Comparison between Subfield and Straightforward Attacks on NTRU

Paul Kirchner[1] and Pierre-Alain Fouque[2]

[1]École normale supérieure & IRISA, paul.kirchner@ens.fr
[2]Université de Rennes 1 and Institut Universitaire de France,
pierre-alain.fouque@univ-rennes1.fr

July 19, 2016

Abstract

Recently in two independent papers, Albrecht, Bai and Ducas and Cheon, Jeong and Lee presented two very similar attacks, that allow to break NTRU with larger parameters and GGH Multinear Map without zero encodings. They proposed an algorithm for recovering the NTRU secret key given the public key which apply for large NTRU modulus, in particular to Fully Homomorphic Encryption schemes based on NTRU. Hopefully, these attacks do not endanger the security of the NTRUENCRYPT scheme, but shed new light on the hardness of this problem. The basic idea of both attacks relies on decreasing the dimension of the NTRU lattice using the multiplication matrix by the norm (resp. trace) of the public key in some subfield instead of the public key itself. Since the dimension of the subfield is smaller, the dimension of the lattice decreases, and lattice reduction algorithm will perform better.
Here, we revisit the attacks on NTRU and propose another variant that is simpler and outperforms both of these attacks in practice. It allows to break several concrete instances of YASHE, a NTRU-based FHE scheme, but it is not as efficient as the hybrid method of Howgrave-Graham on concrete parameters of NTRU. Instead of using the norm and trace, we propose to use the multiplication by the public key in some subring and show that this choice leads to better attacks. We can then show that for power of two cyclotomic fields, the time complexity is polynomial when $q = 2^{\Omega(\sqrt{n \log \log n})}$. Finally, we show that, under heuristics, straightforward lattice reduction is even more efficient, allowing to extend this result to fields without non-trivial subfields, such as NTRU Prime. We insist that the improvement on the analysis applies even for relatively small modulus ; though if the secret is sparse, it may not be the fastest attack. We also derive a tight estimation of security for (Ring-)LWE and NTRU assumptions.

1 Introduction

NTRU has been introduced by Hoffstein, Pipher and Silverman in [27] and has since resisted many attacks [15, 24, 23, 28]. In [30], Kirchner and Fouque describe a new subexponential-time attack on NTRU with complexity $2^{(n/2+o(n))/\log\log q}$, but the $o(n)$ is too large to lead to attack for practical parameters. To date, the most efficient attack on NTRU is the hybrid attack described by Howgrave-Graham in [28]. NTRU is one of the most attractive lattice-based cryptosystems since it is very efficient, and many Ring-LWE cryptosystems have a NTRU equivalent with better performance. For instance, Ducas, Lyubashevsky and Prest propose an Identity Based Encryption scheme based on NTRU [22] (albeit with a much larger standard deviation), López-Alt, Tromer and Vaikuntanathan describe a Fully Homomorphic Encryption scheme [33], which is improved in a scheme called YASHE [7, 32], and Ducas *et al.* propose a very fast signature scheme called BLISS [21].

The key recovery problem of NTRU is the following problem: given a public key $\mathbf{h} = \mathbf{f}/\mathbf{g}$ in some polynomial ring $\mathbb{Z}_q[X]/(X^n+1)$ for n prime, q a small integer and the euclidean norms of $\mathbf{f}, \mathbf{g}$ are small, recover $\mathbf{f}$ and $\mathbf{g}$ or a small multiple of them. In NTRUENCRYPT, $\mathbf{f}$ and $\mathbf{g}$ are two sparse polynomials of degrees $< n$ and coefficients $\{-1, 0, 1\}$. It is easy to see that the public key cannot be uniformly distributed in the whole ring, since the entropy is too small. In [42], Stehlé and Steinfeld, show that if $\mathbf{f}$ and $\mathbf{g}$ are generated using a Gaussian distribution of standard deviation $\sigma \approx q^{1/2}$, then the distribution of the public key is statistically indistinguishable from the uniform distribution, but in practice, such recommendation is never used since it has poor performance [10].

Related Work. At CRYPTO 2015, Kirchner and Fouque in [30] proposed a heuristic subexponential-time algorithm on NTRU in time $2^{O(n/\log\log q)}$ using a variant of the Blum, Kalai and Wasserman algorithm [6].

Recently, in [14, 1], Cheon, Jeong and Lee at ANST 2016 on the one hand and Albrecht, Bai and Ducas at CRYPTO 2016 on the other hand, described a new attack on NTRU-like cryptosystems. They use the fact that for cyclotomic number fields, there exist subfields that allow to reduce the dimension of the lattice. The attack recovers the norm of the secret key in this subfield, which is smaller than in the classical NTRU lattice. Consequently, the quality of the lattice reduction algorithm is important to find such small vectors compared to the reduction of NTRU lattice. The main drawback of their technique is that q has to be very large compared to n and we estimate asymptotically $q = 2^{\Omega(\sqrt{n\log\log n})}$ for a polynomial time complexity. This attack on NTRU with large parameters has been first discovered by Jonsson, Nguyen and Stern and was described in [25, Section 6].

Our Results. We show that using the multiplication matrix by the public key in a subring (which has the same size as the subfield), leads to more efficient attacks. In particular, we were able to attack concrete parameters proposed in YASHE based on overstretched NTRU [7, 8, 31, 17, 18, 16, 12, 32], meaning that we can recover a decryption key for smaller modulus q, compared to the previous approaches [1, 14]. The previous attacks use the norm over the subfield in [1] and the trace in [14]. It would be possible for instance to use all the coefficients of the characteristic polynomial and not two of them. Our attack using the subring is better than the two previous ones since in the same configuration, we can choose exactly the size of the subfield as the number of coordinates.

Contrary to [1, 14], we analyze precisely the running time of this attack and we derive tight bounds on the size of the norm. We do not rely in our analysis on the Hermite factor (or approximate factor) but instead we use a lemma due to Pataki and Tural on the volume of sublattices with high rank. This allows us to precisely predict the success probability of all lattice reduction algorithms against NTRU and indicates that reducing the original lattice, we obtain the same result. This lemma allows us to use the fact that in NTRU lattices, all the multiples of the secret key vector are short vectors.

We also make experiments to understand the behaviour of lattice reduction algorithm, which allows us to give some precise predictions when this attack will work. Finally, we show that the subfield attack is not more efficient than the straightforward lattice reduction and that this attack can also be used to break overstretched NTRU Prime scheme. We also provide a tight asymptotical security estimate of NTRU and LWE schemes.

Comparison with [1, 14]. In our work, we consider the lattice generated by $\begin{pmatrix} q\mathbf{I}_n & \mathbf{M}_{\mathbf{h}}^{\mathcal{O}_\mathbb{L}} \\ \mathbf{0} & \mathbf{I}_{n/r} \end{pmatrix}$ while Albrecht *et al.* for instance consider $\begin{pmatrix} q\mathbf{I}_{n/r} & \mathbf{M}_{\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{h})}^{\mathcal{O}_\mathbb{L}} \\ \mathbf{0} & \mathbf{I}_{n/r} \end{pmatrix}$, where $\mathbf{M}_{\mathbf{h}}^{\mathcal{O}_\mathbb{L}}$ represents the multiplication by the element $\mathbf{h}$ in the subring $\mathcal{O}_\mathbb{L}$ of $\mathbb{K}$. The running time of lattice reduction algorithms depends on the dimension of the matrix. That is the reason why we can work in a projected lattice and not on the full $(n+n/r, n+n/r)$-matrix. However, the second important parameter is the approximation factor. This parameter depends on the size of the Gram-Schmidt coefficients. If we use the logarithm of their size, these coefficients draw a decreasing line of slope correlated with the approximation factor. The smaller the approximation factor be, the more horizontal the line will be. However, if we have only a $2n/r$-dimensional matrix, the determinant

is too small to produce large Gram-Schmidt norms. This problem is bypassed with our approach since we can choose the number of coordinates and the size of the subfield.

Also, we show a tight estimation of the parameters broken by lattice reduction, and in particular that working in the original field works well. Experiments were conducted in an extensive way, and over much larger parameters.

2 Preliminaries

We work over a number field $\mathbb{K}$ of dimension n, which has a subfield $\mathbb{L}$ of dimension $m \mid n$. For simplicity, we assume that $\mathbb{K}$ is a Galois extension of $\mathbb{Q}$, with Galois group G ; and H is the subgroup of G fixing $\mathbb{L}$. It is a standard fact that $|H| = n/m$.

When $\mathbb{K} = \mathbb{Q}(X)/(P(X))$ for a monic irreducible polynomial $P(X)$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ its distinct complex roots, each embedding (ring homomorphism) $e_i : \mathbb{K} \to \mathbb{C}$ is the evaluation of $\mathbf{a} \in \mathbb{K}$ at the root α_i, i.e. $e_i : \mathbf{a} \mapsto \mathbf{a}(\alpha_i)$. If we have r real roots and $2s$ complex roots ($n = r + 2s$), we have $\mathbb{K} \otimes \mathbb{R} \simeq \mathbb{R}^r \times \mathbb{C}^s$ so that we can define a norm $\| \cdot \|$ over $\mathbb{K}$ as the canonical euclidean norm of $\mathbb{R}^r \times \mathbb{C}^s$ where the canonical embedding is defined as: $\sigma(\mathbf{x}) = (\sigma_1(\mathbf{x}), \ldots, \sigma_{r+s}(\mathbf{x})) \in \mathbb{R}^r \times \mathbb{C}^s$, where $\sigma_1, \ldots, \sigma_r$ are the real embeddings and $\sigma_{r+1}, \ldots, \sigma_n$ are the complex embeddings and σ_{r+j} is paired with its complex conjugate σ_{r+s+j}. The number field $\mathbb{K}$ is viewed as an euclidean $\mathbb{Q}$-vector space endowed with the inner product $\langle \mathbf{a}, \mathbf{b} \rangle = \sum_e e(\mathbf{a})\bar{e}(\mathbf{b})$ where e ranges over all the $r + 2s$ embeddings $\mathbb{K} \to \mathbb{C}$. This defines the euclidean norm denoted $\| \cdot \|$. Notice that elements of the Galois group permute or conjugate the coordinates in $\mathbb{R}^r \times \mathbb{C}^s$, and therefore the norm is invariant by elements of G:

$$\forall \sigma \in G, \|\sigma(\mathbf{x})\| = \|\mathbf{x}\|.$$

We call $\mathsf{N}_{\mathbb{K}/\mathbb{L}} : \mathbb{K} \to \mathbb{L}$ the relative norm, with $\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{a})$ the determinant of the $\mathbb{L}$-linear endomorphism $\mathbf{x} \mapsto \mathbf{a}\mathbf{x}$. It is known that we have :

$$\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{a}) = \prod_{\sigma \in H} \sigma(\mathbf{a}).$$

We can bound the norm using the inegality of arithmetic and geometric means :

$$|\mathsf{N}_{\mathbb{K}/\mathbb{Q}}(\mathbf{a})| \leq \left(\frac{\|\mathbf{a}\|}{\sqrt{n}} \right)^n$$

The operator norm for the euclidean norm is denoted $|||\cdot|||$ and is defined as $|||\mathbf{a}||| = \sup_{\mathbf{x} \in \mathbb{K}^*} \|\mathbf{a}\mathbf{x}\| / \|\mathbf{x}\|$. Remark that it is simply the maximum of the norm of the coordinates in $\mathbb{R}^r \times \mathbb{C}^s$. Also, it is sub-multiplicative and $\|\mathbf{x}\| \leq \sqrt{n}|||\mathbf{x}|||$.

Let $\mathcal{O}$ be an *order* of $\mathbb{K}$, that is $\mathcal{O} \subset \mathbb{K}$ and $\mathcal{O}$ is a commutative group which is isomorphic as an abelian group to $\mathbb{Z}^n$. We define $\mathcal{O}_\mathbb{L}$ as $\mathcal{O} \cap \mathbb{L}$, and is an order of $\mathbb{L}$. We denote by $\mathrm{Vol}(\mathcal{L})$ the volume of the lattice $\mathcal{L}$, which is the square root of the determinant of the Gram matrix corresponding to any basis of $\mathcal{L}$. We define Δ to be the square of the volume of $\mathcal{O}$, and likewise for $\Delta_\mathbb{L}$ with respect to $\mathcal{O}_\mathbb{L}$.

We definte

$$\mathbf{M}_\mathbf{a}^\mathcal{L} : \begin{array}{ccc} \mathcal{L} & \longrightarrow & \mathcal{O} \\ \mathbf{x} & \longmapsto & \mathbf{a}\mathbf{x} \end{array}$$

for any lattice $\mathcal{L} \subset \mathcal{O}$ and $\mathbf{a} \in \mathcal{O}$; and we also denote $\mathbf{M}_\mathbf{a}^\mathcal{L}$ the corresponding matrix for some basis of $\mathcal{L}$.

When $\mathbb{K}$ is a cyclotomic field [43], we have more precise results about the ring of integers. We define $\zeta_\mathfrak{f} = \exp(2i\pi/\mathfrak{f})$ and $\phi(\mathfrak{f})$ is the cardinal of $(\mathbb{Z}/\mathfrak{f}\mathbb{Z})^*$, and also the dimension of $\mathbb{Q}[\zeta_\mathfrak{f}]$. It is well known that

$$\mathrm{Vol}(\mathbb{Z}[\zeta_\mathfrak{f}])^2 = \frac{\mathfrak{f}^{\phi(\mathfrak{f})}}{\prod_{p|\mathfrak{f}} p^{\phi(\mathfrak{f})/(p-1)}}.$$

In particular, if $\mathfrak{f}$ is a power of two, $\mathrm{Vol}(\mathbb{Z}[\zeta_\mathfrak{f}]) = (\mathfrak{f}/2)^{\mathfrak{f}/4}$. In this case, we also have that $(\zeta_\mathfrak{f}^i)_{i=0}^{\mathfrak{f}/2-1}$ is an orthogonal basis for the norm $\|\cdot\|$.

The discrete Gaussian distribution over a lattice $\mathcal{L}$ is noted $D_{\mathcal{L},s}$, where the probability of sampling $\mathbf{x} \in \mathcal{L}$ is proportional to $\exp(-\pi\|\mathbf{x}\|^2/s^2)$. The continuous Gaussian distribution over $\mathbb{K}$ is noted D_s, and its density in $\mathbf{x}$ is proportional to $\exp(-\pi\|\mathbf{x}\|^2/s^2)$. We define

$$\rho_s(E) = \sum_{\mathbf{x}\in E} \exp(-\pi\|\mathbf{x}\|^2/s^2).$$

We will denote by $\mathbb{E}[X]$ the expectation of a random variable X.

We now prove a standard bound on ideal lattices, which indicates that they do not have very short vectors :

Lemma 1. *Let $M \subset (\mathbb{K}\otimes\mathbb{R})^d$ be an $\mathcal{O}$ module of rank 1. Then, for any $0 \neq \mathbf{v} \in M$, we have $\mathrm{Vol}(M) \leq \sqrt{\Delta}\|\mathbf{v}/\sqrt{n}\|^n$.*

Proof. Since we can build a $\mathbb{K}$-linear isometry from $\mathbb{R}\otimes M$ to $\mathbb{K}\otimes\mathbb{R}$, we can assume $d = 1$. Then,

$$\mathrm{Vol}(M) \leq \mathrm{Vol}(\mathbf{v}\mathcal{O}) = \mathsf{N}_{\mathbb{K}/\mathbb{Q}}(\mathbf{v})\sqrt{\Delta} \leq \|\mathbf{v}/\sqrt{n}\|^n\sqrt{\Delta}.$$

□

We recall Minkowski's theorem :

Theorem 1. *For any lattice $\mathcal{L}$ of dimension n, there exists $\mathbf{0} \neq \mathbf{x} \in \mathcal{L}$ with $\|\mathbf{x}\| \leq \sqrt{n}\mathrm{Vol}(\mathcal{L})^{1/n}$.*

3 Projection over a sub-ring

3.1 Description of the attack

We first make sure that $\mathcal{O}$ is stable by all elements of H. This can be done by computing the Hermite normal form of the concatenation of the basis of $\sigma(\mathcal{O})$ for all $\sigma \in H$. We may then call $\mathcal{O}$ the order generated by this matrix.

The attack consists in finding short vectors of the lattice generated by

$$\mathbf{A} = \begin{pmatrix} q\mathbf{I}_n & \mathbf{M}_\mathbf{h}^{\mathcal{O}_\mathbb{L}} \\ \mathbf{0} & \mathbf{I}_m \end{pmatrix}$$

by using lattice reduction. We recall that $\mathbf{h}$ is the public key, so that a basis of this lattice can be built. We want to show that $\begin{pmatrix} \mathbf{f}\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})/\mathbf{g} \\ \mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g}) \end{pmatrix}$ is a short vector of this lattice.

The quadratic form we reduce is actually the one induced by $\|\cdot\|$, i.e. $\|(\mathbf{x},\mathbf{y})\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$, on this lattice.

Lemma 2. *For any $\mathbf{g} \in \mathcal{O}$, we have*

$$\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g}) \in \mathbf{g}\mathcal{O} \cap \mathcal{O}_\mathbb{L}.$$

Proof. We have

$$\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g}) = \mathbf{g}\prod_{\sigma\in H-\{1\}} \sigma(\mathbf{g})$$

so that $\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g}) \in \mathbf{g}\mathcal{O}$. By definition of $\mathsf{N}_{\mathbb{K}/\mathbb{L}}$, we have $\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g}) \in \mathbb{L}$. Therefore, $\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g}) \in \mathbf{g}\mathcal{O} \cap \mathcal{O}_\mathbb{L}$. □

We now recall two results from [36] and Banaszczyk's lemma [4] about discrete gaussian sampling over a lattice. For completeness, the proofs are in appendix.

Lemma 3. *Given a lattice $\mathcal{L} \subset \mathbb{R}^n$, for any s and $\mathbf{c} \in \mathbb{R}^n$, we have*

$$\rho_s(\mathcal{L} + \mathbf{c}) \leq \rho_s(\mathcal{L}).$$

Lemma 4. *For a lattice $\mathcal{L}$, any $t \geq 1$, the probability that $\mathbf{x}$ sampled according to $D_{\Lambda,s}$ verifies $\|\mathbf{x}\| > st\sqrt{\frac{n}{2\pi}}$ is at most*

$$\exp\big(-n(t-1)^2/2\big).$$

We now show that integers sampled from a discrete Gaussian distribution behaves in a way similar to a continuous Gaussian distribution.

Lemma 5. *Let $\mathbf{x}$ be sampled according to $D_{\mathcal{O},s}$. Then, the probability that*

$$|||\mathbf{x}||| \geq s\sqrt{2\ln(2n/\epsilon)/\pi}$$

is at most ϵ.

Proof. Let $\mathbf{u}$ be a unit vector, i.e. $\|\mathbf{u}\| = 1$. Then,

$$\begin{aligned}\rho_s(\mathcal{O})\mathbb{E}[\exp(2\pi t\langle \mathbf{x}, \mathbf{u}\rangle/s^2)] &= \sum_{\mathbf{x}\in\mathcal{O}} \exp(-\pi(\langle \mathbf{x}, \mathbf{x}\rangle - 2\langle \mathbf{x}, t\mathbf{u}\rangle)/s^2) \\ &= \exp(\pi t^2/s^2)\sum_{\mathbf{x}\in\mathcal{O}} \exp(-\pi\|\mathbf{x} - t\mathbf{u}\|^2/s^2) \\ &= \exp(\pi t^2/s^2)\rho_s(\mathcal{O} - t\mathbf{u}).\end{aligned}$$

We deduce with the previous lemma

$$\mathbb{E}[\exp(2\pi t\langle \mathbf{x}, \mathbf{u}\rangle/s^2)] \leq \exp(\pi t^2/s^2).$$

Using Markov's inequality and the union bound with $-\mathbf{u}$, we have that the probability of $|\langle \mathbf{x}, \mathbf{u}\rangle| \geq t$ is at most $2\exp(-\pi t^2/s^2)$.

We now use $t = s\sqrt{\ln(2n/\epsilon)/\pi}$, so that the probability of any real or imaginary part of a coordinate of $\mathbf{x}$ in $\mathbb{R}^r\mathbb{C}^s$ is larger than

$$s\sqrt{\ln(2n/\epsilon)/\pi}$$

is at most ϵ. □

Theorem 2. *Let $\mathbf{f}$ be sampled according to $D_{\mathcal{O},\sigma}$, $\mathbf{g}$ according to $D_{\mathcal{O},s}$ and set $\mathbf{h} = \mathbf{f}/\mathbf{g}$. Assume $\mathbf{h}$ is well defined, except with probability at most $\epsilon/3$. Then, there exists $\mathbf{x} \neq \mathbf{0}$ where $\mathbf{x}$ is an integer vector, such that*

$$\|\mathbf{A}\mathbf{x}\| \leq \sqrt{n(1+\sigma^2/s^2)}(s\sqrt{2\ln(6n/\epsilon)/\pi})^{n/m}$$

except with probability at most ϵ.

Proof. With probability at least $1 - \epsilon$, we have

$$|||\mathbf{f}||| \leq \sigma\sqrt{2\ln(6n/\epsilon)/\pi}$$

and

$$|||\mathbf{g}||| \leq s\sqrt{2\ln(6n/\epsilon)/\pi}.$$

In this case, we consider $\mathbf{y}$ such that $\mathbf{h}\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g}) + q\mathbf{y} = \mathbf{f}\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})/\mathbf{g}$ and consider

$$\mathbf{x} = \begin{pmatrix} \mathbf{y} \\ \mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g}) \end{pmatrix}.$$

Using the multiplicativity of operator norms, we have

$$|||\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})||| \leq \left(s\sqrt{2\ln(6n/\epsilon)/\pi}\right)^{|H|}$$

and

$$|||\mathbf{f}\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})/\mathbf{g}||| \leq \sigma/s\left(s\sqrt{2\ln(6n/\epsilon)/\pi}\right)^{|H|}.$$

Finally,

$$\|\mathbf{A}\mathbf{x}\|^2 = \|\mathbf{f}\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})/\mathbf{g}\|^2 + \|\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})\|^2 \leq n\big(|||\mathbf{f}\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})/\mathbf{g}|||^2 + |||\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})|||^2\big).$$

□

We now try to get rid of the factor $\Theta(\ln(6n/\epsilon))^{n/2m}$ which is significant when s is small and n/m is large. To do so, we *h*euristically assume that $D_{\mathcal{O},\sigma}$ has properties similar to a *continuous* Gaussian here.

Theorem 3. *Let* $\mathbf{f}$ *be sampled according to* D_s *and* $E \subset G$*. Then, except with probability at most* ϵ *and under heuristics, we have :*

$$|||\prod_{\sigma\in E}\sigma(\mathbf{f})||| \leq \Theta(s)^{|E|}\exp\left(\Theta(\sqrt{|E|\log(n/\epsilon)})\right)$$

under the condition $|E| = \Omega\big(\log(n/\epsilon)\log^2(\log(n/\epsilon))\big)$

Proof. Let X be a random variable over $\mathbb{R}^+$, with a probability density function proportional to $\exp(-\pi x^2/s^2)$; and $Y = \sqrt{X_0^2 + X_1^2}$ where X_0 and X_1 are independent copies of X.

We have $\mathbb{E}[\log(X)] = \log(s) + \Theta(1)$ and $\mathrm{Var}[\log(X)] = \Theta(1)$ and $\log(X) < \log(s) + \Theta(\log(\log(n/\epsilon)))$ except with probability $\epsilon/(2n^2)$, due to standard bounds on Gaussian tails. Also, the same is true for Y.

We can now use the one-sided version of Bernstein's inequality [9, Theorem 3] : for Z the average of $|E|$ independent copies of $\log(X)$ or $\log(Y)$, we have :

$$\Pr[Z > t + \log(s)] \leq \epsilon/(2n) + \exp\left(-\frac{|E|t^2}{2(\Theta(1) + \Theta(\log(\log(n/\epsilon)))t/3)}\right).$$

We then choose some $t = \Theta\big(\sqrt{\log(n/\epsilon)/|E|}\big)$, so that with our lower bound on $|E|$, this probability is at most ϵ/n.

The result follows from the union bound over the coordinates in the canonical embedding of $\prod_{\sigma\in E}\sigma(\mathbf{f})$.
□

For some parameters, the norm may not be the shortest element, as demonstrated by the following theorem.

Theorem 4. *There exists an element* $\mathbf{v} \in \mathbf{g}\mathcal{O} \cap \mathcal{O}_{\mathbb{L}}$ *with*

$$0 < \|\mathbf{v}\| \leq \sqrt{m}\Delta^{1/2n}\sigma^{n/m}$$

with probability $1 - 2^{-\Omega(n)}$.

Proof. We use Banaszczyk's lemma with $t = 2$, so that $\|\mathbf{g}\| \leq \sigma\sqrt{2n/\pi}$ except with probability $\exp(-n/2)$. Then, the determinant of $\mathbf{v} \in \mathbf{g}\mathcal{O} \cap \mathcal{O}_{\mathbb{L}}$ is smaller than the determinant of $\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})\mathcal{O}_{\mathbb{L}}$, which is $\mathsf{N}_{\mathbb{K}/\mathbb{Q}}(\mathbf{g})\sqrt{\Delta_{\mathbb{L}}}$. But we have $\mathsf{N}_{\mathbb{K}/\mathbb{Q}}(\mathbf{g}) \leq \big(\frac{\|\mathbf{g}\|}{\sqrt{n}}\big)^n$ and $\Delta_{\mathbb{L}} \leq \Delta^{m/n}$ so we conclude with Minkowski's theorem. □

This implies that for most parameters, the norm of the shortest non-zero vector is around $O(\sigma)^{n/m}$. Since this is smaller than the previous value as soon as n/m is a bit large, it explains why [1] found vectors shorter than the solution.

3.2 Asymptotical analysis for power of two cyclotomic fields

We set here $\mathbb{K} = \mathbb{Q}[X]/(X^n+1) \simeq \mathbb{Q}[\zeta_{2n}]$ for n a power of two, and $\mathcal{O} = \mathbb{Z}[X]/(X^n+1) \simeq \mathbb{Z}[\zeta_{2n}]$ which is popular in cryptosystems. For some $r \mid n$ (any such r works), we select $\mathbb{L} = \mathbb{Q}[X^r]$ so that $\mathcal{O}_\mathbb{L} = \mathbb{Z}[X^r]$ and $|H| = r$, so that m, the dimension of $\mathbb{L}$ is n/r. Since the $\mathbf{X}^i$ forms an orthogonal basis, we have that the coordinates of $\mathbf{f}$ and $\mathbf{g}$ are independent discrete Gaussians of parameter $s/\sqrt{n}$. Also, we can directly reduce the lattice generated by $\mathbf{A}$ with the canonical quadratic form.

We restrict our study to power of two cyclotomic fields because $\mathcal{O}$ has a known orthogonal basis, so that we can derive a closed-form expression of the results. In more complicated cases, it is clear that we can deduce the result using a polynomial time algorithm.

Theorem 5. *Given a lattice $\mathcal{L}$ of dimension k, we can find a non-zero vector in $\mathcal{L}$ of norm less than $\beta^{k/\beta}\mathrm{Vol}(\mathcal{L})^{1/k}$ in deterministic time smaller than $2^{O(\beta)}$ times the size of the description of $\mathcal{L}$, for any $\beta < n/2$. With b_i^* the Gram-Schmidt norms of the output basis, we have $b_i^*/b_j^* \leq \beta^{O((j-i)/\beta + \log\beta)}$. Furthermore, the maximum of the Gram-Schmidt norms of the output basis is at most the maximum of the Gram-Schmidt norms of the input basis.*

Proof. Combine the semi-block Korkin-Zolotarev reduction [41] and the efficient deterministic shortest vector algorithm [37] with block size $\Theta(\beta)$ for the first point. Schnorr's algorithm combines the use of LLL reduction on a (possibly) linearly dependent basis, which is known to not increase the maximum of the Gram-Schmidt norms, and the insertion of a vector in position i whose projected norm is less than b_i^*. Also, the b_i^* decrease by a factor of at most $\beta^{O(\log\beta)}$ in a block, and the first Gram-Schmidt norms of blocks decrease by a factor of at most $\beta^{O(\beta)}$. □

For the rest of this section, we assume that when the previous algorithm is used on our orthogonal projection of $\mathbf{A}\mathbb{Z}^{n+m}$, and finds a vector shorter than $\sqrt{n}\mathrm{Vol}(\mathcal{L})^{1/k}$ (which is about the size of the shortest vector of a random lattice), then it must be a short multiple of the key. This assumption is backed by all experiments in the literature, including ours, and can be justified by the fact that decisional problems over lattices are usually as hard as their search counterpart (see [35] for example).

We also assume the size of the input is in $n^{O(1)}$, which is the usual case.

Theorem 6. *Let $nB^2 = \|\mathbf{f}\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})/\mathbf{g}\|^2 + \|\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})\|^2$. Assume $\frac{\log(qB)}{\log(q/B)} \leq r$. Then, for*

$$\frac{\beta}{\log\beta} = \frac{2m\log q}{\log(q/B)^2}$$

we can find a non-zero element $\mathbf{Ax}$ such that $\|\mathbf{Ax}\|^2 = O(nB^2)$ in time $2^{O(\beta+\log n)}$.

Proof. We extract the last $d \approx m\frac{\log(q^2)}{\log(q/B)} \leq n+m$ rows and columns of

$$\mathbf{A} = \begin{pmatrix} q\mathbf{I} & \mathbf{M}_\mathbf{h}^{\mathcal{O}_\mathbb{L}} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

and call the generated lattice $\mathcal{L}$. Note that it is the lattice generated by $\mathbf{A}$ projected orthogonally to the first columns, so that it contains a non-zero vector $\mathbf{y}$ such that $\|\mathbf{y}\|^2 \leq nB^2$. Then, we can compute the needed β

by

$$\begin{aligned}\frac{1}{d}\log\left(\frac{\sqrt{n}\mathrm{Vol}(\mathcal{L})^{1/d}}{\sqrt{n}B}\right) =& \frac{d-m}{d^2}\log(q) - \frac{1}{d}\log(B)\\ \approx& \frac{\log(q/B)}{m\log(q^2)}\left(\frac{\log(qB)\log(q/B)\log(q)}{\log(q/B)\log(q^2)} - \log(B)\right)\\ =& \frac{\log(q/B)}{m\log(q^2)}\left(\frac{\log(qB)}{2} - \log(B)\right)\\ =& \frac{\log^2(q/B)}{2m\log(q)}.\end{aligned}$$

The previous theorem indicates we can recover a short vector $\mathbf{z} \neq \mathbf{0}$ in $\mathcal{L}$ with $\|\mathbf{z}\| \leq nB^2$ in time $2^{\Theta(\beta+\log n)}$, and our assumption implies it is in fact a short vector in $\mathbf{A}\mathbb{Z}^{n+m}$. □

Notice that for $B \leq q$, a necessary condition for the problem to be solvable, we have $d \geq 2m$. It implies that the optimal dimension d cannot be reached by previous algorithms.

Theorem 7. *Let* $\mathbf{f}$ *and* $\mathbf{g}$ *be sampled according to* $D_{\mathcal{O},\sigma}$*, and* $\mathbf{h} = \mathbf{f}/\mathbf{g}$ *mod* q *which is well defined with probability at least* $1-\epsilon$*. Assume* $\sigma = n^{\Omega(1)}$ *and* $\sigma < q^{1/4}$*. Then, we can recover a non-zero multiple of* $(\mathbf{f}, \mathbf{g})$ *of norm at most* $\sqrt{q}$ *in time*

$$\exp\left(O\left(\max\left(\log n, \frac{n\log\sigma}{\log^2 q}\log\left(\frac{n\log\sigma}{\log^2 q}\right)\right)\right)\right)$$

with a probability of failure of at most $\epsilon + 2^{-n}$*.*

This is polynomial time for

$$\log\sigma = O\left(\frac{\log^2 q\log n}{n\log\log n}\right).$$

Proof. We choose $m = \Theta(\max(1, \frac{n\log\sigma}{\log q})) \leq n$ so that we can set $B = \sqrt{q}$, except with probability ϵ. The corresponding β is given by

$$\frac{\beta}{\log\beta} = \frac{2m\log q}{\log(q/B)^2} = \Theta(m/\log(q)) = \Theta\left(\frac{n\log\sigma}{\log^2 q}\right).$$

□

If we use $\log\sigma = \Theta(\log n)$ as in many applications, we are in polynomial time when

$$q = 2^{\Omega(\sqrt{n\log\log n})}.$$

If $\sigma = \Theta(\sqrt{n})$, the best generic algorithm runs in time $2^{\Theta(n/\log\log q)}$, which is slower for any $q \geq n^{\Theta(\sqrt{\log\log n})}$.

4 Precise prediction, simplification and generalization

We now show how to predict when this attack will work, and compare our theoretical analysis with experiments.

The analysis hinges on the fact that the difficulty for lattice reduction to find a vector in a sublattice of low volume depends on the rank of the sublattice. Previous analysis relied on its special case where the rank is one, so that the volume is the length of the generator.

Pataki and Tural [39] proved that the volume of the sublattice generated by r vectors is larger than the product of the r smallest Gram-Schmidt norms. We now prove the quadratic form version of this result, and study its consequences.

Lemma 6. *Let* $\mathbf{A} = \mathbf{B} + \mathbf{C} \in M_{n,m}(\mathbb{R})$ *with* $\mathbf{B}^t\mathbf{C}$ *a strictly upper triangular matrix, and* $\mathbf{B}^t\mathbf{B}$ *a diagonal matrix. Then* $\det(\mathbf{A}^t\mathbf{A}) \geq \det(\mathbf{B}^t\mathbf{B})$.

Proof. We prove the result by induction. If $n = 1$, we have $\det(\mathbf{A}^t\mathbf{A}) = (\mathbf{B}^t + \mathbf{C}^t)(\mathbf{B} + \mathbf{C}) = \mathbf{B}^t\mathbf{B} + \mathbf{C}^t\mathbf{C} \geq \mathbf{B}^t\mathbf{B}$. Else, we let $\begin{pmatrix}\mathbf{A} & \mathbf{a}\end{pmatrix} = \begin{pmatrix}\mathbf{B} & \mathbf{b}\end{pmatrix} + \begin{pmatrix}\mathbf{C} & \mathbf{c}\end{pmatrix}$, and we have $\mathbf{B}^t\mathbf{b} = \mathbf{C}^t\mathbf{b} = \mathbf{0}$ and $\mathbf{b}^t\mathbf{c} = 0$. Thus :

$$\begin{aligned}\det\left(\begin{pmatrix}\mathbf{A}^t \\ \mathbf{b}^t X + \mathbf{c}^t\end{pmatrix}\begin{pmatrix}\mathbf{A} & \mathbf{b}X + \mathbf{c}\end{pmatrix}\right) &= \det\begin{pmatrix}\mathbf{A}^t\mathbf{A} & \mathbf{B}^t\mathbf{c} + \mathbf{C}^t\mathbf{c} \\ \mathbf{c}^t\mathbf{B} + \mathbf{c}^t\mathbf{C} & \mathbf{b}^t\mathbf{b}X^2 + \mathbf{c}^t\mathbf{c}\end{pmatrix} \\ &= \det\begin{pmatrix}\mathbf{A}^t\mathbf{A} & \mathbf{B}^t\mathbf{c} + \mathbf{C}^t\mathbf{c} \\ \mathbf{c}^t\mathbf{B} + \mathbf{c}^t\mathbf{C} & \mathbf{c}^t\mathbf{c}\end{pmatrix} + \mathbf{b}^t\mathbf{b}X^2\det(\mathbf{A}^t\mathbf{A})\end{aligned}$$

and this value is non-negative for all X, since it is the determinant of a positive semi-definite matrix (for any $\mathbf{D}$, $\mathbf{x}^t\mathbf{D}^t\mathbf{D}\mathbf{x} = \langle \mathbf{D}\mathbf{x}, \mathbf{D}\mathbf{x}\rangle \geq 0$ so that $\mathbf{D}^t\mathbf{D}$ is positive semi-definite). We deduce

$$\det\left(\begin{pmatrix}\mathbf{A}^t \\ \mathbf{a}^t\end{pmatrix}\begin{pmatrix}\mathbf{A} & \mathbf{a}\end{pmatrix}\right) \geq \mathbf{b}^t\mathbf{b}\det(\mathbf{A}^t\mathbf{A})$$

and the result follows. □

Lemma 7. *Let* $\mathbf{G} = \mathbf{L}^t\mathbf{L} \in M_{n,n}(\mathbb{R})$ *the Cholesky decomposition of the positive-definite matrix* $\mathbf{G}$*, so that* $\mathbf{L}$ *is upper-triangular. For any* $\mathbf{U} \in M_{n,r}(\mathbb{Z})$ *of rank* $r \leq n$*, we have*

$$\det(\mathbf{U}^t\mathbf{G}\mathbf{U}) \geq \min_{0 \leq t_0 < \cdots < t_{r-1} < n} \prod_i L_{t_i,t_i}^2.$$

Proof. Due to the existence of the Hermite normal form[1], there exists $\mathbf{P} \in GL_r(\mathbb{Z})$ such that for some $0 \leq t_0 < \cdots < t_{r-1} < n$, we have for all j $(UP)_{t_j,j} \geq 1$, and for all $i > t_j$, $(UP)_{i,j} = 0$.

We now let $\mathbf{A} = \mathbf{L}\mathbf{U}\mathbf{P}$ and $\mathbf{B}$ with $B_{i,j}$ equal to $L_{i,i}(UP)_{i,j}$ if $i = t_j$, and 0 else. We then have that $\mathbf{B}^t(\mathbf{A} - \mathbf{B})$ is a strictly upper triangular matrix and $\mathbf{B}^t\mathbf{B}$ is diagonal so using the previous lemma :

$$\det(\mathbf{U}^t\mathbf{G}\mathbf{U}) = \det(\mathbf{P}^t\mathbf{U}^t\mathbf{G}\mathbf{U}\mathbf{P}) \geq \prod_j L_{t_j,t_j}^2.$$

□

We now show that using the previous lemma and the above heuristics, we can actually achieve the same efficiency regardless of the presence of a subfield, as long as we know an orthogonal basis of $\mathcal{O}$.

The following theorem, identical to [1, Theorem 2], indicates that short vectors are multiples of the secret key.

Theorem 8. *Let* $\mathbf{f}, \mathbf{g} \in \mathcal{O}$ *with* $\mathbf{g}$ *invertible modulo* q *and* $\mathbf{f}$ *coprime to* $\mathbf{g}$*. Then, any vector shorter than* $\frac{nq}{\|(\mathbf{f},\mathbf{g})\|}$ *in*

$$\begin{pmatrix}q\mathbf{I}_n & \mathbf{M}_{\mathbf{f}/\mathbf{g}}^{\mathcal{O}} \\ 0 & \mathbf{I}_n\end{pmatrix}\mathcal{O}^2$$

is in $\begin{pmatrix}\mathbf{f} \\ \mathbf{g}\end{pmatrix}\mathcal{O}$.

[1] They are generally defined with an other order on the columns and rows, but this is irrelevant.

Proof. By coprimality, there exists $\mathbf{F}, \mathbf{G}$ such that $\mathbf{fG} - \mathbf{gF} = q$. Then,

$$\begin{pmatrix} \mathbf{f} & \mathbf{F} \\ \mathbf{g} & \mathbf{G} \end{pmatrix}$$

generates the same lattice. We let $\Lambda = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix} \mathcal{O} \subset (\mathbb{R} \otimes \mathbb{K})^2$ and Λ^* the projection of $\begin{pmatrix} \mathbf{F} \\ \mathbf{G} \end{pmatrix} \mathcal{O}$ orthogonally to Λ. We have $\mathrm{Vol}(\Lambda)\mathrm{Vol}(\Lambda^*) = q^n\Delta$. Finally, let $0 \neq \mathbf{x} \in \Lambda^*$. Using twice Lemma 1, we have

$$\|\mathbf{x}/\sqrt{n}\|^n \geq \frac{q^n\Delta}{\sqrt{\Delta}\mathrm{Vol}(\Lambda^*)} = \left(\frac{q\sqrt{n}}{\|(\mathbf{f},\mathbf{g})\|}\right)^n.$$

□

Theorem 9. *Let $\mathbf{f}, \mathbf{g}$ sampled according $D_{\mathcal{O},s}$ such that $\mathbf{g}$ is invertible with probability $1-\epsilon$, and an orthogonal basis of $\mathcal{O}$ is known. Reducing the lattice generated by $\begin{pmatrix} q\mathbf{I}_n & \mathbf{M}_{\mathbf{h}}^{\mathcal{O}_\mathbb{K}} \\ \mathbf{0} & \mathbf{I}_n \end{pmatrix}$ using the algorithm of Theorem 5, assuming the minimum of the Gram-Schmidt norms does not decrease, with*

$$\beta = \Theta\left(\frac{n\log\sigma}{\log^2 q}\log\left(\frac{n\log\sigma}{\log^2 q}\right)\right),$$

we can recover at least $n/2$ vectors of a basis of $\mathbf{f}\mathcal{O}$ and $\mathbf{g}\mathcal{O}$, when $\Delta^{1/2n}\sigma = q^{O(1)}$ and

$$\log q = \Omega\left(\log^2\left(\frac{n\log\sigma}{\log q}\right)\right),$$

with probability $1 - \epsilon - 2^{-\Omega(n)}$.

Proof. At the beginning of the algorithm, the Gram-Schmidt norms are $q\Delta^{1/2n}$ for the first n vectors, and $\Delta^{1/2n}$ for the next n vectors. The lattice contains $\begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix} \mathcal{O}$ so that the lattice spanned has a volume of $\sigma^n\sqrt{\Delta}$ except with probability $2^{-\Omega(n)}$, thanks to Lemma 4.

We consider now the basis outputted by the reduction algorithm (Theorem 5), and call b_i^* 'small' when it is amongst the n smallest Gram-Schmidt norms, and else, 'large'. Let $\ell = \frac{n\log\sigma}{\log q} \leq n/2$. Assume first that there is a $b_i^* \geq \frac{\sqrt{n}q}{\sigma} \geq q^{1/4}\Delta^{1/2n}$ with $i \leq n/2$. Suppose then that there is a $b_j^* \geq q^{1/4}\Delta^{1/2n}$ which is small. Then,

$$b_k^* \geq q^{1/4}\Delta^{1/2n}\beta^{-O(\ell/\beta+\log\beta)} \geq q^{1/4}\Delta^{1/2n}\beta^{-O(\ell/\beta)}$$

for all the ℓ first $k \geq i$ such that b_k^* is small. Hence, the product of the n smallest b_i^* is at least $\Delta^{1/2}q^{\ell/4}\beta^{-O(\ell^2/\beta)}$. We deduce that for small enough constants, this is impossible.

Else, we let $j \geq i$ be the smallest such that b_j^* is small. Then, we have

$$b_k^* \leq q^{1/4}\Delta^{1/2n}\beta^{O(\ell/\beta+\log\beta)}$$

for all the last ℓ $k \leq j$ such that b_k^* is large. Thus, the product of the large Gram-Schmidt norms is at most $\Delta^{1/2}q^{n-\ell/4}\beta^{O(\ell^2/\beta)}$, so that the product of the small Gram-Schmidt norms is at least $\Delta^{1/2}q^{\ell/4}\beta^{-O(\ell^2/\beta)}$, which is impossible.

Finally, the previous lemma implies that all the first $n/2$ vectors are in $\begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix} \mathcal{O}$. □

As we can see from the formula, considering a subfield is not helpful since the quantity $n \log \sigma$ is essentially constant ; unless we have reasons to believe there are huge factors of $\mathbf{g}\mathcal{O}$ which are in the subfield. Even worse, it actually decreases the efficiency when $\sigma \geq \sqrt{q}$ because the value of ℓ is forced to a suboptimal $2n$. We also observe that the significant reduction in the dimension due to the use of subfields, allowing to break instances of high dimension is also present here : indeed, we can project orthogonally to the first $n - \ell/2$ vectors the next ℓ vectors so that we reduce a lattice of dimension ℓ instead of $2n$.

Also, when we choose to work with $\mathcal{O} = \mathbb{Z}[X]/(X^n - X - 1)$ as in NTRU Prime [5], where we can use $(X^i)_{i=0}^{n-1}$ as an orthogonal basis due to the choice of the error distribution made by the authors (the coordinates are almost independent and uniform over $\{-1, 0, 1\}$), the same result applies.

We stress that while our theorem does not prove much - assuming the maximum of the Gram-Schmidt norms decreases is wrong, except for LLL - experiments indicate that either the middle part of the lattice behaves as a 'random' lattice, or the first n vectors are a basis of $\begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix} \mathcal{O}$. Furthermore, the phase transition between the two possible outputs is almost given by the impossibility of the first case. As lattice reduction algorithms are well understood (see [26, 13]), it is thus easy to compute the actual β.

5 Implementation

Heuristically, we have that for reduced random lattices, the sequence $L_{i,i}$ is (mostly) decreasing and therefore the relevant quantity is $\prod_{i=n-r}^{n-1} L_{i,i}$. It means that when the $L_{i,i}$s decrease geometrically and $\det(\mathbf{U}^t\mathbf{G}\mathbf{U})^{1/r}$ is about the squared length of the shortest vector, we need $L_{\lfloor n-r/2 \rfloor, \lfloor n-r/2 \rfloor}$ to be larger than the shortest vector instead of the $L_{n-1,n-1}$ given by a standard analysis. We now remark that for $r = 1$, this is pessimistic. Indeed, for a "random" short vector, we expect the projection to reduce its length by a $\simeq \sqrt{n}$ factor. In our case, we can expect the projection to reduce the length by a $\simeq \sqrt{n/(n-r)}$ factor.

For our predictions, we assumed that the determinant of the quadratic form

$$\mathbf{x} \mapsto \mathbf{f}\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})/\mathbf{g}\mathbf{x}\overline{\mathbf{f}\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})/\mathbf{g}\mathbf{x}} + \mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})\mathbf{x}\overline{\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})\mathbf{x}},$$

which corresponds to the $\det(\mathbf{U}^t\mathbf{G}\mathbf{U})$ above, is about the square of the norm over $\mathbb{Z}$ of $\mathbf{g}$. This quantity can be evaluated in quasi-linear time when we work within a cyclotomic field with smooth conductor by repeatedly computing the norm over a subfield, instead of the generic quadratic algorithm, or its variants such as in [2, section 5.2]. We observe a very good agreement between the experiments and the prediction, while considering only the fact that the lattice has a short vector would lead a much higher bound. Also, while $\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g})$ has a predicted size of $n^{r/2}\exp(\sqrt{r\log(n/r)})$ with $\sigma = \sqrt{n}$, we expect LLL to find a multiple of size $n^{r/2}\exp(n/r)$ (possibly smaller) but none of these quantities are actually relevant for determining whether or not LLL will recover a short element.

Finally, we may have $(\mathsf{N}_{\mathbb{K}/\mathbb{L}}(\mathbf{g}))/((\mathbf{g}) \cap \mathcal{O}_{\mathbb{L}})$ which is non-trivial. However, if it is an ideal of norm κ, we have that κ^2 divides the norm over $\mathbb{Z}$ of $\mathbf{g}$, which is exceedingly unlikely for even small values of $\kappa^{r/n}$.

Our predictions indicate all proposed parameters of [7, Table 1] are broken by LLL. We broke the first three using fplll and about three weeks of computation. The last three where broken in a few days over a 16-core processor (Intel Xeon E5-2650) using a new algorithm ; while the asymptotical complexity of fplll predicts a running time of several years.

The parameters proposed for schemes using similar overstretched NTRU assumption, such as in homomorphic encryption [8, 31, 17, 18, 16, 12, 32, 20] or in private information retrieval [19], are also broken in practical time using LLL. For example, we recovered a decryption key of the FHE described in [17] in only 10 hours. For comparison, they evaluated AES in 29 h: that means that we can more efficiently than the FHE evalution, recover the secret, perform the AES evaluation, and then re-encrypt the result! A decryption key was recovered for [20] in 4 h. Other instanciations such as [11, 29] are harder, but within range of practical cryptanalysis, using BKZ with moderate block-size [13].

$\log n$	$\log q$	$\log r$	Success	Method	Coordinates used	Origin
11	165	4	Yes	[1]	128	-
11	115	4	Yes	Ours	510	-
11	114	4	No	Ours	630	-
11	95	3	Yes	[1]	256	-
11	81	3	Yes	Ours	600	-
11	80	3	No	Ours	600	-
11	79	3	No	Ours	860	YASHE[7]
11	70	2	Yes	Ours	600	-
11	69	2	No	Ours	600	-
12	190	4	Yes	[1]	256	-
12	157	4	Yes	Ours	430	YASHE[7]
12	144	4	Yes	Ours	850	-
12	143	4	No	Ours	850	-
13	383	4	Yes	Ours	512	[20]
13	312	5	Yes	Ours	470	YASHE[7]
14	622	5	Yes	Ours	470	YASHE[7]
15	1271	5	Yes	Ours	512	[17]
15	1243	6	Yes	Ours	660	YASHE[7]
16	2485	7	Yes	Ours	820	YASHE[7]

$\log n$	Prediction	$\log r$
11	116	4
11	82	3
11	71	2
12	146	4
12	105	1
13	271	5
13	155	1
14	525	6
14	228	1
15	1045	7
15	335	1
16	2121	8
16	491	1

Figure 1: Experiments with LLL for solving the NTRU problem in the ring $\mathbb{Z}[X]/(q, X^n + 1)$, where the coefficients of the polynomials are uniform in $\{-1, 0, 1\}$. The lattice dimension used is equal to the number of coordinates used added to n/r. The values of [1] are the smallest moduli for which their algorithm works, up to one, one and five bits. The prediction is the minimum $\log q$ an LLL reduction can solve assuming we use all the (necessary) coordinates.

$\log n$	$\log q$	ℓ	Success
11	72	1116	Yes
11	70	1200	Yes
11	69	1200	No
12	118	1024	Yes
12	117	1024	No
12	105	1700	Yes
12	104	1700	No
13	230	1024	Yes
14	450	1024	Yes
15	930	1024	Yes

$\log n$	ℓ	Prediction
11	1033	71
12	1472	106
13	2275	156
14	3357	230
15	5127	337
16	7124	477

Figure 2: Experiments with LLL for solving the NTRU problem in the ring $\mathbb{Z}[X]/(q, X^p - X - 1)$, where the coefficients of the polynomials are uniform in $\{-1, 0, 1\}$ and p is the smallest prime larger than n. The lattice dimension used is ℓ. The prediction is the minimum $\log q$ an LLL reduction can solve.

6 Explicit complexity

We now turn towards the problem of deriving the first order of the asymptotical complexity of *heuristic* algorithms. Before the dual BKZ algorithm [38], simple derivations (as in [30, Appendix B]) could only be done using the Geometric Series Assumption, since the heuristic Gram-Schmidt norms outputted by the BKZ algorithm have a fairly complicated nature (see [26]), making an exact derivation quite cumbersome if not intractable. We are only interested in the part of the Gram-Schmidt norms known to be geometrically decreasing, which simplifies the computations [2].

We emphasize that we are only using standard heuristics, checked in practice, and *tight* at the first order. We compute the necessary block-size β to solve the problems and assume $\log \beta \approx \log n$. More precisely, if $\log \beta = (1 + o(1)) \log n$, then the exponent in the running time is within $1 + o(1)$ of its actual value.

For more information on the dual BKZ algorithm and dual lattices, see [38]. We denote by dual BKZ algorithm their algorithm 1 followed by a forward (i.e. primal) round, so that it attempts to *minimize* the *first* Gram-Schmidt norm (as the previous algorithms), rather than *maximizing* the *last* Gram-Schmidt norm.

We remark that all uses of NTRU for "standard" cryptography (key-exchange, signature and IBE) are instantiated with a modulus below n^2, so that the lattice reduction algorithms are *not* affected by the property.

6.1 Security of Learning With Errors

The following heuristic analysis applies for NTRU, but also for any LWE problem with dimension n and exactly $2n$ samples [3], or Ring-LWE with two samples. The primal algorithm searches for a short vector in a lattice.

As usual, we build the lattice

$$\mathbf{A} = \begin{pmatrix} q\mathbf{I}_n & \mathbf{M}_{\mathbf{h}}^{\mathcal{O}_{\mathbb{L}}} \\ \mathbf{0} & \mathbf{I}_m \end{pmatrix}$$

and apply the dual BKZ algorithm on its dual. We assume it did not find the key, and suppose the projection of $(\mathbf{f}, \mathbf{g})$ orthogonally to the first $2n-1$ vector has a norm of $\sigma/\sqrt{n}$. Then, the last Gram-Schmidt norm must be smaller than $\sigma/\sqrt{n}$ and we compute the smallest block-size β such that it is not the case. Hopefully, this means that applying the algorithm with a block-size β will find the key.

Once the dual BKZ algorithm has converged, the $2n-\beta$ first Gram-Schmidt norms are decreasing with a rate of $\approx \beta^{-1/\beta}$ and the $2n-\beta$th norm is about $\sqrt{\beta} V^{1/\beta}$ where V is the product of the last β norms. We deduce that the volume of the dual lattice is

$$q^{-n} = \left(\frac{\sigma}{\sqrt{n}}\right)^{-2n} \beta^{-(2n-\beta)^2/2\beta - n} = \left(\frac{\sigma}{\sqrt{n}}\right)^{-2n} \beta^{-2n^2/\beta}$$

so with $q = n^a$, $\sigma = n^b$ and $\beta = nc$ we have

$$-a \approx 1 - 2b - 2/c$$

and we deduce $c = 2/(a + 1 - 2b)$.

Another possibility is to apply the dual BKZ algorithm on the basis. If it reduces the last $m+n$ vectors, then the $m+n-\beta$th Gram-Schmidt norm cannot be smaller than the size of the key, σ. Now, if $m = n$ this norm is $\sqrt{q}\beta^{n/\beta - (2n-\beta)/\beta}$, and we deduce $a/2 - 1/c + 1 = b$ or $c = 2/(a + 2 - 2b)$ which happens when $c \geq 2/a$ (iff $b \geq 1$). Else, we take m maximum so that $q^{m/(m+n)}\beta^{(m+n)/2\beta} = q$ or $m = n(\sqrt{2ca} - 1)$ which gives $q\beta^{-(m+n-\beta)/\beta} = \sigma$ or $a - (\sqrt{2ca} - 1 + 1 - c)/c = b$ and hence $c = 2a/(a + 1 - b)^2$ when $b \leq 1$.

The dual algorithm searches for $2^{o(n)}$ short vectors in the dual lattice, so that the inner product with a gaussian of standard deviation σ can be distinguished. Applying the dual BKZ algorithm on the dual

[2] We remark that the last Gram-Schmidt norms have no constraints in the original algorithm. However, we can always assume they are HKZ-reduced, so that their logarithms are a parabola.

[3] Beware that an element sampled in the ring with standard deviation σ has coordinates of size only $\sigma/\sqrt{n}$.

lattice gives a vector of norm $\beta^{n/\beta}q^{-m/(n+m)} = \sigma/n$. The norm is minimized for $m = \sqrt{2ac} - 1$ or $m = n$, which gives $c = 2a/(a+1-b)^2$ when $b < 1$, and $2/(a+2-2b)$ else.

In all cases, the best complexity is given by $c = \max(2a/(a+1-b)^2, 2/(a+2-2b))$ (and when the number of samples is unlimited, this is $2a/(a+1-b)^2$).

6.2 Security of NTRU

Here, the analysis is specific to NTRU. We apply the dual BKZ algorithm to the same lattice, and compute the β such that the product of the n last Gram-Schmidt norms is equal to σ^n. Note that it is equivalent to having the product of the n first Gram-Schmidt norms equal to q/σ^n.

We first compute m such that the dual BKZ algorithm changes only the $2m$ middle norms. This is given by :

$$q = \sqrt{q}\beta^{m/\beta}$$

so that $m \approx a\beta/2$. For $a \geq 2$, we have $\beta \leq m$ so that, assuming $m \leq n$, the product of the m first norms is $q^m\beta^{-m^2/2\beta}$. Hence, we need $\beta^{m^2/2\beta} = \sigma^n$. We deduce

$$a^2c^2/8c = b$$

so that $c = 8b/a^2$.

When $m > n$, the first vector is of norm only $\sqrt{q}\beta^{n/\beta}$, so that for $c \leq 1$, we must have

$$q^{n/2}\beta^{n^2/2\beta - n^2/\beta} = \sigma^n$$

so that $a/2 - 1/2c \approx b$ and $c = 1/(a-2b)$. For this formula to be correct, we need $8b/a^2 a/2 \geq 1$, or $4b \geq a$.

We can show that this is better than the algorithms against Ring-LWE when $b = 1/2$ ($\approx$ binary errors) when $a \geq (4 + \sqrt[3]{262 - 6\sqrt{129}} + \sqrt[3]{262 + 6\sqrt{129}})/6 \approx 2.783$. When $b \geq 1$ which is the proven case, it is better for all $a > 4$ and $b < a/2 - 1$.

We again remark that going to a subfield, so that nb is constant, does not improve the complexity.

7 Conclusion

We conclude that the shortest vector problem over module lattices seems strictly easier than the bounded distance decoding. Since the practical cost of transforming a NTRU-based cryptosystem into a Ring-LWE-based cryptosystem is usually small, especially for key-exchange (e.g. [3]), we recommend to dismiss the former, in particular since it is known to be weaker (see [40, Section 4.4.4]). One important difference between NTRU and Ring-LWE instances is the fact that in NTRU lattices, there exists many short vectors. This has been used by May and Silverman in [34] and in our case, the determinant of the sublattice generated by these short vectors is an important parameter to predict the behaviour of our algorithm.

We remark that the only proven way to use NTRU is to use $\sigma \approx \sqrt{n^3 q}$ [42]. We showed here that attacks are more efficient against NTRU than on a Ring-LWE lattice until $\sigma \approx n^{-1}\sqrt{q}$, which suggests their result is essentially optimal. Furthermore, the property we use is present until $\sigma \approx \sqrt{nq}$, i.e. until h is (heuristically) indistinguishable from uniform.

Our results show that the root approximation factor is a poor indicator in the NTRU case : indeed, we reached 1.0059 using a mere LLL. We suggest to switch the complexity measure to the maximum dimension used in shortest vector routines (i.e. the block size of the lattice reduction algorithm) of a successful attack. While there are less problems with LWE-based cryptosystems, the root approximation factor has also several shortcomings which are corrected by this modification. Indeed, highly reduced basis do not obey to the Geometric Series Assumption, so that the root approximation factor also depends on the dimension of the lattice. Even when the dimension is much larger than the block-size, converting the factor into a

block-size - which is essentially inverting the function $\beta \mapsto \left(\frac{(\beta/2)!}{\pi^{\beta/2}}\right)^{1/\beta^2}$ - is very cumbersome. Finally, the complexity of shortest vector algorithms is more naturally expressed as a function of the dimension than the asymptotical root approximation factor they can achieve.

Acknowledgments.

We would like to thank the Crypto Team at ENS for providing us computational ressources to perform our experimentations.

References

[1] Martin Albrecht, Shi Bai, and Léo Ducas. A subfield lattice attack on overstretched NTRU assumptions: Cryptanalysis of some FHE and Graded Encoding Schemes. Cryptology ePrint Archive, Report 2016/127, 2016. http://eprint.iacr.org/.

[2] Martin R. Albrecht, Catalin Cocis, Fabien Laguillaumie, and Adeline Langlois. Implementing candidate graded encoding schemes from ideal lattices. In Tetsu Iwata and Jung Hee Cheon, editors, *Advances in Cryptology – ASIACRYPT 2015, Part II*, volume 9453 of *Lecture Notes in Computer Science*, pages 752–775, Auckland, New Zealand, November 30 – December 3, 2015. Springer, Heidelberg, Germany.

[3] Erdem Alkim, Léo Ducas, Thomas Pöppelmann, and Peter Schwabe. Post-quantum key exchange - a new hope. Cryptology ePrint Archive, Report 2015/1092, 2015. http://eprint.iacr.org/2015/1092.

[4] Wojciech Banaszczyk. New bounds in some transference theorems in the geometry of numbers. *Mathematische Annalen*, 296(1):625–635, 1993.

[5] Daniel J Bernstein, Chitchanok Chuengsatiansup, Tanja Lange, and Christine van Vredendaal. Ntru prime. 2016. http://eprint.iacr.org/.

[6] Avrim Blum, Adam Kalai, and Hal Wasserman. Noise-tolerant learning, the parity problem, and the statistical query model. *J. ACM*, 50(4):506–519, 2003.

[7] Joppe W. Bos, Kristin Lauter, Jake Loftus, and Michael Naehrig. Improved security for a ring-based fully homomorphic encryption scheme. In Martijn Stam, editor, *14th IMA International Conference on Cryptography and Coding*, volume 8308 of *Lecture Notes in Computer Science*, pages 45–64, Oxford, UK, December 17–19, 2013. Springer, Heidelberg, Germany.

[8] Joppe W Bos, Kristin Lauter, and Michael Naehrig. Private predictive analysis on encrypted medical data. *Journal of biomedical informatics*, 50:234–243, 2014.

[9] Stéphane Boucheron, Gábor Lugosi, and Olivier Bousquet. Concentration inequalities. In *Advanced Lectures on Machine Learning*, pages 208–240. Springer, 2004.

[10] Daniel Cabarcas, Patrick Weiden, and Johannes A. Buchmann. On the efficiency of provably secure NTRU. In *Post-Quantum Cryptography - 6th International Workshop, PQCrypto 2014, Waterloo, ON, Canada, October 1-3, 2014. Proceedings*, pages 22–39, 2014.

[11] Gizem S. Çetin, Wei Dai, Yarkın Doröz, and Berk Sunar. Homomorphic autocomplete. Cryptology ePrint Archive, Report 2015/1194, 2015. http://eprint.iacr.org/2015/1194.

[12] Gizem S. Çetin, Yarkin Doröz, Berk Sunar, and Erkay Savas. Depth optimized efficient homomorphic sorting. In Kristin E. Lauter and Francisco Rodríguez-Henríquez, editors, *Progress in Cryptology - LATINCRYPT 2015: 4th International Conference on Cryptology and Information Security in Latin America*, volume 9230 of *Lecture Notes in Computer Science*, pages 61–80, Guadalajara, Mexico, August 23–26, 2015. Springer, Heidelberg, Germany.

[13] Yuanmi Chen and Phong Q. Nguyen. BKZ 2.0: Better lattice security estimates. In Dong Hoon Lee and Xiaoyun Wang, editors, *Advances in Cryptology – ASIACRYPT 2011*, volume 7073 of *Lecture Notes in Computer Science*, pages 1–20, Seoul, South Korea, December 4–8, 2011. Springer, Heidelberg, Germany.

[14] Jung Hee Cheon, Jinhyuck Jeong, and Changmin Lee. An Algorithm for NTRU Problems and Cryptanalysis of the GGH Multilinear Map without an encoding of zero. Cryptology ePrint Archive, Report 2016/139, 2016. `http://eprint.iacr.org/`.

[15] Don Coppersmith and Adi Shamir. Lattice attacks on NTRU. In Walter Fumy, editor, *Advances in Cryptology – EUROCRYPT'97*, volume 1233 of *Lecture Notes in Computer Science*, pages 52–61, Konstanz, Germany, May 11–15, 1997. Springer, Heidelberg, Germany.

[16] Wei Dai, Yarkin Doröz, and Berk Sunar. Accelerating SWHE based PIRs using GPUs. In Michael Brenner, Nicolas Christin, Benjamin Johnson, and Kurt Rohloff, editors, *FC 2015 Workshops*, volume 8976 of *Lecture Notes in Computer Science*, pages 160–171, San Juan, Puerto Rico, January 30, 2015. Springer, Heidelberg, Germany.

[17] Yarkın Doröz, Yin Hu, and Berk Sunar. Homomorphic aes evaluation using the modified ltv scheme. *Designs, Codes and Cryptography*, pages 1–26, 2015.

[18] Yarkin Doröz, Aria Shahverdi, Thomas Eisenbarth, and Berk Sunar. Toward practical homomorphic evaluation of block ciphers using prince. In Rainer Böhme, Michael Brenner, Tyler Moore, and Matthew Smith, editors, *FC 2014 Workshops*, volume 8438 of *Lecture Notes in Computer Science*, pages 208–220, Christ Church, Barbados, March 7, 2014. Springer, Heidelberg, Germany.

[19] Yarkin Doröz, Berk Sunar, and Ghaith Hammouri. Bandwidth efficient PIR from NTRU. In Rainer Böhme, Michael Brenner, Tyler Moore, and Matthew Smith, editors, *FC 2014 Workshops*, volume 8438 of *Lecture Notes in Computer Science*, pages 195–207, Christ Church, Barbados, March 7, 2014. Springer, Heidelberg, Germany.

[20] Nathan Dowlin, Ran Gilad-Bachrach, Kim Laine, Kristin Lauter, Michael Naehrig, and John Wernsing. Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. 2015.

[21] Léo Ducas, Alain Durmus, Tancrède Lepoint, and Vadim Lyubashevsky. Lattice signatures and bimodal gaussians. In Ran Canetti and Juan A. Garay, editors, *Advances in Cryptology – CRYPTO 2013, Part I*, volume 8042 of *Lecture Notes in Computer Science*, pages 40–56, Santa Barbara, CA, USA, August 18–22, 2013. Springer, Heidelberg, Germany.

[22] Léo Ducas, Vadim Lyubashevsky, and Thomas Prest. Efficient identity-based encryption over NTRU lattices. In Palash Sarkar and Tetsu Iwata, editors, *Advances in Cryptology – ASIACRYPT 2014, Part II*, volume 8874 of *Lecture Notes in Computer Science*, pages 22–41, Kaoshiung, Taiwan, R.O.C., December 7–11, 2014. Springer, Heidelberg, Germany.

[23] Nicolas Gama, Nick Howgrave-Graham, and Phong Q. Nguyen. Symplectic lattice reduction and NTRU. In Serge Vaudenay, editor, *Advances in Cryptology – EUROCRYPT 2006*, volume 4004 of *Lecture Notes in Computer Science*, pages 233–253, St. Petersburg, Russia, May 28 – June 1, 2006. Springer, Heidelberg, Germany.

[24] Craig Gentry. Key recovery and message attacks on NTRU-composite. In Birgit Pfitzmann, editor, *Advances in Cryptology – EUROCRYPT 2001*, volume 2045 of *Lecture Notes in Computer Science*, pages 182–194, Innsbruck, Austria, May 6–10, 2001. Springer, Heidelberg, Germany.

[25] Craig Gentry and Michael Szydlo. Cryptanalysis of the revised NTRU signature scheme. In Lars R. Knudsen, editor, *Advances in Cryptology – EUROCRYPT 2002*, volume 2332 of *Lecture Notes in Computer Science*, pages 299–320, Amsterdam, The Netherlands, April 28 – May 2, 2002. Springer, Heidelberg, Germany.

[26] Guillaume Hanrot, Xavier Pujol, and Damien Stehlé. Analyzing blockwise lattice algorithms using dynamical systems. In Phillip Rogaway, editor, *Advances in Cryptology – CRYPTO 2011*, volume 6841 of *Lecture Notes in Computer Science*, pages 447–464, Santa Barbara, CA, USA, August 14–18, 2011. Springer, Heidelberg, Germany.

[27] Jeffrey Hoffstein, Jill Pipher, and Joseph H. Silverman. NTRU: A ring-based public key cryptosystem. In *Algorithmic Number Theory, Third International Symposium, ANTS-III, Portland, Oregon, USA, June 21-25, 1998, Proceedings*, pages 267–288, 1998.

[28] Nick Howgrave-Graham. A hybrid lattice-reduction and meet-in-the-middle attack against NTRU. In Alfred Menezes, editor, *Advances in Cryptology – CRYPTO 2007*, volume 4622 of *Lecture Notes in Computer Science*, pages 150–169, Santa Barbara, CA, USA, August 19–23, 2007. Springer, Heidelberg, Germany.

[29] Miran Kim and Kristin Lauter. Private genome analysis through homomorphic encryption. *BMC medical informatics and decision making*, 15(Suppl 5):S3, 2015.

[30] Paul Kirchner and Pierre-Alain Fouque. An improved BKW algorithm for LWE with applications to cryptography and lattices. In Rosario Gennaro and Matthew J. B. Robshaw, editors, *Advances in Cryptology – CRYPTO 2015, Part I*, volume 9215 of *Lecture Notes in Computer Science*, pages 43–62, Santa Barbara, CA, USA, August 16–20, 2015. Springer, Heidelberg, Germany.

[31] Kristin E. Lauter, Adriana López-Alt, and Michael Naehrig. Private computation on encrypted genomic data. In Diego F. Aranha and Alfred Menezes, editors, *Progress in Cryptology - LATINCRYPT 2014: 3rd International Conference on Cryptology and Information Security in Latin America*, volume 8895 of *Lecture Notes in Computer Science*, pages 3–27, Florianópolis, Brazil, September 17–19, 2015. Springer, Heidelberg, Germany.

[32] Tancrède Lepoint and Michael Naehrig. A comparison of the homomorphic encryption schemes FV and YASHE. In David Pointcheval and Damien Vergnaud, editors, *AFRICACRYPT 14: 7th International Conference on Cryptology in Africa*, volume 8469 of *Lecture Notes in Computer Science*, pages 318–335, Marrakesh, Morocco, May 28–30, 2014. Springer, Heidelberg, Germany.

[33] Adriana López-Alt, Eran Tromer, and Vinod Vaikuntanathan. On-the-fly multiparty computation on the cloud via multikey fully homomorphic encryption. In Howard J. Karloff and Toniann Pitassi, editors, *44th Annual ACM Symposium on Theory of Computing*, pages 1219–1234, New York, NY, USA, May 19–22, 2012. ACM Press.

[34] Alexander May and Joseph H. Silverman. Dimension Reduction Methods for Convolution Modular Lattices. In *Cryptography and Lattices, International Conference, CaLC 2001, Providence, RI, USA, March 29-30, 2001, Revised Papers*, pages 110–125, 2001.

[35] Daniele Micciancio and Shafi Goldwasser. *Complexity of lattice problems: a cryptographic perspective*, volume 671. Springer Science & Business Media, 2012.

[36] Daniele Micciancio and Oded Regev. Worst-case to average-case reductions based on gaussian measures. *SIAM J. Comput.*, 37(1):267–302, 2007.

[37] Daniele Micciancio and Panagiotis Voulgaris. Faster exponential time algorithms for the shortest vector problem. In Moses Charika, editor, *21st Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1468–1480, Austin, Texas, USA, January 17–19, 2010. ACM-SIAM.

[38] Daniele Micciancio and Michael Walter. Practical, predictable lattice basis reduction. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 820–849. Springer, 2016.

[39] Gábor Pataki and Mustafa Tural. On sublattice determinants in reduced bases. *arXiv preprint arXiv:0804.4014*, 2008.

[40] Chris Peikert. A decade of lattice cryptography. Cryptology ePrint Archive, Report 2015/939, 2015. `http://eprint.iacr.org/2015/939`.

[41] Claus-Peter Schnorr. A hierarchy of polynomial time lattice basis reduction algorithms. *Theoretical computer science*, 53(2):201–224, 1987.

[42] Damien Stehlé and Ron Steinfeld. Making NTRU as secure as worst-case problems over ideal lattices. In Kenneth G. Paterson, editor, *Advances in Cryptology – EUROCRYPT 2011*, volume 6632 of *Lecture Notes in Computer Science*, pages 27–47, Tallinn, Estonia, May 15–19, 2011. Springer, Heidelberg, Germany.

[43] Lawrence C. Washington. *Introduction to Cyclotomic Fields*, volume 83. Springer, 1997.

A Proofs of Banaszczyk Lemma

We now recall two results from [36] and Banaszczyk's lemma [4] about discrete gaussian sampling over a lattice.

Lemma 3. *Given a lattice $\mathcal{L} \subset \mathbb{R}^n$, for any s and $\mathbf{c} \in \mathbb{R}^n$, we have*

$$\rho_s(\mathcal{L} + \mathbf{c}) \leq \rho_s(\mathcal{L}).$$

Proof. Using Poisson summations, with $\mathcal{L}^* = \{\mathbf{x} \in \mathbb{R}^n; \langle \mathbf{x}, \mathcal{L} \rangle \subset \mathbb{Z}\}$ the dual lattice, we have

$$\begin{aligned}
\rho_s(\mathcal{L} + \mathbf{c}) =& s^n \mathrm{Vol}(\mathcal{L}^*) \sum_{\mathbf{x} \in \mathcal{L}^*} \exp(2\pi i \langle \mathbf{c}, \mathbf{x} \rangle) \exp(-\pi s^2 \|\mathbf{x}\|^2) \\
=& s^n \mathrm{Vol}(\mathcal{L}^*) \sum_{\mathbf{x} \in \mathcal{L}^*} \cos(2\pi \langle \mathbf{c}, \mathbf{x} \rangle) \exp(-\pi s^2 \|\mathbf{x}\|^2) \\
\leq& s^n \mathrm{Vol}(\mathcal{L}^*) \sum_{\mathbf{x} \in \mathcal{L}^*} \exp(-\pi s^2 \|\mathbf{x}\|^2)
\end{aligned}$$

and

$$\rho_s(\mathcal{L}) = s^n \mathrm{Vol}(\mathcal{L}^*) \sum_{\mathbf{x} \in \mathcal{L}^*} \exp(-\pi s^2 \|\mathbf{x}\|^2).$$

□

Lemma 4. *For a lattice $\mathcal{L}$, any $t \geq 1$, the probability that $\mathbf{x}$ sampled according to $D_{\Lambda,s}$ verifies $\|\mathbf{x}\| > st\sqrt{\frac{n}{2\pi}}$ is at most*

$$\exp\big(-n(t-1)^2/2\big).$$

Proof. Without loss of generality, we assume $s = 1$. We first have, using Poisson summation :

$$\frac{\rho_t(\mathcal{L})}{\rho_1(\mathcal{L})} = t^n \frac{\rho_{1/t}(\mathcal{L}^*)}{\rho_1(\mathcal{L}^*)} \leq t^n.$$

Then, with $\mathcal{B}(\mathbf{0}, r)$ the ball of radius r centered on the origin,

$$\begin{aligned} t^n \rho_1(\mathcal{L}) \geq & \rho_t\left(\mathcal{L} \setminus \mathcal{B}\left(\mathbf{0}, t\sqrt{\frac{n}{2\pi}}\right)\right) \\ & \geq \exp((t^2-1)n/2)\rho_1\left(\mathcal{L} \setminus \mathcal{B}\left(\mathbf{0}, t\sqrt{\frac{n}{2\pi}}\right)\right). \end{aligned}$$

And therefore :

$$\begin{aligned} \frac{\rho_1(\mathcal{L} \setminus \mathcal{B}(\mathbf{0}, t\sqrt{\frac{n}{2\pi}}))}{\rho(\mathcal{L})} & \leq \exp(-n(t^2 - 2\ln t - 1)/2) \\ & \leq \exp(-n(t-1)^2/2). \end{aligned}$$

□