An Improved BKW Algorithm for LWE with Applications to Cryptography and Lattices

Paul Kirchner[1] and Pierre-Alain Fouque[2(✉)]

[1] École Normale Supérieure, Paris, France
paul.kirchner@ens.fr
[2] Université de Rennes 1 and Institut Universitaire de France, Rennes, France
pierre-alain.fouque@ens.fr

Abstract. In this paper, we study the Learning With Errors problem and its binary variant, where secrets and errors are binary or taken in a small interval. We introduce a new variant of the Blum, Kalai and Wasserman algorithm, relying on a quantization step that generalizes and fine-tunes modulus switching. In general this new technique yields a significant gain in the constant in front of the exponent in the overall complexity. We illustrate this by solving within half a day a LWE instance with dimension $n = 128$, modulus $q = n^2$, Gaussian noise $\alpha = 1/(\sqrt{n/\pi}\log^2 n)$ and binary secret, using 2^{28} samples, while the previous best result based on BKW claims a time complexity of 2^{74} with 2^{60} samples for the same parameters.

We then introduce variants of BDD, GapSVP and UniqueSVP, where the target point is required to lie in the fundamental parallelepiped, and show how the previous algorithm is able to solve these variants in subexponential time. Moreover, we also show how the previous algorithm can be used to solve the BinaryLWE problem with n samples in subexponential time $2^{(\ln 2/2+o(1))n/\log\log n}$. This analysis does not require any heuristic assumption, contrary to other algebraic approaches; instead, it uses a variant of an idea by Lyubashevsky to generate many samples from a small number of samples. This makes it possible to asymptotically and heuristically break the NTRU cryptosystem in subexponential time (without contradicting its security assumption). We are also able to solve subset sum problems in subexponential time for density $o(1)$, which is of independent interest: for such density, the previous best algorithm requires exponential time. As a direct application, we can solve in subexponential time the parameters of a cryptosystem based on this problem proposed at TCC 2010.

1 Introduction

The Learning With Errors (LWE) problem has been an important problem in cryptography since its introduction by Regev in [34]. Many cryptosystems have been proven secure assuming the hardness of this problem, including Fully Homomorphic Encryption schemes [11,16]. The decision version of the problem can be described as follows: given m samples of the form $(\mathbf{a}, b) \in (\mathbb{Z}_q)^n \times \mathbb{Z}_q$, where $\mathbf{a}$ are

R. Gennaro and M. Robshaw (Eds.): CRYPTO 2015, Part I, LNCS 9215, pp. 43–62, 2015.
DOI: 10.1007/978-3-662-47989-6_3

uniformy distributed in $(\mathbb{Z}_q)^n$, distinguish whether b is uniformly chosen in $\mathbb{Z}_q$ or is equal to $\langle \mathbf{a}, \mathbf{s} \rangle + e$ for a fixed secret $\mathbf{s} \in (\mathbb{Z}_q)^n$ and e a noise value in $\mathbb{Z}_q$ chosen according to some probability distribution. Typically, the noise is sampled from some distribution concentrated on small numbers, such as a discrete Gaussian distribution with standard deviation αq for $\alpha = o(1)$. In the search version of the problem, the goal is to recover $\mathbf{s}$ given the promise that the sample instances come from the latter distribution. Initially, Regev showed that if $\alpha q \geq 2\sqrt{n}$, solving LWE on average is at least as hard as approximating lattice problems in the worst case to within $\tilde{\mathcal{O}}(n/\alpha)$ factors with a quantum algorithm. Peikert shows a classical reduction when the modulus is large $q \geq 2^n$ in [32]. Finally, in [10], Brakerski *et al.* prove that solving LWE instances with polynomial-size modulus in polynomial time implies an efficient solution to GapSVP.

There are basically three approaches to solving LWE: the first relies on lattice reduction techniques such as the LLL [23] algorithm and further improvements [12] as exposed in [25,26]; the second uses combinatorial techniques [9,35]; and the third uses algebraic techniques [6]. According to Regev in [1], the best known algorithm to solve LWE is the algorithm by Blum, Kalai and Wasserman in [9], originally proposed to solve the Learning Parities with Noise (LPN) problem, which can be viewed as a special case of LWE where $q = 2$. The time and memory requirements of this algorithm are both exponential for LWE and subexponential for LPN in $2^{\mathcal{O}(n/\log n)}$. During the first stage of the algorithm, the dimension of $\mathbf{a}$ is reduced, at the cost of a (controlled) decrease of the bias of b. During the second stage, the algorithm distinguishes between LWE and uniform by evaluating the bias.

Since the introduction of LWE, some variants of the problem have been proposed in order to build more efficient cryptosystems. Some of the most interesting variants are Ring-LWE by Lyubashevsky, Peikert and Regev in [29], which aims to reduce the space of the public key using cyclic samples; and the cryptosystem by Döttling and Müller-Quade [14], which uses short secret and error. In 2013, Micciancio and Peikert [30] as well as Brakerski *et al.* [10] proposed a binary version of the LWE problem and obtained a hardness result.

Related Work. Albrecht *et al.* have presented an analysis of the BKW algorithm as applied to LWE in [3,4]. It has been recently revisited by Duc *et al.*, who use a multi-dimensional FFT in the second stage of the algorithm [15]. However, the main bottleneck is the first BKW step and since the proposed algorithms do not improve this stage, the overall asymptotic complexity is unchanged.

In the case of the BinaryLWE variant, where the error and secret are binary (or sufficiently small), Micciancio and Peikert show that solving this problem using $m = n(1 + \Omega(1/\log(n)))$ samples is at least as hard as approximating lattice problems in the worst case in dimension $\Theta(n/\log(n))$ with approximation factor $\tilde{\mathcal{O}}(\sqrt{n}q)$. We show in the full version that existing lattice reduction techniques require exponential time. Arora and Ge describe a $2^{\tilde{\mathcal{O}}(\alpha q)^2}$-time algorithm when $q > n$ to solve the LWE problem [6]. This leads to a subexponential time algorithm when the error magnitude αq is less than $\sqrt{n}$. The idea is to transform this system into a noise-free polynomial system and then use root finding

algorithms for multivariate polynomials to solve it, using either relinearization in [6] or Gröbner basis in [2]. In this last work, Albrecht *et al.* present an algorithm whose time complexity is $2^{\frac{(\omega+o(1))n \log\log\log n}{8 \log\log n}}$ when the number of samples $m = (1 + o(1))n \log\log n$ is super-linear, where $\omega < 2.3728$ is the linear algebra constant, under some assumption on the regularity of the polynomial system of equations; and when $m = \mathcal{O}(n)$, the complexity becomes exponential.

Contribution. Our first contribution is to present in a unified framework the BKW algorithm and all its previous improvements in the binary case [8,18,21,24] and in the general case [4]. We introduce a new quantization step, which generalizes modulus switching [4]. This yields a significant decrease in the constant of the exponential of the complexity for LWE. Moreover our proof does not require Gaussian noise, and does not rely on unproven independence assumptions. Our algorithm is also able to tackle problems with larger noise.

We then introduce generalizations of the BDD, GapSVP and UniqueSVP problems, and prove a reduction from these variants to LWE. When particular parameters are set, these variants impose that the lattice point of interest (the point of the lattice that the problem essentially asks to locate: for instance, in the case of BDD, the point of the lattice closest to the target point) lie in the fundamental parallelepiped; or more generally, we ask that the coordinates of this point relative to the basis defined by the input matrix $\mathbf{A}$ has small infinity norm, bounded by some value B. For small B, our main algorithm yields a subexponential-time algorithm for these variants of BDD, GapSVP and UniqueSVP.

Through a reduction to our variant of BDD, we are then able to solve the subset-sum problem in subexponential time when the density is $o(1)$, and in time $2^{(\ln 2/2+o(1))n/\log\log n}$ if the density is $\mathcal{O}(1/\log n)$. This is of independent interest, as existing techniques for density $o(1)$, based on lattice reduction, require exponential time. As a consequence, the cryptosystems of Lyubashevsky, Palacio and Segev at TCC 2010 [28] can be solved in subexponential time.

As another application of our main algorithm, we show that BinaryLWE with reasonable noise can be solved in time $2^{(\ln 2/2+o(1))n/\log\log n}$ instead of $2^{\Omega(n)}$; and the same complexity holds for secret of size up to $2^{\log^{o(1)} n}$. As a consequence, we can heuristically recover the secret polynomials $\mathbf{f}, \mathbf{g}$ of the NTRU problem in subexponential time $2^{(\ln 2/2+o(1))n/\log\log n}$ (without contradicting its security assumption). The heuristic assumption comes from the fact that NTRU samples are not random, since they are rotations of each other: the heuristic assumption is that this does not significantly hinder BKW-type algorithms. Note that there is a large value hidden in the $o(1)$ term, so that our algorithm does not yield practical attacks for recommended NTRU parameters.

Our results are extended to the case where the secret is small with respect to the L2 norm in the full version.

2 Preliminaries

We identify any element of $\mathbb{Z}/q\mathbb{Z}$ to the smallest of its equivalence class, the positive one in case of tie. Any vector $\mathbf{x} \in \left(\mathbb{Z}/q\mathbb{Z}\right)^n$ has an Euclidean norm

$||\mathbf{x}|| = \sqrt{\sum_{i=0}^{n-1} x_i^2}$ and $||\mathbf{x}||_\infty = \max_i |x_i|$. A matrix $\mathbf{B}$ can be Gram-Schmidt orthogonalized in $\widetilde{\mathbf{B}}$, and its norm $||\mathbf{B}||$ is the maximum of the norm of its columns. We denote by $(\mathbf{x}|\mathbf{y})$ the vector obtained as the concatenation of vectors $\mathbf{x}, \mathbf{y}$. Let $\mathbf{I}$ be the identity matrix and we denote by ln the neperian logarithm and log the binary logarithm. A lattice is the set of all integer linear combinations $\Lambda(\mathbf{b}_1, \ldots, \mathbf{b}_n) = \sum_i \mathbf{b_i} \cdot x_i$ (where $x_i \in \mathbb{Z}$) of a set of linearly independent vectors $\mathbf{b}_1, \ldots, \mathbf{b}_n$ called the basis of the lattice. If $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_n]$ is the matrix basis, lattice vectors can be written as $\mathbf{Bx}$ for $\mathbf{x} \in \mathbb{Z}^n$. Its dual Λ^* is the set of $\mathbf{x} \in \mathbb{R}^n$ such that $\langle \mathbf{x}, \Lambda \rangle \subset \mathbb{Z}^n$. We have $\Lambda^{**} = \Lambda$. We borrow Bleichenbacher's definition of bias [31].

Definition 1. *The bias of a probability distribution ϕ over $\mathbb{Z}/q\mathbb{Z}$ is*

$$\mathbb{E}_{x \sim \phi}[\exp(2i\pi x/q)].$$

This definition extends the usual definition of the bias of a coin in $\mathbb{Z}/2\mathbb{Z}$: it preserves the fact that any distribution with bias b can be distinguished from uniform with constant probability using $\Omega(1/b^2)$ samples, as a consequence of Hoeffding's inequality; moreover the bias of the sum of two independent variable is still the product of their biases. We also have the following simple lemma:

Lemma 1. *The bias of the Gaussian distribution of mean 0 and standard deviation $q\alpha$ is $\exp(-2\pi^2\alpha^2)$.*

Proof. The bias is the value of the Fourier transform at $-1/q$. □

We introduce a non standard definition for the LWE problem. However as a consequence of Lemma 1, this new definition naturally extends the usual Gaussian case (as well as its standard extensions such as the bounded noise variant [10, Definition 2.14]), and it will prove easier to work with.

Definition 2. *Let $n \geq 0$ and $q \geq 2$ be integers. Given parameters α and ϵ, the* LWE *distribution is, for $\mathbf{s} \in (\mathbb{Z}/q\mathbb{Z})^n$, a distribution on pairs $(\mathbf{a}, b) \in (\mathbb{Z}/q\mathbb{Z})^n \times (\mathbb{R}/q\mathbb{Z})$ such that $\mathbf{a}$ is sampled uniformly, and for all $\mathbf{a}$,*

$$|\mathbb{E}[\exp(2i\pi(\langle \boldsymbol{a}, \boldsymbol{s} \rangle - b)/q)|\mathbf{a}] \exp(\alpha'^2) - 1| \leq \epsilon$$

for some universal $\alpha' \leq \alpha$.

For convenience, we define $\beta = \sqrt{n/2}/\alpha$. In the remainder, α is called the noise *parameter*[1]*, and ϵ the* distortion *parameter. Also, we say that a* LWE *distribution has a noise distribution ϕ if b is distributed as $\langle \mathbf{a}, \mathbf{s} \rangle + \phi$.*

Definition 3. *The* Decision-LWE *problem is to distinguish a* LWE *distribution from the uniform distribution over $(\mathbf{a}, b)$. The* Search-LWE *problem is, given samples from a* LWE *distribution, to find $\mathbf{s}$.*

[1] Remark that it differs by a constant factor from other authors' definition of α.

Definition 4. *The real λ_i is the radius of the smallest ball, centered in $\mathbf{0}$, such that it contains i vectors of the lattice Λ which are linearly independent.*

We define $\rho_s(\mathbf{x}) = \exp(-\pi||\mathbf{x}||^2/s^2)$ and $\rho_s(S) = \sum_{\mathbf{x}\in S} \rho_s(\mathbf{x})$ (and similarly for other functions). The discrete Gaussian distribution $D_{E,s}$ over a set E and of parameter s is such that the probability of $D_{E,s}(\mathbf{x})$ of drawing $\mathbf{x} \in E$ is equal to $\rho_s(\mathbf{x})/\rho_s(E)$. To simplify notation, we will denote by D_E the distribution $D_{E,1}$.

Definition 5. *The smoothing parameter η_ϵ of the lattice Λ is the smallest s such that $\rho_{1/s}(\Lambda^*) = 1 + \epsilon$.*

Now, we will generalize the BDD, UniqueSVP and GapSVP problems by using another parameter B that bounds the target lattice vector. For $B = 2^n$, we recover the usual definitions if the input matrix is reduced.

Definition 6. *The $\mathsf{BDD}^{||.||_\infty}_{B,\beta}$ (resp. $\mathsf{BDD}^{||.||}_{B,\beta}$) problem is, given a basis $\mathbf{A}$ of the lattice Λ, and a point $\mathbf{x}$ such that $||\mathbf{As} - \mathbf{x}|| \leq \lambda_1/\beta < \lambda_1/2$ and $||\mathbf{s}||_\infty \leq B$ (resp. $||\mathbf{s}|| \leq B$), to find $\mathbf{s}$.*

Definition 7. *The $\mathsf{UniqueSVP}^{||.||_\infty}_{B,\beta}$ (resp. $\mathsf{UniqueSVP}^{||.||}_{B,\beta}$) problem is, given a basis $\mathbf{A}$ of the lattice Λ, such that $\lambda_2/\lambda_1 \geq \beta$ and there exists a vector $\mathbf{s}$ such that $||\mathbf{As}|| = \lambda_1$ with $||\mathbf{s}||_\infty \leq B$ (resp. $||\mathbf{s}|| \leq B$), to find $\mathbf{s}$.*

Definition 8. *The $\mathsf{GapSVP}^{||.||_\infty}_{B,\beta}$ (resp. $\mathsf{GapSVP}^{||.||}_{B,\beta}$) problem is, given a basis $\mathbf{A}$ of the lattice Λ to distinguish between $\lambda_1(\Lambda) \geq \beta$ and if there exists $\mathbf{s} \neq \mathbf{0}$ such that $||\mathbf{s}||_\infty \leq B$ (resp. $||\mathbf{s}|| \leq B$) and $||\mathbf{As}|| \leq 1$.*

Definition 9. *Given two probability distributions P and Q on a finite set S, the Kullback-Leibler (or KL) divergence between P and Q is*

$$D_{\mathsf{KL}}(P||Q) = \sum_{x\in S} \ln\left(\frac{P(x)}{Q(x)}\right) P(x) \quad \textit{with} \quad \ln(x/0) = +\infty \quad \textit{if} \quad x > 0.$$

The following two lemmata are proven in [33]:

Lemma 2. *Let P and Q be two distributions over S, such that for all x, $|P(x) - Q(x)| \leq \delta(x)P(x)$ with $\delta(x) \leq 1/4$. Then:*

$$D_{\mathsf{KL}}(P||Q) \leq 2\sum_{x\in S} \delta(x)^2 P(x).$$

Lemma 3. *Let A be an algorithm which takes as input m samples of S and outputs a bit. Let x (resp. y) be the probability that it returns 1 when the input is sampled from P (resp. Q). Then:*

$$|x - y| \leq \sqrt{mD_{\mathsf{KL}}(P||Q)/2}.$$

Finally, we say that an algorithm has a negligible probability of failure if its probability of failure is $2^{-\Omega(n)}$.[2]

[2] Some authors use another definition.

2.1 Secret-Error Switching

At a small cost in samples, it is possible to reduce any LWE instance to an instance where the secret follows the same distribution as the error [5,10].

Theorem 1. *Given an oracle that solves* LWE *with m samples in time t with the secret coming from the rounded error distribution, it is possible to solve* LWE *with $m+\mathcal{O}(n \log\log q)$ samples with the same error distribution (and any distribution on the secret) in time $t + \mathcal{O}(mn^2 + (n \log\log q)^3)$, with negligible probability of failure.*

Furthermore, if q is prime, we lose $n+k$ samples with probability of failure bounded by q^{-1-k}.

Proof. First, select an invertible matrix $\mathbf{A}$ from the vectorial part of $\mathcal{O}(n \log\log q)$ samples in time $\mathcal{O}((n \log\log q)^3)$ [10, Claim 2.13].

Let $\boldsymbol{b}$ be the corresponding rounded noisy dot products. Let $\mathbf{s}$ be the LWE secret and $\mathbf{e}$ such that $\mathbf{As} + \mathbf{e} = \mathbf{b}$. Then the subsequent m samples are transformed in the following way. For each new sample $(\mathbf{a'}, b')$ with $b' = \langle \mathbf{a'}, \mathbf{s} \rangle + e'$, we give the sample $(-{}^t\mathbf{A}^{-1}\mathbf{a'}, b' - \langle {}^t\mathbf{A}^{-1}\mathbf{a'}, \mathbf{b} \rangle)$ to our LWE oracle.

Clearly, the vectorial part of the new samples remains uniform and since

$$b' - \langle {}^t\mathbf{A}^{-1}\mathbf{a'}, \mathbf{b} \rangle = \langle -{}^t\mathbf{A}^{-1}\mathbf{a'}, \mathbf{b} - \mathbf{As} \rangle + b' - \langle \mathbf{a'}, \mathbf{s} \rangle = \langle -{}^t\mathbf{A}^{-1}\mathbf{a'}, \mathbf{e} \rangle + e'$$

the new errors follow the same distribution as the original, and the new secret is $\mathbf{e}$. Hence the oracle outputs $\mathbf{e}$ in time t, and we can recover $\mathbf{s}$ as $\mathbf{s} = \mathbf{A}^{-1}(\mathbf{b}-\mathbf{e})$.

If q is prime, the probability that the $n+k$ first samples are in some hyperplane is bounded by $q^{n-1}q^{-n-k} = q^{-1-k}$. □

2.2 Low Dimension Algorithms

Our main algorithm will return samples from a LWE distribution, while the bias decreases. We describe two fast algorithms when the dimension is small enough.

Theorem 2. *If $n = 0$ and $m = k/b^2$, with b smaller than the real part of the bias, the* Decision-LWE *problem can be solved with advantage $1 - 2^{-\Omega(k)}$ in time $\mathcal{O}(m)$.*

Proof. The algorithm DISTINGUISH computes $x = \frac{1}{m}\sum_{i=0}^{m-1} \cos(2i\pi b_i/q)$ and returns the boolean $x \geq b/2$. If we have a uniform distribution then the average of x is 0, else it is larger than $b/2$. The Hoeffding inequality shows that the probability of $|x - \mathbb{E}[x]| \geq b/2$ is $2^{-k/8}$, which gives the result. □

Lemma 4. *For all $\mathbf{s} \neq \mathbf{0}$, if $\mathbf{a}$ is sampled uniformly, $\mathbb{E}[\exp(2i\pi\langle \mathbf{a}, \mathbf{s} \rangle/q)] = 0$.*

Proof. Multiplication by s_0 in $\mathbb{Z}_q$ is a $\gcd(s_0, q)$-to-one map because it is a group morphism, therefore $a_0 s_0$ is uniform over $\gcd(s_0, q)\mathbb{Z}_q$. Thus, by using $k = \gcd(q, s_0, \ldots, s_{n-1}) < q$, $\langle \mathbf{a}, \mathbf{s} \rangle$ is distributed uniformly over $k\mathbb{Z}_q$ so

$$\mathbb{E}[\exp(2i\pi\langle \mathbf{a}, \mathbf{s} \rangle/q)] = \frac{q}{k} \sum_{j=0}^{q/k-1} \exp(2i\pi jk/q) = 0.$$ □

Algorithm 1. FindSecret

function FindSecret($\mathcal{L}$)
  for all $(\mathbf{a}, b) \in \mathcal{L}$ **do**
    $f[\mathbf{a}] \leftarrow f[\mathbf{a}] + \exp(2i\pi b/q)$
  end for
  $t \leftarrow$ FastFourierTransform(f)
  return $\arg\max_{\mathbf{s}\in(\mathbb{Z}/q\mathbb{Z})^n} \Re(t[\mathbf{s}])$
end function

Theorem 3. *The algorithm* FindSecret*, when given* $m > (8n\log q + k)/b^2$ *samples from a* LWE *problem with bias whose real part is superior to* b *returns the correct secret in time* $\mathcal{O}(m + n\log^2(q)q^n)$ *except with probability* $2^{-\Omega(k)}$*.*

Proof. The fast Fourier transform needs $\mathcal{O}(nq^n)$ operations on numbers of bit size $\mathcal{O}(\log(q))$. The Hoeffding inequality shows that the difference between $t[\mathbf{s'}]$ and $\mathbb{E}[\exp(2i\pi(b - \langle \mathbf{a}, \mathbf{s'}\rangle)/q)]$ is at most $b/2$ except with probability at most $2\exp(-mb^2/2)$. Consequently, it holds for all $\mathbf{s'}$ except with probability at most $2q^n \exp(-mb^2/2) = 2^{-\Omega(k)}$ using the union bound. Then $t[\mathbf{s}] \geq b - b/2 = b/2$ and for all $\mathbf{s'} \neq \mathbf{s}$, $t[\mathbf{s'}] < b/2$ so the algorithm returns $\mathbf{s}$. □

3 Main Algorithm

In this section, we present our main algorithm, prove its asymptotical complexity, and present practical results in dimension $n = 128$.

3.1 Rationale

A natural idea in order to distinguish between an instance of LWE (or LPN) and a uniform distribution is to select some k samples that add up to zero, yielding a new sample of the form $(\mathbf{0}, e)$. It is then enough to distinguish between e and a uniform variable. However, if δ is the bias of the error in the original samples, the new error e has bias δ^k, hence roughly δ^{-2k} samples are necessary to distinguish it from uniform. Thus it is crucial that k be as small a possible.

The idea of the algorithm by Blum, Kalai and Wasserman BKW is to perform "blockwise" Gaussian elimination. The n coordinates are divided into k blocks of length $b = n/k$. Then, samples that are equal on the first b coordinates are substracted together to produce new samples that are zero on the first block. This process is iterated over each consecutive block. Eventually samples of the form $(\mathbf{0}, e)$ are obtained.

Each of these samples ultimately results from the addition of 2^k starting samples, so k should be at most $\mathcal{O}(\log(n))$ for the algorithm to make sense. On the other hand $\Omega(q^b)$ data are clearly required at each step in order to generate enough collisions on b consecutive coordinates of a block. This naturally results in a complexity roughly $2^{(1+o(1))n/\log(n)}$ in the original algorithm for LPN. This algorithm was later adapted to LWE in [3], and then improved in [4].

The idea of the latter improvement is to use so-called "lazy modulus switching". Instead of finding two vectors that are equal on a given block in order to generate a new vector that is zero on the block, one uses vectors that are merely close to each other. This may be seen as performing addition modulo p instead of q for some $p < q$, by rounding every value $x \in \mathbb{Z}_q$ to the value nearest xp/q in $\mathbb{Z}_p$. Thus at each step of the algorithm, instead of generating vectors that are zero on each block, small vectors are produced. This introduces a new "rounding" error term, but essentially reduces the complexity from roughly q^b to p^b. Balancing the new error term with this decrease in complexity results in a significant improvement.

However it may be observed that this rounding error is much more costly for the first few blocks than the last ones. Indeed samples produced after, say, one iteration step are bound to be added together 2^{a-1} times to yield the final samples, resulting in a corresponding blowup of the rounding error. By contrast, later terms will undergo less additions. Thus it makes sense to allow for progressively coarser approximations (i.e. decreasing the modulus) at each step. On the other hand, to maintain comparable data requirements to find collisions on each block, the decrease in modulus is compensated by progressively longer blocks.

What we propose here is a more general view of the BKW algorithm that allows for this improvement, while giving a clear view of the different complexity costs incurred by various choice of parameters. Balancing these terms is the key to finding an optimal complexity. We forego the "modulus switching" point of view entirely, while retaining its core ideas. The resulting algorithm generalizes several variants of BKW, and will be later applied in a variety of settings.

3.2 Quantization

The goal of quantization is to associate to each point of $\mathbb{R}^k$ a center from a *small* set, such that the expectancy of the distance between a point and its center is small. We will then be able to produce small vectors by substracting vectors associated to the same center.

Modulus switching amounts to a simple quantizer which rounds every coordinate to the nearest multiple of some constant. Our proven algorithm uses a similar quantizer, except the constant depends on the index of the coordinate.

It is possible to decrease the average distance from a point to its center by a constant factor for large moduli [17], but doing so would complicate our proof without improving the leading term of the complexity. When the modulus is small, it might be worthwhile to use error-correcting codes as in [18].

3.3 Main Algorithm

Let us denote by $\mathcal{L}_0$ the set of starting samples, and $\mathcal{L}_i$ the sample list after i reduction steps. The numbers $d_0 = 0 \leq d_1 \leq \cdots \leq d_k = n$ partition the n coordinates of sample vectors into k buckets. Let $\mathbf{D} = (D_0, \ldots, D_{k-1})$ be the vector of quantization coefficients associated to each bucket.

Algorithm 2. Main resolution

```
function Reduce(L_in, D_i, d_i, d_{i+1})
    L_out ← ∅
    t[] ← ∅
    for all (a, b) ∈ L_in do
        r = ⌊(a_{d_i}, ..., a_{d_{i+1}-1}) / D⌉
        if t[r] = ∅ then
            t[r] ← (a, b)
        else
            L_out ← L_out::{(a, b) − t[r]}
            t[r] ← ∅
        end if
    end for
    return L_out
end function
function Solve(L_0, B, (d_i))
    for 0 ≤ i < k do
        L_{i+1} ← ReduceL_i, D_i, d_i, d_{i+1}
    end for
    return Distinguish{b|(a, b) ∈ L_k}
end function
```

In order to allow for a uniform presentation of the BKW algorithm, applicable to different settings, we do not assume a specific distribution on the secret. Instead, we assume there exists some *known* $\mathbf{B} = (B_0, \ldots, B_{n-1})$ such that $\sum_i (s_i/B_i)^2 \leq n$. Note that this is in particular true if $|s_i| \leq B_i$. We shall see how to adapt this to the standard Gaussian case later on. Without loss of generality, $\mathbf{B}$ is non increasing.

There are a phases in our reduction: in the i-th phase, the coordinates from d_i to d_{i+1} are reduced. We define $m = |\mathcal{L}_0|$.

Lemma 5. Solve *terminates in time* $\mathcal{O}(mn \log q)$.

Proof. The Reduce algorithm clearly runs in time $\mathcal{O}(|\mathcal{L}| n \log q)$. Moreover, $|\mathcal{L}_{i+1}| \leq |\mathcal{L}_i|/2$ so that the total running time of Solve is $\mathcal{O}(n \log q \sum_{i=0}^{k} m/2^i) = \mathcal{O}(mn \log q)$. □

Lemma 6. *Write* $\mathcal{L}'_i$ *for the samples of* $\mathcal{L}_i$ *where the first* d_i *coordinates of each sample vector have been truncated. Assume* $|s_j| D_i < 0.23q$ *for all* $d_i \leq j < d_{i+1}$. *If* $\mathcal{L}'_i$ *is sampled according to the* LWE *distribution of secret* $\mathbf{s}$ *and noise parameters* α *and* $\epsilon \leq 1$, *then* $\mathcal{L}'_{i+1}$ *is sampled according to the* LWE *distribution of the truncated secret with parameters:*

$$\alpha'^2 = 2\alpha^2 + 4\pi^2 \sum_{j=d_i}^{d_{i+1}-1} (s_j D_i/q)^2 \quad \text{and} \quad \epsilon' = 3\epsilon.$$

On the other hand, if $D_i = 1$, *then* $\alpha'^2 = 2\alpha^2$.

Proof. The independence of the outputted samples and the uniformity of their vectorial part are clear. Let $(\mathbf{a}, b)$ be a sample obtained by substracting two samples from $\mathcal{L}_i$. For $\mathbf{a'}$ the vectorial part of a sample, define $\epsilon(\mathbf{a'})$ such that $\mathbb{E}[\exp(2i\pi(\langle \mathbf{a'}, \mathbf{s}\rangle - b')/q)|\mathbf{a'}] = (1 + \epsilon(\mathbf{a'}))\exp(-\alpha^2)$. By definition of LWE, $|\epsilon(\mathbf{a'})| \leq \epsilon$, and by independence:

$$\mathbb{E}[\exp(2i\pi(\langle \mathbf{a}, \mathbf{s}\rangle - b)/q)|\mathbf{a}] = \exp(-2\alpha^2)\mathbb{E}_{\mathbf{a'}-\mathbf{a''}=\mathbf{a}}[(1 + \epsilon(\mathbf{a'}))(1 + \epsilon(\mathbf{a''}))],$$

with $|\mathbb{E}_{\mathbf{a'}-\mathbf{a''}=\mathbf{a}}[(1 + \epsilon(\mathbf{a'}))(1 + \epsilon(\mathbf{a''}))] - 1| \leq 3\epsilon$.

Thus we computed the noise corresponding to adding two samples of $\mathcal{L}_i$. To get the noise for a sample from $\mathcal{L}_{i+1}$, it remains to truncate coordinates from d_i to d_{i+1}. A straightforward induction on the coordinates shows that this noise is:

$$\exp(-2\alpha^2)\mathbb{E}_{\mathbf{a'}-\mathbf{a''}=\mathbf{a}}[(1 + \epsilon(\mathbf{a'}))(1 + \epsilon(\mathbf{a''}))] \prod_{j=d_i}^{d_{i+1}-1} \mathbb{E}[\exp(2i\pi \boldsymbol{a}_j \boldsymbol{s}_j/q)].$$

Indeed, if we denote by $\mathbf{a}^{(j)}$ the vector $\mathbf{a}$ where the first j coordinates are truncated and α_j the noise parameter of $\mathbf{a}^{(j)}$, we have:

$$\begin{aligned}
&|\mathbb{E}[\exp(2i\pi(\langle \mathbf{a}^{(j+1)}, \mathbf{s}^{(j+1)}\rangle - b)/q)|\mathbf{a}^{(j+1)}] - \exp(-\alpha_n^2)\mathbb{E}[\exp(2i\pi \boldsymbol{a}_j \boldsymbol{s}_j/q)]| \\
&= |\mathbb{E}[\exp(-2i\pi \boldsymbol{a}_j \boldsymbol{s}_j/q)(\exp(2i\pi(\langle \mathbf{a}^{(j)}, \mathbf{s}^{(j)}\rangle - b)/q) - \exp(-\alpha_j^2))]| \\
&\leq \epsilon' \exp(-\alpha_j^2)\mathbb{E}[\exp(2i\pi \boldsymbol{a}_j \boldsymbol{s}_j/q)].
\end{aligned}$$

It remains to compute $\mathbb{E}[\exp(2i\pi \boldsymbol{a}_j \boldsymbol{s}_j/q)]$ for $d_i \leq j < d_{i+1}$. Let $D = D_i$. The distribution of $\boldsymbol{a}_j$ is even, so $\mathbb{E}[\exp(2i\pi \boldsymbol{a}_j \boldsymbol{s}_j)]$ is real. Furthermore, since $|\boldsymbol{a}_j| \leq D$,

$$\mathbb{E}[\exp(2i\pi \boldsymbol{a}_j \boldsymbol{s}_j/q)] \geq \cos(2\pi \boldsymbol{s}_j D/q).$$

Assuming $|\boldsymbol{s}_j|D < 0.23q$, simple function analysis shows that

$$\mathbb{E}[\exp(2i\pi \boldsymbol{a}_j \boldsymbol{s}_j/q)] \geq \exp(-4\pi^2 \boldsymbol{s}_j^2 D^2/q^2).$$

On the other hand, if $D_i = 1$ then $\boldsymbol{a}_j = 0$ and $\mathbb{E}[\exp(2i\pi \boldsymbol{a}_j \boldsymbol{s}_j/q)] = 1$. □

Finding optimal parameters for BKW amounts to balancing various costs: the baseline number of samples required so that the final list $\mathcal{L}_k$ is non-empty, and the additional factor due to the need to distinguish the final error bias. This final bias itself comes both from the blowup of the original error bias by the BKW additions, and the "rounding errors" due to quantization. Balancing these costs essentially means solving a system.

For this purpose, it is convenient to set the overall target complexity as $2^{n(x+o(1))}$ for some x to be determined. The following auxiliary lemma essentially gives optimal values for the parameters of Solve assuming a suitable value of x. The actual value of x will be decided later on.

Lemma 7. *Pick some value x (dependent on* LWE *parameters). Choose:*

$$k \leq \left\lfloor \log\left(\frac{nx}{6\alpha^2}\right)\right\rfloor \qquad m = n2^k 2^{nx}$$
$$D_i \leq \frac{q\sqrt{x/6}}{\pi B_{d_i} 2^{(a-i+1)/2}} \qquad d_{i+1} = \min\left(d_i + \left\lfloor \frac{nx}{\log(1+q/D_i)}\right\rfloor, n\right).$$

Assume $d_k = n$ and $\epsilon \leq 1/(\beta^2 x)^{\log 3}$, and for all i and $d_i \leq j < d_{i+1}$, $|s_j|D_i < 0.23q$. Solve *runs in time $\mathcal{O}(mn)$ with negligible failure probability.*

Proof. Remark that for all i,

$$|\mathcal{L}_{i+1}| \geq (|\mathcal{L}_i| - (1+q/D_i)^{d_{i+1}-d_i})/2 \geq (|\mathcal{L}_i| - 2^{nx})/2.$$

Using induction, we then have $|\mathcal{L}_i| \geq (|\mathcal{L}_0| + 2^{nx})/2^i - 2^{nx}$ so that $|\mathcal{L}_k| \geq n2^{nx}$.

By induction and using the previous lemma, the input of Distinguish is sampled from a LWE distribution with noise parameter:

$$\alpha'^2 = 2^k\alpha^2 + 4\pi^2 \sum_{i=0}^{k-1} 2^{k-i-1} \sum_{j=d_i}^{d_{i+1}-1} (s_j D_i/q)^2.$$

By choice of k the first term is smaller than $nx/6$. As for the second term, since B is non increasing and by choice of D_i, it is smaller than:

$$4\pi^2 \sum_{i=0}^{k-1} 2^{k-i-1} \frac{x/6}{\pi^2 2^{k-i+1}} \sum_{j=d_i}^{d_{i+1}-1} \left(\frac{s_j}{B_j}\right)^2 \leq (x/6) \sum_{j=0}^{n-1} \left(\frac{s_j}{B_j}\right)^2 \leq nx/6.$$

Thus the real part of the bias is superior to $\exp(-nx/3)(1-3^a\epsilon) \geq 2^{-nx/2}$, and hence by Theorem 2.2, Distinguish fails with negligible probability. □

Theorem 4. *Assume that for all i, $|s_i| \leq B$, $B \geq 2$, $\max(\beta, \log(q)) = 2^{o(n/\log n)}$, $\beta = \omega(1)$, and $\epsilon \leq 1/\beta^4$. Then* Solve *takes time $2^{(n/2+o(n))/\ln(1+\log\beta/\log B)}$.*

Proof. We apply Lemma 7, choosing

$$k = \lfloor \log(\beta^2/(12\ln(1+\log\beta)))\rfloor = (2-o(1))\log\beta \in \omega(1)$$

and we set $D_i = q/(Bk2^{(k-i)/2})$. It now remains to show that this choice of parameters satisfies the conditions of the lemma.

First, observe that $BD_i/q \leq 1/k = o(1)$ so the condition $|s_j|D_i < 0.23q$ is fulfilled. Then, $d_k \geq n$, which amounts to:

$$\sum_{i=0}^{k-1} \frac{x}{(k-i)/2 + \log\mathcal{O}(kB)} \geq 2x\ln(1 + k/2/\log\mathcal{O}(kB)) \geq 1 + k/n = 1 + o(1)$$

If we have $\log k = \omega(\log\log B)$ (so in particular $k = \omega(\log B)$), we get $\ln(1 + k/2/\log\mathcal{O}(kB)) = (1+o(1))\ln(k) = (1+o(1))\ln(1+\log\beta/\log B)$.

Else, $\log k = \mathcal{O}(\log\log B) = o(\log B)$ (since necessarily $B = \omega(1)$ in this case), so we get $\ln(1 + k/2/\log \mathcal{O}(kB)) = (1 + o(1))\ln(1 + \log\beta/\log B)$.

Thus our choice of x fits both cases and we have $1/x \leq 2\ln(1+\log\beta)$. Second, we have $1/k = o(\sqrt{x})$ so D_i, ϵ and k are also sufficiently small and the lemma applies. Finally, note that the algorithm has complexity $2^{\Omega(n/\log n)}$, so a factor $n2^k \log(q)$ is negligible. □

This theorem can be improved when the use of the given parameters yields $D < 1$, since $D = 1$ already gives a lossless quantization.

Theorem 5. *Assume that for all i, $|s_i| \leq B = n^{b+o(1)}$. Let $\beta = n^c$ and $q = n^d$ with $d \geq b$ and $c + b \geq d$. Assume $\epsilon \leq 1/\beta^4$. Then* Solve *takes time $2^{n/(2(c-d+b)/d+2\ln(d/b)-o(1))}$.*

Proof. Once again we aim to apply Lemma 7, and choose k as above:

$$k = \log(\beta^2/(12\ln(1+\log\beta))) = (2c - o(1))\log n$$

If $i < \lceil 2(c-d+b)\log n\rceil$, we take $D_i = 1$, else we choose $q/D_i = \Theta(B2^{(a-i)/2})$. Satisfying $d_a \geq n - 1$ amounts to:

$$\begin{aligned}
&2x(c-d+b)\log n/\log q + \sum_{i=\lceil 2(c-d+b)\log n\rceil}^{a-1} \frac{x}{(a-i)/2 + \log \mathcal{O}(B)} \\
&\geq 2x(c-d+b)/d + 2x\ln((a - 2(c-d+b)\log n + 2\log B)/2/\log \mathcal{O}(B)) \\
&\geq 1 + a/n = 1 + o(1)
\end{aligned}$$

So that we can choose $1/x = 2(c-d+b)/d + 2\ln(d/b) - o(1)$. □

Corollary 1. *Given a* LWE *problem with $q = n^d$, Gaussian errors with $\beta = n^c$, $c > 1/2$ and $\epsilon \leq n^{-4c}$, we can find a solution in $2^{n/(1/d+2\ln(d/(1/2+d-c))-o(1))}$ time.*

Proof. Apply Theorem 1: with probability 2/3, the secret is now bounded by $B = \mathcal{O}(q\sqrt{n}/\beta\sqrt{\log n})$. The previous theorem gives the complexity of an algorithm discovering the secret, using $b = 1/2 - c + d$, and which works with probability $2/3 - 2^{-\Omega(n)}$. Repeating n times with different samples, the correct secret will be outputted at least $n/2+1$ times, except with negligible probability. By returning the most frequent secret, the probability of failure is negligible. □

In particular, if $c \leq d$, it is possible to quantumly approximate lattice problems within factor $\mathcal{O}(n^{c+1/2})$ [34]. Setting $c = d$, the complexity is $2^{n/(1/c+2\ln(2c)-o(1))}$, so that the constant slowly converges to 0 when c goes to infinity.

A simple BKW using the bias would have a complexity of $2^{d/cn+o(n)}$, the analysis of [4] or [3] only conjectures $2^{dn/(c-1/2)+o(n)}$ for $c > 1/2$. In [4], the authors incorrectly claim a complexity of $2^{cn+o(n)}$ when $c = d$, because the blowup in the error is not explicitely computed.

Finally, if we want to solve the LWE problem for different secrets but with the same vectorial part of the samples, it is possible to be much faster if we work with a bigger final bias, since the Reduce part needs to be called only once.

3.4 Experimentation

We have implemented our algorithm, in order to test its efficiency in practice, as well as that of the practical improvements in the appendix of the full version. We have chosen dimension $n = 128$, modulus $q = n^2$, binary secret, and Gaussian errors with noise parameter $\alpha = 1/(\sqrt{n/\pi}\log^2 n)$. The previous best result for these parameters, using a BKW algorithm with lazy modulus switching, claims a time complexity of 2^{74} with 2^{60} samples [4].

Using our improved algorithm, we were able to recover the secret using $m = 2^{28}$ samples within 13 hours on a single PC equipped with a 16-core Intel Xeon. The computation time proved to be devoted mostly to the computation of $9 \cdot 10^{13}$ norms, computed in fixed point over 16 bits in SIMD.

In appendix of the full version, we compare the different techniques to solve the LWE problem when the number of samples is large or small. We were able to solve the same problem using BKZ with block size 40 followed by an enumeration in two minutes.

4 Applications to Lattice Problems

We first show that $\mathsf{BDD}_{B,\beta}^{||\cdot||_\infty}$ is easier than $\mathsf{LWE}_{B,\beta}$ for some large enough modulus and then that $\mathsf{UniqueSVP}_{B,\beta}^{||\cdot||_\infty}$ and $\mathsf{GapSVP}_{B,\beta}^{||\cdot||_\infty}$ are easier than $\mathsf{BDD}_{B,\beta}^{||\cdot||_\infty}$. In appendix of the full version, we prove the same result for $\mathsf{BDD}_{B,\beta}^{||\cdot||}$.

4.1 Variant of Bounding Distance Decoding

The main result of this subsection is close to the classic reduction of [34]. However, our definition of LWE allows to simplify the proof, and gain a constant factor in the decoding radius. The use of the KL divergence instead of the statistical distance also allows to gain a constant factor, when we need an exponential number of samples, or when λ_n^* is really small.

The core of the reduction lies in Lemma 8, assuming access to a Gaussian sampling oracle. This hypothesis will be taken care of in Lemma 9.

Lemma 8. *Let* $\mathbf{A}$ *be a basis of the lattice* Λ *of full rank* n*. Assume we are given access to an oracle outputting a vector sampled under the law* $D_{\Lambda^*,\sigma}$ *and* $\sigma \geq q\eta_\epsilon(\Lambda^*)$*, and to an oracle solving the* LWE *problem in dimension* n*, modulus* $q \geq 2$*, noise parameter* α*, and distortion parameter* ξ *which fails with negligible probability and use* m *vectors if the secret* $\mathbf{s}$ *verifies* $|s_i| \leq B_i$*.*

Then, if we are given a point $\mathbf{x}$ *such that there exists* $\mathbf{s}$ *with* $\mathbf{v} = \mathbf{As} - \mathbf{x}$*,* $||\mathbf{v}|| \leq \sqrt{1/\pi}\alpha q/\sigma$*,* $|s_i| \leq B_i$ *and* $\rho_{\sigma/q}(\Lambda \setminus \{\mathbf{0}\} + \mathbf{v}) \leq \xi \exp(-\alpha^2)/2$*, we are able to find* $\mathbf{s}$ *in at most* mn *calls to the Gaussian sampling oracle,* n *calls to the* LWE *solving oracle, with a probability of failure* $n\sqrt{m}\epsilon + 2^{-\Omega(n)}$ *and complexity* $\mathcal{O}(mn^3 + n^c)$ *for some* c*.*

In the previous lemma, we required access to a $D_{\Lambda^*,\sigma}$ oracle. However, for large enough σ, this hypothesis comes for free, as shown by the following lemma, which we borrow from [10].

Lemma 9. *If we have a basis* $\mathbf{A}$ *of the lattice* Λ*, then for* $\sigma \geq \mathcal{O}(\sqrt{\log n}||\widetilde{\mathbf{A}}||)$*, it is possible to sample in polynomial time from* $D_{\Lambda,\sigma}$*.*

We will also need the following lemma, due to Banaszczyk [7]. For completeness, a proof is provided in the appendix of the full version.

Lemma 10. *For a lattice* Λ*,* $\mathbf{c} \in \mathbb{R}^n$*, and* $t \geq 1$*,*

$$\frac{\rho\big((\Lambda+\mathbf{c}) \setminus \mathcal{B}\big(0, t\sqrt{\frac{n}{2\pi}}\big)\big)}{\rho(\Lambda)} \leq \exp\big(-n(t^2 - 2\ln t - 1)/2\big) \leq \exp\big(-n(t-1)^2/2\big).$$

Theorem 6. *Assume we have a* LWE *solving oracle of modulus* $q \geq 2^n$*, parameters* β *and* ξ *which needs* m *samples.*

If we have a basis $\mathbf{A}$ *of the lattice* Λ*, and a point* $\mathbf{x}$ *such that* $\mathbf{As}-\mathbf{x} = \mathbf{v}$ *with* $||\mathbf{v}|| \leq (1-1/n)\lambda_1/\beta/t < \lambda_1/2$ *and* $4\exp(-n(t-1/\beta-1)^2/2) \leq \xi\exp(-n/2/\beta^2)$*, then with* n^2 *calls to the* LWE *solving oracle with secret* $\mathbf{s}$*, we can find* $\mathbf{s}$ *with probability of failure* $2\sqrt{m}\exp(-n(t^2 - 2\ln t - 1)/2)$ *for any* $t \geq 1 + 1/\beta$*.*

Proof. Using Lemma 10, we can prove that $\sigma = t\sqrt{n/2/\pi}/\lambda_1 \leq \eta_\epsilon(\Lambda^*)$ for $\epsilon = 2\exp(-n(t^2 - 2\ln t - 1)/2)$ and

$$\rho_{1/\sigma}\big(\Lambda \setminus \{\mathbf{0}\} + \mathbf{v}\big) \leq 2\exp\big(-n(t(1-1/\beta/t)-1)^2/2\big).$$

Using LLL, we can find a basis $\mathbf{B}$ of Λ so that $||\widetilde{\mathbf{B}^*}|| \leq 2^{n/2}/\lambda_1$, and therefore, it is possible to sample in polynomial time from $D_{\Lambda,q\sigma}$ since $q \geq 2^n$ for sufficiently large n.

The LLL algorithm also gives a non zero lattice vector of norm $\ell \leq 2^n\lambda_1$. For i from 0 to n^2, we let $\lambda = \ell(1-1/n)^i$, we use the algorithm of Lemma 8 with standard deviation $tq\sqrt{n/2/\pi}/\lambda$, which uses only one call to the LWE solving oracle, and return the closest lattice vector of $\mathbf{x}$ in all calls.

Since $\ell(1-1/n)^{n^2} \leq 2^n\exp(-n)\lambda_1 \leq \lambda_1$, with $0 \leq i \leq n^2$ be the smallest integer such that $\lambda = \ell(1-1/n)^i \leq \lambda_1$, we have $\lambda \geq (1-1/n)\lambda_1$. Then the lemma applies since

$$||\mathbf{v}|| \leq (1-1/n)\lambda_1/\beta/t \leq \sqrt{1/\pi}\sqrt{n/2}/\beta q/(tq\sqrt{n/2/\pi}/\lambda) = \lambda/t/\beta.$$

Finally, the distance bound makes $\mathbf{As}$ the unique closest lattice point of $\mathbf{x}$. □

Using self-reduction, it is possible to remove the $1-1/n$ factor [27].

Corollary 2. *It is possible to solve* $\mathsf{BDD}^{||.||_\infty}_{B,\beta}$ *in time* $2^{(n/2+o(n))/\ln(1+\log\beta/\log B)}$ *if* $\beta = \omega(1)$*,* $\beta = 2^{o(n/\log n)}$ *and* $\log B = \mathcal{O}(\log\beta)$*.*

Proof. Apply the previous theorem and Theorem 4 with some sufficiently large constant for t, and remark that dividing β by some constant does not change the complexity. □

Note that since we can solve LWE for many secrets in essentially the same time than for one, we have the same property for BDD.

4.2 UniqueSVP and GapSVP

In this section, we show how $\mathsf{GapSVP}^{||\cdot||_\infty}_{B,\beta}$ and $\mathsf{UniqueSVP}^{||\cdot||_\infty}_{B,\beta}$ can be reduced to $\mathsf{BDD}^{||\cdot||_\infty}_{B,\beta}$, and hence to LWE. Proofs are provided in the appendix of the full version.

Theorem 7. *Given a* $\mathsf{BDD}^{||\cdot||_\infty}_{B,\beta}$ *oracle, it is possible to solve* $\mathsf{UniqueSVP}^{||\cdot||_\infty}_{B,\beta}$ *in polynomial time of* n *and* β.

Theorem 8. *We can solve any* $\mathsf{GapSVP}^{||\cdot||_\infty}_{o(B\sqrt{\log\log\log\beta/\log\log\beta}),\beta}$ *instances in time* $2^{(n/2+o(n))/\ln(1+\log\beta/\log B)}$ *for* $\beta = 2^{o(n/\log n)}$, $\beta = \omega(1)$, $B \geq 2$.

Corollary 3. *It is possible to solve any* $\mathsf{GapSVP}^{||\cdot||_\infty}_{2^{\sqrt{\log n}},n^c}$ *with* $c > 0$ *in time* $2^{(n+o(n))/\ln\ln n}$.

Proof. Use Theorem 8 with $B = 2^{\sqrt{\log n}}\log\log n$ and $\beta = n^c$. □

Theorem 9. *If it is possible to solve* $\mathsf{BDD}^{||\cdot||_\infty}_{B,\beta}$ *in polynomial time, then it is possible to solve in randomized polynomial time* $\mathsf{GapSVP}^{||\cdot||_\infty}_{B/\sqrt{n},\beta\sqrt{n/\log n}}$.

5 Other Applications

5.1 Low Density Subset-Sum Problem

Definition 10. *We are given a vector* $\mathbf{a} \in \mathbb{Z}^n$ *whose coordinates are sampled independently and uniformly in* $[0; M)$, *and* $\langle \mathbf{a}, \mathbf{s}\rangle$ *where the coordinates of* $\mathbf{s}$ *are sampled independently and uniformly in* $\{0, 1\}$. *The goal is to find* $\mathbf{s}$. *The density is defined as* $d = \frac{n}{\log M}$.

Note that this problem is trivially equivalent to the *modular* subset-sum problem, where we are given $\langle \boldsymbol{a}, \boldsymbol{s}\rangle \bmod M$ by trying all possible $\lfloor \langle \boldsymbol{a}, \boldsymbol{s}\rangle / M \rfloor$.

In [13,22], Lagarias *et al.* reduce the subset sum problem to UniqueSVP, even though this problem was not defined at that time. We will show a reduction to $\mathsf{BDD}^{||\cdot||_\infty}_{1,\Omega(2^{1/d})}$, which is essentially the same. First, we need two geometric lemmata.

Lemma 11. *Let* $\mathcal{B}_n(r)$, *the number of points of* $\mathbb{Z}^n$ *of norm smaller than* r, *and* V_n *the volume of the unit ball. Then,*

$$\mathcal{B}_n(r) \leq V_n\left(r + \frac{\sqrt{n}}{2}\right)^n.$$

Proof. For each $\mathbf{x} \in \mathbb{Z}^n$, let $E_{\mathbf{x}}$ be a cube of length 1 centered on $\mathbf{x}$. Let E be the union of all the $E_{\mathbf{x}}$ which have a non empty intersection with the ball of center $\mathbf{0}$ and radius r. Therefore $\mathrm{vol}(E) \geq \mathcal{B}_n(r)$ and since E is included in the ball of center $\mathbf{0}$ and radius $r + \frac{\sqrt{n}}{2}$, the claim is proven. □

Lemma 12. *For $n \geq 4$ we have*

$$V_n = \frac{\pi^{n/2}}{(n/2)!} \leq (\sqrt{\pi \mathrm{e}/n})^n.$$

Theorem 10. *Using one call to a $\mathsf{BDD}^{||\cdot||_\infty}_{1,c2^{1/d}}$ oracle with any $c < \sqrt{2/\pi/\mathrm{e}}$ and $d = o(1)$, and polynomial time, it is possible to solve a subset-sum problem of density d, with negligible probability of failure.*

Proof. With the matrix:

$$\mathbf{A} = \begin{pmatrix} \mathbf{I} \\ C\mathbf{a} \end{pmatrix}$$

for some $C > c2^{1/d}\sqrt{n}/2$ and $\mathbf{b} = (1/2, \ldots, 1/2, C\langle \mathbf{a}, \mathbf{s} \rangle)$, return $\mathsf{BDD}(\mathbf{A}, \mathbf{b})$. It is clear that $||\mathbf{As} - \mathbf{b}|| = \sqrt{n}/2$. Now, let $\mathbf{x}$ such that $||\mathbf{Ax}|| = \lambda_1$. If $\langle \mathbf{a}, \mathbf{x} \rangle \neq 0$, then $\lambda_1 = ||\mathbf{Ax}|| \geq C$ therefore $\beta \geq c2^{1/d}$. Else, $\langle \mathbf{a}, \mathbf{x} \rangle = 0$. Without loss of generality, $x_0 \neq 0$, we let $y = -(\sum_{i>0} a_i x_i)/x_0$ and the probability over $\mathbf{a}$ that $\langle \mathbf{a}, \mathbf{x} \rangle = 0$ is:

$$\Pr[\langle \mathbf{a}, \mathbf{x} \rangle = 0] = \Pr[a_0 = y] = \sum_{z=0}^{M-1} \Pr[y = z] \Pr[a_0 = z] \leq \frac{1}{M}.$$

Therefore, the probability of failure is at most, for sufficiently large n,

$$\begin{aligned} \mathcal{B}_n(\beta\sqrt{n}/2)/M \leq & (\sqrt{\pi \mathrm{e}/n})^n (c2^{1/d}\sqrt{n}/2 + \sqrt{n}/2)^n / 2^{n/d} \\ = & \left(\sqrt{\pi \mathrm{e}/2}(c + 2^{-1/d})\right)^n = 2^{-\Omega(n)}. \end{aligned}$$

□

Corollary 4. *For any $d = o(1)$ and $d = \omega(\log n/n)$, we can solve the subset-sum problem of density d with negligible probability of failure in time $2^{(n/2+o(n))/\ln(1/d)}$.*

The cryptosystem of Lyubashevsky *et al.* [28] uses $2^{1/d} > 10n\log^2 n$ and is therefore broken in time $2^{(\ln 2/2+o(1))n/\log\log n}$. Current lattice reduction algorithms are slower than this one when $d = \omega(1/(\log n \log\log n))$.

5.2 Sample Expander and Application to LWE with Binary Errors

Definition 11. *Let q be a prime number. The problem* Small-DecisionLWE *is to distinguish $(\mathbf{A}, \mathbf{b})$ with $\mathbf{A}$ sampled uniformly with n columns and m rows, $\mathbf{b} = \mathbf{As} + \mathbf{e}$ such that $||\mathbf{s}||^2 + ||\mathbf{e}||^2 \leq nk^2$ and $||\mathbf{s}||_\infty \leq B$ from $(\mathbf{A}, \mathbf{b})$ sampled uniformly. Also, the distribution $(\mathbf{s}, \mathbf{e})$ is efficiently samplable.*

The problem Small-SearchLWE *is to find $\mathbf{s}$ given $(\mathbf{A}, \mathbf{b})$ with $\mathbf{A}$ sampled uniformly and $\mathbf{b} = \mathbf{As} + \mathbf{e}$ with the same conditions on $\mathbf{s}$ and $\mathbf{e}$.*

These problems are generalizations of BinaryLWE where $\mathbf{s}$ and $\mathbf{e}$ have coordinates sampled uniformly in $\{0, 1\}$. In this case, remark that each sample is a root of a known quadratic polynomial in the coordinates of $\mathbf{s}$. Therefore, it is easy

to solve this problem when $m \geq n^2$. For $m = \mathcal{O}(n)$, a Gröbner basis algorithm applied on this system will (heuristically) have a complexity of $2^{\Omega(n)}$ [2]. For $m = \mathcal{O}(n/\log n)$ and $q = n^{\mathcal{O}(1)}$, it has been shown to be harder than a lattice problem in dimension $\Theta(n/\log n)$ [30].

In appendix of the full version, we prove the following theorem[3], with the coordinates of $\mathbf{x}$ and $\mathbf{y}$ distributed according to a samplable $\mathcal{D}$:

Theorem 11. *Assume there is an efficient distinguisher which uses k samples for* Decision-LWE *(respectively a solver for* Search-LWE*) with error distribution $\langle \mathbf{s}, \mathbf{y} \rangle + \langle \mathbf{e}, \mathbf{x} \rangle$ of advantage (resp. success probability) ϵ.*

Then, either there is an efficient distinguisher for Decision-LWE *with samples and secret taken uniformly, and error distribution $\mathcal{D}$ in dimension $m-1$ and with $n+m$ samples of advantage $\frac{\xi}{4qk} - q^{-n} - q^{-m}$; or there is an efficient distinguisher of advantage $\epsilon - \xi$ for* Small-Decision-LWE *(resp. solver of success probability $\epsilon - \xi$ for* Small-Search-LWE*).*

Lemma 13. *Let $\mathcal{D} = D_{\mathbb{Z},\sigma}$ for $\sigma \geq 1$. Then, the advantage of a distinguisher for* Decision-LWE *of dimension m with $m+n$ samples of noise distribution $\mathcal{D}$ is at most $\sqrt{q^n/\sigma^{n+m}}$. Furthermore, the bias of $\langle (\mathbf{s}|\mathbf{e}), (\mathbf{x}|\mathbf{y}) \rangle$, for fixed $\mathbf{s}$ and $\mathbf{e}$, is at least $\exp(-\pi(||\mathbf{s}||^2 + ||\mathbf{e}||^2)\sigma^2/q^2)$.*

Proof. We have $\mathcal{D}^{m+n}(\mathbf{a}) \leq \mathcal{D}(0)^{m+n} = 1/\rho_\sigma(\mathbb{Z})^{m+n}$ and $\rho_\sigma(\mathbb{Z}) = \sigma \rho_{1/\sigma}(\mathbb{Z}) \geq \sigma$ using a Poisson summation. The first property is then a direct application of the leftover hash lemma, since q is prime.

The bias of $\lambda\mathcal{D}$ can be computed using a Poisson summation as:

$$\sum_{a \in \mathbb{Z}} \rho_\sigma(a) \cos(2\pi\lambda a/q) = \rho_{1/\sigma}(\mathbb{Z} + \lambda/q) \geq \exp(-\pi\lambda^2\sigma^2/q^2).$$

Therefore, the second property follows from the independency of the coordinates of $\mathbf{x}$ and $\mathbf{y}$. □

Corollary 5. *Let q, n and m such that $m \log q/(n+m) = o(n/\log n)$, then $(m-3)\log q/(n+m) - \log k = \omega(\log B)$ and $m = \omega(1)$. Then, we can solve the* Small-Decision-LWE *problem in time*

$$2^{(n/2+o(n))/\ln((m \log q/(n+m) - \log k)/\log B)}$$

with negligible probability of failure.

Proof. We use the previous lemma with $\sigma = 2q^{(n+2)/(n+m-1)}$, so that we have $\beta = \Omega(q^{(m-3)/(n+m)}/k)$. The algorithm from Theorem 4 needs $2^{o(n)}$ samples, so the advantage of the potential distinguisher for Decision-LWE is $2^{-(1/4+o(1))n}/q$ for $\xi = 2^{-n/4}$; while the previous lemma proves it is less than $2^{-n/2}/q$. □

[3] The authors of [15] gave a short justification of a similar claim which is far from proven.

The NTRU cryptosystem [20] is based on the hardness of finding two polynomials f and g whose coefficients are bounded by 1 given $h = f/g \bmod (X^n - 1, q)$. Since $hg = 0$ with an error bounded by 1, we can apply previous algorithms in this section to *heuristically* recover f and g in time $2^{(n/2+o(n))/\ln\ln q}$. This is the first subexponential time algorithm for this problem since it was introduced back in 1998.

Corollary 6. *Assume we have a* Search-LWE *problem with* $n \log q + \Omega(n/\log q)$ *samples and Gaussian noise with* $\alpha = n^{-c}$ *and* $q = n^d$. *Then, we can solve it in time* $2^{n/(2\ln(d/(d-c))-o(1))}$ *for any failure probability in* $2^{-n^{o(1)}}$.

Proof. First, apply a secret-error switching (Theorem 1). Apply the previous corollary with $B = n^{d-c+o(1)}$ which is a correct bound for the secret, except with probability $2^{-n^{o(1)}}$. Lemma 10 shows that $k^2 \leq \log q\sigma^2$, except with probability $2^{-\Omega(n)}$, so that $\beta = n^{c+o(1)}$. We can then use $\sigma = \Theta(1)$ and apply Theorem 4. □

Note that this corollary can in fact be applied to a very large class of distributions, and in particular to the learning with rounding problem, while the distortion parameter is too large for a direct application of Theorem 4.

Also, if the reduction gives a fast (subexponential) algorithm, one may use $\sigma = 2\sqrt{n}$ and assume that there is no quantum algorithm solving the corresponding lattice problem in dimension m.

Even more heuristically, one can choose σ to be the lowest such that if the reduction does not work, we have an algorithm faster than the best *known* algorithm for the same problem.

References

1. Proceedings of the 25th Annual IEEE Conference on Computational Complexity, CCC 2010, Cambridge, Massachusetts, 9–12 June 2010. IEEE Computer Society (2010)
2. Albrecht, M.R., Cid, C., Faugère, J., Fitzpatrick, R., Perret, L.: Algebraic algorithms for LWE problems. IACR Cryptology ePrint Arch. **2014**, 1018 (2014)
3. Albrecht, M.R., Cid, C., Faugère, J., Fitzpatrick, R., Perret, L.: On the complexity of the BKW algorithm on LWE. Des. Codes Crypt. **74**(2), 325–354 (2015)
4. Albrecht, M.R., Faugère, J.-C., Fitzpatrick, R., Perret, L.: Lazy modulus switching for the BKW algorithm on LWE. In: Krawczyk, H. (ed.) PKC 2014. LNCS, vol. 8383, pp. 429–445. Springer, Heidelberg (2014)
5. Applebaum, B., Cash, D., Peikert, C., Sahai, A.: Fast cryptographic primitives and circular-secure encryption based on hard learning problems. In: Halevi [19], pp. 595–618
6. Arora, S., Ge, R.: New algorithms for learning in presence of errors. In: Aceto, L., Henzinger, M., Sgall, J. (eds.) ICALP 2011, Part I. LNCS, vol. 6755, pp. 403–415. Springer, Heidelberg (2011)
7. Banaszczyk, W.: New bounds in some transference theorems in the geometry of numbers. Math. Ann. **296**(1), 625–635 (1993)
8. Bernstein, D.J., Lange, T.: Never trust a bunny. In: Hoepman, J.-H., Verbauwhede, I. (eds.) RFIDSec 2012. LNCS, vol. 7739, pp. 137–148. Springer, Heidelberg (2013). https://eprint.iacr.org/2012/355.pdf

9. Blum, A., Kalai, A., Wasserman, H.: Noise-tolerant learning, the parity problem, and the statistical query model. J. ACM **50**(4), 506–519 (2003)
10. Brakerski, Z., Langlois, A., Peikert, C., Regev, O., Stehlé, D.: Classical hardness of learning with errors. In: Symposium on Theory of Computing Conference, STOC 2013, pp. 575–584 (2013. http://perso.ens-lyon.fr/damien.stehle/downloads/LWE.pdf
11. Brakerski, Z., Vaikuntanathan, V.: Efficient fully homomorphic encryption from (standard) LWE. SIAM J. Comput. **43**(2), 831–871 (2014)
12. Chen, Y., Nguyen, P.Q.: BKZ 2.0: better lattice security estimates. In: Lee, D.H., Wang, X. (eds.) ASIACRYPT 2011. LNCS, vol. 7073, pp. 1–20. Springer, Heidelberg (2011)
13. Coster, M.J., Joux, A., LaMacchia, B.A., Odlyzko, A.M., Schnorr, C., Stern, J.: Improved low-density subset sum algorithms. Comput. Complex. **2**, 111–128 (1992)
14. Döttling, N., Müller-Quade, J.: Lossy codes and a new variant of the learning-with-errors problem. In: Johansson, T., Nguyen, P.Q. (eds.) EUROCRYPT 2013. LNCS, vol. 7881, pp. 18–34. Springer, Heidelberg (2013)
15. Duc, A., Tramèr, F., Vaudenay, S.: Better algorithms for lwe and LWR. Cryptology ePrint Archive, Report 2015/056 (2015). http://eprint.iacr.org/
16. Gentry, C., Sahai, A., Waters, B.: Homomorphic encryption from learning with errors: conceptually-simpler, asymptotically-faster, attribute-based. In: Canetti, R., Garay, J.A. (eds.) CRYPTO 2013, Part I. LNCS, vol. 8042, pp. 75–92. Springer, Heidelberg (2013)
17. Gray, R.M., Neuhoff, D.L.: Quantization. IEEE Trans. Inf. Theor. **44**(6), 2325–2383 (1998)
18. Guo, Q., Johansson, T., Löndahl, C.: Solving LPN using covering codes. In: Sarkar, P., Iwata, T. (eds.) ASIACRYPT 2014. LNCS, vol. 8873, pp. 1–20. Springer, Heidelberg (2014)
19. Halevi, S. (ed.): CRYPTO 2009. LNCS, vol. 5677. Springer, Heidelberg (2009)
20. Hoffstein, J., Pipher, J., Silverman, J.H.: NTRU: a ring-based public key cryptosystem. In: Buhler, J.P. (ed.) ANTS 1998. LNCS, vol. 1423, pp. 267–288. Springer, Heidelberg (1998)
21. Kirchner, P.: Improved generalized birthday attack. IACR Cryptology ePrint Arch. **2011**, 377 (2011). http://eprint.iacr.org/2011/377.pdf
22. Lagarias, J.C., Odlyzko, A.M.: Solving low-density subset sum problems. J. ACM **32**(1), 229–246 (1985)
23. Lenstra, A., Lenstra, J.H., Lovász, L.: Factoring polynomials with rational coefficients. Mathe. Ann. **261**, 515–534 (1982)
24. Levieil, É., Fouque, P.-A.: An improved LPN algorithm. In: De Prisco, R., Yung, M. (eds.) SCN 2006. LNCS, vol. 4116, pp. 348–359. Springer, Heidelberg (2006)
25. Lindner, R., Peikert, C.: Better key sizes (and attacks) for LWE-based encryption. In: Kiayias, A. (ed.) CT-RSA 2011. LNCS, vol. 6558, pp. 319–339. Springer, Heidelberg (2011)
26. Liu, M., Nguyen, P.Q.: Solving BDD by enumeration: an update. In: Dawson, E. (ed.) CT-RSA 2013. LNCS, vol. 7779, pp. 293–309. Springer, Heidelberg (2013)
27. Lyubashevsky, V., Micciancio, D.: On bounded distance decoding, unique shortest vectors, and the minimum distance problem. In: Halevi [19], pp. 577–594
28. Lyubashevsky, V., Palacio, A., Segev, G.: Public-key cryptographic primitives provably as secure as subset sum. In: Micciancio, D. (ed.) TCC 2010. LNCS, vol. 5978, pp. 382–400. Springer, Heidelberg (2010)
29. Lyubashevsky, V., Peikert, C., Regev, O.: On ideal lattices and learning with errors over rings. J. ACM **60**(6), 43 (2013)

30. Micciancio, D., Peikert, C.: Hardness of SIS and LWE with small parameters. In: Canetti, R., Garay, J.A. (eds.) CRYPTO 2013, Part I. LNCS, vol. 8042, pp. 21–39. Springer, Heidelberg (2013)
31. Mulder, E.D., Hutter, M., Marson, M.E., Pearson, P.: Using Bleichenbacher's solution to the hidden number problem to attack nonce leaks in 384-bit ECDSA: extended version. J. Crypt. Eng. **4**(1), 33–45 (2014)
32. Peikert, C.: Public-key cryptosystems from the worst-case shortest vector problem: extended abstract. In: Mitzenmacher, M. (ed.) Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31–June 2 2009, pp. 333–342. ACM (2009)
33. Pöppelmann, T., Ducas, L., Güneysu, T.: Enhanced lattice-based signatures on reconfigurable hardware. IACR Cryptology ePrint Arch. **2014**, 254 (2014). https://eprint.iacr.org/2014/254.pdf
34. Regev, O.: On lattices, learning with errors, random linear codes, and cryptography. J. ACM (JACM) **56**(6), 34 (2009). http://www.cims.nyu.edu/regev/papers/qcrypto.pdf
35. Wagner, D.: A generalized birthday problem. In: Yung, M. (ed.) CRYPTO 2002. LNCS, vol. 2442, pp. 288–303. Springer, Heidelberg (2002)