基于列表译码方法在查询访问模型下含错学习问题的分析

王明强[①②] 庄金成*[①③]

[①](山东大学密码技术与信息安全教育部重点实验室 青岛 266237)

[②](山东大学数学学院 济南 250100)

[③](山东大学网络空间安全学院 青岛 266237)

摘 要：Regev在2005年提出了含错学习问题(LWE)，这个问题与随机线性码的译码问题密切相关，并且在密码学特别是后量子密码学中应用广泛。原始的含错学习问题是在随机访问模型下提出的，有证据证明该问题的困难性。许多研究者注意到的一个事实是当攻击者可以选择样本时，该问题是容易的。但是目前据作者所知并没有一个完整的求解算法。该文分析了查询访问模型下的带有错误学习问题，给出了完整的求解算法。分析采用的工具是将该问题联系到隐藏数问题，然后应用傅里叶学习算法进行列表译码。

关键词：含错学习问题；查询访问模型；隐藏数问题；傅里叶学习；列表译码

中图分类号：TN918; TP309 **文献标识码**：A **文章编号**：1009-5896(2020)02-0322-05

DOI: 10.11999/JEIT190624

Analysis of Learning With Errors in Query Access Model: A List Decoding Approach

WANG Mingqiang[①②] ZHUANG Jincheng[①③]

[①](*Key Laboratory of Cryptologic Technology and Information Security*, *Ministry of Education Shandong University*, *Qingdao* 266237, *China*)

[②](*School of Mathematics*, *Shandong University*, *Jinan* 250100, *China*)

[③](*School of Cyber Science and Technology*, *Shandong University*, *Qingdao* 266237, *China*)

Abstract: Regev introduced the Learning With Errors (LWE) problem in 2005, which has close connections to random linear code decoding and has found wide applications to cryptography, especially to post-quantum cryptography. The LWE problem is originally introduced in random access model, and there are evidences that indicate the hardness of this problem. It is well known that the LWE problem is vulnerable if the attacker is allowed to choose samples. However, to the best of the author's knowledge, a complete algorithm has not been published. In this paper, the LWE problem in query samples access model is analyzed. The technique is to relate the problem to the hidden number problem, and then Fourier learning method is applied to the list decoding.

Key words: Learning With Errors (LWE) problem; Query access model; Hidden number problem; Fourier learning; List decoding

1 引言

计算学习理论[1]是人工智能理论的一个分支，旨在研究学习算法的设计与分析。给定样本源$(x, f(x))$的访问权限，计算学习理论中的一个基本问题是设计一个学习算法A来构造一个依据某种测度逼近f的函数。根据A获取样本方式的不同，有很多不同的访问模型。其中一类模型是随机访问模型(Random Access Model, RAM)，在此模型中A可以获得样本$(x, f(x))$，这里x是均匀分布的。另外一个模型是查询访问模型(query access model)，在此模型中A可以自己选择输入x获得样本$(x, f(x))$。

含噪的奇偶性学习问题(Learning Parity with Noise，LPN)是学习理论中的一个经典问题。该问题是在随机访问模型中定义的，即给定随机样本访问权限，求解秘密向量。LPN问题和随机二元线性码的译码问题密切相关，目前最快的求解算法是Blum等人[2]设计的算法。LPN问题在密码学中有广泛的应用[3]。

收稿日期：2019-08-14；改回日期：2019-12-05；网络出版：2019-12-09

*通信作者： 庄金成 jzhuang@sdu.edu.cn

基金项目：国家自然科学基金(61672019)

Foundation Item: The National Natural Science Foundation of China (61672019)

Regev[4]提出了含错学习问题(Learning With Errors, LWE)，该问题也是在随机访问模型下定义的，LPN问题是LWE问题的一个特例。LWE问题具有一些良好的特性，例如特定参数的格中最短向量问题的最坏情形的求解可以规约到LWE平均情形的求解，目前没有发现效率优于经典算法的量子求解算法等。LWE问题已经在密码学，特别是后量子密码学中获得了广泛的应用。

为了研究Diffie-Hellman密钥交换协议的比特安全性，Boneh和Venkatesan[5]提出了隐藏数问题。Galbraith和Shani[6]将隐藏数问题进行推广，定义了多变量隐藏数问题。Goldreich和Levin[7]构造了一般值域为二元的单向函数的困难谓词，证明的主要方法可以描述成通过傅里叶学习算法给出了Hadamard编码的列表译码算法。在查询访问模型下，可以应用Goldreich-Levin算法有效地求解LPN问题。Kushilevitz和Mansour[8]推广了Goldreich-Levin算法。Akavia等人[9]基于Goldreich-Levin算法和Kushilevitz-Mansour算法，将列表译码的方法进行了形式化，应用于许多单向函数困难比特的证明上。

许多研究者注意到的一个事实是如果允许攻击者选取样本，那么LWE问题是容易求解的。例如，Galbraith等人[10]提到稀疏傅里叶学习算法可以用于求解LWE。但是目前没有一个完整的求解算法。本文考虑在查询访问模型下的含错学习问题(LWE in query access model, qLWE)，建立qLWE问题和隐藏数问题之间的联系，考察了两类密切相关的最重要比特的概率分布关系，在此基础上构造一类编码和该类编码的扰动版本使得表示二者的函数有共同的重的傅里叶系数，然后应用傅里叶学习算法给出qLWE的完整求解算法。

本文后续部分安排如下：第2节介绍了基本的记号，离散傅里叶变换，纠错码的基本知识；第3节建立了qLWE问题和隐藏数问题之间的联系；第4节给出qLWE问题求解算法的具体设计和分析；第5节总结全文。

2　准备知识

2.1　记号

本文用标准的符号$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{C}$来分别表示自然数，整数，实数，复数。用$\mathbb{Z}_+$和$\mathbb{R}_+$分别表示正整数和正实数。一个函数$\nu(l):\mathbb{N}\to\mathbb{R}$称作是可忽略的如果对每一个常数$\mathrm{c}\in\mathbb{R}_+$存在$l_{\mathrm{c}}\in\mathbb{N}$使得对所有$l>l_{\mathrm{c}}$有$\nu(l)<l^{-\mathrm{c}}$。一个函数$\rho(l):\mathbb{N}\to\mathbb{R}$是非可忽略的如果存在常数$\mathrm{c}\in\mathbb{R}_+$和$l_c\in\mathbb{N}$使得对所有的$l>l_{\mathrm{c}}$有$\rho(l)>l^{-\mathrm{c}}$。给定一个定义域为$\mathcal{D}$的布尔函数$f:\mathcal{D}\to\{\pm1\}$，记$\mathrm{maj}_f=\max_{\{b=\pm1\}}\Pr_{\alpha\in\mathcal{D}}[f(\alpha)=b]$为$f$的偏差。$\mathbb{Z}_m$表示模$m$的整数等价类集合，代表元取自$[0,m-1]$，一个整数加上下标$m$表示该整数对应的代表元。$\mathrm{lb}n$表示以2为底的对数。

2.2　离散傅里叶变换

令$\mathbb{G}$为一个有限交换群。记$\mathbb{C}(\mathbb{G})=\{f:\mathbb{G}\to\mathbb{C}\}$。可知$\mathbb{C}(\mathbb{G})$是一个$\mathbb{C}$上的维数为$|\mathbb{G}|$的向量空间。对任意两个函数$f,g\in\mathbb{C}(\mathbb{G})$，它们的内积定义为$\langle f,g\rangle=(1/|\mathbb{G}|)\sum_{x\in\mathbb{G}}f(x)\overline{g(x)}$。$f$在向量空间$\mathbb{C}(\mathbb{G})$上的$\mathrm{l}_2$范数定义为$\|f\|_2=\sqrt{\langle f,f\rangle}$。一个$\mathbb{G}$的特征是指一个同态映射$\chi:\mathbb{G}\to\mathbb{C}^*$，即对任意的$x,y\in\mathbb{G}$有$\chi(x+y)=\chi(x)\chi(y)$。$\mathbb{G}$的特征的集合构成一个特征群$\widehat{\mathbb{G}}$，该群的元素构成向量空间$\mathbb{C}(\mathbb{G})$的一组基，称为傅里叶基底。任意函数$f\in\mathbb{C}(\mathbb{G})$有傅里叶展开$\sum_{\chi\in\widehat{\mathbb{G}}}\hat{f}(\chi)\chi$，其中$\hat{f}:\widehat{\mathbb{G}}\to\mathbb{C}$是$f$的傅里叶变换，并且有$\hat{f}(\chi)=\langle f,\chi\rangle$。相对于傅里叶基底$\{\chi\}_{\chi\in\widehat{\mathbb{G}}}$展开的系数$\hat{f}(\chi)$被称作傅里叶系数。一个傅里叶系数的重量定义为$|\hat{f}(\chi)|^2$。当$\mathbb{G}=\mathbb{Z}_n$为整数模$n$构成的加群时，$\widehat{\mathbb{G}}=\widehat{\mathbb{Z}}_n$。对每一个$\alpha\in\mathbb{Z}_n$, α-特征被定义为函数$\chi_\alpha:\mathbb{Z}_n\to\mathbb{C}$使得$\chi_\alpha(x)=\omega_n^{\alpha x}$，其中$\omega_n=\mathrm{e}^{2\pi\mathrm{i}/n}$。

如果$A\subset\mathbb{Z}_n$，考虑f在子集合A中的限制，即$f_{|A}=\sum_{\alpha\in A}\hat{f}(\alpha)\chi_\alpha$，其中$\hat{f}(\alpha)=\langle f,\chi_\alpha\rangle$。因为特征是正交的，有$\|f\|_2^2=\sum_{\alpha\in\mathbb{Z}_n}|\hat{f}(\alpha)|^2$以及$\left\|f_{|A}\right\|_2^2=\sum_{\alpha\in A}|\hat{f}(\alpha)|^2$。Parseval等式描述了$f$和$\hat{f}$的范数之间的关系

$$\|f\|_2=\sum_{\alpha\in\mathbb{Z}}|\hat{f}(\alpha)|^2=|G|\cdot\left\|\hat{f}\right\|_2^2 \tag{1}$$

下面引述文献[9]中的一些定义。

定义1　一个函数$f:\mathbb{Z}_n\to\mathbb{C}$是傅里叶$\epsilon$-集聚的，如果存在一个子集合$A\subseteq\mathbb{Z}_n$包含$\mathrm{poly}(\mathrm{lb}n,1/\epsilon)$个特征使得

$$\left\|f-f_{|A}\right\|_2^2=\sum_{\alpha\notin A}|\hat{f}(\alpha)|^2\le\epsilon \tag{2}$$

一个函数f被称作傅里叶集聚的如果对任意的$\epsilon<0$, f都是傅里叶ϵ-集聚的。

定义2　给定一个阈值$\tau>0$和一个任意的函数$f:\mathbb{Z}_n\to\mathbb{C}$，称一个特征$\chi_\alpha$对于$f$是$\tau$-重的，如果该特征对应的傅里叶系数不小于$\tau$。所有$\tau$-重的特征的集合记为

$$\mathrm{Heavy}_\tau(f)=\{\chi_\alpha:|\hat{f}(\alpha)|^2\ge\tau\} \tag{3}$$

2.3　纠错码的定义和性质

纠错码通过在编码的过程中加入冗余信息从而解决在有噪信道传输的信号恢复问题。考虑将每个

元素$\alpha\in\mathbb{Z}_n$编码成一个长度为n的码字C_α。每个码字C_α可以被表示为一个函数$C_\alpha:\mathbb{Z}_n\to\{\pm1\}$。在不引起歧义的情况下对于码字以及其表示函数不加区分。下面引述一些$\mathbb{Z}_n$上编码的定义和引理[9]。

定义3 一个编码$\mathcal{C}=\{C_\alpha:\mathbb{Z}_n\to\{\pm1\}\}$是集聚的，如果每一个函数$C_\alpha$是傅里叶集聚的。

定义4 一个编码$\mathcal{C}=\{C_\alpha:\mathbb{Z}_n\to\{\pm1\}\}$是可恢复的，如果给定特征$\chi\in\widehat{\mathbb{Z}}_n$和一个阈值$\tau$，存在一个恢复算法在$\mathrm{poly}(\mathrm{lb}n,1/\tau)$时间之内输出所有元素$\alpha$的列表，使得$\chi$对于$C_\alpha$是$\tau$-重的系数，即输出集合$\{\alpha\in\mathbb{Z}_n:\chi\in\mathrm{Heavy}_\tau(C_\alpha)\}$。

下述引理1表明对于一个集聚码$\mathcal{C}$的码字C_α，加了噪声的的相近版本$\tilde{C}_\alpha$与之共享至少一个重的特征。下述引理2表明给定函数f的查询访问权限，可以有效地得到它的所有重的特征。

引理1[9] 给定函数$f,g:\mathbb{Z}_n\to\{\pm1\}$，其中$f$是傅里叶集聚的，并且对于某个常数$\epsilon>0$有

$$\Pr_{\alpha\in\mathbb{Z}_n}[f(\alpha)=g(\alpha)]\geq\mathrm{maj}_f+\epsilon \tag{4}$$

那么存在一个阈值τ满足$1/\tau\in\mathrm{poly}(1/\epsilon,\mathrm{lb}n)$，存在一个非平凡的特征$\chi\neq0$(即存在$\alpha\in\mathbb{Z}_n$，$\alpha\neq0$，使得$\chi=\chi_\alpha$)，使得

$$\chi\in\mathrm{Heavy}_\tau(f)\cap\mathrm{Heavy}_\tau(g) \tag{5}$$

引理2[9] 给定函数$w:\mathbb{Z}_n\to\{\pm1\}$，常数$\tau>0$和$0<\delta<1$，存在一个概率算法，通过查询访问该函数，以不低于$1-\delta$的概率输出一个包含$\mathrm{Heavy}_\tau(w)$的个数为$O(1/\tau)$的特征列表$L$，算法的运行时间为$\tilde{O}\left(\mathrm{lb}n\cdot\ln^2\dfrac{(1/\delta)}{\tau^{5.5}}\right)$。

Akavia等人[9]首先提出了针对单向函数困难谓词证明的一般的列表译码方法。简言之，该方法包括下述步骤。给定一个单向函数$f:\mathcal{D}\to\mathcal{R}$和一个谓词$\pi$，构造一个纠错码$\mathcal{C}^\pi=\{C_\alpha:\mathcal{D}\to\{\pm1\}\}_{\alpha\in\mathcal{D}}$，使得每一个函数的输入$\alpha$对应于一个码字$C_\alpha$，并且满足如下条件：

(1) 可访问性：对于码字C_α，可以访问一个与之相近的含噪版本$\tilde{C}_\alpha$。

(2) 集聚性：编码函数C_α是一个傅里叶集聚函数。

(3) 可恢复性：存在有效的算法，给定输入特征χ和阈值τ，输出一个α的列表，使得χ对于C_α是τ-重的。

结合引理1，引理2，给定含噪版本的访问，可以通过学习算法对函数f求逆，从而可以用来证明函数的困难谓词。

3 两类最重要比特的关系

本节建立两类最重要比特(Most Significant Bit, MSB)的关系，之后会用来构造码字和含噪版本。

根据定义，qLWE的样本具有形式

$$(\boldsymbol{a},[\langle\boldsymbol{a},\boldsymbol{s}\rangle+e]_q)\in\mathbb{Z}_q^n\times\mathbb{Z}_q$$

其中，$\boldsymbol{a}\in\mathbb{Z}_q^n$根据询问选择，$e$根据错误分布$\chi$选择。在文献[4]中，$\chi$被设定为中心为0的$\mathbb{Z}_q$上的离散高斯分布。下面估计错误$e$不是太大的概率。为此，考虑整数环上的离散高斯分布$D_{\mathbb{Z},\sigma}$，也就是考虑整数环$\mathbb{Z}$上的概率分布，使得对于一个整数$x$赋与一个概率

$$\Pr(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{6}$$

其中，σ是标准差。

令$\boldsymbol{s}=(s_1,s_2,\cdots,s_n)\in\mathbb{Z}_q^n$,$\boldsymbol{a}=(a_1,a_2,\cdots,a_n)\in\mathbb{Z}_q^n$。定义$\mathbb{Z}_q$上的最重要比特函数为$\mathrm{MSB}(x)=1$，如果$0\leq x<q/2$; $\mathrm{MSB}(x)=-1$，如果$q/2\leq x<q$。

不失一般性可以假设$\boldsymbol{s}\neq0$，即并非所有分量$s_i=0$。因为$\boldsymbol{a}$在$\mathbb{Z}_q^n$中是平均分布的，所以$[\langle\boldsymbol{a},\boldsymbol{s}\rangle+e]_q$也是平均分布的。从而当$q$为偶数时，$\mathrm{MSB}([\langle\boldsymbol{a},\boldsymbol{s}\rangle+e]_q)$的偏差是$1/2$；当$q$为奇数时，$\mathrm{MSB}([\langle\boldsymbol{a},\boldsymbol{s}\rangle+e]_q)$的偏差是$1/2+1/2q$。

给定一个qLWE的样本，容易计算$\mathrm{MSB}([\langle\boldsymbol{a},\boldsymbol{s}\rangle+e]_q)$。因为$e$是服从均值为0的离散高斯分布，所以$\mathrm{MSB}([\langle\boldsymbol{a},\boldsymbol{s}\rangle]_q)$可以从$\mathrm{MSB}([\langle\boldsymbol{a},\boldsymbol{s}\rangle+e]_q)$的值中以较高的概率获得。因此令$C_{\boldsymbol{s}}(\boldsymbol{x})=\mathrm{MSB}([\langle\boldsymbol{x},\boldsymbol{s}\rangle]_q)$为一个编码，可以自然的得到一个可访问的含噪版本。

回顾如下引理，估计了一个服从$D_{\mathbb{Z},\sigma}$分布变量的绝对值超过一个给定上界的概率。

引理3 令$B\geq\sigma$，那么

$$\Pr[|D_{\mathbb{Z},\sigma}|\geq B]\leq\frac{B}{\sigma}\exp\left(\frac{1}{2}-\frac{B^2}{2\sigma^2}\right) \tag{7}$$

上述引理推出对于$\delta\geq\sigma$，有

$$\Pr[|e|\leq\delta]>1-\frac{\delta}{\sigma}\exp\left(\frac{1}{2}-\frac{\delta^2}{2\sigma^2}\right) \tag{8}$$

令$\boldsymbol{a}$为在$\mathbb{Z}_q^n$上平均分布的变量，$\boldsymbol{s}$是一个固定向量，那么内积$\langle\boldsymbol{a},\boldsymbol{s}\rangle$是$\mathbb{Z}_q$上平均分布的变量。下面考察两类最重要比特的关系。

引理4 令$\delta\geq\sigma$, X是$\mathbb{Z}_q$上平均分布的变量，e的分布服从$D_{\mathbb{Z},\sigma}$。$\mathrm{MSB}(X)=\mathrm{MSB}(X+e)$的概率至少是$\left(1-\dfrac{4\delta}{q}\right)\left(1-\dfrac{\delta}{\sigma}\exp\left(\dfrac{1}{2}-\dfrac{\delta^2}{2\sigma^2}\right)\right)$。

证明 根据条件概率公式，有

$$\begin{aligned}\Pr[\mathrm{MSB}(X)=&\mathrm{MSB}(X+e)]\\ &\geq \Pr[\mathrm{MSB}(X)=\mathrm{MSB}(X+e)\mid |e|\leq\delta]\\ &\quad\times\Pr[|e|\leq\delta]\\ &=\Pr\Big[0\leq X+e<\frac{q}{2},0\leq X<\frac{q}{2}\\ &\quad\mid |e|\leq\delta\Big]\times\Pr[|e|\leq\delta]\\ &\quad+\Pr\Big[\frac{q}{2}\leq X+e<q,\frac{q}{2}\leq X<q\\ &\quad\mid |e|\leq\delta\Big]\times\Pr[|e|\leq\delta]\\ &\geq\Big(\Pr\Big[\delta\leq X<\frac{q}{2}-\delta\Big]\\ &\quad+\Pr\Big[\frac{q}{2}+\delta\leq X<q-\delta\Big]\Big)\times\Pr[|e|\leq\delta]\\ &\geq\left(\frac{q-4\delta}{q}\right)\times\Pr[|e|\leq\delta]\end{aligned}\tag{9}$$

所以，有

$$\begin{aligned}\Pr[\mathrm{MSB}(X)=\mathrm{MSB}(X+e)]\geq&\left(1-\frac{4\delta}{q}\right)\\ &\cdot\left(1-\frac{\delta}{\sigma}\exp\left(\frac{1}{2}-\frac{\delta^2}{2\sigma^2}\right)\right)\end{aligned}\tag{10}$$

证毕

推论1　设$s_1\neq 0$, $\boldsymbol{a}$在$\mathbb{Z}_q^n$中平均分布，x在$\mathbb{Z}_q$中平均分布，e按照$\mathbb{Z}_q$上均值为0标准差为σ的离散高斯分布，$\delta\geq\sigma$。记

$$P_1=\Pr[\mathrm{MSB}([\langle\boldsymbol{a},\boldsymbol{s}\rangle]_q)=\mathrm{MSB}([\langle\boldsymbol{a},\boldsymbol{s}+e]_q)\rangle]\tag{11}$$

$$\begin{aligned}P_2&=\Pr[\mathrm{MSB}([\langle(x,0,\cdots,0),\boldsymbol{s}\rangle]_q)\\ &=\mathrm{MSB}([\langle(x,0,\cdots,0),\boldsymbol{s}+e\rangle]_q)]\end{aligned}\tag{12}$$

那　么　$\Pr(P_1=P_2)\geq\left(1-\frac{4\delta}{q}\right)\left(1-\frac{\delta}{\sigma}\exp\left(\frac{1}{2}-\frac{\delta^2}{2\sigma^2}\right)\right)$　。

推论1推出如果σ相较于q足够小，那么$P_1=P_2$的概率就会足够大。

4　qLWE的分析

本节给出求解qLWE问题的算法。首先，将qLWE问题联系到一个隐藏数问题。具体地，用最重要比特函数来构造编码方案，自然得到一个可访问的含噪版本。然后，通过傅里叶学习算法来恢复目标向量。

4.1　隐藏数问题

本节考虑如下隐藏数问题。

定义 5　令$\alpha\neq 0$为一个$\mathbb{F}_p^*$中的秘密元素，$f:\mathbb{F}_p^*\to\{-1,1\}$为一个函数。隐藏数问题是通过对函数$C_\alpha(x)=f(x\alpha)$的查询访问，在多项式时间内计算$\alpha$。

Akavia[11]指出如果函数是傅里叶集聚的，那么存在傅里叶学习算法可以有效地求解该隐藏数问题。

定理1[11]　令A是一个计算定义在$\mathbb{F}_p$上函数的τ-重傅里叶系数的算法。那么对任意傅里叶集聚函数$f:\mathbb{F}_p\to\{-1,1\}$，存在一个调用A的算法有效求解$\mathbb{F}_p^*$上的隐藏数问题。

4.2　主要结果

按照一般的列表译码的框架，给出算法求解qLWE问题。首先，构造合适的纠错码，并且说明它是傅里叶集聚的。然后，说明如何访问含噪版本。最后，用列表译码算法求解。

定理2　令$\boldsymbol{s}\in\mathbb{Z}_q^n$。如果对任意输入$\boldsymbol{x}\in\mathbb{Z}_q^n$，可以通过谕示$\mathcal{O}$获得qLWE样本$\langle\boldsymbol{x},\boldsymbol{s}\rangle+e$，这里$e$服从$\mathbb{Z}_q$上期望为0标准差为$\sigma$的离散高斯分布。并且$\sigma$相比于$q$足够小。那么，存在一个概率多项式时间算法求解$\mathbb{Z}_q^n$上的qLWE问题。

证明　根据谕示可以判定是否$s_i=0$，这里$i=1,2,\cdots,n$。只需选取$\mathbb{Z}_q$上的平均分布的元素x，令$\boldsymbol{v}$是$\mathbb{Z}_q^n$中第i个分量是x其余分量为0的向量。然后通过谕示计算$\langle\boldsymbol{v},\boldsymbol{s}\rangle+e$的结果即可判断。

下面，逐个分量恢复向量$\boldsymbol{s}$。不失一般性，不妨假设$s_1\neq 0$并求解。

首先，通过隐藏数问题定义合适的纠错码。对任意元素$x\in\mathbb{Z}_q$，定义编码

$$\begin{aligned}&C_{\boldsymbol{s}}:\mathbb{Z}_q\to\{\pm1\}\\ &x\mapsto\mathrm{MSB}([\langle(x,0,\cdots,0),\boldsymbol{s}\rangle]_q)\end{aligned}\tag{13}$$

对$x\in\mathbb{Z}_q$，定义$C_{\boldsymbol{s}}$的含噪版本$C'_{\boldsymbol{s}}$为

$$\begin{aligned}&C'_{\mathrm{s}}:\mathbb{Z}_q\to\{\pm1\}\\ &x\mapsto\mathrm{MSB}(\mathcal{O}(x,0,\cdots,0))\end{aligned}\tag{14}$$

令ϵ是$\mathrm{MSB}([\langle(x,0,\cdots,0),\boldsymbol{s}\rangle+e]_q)=\mathrm{MSB}([\langle(x,0,\cdots,0),\boldsymbol{s}\rangle]_q)$的概率优势，$\beta$是$\mathrm{MSB}([\langle(x,0,\cdots,0),\boldsymbol{s}\rangle]_q)$的偏差。那么，$\left|\Pr_x[C'_{\boldsymbol{s}}(x)=C_{\boldsymbol{s}}(x)]\right|\geq\beta+\epsilon$。

Akavia[11]证明了MSB函数是傅里叶集聚的，从而$C_{\boldsymbol{s}}$是傅里叶集聚的。根据定理1，存在概率多项式时间算法计算s_1。其余分量类似可得。　证毕

4.3　一些讨论

4.3.1　主定理的另一个证明

Galbraith等人[6]提出可以用多变量的隐藏数问题及相应的列表译码算法来得出同样的结论。对任意元素$\boldsymbol{x}\in\mathbb{Z}_q^n$，构造编码

$$\left.\begin{aligned}&C_{\boldsymbol{s}}:\mathbb{Z}_q^n\to\{\pm1\}\\ &\boldsymbol{x}\mapsto\mathrm{MSB}([\langle\boldsymbol{x},\boldsymbol{s}\rangle]_q)\end{aligned}\right\}\tag{15}$$

根据谕示$\mathcal{O}$的输出，对于$\boldsymbol{x}\in\mathbb{Z}_q^n$，定义可访问的含噪版本

$$\left.\begin{aligned}&C'_{\boldsymbol{s}}:\mathbb{Z}_q\to\{\pm1\}\\ &\boldsymbol{x}\mapsto\mathrm{MSB}(\mathcal{O}(\boldsymbol{x}))\end{aligned}\right\}\tag{16}$$

文献[6]中的下述引理表明$C_{\boldsymbol{s}}$是傅里叶集聚的。

引理5[6] 令$f:\mathbb{F}_q\to\{-1,1\}$, $\boldsymbol{s}=(s_1,s_2,\cdots,s_n)\in\mathbb{Z}_q^n$并且并非所有分量$s_i=0$, $f_{\boldsymbol{s}}:\mathbb{Z}_q^n\to\{\pm1\}$是函数使得$f_{\boldsymbol{s}}(\boldsymbol{x})=f(\boldsymbol{s}\cdot\boldsymbol{x})$。那么$\text{Heavy}_\tau(f)=\{c_1,c_2,\cdots,c_t\}$当且仅当$\text{Heavy}_\tau(f_{\text{s}})=\{(c_is_1,c_is_2,\cdots,c_is_n)|1\le i\le t\}$。并且，$f$是傅里叶集聚的当且仅当$f_{\boldsymbol{s}}$是傅里叶集聚的。

根据上述引理，可以应用相应的傅里叶学习算法来恢复未知向量。

4.3.2 主要结果的另一个解读

Micciancio等人[12]注意到了Goldreich-Levin算法与LWE和其判定版本之间延续样本的规约的关系。4.2节的主要结果还可以用来构造单向函数的困难比特。给定单向函数$f:\mathbb{Z}_q^n\to\mathbb{Z}_q^*$，考虑衍生的单向函数

$$\begin{aligned}&g:\mathbb{Z}_q^n\times\mathbb{Z}_q^n\to\mathbb{Z}_q^*\times\mathbb{Z}_q^n\\&(\boldsymbol{x},\boldsymbol{s})\mapsto(f(\boldsymbol{x}),\boldsymbol{s})\end{aligned}\tag{17}$$

那么，$\text{MSB}([\langle\boldsymbol{x},\boldsymbol{s}\rangle]_q)$是函数$g$的一个困难比特。即，如果存在谕示以不可忽略的优势计算$\text{MSB}([\langle\boldsymbol{x},\boldsymbol{s}\rangle]_q)$，那么可以采用列表译码算法对$g$求逆。

5 结论

许多研究者注意到当攻击者可以选择样本时，LWE问题是容易求解的。但是，目前没有见到一个完整的求解算法。本文分析了在查询访问模型中LWE问题的变形qLWE问题，并给出了完整的求解算法。一方面，有规约证据表明LWE问题的困难性，另一方面，qLWE问题是可以有效求解的。求解算法的关键步骤是应用傅里叶学习方法解决隐藏数问题。该结果也可以解读为建立了一类单向函数的困难比特。

参 考 文 献

[1] KEARNS M J and VAZIRANI U V. An Introduction to Computational Learning Theory[M]. Cambridge, London England: The MIT Press, 1994.

[2] BLUM A, KALAI A, and WASSERMAN H. Noise-tolerant learning, the parity problem, and the statistical query model[J]. *Journal of the ACM*, 2003, 50(4): 506–519. doi: 10.1145/792538.792543.

[3] PIETRZAK K. Cryptography from learning parity with noise[C]. The 38th International Conference on Current Trends in Theory and Practice of Computer Science, Špindlerův Mlýn, Czech Republic, 2012: 99–114.

[4] REGEV O. On lattices, learning with errors, random linear codes, and cryptography[C]. The 37th Annual ACM Symposium on Theory of Computing, Baltimore, USA, 2005: 84–93.

[5] BONEH D and VENKATESAN R. Hardness of computing the most significant bits of secret keys in Diffie-Hellman and related schemes[C]. The 16th Annual International Cryptology Conference on Advances in Cryptology, Santa Barbara, USA, 1996: 129–142.

[6] GALBRAITH S D and SHANI B. The multivariate hidden number problem[C]. The 8th International Conference on Information Theoretic Security, Lugano, Switzerland, 2015: 250–268.

[7] GOLDREICH O and LEVIN L A. A hard-core predicate for all one-way functions[C]. The 21st Annual ACM Symposium on Theory of Computing, Seattle, USA, 1989: 25–32.

[8] KUSHILEVITZ E and MANSOUR Y. Learning decision trees using the Fourier spectrum[C]. The 23rd Annual ACM Symposium on Theory of Computing, New Orleans, USA, 1991: 455–464.

[9] AKAVIA A, GOLDWASSER S, and SAFRA S. Proving hard-core predicates using list decoding[C]. The 44th Annual IEEE Symposium on Foundations of Computer Science, Cambridge, USA, 2003: 146–157.

[10] GALBRAITH S D, LAITY J, and SHANI B. Finding significant Fourier coefficients: Clarifications, simplifications, applications and limitations[J]. *Chicago Journal of Theoretical Computer Science*, 2018, 6: 1–38.

[11] AKAVIA A. Solving hidden number problem with one bit oracle and advice[C]. The 29th Annual International Cryptology Conference on Advances in Cryptology, Santa Barbara, USA, 2009: 337–354.

[12] MICCIANCIO D and MOL P. Pseudorandom knapsacks and the sample complexity of LWE search-to-decision reductions[C]. The 31st Annual Conference on Advances in Cryptology, Santa Barbara, USA, 2011: 465–484.

王明强：男，1970年生，教授，研究方向为算法数论和公钥密码学.
庄金成：男，1987年生，教授，研究方向为算法数论和公钥密码学.