Factorisation des entiers à l'aide des courbes elliptiques.

Florian CARO et Alexandre POPIER

Table des matières

Introduction

L'objet de ce travail est d'étudier un algorithme, créé par H. Lenstra de factorisation des entiers qui utilise des courbes elliptiques. On va donc commencer par définir la notion de courbe elliptique sur un corps et sur un type d'anneaux particuliers, à savoir $\mathbb{Z}/n\mathbb{Z}$. Ensuite on va décrire l'algorithme en lui-même qui repose sur les propriétés de groupe des points d'une courbe elliptique.

I Définition des courbes elliptiques

On va commmencer par définir une courbe elliptique sur un corps, donner un exemple dans le cas des nombres réels et ensuite étendre cette définition au cas de $\mathbb{Z}/n\mathbb{Z}$.

I.1 Courbe elliptique sur un corps

Soit K un corps, de caractéristique différente de 2 et de 3.

Définition. 1 *On appelle courbe elliptique sur K un ensemble*

$$E_{a,b}(K) = \{(x : y : z) \in \mathbb{P}^2(K) \mid y^2z = x^3 + axz^2 + bz^3\}$$

avec a,b dans K tels que $6(4a^3 + 27b^2) \neq 0$. $\mathbb{P}^2(K)$ *est le plan projectif sur K. C'est l'ensemble des classes d'équivalence des triplets* $(x, y, z) \in K^3 \setminus \{(0, 0, 0)\}$*, deux triplets (x,y,z) et (x',y',z') étant equivalents s'il existe* $\lambda \in K^*$ *tel que (x',y',z') = (λx,λy,λz). La classe d'équivalence de (x,y,z) sera notée (x :y :z).*

L'équation $y^2 = x^3 + ax + b$ *est appelée équation de Weierstrass de la courbe E(K).*

Soit $E_{a,b}$ une courbe elliptique sur K. E(K) contient exactement un point $(x : y : z) \in \mathbb{P}^2(K)$ tel que z = 0. C'est le point (0 :1 :0) ; ce point est appelé *l'origine* de la courbe elliptique et est noté O. Les autres points de E(K) sont les points (x :y :1) qui vérifient l'équation de Weierstrass de la courbe : $y^2 = x^3 + ax + b$. L'ensemble E(K) est muni d'une structure de groupe abélien, définie de la manière suivante :

On pose O + P = P + O = P pour tout P ∈ E(K).

Soient P = (x_1 :y_1 :1) et Q = (x_2 :y_2 :1) deux points différents de l'origine. Alors P + Q = O si et seulement si $x_1 = x_2$ et $y_1 = -y_2$.

Autrement on pose $\lambda = (y_1 - y_2)/(x_1 - x_2)$ si $x_1 \neq x_2$ et $\lambda = (3x_1^2 + a)/(y_1 + y_2)$ si $x_1 = x_2$. Soit $\gamma = y_1 - \lambda x_1$. Alors P + Q = R, où R = (x_3 :y_3 :1) avec $x_3 = \lambda^2 - x_1 - x_2$ et $y_3 = -\lambda x_3 - \gamma$. Remarquons que O est l'élément neutre du groupe, et que -(x :y :z) = (x :-y :z).

Géométriquement étant donnés deux points $P_1 = (x_1, y_1)$ et $P_2 = (x_2, y_2)$, la droite D passant par P_1 et P_2 (la tangente si P_1 = P_2) recoupe E en un troisième point (x_3,-y_3), et si on pose $P_3 = (x_3, y_3) = P_1 + P_2$, on obtient la structure de groupe abélien précédente.

I.2 Exemple graphique sur $\mathbb{R}$

On va s'attacher ici à décrire ce qui se passe dans le cas des réels. L'équation de Weierstrass s'écrit : $y^2 = x^3 + ax + b$. Soit $Q(x) = x^3 + ax + b$. Le discriminant Δ vaut $-(4a^3 + 27b^2)$. Plusieurs cas peuvent se produire.

(1) $\Delta < 0$. Alors Q a une seule racine rélle et le graphe de la courbe a une seule composante connexe. Sous Maple cela correspont aux exemples suivants : a = b = 2 et a = -1.5, b = 1.

(2) $\Delta > 0$. Alors Q a trois racines réelles distinctes et le graphe a deux composantes connexes : une non-compacte, qui est la composante du point O de la courbe (i.e. le point à l'infini), et une compacte. Elle correspond dans nos exemples à a = -4 et b = 2.

(3) $\Delta = 0$. Ce n'est plus une courbe elliptique. En effet Q a alors une racine double, i.e. Q est de la forme $(x - c)^2(x - d)$ avec 2c + d = 0. Si c = d = 0 (ou a = b = 0), la courbe résente un point de rebroussement en zéro. D'o/'u un problème pour calculer certaines sommes (géométriquement on voit qu'en zéro on a un problème). Si $c > d$, la courbe résente un point singulier en (c,0). Si $c < d$, la courbe a un point double en (d,0) et les tangentes en ce point sont différentes de par et d'autre de l'axe des abcisses). Ainsi on voit bien la nécessité de prendre $\Delta \neq 0$.

cf. Annexe 1.

I.3 Courbe elliptique sur $\mathbb{Z}/n\mathbb{Z}$

Pour notre travail on a besoin de définir la notion de courbe elliptique sur un anneau, à savoir $\mathbb{Z}/n\mathbb{Z}$. Tout d'abord le plan projectif $\mathbb{P}^2(\mathbb{Z}$/n $\mathbb{Z})$ peut être défini comme l'ensemble des classes d'équivalence des triplets (x,y,z) premiers entre eux dans leur ensemble pour la relation d'homothétie par un élément de $(\mathbb{Z}/n\mathbb{Z})$*.

Définition. 2 *On appelle courbe elliptique sur $\mathbb{Z}/n\mathbb{Z}$ un ensemble*

$$E(\mathbb{Z}/n\mathbb{Z}) = \{(x : y : z) \in \mathbb{P}^2(\mathbb{Z}/n\mathbb{Z}) \mid y^2z = x^3 + axz^2 + bz^3\}$$

où a,b sont dans $\mathbb{Z}/n\mathbb{Z}$ tels que $6(4a^3 + 27b^2)$ soit inversible dans $\mathbb{Z}/n\mathbb{Z}$.

On note le point (0 :1 :0) de $\mathbb{Z}/n\mathbb{Z}$ par O. On pose $V_n = \{(x : y : 1) \mid x, y \in \mathbb{Z}/n\mathbb{Z}\} \cup O$. V_n est un sous ensemble de $\mathbb{P}^2(\mathbb{Z}/n\mathbb{Z})$. On note $E_{aff}(\mathbb{Z}/n\mathbb{Z})$ l'ensemble $V_n \cap E(\mathbb{Z}/n\mathbb{Z})$. Dans le cas où n est premier, $E_{aff}(\mathbb{Z}/n\mathbb{Z})$ = $E(\mathbb{Z}/n\mathbb{Z})$. Par contre ceci est faux si n n'est pas premier.

L'ensemble $E(\mathbb{Z}/n\mathbb{Z})$ est muni d'une loi de groupe naturelle qui généralise la loi vue plus haut dans le cas n premier. En fait pour la factorisation de n on n'a pas besoin de connaitre cette loi mais il est bon de savoir qu'elle existe.

Par exemple, pour n = 5, a = 1, et b = -1, les points $P_1 = (1, 1)$ et $P_2 = (2, 2)$ sont des points de la courbe elliptique. Pour calculer les coordonnées (x_3, y_3) du point $P_3 = P_1 + P_2$, on applique les formules précédentes, mais avec les règles de la congruence, on obtient $x_3 = 3$ et $y_3 = 2$. L'ensemble $E(\mathbb{Z}/n\mathbb{Z})$ des points des courbes elliptiques modulo n n'a qu'un nombre fini d'éléments, compris entre $n - 2\sqrt{n} + 1$ et $n + 2\sqrt{n} + 1$. En essayant tous les couples de nombres possibles, on trouve que la courbe elliptique $E(\mathbb{Z}/5\mathbb{Z})$ pour a = 1 et b = 4 est constitué de neuf éléments (3,3) (3,2) (0,3) (0,2) (1,4) (1,1) (2,3) (2,2) et O.

II L'algorithme de factorisation

On va faire des calculs sur $E(\mathbb{Z}/n\mathbb{Z})$ en considérant $\mathbb{Z}/n\mathbb{Z}$ comme un corps, en espérant une erreur de calcul au moment de calculer certains inverses.

II.1 L'addition

On décrit un algorithme qui étant donné $n \in \mathbb{N}$, $a \in \mathbb{Z}/n\mathbb{Z}$, et $P, Q \in E_{aff}(\mathbb{Z}/n\mathbb{Z})$, soit calcule un diviseur non trivial de n, soit calcule P + Q = R.

Si P = O (respectivement Q = O), on prend R = Q (resp. R = P).

Sinon $P \neq O$ et $Q \neq O$, en notant $P = (x_1 : y_1 : 1)$ et $Q = (x_2 : y_2 : 1)$, on calcule pgcd($x_1 - x_2$, n), à l'aide de l'algorithme d'Euclide. Si ce pgcd est différent de 1 et de n, on l'appelle d qui est un diviseur non trivial de n. S'il est égal à 1, l'algorithme d'Euclide nous donne aussi $(x_1 - x_2)^{-1}$, et alors en posant

$$\lambda = (y_1 - y_2)/(x_1 - x_2)$$

$$x_3 = \lambda^2 - x_1 - x_2$$

$$y_3 = \lambda(x_1 - x_3) - y_1$$

on obtient R = P + Q =(x_3 :y_3 :1).

Enfin si pgcd($x_1 - x_2$, n) = n, alors $x_1 = x_2$ et on calcule pgcd($y_1 + y_2$, n). S'il est différent de 1 ou de n, on l'appelle d qui est un diviseur non trivial de n. S'il est égal à n, alors $y_1 = -y_2$ et R = P + Q = O. Enfin s'il est égal à 1, l'algorithme d'Euclide fournit aussi $(y_1 + y_2)^{-1}$. et alors en posant

$$\lambda = (3x_1^2 + a)/(y_1 + y_2)$$

$$x_3 = \lambda^2 - x_1 - x_2$$
$$y_3 = \lambda(x_1 - x_3) - y_1$$

on obtient R = P + Q =(x_3 :y_3 :1).

II.2 La méthode de factorisation

On suppose que l'on veut trouver un diviseur de n. On tire au hasard trois entiers a,x,y compris entre 0 et n-1. On pose $b = y^2 - x^3 - ax$ modulo n. On calcule le pgcd de $6(4a^3 + 27b^2)$ et de n. S'il est égal à n, on recommence. S'il est différent de 1 et de n, on a un diviseur non trivial de n. S'il est égal à 1,

$$E(\mathbb{Z}/n\mathbb{Z}) = \{(x : y : z) \in \mathbb{P}^2(\mathbb{Z}/n\mathbb{Z}) \mid y^2z = x^3 + axz^2 + bz^3\}$$

est bien une courbe elliptique et $P = (x : y : 1) \in E_{aff}(\mathbb{Z}/n\mathbb{Z})$. On calcule kP = P + P + ... + P k fois, k étant un entier dépendant de n (voir plus loin), grâce à l'addition précédente. D'après le théorème de Lagrange, quand on additionne un élément d'un groupe fini à lui-même autant de fois qu'il y a d'élément dans le groupe, on obtient l'élément neutre. Dans l'exemple de la définition 2, neuf fois n'importe quel élément de $E(\mathbb{Z}/5\mathbb{Z})$ est égal à O.

Ou on obtient un diviseur de n et l'algorithme s'arrête, ou bien on réussit à calculer jusqu'au bout kP. On réessait alors avec d'autres valeurs de a,x,y (autre courbe) et ceci h fois (nombre de courbes que l'on décide de prendre). Au bout de h courbes (ou avant), on espère avoir trouvé un diviseur de n. Pour cela se pose le problème du choix des paramètres k et h.

III Choix des paramètres et résultats

III.1 Choix de k : nombre de fois où on réitère la somme

On commence par se fixer un paramètre v qui est de l'ordre d'un majorant des facteurs premiers de n. On peut prendre la partie entière de $\sqrt{n}$. Ensuite on fixe w qui est le temps passé sur une courbe. w est pris égal à $L(v)^{\frac{1}{\sqrt{2}}}$ avec :

$$L(x) = e^{\sqrt{ln(x)\ ln(ln(x))}}$$

En effet si u = card $\{s \in \mathbb{Z} || s - (p + 1)| < \sqrt{p}, et\ tout\ nombre\ premier\ divisant\ s\ est \leq w\}$,

$$f(w) = \frac{u}{2\sqrt{p} + 1}$$

est la probabilité qu'un entier choisi au hasard dans l'intervalle (p + 1 - $\sqrt{p}$, p + 1 + $\sqrt{p}$) ait tous ses facteurs premiers inférieurs ou égaux à w. Alors pour tout $h \in \mathbb{N}, > 1$, il existe une constante calculable c > 1 telle que la probabilité de succès de l'algorithme est au moins

$$1 - c^{-h\ \frac{f(w)}{ln(v)}}$$

Ainsi pour avoir une chance raisonnable de succès on doit choisir h de même ordre de grandeur que ln(v)/f(w) ; h devant être minimal on doit choisir w tel que w/f(w) soit minimal.

Un théorème de Canfield, Erdös et Pomerance dit que si α est un réel positif, la probabilité qu'un entier positif $s \leq x$ choisi au hasard ait tous ses facteurs premiers inférieurs ou égaux à $L(x)^{\alpha}$ est $L(x)^{-1/2\alpha + o(1)}$. Une conjecture non prouvée dit que le résultat est encore vrai si s est un entier pris au hasard dans l'intervalle (x + 1 - $\sqrt{x}$, x + 1 - $\sqrt{x}$). Ainsi $f(L(x)^{\alpha}) = L(x)^{-1/2\alpha + o(1)}$. Donc pour x = v, et w = $L(v)^{\alpha}$, on a :

$$w/f(w) = L(v)^{1/2\alpha + \alpha + o(1)}$$

ce qui est minimal pour $\alpha = 1/\sqrt{2}$.

A partir de là on choisit k de la manière suivante : pour tout entier $r \geq 2$ on note e(r) le plus grand entier m tel que $r^m \leq v + 2\sqrt{v} + 1$. On prend alors :

$$k = \prod_{r=2}^{w} r^{e(r)}$$

Ce choix résulte d'une simple conjecture mais il semble que ce soit la meilleure estimation possible.

III.2 Choix de h

D'après le paragraphe précédent h doit être de l'ordre de grandeur de ln(v)/f(w) et comme f(w) est du même ordre de grandeur que $L(v)^{-1/\sqrt{2}}$, on choisit h égal à $ln(v)L(v)^{1/\sqrt{2}}$.

III.3 Résultats

Le programme étant écrit en turbo pascal, on a été limité par la taille des nombres à factoriser ; à savoir des entiers inférieurs à 2147483647 (il faut aussi tenir compte du choix de v : s'il est trop grand on ne peut calculer w). Par contre en testant le programme, on voit que les nombres non premiers sont factorisés de manière immédiate : 2304167 = 2089 * 1103 ou 1037929037 = 27449 * 37813. Dans les cas les plus défavorables (lorsque n=pq avec p et q premiers et assez proches), on obtient de bons résultats. Par exemple le test avec 3196943 = 1787*1789 (277-ème et 278-ème nombres premiers) ou 753667193 = 27449 * 27457 (3000-ème et 3001-ème nombres premiers) renvoie un résultat immédiat.

On a aussi fait quelques tests avec des nombres premiers et l'ordinateur renvoie un message de non factorisation (il n'a pas trouvé de facteur) au bout de 5 secondes avec 3529, 30 secondes avec 11497 et 2 minutes avec 479909.

cf. Annexe 2.

Conclusion

La rapidité d'exécution de l'algorithme dépend surtout des facteurs premiers du nombre à factoriser et non du nombre lui-même. Il convient pour des recherches de facteurs premiers à environ une trentaine de chiffres. Par exemple 2436228587592235686865502196147 divise 6^{167} + 1 (trouvé par Silverman). A ce jour le plus grand facteur premier trouvé ainsi est de 47 chiffres par Peter Montgomery, au Centre de mathématiques et d'informatique d'Amsterdam. Notons que le temps d'execution de l'algorithme augmente très vite pour des facteurs très grands. On a alors recours à d'autres méthodes.

Références

[1] H.W. Lenstra, "Factoring integers with elliptic curves" report numéro 68-18, Math. Int., Univ. of Amsterdam, 1986.

[2] Journée annuelle de la société mathématique de France "Courbes elliptiques et applications", samedi 24 janvier 1987.

[3] Johannes Buchmann, "La factorisation des grands nombres" Pour la science, septembre 1998.

Annexe 1 : Exemples graphiques dans le cas des réels

Annexe 2 : Programme

Ce programme est écrit pour TurboPascal et permet de trouver pour un nombre entier de la taille ”longint” (inférieur ou égal à 2147483647) un facteur (premier ou non) ou bien renvoie un message pour dire qu'il n'a pas pu le factoriser.