grand que $(nB)^n$. Il nous reste encore, toutefois, à trouver l'entier Q, ce qui peut se faire par réduction d'un réseau bien choisi.

Notons, en particulier, que la nouvelle borne pour $\max_i |b_i|$ sera $O(\log Q/(d(Qx_{n+1}, \mathbb{Z}) - nB\varepsilon))$, dont l'ordre de grandeur attendu est $O(n \log B)$. Cela donne une idée de l'efficacité du processus de réduction : la nouvelle borne attendue est logarithmique en la précédente.

La minoration de $\sum b_i x_i + x_{n+1}$ sachant $\max_i |b_i| \leqslant B$ est toutefois une tâche qui peut, plus directement, être traitée par la réduction des réseaux.

Rappelons que l'algorithme LLL permet, étant donné des vecteurs de $\mathbb{R}^n$ linéairement indépendants sur $\mathbb{R}$, de trouver des combinaisons linéaires entières de ces vecteurs de petite norme euclidienne, et même peu supérieures à la plus petite norme euclidienne possible. En particulier, LLL permet de minorer la longueur de la plus petite combinaison linéaire (à coefficients bornés !) d'une famille de n vecteurs. Toutes les normes étant équivalentes, on va pouvoir en déduire une borne inférieure sur la longueur du vecteur le plus court pour la norme $\|\cdot\|_1$, qui est ce qui nous intéresse dans le cas présent.

Commençons dans ce cas par étudier le cas homogène.

Lemme 4. — *Soit* $(b_i)_{1\leqslant i\leqslant n}$ *des entiers,* $(x_i)_{1\leqslant i\leqslant n}$ *des réels,* $B = \max_i |b_i|$.

On suppose que le vecteur renvoyé par LLL sur le réseau engendré par les colonnes de la matrice

$$M(x_1,\ldots,x_n,C)=\begin{pmatrix} 1 & 0 & \ldots & 0 & 0\\ 0 & 1 & \ldots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \ldots & 1 & 0\\ \lfloor Cx_1\rceil & \lfloor Cx_2\rceil & \ldots & \lfloor Cx_{n-1}\rceil & \lfloor Cx_n\rceil \end{pmatrix}$$

a pour longueur l_0, avec

$$l_0 \geqslant 2^{(n-1)/2}B\sqrt{(n^2/4+n-1)}.$$

Alors pour tout n-uplet d'entiers $(b_1,\ldots,b_n)$, on a

$$\Big|\sum_{i=1}^n b_ix_i\Big| \geqslant \frac{1}{C}\bigg(\sqrt{2^{1-n}l_0^2-(n-1)B^2}-\frac{nB}{2}\bigg).$$

Démonstration. — Les propriétés générales d'une base LLL-réduite nous montrent que pour tout vecteur $\boldsymbol{b}=(b_1,\ldots,b_n)$, on a

$$\|A\boldsymbol{b}\|_2 \geqslant 2^{-(n-1)/2}l_0.$$

Il nous reste donc à majorer $\|A\boldsymbol{b}\|_2$.

On a

$$A\boldsymbol{b}=\begin{pmatrix} b_1\\ b_2\\ \vdots\\ b_{n-1}\\ \displaystyle\sum_{1\leqslant i\leqslant n} b_i\lfloor Cx_i\rceil \end{pmatrix},$$

d'où

$$\|A\boldsymbol{b}\|_2^2=\sum_{1\leqslant i\leqslant n-1} b_i^2+\Big(\sum_{1\leqslant i\leqslant n} b_i\lfloor Cx_i\rceil\Big)^2.$$

On majore alors b_i^2 par B^2, il vient :

$$\Big|\sum_{1\leqslant i\leqslant r} b_i\lfloor Cx_i\rceil\Big| \geqslant \sqrt{2^{1-n}l_0^2-(n-1)B^2}.$$

Mais

$$\Big|\sum_{1\leqslant i\leqslant n} b_i\lfloor Cx_i\rceil - C\sum_{1\leqslant i\leqslant n} b_ix_i\Big| \leqslant \sum_{1\leqslant i\leqslant n}|b_i|/2 \leqslant \frac{nB}{2},$$

d'où

$$\left|\sum_{1\leqslant i\leqslant n} b_i x_i\right| \geqslant \frac{1}{C}\left(\sqrt{2^{1-n}l_0^2-(n-1)B^2}-\frac{nB}{2}\right). \qquad \square$$

On peut aussi, par un argument un peu plus fin, contrôler la situation dans le cas inhomogène, où l'on chercher à minorer $\sum_{i=1}^n b_i x_i + x_{n+1}$.

Lemme 5 **([20]).** — *Soit* $\boldsymbol{x} = (x_i)$ *un vecteur de* $\mathbb{Z}^n$, *et* $A = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ *une base LLL-réduite d'un réseau* Λ. *Posons* $\boldsymbol{s} = (s_i) = A^{-1}(x_i)$. *Alors*

$$d(\boldsymbol{x}, \Lambda) \geqslant 2^{(1-n)/2} d(s_{i^*}, \mathbb{Z}) \|\boldsymbol{v}_1\|_2,$$

où i^* *est le plus grand entier* i *tel que* $s_i \notin \mathbb{Z}$.

Démonstration. — On considère la base $(\boldsymbol{v}_1^*, \ldots, \boldsymbol{v}_n^*)$, orthogonalisée de Gram-Schmidt de la base $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$. Les $\boldsymbol{v}_i$ sont donnés dans la base $(\boldsymbol{v}_i^*)$ par

$$\boldsymbol{v}_i = \boldsymbol{v}_i^* + \sum_{j=1}^{i-1} \mu_{ij} \boldsymbol{v}_j^*,$$

où $\mu_{ij} = (\boldsymbol{v}_i, \boldsymbol{v}_j^*)/\|\boldsymbol{v}_j^*\|_2^2$. On définira dans la suite $\mu_{ii} = 1$.

Soit $\boldsymbol{k} \in \mathbb{Z}^n$, et formons $\|A\boldsymbol{k} - \boldsymbol{x}\|_2 = \|A(\boldsymbol{k} - \boldsymbol{s})\|_2$:

$$\begin{aligned} A(\boldsymbol{k}-\boldsymbol{s}) &= \sum_{i=1}^n (k_i - s_i)\boldsymbol{v}_i, \\ &= \sum_{i=1}^n \sum_{j=1}^i (k_i - s_i)\mu_{ij}\boldsymbol{v}_j^*, \\ &= \sum_{j=1}^n \left(\sum_{i=j}^n \mu_{ij}(k_i - s_i)\right)\boldsymbol{v}_j^*. \end{aligned}$$

Par suite,

$$\|A(\boldsymbol{k}-\boldsymbol{s})\|_2^2 = \sum_{j=1}^n \left(\sum_{i=j}^n \mu_{ij}(k_i - s_i)\right)^2 \|\boldsymbol{v}_j^*\|_2^2.$$

Posons $i_1 := \max\{i : k_i \neq s_i\}$. Il est clair que $i_1 \geqslant i^*$. Si $i_1 > i^*$, on a $|k_{i_1} - s_{i_1}| \geqslant 1 \geqslant d(s_{i^*}, \mathbb{Z})$, et sinon $|k_{i_1} - s_{i_1}| \geqslant d(s_{i^*}, \mathbb{Z})$. Dans tous les cas, on a donc

$$\|A(\boldsymbol{k}-\boldsymbol{s})\|_2^2 \geqslant d(s_{i^*}, \mathbb{Z})^2 \|\boldsymbol{v}_{i_1}^*\|_2^2.$$

Le résultat découle alors directement d'une propriété classique des bases LLL-réduites. □

Pour minorer $|\sum_{i=1}^{n} b_i x_i + x_{n+1}|$, on applique cette proposition à la base donnée par les colonnes de la matrice $M(x_1, \dots, x_n, C)$ et au vecteur $(0, \dots, 0, \lfloor C x_{n+1} \rceil)$.

Ces deux propositions nous fournissent une solution complète pour minorer une forme linéaire de logarithmes, une borne sur les coefficients étant connue. Il nous reste simplement à discuter le choix de la constante C.

Heuristiquement, un réseau défini par une base $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ « sans vecteur court » (*i.e.*, dans le cas présent, correspondant à une équation diophantienne sans grande solution) a tous ses vecteurs minimaux successifs de norme du même ordre de grandeur et presque orthogonaux ; on s'attend donc à ce que les normes de ces minima soient de l'ordre de $\det(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)^{1/n}$. On peut donc s'attendre à ce que le vecteur le plus court soit de l'ordre de $C^{1/n}$, qui doit donc être de l'ordre de grandeur de $2^{(1-n)/2} nB$. Ceci montre qu'il faut choisir C de l'ordre de $(nB)^n$, estimation en pratique un peu optimiste et qu'il faudra augmenter un peu pour obtenir le résultat attendu.

Là encore, la nouvelle borne inférieure attendue est de l'ordre de $-\log C$, soit encore $n \log B$.

3.4.2. Cas d'échec de la réduction. — Il y a en pratique deux types de cas où la réduction est peu efficace. Pour comprendre le premier type de cas, il faut remarquer que l'esprit de la réduction, et même de toute la méthode, est de prouver qu'il n'y a pas de grande solution. Tout l'objectif est en effet d'améliorer des bornes sur la taille des solutions. En particulier, l'existence d'une grande solution limitera la qualité de la réduction de la borne, puisque la borne ne saurait descendre en deçà de cette solution. Dans ce cas (très rare, puisqu'en pratique, comme les b_i ne dépendent déjà que logarithmiquement des inconnues initiales, leur valeur est toujours très petite), on peut utiliser des bornes sur le deuxième vecteur le plus court pour estimer toutes les solutions distinctes de la première.

Le second type de situation est le cas où les x_i sont $\mathbb{Z}$-linéairement dépendants. Cette situation d'apparence surprenante se produit parfois. Dans ce cas, toutefois, il suffit de résoudre la dépendance, et d'éliminer une des variables pour réappliquer la méthode générale. Il n'est même pas nécessaire que la dépendance soit prouvée : supposons que l'on dispose d'une relation $|x_{n+1} + \sum_{i=1}^n \lambda_i x_i| \leqslant \tau$. Si l'on montre alors que

$$\left|\sum_{i=1}^{n+1} b_i x_i\right| = \left|\sum_{i=1}^{n} (b_i - \lambda_i b_{n+1}) x_i\right| \geqslant \theta,$$

on en déduit aussitôt que

$$\left|\sum_{i=1}^{n+1} b_i x_i\right| \geqslant \theta - B\tau,$$

ce qui répond au problème de départ si τ est assez petit. On peut ainsi éliminer les variables « numériquement $\mathbb{Z}$-dépendantes » les unes après les autres.

3.4.3. Retour au problème de départ. — La borne initiale, dans le cas présent, est réduite de façon très efficace à une taille raisonnable. On procède ensuite à une énumération exhaustive, qui peut être plus ou moins astucieuse ; essentiellement, si la borne finale est assez petite, on se contente d'énumérer les b_i et c_i en deçà de ladite borne.

3.5. Systèmes de deux équations de Pell. — Toujours dans l'esprit d'introduire les différents ingrédients indépendamment, nous allons étudier maintenant le cas où le traitement algorithmique est simple (fractions continues comme dans $3^x - 2^y = 1$, ou au pire lemme de Baker-Davenport en dimension 2), mais où il y a une part de traitement algébrique.

Le type d'équation traité ici remonte à l'article [**3**], qui date de 1969, et constitue la première application « pratique » de la borne de Baker à la résolution complète d'une équation diophantienne.

On étudie le système

$$\begin{cases} x^2 - ay^2 = 1 \\ x^2 - bz^2 = 1 \end{cases}$$

La partie algébrique du traitement est à peine plus sophistiquée que précédemment. On sait que les solutions d'une équation de Pell sont données par les puissances de l'unité fondamentale du corps quadratique $\mathbb{Q}(\sqrt{a})$, soit, plus précisément, si ε_a est l'unité de $\mathbb{Q}(\sqrt{a})$ et ε_b

celle de $\mathbb{Q}(\sqrt{b})$ (dans les deux cas, on choisit le conjugué plus grand que 1 en valeur absolue) :

$$\exists\, n, m \geqslant 0 \text{ t.q. } 2x = \varepsilon_a^n + \varepsilon_a^{-n} = \varepsilon_b^m + \varepsilon_b^{-m},$$

On a, en particulier,

$$|\varepsilon_a^n \varepsilon_b^{-m} - 1| \leqslant |\varepsilon_b|^{-m} \max(|\varepsilon_a|^{-n}, |\varepsilon_b|^{-m}).$$

Toujours de la même façon, l'identité $\varepsilon_a^n \approx \varepsilon_b^m$ montre que $m \asymp n$, et donc on peut encore réécrire notre inégalité comme :

$$(8) \qquad |n \log|\varepsilon_a| - m \log|\varepsilon_b|| \ll \exp(O(\max(m, n))).$$

À nouveau, la théorie de Baker garantit que

$$|n \log|\varepsilon_a| - m \log|\varepsilon_b|| \geqslant \exp(-C \log \max(m, n))$$

(éventuellement $(\log \max(m, n))^2$ dans le cas de formes en deux logarithmes), et la comparaison des deux bornes fournit à nouveau une borne supérieure sur $\max(m, n)$, et un processus de réduction du même type, puisque (8) montre encore que pour m, n assez grands, n/m est une réduite du développement en fractions continues de $\log|\varepsilon_a/\varepsilon_b|$.

Le problème historique de Baker-Davenport était légèrement différent : il s'agissait du système $3x^2 - y^2 = 2$, $8x^2 - z^2 = 7$. Dans ce cas, on trouve une identité du type

$$|\mu_1 \varepsilon_1^n - \mu_2 \varepsilon_2^m| = O(\exp(-C \max(m, n))),$$

ce qui conduit au problème inhomogène

$$|n \log \varepsilon_1 - m \log \varepsilon_2 + \log(\mu_1/\mu_2)| = O(\exp(-C \max(m, n))),$$

qui, d'une part ne permet pas l'utilisation de formes linéaires en deux logarithmes, d'autre part nécessite cette fois l'application du lemme de Baker-Davenport (d'où le nom) en dimension 2.

3.6. Un exemple récréatif : le problème du canonnier. — Un canonnier veut ranger de façon organisée ses boulets de canon. Il commence par les arranger à même le sol, sous la forme d'un carré, mais s'avise qu'ils occupent une place trop importante. Après réflexion, il les range alors sous la forme d'une pyramide à base carrée. Dans les deux cas, les figures (carré/pyramide) sont complètes. Combien le canonnier avait-il de boulets ?

Si l'on note n le côté du carré, et m le côté de la pyramide, on arrive aisément à l'équation $n^2 = m(m+1)(2m+1)/6$, soit encore $6n^2 = m(m+1)(2m+1)$.

Notons que m, $m+1$ et $2m+1$ sont premiers entre eux deux-à-deux. En particulier, ils n'ont aucun facteur premier en commun, et leurs décompositions en facteurs premiers offrent donc les possibilités suivantes (puisque $2m+1$ est impair) :

(1) $m = u^2$, $m+1 = 2v^2$, $2m+1 = 3w^2$;
(2) $m = u^2$, $m+1 = 6v^2$, $2m+1 = w^2$;
(3) $m = 2u^2$, $m+1 = 3v^2$, $2m+1 = w^2$;
(4) $m = 2u^2$, $m+1 = v^2$, $2m+1 = 3w^2$;
(5) $m = 3u^2$, $m+1 = 2v^2$, $2m+1 = w^2$;
(6) $m = 6u^2$, $m+1 = v^2$, $2m+1 = w^2$.

Une étude modulo 8 montre que seule la première et la dernière solution sont possibles. On est donc amené à résoudre les deux systèmes diophantiens $2v^2 - u^2 = 1, 3w^2 - 2u^2 = 1$ d'une part, et $v^2 - 6u^2 = 1, w^2 - 12u^2 = 1$ d'autre part. L'étude conduit à l'unique solution $(1,1,1)$, qui correspond à $m = 1$ (1 boulet de canon) dans le premier cas, aux deux solutions $(0,1,1)$ (qui correspond à $m = 0$, pas de boulet de canon) et $(2,5,7)$ ($m = 24$, 4900 boulets de canon) dans le second.

4. L'équation de Thue

Il s'agit du point culminant de ce cours, même si quelques prolongements et questions plus générales seront indiquées dans une dernière partie. Cette équation généralise l'équation de Pell dans sa forme, et combine les différents ingrédients (traitement algébrique, construction de forme linéaire, réduction de borne) dans le cas général.

On appelle équation de Thue une équation de la forme

$$P(x,y) = a, \tag{9}$$

avec P homogène, irréductible, de degré $\geqslant 3$. Il est bon de noter que le cas du degré 2 est le cas de l'équation de Pell, que le cas du degré 1 est une recherche de coefficients de Bezout ; enfin, que si le polynôme $P = P_1P_2$ est réductible, on trouve un nombre fini de systèmes $P_1(x,y) = a_1, P_2(x,y) = a_2$, qui se résolvent de façon banale par élimination, par exemple par un calcul de résultant.

Ces remarques étant faites, consacrons-nous à l'équation de Thue. Dans un souci de simplification, on supposera P unitaire en X, ce que l'on peut toujours faire en pratique, quitte à effectuer un changement de variable (bien choisi de façon à conserver le caractère entier des racines ; on ajoutera souvent ce faisant des solutions parasites qui sont éliminées par vérification *a posteriori*).

D'un point de vue algébrique, si l'on prend une racine α du polynôme P, et que l'on écrit $\alpha_1, \dots, \alpha_n$ les différents conjugués de α, on peut réécrire l'équation sous la forme

$$\prod_{i=1}^{n}(X - Y\alpha_i) = a \tag{10}$$

soit encore $N_{\mathbb{Q}(\alpha)/\mathbb{Q}}(X - \alpha Y) = a$.

Dans la suite, on notera σ_i le plongement de $\mathbb{Q}(\alpha)$ dans $\mathbb{C}$ qui envoie α sur α_i. On note en outre s le nombre de α_i réels, $2t$ le nombre de $\alpha_i \in \mathbb{C} - \mathbb{R}$, et on suppose les α_i ordonnés en commençant par les réels.

4.1. Équation aux unités. — Comme dans l'équation de Pell, on construit alors E, un ensemble de solutions non-associées modulo l'action du groupe des unités de cette équation aux normes. Notons μ un élément de E. La suite de l'algorithme devra être répétée une fois pour chaque élément $\mu \in E$.

On va chercher une solution sous la forme $X - \alpha Y = \mu\varepsilon$, où ε est une unité du corps $\mathbb{Q}(\alpha)$. Le groupe des unités étant de type fini et de rang r, on peut chercher ε sous la forme $\varepsilon = \prod \eta_i^{b_i}$, où η_i est un système d'unités fondamentales de $\mathbb{Q}(\alpha)$. Nous avons donc nos inconnues exponentielles.

4.2. L'inégalité fondamentale. — Il nous reste, pour pouvoir appliquer la machinerie générale, à construire une quantité voisine de 1, de façon à obtenir une forme en logarithmes voisine de 0. Pour ce faire, il nous faut contrôler les ordres de grandeur, ce qui va se faire *via* une observation importante sur l'équation initiale. En substance, l'équation sous la forme (10) montre qu'on a un produit de n termes de la forme $X - \alpha Y$ de taille bornée. Comme au plus un de ces termes peut être petit (car, alors, $X \approx \alpha_k Y$ et pour $j \neq k$, $X - \alpha_j Y \approx Y(\alpha_k - \alpha_j)$ est de l'ordre de grandeur de Y), il doit être très petit.

Théorème 7. — *On pose* $f(X) = P(X, 1)$. *Soit*

$$Y_0 = \begin{cases} \Big(\dfrac{2^{n-1}|a|}{\min_{1\leqslant i\leqslant t}|f'(\alpha_{s+i})| \cdot \min_{1\leqslant i\leqslant t}|\operatorname{Im}\alpha_{s+i}|}\Big)^{1/n} & \textit{si } t \geqslant 1, \\ 1 & \textit{si } t = 0, \end{cases}$$

$$c_1 = \frac{2^{n-1}|a|}{\min_{1\leqslant i\leqslant s}|f'(\alpha_i)|}.$$

Soit (x, y) *une solution entière de* (9). *Si* $|y| > Y_0$ *alors, pour un* $i_0 \in \{1, \ldots, s\}$ *on a*

$$|x - \alpha_{i_0} y| \leqslant \frac{c_1}{|y|^{n-1}}. \tag{11}$$

Démonstration. — On définit i_0 par $|x - \alpha_{i_0} y| = \min_i |x - \alpha_i y|$. On a alors

$$\prod_i |x - \alpha_i y| = |a| \tag{12}$$

par ailleurs $|x - \alpha_i y| \geqslant |y||\alpha_{i_0} - \alpha_i| - |x - \alpha_{i_0} y|$, soit encore, par définition de i_0,

$$|x - \alpha_i y| \geqslant \frac{|y|}{2}|\alpha_{i_0} - \alpha_i|.$$

On obtient le résultat en reportant cette minoration dans (10). Supposons alors que pour un certain x, on a $i_0 > s$. Il vient,

$$\frac{c_1}{|y|^{n-1}} \geqslant |x - \alpha_{i_0} y| \geqslant |y||\operatorname{Im}\alpha_{i_0}|,$$

c'est-à-dire que $|y| \leqslant Y_0$. □

Ce théorème est une mine d'informations. Tout d'abord, il indique que si P n'a que des racines imaginaires, l'équation se résout très rapidement par énumération exhaustive.

Ensuite, il montre que dans le cas contraire, hormis pour quelques petites valeurs de y, pour toute solution (x, y), x/y est une réduite du développement en fractions continues d'une des racines réelles de l'équation. Cette remarque est souvent cruciale pour l'énumération finale des petites solutions.

Enfin, c'est ce résultat qui a permis à Thue de prouver, en 1909, la finitude du nombre de solutions à l'équation qui porte maintenant son nom, comme corollaire du résultat d'approximation des algébriques par des rationnels mentionné plus haut.

Mais ce qui nous intéresse surtout, c'est qu'il nous fournit le résultat d'approximation qui va nous permettre de construire la quantité souhaitée. À ce point, il faut ajouter une nouvelle combinatoire à celle des différentes solutions des équations aux normes, *i.e.*, il faut faire ce qui suit pour toutes les racines réelles de l'équation de départ, de façon à énumérer tous les choix possibles pour i_0. Dans la suite, on supposera sans perte de généralité que $\alpha = \alpha_1$ est la racine telle que $X - \alpha Y$ est petit.

4.3. Une quantité voisine de 1. — Il nous devient aisé de construire une quantité de type exponentielle qui soit voisine de 1 : dans la mesure où $X - \alpha_j Y$ est très proche de $(\alpha - \alpha_j)Y$, il vient que $(X - \alpha_j Y)/(\alpha - \alpha_j)$ est très proche de Y, et donc que

$$\frac{X - \alpha_j}{X - \alpha_k} \cdot \frac{\alpha - \alpha_k}{\alpha - \alpha_j}$$

est très proche de 1. Plus précisément, on a alors :

Proposition 3. — *On pose*

$$c_2 = \min_{1 \leqslant i < j \leqslant n} |\alpha_i - \alpha_j|\,,\ c_3 = 1.39 c_1 c_2^{-1},\ X_1 = \max\left(X_0, \left(2c_1 c_2^{-1}\right)^{1/n}\right).$$

Alors si $j, k \neq i_0$, *on a*

$$\left|\operatorname{Log} \frac{X - \alpha_j Y}{X - \alpha_k Y} \cdot \frac{\alpha - \alpha_k}{\alpha - \alpha_j}\right| \leqslant \frac{2c_3}{|Y|^n}.$$

Enfin,

$$\left|\operatorname{Log} \frac{(X - \alpha_{i_0} Y) f'(\alpha_{i_0})}{|a| Y^{n-1}}\right| \leqslant \frac{(n-1)c_3}{|Y|^n}.$$

Démonstration. — Il suffit d'estimer $|\log\{(X - \alpha_j Y)/(\alpha Y - \alpha_j Y)\}|$; On écrit tout simplement que

$$\frac{X/Y - \alpha_i}{\alpha_{i_0} - \alpha_i} - 1 = \frac{X/Y - \alpha_{i_0}}{\alpha_{i_0} - \alpha_i},$$

et on utilise le lemme fondamental et la définition de c_2. On obtient alors

$$\left|\frac{X - \alpha_i Y}{Y(\alpha_{i_0} - \alpha_i)} - 1\right| \leqslant \frac{c_1}{c_2 |Y|^n}$$

Il suffit alors de remarquer que si $|z-1| \leqslant 1/2$, on a alors $|\operatorname{Log} z| \leqslant 1.39 \cdot |z-1|$, où Log est la détermination principale du logarithme complexe ; noter que le choix de X_1 garantit que le majorant est plus petit que 1/2.

Les résultats s'ensuivent en combinant l'inégalité obtenue pour j et pour k (premier point) ou en ajoutant toutes les inégalités (dernier point). □

Prenons un système $\eta_1, \ldots, \eta_r$ d'unités fondamentales du corps $\mathbb{Q}(\alpha)$. On peut alors, par l'étude algébrique, écrire $X - \alpha Y = \mu\eta_1^{b_1} \ldots \eta_r^{b_r}$. Il vient

$$\left|\sum b_i \operatorname{Log} \frac{\sigma_j(\eta_i)}{\sigma_k(\eta_i)} + \operatorname{Log} \frac{\sigma_j(\mu)(\alpha-\alpha_k)}{\sigma_k(\mu)(\alpha-\alpha_j)}\right| \leqslant \frac{2c_3}{|Y|^n}.$$

Il nous reste, pour mettre en œuvre complètement le programme général, à montrer que la borne supérieure se réécrit comme $\exp(O(\max_i |b_i|))$, soit encore que $\max_i |b_i|$ et $\log Y$ sont du même ordre de grandeur. Pour ce faire, on remarque que

$$|\log Y + \log|\alpha-\alpha_i| - \log|\sigma_i(\mu)| + \sum_{j=1}^{r} b_j \log|\sigma_i(\eta_j)|| \leqslant \frac{c_3}{|Y|^n} \leqslant 1/2$$

pour $|Y|$ assez grand, soit un système de $r-1$ équations indépendantes en les r inconnues b_i.

On peut en fait compléter ce système par une équation analogue pour $i = i_0$, obtenue grâce au dernier point de la proposition 3. En résolvant ce système, on voit que $\max_i |b_i|$ dépend linéairement de $\log Y$, et donc que l'on a bien

$$\left|\sum b_i \operatorname{Log} \frac{\sigma_j(\eta_i)}{\sigma_k(\eta_i)} + \operatorname{Log} \frac{\sigma_j(\mu)(\alpha-\alpha_k)}{\sigma_k(\mu)(\alpha-\alpha_j)}\right| \leqslant \exp(-C \max_i |b_i|),$$

pour une constante C explicitement déterminée par l'équation.

On rentre alors dans le cadre général développé précédemment ; la borne de Baker nous fournit une borne (immense) pour $\max_i |b_i|$, et la méthode de réduction permet de réduire ladite borne, conduisant à une borne très raisonnable. Dans une dernière phase, on peut soit choisir (r petit) d'énumérer les r-uplets $(b_1, \ldots, b_r)$, soit traduire la borne en borne sur X, Y et utiliser la caractérisation de x/y en terme de réduite du développement en fraction continue de α_{i_0}, pour y pas trop petit.

4.4. Un exemple. — Cette méthode générale (en fait, une version dont l'algorithmique a été assez largement optimisée pour fonctionner dans le cas de grands degrés) est implantée dans `GP` (commandes `thueinit`, `thue`, la première effectuant le travail préparatoire indépendant du second membre a, la seconde résolvant étant donné une valeur particulière de a et le résultat de `thueinit`).

Pour l'équation $x^4 - 2y^4 = 1$, on trouve que dans tous les cas (un élément dans E, deux racines réelles, donc deux cas à traiter), on a $\max_i |b_i| \leqslant 5.1 \cdot 10^{26}$. Après réduction (légèrement différente de ce qui est présenté ici), on trouve successivement $\max_i |b_i| \leqslant 34, 6, 5$. On finit par en déduire que les seules solutions sont $(1, 0)$ et $(-1, 0)$.

5. Conclusion

Diverses généralisations pourraient à ce point être obtenues à peu de frais algorithmique, mais à un coût algébrique assez lourd.

5.1. Méthode des logarithmes elliptiques. — Je pense à la méthode des logarithmes elliptiques pour résoudre, en particulier, les équations du type $Y^2 = X^3 + aX + b$ (avec $4a^3 + 27b^2 \neq 0$) ; dans ce cas, le groupe sous-jacent est le groupe des points rationnels de la courbe elliptique correspondante, il est bien de type fini (donc tout point s'écrit $P_0 + \sum n_i P_i$, avec P_0 dans un ensemble fini et les P_i fixés), et on a une fonction « logarithme » (le logarithme elliptique) ayant de plus les bonnes propriétés. La même démarche générale, moyennant l'utilisation d'une généralisation de la borne de Baker à ce contexte, permet de résoudre complètement le problème.

Il convient de noter qu'en l'état, la procédure de détermination des générateurs du groupe sous-jacent (les P_i) n'est pas à proprement parler un algorithme, même si dans la plupart des cas raisonnables, elle termine en renvoyant un résultat prouvé. Pour cette méthode, je renvoie par exemple à [**16**].

5.2. Équations superelliptiques. — On pourrait aussi discuter de l'approche de Bilu pour les équations superelliptiques $y^p = f(x)$ [**5**], raffinée par Bilu et l'orateur [**6**]. Cependant, là encore, cela impose un traitement algébrique à la fois soigneux et conséquent, qui dépasse le cadre – assez informel – de ce cours.

5.3. Développements récents du sujet. — On peut considérer, de façon récente, l'algorithmique de ces différents types d'équations diophantiennes comme bien comprises. Le point limitant est en effet devenu le calcul des générateurs du groupe sous-jacent, à côté duquel la résolution proprement dite est souvent très peu coûteuse. La tendance moderne est donc à s'intéresser à deux questions générales :

– les équations diophantiennes pour lesquelles les inconnues vivent dans l'anneau des entiers d'un corps de nombres ;

– les familles d'équations diophantiennes à un ou plusieurs paramètres.

Dans les deux cas, les méthodes et idées de départ sont les mêmes qui sont exposées ici, mais le traitement complet requiert des ingrédients plus subtils.

5.4. Diviseurs primitifs des suites de Lucas. — Pour conclure, on citera une belle application [**7**] provenant de la résolution complète de la famille d'équations de Thue correspondant aux sous-corps réels maximaux des corps cyclotomiques :

Soit α, β deux entiers algébriques tels que $\alpha+\beta$ et $\alpha\beta$ soient dans $\mathbb{Z}-\{0\}$, et α/β n'est pas une racine de l'unité. On pose $U_n(\alpha,\beta) = (\alpha^n-\beta^n)/(\alpha-\beta)$.

On dit qu'un premier p est un diviseur primitif de $U_n(\alpha,\beta)$ si p divise $U_n(\alpha,\beta)$, mais p ne divise pas $(\alpha-\beta)^2\prod_{1\leqslant i\leqslant n-1} U_i(\alpha,\beta)$. Alors on a :

Théorème 8. — *Pour tout $n>30$, $U_n(\alpha,\beta)$ a un diviseur primitif.*

Ce résultat est optimal, comme le montre le cas $(\alpha,\beta)=(1\pm i\sqrt{7})/2$.

Références

[1] A. Baker – « Linear forms in the logarithms of algebraic numbers I », *Mathematika* **13** (1966), p. 204–216 ; II, *ibid.* **14** (1967), p. 102–107 ; III, *ibid.* **14** (1967), p. 220–228 ; IV, *ibid.* **15** (1968), p. 204–216.

[2] ———, « Contributions to the theory of Diophantine equations. I. On the representation of integers by binary forms », *Philos. Trans. Roy. Soc. London Ser. A* **263** (1968), p. 173–191 ; II. The Diophantine equation $y^2=x^3+k$, *ibid.* **263** (1968), p. 193–208.

[3] A. Baker & H. Davenport – « The equations $3x^2-2=y^2$ and $8x^2-7=z^2$ », *Quart. J. Math. Oxford (2)* **20** (1969), p. 129–137.

[4] A. Baker & G. Wüstholz – « Logarithmic forms and group varieties », *J. Reine Angew. Math.* **442** (1993), p. 19–62.
[5] Yu. Bilu – « Solving Superelliptic Diophantine Equations by Baker's method », prépublication, 1994.
[6] Yu. Bilu & G. Hanrot – « Solving Superelliptic Diophantine Equations by Baker's method », *Compositio Math.* **112** (1998), p. 273–312.
[7] Yu. Bilu, G. Hanrot & P. Voutier – « Existence of Primitive Divisors of Lucas and Lehmer Sequences (with an appendix by M. Mignotte) », *J. Reine Angew. Math.* **539** (2001), p. 75–122.
[8] F. Dyson – « The approximation to algebraic numbers by rationals », *Acta Math.* **79** (1947), p. 225–240.
[9] A.O. Gelfond – *Transcendental and Algebraic Numbers*, Dover Publications, New York, 1960, traduction anglaise.
[10] M. Laurent, M. Mignotte & Y. Nesterenko – « Formes linéaires en deux logarithmes et déterminants d'interpolation », *J. Number Th.* **65** (1995), p. 285–321.
[11] A.K. Lenstra, H.W. Lenstra, Jr. & L. Lovász – « Factoring polynomials with rational coefficients », *Math. Ann.* **261** (1982), p. 515–534.
[12] Yu. Matijasevič – « Enumerable sets are diophantine », *Dokl. Akad. Nauk. SSSR* **191** (1970), en russe ; version anglaise complétée : *Soviet Math. Doklady*, **12** (1971), p. 249-54.
[13] K.F. Roth – « Rational approximations to algebraic numbers », *Mathematika* **2** (1955), p. 1–20.
[14] C.L. Siegel – « Approximation algebraischer Zahlen », *Math. Zeit.* **10** (1921), p. 173–213.
[15] N.P. Smart – *The Algorithmic Resolution of Diophantine Equations : A Computational Cookbook*, London Mathematical Society Student Texts, vol. 41, Cambridge University Press.
[16] R. Stroeker & N. Tzanakis – « Solving elliptic Diophantine equations by estimating linear forms in elliptic logarithms », *Acta Arith.* **67** (1994), p. 177–196.
[17] A. Thue – « Über Annäherungswerte algebraischer Zahlen », *J. Reine Angew. Math.* **135** (1909), p. 284–305.
[18] N. Tzanakis & B.M.M. de Weger – « On the Practical Solution of the Thue Equation », *J. Number Th.* **31** (1989), p. 99–132.
[19] I. Vardi – « Archimedes' Cattle Problem », *Amer. Math. Monthly* **105** (1998), p. 305–319.
[20] B.M.M. de Weger – « Solving exponential diophantine equations using lattice basis reduction algorithms », *J. Number Th.* **26** (1987), p. 325–367.

G. Hanrot, Projet spaces, INRIA Lorraine, 615, rue du Jardin Botanique, F-54602 Villers-les-Nancy Cedex • *E-mail :* `hanrot@loria.fr`