On the Dual Attack of LWE Schemes in the Presence of Hints

Han Wu[1,2], Xiaoyun Wang[1,2,3], and Guangwu Xu[1,2](✉)

[1] School of Cyber Science and Technology, Shandong University, Qingdao 266237, Shandong, China
hanwu97@mail.sdu.edu.cn

[2] Key Laboratory of Cryptologic Technology and Information Security of Ministry of Education, Shandong University, Qingdao 266237, Shandong, China
gxu4sdq@sdu.edu.cn

[3] Institute for Advanced Study, Tsinghua University, Beijing 100084, China
xiaoyunwang@mail.tsinghua.edu.cn

Abstract. Combining theoretical-based traditional attack method with practical-based side-channel attack method provides more accurate security estimations for post-quantum cryptosystems. In CRYPTO 2020, Dachman-Soled et al. integrated hints from side-channel information to the primal attack against LWE schemes. This paper develops a general Fourier analytic framework to work with the dual attack in the presence of hints. Distinguishers that depend on specific geometric properties related to hints are established. The Fourier transform of discretized multivariate conditional Gaussian distribution on $\mathbb{Z}_q^d$ is carefully computed and estimated, some geometric characteristics of the resulting distinguisher are explored and a new model of dual attack is proposed. In our framework, an adversary performs the BKZ algorithm directly in a projected lattice to find short projection components, and then recovers them by MLLL algorithm to make a distinction. This method relies on a reasonable assumption and is backed up by naturally formed mathematical arguments. The improvements and the assumption are validated by experiments. For examples, for a Kyber768 instance, with 200 hints, the blocksize can be reduced by at least 188 and the time complexity can be reduced by a factor of greater than 2^{55}. After adding 300 hints to a FireSaber instance, even in the worst case, the blocksize drops from 819 to 542, and the cost drops from $2^{255.61}$ to $2^{174.72}$.

1 Introduction

We are facing an urgent task of gradually replacing classical public key systems with new ones because of the rapid advance of computing technology. Especially for the coming era of quantum computer, concerns of threats to break several widely used public key cryptographic schemes promote the innovations on post-quantum cryptography. In 2017, the US National Institute of Standards and Technology (NIST) solicited proposals for post-quantum cryptography (PQC) primitives including public key encryption/key encapsulation mechanisms and digital signatures to prepare for the reality of practical quantum computing. Proposals to PQC have come from several different areas. For examples, there were submissions of code-based, lattice-based and multivariate-based schemes. In July 2022, NIST identified four candidate algorithms for standardization. Among them, three candidates are from the family of lattice-based cryptography. In particular, the two primary algorithms recommended by NIST – Kyber and Dilithium – are both built on lattice hard problems.

There are two classes of well studied mathematical hard problems to support lattice-based cryptography, one class is the NTRU problem and the other is the learning with error (LWE) problem (as well as its variants). We shall say a bit more on the latter as our discussion falls into this category. The LWE problem, proposed by Regev [29], is one of the most important computational problems and is proved to be at least as hard as (quantumly) solving some (approximate) shortest vector problems. There are many variants of LWE such as RLWE (Ring-LWE) and MLWE (Module-LWE). It is mentioned that learning with rounding (LWR) problem is another LWE variant with determined errors. We have seen many exciting applications of LWE in post-quantum cryptography, for examples, NIST PQC standard Kyber, the third round candidate algorithm Saber and the second round candidate algorithms FrodoKEM, Newhope are all based on LWE.

In cryptanalysis, there are four notable types of attacks for public key systems based on LWE. They are lattice attack, algebraic attack, BKW attack and attack based on the failure of decryption. In practice, the fact that the number of available samples is restricted makes lattice attack method an effective choice. This type of attack includes dual attack and primal attack. The basic idea of both is to transform the problem into searching short vectors in certain lattice, and then solve it by lattice reduction algorithms. It is noted that the complexities for dual attack and primal attack are quite similar for most cryptosystems.

This paper will mainly work with dual attack. This kind of attack was first proposed by Micciancio and Regev [25] in 2009. In a dual attack, for a given instance $(A, b) \in \mathbb{Z}_q^{m\times n} \times \mathbb{Z}_q^m$, the adversary looks for a large number of short vectors in certain lattice and then calculates the inner products of b with (the first m entries of) these vectors respectively. A distinguisher will be used to determine whether b is from an LWE instance or a uniform instance by revealing the difference in the distributions of these inner products in the two cases. Using this distinguisher, the secret can also be obtained easily. Dual attack has been studied extensively since it was proposed. An optimization of dual attack for LWE with small secret was suggested by Albrecht et al. [3] in 2014. A further analysis of dual attack was given by Alkim et al. [6] in 2016. In [1], Albrecht introduced the "scaling factor" and achieved further optimization for "sparse" small secret. In 2021, Li et al. [21] proved that the cost function of dual attack is actually a U-shape function and applied binary search to predict the minimum cost. In the same year, Guo et al. [17] presented an improved distinguisher that in combination with a guessing step. A new two-step lattice reduction strategy was also given.

With the rapid development of lattice-based cryptography, it is natural to consider whether sensitive side-channel information can be extracted and how to assess the threat of side-channel attacks. There have been many researches in this regard, see, for example, [2, 7, 10, 12, 18, 27]. In 2020, Dachman-Soled et al. [15] initiated a study of using pieces of side-channel information about secret/error as "hints", and integrating them into the primal attack. The reduction of the cost of primal attack after adding those hints was discussed. This opens a new direction of mixing the theoretical-based lattice attack method and practical-based side-channel attack method to advance the cryptanalysis of LWE schemes.

In this paper, we consider the idea of integrating hints into the dual attack. The mathematical formulation and characterization of hints seem to make it a natural and appropriate combination with dual attack. We shall use the Fourier transform in this setting in an extensive manner.

The Fourier transform has always been a powerful tool as long as a right model setup is provided. The dual attack against LWE schemes fits in such a situation well. It has been proven that the Fourier transform of the so called discrete normal (or Gaussian) distribution on the group $\mathbb{Z}_q^d$ can be served as a distinguisher to identify the error distribution from the uniform distribution [32]. Such a distribution is essentially a discretized version of the normal distribution $N(0, \sigma_\chi^2 I_d)$. Some *invariant property* of Gaussian function under the (continuous) Fourier transform enables one to use the classical Poisson summation formula in a neat manner to achieve the desired result. [32] suggested a refined Fourier analytic method by introducing and characterizing *local widths* and calculated Fourier transforms of several more distributions to investigate the local behaviors of a distinguisher. Some of their ideas are influential and generalized further in this work.

In the first part of this paper, we present an extensive study of dual attack using Fourier distinguisher. We establish a general framework to deal with the discretization of any type of multivariate normal distribution $N(\mu, \Sigma)$. The particular interest here is to cover the case of the covariance matrix Σ being degenerate. Therefore the conditional multivariate normal distribution derived from hints can be accommodated. It is noted that the (continuous) Fourier transform of a general Gaussian function in an r-dimensional space is still a Gaussian function, but due to the measure theoretical nature of integration, care must be taken when we perform the Fourier transform of the Gaussian function (i.e. pdf of $N(\mu, \Sigma)$) with a degenerate matrix Σ. We are able to obtain a nice expression of the Fourier transform of a discrete (multivariate) normal distribution on $\mathbb{Z}_q^d$ after a smooth application of the Poisson summation formula. A distinguisher is thus implied.

The generality of the distinguisher is reflected by the dependency on more aspects of each short vector and we can work with more geometrical features of the hints. More precisely, it is only the length of some projected component of the short vector that determines the distinguish advantage. This leads to a natural idea that we could simply look for short components in the projected lattice and then recover them to the ones belong to the original lattice. To make a distinction, some transformation may be necessary in theoretical analysis. The above ideas indicate a new model of the dual attack with hints, which is proposed in the second part of this paper. This new approach relies on a reasonable assumption and is backed up by naturally formed mathematical arguments. The lattice used in this approach has a lower dimension and a smaller volume. By analyzing the relationship between the complexity of dual attack and the parameters (i.e. volume and dimension) of the lattice, the performance of our model is proved theoretically. Some benefits have been discovered, making it seems natural and convenient to add hints to a dual attack. Experiments verify the efficiency of our new model as well as the assumption we based on. Multiple hints can indeed significantly reduce the cost of the dual attack. For examples, for a Kyber768 instance, with 200 hints, the blocksize can be reduced by at least 188 and the time complexity can be reduced by a factor of greater than 2^{55}. After adding 300 hints to a FireSaber instance, even in the worst case, the blocksize drops from 819 to 542, and the cost drops from $2^{255.61}$ to $2^{174.72}$.

The paper is organized into 5 sections. Necessary preparations together with some relevant mathematical background and useful algorithms are given in Section 2. In Section 3, we develop a Fourier analytic framework for deriving distinguish advantages in the presence of hints. The corresponding distinguisher is also given. Based on this, we propose a new model of the dual attack with hints in Section 4, in which short vectors are searched in a new lattice of a lower dimension and a smaller volume, and then restored to make a distinction. This new approach relies on a reasonable assumption, some explanations for it are also given. Experiments in Section 5 verify this assumption as well as the efficiency of our new model. Some additional benefits of adding hints to dual attack are also listed.

2 Preliminaries

In this section, we provide necessary preparations for the discussion of the integration of hints to the dual attack against LWE-based encryption schemes. These include some relevant mathematical background and useful algorithms.

2.1 BKZ

A lattice basis reduction algorithm transforms a lattice basis to a new one that consists of short lattice vectors. This paper will mainly use the BKZ algorithm and its variants.

The Hermite factor The quality of the output of a lattice reduction algorithm can be characterized by the *Hermite factor*. Let $d \leq k$ be two positive integers and $B \in \mathbb{R}^{k\times d}$ be a matrix whose column vectors form a basis of a d-dimensional lattice $L \subseteq \mathbb{R}^k$. Then the *volume/determinant* of L is defined as $\mathrm{vol}(L) = \det(L) = \sqrt{\det(B^T B)}$.

Definition 1. *We say that a lattice reduction algorithm has a Hermite factor δ_0, if its output basis satisfies the following condition:*

$$\|b_1\| \leq \delta_0^d \det(L)^{\frac{1}{d}},$$

where the input is a basis of a d-dimensional lattice L and b_1 is the first output vector.

The BKZ algorithm and its variants are generally regarded as the most common and efficient lattice reduction algorithms. There is a blocksize parameter β in BKZ, which determines the quality of the output vectors. To be specific, for a β that is not too small (for example, $\beta \geq 50$), the Hermite factor is predictable by the following heuristic, which has been experimentally verified in [13].

Heuristic 1 *BKZ-$\beta(\beta \geq 50)$ achieves Hermite factor $\delta_0(\beta) \approx \left(\frac{\beta}{2\pi e}(\pi\beta)^{\frac{1}{\beta}}\right)^{\frac{1}{2(\beta-1)}}$.*

The sieving algorithm One could choose to use sieving or enumeration as the SVP oracle when performing BKZ. In this paper, we will be focusing on the former. As assumed in [6], in this case, a typical run of BKZ produces a large number of short vectors whose norms are all close to that of the shortest output vector.

Assumption 1 *For a d-dimensional lattice L, given any of its basis as input, BKZ-β provides $2^{0.2075\beta}$ short vectors in one run when using sieving as the SVP oracle, whose norms are all close to $\delta_0(\beta)^d \cdot \det(L)^{\frac{1}{d}}$.*

The cost of BKZ is usually estimated using some heuristic assumptions. When the blocksize is set to β, its runtime is generally considered as $2^{c\beta+o(\beta)}$, where c is a constant. In 2016, Becker et al. [11] showed that applying spherical LSF to sieving algorithms leads to $c = 0.292$ in the classical case, while when considering quantum situation, the search process can be accelerated so that c is reduced to 0.265. As done in [4, Footnote 5], in this paper, we shall calculate the cost of BKZ in a relatively accurate way without ignoring $o(\beta)$ as follows.

Assumption 2 *When using sieving as the SVP oracle, the runtime of BKZ-β is*

$$T_{BKZ}(\beta) = \begin{cases} 2^{0.292\beta+16.4} & \textit{classical case} \\ 2^{0.265\beta+16.4} & \textit{quantum case} \end{cases}.$$

It is generally considered that the short vectors found by BKZ are non-directional, that is, they have balanced coefficients. More precisely, each entry of the obtained short vectors is assumed to be subject to the same Gaussian distribution independently.

Assumption 3 *([16]) 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 $\frac{\|v\|}{\sqrt{d}}$.*

2.2 Fourier Transform

The Fourier transform characterizes mathematical duality in that a function localized in the time domain can be also viewed to spread out across the frequency domain. It has been shown to be a very powerful tool for lattice theory. We will use the Fourier transforms on the abelian groups $\mathbb{R}^d$ and $\mathbb{Z}_q^d$ respectively. The latter is also called the discrete Fourier transform.

Definition 2. *(1) For a rapidly decreasing smooth function $f : \mathbb{R}^d \to \mathbb{C}$ [4], its Fourier transform $\widehat{f} : \mathbb{R}^d \to \mathbb{C}$ is given by*

$$\widehat{f}(y) = \int_{\mathbb{R}^d} e^{-2\pi i<x,y>} \cdot f(x)\, dx,\ \forall y \in \mathbb{R}^d.$$

(2) For a function $f : \mathbb{Z}_q^d \to \mathbb{C}$, its discrete Fourier transform $\widehat{f} : \mathbb{Z}_q^d \to \mathbb{C}$ is given by

$$\widehat{f}(y) = \sum_{x\in\mathbb{Z}_q^d} e^{\frac{-2\pi i<x,y>}{q}} \cdot f(x),\ \forall y \in \mathbb{Z}_q^d.$$

[4] I.e., f and all its (partial) derivatives $D^\beta f$ satisfy $\sup_{x\in\mathbb{R}^d} |x^\alpha D^\beta f(x)| < \infty$ for every $\alpha, \beta \in \mathbb{N}^d$. Such a function is said to be in the Schwartz space. The Fourier transform can be extended to a larger family of functions, including probability density functions.

The following is a useful property of the Fourier transform of a function composed with an affine transform.

Lemma 1. *([26, Section 8.2.3]) For a rapidly decreasing smooth function $f : \mathbb{R}^d \to \mathbb{C}$, let $M \in \mathbb{R}^{d\times d}$ be an invertible matrix and $h \in \mathbb{R}^d$, then the Fourier transform of $f(Mx + h)$ is $\frac{1}{|\det(M)|} \cdot e^{2\pi i \langle h, M^{-T}y\rangle} \widehat{f}(M^{-T}y)$.*

The next classical Poisson summation formula provides a fundamental way to link a function with its Fourier transform in terms of periodic summations of the function and its Fourier transform. This naturally involves a lattice and its dual.

Lemma 2. *([30, Proposition 15]) Let L be a d-dimensional lattice and L^* be its dual lattice. For a rapidly decreasing smooth function $f : \mathbb{R}^d \to \mathbb{C}$,*

$$\sum_{x\in L} f(x) = \det(L^*) \sum_{y\in L^*} \widehat{f}(y).$$

In our latter discussion, the Poisson summation formula will be used in a slightly different form. For $a \in \mathbb{R} \setminus \{0\}$ and $h \in \mathbb{R}^d$, we define $g(x) = f(ax + h)$, and it is easy to see that $\widehat{g}(y) = \frac{1}{|a|^d} \cdot e^{\frac{2\pi i <h,y>}{a}} \cdot \widehat{f}\left(\frac{y}{a}\right)$ according to lemma 1. Combining this with lemma 2, the following corollary is obtained.

Corollary 1. *Let L be a d-dimensional lattice and L^* be its dual lattice. Given $a \in \mathbb{R} \setminus \{0\}$, $h \in \mathbb{R}^d$. For a rapidly decreasing smooth function $f : \mathbb{R}^d \to \mathbb{C}$,*

$$\sum_{x\in L} f(ax + h) = \frac{\det(L^*)}{|a|^d} \cdot \sum_{y\in L^*} e^{\frac{2\pi i <h,y>}{a}} \cdot \widehat{f}\left(\frac{y}{a}\right).$$

2.3 Multivariate Normal Distribution

Definition 3. *The singular value decomposition (SVD) of a matrix $M \in \mathbb{R}^{m\times n}$ $(m \geq n)$ of rank r is given by*

$$M = UDV^T,$$

where $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ are both orthogonal matrices, D is an $m\times n$ rectangular diagonal matrix of the form $D = \left(\begin{array}{ccc} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \hline & O_{(m-n)\times n} & \end{array}\right)$ with $\sigma_1 \geq \cdots \geq \sigma_r > 0$ and $\sigma_{r+1} = \cdots = \sigma_n = 0$. The SVD for the case of $m < n$ is similar.

When $M \in \mathbb{R}^{m\times m}$ is a symmetric matrix, we can simply consider its *eigenvalue decomposition* (EVD). In this case, $M = QDQ^T$ for an orthogonal matrix Q and a diagonal matrix D. More precisely, $D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{pmatrix}$ with $\{\lambda_i\}_{i=1}^m$ being the eigenvalues of M, and the column vectors of Q are the eigenvectors of M.

SVD makes it easier to give definitions of *pseudo inverse* and *pseudo determinant* of a matrix. Specifically, let $M \in \mathbb{R}^{m\times n}(m \geq n)$ be a matrix of rank r, suppose its

SVD is $M = UDV^T$, where $D = \left(\begin{array}{ccccc} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & 0 \\ \hline \multicolumn{5}{c}{O_{(m-n)\times n}} \end{array}\right)$, $\sigma_1 \geq \cdots \geq \sigma_r > 0$. We define the *pseudo inverse* of D as $D^{\sim} = \left(\begin{array}{ccccc} \frac{1}{\sigma_1} & & & & \\ & \ddots & & & \\ & & \frac{1}{\sigma_r} & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & 0 \\ \hline \multicolumn{5}{c}{O_{(m-n)\times n}} \end{array}\right)$ and the *pseudo inverse* of M as $M^{\sim} = VD^{\sim}U^T$. We define the *pseudo determinant* of M as $\mathrm{rdet}(M) = \sigma_1 \cdot \sigma_2 \cdots \sigma_r$. These definitions will be very useful when considering degenerate multivariate normal distribution.

Definition 4. *Let d be a positive integer. For $\mu \in \mathbb{R}^d$ [5] and a symmetric matrix $\Sigma \in \mathbb{R}^{d\times d}$ of rank r, we denote the (continuous) multivariate normal distribution with mean μ and covariance matrix Σ by $N_d(\mu, \Sigma)$, whose probability density function(pdf) is*

$$f^d_{\mu,\Sigma}(x) = \begin{cases} \frac{1}{(2\pi)^{\frac{r}{2}} \cdot \sqrt{rdet(\Sigma)}} \cdot e^{-\frac{1}{2}(x-\mu)^T \Sigma^{\sim}(x-\mu)}, & x \in \mu + \mathit{Span}(\Sigma) \\ 0, & \mathit{else} \end{cases}.$$

As the entries of the secret and error are usually selected from $\mathbb{Z}_q$ in the actual schemes, the finite discrete version constrained on $\mathbb{Z}_q^d$ needs to be considered. We give the following definition.

Definition 5. *Let d be a positive integer. For $\mu \in \mathbb{R}^d$ and a symmetric matrix $\Sigma \in \mathbb{R}^{d\times d}$, we denote $G_{d,q}(\mu, \Sigma)$ to be the (discrete) multivariate normal distribution derived from $N_d(\mu, \Sigma)$ with the probability mass function (pmf) being*

$$g^{d,q}_{\mu,\Sigma}(x) = \frac{\sum_{t\in\mathbb{Z}^d} f^d_{\mu,\Sigma}(x + tq)}{f^d_{\mu,\Sigma}(\mathbb{Z}^d)}, \quad x \in \mathbb{Z}^d_q,$$

where $f^d_{\mu,\Sigma}$ is the pdf of $N_d(\mu, \Sigma)$.

Conditional multivariate normal distribution plays an important role in our analysis. The following result is a straightforward derivation from properties of the standard multivariate normal distributions (see also in [15]).

Lemma 3. *Let $r < d$ be two positive integers. Let $x \sim N_d(\mu_x, \Sigma_x)$ be a random vector. For a matrix $M \in \mathbb{R}^{r\times d}$ of rank r and a random vector $g \sim N_r(0, \Sigma_g)$, we denote $y = Mx + g$. Then the conditional multivariate normal distribution $(x|y)$ also follows a multivariate normal distribution $N_d(\mu_{x|y}, \Sigma_{x|y})$, where*

$$\begin{cases} \mu_{x|y} = \mu_x + \Sigma_x M^T(M\Sigma_x M^T + \Sigma_g)^{-1}(y - M\mu_x) \\ \Sigma_{x|y} = \Sigma_x - \Sigma_x M^T(M\Sigma_x M^T + \Sigma_g)^{-1}M\Sigma_x \end{cases}.$$

[5] Sometimes the definition may need to be extended to $\mu \in \mathbb{C}$.

In addition, we also need to use some notations about *orthogonal projection* when analyzing the conditional distribution derived from hints.

Definition 6. *Let $X \in \mathbb{R}^{d\times t}(t \leq d)$ be a matrix of rank t and $F = Span(X)$. We denote the orthogonal projection matrix onto F by Π_X or Π_F, and its complement by $\Pi_X^\perp = I - \Pi_X$ or $\Pi_F^\perp = I - \Pi_F$. More specifically, $\Pi_X = X \cdot (X^T X)^{-1} \cdot X^T$.*

It is easy to see that the orthogonal projection matrix is symmetric and idempotent.

2.4 LWE

The Learning With Error (LWE) problem has been a popular problem, especially with its exciting application in post-quantum cryptography. In the original definition of LWE, the secret s is uniformly picked in $\mathbb{Z}_q^n$. Later, LWE in Hermite Normal Form (HNF) was developed, in which the entries of s subject to the same distribution as those of e. In 2009, Applebaum et al. [8] showed that LWE in HNF does not lose security compared with standard LWE and they gave a way of transforming the distribution of the secret to be that of the error through Gaussian elimination.

Definition 7. *For positive integers n, m, q, let χ be a distribution over $\mathbb{Z}_q$ with mean 0 and a small standard deviation of σ_χ, then the Decision-LWE (in Hermite Normal Form) with parameters (m, n, q, χ) is to distinguish pair*

$$(A, b \leftarrow U\left(\mathbb{Z}_q^m\right)) \text{ and } (A, b = As + e \pmod q), \text{ where } A \leftarrow U(\mathbb{Z}_q^{m\times n}), s \leftarrow \chi^n,\ e \leftarrow \chi^m.$$

LWE in HNF is widely used in building cryptographic schemes. For example, NIST PQC algorithms Kyber [9], LAC [24] and Newhope [5] all use the same distribution to sample the entries of the secret and error. In this paper, we also focus on this case.

As mentioned earlier, lattice attacks are usually regarded as the most practical choices due to the limited number of available samples. We consider dual attack in this paper.

Dual Attack In a dual attack, the adversary aims to distinguish between the LWE instance and uniform instance by a distinguisher, with which the secret can also be easily obtained. The steps of dual attack are described below. For a target instance $(A, b) \in \mathbb{Z}_q^{m\times n} \times \mathbb{Z}_q^m$, firstly, the attacker constructs a lattice $\mathscr{L}$ of dimension $d = m+n$ and volume q^n as follows:

$$\mathscr{L} = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{Z}^d : \ A^T x \equiv y \pmod q \right\}.$$

It has the following lattice basis:

$$\mathscr{B} = \begin{pmatrix} I_m & O_{m\times n} \\ A^T & qI_n \end{pmatrix} \in \mathbb{Z}^{d\times d}.$$

Then, the attacker looks for short vectors in $\mathscr{L}$ and then uses them to make a distinction. Broadly speaking, for each short vector $w = \begin{pmatrix} u \\ v \end{pmatrix}$, he/she calculates the value of $\langle u, b\rangle$

$\pmod q$. When $b \leftarrow U(\mathbb{Z}_q^m)$, $\langle u, b\rangle \pmod q$ also follows the uniform distribution over $\mathbb{Z}_q$, while if $b = As + e \pmod q$, we denote $S = \begin{pmatrix} e \\ s \end{pmatrix}$, then

$$\langle u, b\rangle = u^T(As + e) = v^T s + u^T e = \langle S, w\rangle \pmod q$$

will be relatively small as S, w are both short vectors. The difference in the distributions of $\langle u, b\rangle \pmod q$ in the two cases is key to the distinction.

A specific idea of constructing a distinguisher is given in [19]. For any pmf ϕ over $\mathbb{Z}_q$, its *bias* is defined as $\mathbf{B}(\phi) = \mathbb{E}_{x\sim\phi}\left[e^{-\frac{2\pi i x}{q}}\right]$ and it is easy to see that $\mathbf{B}(\phi) = \widehat{\phi}(1)$. Suppose that M short vectors $w_j = \begin{pmatrix} u_j \\ v_j \end{pmatrix}$, $j = 1, 2, \cdots, M$ are used in a dual attack. Let $f_{\langle u_j, b\rangle}$ be the pmf of $\langle u_j, b\rangle \pmod q$, $j = 1, 2, \cdots, M$. The attacker calculates the sample average $\frac{\sum_{j=1}^{M} e^{-\frac{2\pi i \langle u_j, b\rangle}{q}}}{M}$ and it becomes closer to the true mean $\frac{\sum_{j=1}^{M} \mathbf{B}\left(f_{\langle u_j, b\rangle}\right)}{M}$ as M increases. In fact, the so called "*distinguish advantage*" is relevant to the difference between the values of $\frac{\sum_{j=1}^{M} \mathbf{B}\left(f_{\langle u_j, b\rangle}\right)}{M}$ in the two cases. To be more precisely, it is 0 when b is from a uniform instance, and the larger $\left|\frac{\sum_{j=1}^{M} \mathbf{B}\left(f_{\langle u_j, b\rangle}\right)}{M}\right| = \left|\frac{\sum_{j=1}^{M} \mathbf{B}\left(f_{\langle S, w_j\rangle}\right)}{M}\right|$ is in the other case, the better the distinction will be.

Suppose that $\|w_j\| \leq L, j = 1, 2, \cdots, M$. In 2011, an estimation $\mathbf{B}\left(f_{\langle S, w_j\rangle}\right) \geq e^{-\frac{2\pi^2 \cdot \|w_j\|^2 \cdot \sigma_\chi^2}{q^2}} \geq e^{-\frac{2\pi^2 \cdot L^2 \cdot \sigma_\chi^2}{q^2}} := \epsilon$ was given in [22], this means ϵ can be viewed as the advantage from w_j. This method of calculating advantages is widely used, for example, in [6,17,32]. According to the Chernoff-Hoeffding inequality as shown below, $M = O\left(\frac{1}{\epsilon^2}\right)$ samples are sufficient to amplify the success rate of the attack to a constant.

Lemma 4. *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 all $\epsilon \geq 0$,*

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

3 Analyzing the Distinguish Advantage of Dual Attack with Hints by Fourier Transform

As indicated in [15], during an actual attack against an LWE scheme, an attacker might be able to get some "hints" about the secret and/or the error from sources such as side channel information and decryption failures. Combining these hints with a primal attack may leads to a more effective attack. In this section, we investigate the Fourier analytic method in the dual attack in a great detail, some ideas in [32] are pushed further. Some

Fourier transform on $\mathbb{Z}_q^d$ produces a better distinguish advantage. This is done by integrating hints into the dual attack and then applying the (continuous) Fourier transform on a suitable space of a lower dimension. The latter step is mathematically critical. These rigorous theoretical results support the accurate disclosure of the situation after integrating hints to the dual attack. To be specific, we get the following conclusions.

1. Corollary 2 tells us that, in the presence of hints, the distinguish advantage from each short vector found in $\mathscr{L}$ is larger (could be significantly larger provided that sufficiently many hints are available).
2. As described in remark 3, adding hints makes the direction of a short vector in $\mathscr{L}$ an indicator that affects its distinguish advantage. More precisely, unlike the previous case where the advantage depends only on the length of the short vector, now it is also affected by the distance between the short vector and the span of all the hint description vectors.
3. A new distinguisher for the dual attack matching the case with hints is proposed in algorithm 1. Some tricks are used to settle the problem that the argument of the Fourier transform is different for each of the short vectors. Moreover, the effect of this new distinguisher is shown in Section 3.4.

3.1 Integrating Hints into a Dual Attack

Given an LWE instance $(A, b = As + e \pmod q)$ and a short vector $w = \begin{pmatrix} u \\ v \end{pmatrix}$ found in $\mathscr{L}$. As mentioned earlier, in a dual attack, the attacker calculates $\langle u, b\rangle \pmod q$ to distinguish the LWE instance from the uniform instance.

It is noted that if (A, b) is an LWE instance, then $\langle u, b\rangle = \langle S, w\rangle \pmod q$. We denote the pmfs of S and $\langle S, w\rangle \pmod q$ by f_S and $f_{\langle S,w\rangle}$ respectively. [32] contains some indiction that $\widehat{f_{\langle S,w\rangle}}(1) = \widehat{f_S}(w)$ can be used to estimate the distinguish advantage from w, for S from discrete Gaussian (unconditionally). The next proposition extends a conclusion in [32] [6].

Proposition 1. *(1) Let x be a d-dimensional continuous random vector over $\mathbb{R}^d$ with pdf f. For any $v \in \mathbb{R}^d$, we denote the pdf of (the random variable over $\mathbb{R}$) $\langle v, x\rangle$ by $f_{\langle v,x\rangle}$, then*

$$\widehat{f_{\langle v,x\rangle}}(y) = \widehat{f}(yv),\ \forall y \in \mathbb{R}.$$

(2) Let x be a d-dimensional discrete random vector over $\mathbb{Z}_q^d$ with pmf f. 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

$$\widehat{f_{\langle v,x\rangle}}(y) = \widehat{f}(yv),\ \forall y \in \mathbb{Z}_q.$$

[6] Our extension consists of four parts. Firstly, the result is generalized to the continuous case. Secondly, the independence among the coefficients of the random vector is no longer required. Thirdly, the pdf/pmf of each entry can be different. Finally, we prove that the case of $v = 0$ also applies to the proposition.

A proof of proposition 1 is provided in appendix A. Taking $y = 1$ in the above proposition, we get $\widehat{f_{\langle v,x\rangle}}(1) = \widehat{f}(v)$. Hence, when b follows a uniform distribution over $\mathbb{Z}_q^d$, the pmf of $\langle u, b\rangle$ is constant over $\mathbb{Z}_q$ and thus its Fourier transform takes value 0 at 1, i.e., $\widehat{f_{\langle u,b\rangle}}(1) = 0$. While in the other case, it becomes a non-zero complex number $\widehat{f_{\langle u,b\rangle}}(1) = \widehat{f_{\langle S,w\rangle}}(1) = \widehat{f_S}(w)$. Therefore, the difference of the values of $\widehat{f_{\langle u,b\rangle}}(1)$ for these two cases can be used to distinguish. In particular, a lower bound of $|\widehat{f_S}(w)|$ will be useful for estimating the advantage with respect to w.

Since we are focusing on LWE in HNF, it can be assumed that $S \sim G_{d,q}\left(0, \sigma_\chi^2 I_d\right)$ in the original setup without any hints, i.e. $f_S = g^{d,q}_{0,\sigma_\chi^2 I_d}$. However, a more accurate posterior distribution of S can be derived if the adversary obtains some hints about S, and the pmf of S also changes accordingly.

In particular, a hint is characterized as the specific value $\langle S, v\rangle$ of the inner product of S with some vector v, without knowing S. We call the vector v *hint description vector*. For the case with multiple hints, the matrix representation is often used. Suppose the attacker obtains t hints about S, let $Y \in \mathbb{Z}^{d\times t}$ [7] be the matrix whose column vectors are t linearly independent hint description vectors, then hints can be written in the form of $R = Y^T S \in \mathbb{Z}^{t\times 1}$. According to lemma 3, the conditional distribution $(S|Y^T S = R)$ still obeys a normal distribution, whose mean and covariance matrix are

$$\begin{cases} \mu_h = Y(Y^T Y)^{-1} R \\ \Sigma_h = \sigma_\chi^2 I_d - \sigma_\chi^2 Y(Y^T Y)^{-1} Y^T \end{cases}.$$

Hence, we assume $S \sim G_{d,q}\left(\mu_h, \Sigma_h\right)$ and $f_S = g^{d,q}_{\mu_h,\Sigma_h}$ after integrating hints $R = Y^T S$. Let V be the span of the column vectors of Y (i.e. $V = \mathrm{Span}(Y)$), we decompose S into $S = S_V + S_{V^\perp}$ with $S_V \in V, S_{V^\perp} \in V^\perp$. One important fact to note is that, after adding hints, μ_h, Σ_h can also be written as

$$\begin{cases} \mu_h = \Pi_V \cdot S = S_V \\ \Sigma_h = \sigma_\chi^2 \cdot (I_d - \Pi_V) = \sigma_\chi^2 \cdot \Pi_V^\perp \end{cases}.$$

In summary, the mean μ_h actually gives the orthogonal projection S_V of S onto the subspace V, while Σ_h is exactly the orthogonal projection matrix onto $V^\perp$ times σ_χ^2, and they satisfy $\mu_h \in \mathrm{Span}(\Sigma_h)^\perp$.

3.2 Estimating Distinguish Advantages by Fourier Analysis

From the above, we can draw the conclusion that after integrating hints $R = Y^T S$, the distinguish advantage brought by each short vector w depends on $\widehat{g^{d,q}_{\mu_h,\Sigma_h}}(w)$. Thus, a calculation of its value is needed and a tighter lower bound for its absolute value is desirable. The (discrete) multivariate normal distribution on $\mathbb{Z}_q^d$ is defined in terms

[7] To simplify the analysis, we assume that Y is an integer matrix here. Actually, we can also handle the case of $Y \in \mathbb{R}^{d\times t}$, as we shall see later, it is Span(Y) that works. By multiplying a large integer, we can always transform Y into an integer matrix. In addition, we notice that the hints that are available in practice usually have integral coefficients.

of (continuous) multivariate normal distribution while its (discrete) Fourier transform involves the (continuous) Fourier transform and the Poisson summation formula. For the former, an interesting fact to note is that the Fourier transform of a (continuous) multivariate normal distribution is still a (continuous) multivariate normal distribution (without normalization).

Proposition 2. *Let d be a positive integer, given $\mu \in \mathbb{R}^d$ and a symmetric matrix $\Sigma \in \mathbb{R}^{d\times d}$ of full rank. Then for any $y \in \mathbb{R}^d$, we have*

(1) $\widehat{f^d_{\mu,\Sigma}}(y) = \frac{e^{-\frac{1}{2}\mu^T\Sigma^{-1}\mu}\cdot(2\pi)^{\frac{d}{2}}}{\sqrt{\det(\Sigma)}} \cdot f^d_{-i\Sigma^{-1}\mu,\Sigma^{-1}}(2\pi y) = e^{-2\pi i<\mu,y>} \cdot e^{-2\pi^2 y^T\Sigma y}$.

(2) $f^d_{\mu,\Sigma}(y) = \frac{e^{-\frac{1}{2}\mu^T\Sigma^{-1}\mu}}{(2\pi)^{\frac{d}{2}}\cdot\sqrt{\det(\Sigma)}} \cdot \widehat{f^d_{i\Sigma^{-1}\mu,\Sigma^{-1}}}\left(\frac{y}{2\pi}\right)$.

For completeness, we give a proof of proposition 2 in appendix B. It is important to note that, the matrix Σ is required to be of full rank in proposition 2. However, as mentioned earlier, the covariance matrix becomes $\Sigma_h = \sigma_\chi^2 \cdot \Pi_V^\perp$ after adding t hints $R = Y^T S$, then $\text{rank}(\Sigma_h) = d - t < d$. Hence, further consideration is needed.

For the case where $\text{rank}(\Sigma) = r < d$, it is known that $f^d_{\mu,\Sigma}$ is supported by $\mu + \text{Span}(\Sigma)$. To avoid integration on a set with zero measure, we should perform the Fourier transform on the abelian group $\text{Span}(\Sigma)$ (an r-dimensional subspace). Now suppose that the EVD of Σ is $\Sigma = QDQ^T$, where $D =$diag$\{\sigma_1, \cdots, \sigma_r, 0, \cdots, 0\}$. Then $\Sigma^\sim = QD^\sim Q^T$. We write $D_h := \text{diag}\{\sigma_1, \cdots, \sigma_r\} \in \mathbb{R}^{r\times r}$, and denote the matrix consisting of the first r column vectors of Q by Q_r. Then for any $x \in \mu + \text{Span}(\Sigma)$,

$$f^d_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{\frac{r}{2}}\cdot\sqrt{\text{rdet}(\Sigma)}} \cdot e^{-\frac{1}{2}(x-\mu)^T\Sigma^\sim(x-\mu)} \overset{t=x-\mu}{=\!=\!=} \frac{1}{(2\pi)^{\frac{r}{2}}\cdot\sqrt{\text{rdet}(\Sigma)}} \cdot e^{-\frac{1}{2}t^T\Sigma^\sim t}$$
$$= \frac{1}{(2\pi)^{\frac{r}{2}}\cdot\sqrt{\text{rdet}(\Sigma)}} \cdot e^{-\frac{1}{2}t^TQD^\sim Q^Tt} \overset{z=Q^Tt}{=\!=\!=} \frac{1}{(2\pi)^{\frac{r}{2}}\cdot\sqrt{\text{rdet}(D)}} \cdot e^{-\frac{1}{2}z^TD^\sim z}$$
$$\overset{u=(z_1\ \cdots\ z_r)^T}{=\!=\!=\!=\!=} \frac{1}{(2\pi)^{\frac{r}{2}}\cdot\sqrt{\det(D_h)}} \cdot e^{-\frac{1}{2}u^TD_h^{-1}u} = f^r_{0,D_h}(u).$$

It can be seen that there is a one-to-one correspondence between $x \in \mu + \text{Span}(\Sigma)$ and $u \in \mathbb{R}^r$ described by the following relationships:

$$\begin{cases} u = Q_r^T(x-\mu) \\ x = \mu + Q_r \cdot u \\ f^d_{\mu,\Sigma}(x) = f^r_{0,D_h}(u) \end{cases}.$$

Thus, when $\text{rank}(\Sigma) = r$, $f^d_{\mu,\Sigma}$ is actually equivalent to the r-dimensional normal distribution function $f^r_{0,D_h}(u)$.

Next, we turn to the discrete case. When $x \in \mathbb{Z}^d \cap (\mu + \text{Span}(\Sigma))$, we have $u = Q_r^T(x-\mu) \in \left(Q_r^T\left(\mathbb{Z}^d - \mu\right)\right) \cap \mathbb{R}^r = L(Q_r^T) - Q_r^T\mu$, where $L(Q_r^T)$ refers to the lattice taking Q_r^T as a set of generating vectors [8]. Let $P \in \mathbb{R}^{r\times r}$ be a lattice basis of $L(Q_r^T)$, i.e. $L(Q_r^T) = L(P)$. In particular, since $\mu_h \in \text{Span}(\Sigma_h)^\perp$ is always true after adding

[8] As Q_r^T contains $d(> r)$ r-dimensional vectors, it does not form a lattice basis.

hints, for the sake of simplicity, we will focus only on the case of $\mu \in \text{Span}(\varSigma)^{\perp}$ in the rest of our discussion. For any $z \in \mathbb{R}^d$, let $y = \varSigma z$ and $w = Q^T z$, then we have

$$0 =< \mu, y >= \mu^T Q D Q^T z = \mu^T Q D w = \mu^T Q \begin{pmatrix} \sigma_1 w_1 \\ \vdots \\ \sigma_r w_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mu^T Q_r \begin{pmatrix} \sigma_1 w_1 \\ \vdots \\ \sigma_r w_r \end{pmatrix}.$$

Since Q is orthogonal and z can be arbitrary, we conclude that $\mu^T Q_r = 0$. This implies that $u = Q_r^T x$. So when $x \in \mathbb{Z}^d \cap (\mu + \text{Span}(\varSigma))$, we have $u \in L(Q_r^T) = L(P)$.

We are now ready to compute the (discrete) Fourier transform of the pmf $g_{\mu,\varSigma}^{d,q}$ of the (discrete) multivariate normal distribution over $\mathbb{Z}_q^d$. We combine the above analysis with the Poisson summation formula to deal with this computation regardless of whether the rank of $\varSigma$ is full or not. The specific calculation process is as follows. For any $y \in \mathbb{Z}_q^d$,

$$\begin{aligned}
\widehat{g_{\mu,\varSigma}^{d,q}}(y) &= \sum_{z \in \mathbb{Z}_q^d} e^{\frac{-2\pi i < z, y >}{q}} \cdot g_{\mu,\varSigma}^{d,q}(z) = \sum_{z \in \mathbb{Z}_q^d} e^{\frac{-2\pi i < z, y >}{q}} \cdot \frac{\sum_{t \in \mathbb{Z}^d} f_{\mu,\varSigma}^d (z + tq)}{f_{\mu,\varSigma}^d(\mathbb{Z}^d)} \\
&= \frac{1}{f_{\mu,\varSigma}^d(\mathbb{Z}^d)} \cdot \sum_{z \in \mathbb{Z}_q^d} \sum_{t \in \mathbb{Z}^d} e^{\frac{-2\pi i < z, y >}{q}} \cdot f_{\mu,\varSigma}^d (z + tq) \\
&= \frac{1}{f_{\mu,\varSigma}^d(\mathbb{Z}^d)} \cdot \sum_{z \in \mathbb{Z}_q^d} \sum_{t \in \mathbb{Z}^d} e^{\frac{-2\pi i < z + tq, y >}{q}} \cdot f_{\mu,\varSigma}^d (z + tq) \\
&\xlongequal{x = z + tq} \frac{1}{f_{\mu,\varSigma}^d(\mathbb{Z}^d)} \cdot \sum_{x \in \mathbb{Z}^d} e^{\frac{-2\pi i < x, y >}{q}} \cdot f_{\mu,\varSigma}^d (x) \\
&= \frac{1}{f_{\mu,\varSigma}^d(\mathbb{Z}^d)} \cdot \sum_{x \in \mathbb{Z}^d \cap (\mu + span(\varSigma))} e^{\frac{-2\pi i < x, y >}{q}} \cdot f_{\mu,\varSigma}^d (x) \\
&\xlongequal[x = \mu + Q_r \cdot u]{u = Q_r^T \cdot x} \frac{1}{f_{\mu,\varSigma}^d(\mathbb{Z}^d)} \cdot \sum_{u \in L(P)} e^{-\frac{2\pi i \langle \mu + Q_r u, y \rangle}{q}} \cdot f_{0,D_h}^r (u) \\
&\xlongequal{Prop.2} \frac{e^{-\frac{2\pi i \langle \mu, y \rangle}{q}}}{f_{\mu,\varSigma}^d(\mathbb{Z}^d)} \cdot \sum_{u \in L(P)} e^{\frac{-2\pi i \left\langle u, Q_r^T y \right\rangle}{q}} \cdot \frac{1}{(2\pi)^{\frac{r}{2}} \cdot \sqrt{\det(D_h)}} \cdot \widehat{f_{0,D_h^{-1}}^r} \left(\frac{u}{2\pi} \right) \\
&= \frac{e^{-\frac{2\pi i \langle \mu, y \rangle}{q}}}{f_{\mu,\varSigma}^d(\mathbb{Z}^d)} \cdot \frac{1}{(2\pi)^{\frac{r}{2}} \cdot \sqrt{\det(D_h)}} \cdot \sum_{u \in L(P)} e^{i \left\langle u, -\frac{2\pi Q_r^T y}{q} \right\rangle} \cdot \widehat{f_{0,D_h^{-1}}^r} \left(\frac{u}{2\pi} \right) \\
&\xlongequal{Cor.1} \frac{e^{-\frac{2\pi i \langle \mu, y \rangle}{q}}}{f_{\mu,\varSigma}^d(\mathbb{Z}^d)} \cdot \frac{1}{(2\pi)^{\frac{r}{2}} \cdot \sqrt{\det(D_h)}} \cdot \frac{(2\pi)^r}{\det\left(L(P)\right)} \sum_{k \in L^*(P)} f_{0,D_h^{-1}}^r \left(2\pi k - \frac{2\pi Q_r^T y}{q} \right) \\
&= \frac{e^{-\frac{2\pi i \langle \mu, y \rangle}{q}}}{f_{\mu,\varSigma}^d(\mathbb{Z}^d)} \cdot \frac{1}{\det(L(P))} \sum_{k \in L^*(P)} \frac{(2\pi)^{\frac{r}{2}}}{\sqrt{\det(D_h)}} \cdot f_{0,D_h^{-1}}^r \left(2\pi \left(k - \frac{Q_r^T y}{q} \right) \right) \\
&\xlongequal{Prop.2} \frac{e^{-\frac{2\pi i \langle \mu, y \rangle}{q}}}{f_{\mu,\varSigma}^d(\mathbb{Z}^d)} \cdot \frac{1}{\det(L(P))} \sum_{k \in L^*(P)} \widehat{f_{0,D_h}^r} \left(k - \frac{Q_r^T y}{q} \right) \\
&= \frac{e^{-\frac{2\pi i \langle \mu, y \rangle}{q}}}{f_{\mu,\varSigma}^d(\mathbb{Z}^d)} \cdot \frac{1}{|\det(P)|} \cdot \widehat{f_{0,D_h}^r} \left(L^*(P) - \frac{Q_r^T y}{q} \right)
\end{aligned}$$

While on the other hand,

$$\begin{aligned}
f^d_{\mu,\Sigma}\left(\mathbb{Z}^d\right) \quad &= \quad \sum_{x\in\mathbb{Z}^d} f^d_{\mu,\Sigma}(x) = \sum_{x\in\mathbb{Z}^d\cap(\mu+span(\Sigma))} f^d_{\mu,\Sigma}(x) \\
&\overset{u=Q_r^T\cdot x}{\underset{x=\mu+Q_r\cdot u}{=\!=\!=\!=}} \sum_{u\in L(P)} f^r_{0,D_h}(u) \overset{Prop.2}{=\!=\!=} \sum_{u\in L(P)} \frac{1}{(2\pi)^{\frac{r}{2}}\cdot\sqrt{\det(D_h)}}\cdot \widehat{f^r_{0,D_h^{-1}}}\left(\frac{u}{2\pi}\right) \\
&\overset{Cor.1}{=\!=\!=} \frac{1}{(2\pi)^{\frac{r}{2}}\cdot\sqrt{\det(D_h)}}\cdot\frac{(2\pi)^r}{\det(L(P))}\cdot\sum_{k\in L^*(P)} f^r_{0,D_h^{-1}}(2\pi k) \\
&= \frac{1}{|\det(P)|}\sum_{k\in L^*(P)}\frac{(2\pi)^{\frac{r}{2}}}{\sqrt{\det(D_h)}}\cdot f^r_{0,D_h^{-1}}(2\pi k) \overset{Prop.2}{=\!=\!=} \frac{1}{|\det(P)|}\sum_{k\in L^*(P)}\widehat{f^r_{0,D_h}}(k) \\
&= \frac{1}{|\det(P)|}\cdot\widehat{f^r_{0,D_h}}(L^*(P)).
\end{aligned}$$

To sum up, we get the following expression for $\widehat{g^{d,q}_{\mu,\Sigma}}$:

$$\widehat{g^{d,q}_{\mu,\Sigma}}(y) = e^{-\frac{2\pi i<\mu,y>}{q}}\cdot\frac{\widehat{f^r_{0,D_h}}\left(L^*(P)-\frac{Q_r^Ty}{q}\right)}{\widehat{f^r_{0,D_h}}\left(L^*(P)\right)} = e^{-\frac{2\pi i<\mu,y>}{q}}\cdot\frac{\widehat{f^r_{0,D_h}}\left(L^*\left(Q_r^T\right)-\frac{Q_r^Ty}{q}\right)}{\widehat{f^r_{0,D_h}}\left(L^*\left(Q_r^T\right)\right)},\ \forall y\in\mathbb{Z}_q^d.$$

Next, we shall derive a lower bound for $\left|\widehat{g^{d,q}_{\mu,\Sigma}}(y)\right|$ to estimate the advantage each short vector can bring. Applying proposition 2 to the above expression for $\widehat{g^{d,q}_{\mu,\Sigma}}$ and using the definition of $f^r_{0,D_h^{-1}}$, we have

$$\begin{aligned}
\widehat{g^{d,q}_{\mu,\Sigma}}(y) &= e^{-\frac{2\pi i<\mu,y>}{q}}\cdot\frac{\sum_{k\in L^*(P)}\frac{(2\pi)^{\frac{r}{2}}}{\sqrt{\det(D_h)}}\cdot f^r_{0,D_h^{-1}}\left(2\pi\left(k-\frac{Q_r^T}{q}\right)\right)}{\sum_{k\in L^*(P)}\frac{(2\pi)^{\frac{r}{2}}}{\sqrt{\det(D_h)}}\cdot f^r_{0,D_h^{-1}}(2\pi k)} \\
&= e^{-\frac{2\pi i<\mu,y>}{q}}\cdot\frac{\sum_{k\in L^*(P)} e^{-\frac{1}{2}\left(2\pi k-2\pi\frac{Q_r^Ty}{q}\right)^T\cdot D_h\cdot\left(2\pi k-2\pi\frac{Q_r^Ty}{q}\right)}}{\sum_{k\in L^*(P)} e^{-\frac{1}{2}(2\pi k)^T\cdot D_h\cdot(2\pi k)}} \\
&= e^{-\frac{2\pi i<\mu,y>}{q}}\cdot\frac{\sum_{k\in L^*(P)} e^{-2\pi^2k^TD_hk}\cdot e^{\frac{4\pi^2y^TQ_rD_hk}{q}}\cdot e^{\frac{-2\pi^2y^TQ_rD_hQ_r^Ty}{q^2}}}{\sum_{k\in L^*(P)} e^{-2\pi^2k^TD_hk}}.
\end{aligned}$$

This means that $e^{\frac{2\pi i<\mu,y>}{q}}\cdot\widehat{g^{d,q}_{\mu,\Sigma}}(y)$ is a positive real number which can be bounded below in the following manner

$$\begin{aligned}
&e^{\frac{2\pi i<\mu,y>}{q}}\cdot\widehat{g^{d,q}_{\mu,\Sigma}}(y) = e^{\frac{-2\pi^2y^TQ_rD_hQ_r^Ty}{q^2}}\cdot\frac{\sum_{k\in L^*(P)} e^{-2\pi^2k^TD_hk}\cdot e^{\frac{4\pi^2y^TQ_rD_hk}{q}}}{\sum_{k\in L^*(P)} e^{-2\pi^2k^TD_hk}} \\
&= e^{\frac{-2\pi^2y^TQ_rD_hQ_r^Ty}{q^2}}\cdot\frac{\sum_{k\in L^*(P)} e^{-2\pi^2k^TD_hk}\cdot\frac{\left(e^{\frac{4\pi^2y^TQ_rD_hk}{q}}+e^{\frac{-4\pi^2y^TQ_rD_hk}{q}}\right)}{2}}{\sum_{k\in L^*(P)} e^{-2\pi^2k^TD_hk}} \\
&\geq e^{\frac{-2\pi^2y^TQ_rD_hQ_r^Ty}{q^2}}\cdot\frac{\sum_{k\in L^*(P)} e^{-2\pi^2k^TD_hk}\cdot 1}{\sum_{k\in L^*(P)} e^{-2\pi^2k^TD_hk}} = e^{\frac{-2\pi^2y^TQ_rD_hQ_r^Ty}{q^2}}
\end{aligned}$$

$$= e^{\frac{-2\pi^2 y^T QDQ^T y}{q^2}} = e^{\frac{-2\pi^2 y^T \Sigma y}{q^2}}.$$

What we have discussed actually proves the following theorem.

Theorem 1. *Let d be a positive integer. Let $\Sigma \in \mathbb{R}^{d\times d}$ be a symmetric matrix and $\mu \in Span(\Sigma)^\perp$. Suppose that $rank(\Sigma) = r$ and the EVD of Σ is $\Sigma = QDQ^T$, where $D = diag\{\sigma_1, \cdots, \sigma_r, 0, \cdots, 0\}$, $\sigma_i > 0, i = 1, 2, \cdots, r$. We define $D_h = diag\{\sigma_1, \cdots, \sigma_r\} \in \mathbb{R}^{r\times r}$, and denote the lattice taking Q_r^T as a set of generating vectors by $L(Q_r^T)$. Let $L^*\left(Q_r^T\right)$ be the dual lattice of $L(Q_r^T)$, then*

$$\widehat{g_{\mu,\Sigma}^{d,q}}(y) = e^{-\frac{2\pi i<\mu,y>}{q}} \cdot \frac{\widehat{f_{0,D_h}^r}\left(L^*\left(Q_r^T\right) - \frac{Q_r^T y}{q}\right)}{\widehat{f_{0,D_h}^r}\left(L^*\left(Q_r^T\right)\right)}, \ \forall y \in \mathbb{Z}_q^d.$$

Furthermore, we have

$$e^{\frac{2\pi i<\mu,y>}{q}} \cdot \widehat{g_{\mu,\Sigma}^{d,q}}(y) = \left|\widehat{g_{\mu,\Sigma}^{d,q}}(y)\right| \geq e^{-\frac{2\pi^2 y^T \Sigma y}{q^2}}, \ \forall y \in \mathbb{Z}_q^d.$$

As we mentioned in Section 2.4, it has been proven that a short vector w in $\mathscr{L}$ gives a distinguish advantage $e^{-\frac{2\pi^2\sigma_\chi^2\|w\|^2}{q^2}}$. This can be interpreted by the above theorem as the case where $\Sigma = \sigma_\chi^2 I_d$ and $\mu = 0$ since $w^T \cdot \sigma_\chi^2 I_d \cdot w = \sigma_\chi^2\|w\|^2$. Theorem 1 improves the argument not only to the case of a general nonsingular symmetric Σ, but also to the case where Σ is degenerated. We note that the latter case requires a careful and non-trivial proof and this proof reveals some useful information.

Besides the extension of Σ, μ is also generalized to be an arbitrary (non-zero) vector in Span$(\Sigma)^\perp$. According to theorem 1, in this case, $\widehat{g_{\mu,\Sigma}^{d,q}}(y)$ is a complex number and its argument changes with y. This leads to the need of making some adjustments on the distinguisher, and the details will be discussed in Section 3.3. Fortunately, as we shall see, the lower bound of $\left|\widehat{g_{\mu,\Sigma}^{d,q}}(y)\right|$ given in theorem 1 still indicates the advantage from w can be $e^{\frac{-2\pi^2 w^T\Sigma_h w}{q^2}}$ after adding hints, recall that Σ_h is the covariance matrix of $(S|Y^TS = R)$. This can be made more explicit in the following corollary.

Corollary 2. *Given hints $R = Y^TS$. Let $V = Span(Y)$. For any short vector $w \in \mathscr{L}\backslash\{0\}$ found by BKZ, we decompose it into $w = w_V + w_{V^\perp}$ with $w_V \in V, w_{V^\perp} \in V^\perp$. Then the dual attack advantage from w is $e^{-\frac{2\pi^2\sigma_\chi^2\|w_{V^\perp}\|^2}{q^2}}$.*

Proof. This is simply because

$$w^T\Sigma_h w = w^T \cdot \sigma_\chi^2 \Pi_V^\perp \cdot w = \sigma_\chi^2 \cdot w_{V^\perp}^T \cdot w_{V^\perp} = \sigma_\chi^2\|w_{V^\perp}\|^2. \qquad \square$$

Remark 1. As mentioned in assumption 3, it is generally believed that the short vectors found by BKZ in $\mathscr{L}$ have balanced coefficients, and hence belong to $\mathbb{Z}_q^d$.

Remark 2. Corollary 2 shows the improvement of hints over dual attack. The advantage with respect to w goes from $e^{-\frac{2\pi^2\sigma_\chi^2\|w\|^2}{q^2}}$ to $e^{-\frac{2\pi^2\sigma_\chi^2\|w_{V^\perp}\|^2}{q^2}}$ thanks to hints $R = Y^TS$.

As $w_{V^\perp}$ is a projection of w and $\|w_{V^\perp}\|^2 = \|w\|^2 - \|w_V\|^2$, the advantage is increased by a factor of $e^{\frac{2\pi^2\sigma_\chi^2\|w_V\|^2}{q^2}}$ after adding hints.

Remark 3. Although short vectors returned by BKZ are of similar lengths, the distinguish advantages with respective these vectors are different. As corollary 2 asserts that the lengths of their orthogonal projections onto $V^\perp$ are the determining factors. To be specific, when the length of w is fixed, the closer the direction of w is to V, the greater the distinguish advantage is. In other words, the direction plays some roles when hints exist. On the other hand, it may be a better option for the adversary to shift his/her focus from the length of w to that of $w_{V^\perp}$.

Remark 4. It can be seen in theorem 1 that μ has no effect on $\left|\widehat{g^{d,q}_{\mu,\Sigma}}\right|$. More precisely, the final advantage $e^{-\frac{2\pi^2 w^T\Sigma_h w}{q^2}} = e^{-\frac{2\pi^2\sigma_\chi^2\|w_{V^\perp}\|^2}{q^2}}$ is independent of μ_h. This seems reasonable because of the independence between S_V and $S_{V^\perp}$. In particular, as $S_V = \mu_h$ has already been given by the hints, the unknown part of $\langle S, w\rangle \pmod q$ is actually

$$\langle S_{V^\perp}, w_{V^\perp}\rangle = \langle S, w\rangle - \langle S_V, w\rangle = \langle u, b\rangle - \langle \mu_h, w\rangle \pmod q.$$

This quantity will replace $\langle u, b\rangle$ to construct a new distinguisher in the next subsection.

3.3 A NEW Distinguisher

Now, let us post further explanations about how to build a distinguisher during an actual dual attack with hints. Some details should to be noted when the case is extended to $\mu \neq 0$. Suppose that M short vectors $w_j = \begin{pmatrix} u_j \\ v_j \end{pmatrix} \in \mathscr{L},\ j = 1, 2, \cdots, M$ are used. In the original distinguisher described in Section 2.4, the advantage of each short vector is to add up before taking the absolute value, that is, the adversary will calculate

$$\frac{\sum_{j=1}^M \mathbf{B}\left(f_{\langle S, w_j\rangle}\right)}{M} = \frac{\sum_{j=1}^M \widehat{f_S}(w_j)}{M} = \frac{\sum_{j=1}^M \widehat{g^{d,q}_{\mu_h,\Sigma_h}}(w_j)}{M} = \frac{\sum_{j=1}^M e^{-\frac{2\pi i\langle \mu_h, w_j\rangle}{q}} \cdot \left|\widehat{g^{d,q}_{\mu_h,\Sigma_h}}(w_j)\right|}{M}.$$

However, this is a sum of M complex numbers with different arguments, a lower bound of its absolute value can no longer be obtained by a common lower bound of $\left\{\left|\widehat{g^{d,q}_{\mu_h,\Sigma_h}}(w_j)\right|\right\}_{j=1}^M$. It should be pointed out that this situation does not occur without hints as the secret/error vector is always picked from a distribution with a mean of 0.

In this subsection, we propose a new distinguisher that can handle this problem by some interesting tricks. Its essential idea is to rotate the Fourier transform with respect to each of the short vectors into a real number respectively, this achieves a similar effect as taking the absolute value.

Another thing to notice is that, as mentioned in remark 3, now $w_{V^\perp}$ is the only component that determines the advantage from w. So in the new distinguisher, its length will be used as a criterion for selecting short vectors, instead of w. To be specific, suppose that the adversary uses M short vectors $\{w_j\}_{j=1}^M$ that satisfy $\|(w_j)_{V^\perp}\| \leq l, 1 \leq j \leq$

M, then as M increases, the "rotated" arithmetic mean $\frac{\sum_{j=1}^{M} e^{\frac{2\pi i \langle \mu_h, w_j \rangle}{q}} \cdot e^{\frac{-2\pi i \langle u_j, b \rangle}{q}}}{M}$ becomes closer to

$$\frac{\sum_{j=1}^{M} e^{\frac{2\pi i \langle \mu_h, w_j \rangle}{q}} \cdot \mathbf{B}\left(f_{\langle u_j, b \rangle}\right)(1)}{M} = \frac{\sum_{j=1}^{M} e^{\frac{2\pi i \langle \mu_h, w_j \rangle}{q}} \cdot \widehat{f_{\langle u_j, b \rangle}}(1)}{M}$$

$$= \begin{cases} \frac{\sum_{j=1}^{M} e^{\frac{2\pi i \langle \mu_h, w_j \rangle}{q}} \cdot \widehat{f_b}(u_j)}{M} = 0 & b \leftarrow U(\mathbb{Z}_q^m) \\ \frac{\sum_{j=1}^{M} e^{\frac{2\pi i \langle \mu_h, w_j \rangle}{q}} \cdot \widehat{f_S}(w_j)}{M} \geq \frac{\sum_{j=1}^{M} e^{-\frac{2\pi^2 \sigma_\chi^2 \left\| (w_j)_{V^\perp} \right\|^2}{q^2}}}{M} \geq e^{-\frac{2\pi^2 \sigma_\chi^2 l^2}{q^2}} := \epsilon & b \leftarrow \text{LWE} \end{cases} .$$

In the actual attack, only the real part is considered, that is, the adversary calculates $\frac{\sum_{j=1}^{M} \cos\left(\frac{2\pi(\langle \mu_h, w_j \rangle - \langle u_j, b \rangle)}{q}\right)}{M}$ and checks if it is closer to 0 or ϵ. According to the Chernoff-Hoeffding inequality (lemma 4), if $b \leftarrow U(\mathbb{Z}_q^m)$, then

$$Pr\left[\left|\sum_{j=1}^{M} \cos\left(\frac{2\pi\left(\langle \mu_h, w_j \rangle - \langle u_j, b \rangle\right)}{q}\right)\right| \geq \frac{\epsilon}{2}\right] \leq 2 \cdot e^{\frac{-M\epsilon^2}{8}} .$$

On the other hand, the probability Pr $\left[\sum_{j=1}^{M} \cos\left(\frac{2\pi(\langle \mu_h, w_j \rangle - \langle u_j, b \rangle)}{q}\right) < \frac{\epsilon}{2}\right]$ is even smaller when b is from an LWE instance. Taking $M = O\left(\frac{1}{\epsilon^2}\right)$, we can make $e^{\frac{-M\epsilon^2}{8}}$ a constant, and hence achieve a high success rate of the dual attack. The specific algorithm of this new distinguisher is given in algorithm 1.

```
Algorithm 1: Distinguish
Input: Short vectors $\{w_j\}_{j=1}^{M}$ of length no more than $l$ found by BKZ in $\mathscr{L}$.
Output: 0 for $b \leftarrow U(\mathbb{Z}_q^m)$ and 1 for $b \leftarrow$ LWE.
$sum \leftarrow 0$;
for $j = 1$ to $M$ do
    $sum \leftarrow sum + \cos\left(\frac{2\pi(\langle \mu_h, w_j \rangle - \langle u_j, b \rangle)}{q}\right)$;
if $sum \geq \frac{\epsilon}{2}$ then
    return 1;
else
    return 0;
```

It is known that, in the original dual attack, the value of $\langle S, w \rangle = \langle u, b \rangle \pmod q$ is used to distinguish, as S, w are both short vectors. However, as mentioned in remark 4, after adding hints $R = Y^T S$, the component $\mu_h = S_V$ is available and it is a better estimation of S than 0. To put it more clearly, $S_{V^\perp} = S - \mu_h$ is shorter than S. Hence, a better distinction will be obtained by calculating $\langle S_{V^\perp}, w \rangle = \langle u, b \rangle - \langle \mu_h, w \rangle \pmod q$. As we shall see, that is exactly what algorithm 1 does. The construction of this new distinguisher is very natural.

Remark 5. It has been proven that this distinguisher works for all $w \in \mathscr{L}$ that belong to $\mathbb{Z}_q^d$. This range can be extended further since if $w' - w \in V$, i.e. $w'_{V^\perp} = w_{V^\perp}$, then

$$e^{\frac{2\pi i\langle \mu_h, w\rangle}{q}} \cdot e^{\frac{-2\pi i\langle S, w\rangle}{q}} = e^{-\frac{2\pi i\left\langle S, w_{V^\perp}\right\rangle}{q}} = e^{-\frac{2\pi i\left\langle S, w'_{V^\perp}\right\rangle}{q}} = e^{\frac{2\pi i\langle \mu_h, w'\rangle}{q}} \cdot e^{\frac{-2\pi i\left\langle S, w'\right\rangle}{q}}.$$

This means that two vectors have the same orthogonal projection onto $V^\perp$ can give the same distinguish advantage. This idea will be used later in Section 4.2.

3.4 The Effect of the New Distinguisher

In the previous subsection, we have established a new distinguisher for the dual attack with hints. Given t hints $R = Y^T S$ and a short vector w in $\mathscr{L}$, an advantage $e^{\frac{-2\pi^2\sigma_\chi^2\left\|w_{V^\perp}\right\|^2}{q^2}}$ can be achieved, where $V = \text{Span}(Y)$ and $w_{V^\perp}$ is the orthogonal projection of w onto $V^\perp$. As it was remarked, this is an $e^{\frac{2\pi^2\sigma_\chi^2\|w_V\|^2}{q^2}}$-times increase compared to the distinguish advantage without any hint.

Now, let us show the difference in the advantages of the same vector before and after the integration of hints. Let m^*, β^* represent the optimal number of samples and the optimal BKZ blocksize respectively when finding w in $\mathscr{L}$ without hints. To get vectors of the same length for comparison, we still use m^* and β^* when hints exist[9]. Following the balance assumption of BKZ, we can estimate the improvement as follows:

$$e^{-\frac{2\pi^2\sigma_\chi^2\|w_{V^\perp}\|^2}{q^2}} \approx e^{-\frac{2\pi^2\sigma_\chi^2\left(\frac{d-t}{d}\right)\|w\|^2}{q^2}} = \left(e^{-\frac{2\pi^2\sigma_\chi^2\|w\|^2}{q^2}}\right)^{\frac{d-t}{d}}.$$

I.e. it is approximately the original value to the power of $\frac{d-t}{d}$. This is used in table 1 to display the distinguish advantages for several LWE schemes, with numbers of hints being $0, 50, 100, 150, 200, 300$ and 400.

Table 1: The effects of different numbers of hints on the enhancement of advantages.

schemes	m^*	β^*	0	50	100	150	200	300	400
Newhope512	569	382.67	$2^{-39.7}$	$2^{-37.9}$	$2^{-36.0}$	$2^{-34.2}$	$2^{-32.4}$	$2^{-28.7}$	$2^{-25.0}$
Kyber768	690	619.80	$2^{-64.3}$	$2^{-62.1}$	$2^{-59.9}$	$2^{-57.7}$	$2^{-55.5}$	$2^{-51.1}$	$2^{-46.7}$
FireSaber	944	819.20	$2^{-85.0}$	$2^{-82.8}$	$2^{-80.7}$	$2^{-78.5}$	$2^{-76.4}$	$2^{-72.0}$	$2^{-67.7}$
Frodo1344	1275	924.91	$2^{-96.0}$	$2^{-94.1}$	$2^{-92.3}$	$2^{-90.5}$	$2^{-88.6}$	$2^{-85.0}$	$2^{-81.3}$

As we can see from table 1, hints produce a large increase in the advantages. That is to say, if the dual attack is still executed in the original lattice $\mathscr{L}$, with multiple hints, each short vector w can bring a significantly improved advantage at the same search cost. However, a more efficient new model of the dual attack with hints will be presented in the next section.

[9] In fact, with hints, the optimal number of samples and blocksize m_h^*, β_h^* will change after rebalancing. But to make the comparison more intuitive, we still use the same sample number and blocksize. The discussion of calculating m_h^*, β_h^* will be given in the next section.

4 A New Model of the Dual Attack with Hints

From Table 1, hints can indeed increase the advantage from each vector, thus reducing the number of vectors needed as well as the blocksize β required. However, on the other hand, for a smaller β, an increase in the length of the short vector found inevitably leads to a decrease in its advantage. Unfortunately, in general, the positive effect of hints on the advantage is relatively small compared to the negative impact of the decrease in β. Moreover, because of the near-orthogonality of vectors in high-dimensional spaces, very few of the short vectors found in $\mathscr{L}$ are close to V to provide high advantages. In conclusion, if we still perform the dual attack in the original lattice $\mathscr{L}$, β cannot be drastically reduced by adding hints, so neither can the attack cost.

In fact, there are two fundamental reasons for the limited cost reduction. The first one is that we do not change the lattice for finding short vectors, which has a high dimension and its volume is not small. So the cost of performing BKZ in it cannot be low. The other cause is that we do not have a limit on the projection lengths of short vectors, and the vectors with high advantages cannot be screened out. For these reasons, in the following, we shall suggest a new model of dual attack with hints that can settle both problems simultaneously and further reduce the complexity. More specifically, firstly, as described in Section 4.1, short vectors will be found in the projected lattice. Then, we recover the corresponding vectors in $\mathscr{L}$ by the method given in Section 4.2. Some transformations in theoretical analysis are performed to extend the input range of algorithm 1 to fit these recovered vectors. Finally, a complete dual attack process and its complexity analysis are summarized in Section 4.3.

4.1 Searching Short Vectors in a New Lattice

As mentioned earlier, for each w, $w_{V^\perp}$ is the only component that determines the advantage. If we still look for w in $\mathscr{L}$, not only is the attack expensive, but we can only control the length of w, not that of $w_{V^\perp}$. One key point is that, it may be more efficient to search $w_{V^\perp}$ directly in the projected lattice by BKZ to make sure $w_{V^\perp}$ is as short as possible, and then find a $w \in \mathscr{L}$ that satisfies $\Pi_V^\perp \cdot w = w_{V^\perp}$. Let us describe the process in detail.

After obtaining t hints $R = Y^T S$, we construct a new lattice as follows:

$$\mathscr{L}_{V^\perp} := \Pi_V^\perp \cdot \mathscr{L}.$$

It is easy to see that $\mathscr{L}_{V^\perp}$ is made up of the projections of the lattice vectors in $\mathscr{L}$ onto $V^\perp$. Compared to the original lattice, its dimension is reduced by t. Using the idea described in [15], a lattice basis $\mathscr{B}_{V^\perp}$ of $\mathscr{L}_{V^\perp}$ can be easily calculated by the LLL algorithm, or more precisely, the MLLL algorithm, which will be discussed later in Section 4.2. The specific steps are as follows:

1. Compute $\mathscr{B}' := \Pi_V^\perp \cdot \mathscr{B}$, where $\mathscr{B}$ is a basis of $\mathscr{L}$ given in Section 2.4.
2. Apply the MLLL algorithm on $\mathscr{B}'$ to eliminate linear dependencies. Delete t zero vectors in the output result and the remaining $d - t$ vectors form $\mathscr{B}_{V^\perp}$.

Given the new basis, the volume of the new lattice can be obtained directly since $\mathrm{vol}\left(\mathscr{L}_{V^\perp}\right) = |\det(\mathscr{B}_{V^\perp})|$. But in fact, a proper estimation of $\mathrm{vol}\left(\mathscr{L}_{V^\perp}\right)$ can be acquired with an overwhelming probability by a mathematical means even without calculating $\mathscr{B}_{V^\perp}$. We shall use a slight extension of the Fact 14 of [15] to predict the volume of the projected lattice.

Proposition 3. *Let L be a d-dimensional lattice and $X = (v_1\ v_2\ \cdots v_t)$ contains linearly independent vectors of $\mathbb{R}^d$. Suppose that $\widetilde{X} = (a_1v_1\ a_2v_2\ \cdots\ a_tv_t)$ forms a primitive set*[10] *of vectors of L, $a_i \in \mathbb{R}^*, i = 1, 2, \cdots t$, then $L' := \Pi_X^\perp \cdot L$ is a $(d-t)$-dimensional lattice and $vol(L') = \frac{vol(L)}{\sqrt{\det(\widetilde{X}^T\widetilde{X})}} = \frac{vol(L)}{|a_1\cdots a_t|\cdot\sqrt{\det(X^TX)}}$.*

We provide a straightforward and elementary proof of this result in appendix C. As can be seen from proposition 3, if L is an integer lattice, then $\widetilde{X} \in \mathbb{Z}^{d\times t}$ and $\sqrt{\det\left(\widetilde{X}^T\widetilde{X}\right)} \geq 1$. This implies that the projection lattice must have a smaller volume. On the other hand, we can find the primitiveness of the hint description matrix Y with respect to $\mathscr{L}$ is a determining factor in $\mathrm{vol}\left(\mathscr{L}_{V^\perp}\right)$. According to the definition and properties of the dual lattice, it is easy to verify that $\mathscr{B}^{-T} = \begin{pmatrix} I_m & -\frac{A}{q} \\ O_{n\times m} & \frac{1}{q}I_n \end{pmatrix}$ is a basis of $\mathscr{L}^*$. We decompose Y into $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$, where $Y_1 \in \mathbb{Z}^{m\times t}, Y_2 \in \mathbb{Z}^{n\times t}$, then

$$Y^T\mathscr{B}^{-T} = (Y_1^T\ \ Y_2^T)\begin{pmatrix} I_m & -\frac{A}{q} \\ O_{n\times m} & \frac{1}{q}I_n \end{pmatrix} = \begin{pmatrix} Y_1^T & \dfrac{-Y_1^TA + Y_2^T}{q} \end{pmatrix}.$$

It is easy to see that for any hint description matrix $Y \in \mathbb{Z}^{d\times t}$, qY consists of t linearly independent vectors of $\mathscr{L}$ as $qY^T\mathscr{B}^{-T} \in \mathbb{Z}^d$. Moreover, for any vector $v \in \mathscr{L} \cap \mathrm{Span}(Y)$, suppose that $v = Y\alpha$ where $\alpha \in \mathbb{R}^t$, then we have

$$v^T\mathscr{B}^{-T} = \alpha^T \cdot Y^T\mathscr{B}^{-T} = \alpha^T\begin{pmatrix} Y_1^T & \dfrac{-Y_1^TA + Y_2^T}{q} \end{pmatrix} \in \mathbb{Z}^d.$$

Hence $(Y_2 - A^TY_1)\cdot\alpha \in q\mathbb{Z}^n$. A large number of experiments show that $\alpha \in q\mathbb{Z}^t$ is always true when taking $Y \leftarrow U(\mathbb{Z}^{d\times t})$. This leads to the fact that qY is a set of primitive vectors of $\mathscr{L}$. Then, from proposition 3, the volume of the new lattice is

$$\mathrm{vol}\left(\mathscr{L}_{V^\perp}\right) = \frac{\mathrm{vol}(\mathscr{L})}{q^t\cdot\sqrt{\det(Y^TY)}} = \frac{q^{n-t}}{\sqrt{\det(Y^TY)}}. \tag{1}$$

Admittedly, there are a few exceptions. Equation (1) may not be true when the greatest common divisor of all entries of $(Y_2 - A^TY_1)$ is not 1. But this only happens with a very small probability. The above analysis is summarized as the following heuristic.

[10] Recall that a primitive set T of lattice vectors is the one that can be extended to a lattice basis of L, namely, $L \cap \mathrm{Span}(T) = L(T)$.

Heuristic 2 *Given hints $R = Y^T S$ with hint description matrix $Y \in \mathbb{Z}^{d\times t}$, the volume of $\mathscr{L}_{V^\perp}$ is vol $(\mathscr{L}_{V^\perp}) = \frac{q^{n-t}}{\sqrt{\det(Y^TY)}}$ with an overwhelming probability.*

Given the dimension and volume of $\mathscr{L}_{V^\perp}$, the cost of the search phase can be calculated. Specific complexity analysis is given in Section 4.3. It is worth noting that, vol $(\mathscr{L}_{V^\perp}) < \text{vol}(\mathscr{L})$ is always true, whether equation (1) is true or not. This is advantageous to the adversary in a dual attack. We give an explanation in the following.

Remark 6. The hints we are addressing here are of the type "perfect hints" in [15], but the way of integrating them is like that of the "short vector hints" in [15]. In fact, for a short target vector ξ in some integer lattice L, if the attacker gets a hint $r = \langle \xi, v\rangle$, then he/she has two ways of optimizing the attack:

Way 1: Let $\overline{\xi} = \begin{pmatrix}\xi\\1\end{pmatrix}, \overline{v} = \begin{pmatrix}v\\-r\end{pmatrix}$, then $\langle \overline{\xi}, \overline{v}\rangle = 0$, i.e. $\overline{\xi} \in \text{Span}(\overline{v})^\perp$. Let $\overline{L}$ be a properly selected lattice containing $\overline{\xi}$, as the Kannan's embedding is used to construct $\overline{L}$ in [15], then we can search $\overline{\xi}$ in the new lattice $\overline{L} \cap \text{Span}(\overline{v})^\perp$.

Way 2: Actually $r = \langle \xi, v\rangle$ gives the orthogonal projection of ξ onto Span(v):

$$\xi_v := \frac{\langle \xi, v\rangle}{\|v\|^2} \cdot v = \frac{r}{\|v\|^2} v,$$

we can just convert the target vector to the rest of the unknown component of ξ:

$$\xi_{v^\perp} = \xi - \xi_v,$$

i.e., the orthogonal projection of ξ onto Span$(v)^\perp$. It will be looked for in the new lattice $\Pi^\perp_{Span(v)} \cdot L$. (Or similarly, the projection components of its dual vectors are searched in $\Pi^\perp_{Span(v)} \cdot L^*$.)

In general, as can be seen in [15], the first way usually increases the volume of the lattice, i.e. vol $\left(\overline{L} \cap \text{Span}(\overline{v})^\perp\right) > \text{vol}(\overline{L})$, while the second method always results in a reduction in the volume of the lattice, that is, vol $\left(\Pi^\perp_{Span(v)} \cdot L\right) < \text{vol}(L)$. The reason the first way is more suitable for a primal attack is that, the increase in the volume of the lattice makes the primal attack easier. To be specific, as the length of the target vector $\begin{pmatrix}e\\s\\1\end{pmatrix}$ is fixed, the longer the other vectors in the lattice, the easier it is to find the target vector[11]. But instead, in a dual attack, the short vectors found in $\mathscr{L}$ (or $\mathscr{L}_{V^\perp}$) will be used to make a distinction by calculating the inner product. The shorter the lattice vectors, the smaller the inner product, the more effective the attack. This explains why it is better to use the first approach in the primal attack, but here, we tend to choose the second one to improve the dual attack.

4.2 Recovering w from $w_{V^\perp}$ to Make a Distinction

As the attacker calculates $\langle u, b\rangle - \langle \mu_h, w\rangle$ to distinguish, the process of recovering w from $w_{V^\perp}$ is necessary. Although theorem 1 and corollary 2 only work for $w \in \mathbb{Z}_q^d$,

[11] As reported in [23,31], the λ_i-gap $\frac{\lambda_i}{\lambda_1}$ among the successive minima of a lattice especially its λ_2-gap often provides more efficient SVP search algorithms.

from some ideas in remark 5, under certain assumption, it is enough to restore any $w \in \mathscr{L}$ that satisfies $\varPi_V^\perp \cdot w = w_{V^\perp}$ [12]. In fact, an idea of recovering such a w is already included in the steps of computing $\mathscr{B}_{V^\perp}$.

The LLL algorithm was proposed by Lenstra et al. [20] in 1982, to reduce the input lattice basis to another one with better orthogonality. One of the disadvantages of LLL is that the input must consist of linearly independent vectors. In 1987, Pohst [28] overcome this limitation. A modification of the LLL algorithm named the MLLL algorithm was given, whose input range was extended to a set of spanning vectors of the lattice. We want to point out that, in addition to its major contribution in terms of eliminating linear dependence, there is also a "relation matrix" H in the algorithm given in [28] (also can be seen in [14, Section 2.6.4]), which records how the output basis is represented by the original generating vectors.

Let us go back to the dual attack with t hints $R = Y^T S$. Let $\alpha_1, \cdots, \alpha_d$ be the d column vectors of $\mathscr{B}$. Recall that $V = \text{Span}(Y)$ and we denote $\gamma_i = \varPi_V^\perp \cdot \alpha_i,\ i = 1, 2, \cdots, d$. It is easy to see that $\{\gamma_i\}_{i=1}^d$ is a set of spanning vectors of $\mathscr{L}_{V^\perp}$. Given this set as input to MLLL, it will output an LLL-reduced basis $\zeta_1, \cdots, \zeta_{d-t}$ of $\mathscr{L}_{V^\perp}$ and a relation matrix $H = (h_{ij}) \in \mathbb{Z}^{d\times(d-t)}$, such that

$$\zeta_j = \sum_{i=1}^{d} h_{ij}\gamma_i,\ j = 1, 2, \cdots, d-t.$$

These $d - t$ equations are the key to recovering w. The recovery process is divided into the following two steps.

· Step 1: Find $z_j \in \mathbb{Z},\ j = 1, 2, \cdots, d-t$, such that $w_{V^\perp} = \sum_{j=1}^{d-t} z_j\zeta_j$.

By applying the Gram-Schmidt orthogonalization to $\zeta_1, \cdots, \zeta_{d-t}$, we could get

$$\zeta_1^*, \cdots, \zeta_{d-t}^* \text{ and } \mu_{j,k} = \frac{\langle \zeta_j, \zeta_k^* \rangle}{\|\zeta_k^*\|^2},\ j = 1, 2, \cdots, d-t;\ k = 1, 2, \cdots, j-1.$$

We represent $w_{V^\perp}$ as a linear combination of $\zeta_1^*, \cdots, \zeta_{d-t}^*$, i.e.

$$w_{V^\perp} = \sum_{j=1}^{d-t} c_j\zeta_j^*, \text{ where } c_j = \frac{\langle w_{V^\perp}, \zeta_j^* \rangle}{\|\zeta_j^*\|^2},\ j = 1, 2, \cdots, d-t.$$

These $\{c_j\}_{j=1}^{d-t}$ will be used to obtain $\{z_j\}_{j=1}^{d-t}$. On the one hand, for $g \in \{1, \cdots, d-t\}$,

$$\langle w_{V^\perp}, \zeta_g^* \rangle = \left\langle \sum_{j=1}^{d-t} c_j\zeta_j^*, \zeta_g^* \right\rangle = c_g \left\|\zeta_g^*\right\|^2.$$

[12] Actually, there are infinitely many vectors in $\mathscr{L}$ that satisfy this condition. We just need to get any one of them, no matter how long it is, the advantage from it is always $e^{\frac{-2\pi^2\sigma_\chi^2\left\|w_{V^\perp}\right\|^2}{q^2}}$. Moreover, according to remark 9, only one of them can be used, otherwise a high correlation will be raised.

On the other hand,

$$\begin{aligned}\langle w_{V^\perp}, \zeta_g^* \rangle = & \left\langle \sum_{j=1}^{d-t} z_j \zeta_j, \zeta_g^* \right\rangle = \sum_{j=g}^{d-t} z_j \left\langle \zeta_j, \zeta_g^* \right\rangle = \sum_{j=g}^{d-t} z_j \left\langle \zeta_j^* + \sum_{k=1}^{j-1} \mu_{j,k} \zeta_k^*, \zeta_g^* \right\rangle \\ = & \; z_g \left\| \zeta_g^* \right\|^2 + \sum_{j=g+1}^{d-t} z_j \mu_{j,g} \left\| \zeta_g^* \right\|^2 = \left(z_g + \sum_{j=g+1}^{d-t} z_j \mu_{j,g} \right) \left\| \zeta_g^* \right\|^2 . \end{aligned} \tag{2}$$

By comparing the above equations, we have

$$c_g = z_g + \sum_{j=g+1}^{d-t} z_j \mu_{j,g}, \; g = 1, 2, \cdots, d-t.$$

Hence, we could calculate $z_{d-t}, z_{d-t-1}, \cdots, z_1$ in sequence:

$$z_g = c_g - \sum_{j=g+1}^{d-t} \mu_{j,g} \cdot z_j, \; g = d-t, d-t-1, \cdots, 1.$$

· Step2 : Recover w from $w_{V^\perp}$.

Now we can represent $w_{V^\perp}$ using $\{\gamma_i\}_{i=1}^d$ as

$$w_{V^\perp} = \sum_{j=1}^{d-t} z_j \zeta_j = \sum_{j=1}^{d-t} z_j \left(\sum_{i=1}^{d} h_{ij} \gamma_i \right) = \sum_{i=1}^{d} \sum_{j=1}^{d-t} z_j h_{ij} \gamma_i.$$

It obviously corresponds to a vector $w \in \mathscr{L}$ as follows:

$$w = \sum_{i=1}^{d} k_i \alpha_i, \quad \text{where } k_i = \sum_{j=1}^{d-t} z_j h_{ij}.$$

Remark 7. It is easy to see that the MLLL algorithm to $\gamma_1, \cdots, \gamma_d$ and the Gram-Schmidt orthogonalization procedure on $\zeta_1, \cdots, \zeta_{d-t}$ do not need to be repeated for each $w_{V^\perp}$. In fact, as can be seen in [28], the MLLL algorithm keeps track of the Gram-Schmidt orthonormal basis, its length and the projection coefficients $\{\mu_{j,k}\}$ of the current lattice basis during the running time. As a result, let $B_j = \|\zeta_j^*\|^2$ and $\mu_{j,k} = \frac{\langle \zeta_j, \zeta_k^* \rangle}{B_k}$, with minor modifications, we could make MLLL directly output

$$\zeta_j^*, \; B_j, \; \mu_{j,k}, \; j = 1, 2, \cdots, d-t; \; k = 1, 2, \cdots, j-1,$$

as well as $\zeta_1, \cdots, \zeta_{d-t}$ instead of performing the Gram-Schmidt orthogonalization procedure again. However, the $\{c_j\}_{j=1}^{d-t}$ and $\{z_j\}_{j=1}^{d-t}$ corresponding to each $w_{V^\perp}$ need to be solved one by one.

Using the method above, we could attach each $w_{V^\perp}$ to a corresponding $w \in \mathscr{L}$ that satisfies $\mathit{\Pi}_V^\perp \cdot w = w_{V^\perp}$. However, since these w's may not belong to $\mathbb{Z}_q^d$, a further transformation is necessary in theoretical analysis. In fact, our distinguish process relies on the following assumption.

Assumption 4 *For each short vector $w_{V^\perp}$ found by BKZ in $\mathscr{L}_{V^\perp}$, there exists a $w' \in \mathbb{Z}_q^d$, such that $w'_{V^\perp} = w_{V^\perp}$.*

We remark that this is a reasonable assumption as explained in the following. From assumption 3, for each vector $w_{V^\perp}$ found by BKZ, its coefficients are balanced, and it will correspond to a w' also of short length, such that $\Pi_V^\perp \cdot w' = w_{V^\perp}$ and $\|w'\| \approx \sqrt{\frac{d}{d-t}} \cdot \|w_{V^\perp}\|$. Since in the dual attack against actual schemes, $\sqrt{\frac{d}{d-t}} \cdot \|w_{V^\perp}\| \ll \sqrt{d}q$ is always true, the probability of $w' \in \mathbb{Z}_q^d$ is extremely high. One can think that each entry of w' follows a normal distribution with mean 0 and standard deviation $\frac{\|w_{V^\perp}\|}{\sqrt{d-t}}$, and then estimate the probability

$$\Pr\left[w' \in \mathbb{Z}_q^d\right] = \left(\text{erf}\left(\frac{\frac{q}{2}}{\sqrt{2} \cdot \frac{\|w_{V^\perp}\|}{\sqrt{d-t}}}\right)\right)^d, \tag{3}$$

where erf is the error function $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy$. Experiments show that this probability is always 1 in all schemes in Section 5.

Under assumption 4, each $w \in \mathscr{L}$ recovered by the above method can still be used in algorithm 1 (even though it may not belong to $\mathbb{Z}_q^d$). This is because

$$e^{\frac{2\pi i \langle \mu_h, w \rangle}{q}} \cdot e^{-\frac{2\pi i \langle S, w \rangle}{q}} = e^{-\frac{2\pi i \left\langle S_{V^\perp}, w \right\rangle}{q}} = e^{-\frac{2\pi i \left\langle S, w_{V^\perp} \right\rangle}{q}} = e^{-\frac{2\pi i \left\langle S, w'_{V^\perp} \right\rangle}{q}} = e^{\frac{2\pi i \left\langle \mu_h, w' \right\rangle}{q}} \cdot e^{-\frac{2\pi i \left\langle S, w' \right\rangle}{q}}.$$

Suppose that M short vectors $\{(w_j)_{V^\perp}\}_{j=1}^M$ that satisfy $\|(w_j)_{V^\perp}\| \leq l, 1 \leq j \leq M$ are found. As discussed above, the MLLL recovers $\{w_j\}_{j=1}^M \subseteq \mathscr{L}$ and there exist $\{w'_j\}_{j=1}^M \subseteq \mathbb{Z}_q^d$ such that $(w'_j)_{V^\perp} = (w_j)_{V^\perp}, 1 \leq j \leq M$. The attacker calculates $\frac{\sum_{j=1}^M e^{\frac{2\pi i \langle \mu_h, w_j \rangle}{q}} \cdot e^{-\frac{2\pi i \langle u_j, b \rangle}{q}}}{M}$ and it is easy to see that it becomes closer to 0 as M increases when b is uniformly random. While on the other hand, if b is from an LWE instance, the value becomes $\frac{\sum_{j=1}^M e^{\frac{2\pi i \left\langle \mu_h, w'_j \right\rangle}{q}} \cdot e^{-\frac{2\pi i \left\langle S, w'_j \right\rangle}{q}}}{M}$. Since theorem 1 and corollary 2 apply to $w'_j \in \mathbb{Z}_q^d, j = 1, 2, \cdots, M$, we know the above value is getting closer to

$$\frac{\sum_{j=1}^M e^{\frac{2\pi i \left\langle \mu_h, w'_j \right\rangle}{q}} \cdot \widehat{f_S}(w'_j)}{M} \geq \frac{\sum_{j=1}^M e^{-\frac{2\pi^2 \sigma_\chi^2 \left\| \left(w'_j\right)_{V^\perp} \right\|^2}{q^2}}}{M} = \frac{\sum_{j=1}^M e^{-\frac{2\pi^2 \sigma_\chi^2 \left\| \left(w_j\right)_{V^\perp} \right\|^2}{q^2}}}{M} \geq e^{-\frac{2\pi^2 \sigma_\chi^2 l^2}{q^2}}$$

as M increases. Therefore, the distinguish procedure can be performed as before.

Furthermore, the processes of recovery and distinguish can be done together. More specifically, for each $(w_j)_{V^\perp}$, the attacker recovers the corresponding w_j, computes the value of $\cos\left(\frac{2\pi(\langle \mu, w_j \rangle - \langle u_j, b \rangle)}{q}\right)$ and adds it to the previous sum just like algorithm 1.

Remark 8. Assumption 4 is only used for the theoretical analysis. The adversary does not need to figure out what exactly $\left\{w'_j\right\}_{j=1}^M$ are when performing an actual attack.

Remark 9. As we know, the independence among random variables is a requirement of Chernoff-Hoeffding inequality. In fact, in the original case, the independence among

$\{\langle S, w_j\rangle\}_{j=1}^{M}$ comes from $\langle w_j, w_k\rangle \approx 0$ for any $1 \leq j < k \leq M$. This approximation is based on two assumptions – the balance assumption of BKZ and the near-orthogonality assumption of high-dimensional spaces. After adding hints, the independence becomes dependent on $\langle (w_j)_{V^\perp}, (w_k)_{V^\perp}\rangle \approx 0$. It should be pointed out that $\{(w_j)_{V^\perp}\}_{j=1}^{M}$ are exactly the outputs of BKZ now, and they still belong to a high-dimensional space (although the dimension goes down by t). To sum up, the independence before or after the integration of hints relies on the same assumptions under different dimensions. Our algorithm may results in a (negligible) loss of independence.

4.3 Summary and the Complexity Analysis

Let us start with a summary. In our new model, we divided the process of dual attack with hints into three parts: search, recovery and distinguish. First, we apply the way given in Section 4.1 to use BKZ to find enough short vectors $w_{V^\perp}$ in $\mathscr{L}_{V^\perp}$. After performing the MLLL algorithm of $\gamma_1, \cdots, \gamma_d$ whose complexity is much lower than BKZ, the recovery&distinguish process for each $w_{V^\perp}$ will be carried out one by one.

Now let us consider the cost of the dual attack with hints under this new model. In the following, we shall give a discussion on the choice of the optimal number of samples m_h^*, the optimal BKZ blocksize β_h^* to search $w_{V^\perp}$, and a total cost model of the dual attack with t hints $R = Y^T S$ under heuristic 2. Suppose that the adversary uses m samples in a dual attack and he/she applies BKZ-β on $\mathscr{L}_{V^\perp}$, then the length of each short vector found is $l(d,\beta) = \frac{\delta_0^{d-t}(\beta)\cdot q^{\frac{n-t}{d-t}}}{(\det(Y^TY))^{\frac{1}{2(d-t)}}}$, where $d = m + n$, thereby bringing an advantage $\epsilon(d,\beta) = e^{-\frac{2\pi^2\sigma_\chi^2 l(d,\beta)^2}{q^2}}$.

To reach a constant success rate, from lemma 4, the attacker needs $O\left(\frac{1}{\epsilon^2(d,\beta)}\right)$ short vectors in $\mathscr{L}_{V^\perp}$. Thus, according to assumption 1, the process of BKZ has to be repeated at least $R(d,\beta)$ times, where $R(d,\beta) = \max\left\{1, \frac{1}{\epsilon^2(d,\beta)\cdot 2^{0.2075\beta}}\right\}$. Then the cost of the search phase is

$$T_s(d,\beta) = TBKZ(\beta) \cdot R(d,\beta).$$

Now the attacker needs to perform the recovery&distinguish procedure on $R(d,\beta)\cdot 2^{0.2075\beta}$ short vectors in $\mathscr{L}_{V^\perp}$ one by one [13]. Since most of the variables involved in this process belong to $\mathbb{R}$, we will use "flop" which could denote one addition, subtraction, multiplication or division of floating point numbers to express the complexity of this process. For each $w_{V^\perp}$, the complexity of each step in this process is as follows:

- calculate $\{c_j\}_{g=1}^{d-t}$: As $w_{V^\perp}, \zeta_j^* \in \mathbb{R}^d$, d multiplications and $d-1$ additions are required for each inner product $\langle w_{V^\perp}, \zeta_j^*\rangle$. Hence, calculating all $\{c_j\}_{g=1}^{d-t}$ costs $2d(d-t)$ flops.
- calculate $\{z_g\}_{g=1}^{d-t}$: $d-t-g$ multiplications, $d-t-g$ additions and one subtraction are required for calculating z_g. Then $\{z_g\}_{g=1}^{d-t}$ take $\frac{(d-t-1)(d-t)}{2} \times 2 + (d-t) = (d-t)^2$ flops in total.

[13] Since the calculations of different $w_{V^\perp}$'s is independent, it is easy to implement this process in parallel for multiple $w_{V^\perp}$'s. We only consider the time complexity of the non-parallel implementation here.

- calculate w: Each k_i needs $d-t$ multiplications and $d-t$ additions, moreover, $d-1$ additions between d-dimensional vectors and d d-dimensional scalar multiplication are required to obtain w. Therefore, the total number of flops spent in this step is $2(d-t)\cdot d + (d-1)\cdot d + d\cdot d = 4d^2 - 2dt - d$.
- calculate sum: Besides two inner products, one subtraction, one multiplication, one division, and one addition, we also need a cosine operation. In practice, a cosine operation is not much slower than a division operation, so we can assume that a cosine operation costs c_0 flops, where c_0 is a small constant. Thus, about $O(d)$ flops are required in this step, which is much lower than other steps.

To sum up, only considering the second-order terms, we can roughly get the total time complexity of the recovery&distinguish processes for all short vectors is

$$T_{r\&d}(d,\beta) = R(d,\beta)\cdot 2^{0.2075\beta}\cdot(7d^2 - 6dt + t^2).$$

As the entries of the basis of the projected lattice are generally not integers, we can also view the complexity of the search phase as the number of floating point operations needed. Therefore, the time complexity of the whole dual attack is

$$T_h(d,\beta) = \begin{cases} R(d,\beta)\cdot\left(2^{0.292\beta} + 2^{0.2075\beta}\cdot\left(7d^2 - 6dt + t^2\right)\right) & \text{classical case} \\ R(d,\beta)\cdot\left(2^{0.265\beta} + 2^{0.2075\beta}\cdot\left(7d^2 - 6dt + t^2\right)\right) & \text{quantum case} \end{cases}.$$

In the following, we mainly consider the case of $t < \min\{m,n\}$, i.e. the number of hints does not exceed the numbers of the entries of s and e [14]. Actually, under this condition, an important fact from our experiments is that, for the parameters in the actual schemes,

$$T_h(d,\beta) \approx T_s(d,\beta) > T_{r\&d}(d,\beta)$$

is always true in both the classical case and the quantum case. With the assumption $T_h(d,\beta) = T_s(d,\beta)$, we can further ease the calculation of the attack cost.

It has been proven by [21,25] that in the original dual attack (i.e. t=0), the optimal dimension d^* of the dual attack is a function of β, to be specific, $d^* = \sqrt{\frac{n\cdot\ln(q)}{\ln(\delta_0(\beta))}}$. We find that after adding t hints $R = Y^T S$, a nice relation between the optimal dimension d_h^* and β is still available:

$$d_h^*(\beta) = \sqrt{\frac{\ln\left(\frac{q^{n-t}}{\sqrt{\det(Y^TY)}}\right)}{\ln(\delta_0(\beta))}} + t. \tag{4}$$

In fact, (4) is the only zero of the derivative of $l(d,\beta)$ with respect to d :

$$\frac{\partial l(d,\beta)}{\partial d} = \frac{\partial\left[\delta_0^{d-t}\cdot\left(\frac{q^{n-t}}{\sqrt{\det(Y^TY)}}\right)^{\frac{1}{d-t}}\right]}{\partial d}$$

$$= \delta_0^{d-t}\cdot\ln(\delta_0)\cdot\left(\frac{q^{n-t}}{\sqrt{\det(Y^TY)}}\right)^{\frac{1}{d-t}} + \delta_0^{d-t}\cdot\left(\frac{q^{n-t}}{\sqrt{\det(Y^TY)}}\right)^{\frac{1}{d-t}}\cdot\ln\left(\frac{q^{n-t}}{\sqrt{\det(Y^TY)}}\right)\cdot\frac{-1}{(d-t)^2}$$

[14] This seems reasonable as it is hard to get so many hint. On the other hand, it may not make sense to analyze security in terms of dual attack when $t \geq m$ or $t \geq n$.

$$= \delta_0^{d-t} \cdot \left(\frac{q^{n-t}}{\sqrt{\det(Y^T Y)}} \right)^{\frac{1}{d-t}} \cdot \left[\ln(\delta_0) - \frac{\ln\left(\frac{q^{n-t}}{\sqrt{\det(Y^T Y)}} \right)}{(d-t)^2} \right].$$

Hence, the adversary just needs to search for the optimal β_h^*, such that

$$\beta_h^* = \min_{\beta} \left\{ T_h \left(d_h^*(\beta), \beta \right) \right\}. \tag{5}$$

Then the number of samples required is

$$m_h^* = d_h^* \left(\beta_h^* \right) - n. \tag{6}$$

The whole process of the dual attack with hints is summarized in algorithm 2.

5 Discussion and Experiments

In 2020, Dachman-Soled et al. [15] put forward the idea of integrating "hints" about the secret and/or error obtained through side channel information to the primal attack. They showed the fact that hints do reduce the cost of primal attack by experiments. In this paper, this idea is extended to the dual attack, some additional benefits of adding hints to the dual attack compared with the primal attack are found.

The "primitive" requirement is a big limitation when integrating perfect hints to a primal attack. In order to estimate the attack cost after adding a perfect hint, it is necessary to ensure the hint description vector is a primitive vector of the dual lattice of the current lattice. The hints will be added one by one following the steps described below. First, the attacker needs to convert the hint description vector to a primitive vector of the dual lattice of the current lattice. Then he/she adds the converted hint to the lattice and gets a new lattice. The basis, volume, mean and covariance matrix of this new lattice are all need to be recalculated [15]. When adding the next hint, it also needs to be converted to a primitive vector of the dual lattice of the new lattice. This process will be repeated t times when adding t hints $R = Y^T S$. However, the case of the dual attack is much simpler. It can be seen from our previous analysis that, the t hints can be added at once. In particular, not only the primitiveness, but even the linear independency is no longer required. As can be seen from corollary 2, it is only the span of Y that matters.

The effect of hints on the dual attack is very intuitive. Even though the adversary searches short vectors in the original lattice $\mathscr{L}$ by BKZ, the total cost will still be reduced since the advantage of each short vector w increases by a factor of $e^{\frac{2\pi^2 \sigma_\chi^2 \|w_V\|^2}{q^2}}$. We have mentioned a more effective attack, in which the search process is performed in the projected lattice $\mathscr{L}_{V^\perp}$ of a lower dimension and a smaller volume. In particular, heuristic 2 is useful for predicting the volume and thus the attack cost without calculating a new basis. Even when heuristic 2 is not used, the calculation of the new basis is required only once. So the dual attack with hints has a simple cost model as discussed in Section 4.3. In addition, the influence of the number of hints on the cost of the dual

[15] As the computation is heavy, two lightweight implementations with limited functionality are also proposed in [15]. They maintain less information, but can only be used under certain conditions.

Algorithm 2: Dual attack with hints $Y^T S = R$.

```
Input: LWE parameters $A, b, q, n, \sigma_\chi$, hint description matrix $Y$.
Output: 0 for $b \leftarrow U(\mathbb{Z}_q^m)$ and 1 for $b \leftarrow$ LWE.
Set $t \leftarrow$ the number of columns in $Y$;
Set $\beta_h^* \leftarrow \min_\beta \{T_h\left(d_h^*(\beta), \beta\right)\}$;
Set $d \leftarrow d_h^*(\beta_h^*)$;
Set $m \leftarrow d - n$;
Set $L \leftarrow l(d, \beta_h^*)$;
Set $\epsilon \leftarrow e^{-\frac{2\pi^2\sigma_\chi^2 L^2}{q^2}}$;
Set $(\alpha_1\ \alpha_2\ \cdots\ \alpha_d) \leftarrow \begin{pmatrix} I_m & O_{m\times n} \\ A^T & qI_n \end{pmatrix}$;
for $i = 1$ to $R(d, \beta_h^*)$ do
    perform BKZ$(\zeta_1, \cdots, \zeta_{d-t})$ to find short vectors in $\mathscr{L}_d^{(V)}$.
Set $\mathit{\Pi}_{V^\perp} \leftarrow I_d - Y(Y^TY)^{-1}Y^T$;
for $i = 1$ to $d$ do
    Set $\gamma_i \leftarrow \mathit{\Pi}_{V^\perp} \cdot \alpha_i$;
$\zeta_j, \zeta_j^*, B_j, \mu_{j,k}\ (j = 1, \cdots, d-t;\ k = 1, \cdots, j-1), H = (h_{i,j}) \leftarrow$ MLLL$(\gamma_1, \cdots, \gamma_d)$;
$sum \leftarrow 0$;
for each $w_{V^\perp}$ found by BKZ do
    for $j = 1$ to $d - t$ do
        $c_j \leftarrow \frac{\langle w_{V^\perp}, \zeta_j^* \rangle}{B_j}$;
    for $g = d - t$ to 1 do
        $s \leftarrow 0$;
        for $j = g + 1$ to $d - t$ do
            $s = s + z_g \cdot \mu_{j,g}$;
        $z_g = c_g - s$;
    $w \leftarrow 0$;
    for $i = 1$ to $d$ do
        $k_i \leftarrow 0$;
        for $j = 1$ to $d - t$ do
            $k_i \leftarrow k_i + z_j \cdot h_{i,j}$;
        $w \leftarrow w + k_i \cdot \alpha_i$;
    $sum \leftarrow sum + \cos\left(\frac{2\pi\left(\langle \mu, w_j \rangle - \langle u_j, b \rangle\right)}{q}\right)$;
if $sum \geq \frac{\epsilon}{2}$ then
    return 1;
else
    return 0;
```

attack is very clear, to be specific, as the number of hints increases, the total cost is bound to get lower.

However, when perfect hints are added to a primal attack, the situation is a bit more complicated. To simplify the analysis, let us consider the case with only one hint first. We denote $\mathscr{L}'$ to be the lattice used by the attacker in a primal attack. As mentioned in remark 6, an increase in volume makes the primal attack easier. But this may not occur after integrating a hint $r =< v, S >$. To be specific, the volume of the new lattice will increase by a factor of $\|\overline{v}\|$, where $\overline{v}$ is the primitive vector transformed from v with respect to $(\mathscr{L}')^*$. It should be pointed out that although the original hint description vector v is usually an integer vector, $\overline{v}$ may not belong to $\mathbb{Z}_q^d$ after being made primitive. This is because $(\mathscr{L}')^*$ is not an integer lattice [16]. So the case of $\|\overline{v}\| < 1$ may occurs. It is even more complex to predict the case when multiple hints are added to a primal attack. The attacker has to add hints in turn and makes actual calculations.

Besides several benefits mentioned above regarding the primitive requirement, the number of hints added each time, and the prediction of the cost after adding hints, another advantage of adding hints to the dual attack is that an additional embedding like the Kannan's embedding is no longer required. Based on this, we believe that the integration of hints to a dual attack may be more natural. Although it depends on assumption 4, the fact that the probability in equation (3) is always 1 in all schemes we test may make us more optimistic about this assumption.

In this section, we will show the effect of hints on the dual attack by experiments. Using multiple hints, the cost of the attack can be greatly reduced. We mainly consider the case of $t < \min\{m, n\}$. Some hints of a similar form to the ones used in [15] are integrated [17] and the substantial reduction in cost is verified.

In table 2, we show the relationship between the total cost of the dual attack and the number of added hints in Kyber768 both in the classical case and in the quantum case. The time complexity of the search phase in the classical case is denoted by $T_s^{(c)}$, while that in the quantum case is written as $T_s^{(q)}$. $T_{r\&d}$ represents the time complexity of the recovery&distinguish process for all short vectors. The optimal number of samples m_h^* and the optimal blocksize β_h^* after adding hints are also given. $\overline{m_h^*}$ is the optimal sample number predicted by equation (6), experiments show that there is only a very small difference between m_h^* and $\overline{m_h^*}$. The situations of other schemes are shown in appendix D. We remark that although β is an integer in practice, regarding it as a real number gives a better estimate of m_h^*. One can simply round it to an integer.

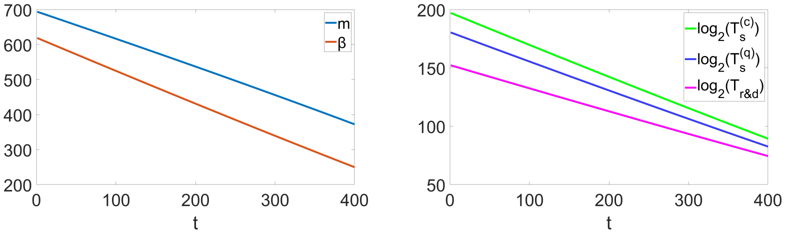
It can be seen from table 2 that, adding multiple hints can significantly reduce the cost of dual attack against a Kyber768 instance. Surprisingly, with 200 hints, even in the worst case, the blocksize can be reduced by 188 and the time complexity can be reduced by a factor of 2^{55}. Further, if more hints are available, such as 250 hints, the instance will be reduced to only 128-bit safe against the dual attack in the classical case and 118-bit safe in the quantum case.

[16] By contrast, the primitiveness of v with respect to the integer lattice $\mathscr{L}$ is considered in dual attack.

[17] This kind of hints is the one with the least volume reduction (i.e. $\det(Y^T Y) = 1$) . So the experimental results given in this section show the effect of hints on the dual attack in the worst case.

Table 2: The relationships between $m_h^*, \overline{m_h^*}, \beta_h^*, \log_2(T_s^{(c)}), \log_2(T_s^{(q)}), \log_2(T_{r\&d})$ and t respectively in Kyber768.

t	m_h^*	$\overline{m_h^*}$	β_h^*	$\log_2(T_s^{(c)})$	$\log_2(T_s^{(q)})$	$\log_2(T_{r\&d})$
0	690	693.79	619.80	197.38	180.65	152.44
50	656	655.79	572.16	183.47	168.02	142.44
100	614	616.67	524.85	169.66	155.49	132.49
150	576	577.15	477.88	155.94	143.04	122.61
200	535	537.21	431.30	142.34	130.69	112.80
250	495	496.81	385.14	128.86	118.46	103.07
300	456	455.90	339.45	115.52	106.36	93.43
400	370	372.31	249.75	89.33	82.58	74.45



Another interesting thing is that, $m_h^*,\ \beta_h^*$ are both the same in the classical and quantum cases, as they are well predicted by equations (5) and (6). It is worth noting that this occurs not only in Kyber768, but in all schemes in appendix D. Moreover, the logarithms of costs $\log_2(T_s)$ and $\log_2(T_{r\&d})$ decrease linearly with respect to t in all schemes, as do m_h^* and β_h^*. It should be pointed out that, to show the fact that $T_h \approx T_s$ in all schemes both in the classical case and in the quantum case, we use a more accurate cost model of BKZ as described in assumption 2 which may leads to a difference of $2^{16.4}$ times compared with the core-SVP model used in most other papers. However, it is the change in cost after adding hints that should be observed.

Finally, attention should be paid to the security with side channel information, the leakage of information should be avoided as much as possible. Especially for those schemes whose original security margins are small, a few hints would put them over the edge. For example, as can be seen in appendix D, just several hints can make a Newhope512 instance be less than 128-bit security against the dual attack.

References

1. Albrecht, M.R.: On dual lattice attacks against small-secret lwe and parameter choices in helib and seal. In: Advances in Cryptology – EUROCRYPT 2017. pp. 103–129. Springer International Publishing (04 2017). https://doi.org/10.1007/978-3-319-56614-6_4
2. Albrecht, M.R., Deo, A., Paterson, K.G.: Cold boot attacks on ring and module lwe keys under the ntt. IACR Transactions on Cryptographic Hardware and Embedded Systems **2018**(3), 173–213. https://doi.org/10.46586/tches.v2018.i3.173-213
3. Albrecht, M.R., Faugère, J.C., Fitzpatrick, R., Perret, L.: Lazy modulus switching for the bkw algorithm on lwe. In: Krawczyk, H. (ed.) Public-Key Cryptography – PKC 2014. pp. 429–445. Springer Berlin Heidelberg, Berlin, Heidelberg (2014)

4. Albrecht, M.R., Göpfert, F., Virdia, F., Wunderer, T.: Revisiting the expected cost of solving usvp and applications to lwe. In: Takagi, T., Peyrin, T. (eds.) Advances in Cryptology – ASIACRYPT 2017. pp. 297–322. Springer International Publishing, Cham (2017)
5. Alkim, E., Avanzi, R., Bos, J., Ducas, L., de la Piedra, A., Pöppelmann, T., Schwabe, P., Stebila, D., Albrecht, M.R., Orsini, E., et al.: Newhope algorithm specifications and supporting documentation. NIST PQC Round **2**, 4–11 (2019)
6. Alkim, E., Ducas, L., Pöppelmann, T., Schwabe, P.: Post-quantum key exchange: A new hope. In: Proceedings of the 25th USENIX Conference on Security Symposium. pp. 327–343. SEC'16, USENIX Association, USA (2016)
7. Amiet, D., Curiger, A., Leuenberger, L., Zbinden, P.: Defeating newhope with a single trace. In: Post-Quantum Cryptography – PQCrypto 2020. Lecture Notes in Computer Science, vol. 12100, pp. 189–205. Springer International Publishing, Cham (04 2020). https://doi.org/10.1007/978-3-030-44223-1_11
8. Applebaum, B., Cash, D., Peikert, C., Sahai, A.: Fast cryptographic primitives and circular-secure encryption based on hard learning problems pp. 595–618 (2009)
9. Avanzi, R., Bos, J., Ducas, L., Kiltz, E., Lepoint, T., Lyubashevsky, V., Schanck, J.M., Schwabe, P., Seiler, G., Stehlé, D.: Crystals-kyber algorithm specifications and supporting documentation. NIST PQC Round **3** (2021)
10. Aysu, A., Tobah, Y., Tiwari, M., Gerstlauer, A., Orshansky, M.: Horizontal side-channel vulnerabilities of post-quantum key exchange protocols. In: 2018 IEEE international symposium on hardware oriented security and trust (HOST). pp. 81–88. IEEE Computer Society (2018). https://doi.org/10.1109/HST.2018.8383894
11. Becker, A., Ducas, L., Gama, N., Laarhoven, T.: New directions in nearest neighbor searching with applications to lattice sieving p. 10–24 (2016)
12. Bos, J.W., Friedberger, S., Martinoli, M., Oswald, E., Stam, M.: Assessing the feasibility of single trace power analysis of frodo. In: Cid, C., Jacobson Jr., M.J. (eds.) Selected Areas in Cryptography – SAC 2018. pp. 216–234. Lecture Notes in Computer Science, Springer International Publishing (2018). https://doi.org/10.1007/978-3-030-10970-7_10
13. Chen, Y.: Réduction de réseau et sécurité concrete du chiffrement completement homomorphe. Ph.D. thesis, Paris 7 (2013)
14. Cohen, H.: A course in computational algebraic number theory. Springer Berlin Heidelberg (06 1993). https://doi.org/10.1007/978-3-662-02945-9
15. Dachman-Soled, D., Ducas, L., Gong, H., Rossi, M.: Lwe with side information: Attacks and concrete security estimation. In: Advances in Cryptology – CRYPTO 2020. pp. 329–358. Springer International Publishing, Cham (08 2020). https://doi.org/10.1007/978-3-030-56880-1_12
16. Espitau, T., Joux, A., Kharchenko, N.: On a dual/hybrid approach to small secret lwe. In: Bhargavan, K., Oswald, E., Prabhakaran, M. (eds.) Progress in Cryptology – INDOCRYPT 2020. pp. 440–462. Springer International Publishing, Cham (2020)
17. Guo, Q., Johansson, T.: Faster dual lattice attacks for solving lwe with applications to crystals. In: Tibouchi, M., Wang, H. (eds.) Advances in Cryptology – ASIACRYPT 2021. Lecture Notes in Computer Science, vol. 4, pp. 33–62. Springer International Publishing, Cham (2021)
18. Guo, Q., Johansson, T., Nilsson, A.: A key-recovery timing attack on post-quantum primitives using the fujisaki-okamoto transformation and its application on frodokem. In: Advances in Cryptology – CRYPTO 2020. pp. 359–386. Lecture Notes in Computer Science, Springer International Publishing (08 2020). https://doi.org/10.1007/978-3-030-56880-1_13
19. Kirchner, P., Fouque, P.A.: An improved bkw algorithm for lwe with applications to cryptography and lattices. In: Gennaro, R., Robshaw, M. (eds.) Advances in Cryptology – CRYPTO 2015. pp. 43–62. Springer Berlin Heidelberg, Berlin, Heidelberg (2015). https://doi.org/10.1007/978-3-662-47989-6_3

20. Lenstra, A., Lenstra, H., László, L.: Factoring polynomials with rational coefficients. Mathematische Annalen **261**, 515–534 (12 1982). https://doi.org/10.1007/BF01457454
21. Li, S., Lu, X., Zhang, J., Li, B., Bi, L.: Predicting the concrete security of lwe against the dual attack using binary search. In: Gao, D., Li, Q., Guan, X., Liao, X. (eds.) Information and Communications Security – ICICS 2021. pp. 265–282. Lecture Notes in Computer Science, Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-88052-1_16
22. Lindner, R., Peikert, C.: Better key sizes (and attacks) for lwe-based encryption. In: Kiayias, A. (ed.) Topics in Cryptology – CT-RSA 2011. pp. 319–339. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
23. Liu, M., Wang, X., Xu, G., Zheng, X.: A note on bdd problems with λ_2-gap. In: Information Processing Letters. vol. 114, pp. 9–12 (2014). https://doi.org/https://doi.org/10.1016/j.ipl.2013.10.004
24. Lu, X., Liu, Y., Zhang, Z., Jia, D., Xue, H., He, J., Li, B., Wang, K.: Lac: Practical ring-lwe based public-key encryption with byte-level modulus. Cryptology ePrint Archive, Report 2018/1009 (2018)
25. Micciancio, D., Regev, O.: Lattice-based cryptography. In: Bernstein, D.J., Buchmann, J., Dahmen, E. (eds.) Post-Quantum Cryptography – PQCrypto 2009. pp. 147–191. Springer Berlin Heidelberg (01 2009). https://doi.org/10.1007/978-3-540-88702-7_5
26. Osgood, B.: Ee 261 the fourier transform and its applications. Electrical Engineering Department: Stanford University pp. 252–253 (2007)
27. Pessl, P., Prokop, L.: Fault attacks on cca-secure lattice kems. IACR Transactions on Cryptographic Hardware and Embedded Systems **2021**(2), 37–60 (Feb 2021). https://doi.org/10.46586/tches.v2021.i2.37-60
28. Pohst, M.: A modification of the lll reduction algorithm. Journal of Symbolic Computation **4**, 123–127 (08 1987). https://doi.org/10.1016/S0747-7171(87)80061-5
29. Regev, O.: On lattices, learning with errors, random linear codes, and cryptography. In: Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing. pp. 84–93. STOC '05, Association for Computing Machinery, New York, NY, USA (2005). https://doi.org/10.1145/1060590.1060603
30. Serre, J.p.: A course in arithmetic. Springer-Verlag (01 1973)
31. Wei, W., Liu, M., Wang, X.: Finding shortest lattice vectors in the presence of gaps. In: Nyberg, K. (ed.) Topics in Cryptology — CT-RSA 2015. pp. 239–257. Springer International Publishing, Cham (2015)
32. Zhao, C., Zheng, Z., Wang, X., Xu, G.: Distinguishing LWE instances using fourier transform: A refined framework and its applications. Cryptology ePrint Archive, Report 2019/1231 (2019), https://eprint.iacr.org/2019/1231

A The proof of proposition 1

The so called *unit impulse function* $\delta(x)$ is given by

$$\delta(x) = \begin{cases} +\infty & x = 0 \\ 0 & \text{else} \end{cases} \quad \text{and} \quad \int_{\mathbb{R}^d} \delta(x)\, dx = 1.$$

It is an important tool in our proof. One of the most commonly used properties of it is, $\delta(x)$ and constant 1 are the Fourier transforms of each other.

In the following, we shall give a proof of proposition 1(1) for the continuous case. This proof can be easily extended to the discrete case using the *unit impulse train* in parallel.

First of all, we give the proof for the case where the coefficients of x are independent of each other. In this case, we can decompose $f(x)$ as $f(x) = f_1(x_1) \cdot f_2(x_2) \cdots f_d(x_d)$. Since $< v, x >= \sum_{j=1}^{d} v_j x_j$, we have $f_{<v,x>} = \otimes_{j=1}^{d} f_{v_j x_j}$, then $\widehat{f_{<v,x>}} = \prod_{j=1}^{d} \widehat{f_{v_j x_j}}$. As the pdf of $v_j x_j$ is

$$f_{v_j x_j}(x) = \begin{cases} \frac{1}{|v_j|} f_j\left(\frac{x}{v_j}\right) & v_j \neq 0 \\ \delta(x) & v_j = 0 \end{cases}, \ \ j = 1, 2, \cdots, d,$$

the Fourier transform of $f_{v_j x_j}(x)$ is

$$\widehat{f_{v_j x_j}}(y) = \begin{cases} \frac{1}{|v_j|} \widehat{f_j\left(\frac{x}{v_j}\right)}(y) = \frac{1}{|v_j|} \cdot |v_j| \cdot \widehat{f}_j(v_j y) = \widehat{f}_j(v_j y) & v_j \neq 0 \\ 1 & v_j = 0 \end{cases}, \ j = 1, 2, \cdots, d.$$

Moreover, because when $v_j = 0$, the pdf f_j also satisfies

$$1 = \int_{\mathbb{R}} f_j(x) dx = \int_{\mathbb{R}} f_j(x) \cdot e^{-2\pi i <x,0>} dx = \widehat{f}_j(0) = \widehat{f}_j(v_j y),$$

the result can be combined to

$$\widehat{f_{v_j x_j}}(y) = \widehat{f}_j(v_j y), \ j = 1, 2, \cdots, d.$$

Hence, we have

$$\widehat{f_{<v,x>}}(y) = \prod_{j=1}^{d} \widehat{f_{v_j x_j}}(y) = \prod_{j=1}^{d} \widehat{f}_j(v_j y) = \widehat{f}(yv).$$

Now let us consider the general case. We denote the covariance matrix of x by Σ_x. By performing EVD on Σ_x, we have

$$Q \Sigma_x Q^T = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{pmatrix} := D,$$

where $\lambda_1, \cdots, \lambda_d$ are the eigenvalues of Σ_x and Q is an orthonormal matrix. Let $y = Qx$ and we denote the pdf of y by g, then

$$g(y) = \frac{1}{|\det(Q)|} \cdot f\left(Q^{-1} y\right) = f(Q^T y) \text{ and } \widehat{g}(z) = \widehat{f}(Q^T z).$$

We denote the covariance matrix of y by Σ_y, then $\Sigma_y = Q \Sigma_x Q^T = D$ is a diagonal matrix, thus the coefficients of y are independent of each other. Then $g(y)$ can be decomposed as $g(y) = g_1(y_1) \cdots g_n(y_n)$. Since $< v, x >= v^T x = v^T Q^T Q x =< Qv, y >$, and we have

$$\widehat{f_{<v,x>}}(z) = \widehat{f_{<Qv,y>}}(z) = \widehat{g}(zQv) = \widehat{f}(Q^T \cdot z \cdot Qv) = \widehat{f}(zv).$$

B The proof of proposition 2

Since $\varSigma$ is a real symmetric matrix, we could perform the EVD of it and obtain

$$S^T \varSigma S = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_d \end{pmatrix} := D,$$

where $\lambda_1, \cdots, \lambda_d$ are the eigenvalues of $\varSigma$, and S is an orthonormal matrix. Then $SDS^T = \varSigma$. We denote $A = \begin{pmatrix} \sqrt{\lambda_1} & & & \\ & \sqrt{\lambda_2} & & \\ & & \ddots & \\ & & & \sqrt{\lambda_d} \end{pmatrix}$ and $B = SA$, it can be easily proven that $BB^T = \varSigma$, so $det(B) = \sqrt{\det(\varSigma)}$, then

$$\begin{aligned}
\widehat{f^d_{\mu,\varSigma}}(y) &= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{\det(\varSigma)}} e^{-\frac{1}{2}(x-\mu)^T \varSigma^{-1}(x-\mu)} \cdot e^{-2\pi i<x,y>} dx \\
&\xlongequal{z=x-\mu} \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{\det(\varSigma)}} e^{-\frac{1}{2}z^T \varSigma^{-1} z} \cdot e^{-2\pi i<z+\mu,y>} dz \\
&= \frac{e^{-2\pi i<\mu,y>}}{(2\pi)^{\frac{d}{2}} \sqrt{\det(\varSigma)}} \int_{\mathbb{R}^n} e^{-\frac{1}{2}z^T \varSigma^{-1} z} \cdot e^{-2\pi i<z,y>} dz \\
&\xlongequal{z=Bu} \frac{e^{-2\pi i<\mu,y>}}{(2\pi)^{\frac{d}{2}} \sqrt{\det(\varSigma)}} \int_{\mathbb{R}^n} e^{-\frac{1}{2}u^T B^T \varSigma^{-1} Bu} \cdot e^{-2\pi i<Bu,y>} \det(B) du \\
&= \frac{e^{-2\pi i<\mu,y>}}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^n} e^{-\frac{1}{2}u^T A^T S^T \varSigma^{-1} SAu} \cdot e^{-2\pi i<Bu,y>} du \\
&= \frac{e^{-2\pi i<\mu,y>}}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^n} e^{-\frac{1}{2}u^T u} \cdot e^{-2\pi i u^T B^T y} du \\
&= \frac{e^{-2\pi i<\mu,y>}}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^n} e^{-\frac{1}{2}\left[\left(u+2\pi i B^T y\right)^T \left(u+2\pi i B^T y\right) + 4\pi^2 y^T B B^T y\right]} du \\
&= \frac{e^{-2\pi i<\mu,y>}}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^n} e^{-\frac{1}{2}\left\|u+2\pi i B^T y\right\|^2} \cdot e^{-2\pi^2 y^T \varSigma y} du \\
&= \frac{e^{-2\pi i<\mu,y>} \cdot e^{-2\pi^2 y^T \varSigma y}}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^n} e^{-\frac{1}{2}\left\|u+2\pi i B^T y\right\|^2} du \\
&= \frac{e^{-2\pi i<\mu,y>} \cdot e^{-2\pi^2 y^T \varSigma y}}{(2\pi)^{\frac{d}{2}}} \cdot (2\pi)^{\frac{d}{2}} = e^{-2\pi i<\mu,y>} \cdot e^{-2\pi^2 y^T \varSigma y} \\
&= e^{-2\pi^2\left(y^T \varSigma y + \frac{i}{\pi}\mu^T y\right)} = e^{-\frac{1}{2}\mu^T \varSigma^{-1}\mu} \cdot e^{-2\pi^2\left[\left(y+\frac{i}{2\pi}\varSigma^{-1}\mu\right)^T \cdot \varSigma \cdot \left(y+\frac{i}{2\pi}\varSigma^{-1}\mu\right)\right]} \\
&= e^{-\frac{1}{2}\mu^T \varSigma^{-1}\mu} \cdot e^{-\frac{1}{2}\left[\left(2\pi y + i\varSigma^{-1}\mu\right)^T \varSigma \left(2\pi y + i\varSigma^{-1}\mu\right)\right]} \\
&= e^{-\frac{1}{2}\mu^T \varSigma^{-1}\mu} \cdot f_{-i\varSigma^{-1}\mu, \varSigma^{-1}}(2\pi y) \cdot (2\pi)^{\frac{d}{2}} \cdot \sqrt{\det(\varSigma^{-1})} \\
&= \frac{e^{-\frac{1}{2}\mu^T \varSigma^{-1}\mu} \cdot (2\pi)^{\frac{d}{2}}}{\sqrt{\det(\varSigma)}} \cdot f_{-i\varSigma^{-1}\mu, \varSigma^{-1}}(2\pi y).
\end{aligned}$$

Then (1) has already been proved. Taking $\varSigma' = \varSigma^{-1}, \mu' = -i\varSigma'\mu$ and $z = 2\pi y$ in (1), we can easily prove (2).

C The proof of proposition 3

Let us start with the proof for a simple case where $a_i = 1, 1 \leq i \leq t$. As $X = (v_1\ v_2\ \cdots\ v_t)$ is a primitive vector set, it can be extended to a basis $B = (v_1\ \cdots\ v_t\ v_{t+1}\ \cdots\ v_d)$ of L. We write $F = \text{Span}(X)$, then for any $j \geq t+1,\ v_j$ can be decomposed into

$$v_j = \sum_{k=1}^{t} a_{jk} v_k + v_j^{\perp},$$

with $v_j^{\perp} \in F^{\perp}$. And it is easy to see that $B_{\perp} := (v_{t+1}^{\perp}\ \ \cdots\ \ v_n^{\perp})$ is a basis of L'. Then for any $j \geq t+1$, we have

$$\langle v_i, v_j \rangle = \left\langle v_i, \sum_{k=1}^{t} a_{jk} v_k + v_j^{\perp} \right\rangle = \sum_{k=1}^{t} a_{jk} \langle v_i, v_k \rangle + \left\langle v_i, v_j^{\perp} \right\rangle .$$

Hence, for $j \geq t+1$, when $1 \leq i \leq t$, $\langle v_i, v_j \rangle = \sum_{k=1}^{t} a_{jk} \langle v_i, v_k \rangle$, while if $t+1 \leq i \leq d$, $\langle v_i, v_j \rangle = \sum_{k=1}^{t} a_{jk} \langle v_i, v_k \rangle + \left\langle v_i^{\perp}, v_j^{\perp} \right\rangle$. Therefore,

$$\begin{aligned}
\text{vol}(L) = \sqrt{\det(B^T B)} &= \sqrt{\det \begin{pmatrix} \langle v_1, v_1 \rangle & \cdots & \langle v_1, v_t \rangle & \langle v_1, v_{t+1} \rangle & \cdots & \langle v_1, v_d \rangle \\ & \cdots & & & \cdots & \\ \langle v_t, v_1 \rangle & \cdots & \langle v_t, v_t \rangle & \langle v_t, v_{t+1} \rangle & \cdots & \langle v_t, v_d \rangle \\ \langle v_{t+1}, v_1 \rangle & \cdots & \langle v_{t+1}, v_t \rangle & \langle v_{t+1}, v_{t+1} \rangle & \cdots & \langle v_t, v_d \rangle \\ & \cdots & & & \cdots & \\ \langle v_d, v_1 \rangle & \cdots & \langle v_d, v_t \rangle & \langle v_d, v_{t+1} \rangle & \cdots & \langle v_d, v_d \rangle \end{pmatrix}} \\
&= \sqrt{\det \begin{pmatrix} \langle v_1, v_1 \rangle & \cdots & \langle v_1, v_t \rangle & 0 & \cdots & 0 \\ & \cdots & & & \cdots & \\ \langle v_t, v_1 \rangle & \cdots & \langle v_t, v_t \rangle & 0 & \cdots & 0 \\ \langle v_{t+1}, v_1 \rangle & \cdots & \langle v_{t+1}, v_t \rangle & \left\langle v_{t+1}^{\perp}, v_{t+1}^{\perp} \right\rangle & \cdots & \left\langle v_t^{\perp}, v_d^{\perp} \right\rangle \\ & \cdots & & & \cdots & \\ \langle v_d, v_1 \rangle & \cdots & \langle v_d, v_t \rangle & \left\langle v_d^{\perp}, v_{t+1}^{\perp} \right\rangle & \cdots & \left\langle v_d^{\perp}, v_d^{\perp} \right\rangle \end{pmatrix}} \\
&= \sqrt{\det \begin{pmatrix} \langle v_1, v_1 \rangle & \cdots & \langle v_1, v_t \rangle \\ & \cdots & \\ \langle v_t, v_1 \rangle & \cdots & \langle v_t, v_t \rangle \end{pmatrix}} \cdot \sqrt{\det \begin{pmatrix} \left\langle v_{t+1}^{\perp}, v_{t+1}^{\perp} \right\rangle & \cdots & \left\langle v_t^{\perp}, v_d^{\perp} \right\rangle \\ & \cdots & \\ \left\langle v_d^{\perp}, v_{t+1}^{\perp} \right\rangle & \cdots & \left\langle v_d^{\perp}, v_d^{\perp} \right\rangle \end{pmatrix}} \\
&= \sqrt{\det(X^T X)} \cdot \sqrt{\det(B_{\perp}^T \cdot B_{\perp})} = \sqrt{\det(X^T X)} \cdot \text{vol}(L').
\end{aligned}$$

The third equal sign holds by adding column $k(1 \leq k \leq t)$ multiplied $(-a_{jk})$ in turn to column j for $t+1 \leq j \leq d$. In summary, we have

$$\text{vol}(L') = \frac{\text{vol}(L)}{\sqrt{\det(X^T X)}} \quad \text{when } a_i = 1, i = 1, 2, \cdots, t.$$

As for the more general case, we define $\widetilde{X} = (a_1 v_1 \ \cdots \ a_t v_t)$, then

$$\mathrm{vol}(L') = \frac{\mathrm{vol}(L)}{\sqrt{\det(\widetilde{X}^T \cdot \widetilde{X})}} = \frac{vol(L)}{|a_1 \cdots a_t| \cdot \sqrt{\det(X^T X)}}.$$

D Experiments on some actual schemes

Table 3: The relationships between various parameters and t in Newhope512.

t	m_h^*	$\overline{m_h^*}$	β_h^*	$\log_2(T_s^{(c)})$	$\log_2(T_s^{(q)})$	$\log_2(T_{r\&d})$
0	569	572.72	382.67	128.14	117.81	102.37
50	520	523.96	336.26	114.59	105.51	92.54
100	472	474.61	290.44	101.21	93.37	82.83
150	423	424.60	245.29	88.02	81.40	73.24
200	373	373.85	200.90	75.06	69.64	63.78
250	319	322.28	157.37	62.35	58.10	54.45
300	268	269.79	114.71	49.90	46.80	45.29

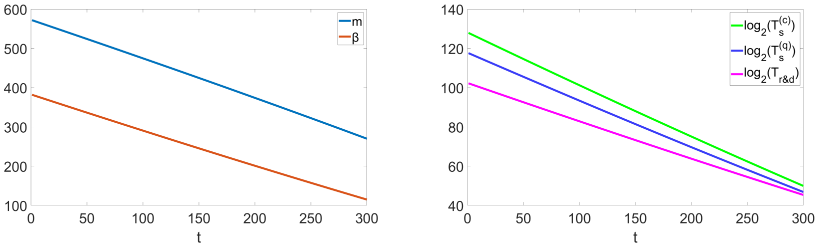
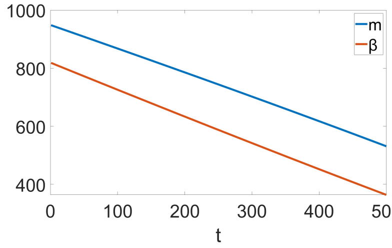


Table 4: The relationships between various parameters and t in FireSaber.

t	m_h^*	$\overline{m_h^*}$	β_h^*	$\log_2(T_s^{(c)})$	$\log_2(T_s^{(q)})$	$\log_2(T_{r\&d})$
0	944	948.70	819.20	255.61	233.49	194.68
50	905	908.29	772.39	241.94	221.08	184.97
100	863	867.65	725.82	228.34	208.74	175.11
150	823	826.72	679.49	214.81	196.46	165.40
200	782	785.53	633.43	201.36	184.26	155.74
300	699	702.22	542.19	174.72	160.08	136.59
400	616	617.53	452.29	148.47	136.26	117.69
500	531	531.21	364.00	122.69	112.96	99.11



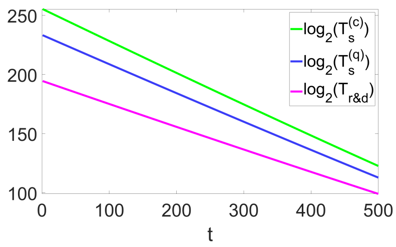
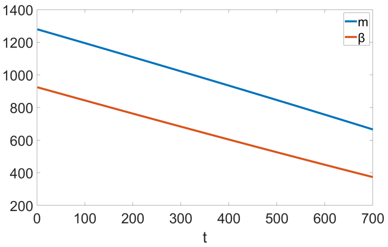
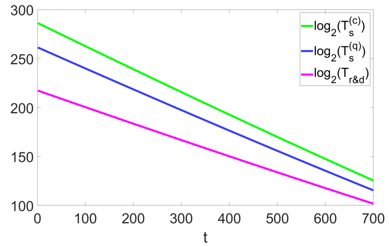


Table 5: The relationships between various parameters and t in Frodo1344.

t	m_h^*	$\overline{m_h^*}$	β_h^*	$\log_2(T_s^{(c)})$	$\log_2(T_s^{(q)})$	$\log_2(T_{r\&d})$
0	1275	1279.91	924.91	286.47	261.50	217.44
100	1192	1194.57	843.43	262.68	239.91	200.39
200	1104	1108.56	762.70	239.11	218.52	183.48
300	1019	1021.80	682.79	215.77	197.34	166.74
400	934	934.23	603.80	192.71	176.41	150.18
500	845	845.74	525.85	169.95	155.75	133.81
600	752	756.23	449.09	147.53	135.41	117.68
700	663	665.52	373.69	125.52	115.43	101.81


1400
1200
1000
800
600
400
200
0
100
200
300
400
500
600
700
t
m
β


t
$\log_2(T_s^{(c)})$
$\log_2(T_s^{(q)})$
$\log_2(T_{r\&d})$