Étude des distributions du problème LWE

Guilhem Repetto, ENS de Rennes

Stage de licence effectué au laboratoire GREYC de Caen,

encadré par Adeline Roux-Langlois (AMACC) et Matthieu Dien (SAFE),

du 15 mai au 30 juin 2023

Table des matières

Introduction

La cryptographie est l'art de sécuriser l'échange d'informations. L'un de ses objectifs principaux est de mettre au point des *schémas de chiffrement* permettant d'échanger des messages entre deux parties, de telle sorte qu'un adversaire qui intercepterait la transmission ne puisse rien apprendre du contenu des messages.

En pratique, la sécurité d'un schéma est garantie par la quantité de calculs nécessaires pour déchiffrer un message sans connaître la clef, ou plus généralement pour obtenir une information. On parle de *sécurité calculatoire*. Pour prouver la sécurité d'un schéma, on démontre mathématiquement qu'obtenir une information sur un message chiffré est plus difficile que de résoudre un problème calculatoire difficile connu. On effectue donc une *réduction* de ce problème réputé difficile vers le schéma de chiffrement. Le problème de la factorisation en nombres premiers en est un exemple, et la sécurité de plusieurs schémas repose dessus, comme par exemple le chiffrement RSA.

Le modèle de calcul quantique permet de faire baisser la complexité de beaucoup de problèmes calculatoires, dont ce dernier. La perspective de construire un ordinateur quantique efficace force donc à chercher de nouveaux chiffrements, dont la sécurité repose sur des problèmes difficiles à résoudre de façon classique et quantique. Un tel schéma est dit *post-quantique*.

En 2017, le NIST [1] a lancé un concours pour définir de nouveaux standards post-quantiques. Six ans plus tard, la sécurité de la plupart des schémas retenus repose notamment sur des problèmes sur les *réseaux euclidiens*.

Un des problèmes calculatoires les plus utilisés dans la cryptographie reposant sur les réseaux est le problème *Learning With Errors* (LWE). De façon informelle, un vecteur secret $\mathbf{s}$ d'un certain espace vectoriel est d'abord choisi au hasard, l'objectif pour l'attaquant étant de deviner $\mathbf{s}$. Pour cela, il dispose de vecteurs $\mathbf{a}_1, \dots, \mathbf{a}_m$ eux aussi tirés au hasard, ainsi que des angles entre $\mathbf{s}$ et ces vecteurs. Si les $\mathbf{a}_i$ n'engendrent pas tout l'espace, le problème est impossible à résoudre, car il admet plusieurs solutions. Dans le cas contraire, il est facile de retrouver $\mathbf{s}$, car on peut calculer sa décomposition dans une sous-base des $\mathbf{a}_i$. C'est pourquoi l'attaquant n'a pas à sa disposition l'angle entre chaque $\mathbf{a}_i$ et $\mathbf{s}$, mais seulement une *approximation* de cette valeur, d'où la notion d'*erreur*. Cette approximation provient de l'ajout d'un bruit provenant d'une distribution de probabilité bien choisie à la valeur de l'angle. Naturellement, plus le bruit est important, plus le problème est difficile.

La sécurité de LWE est bien connue lorsque le bruit est gaussien. En pratique, comme l'échantillonnage de bruit gaussien est relativement lent, et surtout sujet à des attaques par canaux cachés, on préfère utiliser un bruit provenant de distributions plus faciles à échantillonner, comme des lois uniformes ou binomiales.

Dans ce rapport, on se propose dans un premier temps de comprendre les liens entre quelques problèmes sur les réseaux euclidiens. Ensuite, on étudie l'influence de la distribution de probabilité choisie pour le chiffrement LWE, et l'impact sur la sécurité du schéma.

1 Les réseaux euclidiens pour la cryptographie

Nous commençons par quelques notations et définitions à propos des ensembles de matrices, et des lois de probabilité.

Notations 1.1. Soient $n, m \in \mathbb{N}^*$, et A un anneau. On note $A^{n\times m}$ l'ensemble des matrices à n lignes et m colonnes à coefficients dans A. On note $\mathrm{GL}_n(A)$ l'ensemble des matrices inversibles de $A^{n\times n}$ dont l'inverse est dans $A^{n\times n}$. Cet ensemble forme un groupe multiplicatif. En particulier, $\mathrm{GL}_n(\mathbb{Z})$ est formé des matrices de déterminant ± 1, parfois appelé *groupe unimodulaire*. Si q est un nombre premier, on note $\mathbb{Z}_q = \mathbb{Z}/q\mathbb{Z}$ le corps à q éléments.

1. *National Institute of Standards and Technology*, institution américaine dont la mission est de réaliser des mesures physiques, et de définir des normes et standards.

Définition 1.2. Si E est un ensemble non-vide, on note $U(E)$ la loi de probabilité uniforme sur E. Si $\mathcal{L}$ est une loi de probabilité, on note $X \hookleftarrow \mathcal{L}$ pour "X est tiré selon la loi $\mathcal{L}$". On définit aussi la *distribution gaussienne discrète* sur $\mathbb{Z}$ de paramètre α, notée $\mathcal{D}_{\mathbb{Z},\alpha}$, par

$$\forall n \in \mathbb{Z}, \quad \mathcal{D}_{\mathbb{Z},\alpha}(n) = \frac{\rho_\alpha(n)}{\sum_{k=-\infty}^{+\infty} \rho_\alpha(k)} \quad \text{où} \quad \forall x \in \mathbb{R}, \quad \rho_\alpha(x) = \frac{1}{\alpha} \exp\left(-\pi \left(\frac{x}{\alpha}\right)^2\right)$$

Les problèmes considérés ici comme "faciles" sont ceux dont la résolution se fait en un temps au plus polynomial en la taille de l'instance. Une définition précise de l'ensemble **poly** est donnée ci-dessous, ainsi que celle de l'ensemble **negl** des *fonctions négligeables* :

Définition 1.3. Si f est une fonction définie sur $\mathbb{N}$, on note $f(n) = \mathbf{poly}(n)$ s'il existe un polynôme p tel que $f(n) = \mathcal{O}(p(n))$. Si g est une fonction, on dit que g est *négligeable* si elle décroît plus vite que l'inverse de tout polynôme, c'est-à-dire $1/p(n) = o(g(n))$ pour tout polynôme p. On note **negl** l'ensemble des fonctions négligeables.

Les preuves de sécurité étudiées ici sont fondées sur la notion de *réduction*. Ainsi, pour montrer qu'un problème `B` est "difficile", on montre qu'il est au moins aussi difficile qu'un problème `A` de référence, dont la difficulté est supposée. On dit ainsi que l'on réduit `A` à `B`. Les définitions suivantes formalisent cela, pour les problèmes décisionnels et calculatoires.

Définition 1.4. Soient `A` et `B` deux problèmes décisionnels. On dit que `A` *se réduit à* `B` s'il existe une fonction φ qui prend une instance de `A` et renvoie une instance de `B` telle que

— I est une instance positive de `A` si et seulement si $\varphi(I)$ est une instance positive de `B`,

— $\varphi(I)$ est calculable en temps $\mathbf{poly}(|I|)$.

Intuitivement, `A` est "plus facile" que `B`.

Définition 1.5. Soient `A` et `B` deux problèmes calculatoires. On dit que `A` *se réduit à* `B` s'il existe une fonction φ qui prend une instance de `A` et renvoie une instance de `B`, et une fonction ψ qui prend une instance de `B` et une solution à cette instance, et renvoie une solution à une instance de `A`, telles que

— Si s est une solution de $\varphi(I)$, alors $\psi(\varphi(I), s)$ est une solution de I,

— $\varphi(I)$ est calculable en temps $\mathbf{poly}(|I|)$, et $\psi(J, s)$ est calculable en temps $\mathbf{poly}(|J|, |s|)$.

Enfin, pour modéliser un adversaire qui veut effectuer des actions sur des messages chiffrés, comme par exemple obtenir une information ou déchiffrer le message, on adopte le modèle du *distingueur probabiliste polynomial*. Cela modélise un agent dont la capacité de calcul est polynomiale, et qui tente d'effectuer une certaine action.

Définition 1.6. Soient $\{X_n\}_{n\in\mathbb{N}}$ et $\{Y_n\}_{n\in\mathbb{N}}$ deux suites de variables aléatoires. Un *distingueur probabiliste polynomial*, abrégé en PPT pour *probabilistic polynomial time*, est un algorithme $\mathcal{D}$

— qui prend en entrée $n \in \mathbb{N}$, et un élément ω tiré soit selon la loi X_n, soit selon la loi Y_n

— renvoie un bit 0 ou 1

— qui s'exécute en temps $\mathbf{poly}(n)$, éventuellement de façon non-déterministe

L'objectif de $\mathcal{D}$ est d'identifier de quelle distribution X_n ou Y_n provient l'élément ω ; il renvoie 0 pour X et 1 pour Y.

On note $\mathcal{D}(1^n, X_n)$ (respectivement $\mathcal{D}(1^n, Y_n)$) pour une exécution où $\omega \hookleftarrow X_n$ (resp. $\omega \hookleftarrow Y_n$). La notation 1^n fait référence au nombre n écrit en unaire. L'*avantage* de $\mathcal{D}$ est la suite définie par

$$Adv(\mathcal{D}) = \{|\mathbb{P}[\mathcal{D}(1^n, X_n) = 1] - \mathbb{P}[\mathcal{D}(1^n, Y_n) = 1]|\}_{n\in\mathbb{N}}$$

Intuitivement, l'avantage de $\mathcal{D}$ est non-négligeable lorsque $\mathcal{D}$ distingue $\{X_n\}_n$ de $\{Y_n\}_n$ de façon non-négligeable lorsque n tend vers l'infini. Dans le cas contraire, on dit que $\{X_n\}_n$ et $\{Y_n\}_n$ sont *calculatoirement indistinguables*. Dans le cadre de la cryptographie, on voit ces distributions comme des instances de problèmes cryptographiques, avec n comme paramètre de sécurité.

1.1 Définitions des réseaux euclidiens

Définissons le contexte des réseaux euclidiens, et énonçons les problèmes difficiles qui servent de base à la cryptographie basée sur les réseaux.

On considère l'espace vectoriel euclidien $(\mathbb{R}^n, \langle , \rangle)$, où $\langle , \rangle$ est le produit scalaire usuel, de norme notée $\|\|$.

Définition 1.7. Un *réseau* Λ de $\mathbb{R}^n$ est une partie de $\mathbb{R}^n$ telle qu'il existe une famille libre $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_k)$ de $\mathbb{R}^n$ vérifiant $\Lambda = \left\{ x \in \mathbb{R}^n \ / \ \exists (x_1, \ldots, x_k) \in \mathbb{Z}^k, x = \sum_{i=1}^k x_i \mathbf{b}_i \right\}$. L'entier k est appelé *dimension* du réseau, et noté $\dim \Lambda$. On remarque que Λ est toujours un sous-groupe discret de $(\mathbb{R}^n, +)$. Si $\mathbf{B}$ est une famille libre, on note $\Lambda(\mathbf{B}) = \left\{ \sum_{i=1}^k x_i \mathbf{b}_i \ / \ (x_1, \ldots, x_k) \in \mathbb{Z}^k \right\}$ le *réseau engendré* par cette famille. On dit que $\mathbf{B}$ est une *base* de $\Lambda(\mathbf{B})$. L'espace vectoriel engendré par $\mathbf{B}$ de $\Lambda(\mathbf{B})$ est noté $\mathrm{span}(\Lambda(\mathbf{B}))$.

Un réseau non trivial possède une infinité de bases. En effet, si $\mathbf{B} \in \mathbb{R}^{n \times n}$ est une base de Λ, on montre que les bases sont exactement les $U\mathbf{B}$ où U parcourt $\mathrm{GL}_n(\mathbb{Z})$. Cela permet d'obtenir la définition suivante :

Définition 1.8. On définit le *volume* de Λ, noté $\mathrm{vol}(\Lambda)$, par la quantité $\sqrt{\det(\mathbf{B}^T\mathbf{B})}$, où $\mathbf{B}$ est une base quelconque de Λ. Dans le cas où $\mathrm{span}(\Lambda) = \mathbb{R}^n$, on a l'égalité $\mathrm{vol}(\Lambda) = |\det(\mathbf{B})|$.

La difficulté des problèmes que nous allons voir par la suite est liée à la taille des vecteurs d'une base d'un réseau. En particulier, plus la base dont on dispose comporte des vecteurs "courts", plus les problèmes sont faciles. Définissons donc une mesure de la taille de ces vecteurs courts :

Définition 1.9. On note $\lambda_1(\Lambda)$ la taille du plus court vecteur non-nul de Λ : $\lambda_1(\Lambda) = \min_{v \in \Lambda \setminus \{\mathbf{0}\}} \|v\|$. On définit plus généralement $\lambda_i(\Lambda)$ pour $i \in [\![1, n]\!]$ par

$$\lambda_i(\Lambda) = \inf_{\ell \in \mathbb{R}_+} \{\ell, \ \dim(\mathrm{span}(\overline{B(\mathbf{0}, \ell)} \cap \Lambda)) \geq i\},$$

où $\overline{B(\mathbf{0}, \ell)}$ est la boule fermée de $\mathbb{R}^n$ de centre $\mathbf{0}$ et de rayon ℓ. Ainsi, $\lambda_i(\Lambda)$ est la plus petite longueur telle que l'ensemble des vecteurs de Λ de norme inférieure ou égale à $\lambda_i(\Lambda)$ engendre un espace vectoriel de dimension au moins i.

Le théorème suivant donne un ordre de grandeur du vecteur le plus court d'un réseau :

Théorème 1.10. (Minkowski) Pour tout réseau Λ, on a $\lambda_1(\Lambda) \leq \sqrt{n}\,(\mathrm{vol}(\Lambda))^{\frac{1}{n}}$.

1.2 Problèmes difficiles sur les réseaux euclidiens

La complexité des problèmes basés sur les réseaux réside, dans le cadre de cette étude, sur la difficulté d'en trouver des vecteurs courts, c'est-à-dire comparables en un sens à préciser à $\lambda_1(\Lambda)$. En particulier, le "Shortest Vector Problem" (SVP) est un problème standard qui sert de base pour de nombreuses preuves de sécurité par réduction. Nous énonçons ici les différentes variantes de SVP. Dans tous les cas, une base $\mathbf{B} \in \mathbb{Z}^{m \times n}$ d'un certain réseau $\Lambda(\mathbf{B})$ est donnée en entrée.

Search SVP	Trouver un vecteur $\mathbf{v} \in \Lambda(\mathbf{B})$ de norme $\lambda_1(\Lambda(\mathbf{B}))$.
Optimisation SVP	Déterminer $\lambda_1(\Lambda(\mathbf{B}))$.
Decisional SVP	Étant donné $r \in \mathbb{Q}$, déterminer si $\lambda_1(\Lambda(\mathbf{B})) < r$.

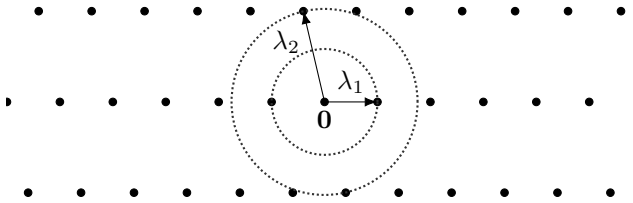


FIGURE 1 – Un réseau de $\mathbb{R}^2$ muni d'une base de vecteurs les plus courts possibles, soit λ_1 et λ_2.

Il se trouve que ces trois problèmes sont calculatoirement équivalents. En pratique, dans les preuves de sécurité de schémas cryptographiques, ce sont des versions approximées qui servent de garantie de sécurité. Certains de ces problèmes sont définis ci-dessous, et leur complexité est discutée en partie 2.2 pour plusieurs paramètres d'approximation. On y ajoute SIVP, qui signifie Shortest Independant Vector Problem, et qui consiste à trouver une famille libre de vecteur de tailles raisonnables. Pour chaque problème, γ est un réel strictement supérieur à 1, qui représente le facteur d'approximation.

Search SVP$_\gamma$	Trouver un vecteur $v \in \Lambda(\mathbf{B}) \setminus \{\mathbf{0}\}$ tel que $\|v\| \leq \gamma \times \lambda_1(\Lambda(\mathbf{B}))$.
Optimisation SVP$_\gamma$	Déterminer $d \in \mathbb{Q}$ tel que $d \leq \lambda_1(\Lambda(\mathbf{B})) \leq \gamma \times d$.
SIVP$_\gamma$	Trouver une famille libre $\mathbf{v}$ telle que $\forall i, \|\mathbf{v}_i\| \leq \gamma \times \lambda_n(\Lambda)$.
GapSVP$_\gamma$	Étant donné $r \in \mathbb{Q}$, déterminer si $\lambda_1(\Lambda(\mathbf{B})) < r$ ou $\lambda_1(\Lambda(\mathbf{B})) > \gamma \times r$.

Plus la base $\mathbf{B}$ donnée en entrée comporte des vecteurs courts, plus il semble facile de construire un vecteur qui satisfasse l'un des problèmes ci-dessus.

2 Le problème Learning With Errors

2.1 Définition

Le problème Learning With Errors (LWE) consiste à résoudre un système d'équations linéaires bruitées. Commençons par définir la distribution de probabilités $\mathcal{D}^{\text{LWE}}_{n,q,\alpha}(\mathbf{s})$, dépendant d'un vecteur $\mathbf{s}$:

Définition 2.1. Étant donné un vecteur $\mathbf{s} \in \mathbb{Z}_q^n$, on note $\mathcal{D}^{\text{LWE}}_{n,q,\alpha}(\mathbf{s})$ la distribution de probabilité sur $\mathbb{Z}_q^n \times \mathbb{Z}_q$ définie par $(\mathbf{a}, \langle \mathbf{a}, \mathbf{s} \rangle + \mathbf{e})$ où $\mathbf{a} \hookleftarrow U(\mathbb{Z}_q^n)$ et $\mathbf{e} \hookleftarrow \mathcal{D}_{\mathbb{Z},\alpha q}$.

Ainsi, un échantillon $(\mathbf{a}, \mathbf{b})$ issu de la distribution $\mathcal{D}^{\text{LWE}}_{n,q,\alpha}(\mathbf{s})$ est composé d'un vecteur $\mathbf{a}$, et d'une approximation du produit scalaire de $\mathbf{a}$ et $\mathbf{s}$. La qualité de l'approximation dépend donc de αq : plus cette quantité est petite, meilleure est l'approximation.

Le problème LWE consiste essentiellement à retrouver $\mathbf{s}$ à partir d'événements issus de la distribution $\mathcal{D}^{\text{LWE}}_{n,q,\alpha}(\mathbf{s})$. Nous pouvons alors définir les versions calculatoires du problème :

sLWE$_{n,q,\alpha}$ Un "challenger" $\mathcal{C}$ génère un vecteur secret $\mathbf{s} \hookleftarrow U(\mathbb{Z}_q^n)$. Un attaquant $\mathcal{A}$ adresse alors des requêtes à $\mathcal{C}$, qui lui envoie des éléments $(\mathbf{a}_i, \mathbf{b}_i) \hookleftarrow \mathcal{D}^{\text{LWE}}_{n,q,\alpha}(\mathbf{s})$. L'objectif de $\mathcal{A}$ est de déterminer $\mathbf{s}$.

sLWE$_{n,q,\alpha,m}$ Identique à LWE$_{n,q,\alpha}$ avec un nombre requêtes m fixé d'avance. Une version matricielle équivalente est la suivante : $\mathcal{C}$ publie $(\mathbf{A}, \mathbf{As}+\mathbf{e})$, où $\mathbf{s} \hookleftarrow U(\mathbb{Z}_q^n)$, $\mathbf{A} \hookleftarrow U(\mathbb{Z}_q^{n \times m})$, et $\mathbf{e} \hookleftarrow \mathcal{D}^m_{\alpha q}$. L'objectif de $\mathcal{A}$ est de déterminer $\mathbf{s}$.

Ainsi, $\mathcal{A}$ reçoit une suite d'égalités bruitées telles que présentées par la figure 2, et tente de retrouver $\mathbf{s}$.

$$\begin{cases} 2\mathbf{s}_1 + 4\mathbf{s}_2 + 4\mathbf{s}_3 & \approx 3 \mod 5 \\ 0\mathbf{s}_1 + 1\mathbf{s}_2 + 2\mathbf{s}_3 & \approx 4 \mod 5 \\ 3\mathbf{s}_1 + 4\mathbf{s}_2 + 0\mathbf{s}_3 & \approx 1 \mod 5 \\ 2\mathbf{s}_1 + 2\mathbf{s}_2 + 3\mathbf{s}_3 & \approx 0 \mod 5 \\ 3\mathbf{s}_1 + 4\mathbf{s}_2 + 1\mathbf{s}_3 & \approx 1 \mod 5 \end{cases} \iff \begin{pmatrix} 2 & 4 & 4 \\ 0 & 1 & 2 \\ 3 & 4 & 0 \\ 2 & 2 & 3 \\ 3 & 4 & 1 \end{pmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \mathbf{s}_3 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \\ \mathbf{e}_4 \\ \mathbf{e}_5 \end{bmatrix} = \begin{pmatrix} 3 \\ 4 \\ 1 \\ 0 \\ 1 \end{pmatrix} \mod 5$$

FIGURE 2 – Requêtes issues d'une instance de $\text{sLWE}_{3,5,\alpha,5}$

Le problème LWE se décline en version décisionnelle. Il s'agit ici de distinguer entre des échantillons provenant de $\mathcal{D}^{\text{LWE}}_{n,q,\alpha}(\mathbf{s})$ et des échantillons tirés selon la loi uniforme, ne donnant aucune information sur le vecteur secret $\mathbf{s}$:

$\text{LWE}_{n,q,\alpha}$ Un "challenger" $\mathcal{C}$ génère un bit $b \hookleftarrow U(\{0;1\})$, et un vecteur secret $\mathbf{s} \hookleftarrow U(\mathbb{Z}_q^n)$. Un attaquant $\mathcal{A}$ adresse alors des requêtes à $\mathcal{C}$, qui renvoie des éléments de $\mathbb{Z}_q^n \times \mathbb{Z}_q$ aléatoires répartis selon

— la distribution $U(\mathbb{Z}_q^n \times \mathbb{Z}_q)$ si $b = 0$;

— la distribution $\mathcal{D}^{\text{LWE}}_{n,q,\alpha}(\mathbf{s})$ si $b = 1$.

L'objectif de $\mathcal{A}$ est de déterminer b, c'est à dire de distinguer les distributions $\mathcal{D}^{\text{LWE}}_{n,q,\alpha}(\mathbf{s})$ et $U(\mathbb{Z}_q^n \times \mathbb{Z}_q)$.

$\text{LWE}_{n,q,\alpha,m}$ Identique, avec un nombre de requêtes m fixé à l'avance.

2.2 Difficulté de LWE

On constate que plus α est petit, plus le problème LWE est facile. En particulier, si $\alpha = 0$, le problème revient à résoudre un système linéaire, ce qui se fait en temps polynomial. Si α est trop grand, le problème devient impossible à résoudre, car le bruit couvre toute l'information de l'angle entre $\mathbf{a}$ et $\mathbf{s}$.

La complexité des problèmes en jeu est ici étudiée de façon asymptotique. Le choix fait dans la littérature est d'étudier en premier un problème en considérant que ses paramètres sont variables, pour démontrer une complexité asymptotique suffisante, puis d'étudier des instances particulières du problème avec des paramètres fixés, en vue d'une application réelle.

La figure 3 représente de façon simpliste différentes réductions entre les problèmes, sans indiquer les paramètres des les problèmes ni ceux des réductions. En particulier, elle représente la réduction donnée par le théorème suivant :

Théorème 2.2. (Brakerski, Langlois, Peikert, Regev, Stehlé. 2013) Le problème GapSVP_γ dans un réseau de dimension $\Theta(\sqrt{n})$ se réduit à $\text{LWE}_{n,q,\alpha}$, pour $\gamma = \mathcal{O}(n^2/\alpha)$.

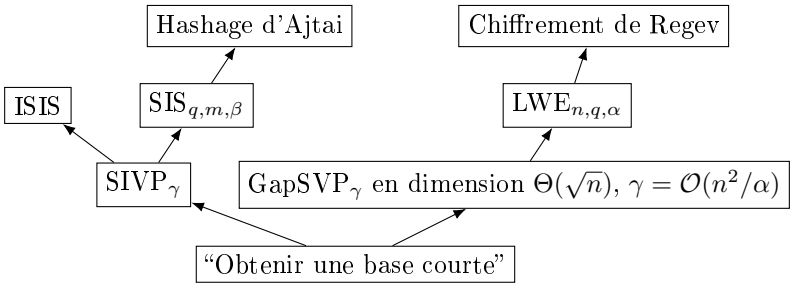


FIGURE 3 – Réductions entre plusieurs problèmes liés aux réseaux euclidiens. On note `A` → `B` pour "`A` se réduit à `B`". Les paramètres des problèmes ne sont pas spécifiés.

Comme GapSVP_γ se réduit à LWE, on peut étudier la complexité de GapSVP_γ pour avoir une garantie de complexité de LWE. La figure 4 résume les complexités du problème GapSVP_γ pour plusieurs γ. La sécurité du chiffrement reposant sur les réseaux vient de la conjecture que $\text{GapSVP}_{\mathbf{poly}(n)}$ est suffisamment difficile.

γ	$\leq \mathcal{O}(1/\log\log n)$	$\mathcal{O}(\sqrt{n})$	$\mathcal{O}(n)$	$\mathbf{poly}(n)$	$\geq 2^{\mathcal{O}(n)}$
Complexité	NP-difficile	NP ∩ coNP	coNP	?	P

FIGURE 4 – Complexité de GapSVP_γ pour plusieurs fonctions γ de n.

2.3 Chiffrement de Regev

Le chiffrement de Regev repose directement sur LWE. Ses paramètres sont, comme pour le problème LWE, des entiers n, m, un nombre premier q, et un réel α, avec m vérifiant $m \geq 4(n+1)\log_2 q$, et $\alpha \in \left]0; \frac{1}{8m}\right[$. Notons que le l'opération de déchiffrement est probabiliste, et peut donc échouer avec une faible probabilité.

Génération de clef Le clef secrète sk est un vecteur $\mathbf{s} \hookleftarrow U(\mathbb{Z}_q^n)$. La clef publique pk est un couple $(\mathbf{A}, \mathbf{As}+\mathbf{e})$ de $\mathbb{Z}_q^{n\times m} \times \mathbb{Z}_q^m$, tirée selon la distribution $\mathcal{D}^{\text{LWE}}_{n,q,\alpha q,m}(\mathbf{s})$.

Chiffrement Pour chiffrer un bit $m \in \{0;1\}$ avec la clef $(\mathbf{A}, \mathbf{b})$, tirer $\mathbf{r} \hookleftarrow U(\{0;1\}^m)$, et renvoyer $\left(\mathbf{r}^T\mathbf{A}, \mathbf{r}^T\mathbf{b} + m \cdot \left\lfloor \frac{q}{2} \right\rfloor\right)$.

Déchiffrement Pour déchiffrer $(\mathbf{c}^T, x)$, calculer $x - \mathbf{c}^T \cdot \mathbf{s}$. Si le résultat est plus proche de 0 que de $\lfloor q/2 \rfloor$, renvoyer 0, sinon renvoyer 1.

On vérifie que ce chiffrement est correct : si $(\mathbf{c}^T, x) = \left(\mathbf{r}^T\mathbf{A}, \mathbf{r}^T\mathbf{b} + m \cdot \left\lfloor \frac{q}{2} \right\rfloor\right)$ est un chiffré, alors $x - \mathbf{c}^T \cdot \mathbf{s} = \mathbf{r}^T\mathbf{e} + m \cdot \left\lfloor \frac{q}{2} \right\rfloor$. Comme la masse sous la gaussienne est majoritairement concentrée dans $[-\sqrt{\alpha q}, \sqrt{\alpha q}]$ et $\sqrt{\alpha q} \leq \frac{q}{8m}$, on a avec probabilité supérieure à $1 - 2^{-n}$ la majoration suivante :

$$\mathbf{r}^T\mathbf{e} \leq \|\mathbf{r}\| \cdot \|\mathbf{e}\| \leq \sqrt{m} \cdot \sqrt{m}\frac{q}{8m} = \frac{q}{8}$$

Ainsi, le terme $m \cdot \left\lfloor \frac{q}{2} \right\rfloor$ est prépondérant dans la majorité des cas, et permet de décider du bit chiffré.

En pratique, ce chiffrement est peu efficace : la taille des clefs est importante, ainsi que la taille des messages chiffrés, et beaucoup de calculs sont nécessaires. Ce sont donc des variantes dites *structurées* de ce chiffrement qui sont utilisées, telles ring-LWE ou module-LWE.

3 Analyse d'avantages et divergence de Rényi

La sécurité du problème LWE repose sur l'utilisation de l'erreur gaussienne. Il existe cependant d'autres choix possibles d'erreur, afin de réduire le temps d'échantillonnage et résister à des attaques à canaux cachés. Il convient donc de comparer la sécurité des dérivés de LWE avec le problème initial. L'objectif est de montrer que si l'attaque réussit avec probabilité non-négligeable[2] sur une variante donnée, alors elle réussit aussi sur le problème LWE standard avec probabilité non-négligeable.

Pour comparer deux distributions de probabilité de réussite d'une attaque, on utilise des notamment la distance statistique et la divergence de Rényi. Dans la suite, il faut voir D_1 comme la distribution de probabilité qui modélise la réussite d'une attaque sur un variant de LWE, et D_2 celle qui modélise la réussite des attaques sur LWE.

2. On considère une suite d'instances avec un paramètre de sécurité n qui tend vers l'infini, et on dit que l'attaque réussit avec probabilité non-négligeable si la suite des probabilités de réussite est une suite non-négligeable.

— La distance statistique entre deux distributions D_1 et D_2 de même support X est définie par $\Delta(D_1, D_2) = \frac{1}{2}\int_X |D_1(x) - D_2(x)|\,\mathrm{d}x$. La propriété utilisée est que si E est un ensemble mesurable, alors on a

$$D_2(E) \geq D_1(E) - \Delta(D_1, D_2). \tag{1}$$

Cette inégalité entraîne que si $\Delta(D_1, D_2)$ est négligeable, et que $D_1(E)$ est non-négligeable, alors $D_2(E)$ est non-négligeable.

— La divergence de Rényi est une alternative multiplicative, qui n'est cependant pas une distance. Pour $a \in]1; +\infty[$, on définit

$$R_a(D_1||D_2) = \left(\int_X \frac{D_1(x)^a}{D_2(x)^{a-1}}\,\mathrm{d}x\right)^{\frac{1}{a-1}}.$$

L'inégalité qui est l'équivalent de (1) est la suivante :

$$D_2(E) \geq D_1(E)^{\frac{a}{a-1}}/R_a(D_1||D_2). \tag{2}$$

En particulier, pour $a = 2$, l'équation devient $D_2(E) \geq D_1(E)^2/R_2(D_1||D_2)$. Une autre propriété importante est la multiplicativité de la divergence : si D_i^m représente un vecteur de m tirages indépendants selon la distribution D_i, alors on a $R_a(D_1^m||D_2^m) = R_a(D_1||D_2)^m$.

L'objectif de ce rapport est donc notamment d'étudier des divergences où D_2 est la gaussienne continue ou discrète, et D_1 une distribution "proche" de D_2 au sens de la divergence de Rényi pour un certain paramètre a.

4 Bibliographie commentée

Pour comparer la sécurité d'un schéma de chiffrement $\mathcal{S}$ à celle d'un schéma de référence $\mathcal{S}_{\text{ref}}$, on utilise usuellement la distance statistique. Elle permet notamment de comparer la probabilité p pour un attaquant de réussir une attaque sur $\mathcal{S}$ à celle p_{ref} de réussir une attaque sur $\mathcal{S}_{\text{ref}}$. L'article [1] montre que la divergence de Rényi peut jouer un rôle similaire, et présente certains avantages par rapport à la distance statistique. Notamment, les auteurs effectuent une réduction de $\text{LWE}_{n,q,D_\alpha,m}$ vers $\text{LWE}_{n,q,\phi,m}$, où ϕ est un bruit uniforme. La divergence de Rényi est utilisée pour montrer que si un attaquant possède un avantage non-négligeable pour $\text{LWE}_{n,q,\phi,m}$, alors il possède aussi un avantage non-négligeable pour $\text{LWE}_{n,q,D_\alpha,m}$. En effet, l'équation (2) est mise à profit, avec la preuve que la divergence de Rényi est polynomiale.

L'article [2] liste les propriétés principales de la divergence de Rényi, notamment des inégalités et des propriétés de continuité. A noter que la définition de divergence choisie dans cet article est, pour des raisons de simplicité dans les formules, le logarithme de celle choisie dans [1].

L'article [3] donne une description des problèmes difficiles sur les réseaux euclidiens, tels que CVP, SVP etc. Il décrit des algorithmes pour résoudre certains problèmes, comme réduire la taille des bases avec l'algorithme LLL. Enfin, une figure montre les différentes relations de réduction entre les problèmes, dont est inspirée la figure 3 de ce rapport.

L'ouvrage [4] présente différentes familles de distributions de probabilités dites *stables*. Ces familles sont composées de lois continues indexées par quatre paramètres. Elles sont telles que si deux variables aléatoires indépendantes suivent des lois appartenant à une famille, alors les combinaisons linéaires de ces variables suit aussi une loi de la famille. En faisant varier les paramètres, on obtient les lois normale, gamma, de Cauchy, de Lévy etc.

Dans son exposé [5], O. Regev présente sa preuve de la difficulté du problème LWE. Il explique une réduction pire-cas / moyen-cas de problèmes sur les réseaux, en particulier BDD_d, vers LWE. Sa preuve comporte des réductions quantiques, c'est-à-dire nécessitant des opérations appartenant au modèle de calcul quantique.

D'après le théorème central limite, une somme de lois indépendantes identiquement distribuées centrées converge en loi vers une loi normale. L'article [7] établit notamment des inégalités et approximations asymptotiques entre une somme de lois uniformes et une gaussienne. Les théorèmes 1.2 et 1.3 ont un intérêt pour notre étude.

5 Contributions

Une des applications de la divergence est de montrer la difficulté de variantes d'une hypothèse standard. Dans l'article [1], les auteurs montrent par réduction qu'une variante de LWE utilisant une distribution de probabilité uniforme est au moins aussi difficile que LWE standard. La réduction se fait en introduisant quatre problèmes intermédiaires et en détaillant les réductions de l'un à l'autre. En particulier, la troisième étape est la réduction de sLWE avec une loi gaussienne+uniforme à sLWE avec loi uniforme. Elle exploite directement l'équation (2) pour montrer qu'un avantage non-négligeable pour sLWE avec loi uniforme donne un avantage non-négligeable pour sLWE avec loi gaussienne+uniforme. On peut donc espérer se servir de la divergence pour étudier des variantes de chiffrement basé sur LWE où un compromis est fait entre rapidité et précision de l'échantillonnage des lois. Par exemple, le chiffrement CRYSTALS-Kyber, présenté dans [6], et actuel candidat à la standardisation NIST, utilise une loi binomiale centrée plutôt qu'une gaussienne afin d'avoir un échantillonnage plus rapide et moins vulnérable aux attaques par canaux cachés, notamment temporelles.

L'objectif du stage est d'essayer d'autres distributions pour comprendre comment elles font évoluer la difficulté des problèmes. Après quelques essais, la distribution triangle a été choisie (section 5.2), et sa distance avec la distribution gaussienne a été étudiée numériquement. Après avoir remarqué que la loi triangle est simplement la loi suivie par la somme de deux lois uniformes indépendantes, l'étude s'est naturellement portée sur la somme de trois lois uniformes, dans le but de trouver les paramètres optimaux pour approximer une gaussienne (section 5.3).

5.1 Tracé de divergences de lois stables

On se propose de représenter graphiquement quelques divergences.

5.1.1 Divergence des lois normales entre elles

Définition 5.1. On note $\rho_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2)$ la loi normale d'écart-type $\sigma > 0$, définie sur $\mathbb{R}$.

On considère la fonction $R_2^\rho : \sigma \mapsto R_2(\rho_\sigma || \rho_1)$. On a l'expression close $R_2^\rho(\sigma) = \frac{1}{\sigma}\frac{1}{\sqrt{2-\sigma^2}}$. On a donc en particulier $R_2^\rho(\sigma) \sim_0 \frac{1}{\sqrt{2}\sigma}$.

5.1.2 Divergence des lois de Cauchy entre elles

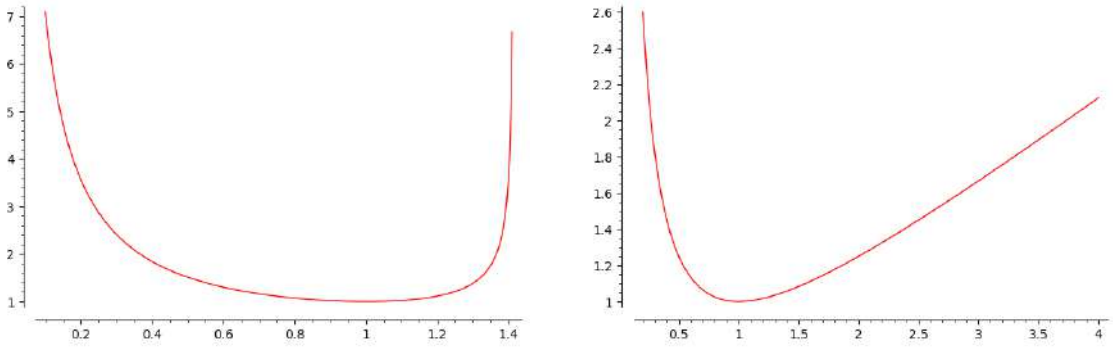
Définition 5.2. On note $C_\gamma(x) = \frac{1}{\pi}\frac{\gamma}{x^2+\gamma^2}$ la loi de Cauchy de paramètre $\gamma > 0$, définie sur $\mathbb{R}$.

On considère la fonction $R_2^C : \gamma \mapsto R_2(C_\gamma || C_1)$. On a l'expression close $R_2^C(\gamma) = \frac{\gamma^2+1}{2\gamma}$. On a donc en particulier $R_2^C(\gamma) \sim_0 \frac{1}{2\gamma}$ et $R_2^C(\gamma) \sim_\infty \frac{\gamma}{2}$. La figure 5 représente le graphe des deux fonctions étudiées.

5.2 Étude de la loi triangle

5.2.1 Somme de deux uniformes : la loi triangle

On note $T(x) = -\alpha^2 x + \alpha$ sur $[0; 1/\alpha]$, $\alpha^2 x + \alpha$ sur $[-1/\alpha; 0]$, et 0 ailleurs. Il s'agit d'un triangle centré en 0. Il s'agit également de la loi suivie par la somme de deux lois uniformes continues centrées.



(a) $R_2^\rho : \sigma \mapsto R_2(\rho_\sigma || \rho_1)$ (b) $R_2^C : \gamma \mapsto R_2(C_\gamma || C_1)$

FIGURE 5 – Représentation des divergences entre des fonctions d'une même famille de lois stables.

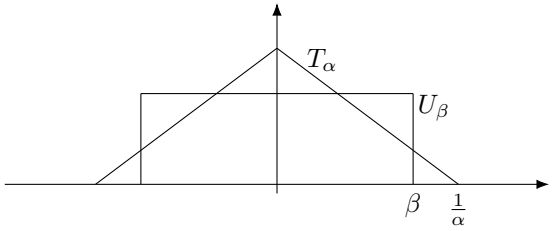


FIGURE 6 – Graphe des distributions U_β et T_α.

5.2.2 Construction de la fonction triangle

Nous allons partir de la fonction $x \mapsto -\alpha^2 |x| + \alpha$ sur $\mathbb{R}$, et lui ajouter une fonction affine par morceaux construite à l'aide de la fonction valeur absolue, pour obtenir une expression close de la distribution triangle.

D'abord, $x \mapsto \frac{|x|-x}{2}$ est la fonction $x \mapsto -x$ sur $\mathbb{R}_-$ et $x \mapsto 0$ sur $\mathbb{R}_+$. Donc $x \mapsto \alpha^2 \frac{|x|-x}{2}$ est la fonction $x \mapsto -\alpha^2 x$ sur $\mathbb{R}_-$ et $x \mapsto 0$ sur $\mathbb{R}_+$. Donc $x \mapsto \alpha^2 \frac{|x+1/\alpha|-x-1/\alpha}{2}$ est la même fonction décalée de $1/\alpha$ vers la gauche. Donc $x \mapsto \alpha^2 \frac{|-x+1/\alpha|+x-1/\alpha}{2}$ est la fonction précédente symétrisée par rapport à l'axe des abscisses. Ainsi, après réarrangement, on obtient

$$T_\alpha : x \mapsto -\alpha^2 |x| + \alpha + \frac{\alpha}{2}(|\alpha x - 1| + |\alpha x + 1| - 2).$$

Une primitive de T_α est donc

$$x \mapsto -\alpha^2 \frac{x|x|}{2} + \alpha x + \frac{\alpha}{2}\left(\frac{(\alpha x - 1)|\alpha x - 1|}{2\alpha} + \frac{(\alpha x + 1)|\alpha x + 1|}{2\alpha} - 2x\right).$$

La distribution et une de ses primitives sont représentées en figure 7.

5.2.3 Approximation d'un triangle par uniforme + triangle

Pour tenter de mettre en œuvre la réduction de l'article [1], on estime $R_2(T_\alpha || U_\beta + T_\alpha)$. La figure 8 représente un triangle T_α, et plusieurs lois $T_\alpha + U_\beta$ pour plusieurs valeurs de β. On constate graphiquement que plus β est petit, meilleure est l'approximation.

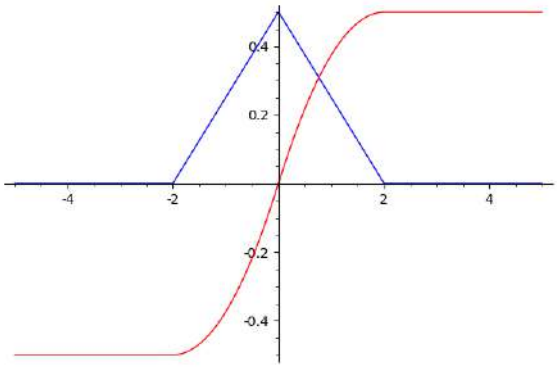


FIGURE 7 – Distribution T_α (bleu), et une de ses primitives (rouge).

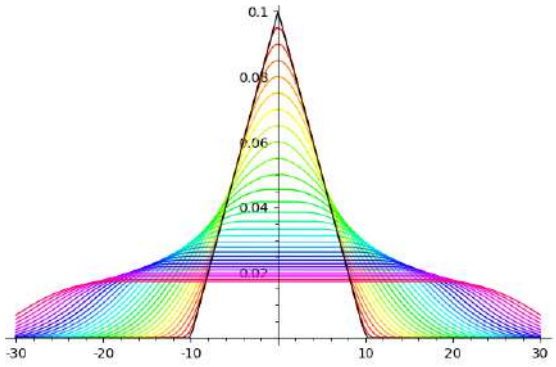


FIGURE 8 – Comparaison entre la loi triangle et des lois triangle + uniforme. Les petites valeurs de β correspondent aux courbes rouges, et les grandes valeurs de β aux courbes violettes.

De la même façon que dans l'article [1], on souhaite que $R_2(T_\alpha^m||(U_\beta + T_\alpha)^m)$ soit majoré par un $n^{\mathcal{O}(1)}$. Il faut d'abord calculer $T_\alpha * U_\beta$:

$$T_\alpha * U_\beta(x) = \frac{\alpha^2 \left|\beta + x\right| (\beta + x) + \alpha^2 \left|\beta - x\right| (\beta - x) - 4\alpha\beta}{4\beta}.$$

Il faut ensuite étudier la quantité suivante pour espérer obtenir un $n^{\mathcal{O}(1)}$:

$$\begin{aligned} R_2(T_\alpha||U_\beta + T_\alpha) &= \int_{-\beta}^{\beta} \frac{T_\alpha^2(x)}{T_\alpha * U_\beta(x)} \,\mathrm{d}x \\ &= -2\int_{-\beta}^{\beta} \frac{\left(\alpha^2|x| - \alpha\right)^2}{\alpha^2(\beta+x)|\beta+x| + \alpha^2(\beta-x)|-\beta+x| - 2\alpha(\beta+x) - 2\alpha(\beta-x)} \,\mathrm{d}x. \end{aligned}$$

5.3 Estimation asymptotique de la distance statistique et de la divergence entre somme d'uniformes et gaussienne

L'article [7] présente des approximations asymptotiques entre une somme de variables uniformes et une loi gaussienne. De telles approximations ouvrent des perspectives intéressantes pour échantillonner rapidement une loi proche de la gaussienne.

On note X une variable aléatoire réelle continue telle que $\mathbb{E}(X) = 0$ et $\mathbb{E}(X^2) = 1$, puis $X_1, \ldots, X_k$ des

copies indépendantes de X, et $Z_k = \frac{X_1+\cdots+X_k}{\sqrt{k}}$. Le théorème 1.2 de l'article s'adapte de la façon suivante au cas de la dimension 1 :

Théorème 5.3. $R_2(Z_k||\rho_1)$ tend vers 1 si et seulement si $R_2(Z_k||\rho_1)$ est fini pour un certain k, et que $\forall t \in \mathbb{R}^*, \mathbb{E}\left(e^{tX}\right) < e^{t^2}$. Dans ce cas, on a $R_2(Z_k||\rho_1) = e^{\mathcal{O}\left(\frac{1}{k}\right)}$, et même $e^{\mathcal{O}\left(\frac{1}{k^2}\right)}$ si la loi X est symétrique par rapport à l'origine.

En particulier, cela implique que $R_2(Z_k^m||\rho_1^m) = e^{\mathcal{O}\left(\frac{m}{k^2}\right)}$, et donc que si $k = \Theta(\sqrt{m})$, alors $R_2(Z_k^m||\rho_1^m) = e^{\mathcal{O}(1)}$. Ainsi, dans ce cas, la réduction présentée dans l'article [1] fonctionne. On obtient donc une réduction de LWE classique à LWE dont la gaussienne est remplacée par $Z_{\Theta(\sqrt{m})}$. Cela n'est cependant pas satisfaisant si l'objectif est de remplacer la gaussienne par une loi rapide à échantillonner.

Le théorème 1.3 de l'article s'adapte de la façon suivante pour calculer la distance statistique :

Théorème 5.4. Si les conditions du théorème précédent sont vérifiées, alors pour k assez grand, on a pour tout x réel l'inégalité $|Z_k(x) - \rho_1(x)| \leq \frac{c}{\sqrt{k}} e^{-\left(\frac{x}{2}\right)^2}$ pour une certaine constante c qui ne dépend que de X.

Ce théorème permet d'obtenir une majoration sur la distance statistique en intégrant l'inégalité :

$$\Delta(Z_k, \rho_1) = \int_{-\sqrt{k}}^{+\sqrt{k}} |Z_k(x) - \rho_1(x)| \, \mathrm{d}x = \mathcal{O}\left(\frac{1}{\sqrt{k}}\right).$$

5.4 Approximation de la gaussienne par une somme de trois uniformes

Une idée est de réaliser des approximations de ρ_σ une somme d'un petit nombre de lois uniformes, par exemple trois. Pour tenter de mettre en œuvre la réduction de l'article [1], on estime donc $R_2(U_\beta + T_\alpha||\rho_\sigma)$. La figure 9 représente à gauche des gaussiennes comparées à des sommes de trois lois uniformes avec plusieurs paramètres, et à droite les fonctions dont l'intégrale est la divergence de Rényi, c'est-à-dire les fonctions $x \mapsto \frac{(T_\alpha * U_\beta)^2(x)}{\rho_\sigma(x)}$. Les calculs des divergences dans le meilleurs cas, ainsi que le meilleur σ en fonction de β sont regroupés dans la figure 10.

Cette fois-ci, le calcul à effectuer est :

$$R_2(U_\beta + T_\alpha||\rho_\sigma) = \int_{-\beta-\frac{1}{\alpha}}^{\beta+\frac{1}{\alpha}} \frac{T_\alpha * U_\beta(x)^2}{\rho_\sigma(x)} \, \mathrm{d}x.$$

Pour obtenir la meilleure gaussienne qu'approxime $U_\beta + T_\alpha$, une piste peut-être de dériver l'expression ci-dessus par rapport à σ, et de tenter d'annuler l'expression obtenue :

$$\frac{\partial R_2(U_\beta + T_\alpha||\rho_\sigma)}{\partial \sigma} = \int_{-\beta-\frac{1}{\alpha}}^{\beta+\frac{1}{\alpha}} \frac{T_\alpha * U_\beta(x)^2}{\rho_\sigma(x)} \left(1 - \left(\frac{x}{\sigma}\right)^2\right) \mathrm{d}x.$$

De la même façon que dans l'article [1], on souhaite que $R_2((U_\beta + T_\alpha)^m||\rho_\sigma^m)$ soit majoré par un $n^{\mathcal{O}(1)}$.

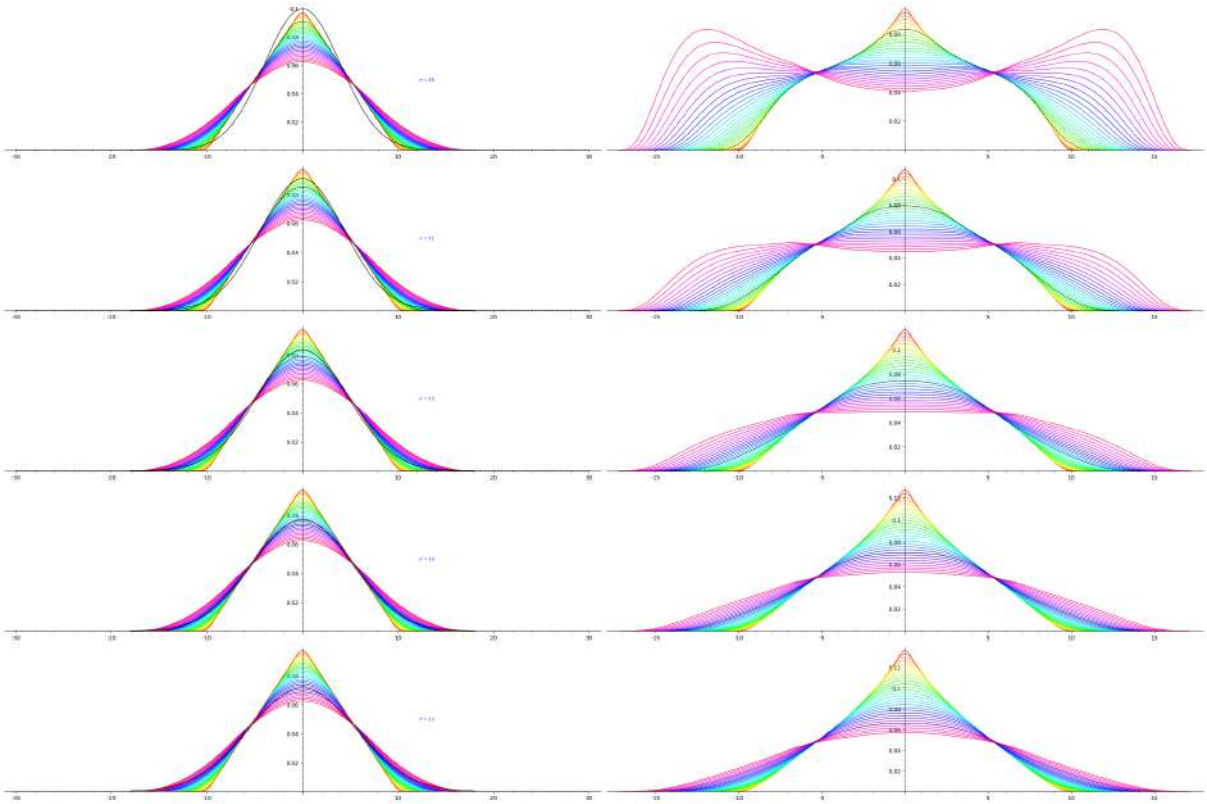
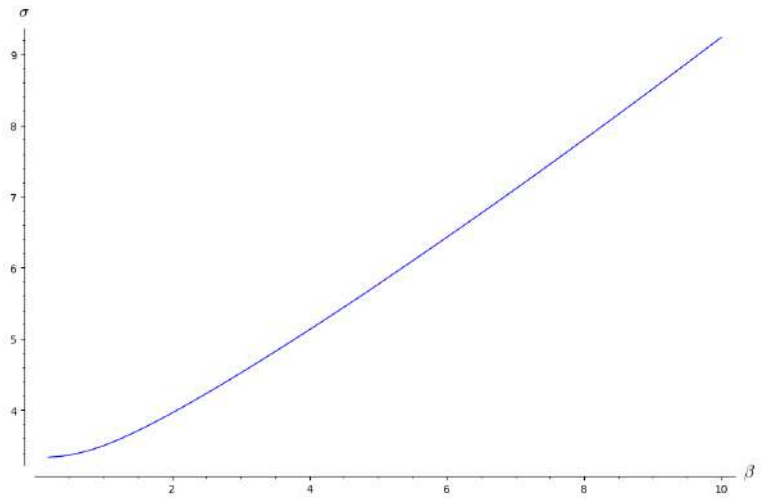


FIGURE 9 – A gauche, tracé de gaussiennes (en noir), et des lois correspondant à la somme de trois lois uniformes, c'est-à-dire $U_\beta + T_\alpha$ (en couleur). A droite, les fonctions dont l'intégrale est la divergence de Rényi. Les fonctions en pointillés sont celles qui réalisent la meilleure approximation des gaussiennes au sens de R_2.

6 Conclusion

L'étude de la cryptographie basée sur les réseaux relève de nombreux domaines mathématiques : algèbre, géométrie, analyse, probabilités, théorie des nombres... Nous avons établi quelques liens entres les problèmes fondamentaux liés aux réseaux euclidiens, le problème Learning With Errors, et le chiffrement de Regev, dont les variantes sont des candidats sérieux pour le chiffrement post-quantique. Nous avons aussi étudié la divergence de Rényi, qui permet de déterminer si une distribution standard peut être approximée par une distribution plus simple. En particulier, on constate numériquement qu'il est possible avec trois lois uniformes d'approximer une loi normale de sorte que la divergence soit relativement petite, soit de l'ordre de 1.009. Comme les spécifications de certains cryptosystèmes comme FrodoKEM annoncent des divergences beaucoup plus petites de 1.00003, il serait intéressant de poursuivre cette étude pour obtenir une expression explicite de $R_2(U_\beta + T_\alpha || \rho_\sigma)$ afin de déterminer le niveau de sécurité réellement atteint.



(a)

σ	β	α	$R_2(U_\beta + T_\alpha \| \rho_\sigma)$
10	1.750	0.1	1.03039
11	2.875	0.1	1.01297
12	4.125	0.1	1.00901
13	5.375	0.1	1.00865
14	6.375	0.1	1.00951

(b)

FIGURE 10 – En haut, le tracé du meilleur σ tel que $U_\beta + T_\alpha$ approxime ρ_σ, en fonction de β, α étant fixé. L'écart type σ semble croître de façon linéaire en β. En bas, quelques valeurs optimales de β pour quelques valeurs de σ, avec α fixé.

Références

[1] *Improved Security Proofs in Lattice-Based Cryptography : Using the Rényi Divergence Rather than the Statistical Distance*, S. Bai, T. Lepoint, A. Roux-Langlois et al., J Cryptol 31, 610–640, (2018)

[2] *Rényi Divergence and Kullback-Leibler Divergence*, T. van Erven, P.Harremoës

[3] *Solving Hard Lattice Problems and the Security of Lattice-Based Cryptosystems*, T. Laarhoven, J. van de Pol, B. de Weger, (2012)

[4] *One-dimensional Stable Distributions*, V.M. Zolotarev, American Mathematical Society, Volume 65, (1986)

[5] *Winter School on Cryptography : Proving Hardness of LWE*, [Vidéo], O. Regev, Bar-Ilan University (31 mai 2012). https://www.youtube.com/watch?v=Z4DM3qhH6pA

[6] *CRYSTALS-Kyber, Algorithm Specifications And Supporting Documentation*, (version 3.02), R. Avanzi, J. Bos, L. Ducas et al., https://pq-crystals.org/

[7] *Rényi Divergence and the Central Limit Theorem*, S. G. Bobkov, G. P. Chistyakov, F. Götze.