Improved Provable Reduction of NTRU and Hypercubic Lattices

Henry Bambury[1,2] and Phong Q. Nguyen[1]

[1] DIENS, École normale supérieure, PSL University, CNRS, Inria, Paris, France
[2] DGA, Paris, France

Abstract. Lattice-based cryptography typically uses lattices with special properties to improve efficiency. We show how blockwise reduction can exploit lattices with special geometric properties, effectively reducing the required blocksize to solve the shortest vector problem to half of the lattice's rank, and in the case of the hypercubic lattice $\mathbb{Z}^n$, further relaxing the approximation factor of blocks to $\sqrt{2}$. We study both provable algorithms and the heuristic well-known primal attack, in the case where the lattice has a first minimum that is almost as short as that of the hypercubic lattice $\mathbb{Z}^n$. Remarkably, these near-hypercubic lattices cover Falcon and most concrete instances of the NTRU cryptosystem: this is the first provable result showing that breaking NTRU lattices can be reduced to finding shortest lattice vectors in halved dimension, thereby providing a positive response to a conjecture of Gama, Howgrave-Graham and Nguyen at Eurocrypt 2006.

Keywords: Lattices · Cryptanalysis · NTRU · ZLIP

1 Introduction

Lattice-based cryptography has emerged as the main alternative to classical public key cryptography based on factoring and discrete logarithm: it can provide resistance to quantum computers and offer new functionalities such as fully-homomorphic encryption. However, for efficiency reasons, the lattices used in concrete cryptosystems are usually not random lattices: they have special properties, to improve keysize and/or speed up operations and/or enable extra operations. For instance, all the lattices used in NIST's new post-quantum standards are special: module lattices for Kyber [7] and Dilithium [18], and NTRU lattices for Falcon [24]. Recently, even hypercubic lattices [46], which are simply rotations of $\mathbb{Z}^n$, have been proposed in [22,19,8] as the basis of concrete cryptosystems, with Hawk [19] being submitted to the new NIST call for post-quantum signatures.

Accordingly, it is crucial to understand if these special properties make the underlying lattice problems easier to solve, and if so, by how much. In the case of module lattices, this remains very much an open problem, except for the case of ideal lattices, for which better algorithms have been found [15,41,9]. For NTRU lattices, it is also an open problem: in fact, Gama, Howgrave-Graham

and Nguyen [26] conjectured at Eurocrypt 2006 that the reduction of a $2n$-dimensional NTRU lattice could be reduced to that of an αn-dimensional lattice for some $\alpha < 2$. The hypercubic lattice $\mathbb{Z}^n$ was first studied by Szydlo [46] twenty years ago, but it was only shown very recently to be significantly easier to reduce than generic lattices: one can recover an orthonormal basis of $\mathbb{Z}^n$ in time $2^{n/2+o(n)}$ using the algorithm of Bennett, Ganju, Peetathawatchai and Stephens-Davidowitz[3] [8], or, as shown by Ducas [17] by using polynomially many calls to an oracle for the shortest vector problem (SVP) in dimension $n/2$, which also leads to an asymptotic running time of $2^{n/2+o(n)}$. In other words, solving SVP for $\mathbb{Z}^n$ can be reduced to solving SVP in dimension $n/2$.

Our results. We introduce a new blockwise reduction algorithm, which is a variant of Ducas's algorithm [17], itself a variant of Gama-Nguyen's slide reduction [27]. The differences with Ducas's approach are twofold.

First, our algorithm is more general, as it is not restricted to $\mathbb{Z}^n$: it also applies to any lattice L such that the product of its first minimum with that of its dual lattice is small, namely $\lambda_1(L)\lambda_1(L^\times) < 1 - \frac{1}{\mathrm{poly}(n)}$, where $\lambda_1(\cdot)$ and $L^\times$ denote respectively the first minimum and the dual lattice. This condition is typically not satisfied for a generic lattice: Minkowski's inequality only implies that $\lambda_1(L)\lambda_1(L^\times) = O(n)$. But it turns out to be satisfied by most instantiations of NTRU, because the symplectic property of NTRU uncovered by Gama *et al.* [26] implies that $\lambda_1(L)\lambda_1(L^\times) = \lambda_1(L)^2/q$ where q is the small modulus of the NTRU cryptosystem, and also equal to $\mathrm{vol}(L)^{2/\mathrm{rank}(L)}$. In the recent NTRU-HPS submission [12] to NIST, we have $\lambda_1(L)^2/q < 1/2$ for all three parameter sets proposed, due to the absence of decryption failures. For the original NTRU [31] from the 90s and for Falcon [24], this does not hold but can be taken care of by a mild heuristic assumption on the projection of secret vectors over random subspaces related to lattice reduction: similar yet stronger assumptions were made and checked in the context of lattice enumeration [29]. Thus, we show that for the NTRU-HPS submission [12], one can provably find a non-zero lattice vector at least as short as the secret key, by solving the shortest vector problem in a lattice of halved dimension. This is the first rigorous result showing that an NTRU lattice can be solved by working with SVP oracles in a smaller dimension than what is required for a generic lattice. It should not be confused with heuristic security estimates where the blocksize required to break the underlying system is heuristically estimated to be a fraction of the lattice dimension.

Second, our algorithm improves that of Ducas in the case of $\mathbb{Z}^n$. Ducas required an exact or nearly-exact algorithm for SVP in dimension $n/2$, whereas our algorithm can tolerate an approximate-SVP algorithm in dimension $n/2$ with an approximation factor essentially $\sqrt{2}$. Intuitively, a factor $\sqrt{2}$ should make the problem easier, and the SVP challenges [43] suggest that the problem is easier in practice. Eisenbrand and Venzin [23] note that the best sieving algorithms give a provable $2^{0.802n+o(n)}$-runtime algorithm for $O(1)$ approximations of the

[3] Note that the *semi-stable* variant of their algorithm [8, Cor. 5.5] also applies to NTRU lattices.

shortest vector, although the constant is larger than $\sqrt{2}$. There is currently no theoretical evidence that approximating SVP to within $\sqrt{2}$ is easier than solving exact SVP, but if ever it is strictly easier, such as solvable in time $2^{\alpha n+o(n)}$ for some $\alpha < 1$, we would immediately obtain an exponentially faster algorithm for the $\mathbb{Z}^n$-Lattice Isomorphism Problem ($\mathbb{Z}$-LIP), running in time $2^{\alpha n/2+o(n)}$.

Finally, we compare the performances of our provable algorithms with heuristic estimates provided by the so-called primal attack [6]. In the case of $\mathbb{Z}^n$, this was done by [19], where the authors state without giving many details that a blocksize of $n/2 + o(n)$ is heuristically sufficient to recover a shortest vector. We show more generally that for any n-dimensional lattice L such that $\lambda_1(L) = O(\mathrm{vol}(L)^{1/n})$, the primal attack heuristically recovers a shortest lattice vector using a blocksize $n/2 + \Theta(n/\log n)$: this result applies to both $\mathbb{Z}^n$ and NTRU lattices. For these lattices, there are actually multiple shortest vectors, even a linear number: somewhat surprisingly, we show that the heuristic asymptotical blocksize required by the primal attack remains $n/2 + \Theta(n/\log n)$, even though in practice, it is somewhat easier. This means there is not much difference between the best theoretical algorithm and the best heuristic algorithm, and that the result depends essentially on the existence of one unusually short lattice vector.

Technical overview. Our algorithm differs from Ducas's algorithm in two main aspects. First, we distinguish the primal and the dual lattice. Second, we change the termination condition: instead of densifying a certain sublattice until it becomes hypercubic, we check whether our current primal and dual sublattices include a shortest vector.

Ducas's analysis [17] is based on a surprising upper bound $\sqrt{1-1/n}$ on the first minimum of projections of $\mathbb{Z}^n$ over certain subspaces. This upper bound is tight when the subspace is a hyperplane corresponding to the dual root lattice $A_{n-1}^{\times}$. However, we show that the upper bound can be improved for certain lower-dimensional subspaces, which might be of independent interest, and allows us to relax the SVP oracle to an approximate-SVP oracle with factor essentially $\sqrt{2}$. More precisely, it is well-known that the expectation of the squared norm of the projection of a unit vector onto a k-dimensional random subspace of $\mathbb{R}^n$ is k/n. We show that the expectation of the squared norm of the projection of a random element of a fixed orthonormal basis of $\mathbb{R}^n$ onto a fixed k-dimensional subspace is also k/n. This allows us to replace the bound $\sqrt{1-1/n}$ by essentially $\sqrt{1/2}$ when $k \approx n/2$.

Our analysis of the primal attack [6] differs a bit from the literature [5,16]. In the primal attack, it is crucial to estimate the projection of a short vector onto random subspaces related to lattice reduction. Previous work [5,16] restricted to a short vector from LWE, whose coordinates are independent Gaussians. However, we argue that this model does not match $\mathbb{Z}^n$ nor NTRU. So instead of the χ^2 distribution, we rely on the Beta distribution related to classical sphere statistics. And we heuristically extend the analysis to the case of linearly many short vectors.

Roadmap. Sect. 2 provides background. In Sect. 3, we present our new blockwise reduction algorithm for near-hypercubic lattices. In Sect. 4, we study the heuristic primal attack on those same lattices, and analyse which blocksize is required.

2 Preliminaries

General notations. Vectors are written in bold lowercase $\mathbf{v}$. The Euclidean norm of a vector $\mathbf{v} \in \mathbb{R}^n$ is denoted $\|\mathbf{v}\|$. The associated scalar product of $\mathbf{a} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^n$ is written $\langle \mathbf{a}, \mathbf{b} \rangle$. Throughout this paper, we use row representation of matrices. For a set of vectors $V \subseteq \mathbb{R}^n$, we write $\mathrm{span}(V)$ the real vector space generated by V. We write $V^\perp$ or $\mathrm{span}(V)^\perp$ for the set of vectors $\mathbf{w} \in \mathbb{R}^n$ such that $\langle \mathbf{w}, \mathbf{v} \rangle = 0$ for all $\mathbf{v}$ in V. π_V denotes the orthogonal projection onto $\mathrm{span}(V)$. We use the standard asymptotic notations $o(\cdot), O(\cdot), \Theta(\cdot)$ and $\omega(\cdot)$. As n goes to infinity, we use the notation $a_n \sim b_n$ as shorthand for $a_n = b_n + o(b_n)$. We use $\ll$ slightly differently to how it might usually be used: $a_n \ll b_n$ if there exists a polynomial P of constant degree such that for any large enough n, $a_n < b_n - \frac{1}{P(n)}$.

Probabilities. We denote the expectation of a random variable by $\mathbb{E}(\cdot)$, and probabilities by $\mathbb{P}(\cdot)$. As proved in [25], the squared norm of the projection of a unit vector of $\mathbb{R}^n$ onto a random k-dimensional subspace of $\mathbb{R}^n$ follows the *Beta distribution* $B(k/2, (n-k)/2)$. In particular, the expected squared norm of this projection is k/n. The cumulative distribution function of $B(a, b)$ is the *regularised incomplete beta function* $I_x(a, b)$. Asymptotic expansions of the regularised incomplete beta function rely on the *complementary error function* $\mathrm{erfc}(z) := \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-t^2} dt$. When z goes to infinity, $\mathrm{erfc}(z) \sim \pi^{-1/2} z^{-1} e^{-z^2}$.

Lattices. A *lattice* L is a discrete subgroup of $\mathbb{R}^m$. Alternatively, we can define a lattice as the set $\mathcal{L}(\mathbf{b}_1, \ldots, \mathbf{b}_n) = \{\sum_{i=1}^n x_i \mathbf{b}_i \ : \ x_i \in \mathbb{Z}\}$ of all integer combinations of n linearly independent vectors $\mathbf{b}_1, \ldots, \mathbf{b}_n \in \mathbb{R}^m$. This sequence of vectors is known as a *basis* of the lattice L. All the bases of L have the same number n of elements, called the dimension or rank of L, and the n-dimensional volume of the parallelepiped $\{\sum_{i=1}^n a_i \mathbf{b}_i \ : \ a_i \in [0, 1)\}$ they generate. We call this volume the volume, or determinant, of L, and denote it by $\mathrm{vol}(L)$. The lattice L is said to be *full-rank* if $n = m$. We denote by $\lambda_1(L)$ the first minimum of L, defined as the norm of a shortest nonzero vector of L.

Orthogonalisation. For a basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of a lattice L, and an index $1 \leq i \leq n$, we denote by π_i the orthogonal projection on $\mathrm{span}(\mathbf{b}_1, \ldots, \mathbf{b}_{i-1})^\perp$. The *Gram-Schmidt orthogonalisation* (GSO) of the basis B is defined as the orthogonal sequence of vectors $B^\star = (\mathbf{b}_1^\star, \ldots, \mathbf{b}_n^\star)$, where $\mathbf{b}_i^\star := \pi_i(\mathbf{b}_i)$. The projection of a lattice is not always a lattice, but for all $i \in \{1, \ldots, n\}$, $\pi_i(L)$ is a lattice of dimension $n + 1 - i$ generated by the basis $\pi_i(\mathbf{b}_i), \ldots, \pi_i(\mathbf{b}_n)$, such that $\mathrm{vol}(\pi_i(L)) = \prod_{j=i}^n \left\|\mathbf{b}_j^\star\right\|$.

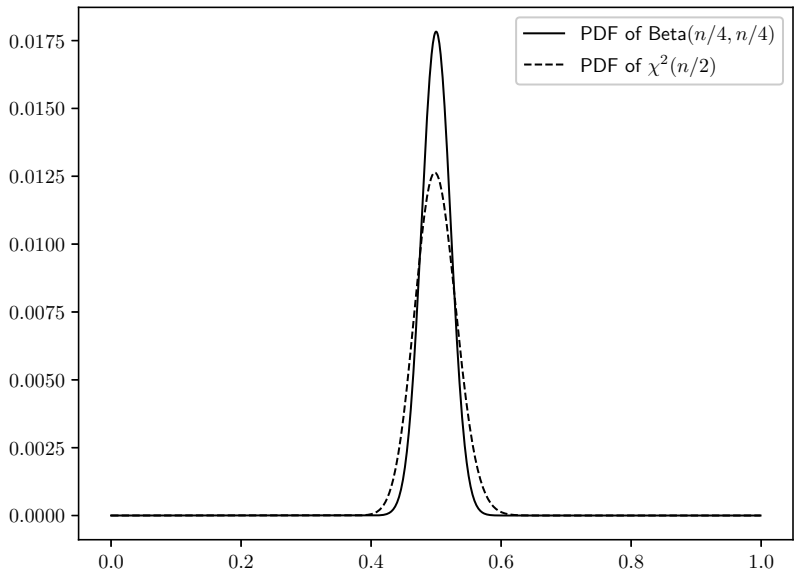


Fig. 1: Comparing the χ^2 and *Beta* distributions for $n = 1000$

Duality. For any lattice L, its *dual lattice* $L^\times$ is defined by

$$L^\times := \{\mathbf{w} \in \mathrm{span}(L) : \langle \mathbf{w}, \mathbf{v} \rangle \in \mathbb{Z} \text{ for all } \mathbf{v} \in L\}.$$

If L has rank $n > 0$, then $L^\times$ also, and $\mathrm{vol}(L) = \mathrm{vol}(L^\times)^{-1}$. If $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ is a basis of L, then there is a unique *dual basis* $(\mathbf{d}_1, \ldots, \mathbf{d}_n)$ of $L^\times$ such that $\langle \mathbf{b}_i, \mathbf{d}_j \rangle = \delta_{i,j}$ (Kronecker symbol) for all i, j. Duality is related to GSO as $\langle \mathbf{b}_i^\star / \|\mathbf{b}_i^\star\|^2, \mathbf{b}_i \rangle = 1$ implies that

$$\frac{\mathbf{b}_i^\star}{\|\mathbf{b}_i^\star\|^2} \in \mathcal{L}(\mathbf{b}_1, \ldots, \mathbf{b}_i)^\times.$$

In particular, $\mathbf{d}_n = \mathbf{b}_n^\star / \|\mathbf{b}_n^\star\|^2$ and $\|\mathbf{d}_n\| = \|\mathbf{b}_n^\star\|^{-1}$.

Primitivity. A sublattice L' of L is called *primitive* if $L' = \mathrm{span}(L') \cap L$ or equivalently L/L' is torsion free. In this case, $L = L' \oplus L/L'$. Equivalently, a sublattice of L is primitive if its bases can be completed into a basis of L. We will make heavy use of the following identity: if L' is a primitive sublattice of L, then

$$L/L' = \pi_{L'^\perp}(L) = (L^\times \cap \mathrm{span}(L')^\perp)^\times.$$

We refer to Chap. 1 of [38] for a proof as well as a more complete presentation of the interconnections between duality and primitivity.

Lattice problems. Let $\gamma \geq 1$. The most famous lattice problem is the *approximate shortest vector problem* (γ-SVP or SVP if $\gamma = 1$), which asks to find a nonzero lattice vector of norm less than $\gamma\lambda_1(L)$. A γ-SVP-oracle (or SVP-oracle when $\gamma = 1$) is an algorithm that takes a lattice L as input, and outputs a nonzero vector of L of norm less than $\gamma\lambda_1(L)$. Currently, the fastest known algorithms for worst-case SVP have runtime $2^{n+o(n)}$ ([1,3]).

Another lattice problem that has recently achieved significant cryptographic interest is the *lattice isomorphism problem* (LIP), and in particular its specialisation to rotations of $\mathbb{Z}^n$ ($\mathbb{Z}$LIP): given the image of $\mathbb{Z}^n$ under a linear orthogonal map (or rotation) $O \in \mathcal{O}_n(\mathbb{R})$, $\mathbb{Z}$LIP asks to recover O. It is not hard to see $\mathbb{Z}$LIP reduces to recovering unit vectors of the rotation, making $\mathbb{Z}$LIP at least as easy as SVP. Indeed, [8] and [17] propose $2^{n/2+o(n)}$ algorithms for $\mathbb{Z}$LIP.

We call *hypercubic* any lattice of $\mathbb{R}^n$ which has a $\mathbb{Z}$-basis consisting of unit vectors which are pairwise orthogonal. Full rank hypercubic lattices of $\mathbb{R}^n$ are exactly isomorphisms of $\mathbb{Z}^n$. In addition, a hypercubic lattice Λ is self-dual: $\Lambda = \Lambda^\times$.

Lattice reduction. The celebrated LLL algorithm [36] solves 2^n-SVP in polynomial time. Blockwise algorithms such as BKZ [45,14] and its variants [27,39,2] approximate SVP within better factors, using polynomially many calls to an exact (or near-exact) SVP oracle in rank less than an input parameter called the *blocksize*. Following [27,39,2], we call γ-SVP-reduction any algorithm which outputs a basis whose first vector solves γ-SVP. Similarly, we call γ-DVSP-reduction (where D stands for dual) any algorithm which outputs a basis whose last Gram-Schmidt vector solves γ-SVP in the dual lattice. Given a γ-SVP-oracle, it is possible to γ-SVP-reduce or γ-DSVP-reduce in polynomial time (see [28,39]).

Reduced bases. Lattice reduction algorithms aim to transform an input basis into a "high quality" basis. There are many ways to quantify the quality of bases produced by lattice reduction algorithms. One popular way is to consider the Gram-Schmidt norms $\|\mathbf{b}_1^\star\|, \ldots, \|\mathbf{b}_n^\star\|$. Intuitively speaking, a good basis is one in which this sequence does not decay too fast. In practice, it turns out that the Gram-Schmidt coefficients of bases produced by the main reduction algorithms (such as LLL or BKZ) have a certain "typical shape", assuming the input basis is sufficiently random. This property was thoroughly investigated in [28,40]. This typical shape is often used to estimate the running time of various algorithms. In particular, many theoretical asymptotic analyses (as introduced by Schnorr [44]) assume for simplicity that this shape is given by $\|\mathbf{b}_i^\star\|/\|\mathbf{b}_{i+1}^\star\| = q$ where q depends on the reduction algorithm; although less precise, this approximation called the *geometric series assumption (GSA)* is close to the shape observed in practice. It is heuristically[4] estimated [14,13,37] that the BKZ algorithm with blocksize β, given as input a basis of an n-rank lattice L outputs a basis whose first vector has norm approximately equal to

[4] By replacing Hermite's constant by a Gaussian heuristic estimate.

$\delta_\beta^n \mathrm{vol}(L)^{1/n}$, where $\delta_\beta = \left(\frac{\beta}{2\pi e}(\pi\beta)^{1/\beta}\right)^{\frac{1}{2(\beta-1)}}$. Combining this with the GSA and the fact that $\mathrm{vol}(L) = \prod_{i=1}^n \|\mathbf{b}_i^\star\|$ gives estimates of the Gram-Schmidt norms: for $1 \leq i \leq n$,

$$\|\mathbf{b}_i^\star\| \approx \delta_\beta^{n-\frac{2n}{n-1}i} \mathrm{vol}(L)^{1/n}.$$

Such a heuristic model is widely used in security estimates of lattice-based NIST submissions.

The primal attack. Parameters of lattice-based cryptosystems are chosen after careful study of known attacks. The most important attack that people consider today is called the *primal attack*, which runs the BKZ blockwise reduction [45,14] with a sufficiently high blocksize. Building upon [28,14], the authors of [6] proposed to heuristically estimate the blocksize required by this attack to recover a short vector $\mathbf{s}$ in a rank n lattice L, by comparing the expected norm of $\pi_{n-\beta+1}(\mathbf{s})$ to the expected value of $\|\mathbf{b}_{n-\beta+1}^\star\|$. Using the GSA, as soon as

$$\sqrt{\frac{\beta}{n}}\|\mathbf{s}\| < \delta_\beta^{2\beta-n-1}\mathrm{vol}(L)^{1/n} \tag{1}$$

holds, the projection $\pi_{n-\beta+1}(\mathbf{s})$ is either 0 and then $\mathbf{s}$ lives in the subspace generated by the first $n-\beta$ vectors of the reduced basis, or it has a high chance of being shorter than $\|\mathbf{b}_{n-\beta+1}^\star\|$, making it such that the SVP oracle on the last block of size β will recover it. Albrecht, Göpfert, Virdia and Wunderer [5] and Dachman-Soled, Ducas, Gong, Rossi [16] refine and experimentally confirm this framework in the case of LWE. It should be stressed that the analysis of the primal attack remains very much heuristic.

The original NTRU cryptosystem. The NTRU cryptosystem [31], proposed by Hoffstein, Pipher and Silverman, works in the ring $\mathcal{R} = \mathbb{Z}[X]/(X^n - 1)$. An element $f = \sum_{i=0}^{n-1} f_i x^i = [f_0, f_1, \ldots, f_{n-1}] \in \mathcal{R}$ is seen as a polynomial or a row vector. To select keys, one uses the set $\mathcal{L}(d_1, d_2)$ of polynomials $F \in \mathcal{R}$ such that d_1 coefficients are equal to 1, d_2 coefficients are equal to -1, and the rest are zero. There are two small coprime moduli $p < q$, such as $q = 128$ and $p = 3$.

Historically, the secret keys were $f \in \mathcal{L}(d_f, d_f - 1)$ and $g \in \mathcal{L}(d_g, d_g)$ for some integers d_f and d_g significantly smaller than n, but other NTRU instantiations [30,32,12] use different parameters for $\mathcal{L}$, such as binary polynomials $\mathcal{L}(d, 0)$. To illustrate, we focus on the NTRU-HPS parameters of NTRU's NIST submission [12], one of the seven finalists: f is a random polynomial in $\{0, \pm 1\}^n$, and $g \in \mathcal{L}(d_g, d_g)$ where $2d_g = q/8 - 2$. With high probability, f is invertible mod q. The public key $h \in \mathcal{R}$ is defined as $h = g/f \mod q$. Thus, in the ring $\mathcal{R}/q\mathcal{R}$ which we represent by $\mathbb{Z}_q^n$, we have $f * h = g$. In this article, there is no need to know how NTRU encryption or signature works. The polynomial h defines the so-called NTRU lattice Λ_h, formed by all pairs of polynomials $(u, v) \in \mathcal{R}^2$ such that $v * h \equiv u \mod q$. Here, we follow the definition of [33], but other papers may use a variant of Λ_h, using a permutation of the coordinates. Λ_h is generated

by the rows of the following lower-triangular matrix, which is its Hermite normal form:

$$\begin{pmatrix} qI_n & 0 \\ H & I_n \end{pmatrix},$$

where H is the circulant matrix for the polynomial $h \equiv g/f = \sum_{i=0}^{n-1} h_i x^i$:

$$H = \begin{pmatrix} h_0 & h_1 & \dots & h_{n-1} \\ h_{n-1} & h_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & h_1 \\ h_1 & \dots & h_{n-1} & h_0 \end{pmatrix}.$$

The lattice Λ_h contains by definition the following set of n secret short vectors $\mathcal{S}_h = \{(x^i * g, x^i * f), 0 \leq i \leq n-1\}$ formed by the secret vector (g, f) and its $n-1$ rotations.

NTRU variants. Variants of the original NTRU [31] choose to use different polynomial rings $\mathcal{R} = \mathbb{Z}[X]/P(X)$, for a unitary degree n polynomial $P \in \mathbb{Z}[X]$. Without giving an exhaustive list, examples of cryptosystems that use such variants include NTRU Prime [10], NTRU+ [34], as well as the Falcon signature scheme [24]. In these cases, the public key $h \in \mathcal{R}/q\mathcal{R}$ is also defined as $h = g/f \mod q$, where $(f, g) \in \mathcal{R}^2$ is the secret key. In the most general case, the NTRU lattice is obtained by embedding the rank 2 $\mathcal{R}$-module that we call the *NTRU module*

$$M_h := \{(u, v) \in \mathcal{R}^2 : hu \equiv v \mod q\mathcal{R}\}$$

into $\mathbb{C}^{2n}$ via an embedding map $\sigma : \mathcal{R} \to \mathbb{C}^n$. The secret key is usually of small norm after embedding, that is $\|(\sigma(g), \sigma(f))\|$ is small. Most commonly, as is the case in the aforementioned cryptosystems, σ is simply the *coefficient embedding*. It has the advantage of being simple as it is easy to implement, as its image is integral. Other embeddings can also be of cryptanalytic interest. Most notably, the *canonical embedding* is obtained by evaluating a polynomial of $\mathcal{R}$ at all complex roots of P. This embedding is more complicated to deal with on a computer, but is a ring homomorphism and therefore behaves well with multiplication, which is usually not the case with the coefficient embedding. In particular if P is irreducible, then $\mathcal{R}$ is the ring of integers of a number field and the *canonical embedding* coincides with the *Minkowski embedding*.

3 Blockwise Reduction of Near-Hypercubic Lattices

In this section, we describe our reduction algorithm, and specialise its analysis to NTRU and hypercubic lattices.

Algorithm 1 Primal/dual reduction with blocksize of halved dimension

Input: A basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of a lattice $\Lambda \subseteq \mathbb{Z}^m$, together with two upper bounds r and $r^\times$ such that $\lambda_1(L) \leq r$ and $\lambda_1(L^\times) \leq r^\times$. L (resp. N) is the sublattice spanned by the first $\lfloor n/2 \rfloor$ (resp. $\lfloor n/2 \rfloor + 1$) vectors of B, *i.e.* $L = \mathcal{L}(\mathbf{b}_1, \ldots, \mathbf{b}_{\lfloor n/2 \rfloor})$. Keep in mind that L and N are updated naturally as B evolves.

Output: A short non-zero vector in Λ of norm $\leq r$ or a short non-zero vector in the dual $\Lambda^\times$ of norm $\leq r^\times$, or a basis B such that $\mathrm{vol}(L)$ is guaranteed to be small.

1: LLL-reduce B.
2: **while** $\mathrm{vol}(L)$ strictly decreases **do**
3: $\quad \mathbf{e} \leftarrow$ SVP-oracle(L) to check for short primal lattice vectors.
4: $\quad$ **if** $\|\mathbf{e}\| \leq r$ **then**
5: $\quad\quad$ Return $\mathbf{e}$.
6: $\quad$ **else**
7: $\quad\quad$ SVP-reduce(Λ/L) to reduce the second half of B modulo its first half.
8: $\quad$ **end if**
9: $\quad \mathbf{e}' \leftarrow$ SVP-oracle$(\Lambda^\times \cap \mathrm{span}(N)^\perp)$ to check for short dual lattice vectors.
10: $\quad$ **if** $\|\mathbf{e}'\| \leq r^\times$ **then**
11: $\quad\quad$ Return $\mathbf{e}'$.
12: $\quad$ **else**
13: $\quad\quad$ SVP-reduce$(N^\times)$ to dual-reduce the first half of B: this is DSVP-reduction of the lattice N.
14: $\quad$ **end if**
15: **end while**
16: Return B.

3.1 Provable Algorithm

Alg. 1 can be viewed as a variant of Gama-Nguyen's slide algorithm [27] and Ducas' algorithm [17, Alg. 1]. However, in differs in a few ways, mainly:

- It is not specialised to $\mathbb{Z}^n$.
- The termination conditions are different: instead of uniquely focusing on the reduction task, our algorithm can also check for the presence of short vectors in the lattice Λ or its dual $\Lambda^\times$. Indeed, the tests at Lines 4 and 10 are parametrised by values r and $r^\times$, which will be specified in the case of NTRU lattices and hypercubic lattices. If the user knows that Λ and/or $\Lambda^\times$ contains a short vector of a prescribed length, then he can change the values of r and $r^\times$ accordingly, for example by setting $r = \lambda_1(\Lambda)$ and/or $r^\times = \lambda_1(\Lambda^\times)$ when the first minima are known.
- Lines 3 and 9 add an extra call to the SVP oracle, which provides a way to prematurely abort if the objective is to find a vector of Λ and/or $\Lambda^\times$ of norm less than a fixed value. This is especially useful in the case of NTRU and hypercubic lattices where the first minimum is well-known.
- Unlike [27,17], our algorithm assumes no requirement on the parity of n.

We make an important remark on Alg. 1, which explains why we view this reduction as a primal/dual reduction: Steps 9-14 are dual to Steps 3-8, in the

sense that they are exactly Steps 3-8 if we replace the lattice Λ by its dual $\Lambda^\times$, and the sublattice L by $\Lambda^\times \cap \mathrm{span}(N)^\perp$.

The efficiency of the algorithm is based on the following key elementary result:

Lemma 1. *Assume that $\Lambda \subseteq \mathbb{Z}^m$. During a loop iteration, the sublattice L (at the beginning of a loop iteration) is transformed into L', after Step 14. Then:*

$$\frac{\mathrm{vol}(L')}{\mathrm{vol}(L)} = \lambda_1(\Lambda/L)\lambda_1(N^\times), \tag{2}$$

where N is from Step 13. Furthermore, if the exact reduction oracles of Steps 7 and 13 are replaced by approximate-reduction with factor respectively γ and γ', then:

$$\frac{\mathrm{vol}(L')}{\mathrm{vol}(L)} \leq \gamma\gamma'\lambda_1(\Lambda/L)\lambda_1(N^\times). \tag{3}$$

Proof. The sublattice L can only be changed by Step 13, which cannot change the sublattice N. Since $\mathrm{vol}(N) = \mathrm{vol}(L)\|\mathbf{b}_{k+1}^\star\|$, we are interested in $\|\mathbf{b}_{k+1}^\star\|$, which can only be changed by Steps 7 and 13. After Step 7, we have $\|\mathbf{b}_{k+1}^\star\| = \lambda_1(\Lambda/L)$. After Step 13, we have $1/\|\mathbf{b}_{k+1}^\star\| = \lambda_1(N^\times)$. So $\|\mathbf{b}_{k+1}^\star\|$ changes from $\lambda_1(\Lambda/L)$ to $1/\lambda_1(N^\times)$, which proves (2). The inequality (3) follows from the definition of approximate reduction.

□

Theorem 1. *Let $\Lambda \subseteq \mathbb{Z}^m$ be a rank n lattice. Assume that $\lambda_1(\Lambda)\lambda_1(\Lambda^\times) < 1-\varepsilon$ for some $\varepsilon = \frac{1}{\mathrm{poly}(n)}$. Then Alg. 1 returns a non-zero vector of Λ with norm $\leq r$, or a non-zero vector of its dual $\Lambda^\times$ with norm $\leq r^\times$. The number of loop iterations from Step 3 till Step 14 is polynomial in the size of the input basis B and $1/\varepsilon$. The number of SVP oracle queries is linear in the number of loop iterations, and the dimension of the lattice in each oracle query is $\leq \lfloor n/2 \rfloor + 1$.*

Proof. $\Lambda \subseteq \mathbb{Z}^m$ implies that $\mathrm{vol}(L)^2 \in \mathbb{Z}$. $\log \mathrm{vol}(L)$ is polynomially bounded by the size of the input basis B, and can only decrease with LLL reduction (Step 1). This means that the number of times $\mathrm{vol}(L)$ decreases by $1-\varepsilon$ is polynomially bounded by the size of the input basis B and $1/\varepsilon$.

If $\|\mathbf{e}\| > r \geq \lambda_1(\Lambda)$, there exists $\mathbf{u} \in \Lambda$ such that $\|\mathbf{u}\| = \lambda_1(\Lambda)$ and $\mathbf{u} \notin L$, therefore $\lambda_1(\Lambda/L) \leq \|\mathbf{u}\| = \lambda_1(L)$. Similarly, if $\|\mathbf{e}'\| > r^\times \geq \lambda_1(\Lambda^\times)$, then $\lambda_1(N^\times) \leq \lambda_1(\Lambda^\times)$. Thus, if both $\|\mathbf{e}\| > r$ and $\|\mathbf{e}'\| > r^\times$, then using our assumption, $\lambda_1(\Lambda/L)\lambda_1(N^\times) \leq \lambda_1(\Lambda)\lambda_1(\Lambda^\times) < 1-\varepsilon$. This implies by Lem. 1 that $\mathrm{vol}(L)$ decreases by at least $1-\varepsilon$, which can only happen polynomially many times.

Thus, we will find, within polynomially many iterations, some $\mathbf{e} \in L \subseteq \Lambda$ such that $\|\mathbf{e}\| \leq r$ or some $\mathbf{e} \in N^\times \subseteq \Lambda^\times$ such that $\|\mathbf{e}'\| \leq r^\times$.

By definition, each loop iteration makes four calls to an SVP oracle, and the underlying lattice has rank $\in \{\lfloor n/2 \rfloor, n - \lfloor n/2 \rfloor, n - \lfloor n/2 \rfloor - 1, \lfloor n/2 \rfloor + 1\}$.

□

3.2 Application to NTRU and Falcon

In 2006, Gama, Howgrave-Graham and Nguyen [26] showed that coordinate embedding NTRU lattices from the ring $\mathbb{Z}[X]/(X^n - 1)$ are proportional to symplectic lattices, which is a special case of isodual lattices, *i.e.* there is an isometry between the lattice and its dual. We derive the following property of NTRU lattices:

Theorem 2. *Let $\mathcal{R}$ be $\mathbb{Z}[X]/(X^n - 1)$ or $\mathbb{Z}[X]/(X^n + 1)$. Let $(f, g) \in \mathcal{R}^2$ be an NTRU secret key corresponding to parameters (q, n) and a lattice Λ obtained from the coefficient embedding. Then there is an explicit bijection $\phi : \Lambda \to q\Lambda^\times$ which preserves the Euclidean norm, and which can be computed in polynomial time (in both directions). In particular,*

$$\lambda_1(\Lambda^\times) = \frac{\lambda_1(\Lambda)}{q},$$

where $\lambda_1(\Lambda)^2 \leq \|\mathbf{f}\|^2 + \|\mathbf{g}\|^2$, where $(\mathbf{f}, \mathbf{g})$ is the coefficient embedding of (f, g).

Proof. Using row notation, it is not hard to show that Λ and $\Lambda^\times$ are respectively generated by the bases B_Λ and $B_{\Lambda^\times}$, where

$$B_\Lambda = \begin{pmatrix} qI_n & 0 \\ H & I_n \end{pmatrix} \text{ and } B_{\Lambda^\times} = \begin{pmatrix} \frac{1}{q}I_n & -\frac{1}{q}H^T \\ 0 & I_n \end{pmatrix},$$

where H is circulant (resp. anti-circulant) in the coefficients of $h \in \mathcal{R}$ the public key corresponding to (f, g) if $\mathcal{R} = \mathbb{Z}[X]/(X^n - 1)$ (resp. $\mathcal{R} = \mathbb{Z}[X]/(X^n + 1)$). We claim that

$$\phi : \begin{cases} \Lambda & \to q\Lambda^\times \\ (\mathbf{u}, \mathbf{v}) & \mapsto (\tilde{\mathbf{v}}, -\tilde{\mathbf{u}}) \end{cases}$$

is the desired isometry, where $\tilde{\mathbf{u}}$ is $\mathbf{u}$ in reverse order. Indeed, because of the circulant or anti-circulant nature of H, the i-th row of h is exactly the same as the $(n + 1 - i)$-th row of H^T in reverse order. The structure of $B_{\Lambda^\times}$ relatively to B_Λ allows us to conclude that ϕ is a suitable candidate. This map ϕ can clearly be computed in polynomial time, in both directions. Finally, the inequality $\lambda_1(\Lambda)^2 \leq \|\mathbf{f}\|^2 + \|\mathbf{g}\|^2$ follows from the fact that (g, f) is a lattice vector.

□

Thus, we can upper bound $\lambda_1(\Lambda^\times)\lambda_1(\Lambda)$ by $\frac{1}{q}(\|\mathbf{f}\|^2 + \|\mathbf{g}\|^2)$. Tab. 1 gives the explicit value of this upper bound for three types of NTRU lattices: the ones of the NTRU submission to NIST [12], the original NTRU cryptosystem [31], and the NIST signature standard Falcon [24]. These three types differ from the distribution used for f and g:

– For the first two, f and g have ternary coefficients $\in \{0, \pm 1\}$ but the number of ± 1 differ for each type.

– For Falcon however, f and g no longer have ternary coefficients: instead, its coefficients follow a discrete Gaussian distribution. We used publicly-available key generation software to compute the typical value of $\|\mathbf{f}\|^2+\|\mathbf{g}\|^2$.

In addition, all of these examples use the coefficient embedding version of NTRU. The first two use the ring $\mathbb{Z}[X]/(X^n-1)$, and the third uses $\mathbb{Z}[X]/(X^n+1)$, both of which fall into the scope of Theorem 2.

Upper bound on $\lambda_1(L)\lambda_1(L^\times)$ for various NTRU parameters						
Lattice	N	q	$\|(\mathbf{f},\mathbf{g})\|^2$	$\lambda_1(L)\lambda_1(L^\times)$	$\frac{1}{2}\lambda_1(L)\lambda_1(L^\times)$	Approx factor
NIST-1 [12]	509	2048	593	.2897	.1449	2.628
NIST-2 [12]	677	2048	705	.3444	.1722	2.410
NIST-3 [12]	821	2048	1057	.2581	.1291	1.969
Original toy [31]	107	64	53	.8281	.4141	1.554
Original [31]	167	128	161	1.258	.6289	1.261
	263	128	147	1.148	.5742	1.320
	503	256	575	2.246	1.123	N/A
Falcon-512 [24]	512	12889	16481	1.341	.6706	1.251
Falcon-1024 [24]	1024	12889	16487	1.342	.6708	1.250

Table 1: NTRU parameters: the two filled-in columns determine whether Th. 1 applies, theoretically or heuristically. The last column illustrates by how much we can relax the SVP-reduction used Steps 7 and 13 of Alg. 1. When $\|(\mathbf{f},\mathbf{g})\|^2$ is not fixed by the specifications, we take the experimental median over 1000 instances.

In Tab. 1, the green colour indicates that the upper bound is $< 1-\varepsilon$ for some constant $\varepsilon > 0$, which makes Th. 1 applicable: this is the case for all parameter sets of NTRU submission to NIST [12], and for the toy parameter set of the original NTRU [31]. If we run Alg. 1 with input $r^2 = \|\mathbf{f}\|^2 + \|\mathbf{g}\|^2$ and $r^\times = \frac{1}{q}\sqrt{\|\mathbf{f}\|^2 + \|\mathbf{g}\|^2}$ (where the exact value may be replaced by a good upper bound): this will return a nonzero vector in the primal lattice at least as short as the secret key, using only an SVP oracle in halved dimension. Indeed, if ever a dual vector is returned, the isometry of Th. 2 allows to transform the short dual vector into a short primal vector. Bare in mind that it is believed that the secret-key vectors are the shortest vectors of the NTRU lattice, but this has not been proved.

We explain the situation of the NTRU submission to NIST [12]. To avoid decryption failures, the generation of $\mathbf{f}$ and $\mathbf{g}$ is such that $\|\mathbf{f}\|^2+\|\mathbf{g}\|^2 \leq q/2$. In

fact, we have $\|\mathbf{f}\|^2 \leq N$ and $\|\mathbf{g}\|^2 = q/8 - 2$. Thus:

$$\lambda_1(\Lambda^\times)\lambda_1(\Lambda) \leq \frac{1}{q}(\|\mathbf{f}\|^2 + \|\mathbf{g}\|^2) \leq \frac{1}{2}.$$

On the other hand, the historical parameters of NTRU allowed decryption failures, which increased $\|\mathbf{f}\|$ and $\|\mathbf{g}\|$.

The red colour in Tab. 1 shows that the bound is not satisfied. However, there is a way to get around this issue, under a mild assumption, except for the largest parameter of original NTRU [31]. Indeed, Th. 1 uses an upper bound on $\lambda_1(\Lambda)\lambda_1(\Lambda^\times)$ to actually upper bound $\lambda_1(\Lambda/L)$ and $\lambda_1(N^\times)$, knowing that none of the n short vectors $\mathbf{s}_1, \ldots, \mathbf{s}_n$ related to the secret key, obtained by coefficient embedding of the $(x^i * f, x^i * g)$, belong to the sublattice L (and similarly for the dual, with respect to $N^\times$). It follows that $\lambda_1(\Lambda/L) \leq \min_{1\leq i\leq n} \|\pi(\mathbf{s}_i)\|$, where π denotes the orthogonal projection over $\mathrm{span}(L)^\perp$. If $\mathrm{span}(L)^\perp$ was a random subspace, the expectation of $\|\pi(\mathbf{s}_i)\|^2$ would be $\|\mathbf{s}_i\|^2 \frac{1}{n} \dim \mathrm{span}(L)^\perp \approx \|\mathbf{s}_i\|^2 \frac{1}{2}$. This suggests to make the mild assumption that:

$$\lambda_1(\Lambda/L)\lambda_1(N^\times) \leq \frac{\lambda_1(\Lambda)\lambda_1(\Lambda^\times)}{2}.$$

If this assumption holds at each loop iteration, then the conclusions of Th. 1 still hold: we will obtain a nonzero vector in the primal lattice at least as short as the secret key. The second to last column of Tab. 1 therefore shows an upper bound of $\frac{1}{2}\lambda_1(\Lambda)\lambda_1(\Lambda^\times)$: it turns out that the upper bound is now always $\ll 1$, except for the largest parameter of original NTRU [31]. If this product is $\ll 1$, then we can heuristically relax the SVP-reductions used in Steps 7 and 13 of Alg. 1 to approximate-SVP-reductions with approximation factor $\ll \sqrt{\frac{2}{\lambda_1(\Lambda)\lambda_1(\Lambda^\times)}}$. The rightmost column of Tab. 1 provides explicit values of the best approximation factors.

To summarise, Alg. 1 provably returns a nonzero lattice vector at least as short as the secret key for all parameter sets of NTRU submission to NIST [12], using only an SVP oracle in halved dimension. And it succeeds heuristically under a mild assumption, for Falcon [24] and all parameter sets of original NTRU [31] except for one. This gives a positive answer to the conjecture of Gama *et al.* [26]: the reduction of a $2n$-dimensional NTRU lattice can be reduced to that of a n-dimensional lattice[5]. In addition, half of the oracle calls of our algorithm still work with approximate reduction, up to constant approximation factors that increase as $\lambda_1(\Lambda)\lambda_1(\Lambda^\times)$ decreases.

We provide an additional result regarding the isodual nature of the NTRU modules, which we believe can be of independent cryptanalytic interest.

Theorem 3. *Any NTRU module is isomorphic to its dual module. Additionally, the canonical embedding NTRU lattice is isometric up to a scalar factor to its dual lattice.*

[5] In this sentence *reduction* and *reduced* have different meanings.

Proof. Let $\mathcal{R} = \mathbb{Z}[X]/P(X)$ for some unitary degree n polynomial $P \in \mathbb{Z}[X]$. Let $h \in \mathcal{R}$ and M_h be a NTRU module as defined in Sec. 2:

$$M_h := \{(u, v) \in \mathcal{R}^2 : hu \equiv v \mod q\mathcal{R}\}.$$

The dual module $M_h^\times$ is defined as the set of module homomorphisms from M_h to $\mathcal{R}$. We have

$$M_h^\times = \{(\alpha, \beta) \in (\mathbb{Q}[X]/P(X))^2 : \forall (u, v) \in M_h, \alpha u + \beta v \in \mathcal{R}\}.$$

Let $(\alpha, \beta) \in M_h^\times$. Observe that $(0, q) \in M_h$. Therefore $q\beta \in \mathcal{R}$, and there exists $\beta' \in \mathcal{R}$ such that $\beta = \frac{1}{q}\beta'$. Now observe that $(1, h) \in M_h$. This gives $\alpha + \frac{1}{q}\beta' h \in \mathcal{R}$, from which we deduce that there also exists $\alpha' \in \mathcal{R}$ such that $\alpha = \frac{1}{q}\alpha'$, and $\frac{1}{q}(\alpha' + \beta' h) \in \mathcal{R}$. Let

$$L_h := \{(x, y) \in \mathcal{R}^2 : hy \equiv -x \mod q\mathcal{R}\},$$

then $(\alpha', \beta') \in L_h$, therefore $qM_h^\times \subseteq L_h$. But clearly L_h and M_h are isomorphic via the map $\psi : (x, y) \mapsto (y, -x)$, so by examining the index of $qM_h^\times$ in M_h we can conclude that M_h and $M_h^\times$ are isomorphic via the map $\frac{1}{q}\psi$. Because the canonical embedding is a ring homomorphism, the second part of the statement follows directly from the shape of the isomorphism.

□

Th. 3 essentially says that any NTRU lattice can be turned in a symplectic version of itself by a simple change of embedding. Note that this isn't a generalisation of Th. 2.

3.3 Reducing Hypercubic Lattices with Approximate-SVP Oracles

In this subsection, we specialise Alg. 1 to the case of $\mathbb{Z}^n$, and allow to relax the exact-SVP oracle into an approximate-SVP oracle: Ducas [17] was only able to relax his oracle for an approximation factor very close to 1, while we allow an approximation factor close to $\sqrt{2}$. Our improvement also leads to a speculative improvement over the $2^{n/2}$ running time, if approximating SVP to within a factor 2 is exponentially faster than solving SVP.

We first present our specialised algorithm: Alg. 2 is basically Alg. 1 with $r = r^\times = 1$ and approximate oracles instead of exact oracles, with a different termination: since we want to obtain an orthonormal basis, we don't stop once a unit vector has been found, we reduce the dimension of Λ by projection, and keep iterating Alg. 1 until the rank becomes trivial.

Algorithm 2 An algorithm for ZLIP with approximate-SVP oracles in dimension $n/2$.

Input: An approximation factor $\gamma \in [1, \sqrt{2-2/n})$. A basis B of $\Lambda \simeq \mathbb{Z}^n$. L (resp. N) is the lattice spanned by the first $\lfloor n/2 \rfloor$ (resp. $\lfloor n/2 \rfloor + 1$) vectors of B.
Output: O an orthonormal basis of Λ.

$O = \{\}$
LLL-reduce B
while $\dim(B) > 0$ **do**
  if γ-SVP-oracle(L) returns a vector $\mathbf{e}$ such that $\|\mathbf{e}\| = 1$ **then**
    $O \leftarrow O \cup \{\mathbf{e}\}$.
    $B \leftarrow \mathrm{LLL}(\pi_{\mathbf{e}^\perp}(B))$ (update L and N accordingly).
  else
    γ-SVP-reduction-oracle(Λ/L) to reduce the second half of B modulo its first half.
  end if
  if γ-SVP-oracle($(\Lambda^\times/N)^\times$) returns a vector $\mathbf{e}'$ such that $\|\mathbf{e}'\| = 1$ **then**
    $O \leftarrow O \cup \{\mathbf{e}'\}$.
    $B \leftarrow \mathrm{LLL}(\pi_{\mathbf{e}'^\perp}(B))$ (update L and N accordingly).
  else
    γ-SVP-reduction-oracle($N^\times$) to dual-reduce the first half of B.
  end if
end while
Return O.

The main result in this subsection is the following:

Theorem 4. *Given as input a basis B of $\Lambda \simeq \mathbb{Z}^n$ and given access to a γ-SVP approximation oracle in dimension $\lfloor n/2 \rfloor + 1$ where $\gamma \in \left[1, \sqrt{2-\frac{2}{n}}\right)$, Alg. 2 returns an orthonormal basis of Λ in polynomial time.*

We briefly compare Alg. 2 with Ducas's algorithm [17]. Ducas's algorithm restricts to a hypercubic lattice of odd dimension: the algorithm keeps reducing until the "half-sublattice" L (the sublattice generated by the first half of the current basis) generates a hypercubic lattice. Instead, Alg. 2 checks using an approximate SVP oracle whether the "half-sublattice" L or its dual counterpart contains a unit vector: if not, the first minimum of L is > 1, which allows us to better upper bound the first minimum of Λ/L or its dual counterpart, compared to [17, Lem. 4]. If ever a unit vector is discovered, we can decrement the lattice rank by projection, which also means that our algorithm must not be sensitive to the parity of the rank. The key to our improvement is the following technical result on random projections, which might be of independent interest.

Projecting an orthonormal basis. It is well-known that the expectation of the squared norm of the projection of a unit vector onto a k-dimensional random subspace of $\mathbb{R}^n$ is $\frac{k}{n}$. The following elementary lemma shows that the expectation

of the squared norm of the projection of a random element of a fixed orthonormal basis of $\mathbb{R}^n$ onto a fixed k-dimensional subspace is also $\frac{k}{n}$.

Lemma 2. *Let $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ be an orthonormal basis of $\mathbb{R}^n$. Let π be the orthogonal projection over a k-dimensional subspace F of $\mathbb{R}^n$. Then:*

$$\sum_{i=1}^{n} \|\pi(\mathbf{e}_i)\|^2 = k.$$

Proof. Let $(\mathbf{f}_1, \ldots, \mathbf{f}_k)$ be an orthonormal basis of F. Then for each $1 \leq i \leq n$:

$$\|\pi(\mathbf{e}_i)\|^2 = \sum_{j=1}^{k} \langle \mathbf{e}_i, \mathbf{f}_j \rangle^2.$$

Therefore:

$$\sum_{i=1}^{n} \|\pi(\mathbf{e}_i)\|^2 = \sum_{i=1}^{n} \sum_{j=1}^{k} \langle \mathbf{e}_i, \mathbf{f}_j \rangle^2 = \sum_{j=1}^{k} \sum_{i=1}^{n} \langle \mathbf{e}_i, \mathbf{f}_j \rangle^2 = \sum_{j=1}^{k} 1 = k,$$

because each $\mathbf{f}_j$ is a unit vector and $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is an orthonormal basis of $\mathbb{R}^n$. □

The previous lemma allows us to upper bound the first minimum of the projection of a hypercubic lattice, as follows:

Corollary 1. *Let L be a primitive sublattice of rank $1 \leq k < n$ of a full-rank hypercubic lattice Λ of $\mathbb{R}^n$ such that $\lambda_1(L) \geq \sqrt{2}$. Then $\lambda_1(\Lambda/L)^2 \leq 1 - \frac{k}{n}$.*

Proof. L is primitive so Λ/L is a lattice and $\lambda_1(\Lambda/L)$ is well-defined. Let π be the orthogonal projection onto the $(n-k)$-dimensional subspace $L^\perp$. We know that Λ has an orthonormal basis $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$: this is also an orthonormal basis of $\mathbb{R}^n$ so the lemma shows that

$$\sum_{i=1}^{n} \|\pi(\mathbf{e}_i)\|^2 = n - k$$

Furthermore, note that all the $\pi(\mathbf{e}_i)$'s are nonzero: if $\pi(\mathbf{e}_i) = 0$ for some i, then $\mathbf{e}_i \in L$ because L is primitive, then $\lambda_1(L) \leq 1$, which contradicts $\lambda_1(L) \geq \sqrt{2}$. Therefore there exists an integer $i \in \{1, \ldots, n\}$ such that $0 < \|\pi(\mathbf{e}_i)\|^2 \leq \frac{n-k}{n}$. Hence $\lambda_1(\Lambda/L)^2 \leq 1 - \frac{k}{n}$. □

In other words, under certain conditions over L, we can decrease Ducas [17]'s upper bound $\sqrt{1 - 1/n}$ to $\sqrt{1 - k/n}$, which is better as soon $k \geq 2$: we note that for $k = 1$, Ducas [17]'s upper bound is actually tight for L spanned by the all-one vector $(1, 1, \ldots, 1)$, which means that Λ/L is the dual root lattice $A_{n-1}^{\times}$. We are now ready for the proof of Th. 4, which is very similar to that of Th. 1: we simply combine Lem. 2 with (3) of Lem. 1.

Proof (of Theorem 4). $\Lambda \simeq \mathbb{Z}^n$ implies that $\mathrm{vol}(L)^2 \in \mathbb{Z}$. Because $\mathrm{vol}(\Lambda) = 1$ and well-known properties of LLL reduction, Step 2 guarantees $\log \mathrm{vol}(L) = O(n^2)$. This means that the number of times $\mathrm{vol}(L)$ decreases by a factor $1-\varepsilon$ (without changing Λ) is $O(n^2/\varepsilon)$.

We have $n = 2k$ or $n = 2k+1$ where $k = \lfloor n/2 \rfloor$. We let L be the primitive lattice spanned by $(\mathbf{b}_1, \ldots, \mathbf{b}_k)$.

Consider Step. 4. if $\|\mathbf{e}\| < \sqrt{2}$, then $\|\mathbf{e}\| = 1$ because Λ has no vector of norm in the interval $(1, \sqrt{2})$. So we recovered a shortest vector $\mathbf{e}$ of Λ, and Step. 6 iterates the algorithm, by projecting Λ over the hyperplane orthogonal to $\mathbf{e}$: this is a hypercubic lattice of rank $n-1$, and we have to recompute an LLL-reduced basis.

Otherwise, $\|\mathbf{e}\| \geq \sqrt{2}$. We deduce that $\lambda_1(L) > 1$, as otherwise $\lambda_1(L) = 1$ because $\lambda_1(\Lambda) = 1$, which would contradict $\|\mathbf{e}\| \leq \gamma$. But $\lambda_1(L) > 1$ implies that $\lambda_1(L) \geq \sqrt{2}$ because Λ has no vector of norm in the interval $(1, \sqrt{2})$. So Cor. 1 shows that $\lambda_1(\Lambda/L)^2 \leq 1 - \frac{k}{n}$.

The remaining steps are the dual counter part. So if $\|\mathbf{e}'\| \geq \sqrt{2}$ in Step. 10, we deduce similarly by applying Cor. 1 to the sublattice $\Lambda^\times \cap \mathrm{span}(N)^\perp$ of rank $n-(k+1)$, that $\lambda_1(N^\times)^2 = \lambda_1(\Lambda^\times/(\Lambda^\times \cap \mathrm{span}(N)^\perp))^2 \leq 1 - \frac{n-(k+1)}{n}$. We thus have proved:

$$\lambda_1(\Lambda/L)\lambda_1(N^\times) \leq \sqrt{1-\frac{k}{n}}\sqrt{1-\frac{n-(k+1)}{n}} = \frac{\sqrt{(n-k)(k+1)}}{n}.$$

If $n = 2k$, then:

$$\frac{\sqrt{(n-k)(k+1)}}{n} = \frac{1}{2}\sqrt{1+\frac{2}{n}} = \frac{1}{2}\left(1 + \frac{1}{n} - \frac{1}{2n^2} + O\left(\frac{1}{n^3}\right)\right).$$

Otherwise, $n = 2k+1$ and:

$$\frac{\sqrt{(n-k)(k+1)}}{n} = \frac{k+1}{n} = \frac{1}{2}\left(1 + \frac{1}{n}\right).$$

Since $\gamma^2 < 2 - \frac{2}{n}$, (3) of Lem. 1 implies that, unless we find a unit vector, $\mathrm{vol}(L)$ decreases by at least $\left(1-\frac{1}{n}\right)\left(1+\frac{1}{n}\right) = 1 - \frac{1}{n^2}$. Thus, within polynomially many iterations, we will find a unit vector $\mathbf{e}$ or $\mathbf{e}'$. Since there are only n unit vectors, we find all of them within polynomially many iterations.

□

A consequence of Th. 4 is the following speculative Corollary, that would break the $n/2$ barrier for $\mathbb{Z}$LIP as long as $\sqrt{2}$-approx SVP is exponentially easier than its exact counterpart.

Corollary 2. *Let $\alpha < 1$ be a constant. If there exists an algorithm for approx-SVP with approximation factor $\sqrt{2-2/n}$ that runs in time $2^{\alpha n + o(n)}$, then there also exists an algorithm for $\mathbb{Z}$LIP that runs in time $2^{\alpha n/2 + o(n)}$.*

Aside from being visibly easier in practice, γ-approx SVP has been shown to be exponentially easier than exact SVP for some larger constant approximation factors (Th. 3.2 of [23]), this gives some evidence as to why the premise of Cor. 2 might be true.

4 The Primal Attack on Near-Hypercubic Lattices

In this section, we derive the asymptotic behaviour of the heuristic minimal blocksizes required to break lattice problems such as ZLIP and NTRU. We then tweak the primal attack framework to incorporate the fact that special lattices like the hypercubic and NTRU lattice have not just one, but many shortest vectors. In both cases, the quantity $\mathrm{vol}(L)^{1/n}/\lambda_1(L)$ is a constant, whereas we would expect it to be $\Theta(n^{-1/2})$ for a generic lattice. Our results using the primal attack approach could be considered folklore, but we think it profitable to write them down clearly. They nicely complement Section 3, because they provide an opportunity to compare the best known provable and heuristic reduction algorithms for ZLIP and NTRU.

4.1 Using a Single Short Vector

Proposition 1. *Let $c = \Theta(1)$ be a positive constant. If $\beta = \omega(1)$ satisfies the equation:*

$$\sqrt{\frac{\beta}{n}} = \delta_\beta^{2\beta-n-1}\sqrt{c},$$

then

$$\beta = \frac{n}{2} - \frac{\log(2c)}{4}\frac{n}{\log n} + o\left(\frac{n}{\log n}\right).$$

Proof. All equivalents denote asymptotics as n goes to infinity. Let $0 < \beta < n$ be a solution to the equation for which $\beta = \omega(1)$. Because

$$\frac{\beta}{n} = \left(\frac{\beta}{2\pi e}(\pi\beta)^{\frac{1}{\beta}}\right)^{\frac{2\beta-n-1}{\beta-1}} c \sim \left(\frac{\beta}{2\pi e}\right)^{\frac{2\beta-n-1}{\beta-1}} c,$$

we obtain

$$\left(\frac{\beta}{n}\right)^{\beta-1} \sim \left(\frac{\beta}{2\pi e}\right)^{2\beta-n-1} c^{\beta-1}.$$

It is clear from $\beta = \omega(1)$ and the above expression that $\beta = n/2 + o(n)$. In what follows we write $\beta = (1/2-\varepsilon)n$, where $\varepsilon = o(1)$. We get

$$(1/2-\varepsilon)^{(1/2-\varepsilon)n-1} \sim \left(\frac{(1/2-\varepsilon)n}{2\pi e}\right)^{-2\varepsilon n-1} c^{(1/2-\varepsilon)n-1},$$

and by taking the log of the ratio, we must have

$$((1/2-\varepsilon)n-1)(\log(1/2-\varepsilon)-\log c) + (2\varepsilon n+1)\log\left(\frac{(1/2-\varepsilon)n}{2\pi e}\right) \longrightarrow 0.$$

The dominating terms of the expression above are $2\varepsilon n\log(n)$ and $-\frac{n}{2}\log(2c)$ so they must cancel out, leaving us with $\varepsilon = \frac{\log(2c)}{4\log(n)} + o(\log(n)^{-1})$.

□

Remark 1. We pay no concern to β having to be an integer. We choose to replace the inequality in Eq. 1 by an equality as we are interested in the largest value of β such that the inequality still holds.

Corollary 3. *Let L be a rank n lattice and $\mathbf{s}$ a short vector of L for which $\|\mathbf{s}\|/\mathrm{vol}(L)^{1/n} =: c^{-1/2} = O(1)$. The primal attack framework heuristically predicts that applying BKZ with blocksize $\beta = n(1/2 - \log(2c)/4\log n + o(1/\log n))$ recovers a vector of norm $\|\mathbf{s}\|$ or less with high probability. In particular, this condition holds for hypercubic and NTRU lattices.*

Proof. The main point follows directly from Prop. 1. In the case of hypercubic lattices, $\mathrm{vol}(L) = \|\mathbf{s}\| = 1$. For NTRU lattices, $\mathrm{vol}(L)^{1/n} = \sqrt{q} = \Theta(\sqrt{n})$, and $\|\mathbf{s}\| = \Theta(n)$, where $\mathbf{s} = (\mathbf{g}, \mathbf{f})$ is the secret key and q is the NTRU modulus. □

The authors of the Hawk signature specifications [19,11] use the primal attack to heuristically evaluate the security of their scheme. They obtain from Eq. 1 that the optimal blocksize for secret key recovery is $n/2 + o(n)$. Cor. 3 helps with understanding the hidden contribution.

Remark 2. In the case of NTRU, the Gram-Schmidt norms after reduction behave differently to the GSA because of the presence of q-vectors, and this could influence our primal attack heuristic analysis (See [4] for a more precise discussion of the difference in shape). Because we can always choose to randomise the input bases, it is sound to presume GSA behaviour.

4.2 Using Many Short Vectors

Hypercubic and NTRU lattices have multiple shortest vectors. The primal attack framework as described by Eq. 1 does not take this into account, as it only relies on the expected value of the norm of the projection of a single vector. We only need one projection to be smaller than the expected Gram-Schmidt norm $\|\mathbf{b}^\star_{n-\beta+1}\|$ for the SVP oracle on the last BKZ block of size β to recover said projection. And because the squared norms of the projections onto random subspaces follow Beta distributions, we can estimate the expected value of the minimal projection and slightly lower the blocksize. See Fig. 2 for an illustration. For smaller dimensions, we observe how considering more short vectors improves the double-intersection phenomenon described in [5].

The *Leaky-LWE estimator* has an option to account for the presence of multiple shortest vectors, however this option is not discussed in detail in [16]. Our new framework (although the same in spirit), addresses this issue differently, offering asymptotic insights as well as specifically isolating the impact of this condition on the blocksize.

In the literature on the primal attack, authors have never used any special property of the Beta distribution other than its mean. The authors of [16] use a probabilistic model in which the squared norms of the projections are approximated using a χ^2 distribution. Even though the χ^2 and the Beta distributions

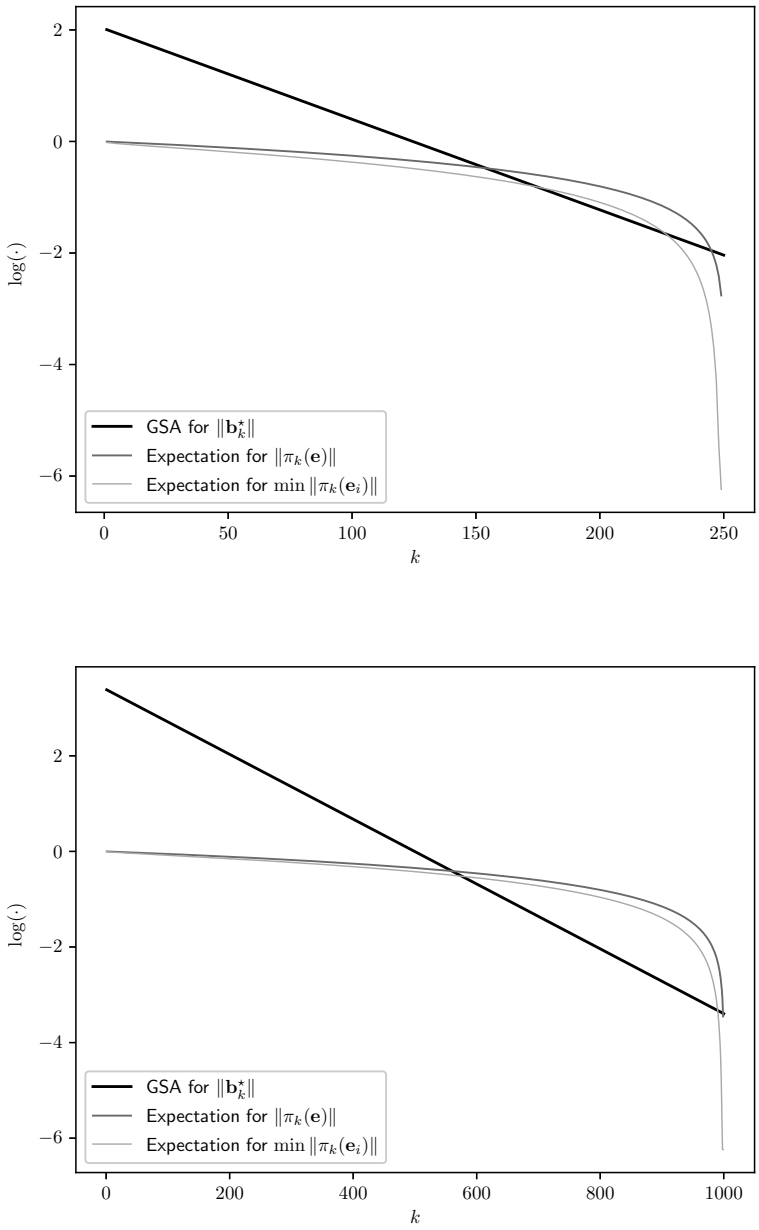


Fig. 2: Comparing the expected norms of randomised Gram-Schmidt vectors of a basis of $\mathbb{Z}^n$ after BKZ reduction with blocksize $n/2$ with the expected projection norms of one and n unit vectors. $n = 250$ above and $n = 1000$ below.

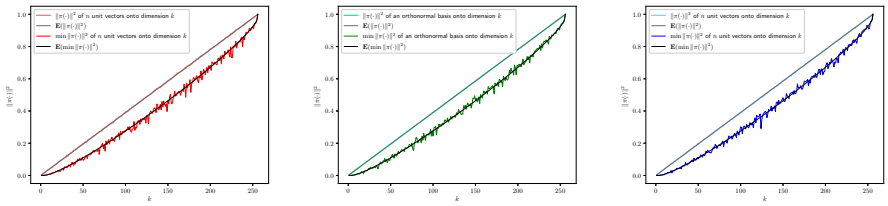
are very good approximations of each other in the small-β context, the difference might become more noticeable for larger blocksizes, so to correct this we choose to work with Beta distributions instead.

We want to emphasise that our framework is not intended for practical use or to supplant existing work. Instead, its purpose is to enhance our comprehension of the components involved in the primary attack. When compared to [16] it simplifies the situation greatly by not taking into account lifting probabilities, or even more precise Gram-Schmidt norm estimates. It also ignores possible fluctuations in the value of $\|\mathbf{b}^\star_{n-\beta+1}\|$. Estimations for hypercubic lattices obtained by both frameworks are compared in Fig. 4.

To estimate the expectation of the minimal norm of the projections, we use the following heuristic.

Heuristic 1 *Let $0 < k < n$. If a lattice L of rank n contains N vectors $\mathbf{s}_1, \ldots, \mathbf{s}_N$ of equal norm r, then the random variables defined by the squared norms of their projections onto a random dimension k subspace of $\mathbb{R}^n$ are independent.*

Heu. 1 is very close to heuristics used in the study of the dual attack [20]. We argue that when N is not too large (we only use $N \leq n$), this heuristic is reasonable for our purposes. See Fig. 3 for a comparison of the average minimal squared norms of the projections of shortest vectors onto random subspaces in the cases of a random set of unit vectors, an orthonormal basis of $\mathbb{R}^n$, and a cirulant set of n vectors.



(a) Random (b) Orthonormal (c) Structured

Fig. 3: Comparing average/minimal norms of projections π of sets of n unit vectors onto random k-dimensional subspaces of $\mathbb{R}^n$. $n = 256$, and k ranges from 0 to n. Theoretical expected values are plotted in black. The sets considered are random on the sphere (3a), orthonormal basis (3b) and structured: all cyclic permutations of a normalised NTRU-like secret vector (3c). Each point correspond to a single random choice of vectors as well as a single random choice of subspace.

Lemma 3. *Let $0 < k < n$. Let $\mathbf{s}_1, \ldots, \mathbf{s}_N$ be vectors of norm r in a lattice that satisfies Heu. 1. Then*

$$\mathbb{E}\left(\min_{1\leq i\leq N} \|\pi(\mathbf{s}_i)\|^2\right) = r^2 \int_0^1 \left(1 - I_x\left(\frac{k}{2}, \frac{n-k}{2}\right)\right)^N dx,$$

where π is the projection onto a random dimension k subspace of $\mathbb{R}^n$, and I is the regularised incomplete beta function.

Proof. All of the $\|\pi(\mathbf{s}_i)\|^2/r^2$ follow the Beta distribution $B(k/2, (n-k)/2)$. Let f denote its probability density function (pdf), and F the associated cumulative distribution function (cdf). Then by independence, the pdf $f_{\min}$ of $\min_{1\leq i\leq N} \|\pi(\mathbf{s}_i)\|^2/r^2$ satisfies $f_{\min}(x) = N(1-F(x))^{N-1}f(x)$. It follows that

$$\mathbb{E}\left(\min_{1\leq i\leq N} \|\pi(\mathbf{s}_i)\|^2/r^2\right) = \int_0^1 x f_{\min}(x)dx = \int_0^1 (1-F(x))^N dx,$$

where we used integration by parts. We conclude using the fact that the cdf of the beta function is equal to the regularised incomplete beta function.

□

While Lem. 3 can be quite practical, we prefer to work with a slightly different quantity that is easier to manipulate.

Lemma 4. *Let $\tau > 0$. Let $0 < k < n$ and π a projection onto a random dimension k subspace of $\mathbb{R}^n$. Let $\mathbf{s}_1, \ldots, \mathbf{s}_N$ be vectors of norm r in a lattice that satisfies Heu. 1. Then*

$$\mathbb{P}\left(\min_{1\leq i\leq N} \|\pi(\mathbf{s}_i)\| < r\tau\right) = 1 - \left(1 - I_{\tau^2}\left(\frac{k}{2}, \frac{n-k}{2}\right)\right)^N$$

Proof. By independence,

$$\mathbb{P}\left(\min_{1\leq i\leq N} \|\pi(\mathbf{s}_i)\| < r\tau\right) = 1 - \prod_{i=1}^{N} \mathbb{P}\left(\|\pi(\mathbf{s}_i)\|^2 \geq r^2\tau^2\right).$$

All of the $\|\pi(\mathbf{s}_i)\|^2/r^2$ follow the Beta distribution of parameters $k/2, (n-k)/2$. Each term of the product is exactly the complement to 1 of the cdf of the previous beta function evaluated at τ^2. We conclude by definition of $I_x(a,b)$.

□

In our study we consider blocksizes that are fractions of n. For this reason we will use the notation $\beta = \alpha n$, where $\alpha \in [0,1]$. Again, we are interested in asymptotic behaviours as n goes to infinity, which means we do not care if β is not integral. In order to get anything meaningful from Lem. 4, we need a precise estimate of $I_x\left(\frac{\alpha n}{2}, \frac{(1-\alpha)n}{2}\right)$. For this we use a result by Temme [47].

Lemma 5 (Derived from [47], Sec. 3). *Let $\varepsilon > 0$, $x \in (0,1)$ and $\alpha \in (\varepsilon, 1-\varepsilon)$. Then*

$$I_x\left(\frac{\alpha n}{2}, \frac{(1-\alpha)n}{2}\right) = \frac{1}{2}\mathrm{erfc}\left(-\frac{\eta\sqrt{n}}{2}\right) + o\left(\mathrm{erfc}\left(-\frac{\eta\sqrt{n}}{2}\right)\right),$$

where $\eta = \mathrm{sign}(x-\alpha)\sqrt{-2\alpha\log\left(\frac{x}{\alpha}\right) - 2(1-\alpha)\log\left(\frac{1-x}{1-\alpha}\right)}$, *and* $\mathrm{erfc} = 1 - \mathrm{erf}$ *is the* complementary error function.

Proof. We are in the second case studied by [47], where $a = \frac{\alpha n}{2}$ and $b = \frac{(1-\alpha)n}{2}$ are such that $a + b = \frac{n}{2} \to \infty$, and both ratios $\frac{a}{b} = \frac{\alpha}{1-\alpha}$ and $\frac{b}{a} = \frac{1-\alpha}{\alpha}$ are bounded away from 0. The Lemma follows directly from Eq. (3.9) in [47]. □

Lem. 5 begs the question: how big can η get? By deriving the asymptotic behaviour of η, we can deduce the asymptotic blocksize required by our variant of the primal attack.

Proposition 2. *Let $0 < p < 1$ be a fixed constant probability. Let $0 < \varepsilon < 1$ and $\varepsilon < \alpha < 1 - \varepsilon$ be a function of n. Let $\pi_{n-\alpha n+1}$ be a projection onto a random dimension αn subspace of $\mathbb{R}^n$. Let $\mathbf{s}_1, \ldots, \mathbf{s}_N$ be $\Theta(n)$ vectors of norm r in a lattice L that satisfies Heu. 1. Suppose also that $c := \mathrm{vol}(L)^{1/n}/r = \Theta(1)$. Then if the asymptotic identity*

$$\mathbb{P}\left(\min_{1\le i\le N} \|\pi_{n-\alpha n+1}(\mathbf{s}_i)\| < \delta_{\alpha n}^{(2\alpha-1)n-1}\sqrt{c}\right) = p + o(1) \tag{4}$$

holds, then

$$\beta := \alpha n = \frac{n}{2} - \frac{\log(2c)}{4}\frac{n}{\log(n)} + o\left(\frac{n}{\log n}\right).$$

Proof. By Lem. 4 with $\tau = \delta_{\alpha n}^{(2\alpha-1)n-1}\sqrt{c}$,

$$\mathbb{P}\left(\min_{1\le i\le N} \|\pi_{n-\alpha n+1}(\mathbf{s}_i)\| < \delta_{\alpha n}^{(2\alpha-1)n-1}\sqrt{c}\right) = 1 - \left(1 - I_x\left(\frac{\alpha n}{2}, \frac{(1-\alpha)n}{2}\right)\right)^N,$$

where $x = \delta_{\alpha n}^{2((2\alpha-1)n-1)}c$ therefore it would suffice to prove that

$$\log\left(1 - I_x\left(\frac{\alpha n}{2}, \frac{(1-\alpha)n}{2}\right)\right) \sim \frac{\log p}{N}. \tag{5}$$

Letting $\eta = \mathrm{sign}(x-\alpha)\sqrt{-2\alpha\log\left(\frac{x}{\alpha}\right) - 2(1-\alpha)\log\left(\frac{1-x}{1-\alpha}\right)}$ as in Lem. 5 and combining the result of this same Lemma with Eq. 5, we get

$$-\frac{\log p}{n} \sim I_x\left(\frac{\alpha n}{2}, \frac{(1-\alpha)n}{2}\right) \sim \frac{1}{2}\mathrm{erfc}\left(-\frac{\eta\sqrt{n}}{2}\right).$$

This yields $x < \alpha$ and $\eta^2 \sim 4\frac{\log n}{n}$ (we used that $N = \Theta(n)$ and the following estimate for large u: $\mathrm{erfc}(u) \sim \pi^{-1/2}u^{-1}e^{-u^2}$. See also [42] for an alternative method). To conclude we look for the most important terms inside of η^2. Looking at

$$4\frac{\log n}{n} \sim 2\alpha \log \alpha + 2(1-\alpha)\log(1-\alpha) - 2\alpha \log x - 2(1-\alpha)\log(1-x), \quad (6)$$

we deduce that $\alpha = \frac{1}{2} - \frac{\xi}{\log n}$, where $\xi = O(1)$. By carefully taking care of the little o terms, x can be expressed using

$$\frac{x}{c} = \delta_{\alpha n}^{2((2\alpha-1)n-1)}c \quad = \left(\frac{\alpha n}{2\pi e}(\alpha n \pi)^{1/(\alpha n)}\right)^{\frac{-2\xi n/\log n - 1}{\alpha n - 1}} \sim \left(\frac{n}{4\pi e}\right)^{-4\frac{\xi}{\log n}} \quad \sim e^{-4\xi}.$$

We can now compute the largest contribution K to the right hand side of Eq. 6:

$$K = \log\left(\frac{e^{4\xi}}{4c(1-ce^{-4\xi})}\right) = \log\left(\frac{(e^{4\xi}-2c)^2 + 4c(e^{4\xi}-c)}{4c(e^{4\xi}-c)}\right).$$

We must have $K + o(K) = 4\frac{\log n}{n}$, therefore the constant term must be 0, and thus $\xi = \frac{\log(2c)}{4} + o(1)$, which concludes our proof.

□

Corollary 4. *Let L be a rank n lattice for which Heu. 1 holds with vectors $\mathbf{s}_1, \ldots, \mathbf{s}_N$ of norm r such that $N = \Theta(n)$ and $r/\mathrm{vol}(L)^{1/n} := c^{-1/2} = O(1)$. The primal attack framework predicts that applying BKZ reduction with blocksize $\beta = n(1/2 - \log(2c)/4\log n + o(1/\log n))$ recovers a vector of norm at most r with high probability. In particular if the heuristic holds for hypercubic and NTRU lattices, then so does this result.*

Proof. The main point follows directly from Prop. 2. For a hypercubic lattice Λ, $\mathrm{vol}(\Lambda) = \|\mathbf{s}\| = 1$. For a NTRU lattice L, $\mathrm{vol}(L)^{1/n} = \sqrt{q} = \Theta(\sqrt{n})$, and $\|\mathbf{s}\| = \Theta(n)$, where $\mathbf{s} = (\mathbf{g}, \mathbf{f})$ is the secret key and q is the NTRU modulus.

□

4.3 Discussion and Illustration

The results of Prop. 1 and Prop. 2 are identical. If we focus uniquely on the primal attack[6], this means that asymptotically, having n short vectors does not offer any advantage over having just one. In fact, we conjecture that for k a constant, if we had a polynomial number N of independent (in the sense of Heu. 1) equally short vectors, then the following k terms of the expansion of the predicted blocksize assuming the presence of these N vectors would match precisely with the next k terms (of the form $a_i n \log^{-i}(n)$) derived in the case

[6] Dense sublattice attacks can asymptotically outperform generic lattice reduction for NTRU with overstretched parameters [35,21], but this is outside the scope of our study.

of a solitary short vector. Indeed, the estimates of Prop. 1 and Prop. 2 are not very good in practice, because the convergence rate is very weak (notice that the term in the erfc function is a $\Theta(\sqrt{\log n})$). This means that the asymptotic regime will only kick in for huge values of n, beyond cryptographic relevance. However, this does not prove that the presence of more short vectors is useless with regards to the primal attack. In fact, the structure of Eq. 1 indicates that having strictly more short vectors is directly advantageous.

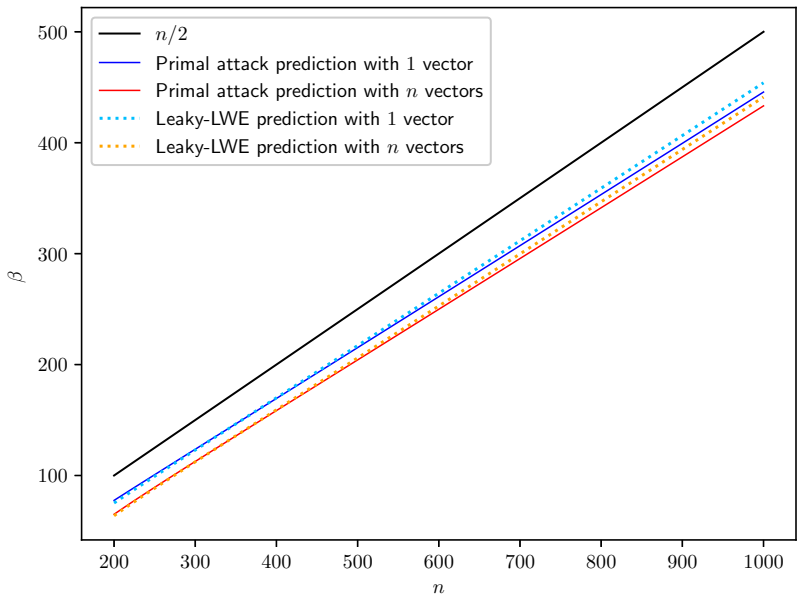


Fig. 4: Blocksizes required to recover unit vectors in dimension n hypercubic lattices. The predictions in dotted lines were generated using the sage script provided in [19]. Our model does not assume progressive-BKZ execution.

Practical alternative. Due to the reasons mentioned above, for practical application of our framework, we recommend directly solving the modified primal attack equation obtained from combining Eq. 1 and Lem. 3 numerically. The results for hypercubic lattices are plotted in Fig. 4, and compared with the predictions of [16]. In the observed range of dimensions, the heuristic blocksize gain is consistently between 11 and 13, compared to simply evaluating the asymptotic formula. Surprisingly, our naive predictions end up being very close to the

more precise predictions of [16]. We provide a proof of concept sage script at `https://github.com/htmb-bot/NTRU-and-Hypercubic`.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 885394). We would like to thank Huck Bennett, Léo Ducas, Noah Stephens-Davidowitz and Wessel van Woerden for insightful discussions.

References

1. Aggarwal, D., Dadush, D., Stephens-Davidowitz, N.: Solving the closest vector problem in 2^n time - the discrete gaussian strikes again! In: Proc. IEEE 56th FOCS. pp. 563–582 (2015)
2. Aggarwal, D., Li, J., Nguyen, P.Q., Stephens-Davidowitz, N.: Slide reduction, revisited - filling the gaps in SVP approximation. In: Micciancio, D., Ristenpart, T. (eds.) Advances in Cryptology - Proc. CRYPTO 2020, Part II. Lecture Notes in Computer Science, vol. 12171, pp. 274–295. Springer (2020)
3. Aggarwal, D., Stephens-Davidowitz, N.: Just take the average! an embarrassingly simple 2^n-time algorithm for SVP (and CVP). In: SOSA (2018), `http://arxiv.org/abs/1709.01535`
4. Albrecht, M.R., Ducas, L.: Lattice Attacks on NTRU and LWE: A History of Refinements, p. 15–40. London Mathematical Society Lecture Note Series, Cambridge University Press (2021)
5. Albrecht, M.R., Göpfert, F., Virdia, F., Wunderer, T.: Revisiting the expected cost of solving uSVP and applications to LWE. In: Proc. ASIACRYPT 2017, Part I. Lecture Notes in Computer Science, vol. 10624, pp. 297–322. Springer (2017)
6. Alkim, E., Ducas, L., Pöppelmann, T., Schwabe, P.: Post-quantum key exchange - A new hope. In: Proc. 25th USENIX. pp. 327–343. USENIX (2016)
7. Avanzi, R., Bos, J., Ducas, L., Kiltz, E., Lepoint, T., Lyubashevsky, V., Schanck, J.M., Schwabe, P., Seiler, G., Stehlé, D.: Crystals-kyber (version 2.0) – submission to round 2 of the nist post-quantum project (3 2019)
8. Bennett, H., Ganju, A., Peetathawatchai, P., Stephens-Davidowitz, N.: Just how hard are rotations of $\mathbb{Z}^n$? algorithms and cryptography with the simplest lattice. In: Hazay, C., Stam, M. (eds.) Advances in Cryptology – EUROCRYPT 2023. pp. 252–281. Springer Nature Switzerland, Cham (2023)
9. Bernard, O., Roux-Langlois, A.: Twisted-phs: Using the product formula to solve approx-svp in ideal lattices. In: Moriai, S., Wang, H. (eds.) Advances in Cryptology – ASIACRYPT 2020. pp. 349–380. Springer International Publishing, Cham (2020)
10. Bernstein, D.J., Chuengsatiansup, C., Lange, T., van Vredendaal, C.: Ntru prime: reducing attack surface at low cost. Cryptology ePrint Archive, Paper 2016/461 (2016)
11. Bos, J.W., Bronchain, O., Ducas, L., Fehr, S., Huang, Y.H., Pornin, T., Postlethwaite, E.W., Prest, T., Pulles, L.N., van Woerden, W.: Hawk signature specification document (6 2023)
12. Chen, C., Danba, O., Hoffstein, J., Hulsing, A., Rijneveld, J., Schanck, J.M., Saito, T., Schwabe, P., Whyte, W., Xagawa, K., Yamakawa, T., Zhang, Z.: Ntru algorithm specifications and supporting documentation (9 2020)

13. Chen, Y.: Réduction de réseau et sécurité concrète du chiffrement complètement homomorphe. Ph.D. thesis, Univ. Paris 7 (2013)
14. Chen, Y., Nguyen, P.Q.: BKZ 2.0: better lattice security estimates. In: Proc. ASIACRYPT 2011, LNCS, vol. 7073, pp. 1–20. Springer (2011)
15. Cramer, R., Ducas, L., Wesolowski, B.: Short stickelberger class relations and application to ideal-svp. In: Coron, J.S., Nielsen, J. (eds.) Advances in Cryptology – EUROCRYPT 2017. pp. 324–348. Springer International Publishing, Cham (2017)
16. Dachman-Soled, D., Ducas, L., Gong, H., Rossi, M.: Lwe with side information: Attacks and concrete security estimation. In: Advances in Cryptology – Proc. CRYPTO 2020. p. 329–358. Springer-Verlag, Berlin, Heidelberg (2020)
17. Ducas, L.: Provable lattice reduction of $\mathbb{Z}^n$ with blocksize $n/2$. Designs, Codes and Cryptography (Nov 2023)
18. Ducas, L., Kiltz, E., Lepoint, T., Lyubashevsky, V., Schwabe, P., Seiler, G., Stehlé, D.: Crystals-dilithium – submission to round 2 of the nist post-quantum project (3 2019)
19. Ducas, L., Postlethwaite, E.W., Pulles, L.N., van Woerden, W.P.J.: Hawk: Module LIP makes lattice signatures fast, compact and simple. In: Advances in Cryptology - Proc. ASIACRYPT 2022. Lecture Notes in Computer Science, vol. 13794, pp. 65–94. Springer (2022)
20. Ducas, L., Pulles, L.N.: Does the dual-sieve attack on learning with errors even work? In: Handschuh, H., Lysyanskaya, A. (eds.) Advances in Cryptology – CRYPTO 2023. pp. 37–69. Springer Nature Switzerland, Cham (2023)
21. Ducas, L., van Woerden, W.: Ntru fatigue: How stretched is overstretched? In: Tibouchi, M., Wang, H. (eds.) Advances in Cryptology – ASIACRYPT 2021. pp. 3–32. Springer International Publishing, Cham (2021)
22. Ducas, L., van Woerden, W.: On the lattice isomorphism problem, quadratic forms, remarkable lattices, and cryptography. In: Dunkelman, O., Dziembowski, S. (eds.) Advances in Cryptology - Proc. EUROCRYPT 2022. Lecture Notes in Computer Science, vol. 13277, pp. 643–673. Springer (2022)
23. Eisenbrand, F., Venzin, M.: Approximate cvp_p in time $2^{0.802n}$. J. Comput. Syst. Sci. **124**, 129–139 (2022)
24. Fouque, P.A., Hoffstein, J., Kirchner, P., Lyubashevsky, V., Pornin, T., Prest, T., Ricosset, T., Seiler, G., Whyte, W., Zhang, Z.: Falcon: Fast-fourier lattice-based compact signatures over ntru (3 2019)
25. Frankl, P., Maehara, H.: Some geometric applications of the beta distribution. Annals of the Institute of Statistical Mathematics **42**, 463–474 (1990)
26. Gama, N., Howgrave-Graham, N., Nguyen, P.Q.: Symplectic lattice reduction and NTRU. In: Vaudenay, S. (ed.) Advances in Cryptology - Proc. EUROCRYPT 2006. Lecture Notes in Computer Science, vol. 4004, pp. 233–253. Springer (2006)
27. Gama, N., Nguyen, P.Q.: Finding short lattice vectors within Mordell's inequality. In: Proc. 40th ACM Symp. on Theory of Computing (STOC) (2008)
28. Gama, N., Nguyen, P.Q.: Predicting Lattice Reduction. In: Proc. of Eurocrypt '08. pp. 31–51. LNCS, Springer - Verlag (2008)
29. Gama, N., Nguyen, P.Q., Regev, O.: Lattice enumeration using extreme pruning. In: Advances in Cryptology - Proc. EUROCRYPT 2010, LNCS, vol. 6110. Springer (2010)
30. Hirschhorn, P.S., Hoffstein, J., Howgrave-Graham, N., Whyte, W.: Choosing NTRUEncrypt parameters in light of combined lattice reduction and MITM approaches. In: Proc. ACNS 2009. Lecture Notes in Computer Science, vol. 5536, pp. 437–455 (2009)

31. Hoffstein, J., Pipher, J., Silverman, J.: NTRU: a ring based public key cryptosystem. In: Proc. of ANTS III. LNCS, vol. 1423, pp. 267–288. Springer-Verlag (1998)
32. Hoffstein, J., Pipher, J., Schanck, J.M., Silverman, J.H., Whyte, W., Zhang, Z.: Choosing parameters for ntruencrypt. In: Handschuh, H. (ed.) Topics in Cryptology – CT-RSA 2017. pp. 3–18. Springer International Publishing, Cham (2017)
33. Howgrave-Graham, N.: A hybrid lattice-reduction and meet-in-the-middle attack against NTRU. In: Proc. CRYPTO 2007. Lecture Notes in Computer Science, vol. 4622, pp. 150–169. Springer (2007)
34. Kim, J., Park, J.H.: Ntru+: Compact construction of ntru using simple encoding method. IEEE Transactions on Information Forensics and Security **18**, 4760–4774 (2023)
35. Kirchner, P., Fouque, P.A.: Revisiting lattice attacks on overstretched ntru parameters. In: Coron, J.S., Nielsen, J.B. (eds.) Advances in Cryptology – EUROCRYPT 2017. pp. 3–26. Springer International Publishing (2017)
36. Lenstra, A.K., Lenstra, Jr., H.W., Lovász, L.: Factoring polynomials with rational coefficients. Mathematische Ann. **261**, 513–534 (1982)
37. Li, J., Nguyen, P.Q.: A complete analysis of the bkz lattice reduction algorithm. Cryptology ePrint Archive, Paper 2020/1237 (2020)
38. Martinet, J.: Perfect Lattices in Euclidean Spaces. Springer Berlin, Heidelberg (2003)
39. Micciancio, D., Walter, M.: Practical, predictable lattice basis reduction. In: Advances in Cryptology - Proc. EUROCRYPT 2016, Part I. Lecture Notes in Computer Science, vol. 9665, pp. 820–849. Springer (2016)
40. Nguyen, P.Q., Stehlé, D.: LLL on the average. In: Proc. ANTS. pp. 238–256 (2006)
41. Pellet-Mary, A., Hanrot, G., Stehlé, D.: Approx-svp in ideal lattices with preprocessing. In: Ishai, Y., Rijmen, V. (eds.) Advances in Cryptology – EUROCRYPT 2019. pp. 685–716. Springer International Publishing (2019)
42. Philip, J.R.: The Function inverfc θ. Australian Journal of Physics **13** (Mar 1960)
43. Schneider, M., Gama, N.: SVP challenge, available at `http://www.latticechallenge.org/svp-challenge/`
44. Schnorr, C.P.: Lattice reduction by random sampling and birthday methods. In: Proc. STACS 2003, LNCS, vol. 2607, pp. 145–156. Springer (2003)
45. Schnorr, C.P., Euchner, M.: Lattice basis reduction: improved practical algorithms and solving subset sum problems. Math. Programming **66**, 181–199 (1994)
46. Szydlo, M.: Hypercubic lattice reduction and analysis of GGH and NTRU signatures. In: Biham, E. (ed.) Advances in Cryptology - Proc. EUROCRYPT 2003. Lecture Notes in Computer Science, vol. 2656, pp. 433–448. Springer (2003)
47. Temme, N.: Asymptotic inversion of the incomplete beta function. Journal of Computational and Applied Mathematics **41**(1), 145–157 (1992)