Enhancing the Dual Attack against MLWE: Constructing More Short Vectors Using Its Algebraic Structure

Han Wu[a,b], Guangwu Xu[a,c,d,*]

[a] School of Cyber Science and Technology, Shandong University, Qingdao, 266237, China
[b] Key Laboratory of Cryptologic Technology and Information Security of Ministry of Education, Shandong University, Qingdao, 266237, China
[c] Quancheng Laboratory, Jinan, 250103, China
[d] Shandong Institute of Block-chain, Jinan, 250101, China

Abstract

Primal attack, BKW attack, and dual attack are three well-known attacks to LWE. To build efficient post-quantum cryptosystems in practice, the structured variants of LWE (i.e. MLWE/RLWE) are often used. Some efforts have been spent on addressing concerns about additional vulnerabilities introduced by algebraic structures and no effective attack method based on ideal lattices or module lattices has been proposed so far; these include refining primal attack and BKW attack to MLWE/RLWE. It is thus an interesting problem to consider how to enhance the dual attack against LWE with the rich algebraic structure of MLWE (including RLWE). In this paper, we present the first attempt to this problem by observing that each short vector found by BKZ generates another $n-1$ vectors of the same length automatically and all of these short vectors can be used to distinguish. To this end, an interesting property which indicates the rotations are consistent with certain linear transformations is proved, and a new kind of intersection lattice is constructed with some tricks. Moreover, we notice that coefficient vectors of different rotations of the same polynomial are near-orthogonal in high-dimensional spaces. This is validated by extensive experiments and is treated as an extension to the assumption under the original dual attack against LWE. Taking Newhope512 as an example, we show that by our enhanced dual attack method, the required blocksize and time complexity (in both classical and quantum cases) all decrease. It is remarked that our improvement is not significant and its limitation is also touched on. Our results do not reveal a severe security problem for MLWE/RLWE compared to that of a general LWE, this is consistent with the findings by the previous work for using primal and BKW attacks to MLWE/RLWE.

Keywords: Lattice-based cryptography, dual attack, Module-LWE

1. Introduction

The rapid development of computing (especially quantum computing) technology has made a huge impact on the widely used cryptographic applications at present. For examples, systems based on factorization and discrete logarithm are no longer secure in quantum computing framework, so gradual replacement is necessary. It is believed that several fundamental lattice problems are resistant to quantum attacks, among them, the learning with error (LWE) problem introduced by Regev [1] in 2005 is considered as one of the most promising choices for establishing secure and reliable post-quantum cryptosystems.

For LWE schemes, the security is not the only goal, a great attention has been paid to its efficiency as well. In 2010, an algebraic variant of LWE called Ring-LWE (RLWE) was proposed by Lyubashevsky et al.[2]. It answered the open question whether extra algebraic structure can be used to promote the efficiency of LWE and its applications. RLWE has many attractive features, for example, the size of the public key is reduced by a factor of n. The Module-LWE (MLWE) problem was first introduced in [3], and then thoroughly studied in [4]. Informally speaking, it replaces the single ring elements with module elements over the same ring. So RLWE can be seen as a special case of MLWE. These two variants have received more and more popularity, and several candidate algorithms based on them have been presented, for examples, RLWE-based public key cryptographic schemes Newhope[5] and LAC[6], MLWE-based key encapsulation mechanism Kyber[7], as well as MLWE-based signature scheme Dilithium [8]. It is worth noting that the last two algorithms have been selected as NIST PQC standard.

The MLWE/RLWE achieve performance improvements, but there are security concerns about additional vulnerabilities with the injection of number-theoretical structure. To date, the theoretical foundation of RLWE has only been shown to be the hardness of approximate-SVP on arbitrary ideals, usually in a cyclotomic ring and for near-polynomial approximation factors, (but not known to be equivalent). It was found in [9, 10, 11, 12] that there exists an asymptotic gap between the search of mildly short vectors in general lattices and in certain structured lattices (i.e. ideal lattices), in other words, the search for short vectors in such lattices (i.e. Ideal-SVP) is easier. However, these proposed algorithms do not lead to practical attacks. On the one

*Corresponding author
Email addresses: hanwu97@mail.sdu.edu.cn (Han Wu), gxu4sdq@sdu.edu.cn (Guangwu Xu)

hand, the approximation factors in these methods are too large to affect any actual RLWE schemes. On the other hand, the gap between Ideal-SVP and RLWE is still unclear. In certain sense, MLWE can be viewed as an interpolation between LWE and RLWE. It is believed that MLWE provides both a better security level than RLWE, and a better performance than LWE.

Since no effective attack method based on ideal lattices or module lattices has been put forward so far, the security of the schemes based on MLWE (including RLWE) is usually evaluated by first converting the underlying MLWE instance into an LWE one and then attacking it by the best known algorithms against LWE. It is thus an interesting problem to consider how to enhance the existing attacks against LWE with the rich algebraic structure of MLWE. Actually, there have been some such attempts. It is noted that there are three major attacks on LWE, namely the primal attack, the dual attack, and the BKW attack. In 2021, Nakamura and Yasuda [13] proposed a new kind of extended lattice which contains multiple short vectors. They found that performing the decoding attack (via Kannan's embedding [14]) or the primal attack (via Bai-Galbraith embedding [15]) on the new lattice increases the probability of finding a target vector by pruned ENUM. However, their increased success rate sometimes is at the cost of enlarging blocksize. Ring-BKW, a version of the Blum-Kalai-Wasserman algorithm which respects the ring structure was presented in [16]. Its primary advantage is that there is no need for back-substitution, and the hypothesis testing phase supports parallel processing. Even so, only a polynomial factor speedup can be expected.

We fill the gap for the dual attack. To be specific, we improve the original dual attack against LWE by exploring special algebraic structure of MLWE (including RLWE). It is observed that each MLWE sample produces n LWE samples. A relatively simple case is discussed first, where the number of LWE samples used (which is denoted by m) is a multiple of n. An interesting property has been proven, which indicates that the rotations are consistent with certain linear transformations. It shows that with any short vector found by BKZ, the adversary could construct $n-1$ additional short vectors automatically (instead of using BKZ). With some tricks, this conclusion is extended to the case of a general m by considering a new kind of intersection lattice. In summary, we propose an enhanced dual attack method in this paper. Short vectors are first searched in the new lattice, and then the number of the vectors increases to n times the original size by "rotation" operation. Finally, all of these short vectors can be used to make a distinction.

Since the coefficient vectors of different rotations of the same polynomial vector are all included, we suppose that they are near-orthogonal, for sample independence. This could be regarded as an extension of the assumption under the original dual attack against LWE. We give some theoretical explanations as well as extensive experimental validation for this assumption. By our improved approach, the required blocksize and time complexity (in both classical and quantum cases) all decrease. It is remarked that our improvement is not significant and its limitation is also touched on. In conclusion, our results do not reveal a severe security problem for MLWE/RLWE compared to that of a general LWE, we notice that this is consistent with the findings by the previous work for using primal and BKW attacks to MLWE/RLWE.

The remaining of the paper is organized as follows. Section 2 contains necessary notations, properties as well as useful algorithms. In the first half of Section 3, we describe a way of constructing more short vectors of the same length using the ones found by BKZ. The complete enhanced dual attack method against MLWE is given in the second half of Section 3, as well as its complexity and correlation analysis. Experiments with this improved algorithm are conducted in Section 4, corresponding results and some explanations are both provided.

2. Preliminaries

For any distribution D, we use $x \leftarrow D$ to express that x is sampled according to D. The uniform distribution over some set X is written as $U(X)$. For any matrix A, we denote the submatrix formed by its i-th to j-th rows by $A_{[i:j]}$. While $v_{[i:j]}$ represents the subvector that contains the i-th to the j-th entries of some vector v. For any matrix $B \in \mathbb{R}^{m\times n}$, $L(B)$ is the lattice generated by B. For two vectors $y, z \in \mathbb{R}^d$, let $\theta(y, z)$ be the angle between them.

2.1. Statistics

The Gaussian distribution is one of the most important distributions in lattice cryptography, as in most schemes based on LWE or its variants, the coefficients of the error (sometimes as well as the secret) are independently picked up from the discrete Gaussian distribution.

Let $\sigma > 0$. For any $x \in \mathbb{R}$, the density of the d-dimensional *centered Gaussian distribution* with standard deviation σ (i.e. N_σ^d) is defined as $\rho_\sigma^d(x) = \frac{1}{(2\pi)^{\frac{d}{2}}\sigma^d} \cdot e^{-\frac{\|x\|^2}{2\sigma^2}}$. Reducing it modulo q gives the *discrete Gaussian distribution* $G_{\sigma,q}^d$ over $\mathbb{Z}_q^d$, whose probability mass function (pmf) is $g_{\sigma,q}^d(x) = \frac{\sum_{t\in\mathbb{Z}^d}\rho_\sigma^d(x+tq)}{\sum_{t\in\mathbb{Z}^d}\rho_\sigma^d(t)}, \forall x \in \mathbb{Z}_q^d$.

Recall that the *discrete Fourier transform* (DFT) of a function $f : \mathbb{Z}_q^d \to \mathbb{C}$ is given by $\widehat{f}(y) = \sum_{x\in\mathbb{Z}_q^d} e^{-\frac{2\pi i\langle x,y\rangle}{q}} f(x), \forall y \in \mathbb{Z}_q^d$. It is a powerful tool in cryptanalysis. For example, for any pmf ϕ over $\mathbb{Z}_q$, its *bias* is defined as $\mathbb{B}(\phi) = \mathbb{E}_{x\leftarrow\phi}\left[e^{-\frac{2\pi ix}{q}}\right] = \widehat{\phi}(1)$. It has been shown in [17] that DFT can be used to derive the *distinguish advantage* in a dual attack, and this idea was further developed in [18].

Lemma 1. *Given two positive integers d, q, let $\sigma > 0$. Suppose that $x \sim G_{\sigma,q}^d$ is a random vector. For any $v \in \mathbb{Z}_q^d$, we denote the pmf of (the random variable over $\mathbb{Z}_q$) $\langle v, x\rangle \pmod q$ by $f_{\langle v,x\rangle}$, then $\widehat{f_{\langle v,x\rangle}}(1) \geq e^{-\frac{2\pi^2\sigma^2\|v\|^2}{q^2}}$.*

Given the distinguish advantage, the well-known Chernoff-Hoeffding inequality is useful for estimating the number of samples required.

Lemma 2. *Let $\xi_1, \cdots, \xi_M$ be real-valued independent bounded random variables with $\xi_j \in [c,d]$ and $E[\xi_j] = \mu_j$, $j = 1, 2, \cdots, M$. Then for any $\epsilon \geq 0$,*

$$Pr\left[\left|\frac{1}{M}\sum_{j=1}^{M}\left(\xi_j - \mu_j\right)\right| \geq \epsilon\right] \leq 2 \cdot e^{-\frac{2M\epsilon^2}{(d-c)^2}}.$$

The following lemma is used to assess the independence (or correlation) of multiple samples, as this is a requirement of the Chernoff-Hoeffding inequality.

Lemma 3. *([19, Example 2.2]) Let X be a random vector in $\mathbb{R}^d$ whose covariance matrix is Σ. Then for any $y, z \in \mathbb{R}^d$,*

$$cov\left(\langle X, y\rangle, \langle X, z\rangle\right) = \langle y, \Sigma z\rangle.$$

It is known that for the X, y, z above, $\mathrm{var}(\langle X, y\rangle) = y^T\Sigma y$ and $\mathrm{var}(\langle X, z\rangle) = z^T\Sigma z$, so we derive the corollary below.

Corollary 1. *Let X be a random vector in $\mathbb{R}^d$ whose covariance matrix is Σ. Then for any $y, z \in \mathbb{R}^d$,*

$$r\left(\langle X, y\rangle, \langle X, z\rangle\right) = \frac{\langle y, \Sigma z\rangle}{\sqrt{y^T\Sigma y} \cdot \sqrt{z^T\Sigma z}},$$

where $r(\cdot,\cdot)$ denotes the correlation coefficient. In particular, if $\Sigma = \sigma^2 I_d$ for some $\sigma > 0$, then we have

$$r\left(\langle X, y\rangle, \langle X, z\rangle\right) = \frac{\langle y, z\rangle}{\|y\| \cdot \|z\|} = \cos\left(\theta(y, z)\right).$$

2.2. Lattice and BKZ

In this subsection, we shall provide some properties and algorithms in lattice cryptography that are useful in our later discussion.

Recall that a d-dimensional *lattice* Λ is a discrete additive subgroup of $\mathbb{R}^d$. The set $\{b_j\}_{j=1}^r \subseteq \mathbb{R}^d$ is a *basis* of Λ, if $b_1, \cdots, b_r$ are linearly independent and $\Lambda = \left\{\sum_{j=1}^r z_j b_j : z_j \in \mathbb{Z}\right\}$. Then r is called the *rank* of Λ. We write $B = (b_1\ b_2\ \cdots\ b_r)$, and the *volume* of Λ is defined as $\mathrm{vol}(\Lambda) = \sqrt{\det(B^T B)}$.

For a lattice Λ with basis B, its *dual lattice* is denoted by $\Lambda^* = \{y \in \mathrm{Span}(B) : \forall x \in \Lambda, \langle x, y\rangle \in \mathbb{Z}\}$. It is easy to verify that the *dual matrix* $B^{\sim T}$ of B is a basis of Λ^*, and thus $\mathrm{vol}(\Lambda^*) = \frac{1}{vol(\Lambda)}$.

The "*primitiveness*" of a set of lattice vectors is related to its ability to be extended to a lattice basis. To be specific, the column vectors of $\Phi = (\phi_1\ \phi_2\ \cdots\ \phi_t)$ are said to be *primitive* with respect to Λ, if they are linearly independent and $\Lambda \cap \mathrm{Span}(\Phi) = L(\Phi)$. The following lemma is a natural extension of lemma 12 in [20] and a proof is given in [21, appendix B]. It predicts the volume of the intersection of a lattice and some subspace, if certain primitiveness conditions are met.

Lemma 4. *Given a lattice Λ. Suppose that $\Phi = (\phi_1\ \cdots\ \phi_t)$ contains a set of primitive vectors of Λ^*, then $\Lambda \cap Span(\Phi)^\perp$ is a lattice of volume $\sqrt{\det\left(\Phi^T\Phi\right)} \cdot vol(\Lambda)$.*

The BKZ algorithm put forward by Schnorr and Euchner [22] in 1991 is a strong tool of the dual attack. Since it was proposed, there have been many variants of it, such as [23]. At present, the BKZ algorithm with sieving [24] as the SVP oracle is regarded as the most common and efficient choice. The "*blocksize*" $\mathfrak{b}$ in it is a core parameter. When applying the BKZ algorithm with blocksize $\mathfrak{b}$ (BKZ-$\mathfrak{b}$), the *Hermite factor* $\delta_0(\mathfrak{b}) \approx \left(\frac{\mathfrak{b}}{2\pi e}(\pi\mathfrak{b})^{\frac{1}{\mathfrak{b}}}\right)^{\frac{1}{2(\mathfrak{b}-1)}}$ [25] is useful for predicting the length of the output vectors. $\mathfrak{b}$ also determines the number of the output and the running time of the algorithm.

Assumption 1. *For a lattice Λ of rank r, given any of its basis as input, BKZ-$\mathfrak{b}$ with sieving as the SVP oracle provides $2^{0.2075\mathfrak{b}}$ short vectors in one run, whose norms are all close to $\delta_0^r(\mathfrak{b}) \cdot vol(\Lambda)^{\frac{1}{r}}$. This costs*

$$T_{BKZ}(\mathfrak{b}) = \begin{cases} 2^{0.292\mathfrak{b}} & classical\ case \\ 2^{0.265\mathfrak{b}} & quantum\ case \end{cases}.$$

Moreover, it is generally assumed that the short vectors found by BKZ are non-directional. To be more precisely, each of their coefficients independently obeys the same Gaussian distribution. This is also known as the *BKZ balance assumption*.

Assumption 2. *Let $v \in \mathbb{R}^d$ be a short vector found by BKZ, then each entry of v follows a Gaussian distribution with mean 0 and standard deviation $\frac{\|v\|}{\sqrt{d}}$.*

2.3. Rotations

Let n, q, k be three positive integers. In MLWE-based schemes, we mainly work on the rings $R = \mathbb{Z}[x]/(x^n + 1)$ and $R_q = \mathbb{Z}_q[x]/(x^n + 1)$. For any polynomial $a = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} \in R_q$, we define its *coefficient vector* as $\overline{a} = (a_0, a_1, \cdots, a_{n-1})^T \in \mathbb{Z}_q^n$. This could be easily extended to the case of a polynomial vector $\alpha = \begin{pmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_k\end{pmatrix} \in R_q^k$, whose *coefficient vector* is $\begin{pmatrix}\overline{\alpha_1}\\ \overline{\alpha_2}\\ \vdots\\ \overline{\alpha_k}\end{pmatrix} \in \mathbb{Z}_q^{n\cdot k}$.

The "*rotations*" are useful for simplifying the inner product in R_q^k. Fix a positive integer r, an r-rotation of a polynomial vector $\alpha \in R_q^k$ is given by

$$\alpha^{(r)} = \left(x^r \cdot \alpha_1(x^{-1}) \pmod{x^n+1}, \cdots, x^r \cdot \alpha_k(x^{-1}) \pmod{x^n+1}\right)^T \in R_q^k.$$

The following properties of rotations are also relevant to our discussion.

Lemma 5. *For any two polynomial vectors $\alpha, \beta \in R_q^k$, we have*

(i) $\alpha^T\beta = \sum_{j=0}^{n-1} \overline{\alpha}^T\overline{\beta^{(j)}}x^j$.

(ii) For any integer r, $\alpha^{(n+r)} = -\alpha^{(r)}$ and $\alpha^{(2n+r)} = \alpha^{(r)}$.

(iii) Fix $r \in \{0, 1, \cdots, n-1\}$ and $k = 1$. Let α_j and $\alpha_j^{(r)}$ be the j-th $(0 \le j \le n-1)$ degree coefficients of α and $\alpha^{(r)}$ respectively, then

$$\alpha_j^{(r)} = \begin{cases} \alpha_{r-j} & 0 \le j \le r \\ -\alpha_{n+r-j} & r+1 \le j \le n-1 \end{cases}.$$

We shall introduce some notations for the sake of the later analysis. For any polynomial $a \in R_q$, we write $\overline{\overline{a}}$ to be the square matrix consisting of the coefficient vectors of the 0-rotation to the $(n-1)$-rotation of a, i.e. $\overline{\overline{a}} = \left(\overline{a^{(0)}}\ \overline{a^{(1)}}\ \cdots\ \overline{a^{(n-1)}}\right) \in \mathbb{Z}_q^{n\times n}$. This concept also applies to polynomial vectors and even matrices. That is, for any polynomial matrix $A = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{t1} & \cdots & a_{tk} \end{pmatrix} \in R_q^{t\times k}$, we define

$$\overline{\overline{A}} = \begin{pmatrix} \overline{\overline{a_{11}}} & \cdots & \overline{\overline{a_{1k}}} \\ \vdots & \ddots & \vdots \\ \overline{\overline{a_{t1}}} & \cdots & \overline{\overline{a_{tk}}} \end{pmatrix} \in \mathbb{Z}_q^{tn\times kn}, \tag{1}$$

and call $\overline{\overline{A}}$ the *H-rotation* of A, as it contains the rotations of A within half cycle.

2.4. Module-LWE

We focus on the case of Module-LWE (including Ring-LWE) in this paper. They are structured variants of LWE that were introduced to achieve higher efficiency.

Definition 1. *Fix positive integers n, q, k and the secret $s \in R_q^k$. Let χ be a distribution over $\mathbb{Z}_q$ with a mean of 0 and a small standard deviation of σ_χ. Then an Module-LWE sample with parameters (n, q, k, χ) is of the form*

$$(a, b) \in R_q^k \times R_q \quad with \quad b = \langle a, s\rangle + e,$$

where $a \leftarrow U\left(R_q^k\right)$ and each coefficient of the error $e \in R_q$ is sampled independently according to χ, i.e., $\overline{e} \leftarrow \chi^n$. Further, the Search-MLWE is to find s when given a limited number of samples, while the Decision-MLWE asks for distinguishing these samples between uniform ones.

Remark 1: When $k = 1$, the problem is reduced to Ring-LWE.

Reduction to LWE Although MLWE features more algebraic structure, no more efficient algorithm has been proposed based on this. The security of actual MLWE-based schemes are usually assessed by first transforming the MLWE instance to an LWE one and then attacking it by methods against LWE.

The following describes a transformation way for an MLWE instance $(A, b = As + e) \in R_q^{t\times k} \times R_q^t$ consisting of t MLWE samples. Some structure is preserved in the sample matrix of the resulting LWE instance. Specifically, let $b = \begin{pmatrix} b_1 \\ \vdots \\ b_t \end{pmatrix}$ and $e = \begin{pmatrix} e_1 \\ \vdots \\ e_t \end{pmatrix}$. We denote the i-th row of $A = (a_{ij})$ by A_i^T. Then from lemma 5 (i), we have

$$b_i = A_i^T s + e_i = \sum_{j=0}^{n-1} \overline{A_i^{(j)}}^T \overline{s}\, x^j + e_i,\ i = 1, 2, \cdots, t. \tag{2}$$

Writing $b_i = \sum_{j=0}^{n-1} b_{ij}x^j, e_i = \sum_{j=0}^{n-1} e_{ij}x^j$, and comparing the j-th $(0 \le j \le n-1)$ degree coefficient of the two sides of equation (2), we have

$$b_{ij} = \overline{A_i^{(j)}}^T \overline{s} + e_{ij},\ i = 1, 2, \cdots, t;\ j = 0, 1, \cdots, n-1. \tag{3}$$

These make up $t \cdot n$ LWE samples that can be reformulated as:

$$\begin{pmatrix} \overline{b_1} \\ \vdots \\ \overline{b_t} \end{pmatrix} = \begin{pmatrix} \overline{A_1^{(0)}}^T \\ \vdots \\ \overline{A_1^{(n-1)}}^T \\ \vdots \\ \overline{A_t^{(1)}}^T \\ \vdots \\ \overline{A_t^{(n-1)}}^T \end{pmatrix} \cdot \overline{s} + \begin{pmatrix} \overline{e_1} \\ \vdots \\ \overline{e_t} \end{pmatrix} \pmod q, \tag{4}$$

or further summarized as (a proof is given in Appendix A):

$$\overline{b} = \overline{\overline{A^T}}^T \cdot \overline{s} + \overline{e} \pmod q. \tag{5}$$

Recall that $\overline{\overline{A^T}}$ is the H-rotation of A^T, and $\overline{b}, \overline{s}, \overline{e}$ are the coefficient vectors of b, s, e respectively.

From the above, $t \cdot n$ LWE samples can be converted from t MLWE samples, and the new $(k \cdot n)$-dimensional secret $\overline{s}$ is the coefficient vector of the original secret s. Some structure has been preserved in the sample matrix $\overline{\overline{A^T}}^T$. More specifically, it is easy to see that for $i = 1, 2, \cdots, t$, the $(n \cdot (i-1) + 1)$-th through the $(n \cdot (i-1) + n)$-th rows of $\overline{\overline{A^T}}^T$ respectively correspond to the coefficient vectors of different rotations of the same polynomial vector A_i. This property is exactly what we will use later to enhance the dual attack against MLWE.

Remark 2: It should be noticed that $\overline{\overline{A^T}}^T \neq \overline{\overline{A}}$ although they are both $tn \times kn$ matrices. Actually, it is easy to verify that $\overline{\overline{A^T}}^T = \begin{pmatrix} \overline{\overline{a_{11}}}^T & \cdots & \overline{\overline{a_{1k}}}^T \\ \vdots & \ddots & \vdots \\ \overline{\overline{a_{t1}}}^T & \cdots & \overline{\overline{a_{tk}}}^T \end{pmatrix}$, and the reason why it is not equal to $\overline{\overline{A}}$ is because $\left\{\overline{\overline{a_{ij}}}\right\}$ and A are not symmetric.

Remark 3: For actual schemes such as MLWE-based Kyber and RLWE-based Newhope, there are only $k+1$ MLWE samples available to the attacker, i.e., $t \le k+1$. So the number of LWE samples used in a dual attack should be limited to $(k+1) \cdot n$.

Remark 4: In the following, we focus on the case where the coefficients of the secret s also obey the error distribution

χ independently, i.e., $\overline{s} \leftarrow \chi^{k\cdot n}$. Then, the resulting LWE instance is referred to as an instance in Hermite Normal Form (HNF). This manner of selecting s is often used in practice, for example, Kyber (with certain parameter sets)[7], Newhope [5], FrodoKEM[26] and LAC [6] all use the same distribution to sample the entries of s and the e. Moreover, a way was given in [27] to transform the distribution of the secret to be that of the error through Gaussian elimination.

3. Dual Attack against MLWE

This section discusses the dual attack against MLWE. The special structure mentioned earlier of the resulting LWE sample matrix gives the possibility of making an enhancement to the dual attack on it. Actually, as we shall see, each short vector found by BKZ contributes much more. To be specific, it is described in Section 3.1 that, using a short vector found in $\mathscr{L}'$ (see equation (8)), the attacker can construct another $n-1$ vectors in $\mathscr{L}$ (see equation (6)) of the same length automatically. All of these short vectors can be used to distinguish. Our enhanced attack method utilizes this interesting property, and is given in Section 3.2. Its complexity and correlation analysis are also provided.

3.1. Constructing more short vectors in $\mathscr{L}$

In this subsection, we describe a method of constructing more short vectors from those returned by the BKZ algorithm.

Suppose that the goal of the adversary is to distinguish whether a given instance $(A, b) \in R_q^{t\times k} \times R_q^t$ is from MLWE or not. After performing the transformation described in Section 2.4, he/she gets an instance $\left(\overline{\overline{A^T}}^T, \overline{b}\right) \in \mathbb{Z}_q^{tn\times kn} \times \mathbb{Z}_q^{tn}$.

There is a noteworthy phenomenon in the dual attack that the optimal performance may not be achieved when all of the obtained samples are used. The reason is that using more samples makes the dimension of the lattice higher. This means that we may not have to use all of the samples in $\left(\overline{\overline{A^T}}^T, \overline{b}\right)$. Let $m = hn + g, 0 < g \le n$. Without loss of generality, we assume that the first $n \cdot h$ LWE samples come from the first h MLWE samples, and the remaining g LWE samples are derived from the $(h+1)$-th MLWE sample.

In the following, let us start with a simple case: assume that m is a multiple of n, i.e., $m = (h+1)\cdot n, 0 \le h < t$. Then, the LWE sample matrix used in the dual attack is of the form

$$\left(\overline{A_1^{(0)}} \cdots \overline{A_1^{(n-1)}} \cdots \overline{A_{h+1}^{(0)}} \cdots \overline{A_{h+1}^{(n-1)}}\right)^T = \left(\overline{\overline{A_1}} \cdots \overline{\overline{A_{h+1}}}\right)^T = \overline{\overline{A_{[1:h+1]}^T}}^T.$$

To perform a dual attack on the instance $(\overline{\overline{A_{[1:h+1]}^T}}^T, \overline{b_{[1:h+1]}})$, the adversary considers the lattice

$$\mathscr{L} = \left\{\begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{Z}^{(h+1)n+kn} : \overline{\overline{A_{[1:h+1]}^T}} \cdot u = v \pmod q\right\}. \quad (6)$$

It is an $((h+1)n+kn)$-dimensional lattice of volume q^{kn} and has a basis

$$\mathscr{B} = \begin{pmatrix} I_{(h+1)n} & O_{(h+1)n\times kn} \\ \overline{\overline{A_{[1:h+1]}^T}} & q_{kn} \end{pmatrix} \in \mathbb{Z}^{((h+1)n+kn)\times((h+1)n+kn)}.$$

Short vectors will be searched in $\mathscr{L}$ by BKZ to distinguish whether the instance $(\overline{\overline{A_{[1:h+1]}^T}}^T, \overline{b_{[1:h+1]}})$ comes from LWE or is uniform. The difference here is that, with any short vector found in $\mathscr{L}$, the adversary could construct $n-1$ additional short vectors in $\mathscr{L}$ by himself/herself (instead of using BKZ). This means that if M short vectors in $\mathscr{L}$ are obtained by BKZ, they will be expanded to $n \cdot M$ short vectors in $\mathscr{L}$, and then all of these short vectors will be used for distinguishing. Therefore, fewer short vectors are required during the BKZ search phase, and the cost is thus reduced.

Before showing this interesting fact, we shall give a useful proposition which indicates that the rotations are consistent with linear transformations whose matrices are H-rotation matrices.

Proposition 1. *Given positive integers z, k. For any polynomial matrix $P \in R_q^{k\times z}$ and polynomial vectors $\alpha \in R_q^z$, $\beta \in R_q^k$, if $\overline{\overline{P}} \cdot \overline{\alpha^{(r)}} = \overline{\beta^{(r)}}$, then $\overline{\overline{P}} \cdot \overline{\alpha^{(r+1)}} = \overline{\beta^{(r+1)}}$, $r \in \{0, 1, \cdots, n-2\}$.*

The proof of proposition 1 is given in Appendix B. From it we know that, for any short vector $\begin{pmatrix} u \\ v \end{pmatrix} \in \mathscr{L}$, it is easy to find polynomial vectors $\alpha \in R_q^{h+1}$ and $\beta \in R_q^k$ [1], such that $\overline{\alpha^{(0)}} = u$ and $\overline{\beta^{(0)}} = v$. Then $\overline{\overline{A_{[1:h+1]}^T}} \cdot \overline{\alpha^{(0)}} = \overline{\beta^{(0)}}$. By applying proposition 1 multiple times for $r = 0, 1, \cdots, n-2$, we have $\overline{\overline{A_{[1:h+1]}^T}} \cdot \overline{\alpha^{(j)}} = \overline{\beta^{(j)}}$, $j = 1, 2, \cdots, n-1$. In other words, the attacker could use $\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \overline{\alpha^{(0)}} \\ \overline{\beta^{(0)}} \end{pmatrix}$ to immediately construct another $n-1$ different short vectors (of the same length) in $\mathscr{L}$:

$$\left(\overline{\begin{matrix}\alpha^{(1)}\\ \beta^{(1)}\end{matrix}}\right), \left(\overline{\begin{matrix}\alpha^{(2)}\\ \beta^{(2)}\end{matrix}}\right), \cdots, \left(\overline{\begin{matrix}\alpha^{(n-1)}\\ \beta^{(n-1)}\end{matrix}}\right).$$

To make it more clear, we give an example below.

Example 1: Assume that $n = k = t = h+1 = 2$. Let $A = (a_{ij})_{2\times 2}$, where $a_{ij} = \sum_{k=0}^{1} a_{ijk}x^k$. Let $A_1 = \begin{pmatrix} a_{11} \\ a_{12} \end{pmatrix}$ and $A_2 = \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix}$, then

$$\overline{\overline{A_{[1:h+1]}^T}} = \overline{\overline{A^T}} = \begin{pmatrix} \overline{\overline{a_{11}}} & \overline{\overline{a_{21}}} \\ \overline{\overline{a_{12}}} & \overline{\overline{a_{22}}} \end{pmatrix} = \begin{pmatrix} \overline{a_{11}^{(0)}} & \overline{a_{11}^{(1)}} & \overline{a_{21}^{(0)}} & \overline{a_{21}^{(1)}} \\ \overline{a_{12}^{(0)}} & \overline{a_{12}^{(1)}} & \overline{a_{22}^{(0)}} & \overline{a_{22}^{(1)}} \end{pmatrix} = \begin{pmatrix} a_{110} & a_{111} & a_{210} & a_{211} \\ -a_{111} & a_{110} & -a_{211} & a_{210} \\ a_{120} & a_{121} & a_{220} & a_{221} \\ -a_{121} & a_{120} & -a_{221} & a_{220} \end{pmatrix}.$$

For any short vector $\begin{pmatrix} u \\ v \end{pmatrix} \in \mathscr{L}$, we denote $u = \begin{pmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} \overline{\alpha_1^{(0)}} \\ \overline{\alpha_2^{(0)}} \end{pmatrix}$ and $v = \begin{pmatrix} v_0 \\ v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} \overline{\beta_1^{(0)}} \\ \overline{\beta_2^{(0)}} \end{pmatrix}$, where $\alpha_1, \alpha_2, \beta_1, \beta_2$ are polynomials in R_q. As

[1] As u, v are both short vectors in the q-ary lattice $\mathscr{L}$, it is natural to think that their coefficients all belong to $\mathbb{Z}_q$.

$\overline{\overline{A^T}} \cdot u = v \pmod q$, *we have,*

$$\begin{cases} a_{110}u_0 + a_{111}u_1 + a_{210}u_2 + a_{211}u_3 = v_0 \\ -a_{111}u_0 + a_{110}u_1 - a_{211}u_2 + a_{210}u_3 = v_1 \\ a_{120}u_0 + a_{121}u_1 + a_{220}u_2 + a_{221}u_3 = v_2 \\ -a_{121}u_0 + a_{120}u_1 - a_{221}u_2 + a_{220}u_3 = v_3 \end{cases}.$$

This could be reformulated as $\overline{\overline{A^T}}\begin{pmatrix} u_1 \\ -u_0 \\ u_3 \\ -u_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ -v_0 \\ v_3 \\ -v_2 \end{pmatrix}$, *i.e.,* $\begin{pmatrix} \overline{\alpha_1^{(1)}} \\ \overline{\alpha_2^{(1)}} \\ \overline{\beta_1^{(1)}} \\ \overline{\beta_2^{(1)}} \end{pmatrix}$ *is also a vector belong to* $\mathscr{L}$ *and it is as short as* $\begin{pmatrix} u \\ v \end{pmatrix}$. □

Now we are ready for the discussion on the case where m is not a multiple of n. This is necessary as in a dual attack against actual LWE schemes, the lowest cost is usually obtained when the number of used samples is not a multiple of n. Let $m = hn + g$, where $0 \le h < t, 0 < g < n$. A naive idea is to construct an analogue lattice to $\mathscr{L}$ of dimension $(hn + g + kn)$ that considers the sample matrix

$$\left(\overline{A_1^{(0)}} \cdots \overline{A_1^{(n-1)}} \cdots \overline{A_h^{(0)}} \cdots \overline{A_h^{(n-1)} A_{h+1}^{(0)}} \cdots \overline{A_{h+1}^{(g-1)}}\right)^T. \tag{7}$$

However, since the last g columns do not form the whole $\overline{\overline{A_{h+1}}}$, proposition 1 is not applicable. From another perspective, because the last g entries of u cannot constitute a complete coefficient vector, rotations can not be applied on it to produce more short vectors.

To deal with this problem, we come up with a new kind of lattice where only m samples are used implicitly, and meanwhile the missing entries of u are successfully padded. Specifically, define the subspace $\mathscr{V} = \{w \in \mathbb{R}^{(h+1)n+kn} : \langle w, \gamma_j \rangle = 0,\ j = hn + g + 1, \cdots, (h+1)n\}$, where γ_j represents the unit vector in $\mathbb{Z}^{(h+1)n+kn}$ with only the j-th coefficient being 1. Then the attacker searches short vectors in

$$\mathscr{L}' = \mathscr{L} \cap \mathscr{V}. \tag{8}$$

Although $\mathscr{L}' \subseteq \mathbb{Z}^{(h+1)n+kn}$, it is easy to see that $\mathrm{rank}(\mathscr{L}') = m + kn$. Properly speaking, for a vector $w = \begin{pmatrix} u \\ v \end{pmatrix} \in \mathscr{L}'$, the $(hn + g + 1)$-th through the $(h + 1)n$-th entries of u are all 0, making the corresponding columns $(hn + g + 1)$ to $(h + 1)n$ of $\overline{\overline{A_{[1:h+1]}^T}}$ (i.e. $\overline{\overline{A_{h+1}^{(g)}}}, \cdots, \overline{\overline{A_{h+1}^{(n-1)}}}$) lose their effects. This implies only m samples mentioned in equation (7) are used. According to [20], a basis $\mathscr{B}'$ of $\mathscr{L}'$ could be computed by the following way:

1. Calculate $D = \Pi_{\mathscr{V}} \cdot \mathscr{B}^{-T}$, where $\Pi_{\mathscr{V}}$ is the orthogonal projection matrix onto $\mathscr{V}$.
2. Apply the MLLL algorithm on D to eliminate linear dependencies. Delete $n - g$ zero vectors in the result and denote the remaining vectors by $D_{\mathscr{V}}$.
3. Output the dual matrix of $D_{\mathscr{V}}$ as $\mathscr{B}'$.

We need to take care of the volume of $\mathscr{L}'$, as it is closely related to the cost of the dual attack (see Section 3.2 for details). Actually, $\mathrm{vol}(\mathscr{L}')$ can be computed even without calculating $\mathscr{B}'$. Since $\mathscr{B}$ is a basis of $\mathscr{L}$, $\mathscr{B}^{-T} = \begin{pmatrix} I_{(h+1)n} & -\frac{\overline{\overline{A_{[1:h+1]}^T}}^T}{q} \\ O_{kn\times(h+1)n} & \frac{1}{q} I_{kn} \end{pmatrix}$ is a basis of $\mathscr{L}^*$. It can be seen that $\left\{\gamma_{hn+g+1}, \cdots, \gamma_{(h+1)n}\right\}$ forms a set of primitive vectors with respect to $\mathscr{L}^*$, as they can be extended to $\mathscr{B}^{-T}$. Then from lemma 4,

$$\mathrm{vol}(\mathscr{L}') = \mathrm{vol}(\mathscr{L}) = q^{kn}.$$

In conclusion, we reduce the rank of $\mathscr{L}'$ but without changing its volume [2]. For any vector $w = \begin{pmatrix} u \\ v \end{pmatrix} \in \mathscr{L}'$, just like before, we could find a polynomial vector $\eta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \in R_q^{(h+1)+k}$, such that $\overline{\alpha^{(0)}} = u, \overline{\beta^{(0)}} = v$. Then $u^{(1)}, \cdots, u^{(n-1)}$ all have at least $n - g$ zero coefficients (but may not necessarily be in the last $n - g$ positions). As a result, all of $\overline{\eta^{(1)}}, \cdots, \overline{\eta^{(n-1)}}$ are short vectors of length $\|w\|$ in $\mathscr{L}$. We should note that these vectors may not belong to $\mathscr{L}'$, but this has no effect on our dual attack as we shall see later.

The above analysis is summarized as the following theorem.

Theorem 1. *For any short vector* $\begin{pmatrix} u \\ v \end{pmatrix} \in \mathscr{L}$ *(or* $\mathscr{L}'$*), let* $\eta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \in R_q^{(h+1)+k}$ *be the polynomial vector that satisfy* $\overline{\alpha^{(0)}} = u$ *and* $\overline{\beta^{(0)}} = v$, *then* $\overline{\eta^{(j)}} \in \mathscr{L},\ j = 0, 1, \cdots, n-1$.

Remark 5: It is not hard to see that $\mathscr{L}$ is the lattice corresponding to $\mathscr{L}'$ when $g = n$. So for simplicity, we just say that short vectors are found in $\mathscr{L}'$ by BKZ in both cases.

3.2. The enhanced algorithm as well as its correlation and complexity analysis

In this subsection, we shall give a detailed description of our enhanced attack. Its correlation and complexity analysis are also provided.

To distinguish the target instance $(A, b) \in R_q^{t\times k} \times R_q^t$ of MLWE from uniform, the adversary constructs the lattice $\mathscr{L}'$ as in equation (8) and looks for short vectors in it.

Suppose that M short vectors $\begin{pmatrix} u_j \\ v_j \end{pmatrix} \in \mathscr{L}', j = 1, 2, \cdots, M$ of length at most l are found. As described in the previous subsection, let $\eta_j = \begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \in R_q^{(h+1)+k}$ be the polynomial vectors

[2] As we know, for the original dual attack against LWE, the volume of the lattice depends only on the dimension of the secret, not on the number of samples used. So q^{kn} is exactly what we expect. It should be noted that the volume of the intersection lattice is usually larger than that of the original lattice. This is disadvantageous to an adversary who performs a dual attack, because it is easier to find (unspecified) short vectors in a lattice of a smaller volume. Fortunately, the new lattice $\mathscr{L}'$ we constructed has the same volume as the original lattice $\mathscr{L}$, because certain primitiveness requirements are met.

that satisfy $\overline{\alpha_j^{(0)}} = u_j$ and $\overline{\beta_j^{(0)}} = v_j, j = 1, 2, \cdots, M$. Then according to theorem 1, the attacker obtains $n \cdot M$ short vectors $\left\{\overline{\eta_j^{(r)}}\right\}_{j,r}$ in $\mathscr{L}$.

All of these short vectors will be used for distinguishing. To be specific, the difference in the distributions of $\left\{\left\langle\overline{\alpha_j^{(r)}}, \overline{b_{[1:h+1]}}\right\rangle \pmod q\right\}_{j,r}$ in the two cases is key to the distinction. If b is uniform, then so is $\overline{b_{[1:h+1]}}$, and thus $\left\langle\overline{\alpha_j^{(r)}}, \overline{b_{[1:h+1]}}\right\rangle \pmod q \leftarrow U\left(\mathbb{Z}_q\right)$. While if $b = As + e$, we have $\overline{b_{[1:h+1]}} = \overline{\overline{A_{[1:h+1]}^T}} \cdot \overline{s} + \overline{e_{[1:h+1]}} \pmod q$. Let $\overline{S} = \begin{pmatrix}\overline{e_{[1:h+1]}} \\ \overline{s}\end{pmatrix}$, then $\left\langle\overline{\alpha_j^{(r)}}, \overline{b_{[1:h+1]}}\right\rangle = \left\langle\begin{pmatrix}\overline{\alpha_j^{(r)}} \\ \overline{\beta_j^{(r)}}\end{pmatrix}, \begin{pmatrix}\overline{e_{[1:h+1]}} \\ \overline{s}\end{pmatrix}\right\rangle = \left\langle\overline{\eta_j^{(r)}}, \overline{S}\right\rangle \pmod q$ is relatively small, as $\overline{\eta_j^{(r)}}, \overline{S}$ are both short vectors.

In consequence, the attacker calculates $\frac{\sum_{j=1}^{M}\sum_{r=0}^{n-1} e^{-\frac{2\pi i\left\langle\overline{\alpha_j^{(r)}}, \overline{b_{[1:h+1]}}\right\rangle}{q}}}{nM}$ and it goes to $\frac{\sum_{j=1}^{M}\sum_{r=0}^{n-1} \widehat{f}^{(1)}_{\left\langle\overline{\alpha_j^{(r)}}, \overline{b_{[1:h+1]}}\right\rangle}}{nM}$ as M increase, where $f_{\left\langle\overline{\alpha_j^{(r)}}, \overline{b_{[1:h+1]}}\right\rangle}$ refers to the pmf of $\left\langle\overline{\alpha_j^{(r)}}, \overline{b_{[1:h+1]}}\right\rangle \pmod q$. The result is 0 when b is uniform, and is

$$\frac{\sum_{r=0}^{n-1}\sum_{j=1}^{M} \widehat{f}^{(1)}_{\left\langle\overline{S}, \overline{\eta_j^{(r)}}\right\rangle}}{nM} \geq e^{-\frac{2\pi^2\sigma_\chi^2\left\|\overline{\eta_j^{(r)}}\right\|^2}{q^2}} \geq e^{-\frac{2\pi^2\sigma_\chi^2 l^2}{q^2}} := \epsilon$$

in the other case from lemma 1. The distinction is made according to whether the calculated value is closer to 0 or ϵ. To achieve a constant success rate, by lemma 2, $nM = \frac{1}{O(\epsilon^2)}$ short vectors are sufficient.

Remark 6: It is noted that lemma 1 is true for any $v \in \mathbb{Z}_q^d$. Hence, the fact that there are some zero coefficients in $\overline{\eta_j^{(r)}}$ does not have any effect on the above analysis.

Now we shall derive the cost model of our improved dual attack method against MLWE, as well as the selection approach of each parameter. From assumption 1, when BKZ-$\mathfrak{b}$ is applied and m LWE samples are used, $2^{0.2075\mathfrak{b}}$ short vectors in $\mathscr{L}'$ of length $l(m, \mathfrak{b}) = \delta_0^{m+kn}(\mathfrak{b}) \cdot q^{\frac{kn}{m+kn}}$ will be found. By the way described in Section 3.1, they can be extended to $n \cdot 2^{0.2075\mathfrak{b}}$ short vectors in $\mathscr{L}$ (of length $l(m, \mathfrak{b})$), each giving a distinguish advantage of $\epsilon(m, \mathfrak{b}) = e^{-\frac{2\pi^2\sigma_\chi^2 l^2(m,\mathfrak{b})}{q^2}}$.

To achieve a constant success rate, $O\left(\frac{1}{\epsilon^2(m,\mathfrak{b})}\right)$ short vectors are needed. So the BKZ search process has to be repeated at least $R(m, \mathfrak{b}) = \max\left\{1, \frac{1}{n \cdot 2^{0.2075\mathfrak{b}} \cdot \epsilon^2(m,\mathfrak{b})}\right\}$ times. Then the time complexity is

$$T(m, \mathfrak{b}) = T_{BKZ}(\mathfrak{b}) \cdot R(m, \mathfrak{b}).$$

Hence, the attacker should figure out the optimal blocksize $\mathfrak{b}^*$ and the optimal number of samples m^*, so that $T(m^*, \mathfrak{b}^*)$ is minimized (i.e. $T(m^*, \mathfrak{b}^*) = \min_{m,\mathfrak{b}}\{T(m, \mathfrak{b})\}$). As we know, for the original dual attack against LWE, m^* is actually a function of $\mathfrak{b}$ [28, 29]. That is, $m^* = \sqrt{\frac{n \ln q}{\ln(\delta_0(\mathfrak{b}))}} - n$. We notice that an analogue of this relationship is still available in the case of MLWE:

$$m^*(b) = \sqrt{\frac{kn \cdot \ln q}{\ln(\delta_0(\mathfrak{b}))}} - kn, \tag{9}$$

as it is the only zero of the derivative of $l(m, \mathfrak{b})$ with respect to m.

To sum up, only the optimal blocksize $\mathfrak{b}^*$ needs to be searched to meet

$$\mathfrak{b}^* = \text{argmin}_{\mathfrak{b}}\left\{T\left(m^*(\mathfrak{b}), \mathfrak{b}\right)\right\},$$

and then the optimal number of samples is $m^*(\mathfrak{b}^*)$.

In the last part of this subsection, we further show the correctness of the above algorithm. As mentioned earlier, the analysis of the dual attack relies on the Chernoff-Hoeffding inequality, which requires the independence between samples. We notice that $\overline{\eta_j^{(0)}}, \overline{\eta_j^{(1)}}, \cdots, \overline{\eta_j^{(n-1)}}$ are coefficient vectors of different rotations of the same polynomial vector, their correlations should be examined for using Chernoff-Hoeffding inequality.

It is a widely used analytic approach for the dual attack against LWE to suppose the so-called near-orthogonality assumption of high-dimensional spaces and the balance assumption of BKZ. These two assumptions will be used in our analysis of MLWE as well. In our case, we only need to address the issue of near-orthogonality.

Properly speaking, what we care about is the correlation between $\left\langle\overline{S}, \overline{\eta_{j_1}^{(r_1)}}\right\rangle$ and $\left\langle\overline{S}, \overline{\eta_{j_2}^{(r_2)}}\right\rangle$. According to corollary 1, their correlation coefficient is $\cos\left(\theta\left(\overline{\eta_{j_1}^{(r_1)}}, \overline{\eta_{j_2}^{(r_2)}}\right)\right)$. When $j_1 \neq j_2$, by the BKZ balance assumption, $\overline{\eta_{j_1}^{(r_1)}}$ and $\overline{\eta_{j_2}^{(r_2)}}$ could be regarded as random vectors whose directions are uniform. Then from the near-orthogonality assumption, $\overline{\eta_{j_1}^{(r_1)}}$ and $\overline{\eta_{j_2}^{(r_2)}}$ are close to be perpendicular and hence the cosine of the angle between them is close to 0. While for the case of $j_1 = j_2$ and $r_1 \neq r_2$, owing to the independence of the coefficients of η_{j_1} and the high dimension, $\cos\left(\theta\left(\overline{\eta_{j_1}^{(r_1)}}, \overline{\eta_{j_2}^{(r_2)}}\right)\right)$ will still have a very small absolute value, meaning that they are less correlated as in the case for independent $\overline{\eta_{j_1}^{(r_1)}}$ and $\overline{\eta_{j_2}^{(r_2)}}$ ($j_1 \neq j_2$). This interesting observation is supported by extended experiments in Section 4.1. The above analysis can be summarized as the following assumption.

Assumption 3. *For $1 \leq j \leq M$, if $r_1, r_2 \in \{0, 1, \cdots, n-1\}$ are chosen randomly, then $\cos\left(\theta\left(\overline{\eta_j^{(r_1)}}, \overline{\eta_j^{(r_2)}}\right)\right) \approx 0$ with an overwhelming probability.*

4. Experiments

We notice that our enhanced dual attack method is more suitable for RLWE-based schemes with small parameter sets. On the one hand, for an MLWE-based cryptosystem, since the coefficient vector of the secret is of dimension $k \cdot n$, at the same security level, $k = 1$ obviously results in a larger n. This means that the attacker is able to construct more short vectors for distinguishing from theorem 1. On the other hand, in a larger parameter set, although n is larger, however, so is the blocksize $\mathfrak{b}$.

	original dual attack	new algorithm in this paper
m^*	569	570
$\mathfrak{b}^*$	382.67	376.86
T_c^*	$2^{111.740}$	$2^{110.043}$
T_q^*	$2^{101.408}$	$2^{99.868}$

Table 1: A comparison between the original dual attack and our improved algorithm when attacking a Newhope512 instance.

As mentioned earlier, $2^{0.2075\mathfrak{b}}$ short vectors will be provided by BKZ in one run. In fact, the increase in n is small compared to that in $2^{0.2075\mathfrak{b}}$, making the improvement effect relatively less obvious. In the following experiments, Newhope512 is taken as an example.

4.1. Newhope512

The RLWE-based Newhope [5] is a second round candidate in NIST's post-quantum standardization effort. In Newhope512, $n = 512, k = 1, q = 12289$. Each coefficient of the secret and the error is independently picked up from the centered binomial distribution ψ_8 [3] with a standard deviation of 2. As done in [5], we assume that this does not lead to much difference in the conclusions we deduce earlier on the discrete Gaussian distribution.

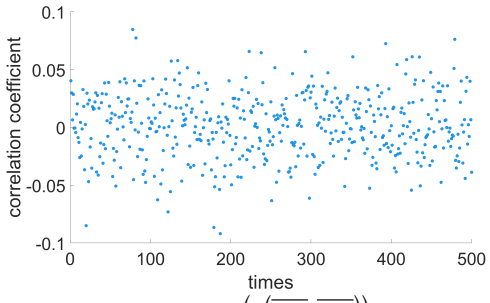
Table 1 compares the original dual attack against Newhope512 with our improved one. The optimal blocksize and the optimal number of samples are denoted by $\mathfrak{b}^*$ and m^* respectively, and then the time complexity is T_c^* (T_q^*) in the classical (quantum) case. Although the blocksize $\mathfrak{b}$ needs to be selected as an integer in practice, setting its precision to 0.01 helps us to give a more accurate prediction.

It can been seen that, by taking advantage of the algebraic structure in Newhope512, we reduce the blocksize $\mathfrak{b}^*$ by 5.81. The time complexity is decreased by a factor of $2^{1.697}$ in the classical case and $2^{1.540}$ in the quantum case. The optimal number of samples is 570, which is very close to the predicted value of 567.16 in equation (9). When taking $\mathfrak{b} = 376.86$, the BKZ algorithm finds $2^{0.2075*376.86} \approx 2^{78}$ short vectors in one run, the results from the previous section imply that the attacker can actually have $512 \cdot 2^{78} = 2^{87}$ short vectors available for distinguishing. We see that the improvement is not that significant in the dual attack against Newhope512 (as well as other MLWE/RLWE schemes), this is due to the fact that $2^{0.2075\mathfrak{b}^*} \gg n$.

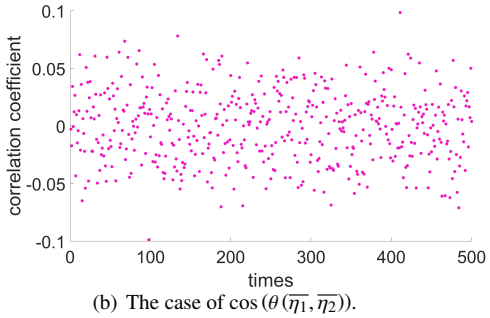
We also conduct some experiments to show the rationality of assumption 3. If $m^* = 549$ LWE samples are used, the lattice is of rank $549 + 512 = 1061$ and dimension $1024 + 512 = 1536$. The length of the short vectors found by BKZ is $l(549, 376.86) \approx 7612.185$, hence each of their non-zero coefficients follows a Gaussian distribution with a mean of 0 and a standard deviation of $\frac{7612.185}{\sqrt{1061}} \approx 233.696$.

The steps below are repeated multiple times:

1. Take $\overline{\eta_1} \leftarrow N_{233.696}^{1061}$ and define $\overline{\eta} = \begin{pmatrix} \overline{\eta_1} \\ 0_{475} \end{pmatrix}$.
2. Pick up $r_1, r_2 \leftarrow U(\{0, 1, \cdots, n-1\})$ independently. If $r_1 \neq r_2$, then goto step 3, otherwise go back to step 2.
3. Compute $\cos\left(\theta\left(\overline{\eta^{(r_1)}}, \overline{\eta^{(r_2)}}\right)\right)$.



(a) The case of $\cos\left(\theta\left(\overline{\eta^{(r_1)}}, \overline{\eta^{(r_2)}}\right)\right)$.



(b) The case of $\cos(\theta(\overline{\eta_1}, \overline{\eta_2}))$.

Figure 1: The dispersion of the cosine values of the angles in the two cases.

Figure 1(a) shows the dispersion of the results of 500-times experiments. We can observe that $\left|\cos\left(\theta\left(\overline{\eta^{(r_1)}}, \overline{\eta^{(r_2)}}\right)\right)\right|$ is always within 0.1. The independence is implied, as it is generally recognized that a coefficient with an absolute value within 0.1 indicates a negligible relationship. For comparison, we also provide figure 1(b). The case of $\cos(\theta(\overline{\eta_1}, \overline{\eta_2}))$ is characterized, where $\overline{\eta_1}, \overline{\eta_2} \leftarrow N_{233.696}^{1061}$ are independent. It should be pointed out that the two situations in figure 1(a) and (b) is quite similar, so it is reasonable to consider the coefficient vectors of different rotations of the same polynomial vector to be independent and nearly-orthogonal.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No.12271306), and by the National Key Research and Development Program of China (No.2018YFA0704702).

[3] It is known that the ideal discrete Gaussian sampler is not available in practical LWE schemes, not only due to the limitation of precisions, but also because of the high cost. Recall that the central binomial distribution ψ_μ is a common substitute defined on the set $X = \{-\mu, -\mu+1, \cdots, \mu-1, \mu\}$ with the probability assignment at $k \in X$ to be $\psi_\mu(k) = \binom{2\mu}{k+\mu}\frac{1}{2^{2\mu}}$. Moreover, it has a variance $\frac{\mu}{2}$.

Appendix A. The sample matrix of the LWE instance

We just have to prove that $\overline{\overline{A^T}} = \left(\overline{A_1^{(0)}} \cdots \overline{A_1^{(n-1)}} \cdots \overline{A_t^{(0)}} \cdots \overline{A_t^{(n-1)}}\right)$.

Let $A_i = \begin{pmatrix} a_{i1} \\ \vdots \\ a_{ik} \end{pmatrix} \in R_q^k,\ i = 1, 2, \cdots, t$, then $\overline{\overline{A^T}} = \begin{pmatrix} \overline{\overline{a_{11}}} & \cdots & \overline{\overline{a_{t1}}} \\ \vdots & \ddots & \vdots \\ \overline{\overline{a_{1k}}} & \cdots & \overline{\overline{a_{tk}}} \end{pmatrix} =$

$$\begin{pmatrix} \overline{a_{11}^{(0)}} & \cdots & \overline{a_{11}^{(n-1)}} & \cdots & \overline{a_{t1}^{(0)}} & \cdots & \overline{a_{t1}^{(n-1)}} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \overline{a_{1k}^{(0)}} & \cdots & \overline{a_{1k}^{(n-1)}} & \cdots & \overline{a_{tk}^{(0)}} & \cdots & \overline{a_{tk}^{(n-1)}} \end{pmatrix} = \left(\overline{A_1^{(0)}} \cdots \overline{A_1^{(n-1)}} \cdots \overline{A_t^{(0)}} \cdots \overline{A_t^{(n-1)}}\right).$$

Appendix B. The proof of proposition 1

From lemma 5 (iii), for $r \in \{0, 1, \cdots, n-2\}$, we have the following equations:

$$\alpha_f^{(r+1)} = \begin{cases} -\alpha_{n-1}^{(r)} & f = 0 \\ \alpha_{f-1}^{(r)} & 1 \le f \le n-1 \end{cases}, \tag{B.1}$$

$$\alpha_f^{(r)} = \begin{cases} \alpha_{f+1}^{(r+1)} & 0 \le f \le n-2 \\ -\alpha_0^{(r+1)} & f = n-1 \end{cases}. \tag{B.2}$$

We denote $P = \left(P_{ij}\right)_{k\times z}, \alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_z \end{pmatrix}$ and $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$. Let $P_{ij,l}, \alpha_{i,l}, \beta_{i,l}, P_{ij,l}^{(r)}, \alpha_{i,l}^{(r)}, \beta_{i,l}^{(r)}$ be the l-th $(0 \le l \le n-1)$ degree coefficients of $P_{ij}, \alpha_i, \beta_i$ and $P_{ij}^{(r)}, \alpha_i^{(r)}, \beta_i^{(r)}$ respectively, then from $\overline{\overline{P}} \cdot \overline{\alpha^{(r)}} = \overline{\beta^{(r)}}$, we have

$$\begin{pmatrix} \overline{\overline{P_{11}}} & \cdots & \overline{\overline{P_{1z}}} \\ \vdots & \ddots & \vdots \\ \overline{\overline{P_{k1}}} & \cdots & \overline{\overline{P_{kz}}} \end{pmatrix} \cdot \begin{pmatrix} \overline{\alpha_1^{(r)}} \\ \vdots \\ \overline{\alpha_z^{(r)}} \end{pmatrix} = \begin{pmatrix} \overline{\beta_1^{(r)}} \\ \vdots \\ \overline{\beta_k^{(r)}} \end{pmatrix}.$$

Then for $i = 1, 2, \cdots, k$,

$$\overline{\beta_i^{(r)}} = \sum_{d=1}^{z} \overline{\overline{P_{id}}} \cdot \overline{\alpha_d^{(r)}} = \sum_{d=1}^{z} \begin{pmatrix} \overline{P_{id}^{(0)}} & \cdots & \overline{P_{id}^{(n-1)}} \end{pmatrix} \begin{pmatrix} \alpha_{d,0}^{(r)} \\ \vdots \\ \alpha_{d,n-1}^{(r)} \end{pmatrix} = \sum_{d=1}^{z} \sum_{c=0}^{n-1} \overline{P_{id}^{(c)}} \cdot \alpha_{d,c}^{(r)}.$$

Compare the entries of the vectors on both sides of the above formula, we have

$$\sum_{d=1}^{z} \sum_{c=0}^{n-1} \alpha_{d,c}^{(r)} P_{id,f}^{(c)} = \beta_{i,f}^{(r)},\ f = 0, 1, \cdots, n-1;\ i = 1, 2, \cdots, k. \tag{B.3}$$

Now we consider the value of $\beta_{i,f}^{(r+1)}$ in two cases of $f = 0$ and $1 \le f \le n-1$.

Case 1 $f = 0$.

$$\begin{aligned} \beta_{i,0}^{(r+1)} &\overset{equ.(B.1)}{=\!=} -\beta_{i,n-1}^{(r)} \overset{equ.(B.3)}{=\!=} -\sum_{d=1}^{z} \sum_{c=0}^{n-1} \alpha_{d,c}^{(r)} P_{id,n-1}^{(c)} \\ &\overset{\substack{equ.(B.2) \\ equ.(B.1)}}{=\!=} -\sum_{d=1}^{z} \left(\sum_{c=0}^{n-2} \alpha_{d,c+1}^{(r+1)} \cdot \left(-P_{id,0}^{(c+1)}\right) + \left(-\alpha_{d,0}^{(r+1)}\right) \cdot P_{id,0}^{(0)} \right) \\ &\overset{c'=c+1}{=\!=} \sum_{d=1}^{z} \left(\sum_{c'=1}^{n-1} \alpha_{d,c'}^{(r+1)} P_{id,0}^{(c')} + \alpha_{d,0}^{(r+1)} \cdot P_{id,0}^{(0)} \right) \\ &= \sum_{d=1}^{z} \sum_{c'=0}^{n-1} \alpha_{d,c'}^{(r+1)} \cdot P_{id,0}^{(c')}. \end{aligned}$$

Case 2 $1 \le f \le n-1$.

$$\begin{aligned} \beta_{i,f}^{(r+1)} &\overset{equ.(B.1)}{=\!=} \beta_{i,f-1}^{(r)} \overset{equ.(B.3)}{=\!=} \sum_{d=1}^{z} \sum_{c=0}^{n-1} \alpha_{d,c}^{(r)} P_{id,f-1}^{(c)} \\ &\overset{\substack{equ.(B.2) \\ equ.(B.1)}}{=\!=} \sum_{d=1}^{z} \left(\sum_{c=0}^{n-2} \alpha_{d,c+1}^{(r+1)} P_{id,f}^{(c+1)} + \alpha_{d,n-1}^{(r)} P_{id,f-1}^{(n-1)} \right) \\ &\overset{c'=c+1}{=\!=} \sum_{d=1}^{z} \left(\sum_{c'=1}^{n-1} \alpha_{d,c'}^{(r+1)} P_{id,f}^{(c')} + \left(-\alpha_{d,0}^{(r+1)}\right) \cdot \left(-P_{id,f}^{(0)}\right) \right) \\ &= \sum_{d=1}^{z} \sum_{c'=0}^{n-1} \alpha_{d,c'}^{(r+1)} P_{id,f}^{(c')}. \end{aligned}$$

Combine case 1 with case 2, we get

$$\sum_{d=1}^{z} \sum_{c=0}^{n-1} \alpha_{d,c}^{(r+1)} P_{id,f}^{(c)} = \beta_{i,f}^{(r+1)},\ i = 1, 2, \cdots, k;\ f = 0, 1, \cdots, n-1.$$

So

$$\begin{aligned} \overline{\beta_i^{(r+1)}} &= \sum_{d=1}^{z} \sum_{c=0}^{n-1} \begin{pmatrix} \alpha_{d,c}^{(r+1)} P_{id,0}^{(c)} \\ \vdots \\ \alpha_{d,c}^{(r+1)} P_{id,n-1}^{(c)} \end{pmatrix} = \sum_{d=1}^{z} \sum_{c=0}^{n-1} \alpha_{d,c}^{(r+1)} \overline{P_{id}^{(c)}} \\ &= \sum_{d=1}^{z} \begin{pmatrix} \overline{P_{id}^{(0)}} & \cdots & \overline{P_{id}^{(n-1)}} \end{pmatrix} \begin{pmatrix} \alpha_{d,0}^{(r+1)} \\ \vdots \\ \alpha_{d,n-1}^{(r+1)} \end{pmatrix} = \sum_{d=1}^{z} \overline{\overline{P_{id}}} \cdot \overline{\alpha_d^{(r+1)}} \\ &= \begin{pmatrix} \overline{\overline{P_{i1}}} & \cdots & \overline{\overline{P_{id}}} \end{pmatrix} \begin{pmatrix} \overline{\alpha_1^{(r+1)}} \\ \vdots \\ \overline{\alpha_z^{(r+1)}} \end{pmatrix} = \overline{\overline{P_i}} \cdot \overline{\alpha^{(r+1)}},\ i = 1, 2, \cdots, k. \end{aligned}$$

Then the conclusion is drawn since

$$\overline{\overline{P}} \cdot \overline{\alpha^{(r+1)}} = \begin{pmatrix} \overline{\overline{P_1}} \\ \vdots \\ \overline{\overline{P_k}} \end{pmatrix} \cdot \overline{\alpha^{(r+1)}} = \begin{pmatrix} \overline{\beta_1^{(r+1)}} \\ \vdots \\ \overline{\beta_k^{(r+1)}} \end{pmatrix} = \overline{\beta^{(r+1)}}.$$

References

[1] O. Regev, On lattices, learning with errors, random linear codes, and cryptography, in: Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing, STOC '05, Association for Computing Machinery, New York, NY, USA, 2005, pp. 84–93. doi:10.1145/1060590.1060603.

[2] V. Lyubashevsky, C. Peikert, O. Regev, On ideal lattices and learning with errors over rings, in: H. Gilbert (Ed.), Advances in Cryptology – EUROCRYPT 2010, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 1–23.

[3] Z. Brakerski, C. Gentry, V. Vaikuntanathan, (leveled) fully homomorphic encryption without bootstrapping, in: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, ITCS '12, Association for Computing Machinery, New York, NY, USA, 2012, p. 309–325. doi:10.1145/2090236.2090262.
URL https://doi.org/10.1145/2090236.2090262

[4] A. Langlois, D. Stehlé, Worst-case to average-case reductions for module lattices, Designs, Codes and Cryptography 75 (3) (2015) 565–599. doi:10.1007/s10623-014-9938-4.

[5] E. Alkim, R. Avanzi, J. Bos, L. Ducas, A. de la Piedra, T. Pöppelmann, P. Schwabe, D. Stebila, M. R. Albrecht, E. Orsini, et al., Newhope algorithm specifications and supporting documentation, NIST PQC Round 2 (2019).

[6] X. Lu, Y. Liu, Z. Zhang, D. Jia, H. Xue, J. He, B. Li, K. Wang, Lac: Practical ring-lwe based public-key encryption with byte-level modulus, Cryptology ePrint Archive, Report 2018/1009 (2018).

[7] R. Avanzi, J. Bos, L. Ducas, E. Kiltz, T. Lepoint, V. Lyubashevsky, J. M. Schanck, P. Schwabe, G. Seiler, D. Stehlé, Crystals-kyber algorithm specifications and supporting documentation (version 3.01), NIST PQC Selected Algorithms (2021) 1–43.

[8] S. Bai, L. Ducas, E. Kiltz, T. Lepoint, V. Lyubashevsky, P. Schwabe, G. Seiler, D. Stehlé, Crystals-dilithium algorithm specifications and supporting documentation (version 3.1), NIST PQC Selected Algorithms (2021) 1–38.

[9] R. Cramer, L. Ducas, C. Peikert, O. Regev, Recovering short generators of principal ideals in cyclotomic rings, in: Proceedings, Part II, of the 35th Annual International Conference on Advances in Cryptology — EUROCRYPT 2016 - Volume 9666, Springer-Verlag, Berlin, Heidelberg, 2016, p. 559–585.

[10] R. Cramer, L. Ducas, B. Wesolowski, Short stickelberger class relations and application to ideal-svp, in: J.-S. Coron, J. B. Nielsen (Eds.), Advances in Cryptology – EUROCRYPT 2017, Springer International Publishing, Cham, 2017, pp. 324–348.

[11] O. Bernard, A. Roux-Langlois, Twisted-phs: Using the product formula to solve approx-svp in ideal lattices, in: S. Moriai, H. Wang (Eds.), Advances in Cryptology – ASIACRYPT 2020, Springer International Publishing, Cham, 2020, pp. 349–380.

[12] Y. Pan, J. Xu, N. Wadleigh, Q. Cheng, On the ideal shortest vector problem over random rational primes, in: A. Canteaut, F.-X. Standaert (Eds.), Advances in Cryptology – EUROCRYPT 2021, Springer International Publishing, Cham, 2021, pp. 559–583.

[13] S. Nakamura, M. Yasuda, An extension of kannan's embedding for solving ring-based lwe problems, in: M. B. Paterson (Ed.), Cryptography and Coding, Springer International Publishing, Cham, 2021, pp. 201–219.

[14] R. Kannan, Minkowski's Convex Body Theorem and Integer Programming, Mathematics of Operations Research 12 (3) (1987) 415–440. doi:10.1287/moor.12.3.415.

[15] S. Bai, S. D. Galbraith, Lattice decoding attacks on binary lwe, in: W. Susilo, Y. Mu (Eds.), Information Security and Privacy, Springer International Publishing, Cham, 2014, pp. 322–337.

[16] K. E. Stange, Algebraic aspects of solving ring-lwe, including ring-based improvements in the blum–kalai–wasserman algorithm, SIAM Journal on Applied Algebra and Geometry 5 (2) (2021) 366–387. doi:10.1137/19M1280442.

[17] Z. Chunhuan, Z. Zhongxiang, W. Xiaoyun, X. Guangwu, Distinguishing lwe instances using fourier transform: A refined framework and its applications, Cryptology ePrint Archive, Paper 2019/1231, https://eprint.iacr.org/2019/1231 (2019).
URL https://eprint.iacr.org/2019/1231

[18] H. Wu, X. Wang, G. Xu, On the dual attack of lwe schemes in the presence of hints, Cryptology ePrint Archive, Paper 2022/1403, https://eprint.iacr.org/2022/1403 (2022).
URL https://eprint.iacr.org/2022/1403

[19] R. Adler, Institute of mathematical statistics lecture notes—monograph series (1990).

[20] D. Dachman-Soled, L. Ducas, H. Gong, M. Rossi, Lwe with side information: Attacks and concrete security estimation, in: D. Micciancio, T. Ristenpart (Eds.), Advances in Cryptology – CRYPTO 2020, Springer International Publishing, Cham, 2020, pp. 329–358.

[21] H. Wu, X. Wang, G. Xu, Reducing an lwe instance by modular hints and its applications to primal attack, dual attack and bkw attack, Cryptology ePrint Archive, Paper 2022/1404, https://eprint.iacr.org/2022/1404 (2022).
URL https://eprint.iacr.org/2022/1404

[22] C. P. Schnorr, M. Euchner, Lattice basis reduction: Improved practical algorithms and solving subset sum problems, in: L. Budach (Ed.), Fundamentals of Computation Theory, Springer Berlin Heidelberg, Berlin, Heidelberg, 1991, pp. 68–85.

[23] Y. Chen, P. Q. Nguyen, Bkz 2.0: Better lattice security estimates, in: D. H. Lee, X. Wang (Eds.), Advances in Cryptology – ASIACRYPT 2011, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 1–20.

[24] A. Becker, L. Ducas, N. Gama, T. Laarhoven, New directions in nearest neighbor searching with applications to lattice sieving, in: Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '16, Society for Industrial and Applied Mathematics, USA, 2016, p. 10–24.

[25] Y. Chen, Réduction de réseau et sécurité concrete du chiffrement completement homomorphe, Ph.D. thesis, Paris 7 (2013).

[26] M. Naehrig, E. Alkim, J. Bos, L. Ducas, K. Easterbrook, B. LaMacchia, P. Longa, I. Mironov, V. Nikolaenko, C. Peikert, et al., Frodokem: Learning with errors key encapsulation–algorithm specifications and supporting documentation, NIST PQC Round 2 (2019).

[27] B. Applebaum, D. Cash, C. Peikert, A. Sahai, Fast cryptographic primitives and circular-secure encryption based on hard learning problems (2009) 595–618.

[28] D. Micciancio, O. Regev, Lattice-based cryptography, in: D. J. Bernstein, J. Buchmann, E. Dahmen (Eds.), Post-Quantum Cryptography – PQCrypto 2009, Springer Berlin Heidelberg, 2009, pp. 147–191. doi:10.1007/978-3-540-88702-7_5.

[29] S. Li, X. Lu, J. Zhang, B. Li, L. Bi, Predicting the concrete security of lwe against the dual attack using binary search, in: D. Gao, Q. Li, X. Guan, X. Liao (Eds.), Information and Communications Security, Springer International Publishing, Cham, 2021, pp. 265–282.