关于非对称含错学习问题的困难性研究

张 江* 范淑琴

(密码科学技术国家重点实验室 北京 100878)

摘 要：由于基于最坏情况困难假设等优点，基于格的密码被认为是最具前景的抗量子密码研究方向。作为格密码的常用的两个主要困难问题之一，含错学习(LWE)问题被广泛用于密码算法的设计。为了提高格密码算法的性能，Zhang等人(2019)提出了非对称含错学习问题，该文将从理论上详细研究非对称含错学习问题和标准含错学习问题关系，并证明在特定错误分布下非对称含错学习问题和含错学习问题是多项式时间等价的，从而为基于非对称含错学习问题设计安全的格密码算法奠定了理论基础。

关键词：抗量子密码；格密码；含错学习问题

中图分类号：TN918, TP309.7 **文献标识码**：A **文章编号**：1009-5896(2020)02-0327-06

DOI: 10.11999/JEIT190685

On the Hardness of the Asymmetric Learning With Errors Problem

ZHANG Jiang FAN Shuqin

(*State Key Laboratory of Cryptology*, *Beijing* 100878, *China*)

Abstract: Due to the advantages such as the worst-case hardness assumption, lattice-based cryptography is believed to the most promising research direction in post-quantum cryptography. As one of the two main hard problems commonly used in lattice-based cryptography, Learning With Errors (LWE) problem is widely used in constructing numerous cryptosystems. In order to improve the efficiency of lattice-based cryptosystems, Zhang *et al*. (2019) introduced the Asymmetric LWE (ALWE) problem. In this paper, the relation between the ALWE problem and the standard LWE problem is studied, and it shows that for certain error distributions the two problems are polynomially equivent, which paves the way for constructing secure lattice-based cryptosystems from the ALWE problem.

Key words: Post-qauntum cryptography; Lattice-based cryptography; Learning With Errors (LWE)

1 引言

当前，公钥密码算法，如公钥加密、数字签名和密钥交换已经大规模地应用于实际生活中，提供着各式各样的信息安全保障。但几乎所有应用中的公钥密码算法的安全性都建立在分解大整数或者求解离散对数问题的困难性之上。换句话说，如果没有多项式时间的算法能够分解大整数或者求解离散对数问题，那么这些公钥密码算法就能够实现安全的功能。不幸的是美国科学家Shor[1]在1996年提出了能够在多项式时间分解大整数或者求解离散对数问题的量子算法。近年来，随着量子计算技术的快速发展，研究能够抵抗量子计算机攻击的公钥密码算法——抗量子密码，已经迫在眉睫。事实上，世界各国政府和组织都相应地发起了重大的研究计划来发展能够抵抗量子计算机攻击的密码算法，如欧盟的SAFEcrypto项目、日本的CryptoMath-CREST密码项目等。特别地，美国国家安全局(NSA)已于2015年8月宣布了抗量子密码算法的迁移计划[2]。同年，美国国家标准与技术研究院(NIST)举行了“后量子世界的网络安全研讨会”，并启动了面向全世界征集抗量子公钥密码算法的计划[3]。2018年5月，中国科协发布了我国面临的60个重大科技难题[4]，“抗量子密码算法设计”是信息科技领域入选的6个重大科技难题之一。

基于格的密码、基于编码的密码、基于多变量的密码和基于杂凑函数的密码是当前国际上公认的

收稿日期：2019-09-14；改回日期：2019-11-20；网络出版：2019-11-29
*通信作者： 张江 jiangzhang09@gmail.com
基金项目：国家重点研发计划(2017YFB0802005, 2018YFB0804105)，国家自然科学基金(61602046, 61932019)，中国科协“青年人才托举工程”(2016QNRC001)

Foundation Items: The National Key Research and Development Program of China (2017YFB0802005, 2018YFB0804105), The National Natural Science Foundation of China (61602046, 61932019), The Young Elite Scientists Sponsorship Program by China Association for Science and Technology (2016QNRC001)

4个主要的抗量子密码研究方向。由于具有基于最坏情况的困难假设、高效率以及极大的多样性等优点，基于格的密码被认为是最具前景的抗量子密码研究方向。含错学习问题(Learning With Errors, LWE)和小整数解问题(Small Integer Solutions, SIS)是基于格的密码中两个常用的困难问题，其中含错学习问题是由Regev[5]在2005年提出，而小整数解问题则可追溯到Ajtai[6]的开创性工作。含错学习问题和小整数解问题的定义都非常简单，且都和解整数模方程有关。特别地，令n,m,q是任意正整数，令χ_α是定义在整数Z上以实数α为参数的概率分布。计算性含错学习问题$\mathrm{LWE}_{n,m,q,\alpha}$要求给定元组$(\boldsymbol{A},\boldsymbol{b}=\boldsymbol{A}\boldsymbol{s}+\boldsymbol{e})\in Z_q^{m\times n}\times Z_q^m$，输出$\boldsymbol{s}\in Z_q^n$，其中$\boldsymbol{A}\leftarrow Z_q^{m\times n},\boldsymbol{s}\leftarrow Z_q^n,\boldsymbol{e}\leftarrow\chi_\alpha^m$。而小整数解问题$\mathrm{SIS}_{n,m,q,\beta}$则要求给定矩阵$\boldsymbol{A}\leftarrow Z_q^{m\times n}$和实数$\beta$，计算使得$\boldsymbol{x}\in Z_q^m$使得$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}=0 \bmod q$且$\|\boldsymbol{x}\|_\infty\le\beta$。显然，以上两个问题非常相似。事实上，在一定意义上含错学习问题和小整数解问题是互为对偶问题。此外，以上两个问题在平均情况下的困难性被严格归约到了格上问题在最坏情况下的困难性之上[5,6]。一般来说，含错学习问题主要用于加密算法的设计，而小整数解问题则用于签名算法的设计。为了提高方案的效率，文献中还经常使用含错学习问题的一个变种问题，即正规形含错学习问题。该类问题要求秘密向量$\boldsymbol{s}\in Z_q^n$取自于分布χ_α。特别地，正规形含错学习问题和标准学习问题在多项式时间的归约下是等价的。文献中常用的两种错误分布为高斯分布和二项分布。

虽然基于含错学习问题和小整数解问题通常都比较简单，且具有较高的计算效率，但像基于编码等数学问题的抗量子方案一样，基于格的密码方案的参数，如公钥、密文等都比较大。为了更好地折中密码算法的安全性和参数大小，Zhang等人[7]提出了非对称含错学习问题(Asymmetric LWE, ALWE)和非对称小整数解问题(Asymmetric SIS, ASIS)。在定义上，非对称含错学习问题修改了标准含错学习问题的实例分布，而非对称小整数解问题则只是修改了标准小整数解问题的解分布。由此造成的差别是可以很容易证明非对称小整数解问题和标准小整数解问题在多项式归约下是等价的，而对于非对称含错学习问题，却不那么容易证明标准含错学习问题和非对称学习问题的困难关系[7]。

本文将正式研究非对称学习问题和标准学习问题的困难关系，并证明对于满足特定“加和”性质的分布χ，标准含错学习问题和非对称含错学习问题在一定参数下是等价的。简单来说，我们说一个分布χ具有“加和”性质。即给定分布参数α_1和$\alpha_2>\alpha_1$，存在多项式时间的算法$S(\cdot,\cdot)$，使得对于$x_1\leftarrow\chi_{\alpha_1}$和$x_2\leftarrow S(\alpha_1,\alpha_2)$，有$x=x_1+x_2$服从分布$\chi_{\alpha_2}$。注意到算法$S(\cdot,\cdot)$并不以$x_1$作为输入，这一点对于本文的证明非常重要。结合以上“加和”性质，以及(非对称)含错学习问题的同态性质，本文证明了对于任何“加和”的分布χ，非对称含错学习问题和标准含错学习问题是多项式时间等价的。特别地，本文证明了两类格密码常用的概率分布(即高斯分布和中心二项分布)均满足“加和”性质。换句话说，对于特定参数的高斯分布和中心二项分布，标准含错学习问题和非对称含错学习问题在多项式时间意义下是等价的。这就为基于高斯分布和二项分布的非对称含错学习问题设计基于格的密码算法奠定了严格的理论基础。

2 基础知识

令κ为全文中默认的安全参数，所有其他的参数都隐含地是关于κ的函数。令符号$\log_\gamma(\cdot)$表示以γ为底的对数函数。当$\gamma=2$时，简写为$\mathrm{lb}(\cdot)$。如果没有特别说明，符号$O(\cdot)$和$\omega(\cdot)$表示标准的渐近函数。符号$\mathrm{poly}(n)$表示关于变量n的任意多项式函数，即存在常数c使得$\mathrm{poly}(n)=O(n^{\mathrm{c}})$。对于函数$f(n)$和函数$g(n)$，如果存在常数c使得$f(n)=O(g(n)\cdot\mathrm{lb}^{\mathrm{c}}n)$，记$f(n)=\tilde{O}(g(n))$。特别地，如果对于任意$\mathrm{c}>0$都存在足够大的整数$N$使得对于所有的$n>N$都有$f(n)<1/n^{\mathrm{c}}$成立，那么称函数$f$关于变量$n$是可忽略的。本文用$\mathrm{negl}(\cdot)$表示未明确定义的可忽略函数。

符号Z和R分别代表整数集合和实数集合。对于分布D和有限集合S，符号$v\leftarrow_r D$表示从分布D中随机选取元素，而符号$v\leftarrow_r S$则表示随机均匀地选取集合S中的元素。如果随机变量v服从分布D，简记为$v\sim D$。向量默认写成列的形式并用小写加粗字母表示(例如，$\boldsymbol{x}$)，而矩阵则视为列向量的集合并用大写加粗字母表示(例如$\boldsymbol{X}$)。符号$\boldsymbol{x}^{\mathrm{T}}$和$\boldsymbol{X}^{\mathrm{T}}$分别表示向量$\boldsymbol{x}$和矩阵$\boldsymbol{X}$的转置。符号$||\boldsymbol{x}||=\sqrt{\sum {x_i}^2}$和$||\boldsymbol{x}||_\infty=\max\{|x_i|\}$分别表示取向量的第2范数$\ell_2$和无穷范数$\ell_\infty$，其中$\boldsymbol{x}=(x_1,x_2,\cdots)$。

3 含错学习问题

令n,m,q是任意正整数，令χ_α是定义在整数Z上以实数α为参数的概率分布。计算性含错学习问题$\mathrm{LWE}_{n,m,q,\alpha}$要求给定元组

$$(\boldsymbol{A},\boldsymbol{b}=\boldsymbol{A}\boldsymbol{s}+\boldsymbol{e})\in Z_q^{m\times n}\times Z_q^m \tag{1}$$

输出$\boldsymbol{s}\in Z_q^n$，其中$\boldsymbol{A}\leftarrow Z_q^{m\times n},\boldsymbol{s}\leftarrow Z_q^n,\boldsymbol{e}\leftarrow\chi_\alpha^m$。类似地，判定性含错学习问题$\mathrm{DLWE}_{n,m,q,\alpha}$则要求将

元组$(\boldsymbol{A},\boldsymbol{b})\in Z_q^{m\times n}\times Z_q^m$和选择于$Z_q^{m\times n}\times Z_q^m$上均分分布的元组区分开来。对于特定的分布和参数，Regev[5]证明了如果存在一个多项式时间的算法能够求解$\mathrm{LWE}_{n,m,q,\alpha}$问题，那么存在多项式时间的量子算法能够求解任意格上最困难的近似最短独立向量问题(Shortest Vector Problem, SVP)，且判定性含错学习问题和计算性含错学习问题是等价的。此外，对于秘密向量$\boldsymbol{s}\in Z_q^n$同样取自于分布χ_α的正规形含错学习问题，Applebaum等人[8]证明了其与标准的含错学习问题是等价的。

为了提高基于含错学习问题的密码算法的效率，Zhang等人[7]提出了非对称含错学习问题作为正规形含错学习问题的变种。令n,m,q是任意正整数。令χ_{α_1}和χ_{α_2}是定义在整数$\mathbb{Z}$上分别以实数α_1和α_2为参数的概率分布。计算性非对称含错学习问题$\mathrm{ALWE}_{n,m,q,\alpha_1,\alpha_2}$要求给定元组

$$(\boldsymbol{A},\boldsymbol{b}=\boldsymbol{A}\boldsymbol{s}+\boldsymbol{e})\in Z_q^{m\times n}\times Z_q^m \tag{2}$$

输出$\boldsymbol{s}\in Z_q^n$，其中$\boldsymbol{A}\leftarrow Z_q^{m\times n},\boldsymbol{s}\leftarrow\chi_{\alpha_1}^n,\boldsymbol{e}\leftarrow\chi_{\alpha_2}^m$。类似地，也可以定义判定性非对称含错学习问题。Zhang等人[7]对已知求解含错学习问题的最高效的算法进行了研究发现以α_1,α_2为高斯参数的非对称含错学习问题和以$\sqrt{\alpha_1,\alpha_2}$为高斯参数的标准含错学习问题的困难性大致相等，即从攻击的角度间接揭示非对称含错学习问题和标准含错学习问题在一定程度上是等价的。接下来，本文将研究含错学习问题和标准学习问题的困难性间的关系。

4　非对称含错学习问题的困难性

这一节中，将考虑定义在具有“加和”性质的概率分布上的非对称含错学习问题。

定义1　对于任意以α为参数的概率分布χ_α和参数$0<\alpha_1\le\alpha_2$，如果存在一个多项式时间的算法$\mathrm{S}_\chi(\cdot,\cdot)$，如果分布$D_{\alpha_1,\alpha_2}=\{x=x_1+x_2\in Z|x_1\leftarrow\chi_{\alpha_1},x_2\leftarrow S_\chi(\alpha_1,\alpha_2)\}$和分布$\chi_\alpha$是统计不可区分的，则称分布$\chi_\alpha$具有“加和”性质。

对于具有“加和”性质的概率分布，有如下结论：

定理1　对于任意实数α_1和α_2，和定义在概率分布χ_{α_1}和χ_{α_2}上的非对称含错学习问题$\mathrm{ALWE}_{n,m,q,\alpha_1,\alpha_2}$，如果分布$\chi_\alpha$具有“加和”性质，那么有$\mathrm{ALWE}_{n,m,q,\alpha_1,\alpha_2}$问题和标准含错学习问题在多项式时间是等价的，即有如下困难性不等式在多项式时间的归约下是成立的

$$\begin{aligned}\mathrm{LWE}_{n,m,q,\min(\alpha_1,\alpha_2)}&\le\mathrm{ALWE}_{n,m,q,\alpha_1,\alpha_2}\\&\le\mathrm{LWE}_{n,m,q,\max(\alpha_1,\alpha_2)}\end{aligned}\tag{3}$$

证明　显然，为了证明定理1，只需要证明如下两个结论成立即可：

(1) 如果存在多项式时间的算法能够解决$\mathrm{ALWE}_{n,m,q,\alpha_1,\alpha_2}$问题，那么存在多项式时间的算法能够解决$\mathrm{LWE}_{n,m,q,\min(\alpha_1,\alpha_2)}$问题；

(2) 如果存在多项式时间的算法能够解决$\mathrm{LWE}_{n,m,q,\max(\alpha_1,\alpha_2)}$问题，那么存在多项式时间的算法能够解决$\mathrm{ALWE}_{n,m,q,\alpha_1,\alpha_2}$问题。

接下来，将分别证明以上两个结论。为了方便证明，不妨先假设$0<\alpha_1\le\alpha_2$。首先，来证明结论(1)成立，即如果存在多项式时间的算法$\mathcal{A}$能够解决$\mathrm{ALWE}_{n,m,q,\alpha_1,\alpha_2}$问题，那么存在多项式时间的算法$\mathcal{B}$能够解决$\mathrm{LWE}_{n,m,q,\min(\alpha_1,\alpha_2)}$问题。特别地，给定$\mathrm{ALWE}_{n,m,q,\alpha_1,\alpha_2}$的安全参数$\mathrm{parms}=(1^\kappa,\alpha_1,\alpha_2)$，实例元组$(\boldsymbol{A},\boldsymbol{b}=\boldsymbol{A}\boldsymbol{s}+\boldsymbol{e})\in Z_q^{m\times n}\times Z_q^m$，算法$\mathcal{A}(\mathrm{parms},\boldsymbol{A},\boldsymbol{b})$能够以不可忽略的概率$\varepsilon$输出$\boldsymbol{s}\in Z_q^n$，其中$\kappa$是安全参数，$\boldsymbol{A}\leftarrow Z_q^{m\times n},\boldsymbol{s}\leftarrow\chi_{\alpha_1}^n,\boldsymbol{e}\leftarrow\chi_{\alpha_2}^m$。现在构造算法$\mathcal{B}$使得给定$\mathrm{LWE}_{n,m,q,\min(\alpha_1,\alpha_2)}=\mathrm{LWE}_{n,m,q,\alpha_1}$的公共参数$\mathrm{parms}_1=(1^\kappa,\alpha_1)$和实例元组$(\boldsymbol{A}_1,\boldsymbol{b}_1=\boldsymbol{A}_1\boldsymbol{s}_1+\boldsymbol{e}_1)\in Z_q^{m\times n}\times Z_q^m$，算法$\mathcal{B}(\mathrm{parms}_1,\boldsymbol{A}_1,\boldsymbol{b}_1)$能够以$\varepsilon-\mathrm{negl}(\kappa)$输出$\boldsymbol{s}_1\in Z_q^n$，$\boldsymbol{A}_1\leftarrow\mathbb{Z}_q^{m\times n},\boldsymbol{s}\leftarrow\chi_{\alpha_1}^n,\boldsymbol{e}\leftarrow\chi_{\alpha_1}^m$。算法$\mathcal{B}$进行如下运算：

(1) 运行m次算法$e_{2,i}\leftarrow\mathrm{S}_\chi(\alpha_1,\alpha_2)$得到向量$\boldsymbol{e}_2=(e_{2,1},e_{2,2},\cdots,e_{2,m})$；

(2) 令$\boldsymbol{A}=\boldsymbol{A}_1\in Z_q^{m\times n},\boldsymbol{b}=\boldsymbol{b}_1+\boldsymbol{e}_2\in Z_q^m$；

(3) 令$\mathrm{parms}=(1^\kappa,\alpha_1,\alpha_2)$；

(4) 运行算法$\boldsymbol{s}\leftarrow\mathcal{A}(\mathrm{parms},\boldsymbol{A},\boldsymbol{b})$，并输出向量$\boldsymbol{s}\in Z_q^n$。

由等式

$$\boldsymbol{b}=\boldsymbol{b}_1+\boldsymbol{e}_2=\boldsymbol{A}_1\boldsymbol{s}_1+\underbrace{\boldsymbol{e}_1+\boldsymbol{e}_2}_{=\boldsymbol{e}}=\boldsymbol{A}_1\boldsymbol{s}_1+\boldsymbol{e}\bmod q\tag{4}$$

和分布χ_α的“加和”性质可知，$\boldsymbol{e}=\boldsymbol{e}_1+\boldsymbol{e}_2$的分布选自于$\chi_{\alpha_2}$中元素的分布是统计不可区分的。换句话说，元组$(\boldsymbol{A}=\boldsymbol{A}_1,\boldsymbol{b}=\boldsymbol{b}_1+\boldsymbol{e}_2)\in Z_q^{m\times n}\times Z_q^m$的分布和$\mathrm{ALWE}_{n,m,q,\alpha_1,\alpha_2}$问题的实例分布是统计不可区分的，这就意味着算法$\mathcal{B}$能够以$\varepsilon-\mathrm{negl}(\kappa)$输出$\boldsymbol{s}=\boldsymbol{s}_1\in\mathbb{Z}_q^n$。结论(1)得证。

现在来证明结论(2)成立，即如果存在多项式时间的算法$\mathcal{A}$能够解决问题$\mathrm{LWE}_{n,m,q,\max(\alpha_1,\alpha_2)}$，那么存在多项式时间的算法$\mathcal{B}$能够解决$\mathrm{ALWE}_{n,m,q,\alpha_1,\alpha_2}$问题。如前述一样，为了便于证明，仍然假设$0<\alpha_1\le\alpha_2$。特别地，给定$\mathrm{LWE}_{n,m,q,\max(\alpha_1,\alpha_2)}=\mathrm{LWE}_{n,m,q,\alpha_2}$的公共参数$\mathrm{parms}=(1^\kappa,\alpha_2)$，实例元组$(\boldsymbol{A},\boldsymbol{b}=\boldsymbol{A}\boldsymbol{s}+\boldsymbol{e})\in Z_q^{m\times n}\times Z_q^m$，算法$\mathcal{A}(\mathrm{parms},\boldsymbol{A},\boldsymbol{b})$能够以不可忽略的概率$\varepsilon$输出$\boldsymbol{s}\in\mathbb{Z}_q^n$，其中$\kappa$是安全参数，$\boldsymbol{A}\leftarrow Z_q^{m\times n},\boldsymbol{s}\leftarrow\chi_{\alpha_2}^n,\boldsymbol{e}\leftarrow\chi_{\alpha_2}^m$。现在构造算法$\mathcal{B}$使得给定$\mathrm{ALWE}_{n,m,q,\alpha_1,\alpha_2}$的公共参数$\mathrm{parms}_1=$

$(1^\kappa, \alpha_1, \alpha_2)$ 和实例元组$(\boldsymbol{A}_1, \boldsymbol{b}_1 = \boldsymbol{A}_1\boldsymbol{s}_1 + \boldsymbol{e}_1) \in Z_q^{m\times n} \times Z_q^m$，算法$\mathcal{B}(\text{parms}_1, \boldsymbol{A}_1, \boldsymbol{b}_1)$能够以$\varepsilon - \text{negl}(\kappa)$输出$\boldsymbol{s}_1 \in Z_q^n$，$\boldsymbol{A}_1 \leftarrow Z_q^{m\times n}, \boldsymbol{s} \leftarrow \chi_{\alpha_1}^n, \boldsymbol{e} \leftarrow \chi_{\alpha_2}^m$。算法$\mathcal{B}$进行如下运算：

(1) 运行n次算法$s_{2,i} \leftarrow \text{S}_\chi(\alpha_1, \alpha_2)$得到向量$\boldsymbol{s}_2 = (s_{2,1}, s_{2,2}, \cdots, s_{2,n})$；

(2) 令$\boldsymbol{A} = \boldsymbol{A}_1 \in Z_q^{m\times n}, \boldsymbol{b} = \boldsymbol{b}_1 + \boldsymbol{A}_1\boldsymbol{s}_2 \in Z_q^m$；

(3) 令$\text{parms} = (1^\kappa, \alpha_2)$；

(4) 运行算法$\boldsymbol{s} \leftarrow \mathcal{A}(\text{parms}, \boldsymbol{A}, \boldsymbol{b})$，并得到向量$\boldsymbol{s} \in Z_q^n$；

(5) 计算并输出$\boldsymbol{s}_1 = \boldsymbol{s} - \boldsymbol{s}_2$。

由等式

$$\boldsymbol{b} = \boldsymbol{b}_1 + \boldsymbol{A}_1\boldsymbol{s}_2 = \boldsymbol{A}_1\underbrace{(\boldsymbol{s}_1 + \boldsymbol{s}_2)}_{=\boldsymbol{s}} + \boldsymbol{e}_1 = \boldsymbol{A}_1\boldsymbol{s} + \boldsymbol{e}_1 \bmod q \tag{5}$$

和分布χ_α的“加和”性质可知，$\boldsymbol{s} = \boldsymbol{s}_1 + \boldsymbol{s}_2$的分布选自于$\chi_{\alpha_2}$中元素的分布是统计不可区分的。换句话说，元组$(\boldsymbol{A} = \boldsymbol{A}_1, \boldsymbol{b} = \boldsymbol{b}_1 + \boldsymbol{A}_1\boldsymbol{s}_2) \in Z_q^{m\times n} \times Z_q^m$的分布和$\text{ALWE}_{n,m,q,\alpha_1,\alpha_2}$问题的实例分布是统计不可区分的，这就意味着算法$\boldsymbol{s} \leftarrow \mathcal{A}(\text{parms}, \boldsymbol{A}, \boldsymbol{b})$能够以$\varepsilon - \text{negl}(\kappa)$输出$\boldsymbol{s} = \boldsymbol{s}_1 + \boldsymbol{s}_2 \in Z_q^n$，从而算法$\mathcal{B}$能够以$\varepsilon - \text{negl}(\kappa)$输出$\boldsymbol{s}_1 = \boldsymbol{s} - \boldsymbol{s}_2 \in Z_q^n$。结论(2)得证。

对于$\alpha_1 \geq \alpha_2 > 0$的情况，仍然可以按照上述方式利用分布$\chi_\alpha$的“加和”性质和(非对称)含错学习问题的“加法同态”性质证明结论(1)和结论(2)成立。由此定理1得证。

5 高斯分布和中心二项分布

自Micciancio等人[9]利用高斯分布定义了一个新的格参数——平滑参数——之后，高斯分布就与格密码的研究密不可分。特别地，许多格密码常用数学问题(例如含错学习问题和小整数解问题)的困难性证明都与高斯分布密切相关。事实上，格密码独有的优点——密码算法的平均情况安全性和数学问题的最坏情况困难性的联系——就依赖于高斯分布的优良性质。然而，尽管高斯分布非常有利于格密码的理论研究，但高斯分布在程序实现过程中却存在一定的技术难点，使得输出的样本常与随机数、时间和能量的消耗，以及计算精度有关，从而导致实现比较复杂，且容易遭受侧信道攻击。为了便于实现和抵抗侧性道攻击，面向实用的格密码算法常选择使用与高斯分布相近的中心二项分布来替换原有的高斯分布(如图1)。因此，研究基于高斯分布或中心二项分布的非对称含错学习的困难性至关重要。幸运的是，高斯分布或中心二项分布在一定条件下都具有“加和”的性质，因此定理1中的结论可直接应用于基于高斯分布或中心二项分布的非对称含错学习。接下来，给出高斯分布和中心二项分布的定义。

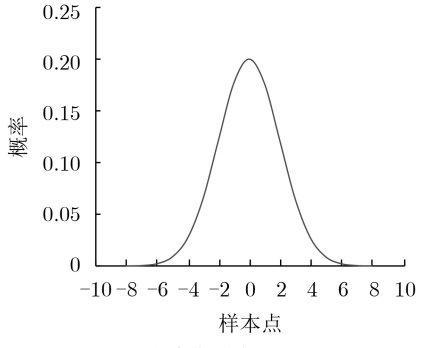
对于任意正实数$s \in R$和向量$\boldsymbol{c} \in R^m$，定义在R^m上以$\boldsymbol{c}$为中心、s为标准差的连续高斯函数为$\rho_{s,\boldsymbol{c}}(\boldsymbol{x}) = \left(\frac{1}{\sqrt{2\pi s^2}}\right)^m \exp\left(\frac{-||\boldsymbol{x} - \boldsymbol{c}||^2}{2s^2}\right)$，记对应的连续高斯分布为$D_{s,\boldsymbol{c}}$。对于任意格$\Lambda \subseteq Z^m$，定义$\rho_{s,\boldsymbol{c}}(\Lambda) = \sum_{\boldsymbol{x}\in\Lambda}\rho_{s,\boldsymbol{c}}(\boldsymbol{x})$。由此，可以诱导出定义在$\Lambda$上以$\boldsymbol{c}$为中心、$s$为标准差的离散高斯分布$D_{\Lambda,s,\boldsymbol{c}}(\boldsymbol{y}) = \rho_{s,\boldsymbol{c}}(\boldsymbol{y})/\rho_{s,\boldsymbol{c}}(\Lambda)$。当下标$s = 1$(或$\boldsymbol{c} = \boldsymbol{0}$)时，通常忽略相应的下标。

对于任意正整数$k \in Z$，定义以$k \in Z$为参数的中心二项分布为

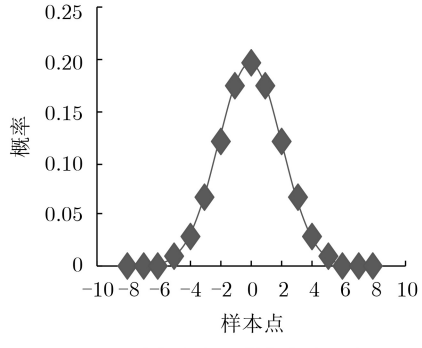
$$B_k = \left\{\sum_{i=1}^k (b_{i,0} - b_{i,1}) \,|\, b_{i,0}, b_{i,1} \leftarrow \{0,1\}, \; i \in \{1,2,\cdots,k\}\right\} \in \{-k, -k+1, \cdots, k\} \tag{6}$$

显然，中心二项分布具有“加和”性质，这是因为总有$B_{k_1+k_2} = B_{k_1} + B_{k_2}$恒成立。换句话说，有引理1。

引理1 中心二项分布B_k具有“加和”性质。



(a) 高斯分布 (s=2, c=0)



(b) 中心二项分布 (k=8)

图 1 高斯分布和二项分布

此外，定义在实数上以原点为中心的连续高斯分布也存在“加和”性质，因为如果向量$\boldsymbol{x}_1 \in R^m$服从以0为中心，s_1为标准差的连续高斯分布，向量$\boldsymbol{x}_2 \in \mathbb{R}^m$服从以0为中心，$s_2$为标准差的连续高斯分布，令$s=\sqrt{{s_1}^2+{s_2}^2}$那么向量$\boldsymbol{x}=\boldsymbol{x}_1+\boldsymbol{x}_2$的密度函数为

$$\begin{aligned}\mathrm{DF}(\boldsymbol{x}) &= \int_{\boldsymbol{x}_1} \rho_{s_1}(\boldsymbol{x}_1)\rho_{s_2}(\boldsymbol{x}-\boldsymbol{x}_1)\mathrm{d}\boldsymbol{x}_1 \\ &= \left(\frac{1}{2\pi s_1 s_2}\right)^m \int_{\boldsymbol{x}_1} \exp\left(\frac{-\|\boldsymbol{x}_1\|^2}{2{s_1}^2}\right) \\ &\quad \cdot \exp\left(\frac{-\|\boldsymbol{x}-\boldsymbol{x}_1\|^2}{2{s_2}^2}\right)\mathrm{d}\boldsymbol{x}_1 = \left(\frac{1}{2\pi s_1 s_2}\right)^m \\ &\quad \cdot \int_{\boldsymbol{x}_1} \exp\left(-\frac{{s_2}^2\|\boldsymbol{x}_1\|^2+{s_1}^2\|\boldsymbol{x}-\boldsymbol{x}_1\|^2}{2{s_1}^2{s_2}^2}\right)\mathrm{d}\boldsymbol{x}_1 \\ &= \left(\frac{1}{2\pi s_1 s_2}\right)^m \\ &\quad \cdot \int_{\boldsymbol{x}_1} \exp\left(-\frac{({s_1}^2+{s_2}^2)\left\|\boldsymbol{x}_1-\frac{{s_1}^2}{{s_1}^2+{s_2}^2}\boldsymbol{x}\right\|^2}{2{s_1}^2{s_2}^2}\right. \\ &\quad \left.-\frac{\|\boldsymbol{x}\|^2}{2({s_1}^2+{s_2}^2)}\right)\mathrm{d}\boldsymbol{x}_1 \\ &= \left(\frac{1}{\sqrt{2\pi({s_1}^2+{s_2}^2)}}\right)^m \exp\left(-\frac{\|\boldsymbol{x}\|^2}{2({s_1}^2+{s_2}^2)}\right) \\ &= \rho_s(\boldsymbol{x})。\end{aligned} \tag{7}$$

但由于考虑的是限制在整数上的离散高斯分布，上述“加和”性质对于研究基于高斯分布的(非对称)含错学习问题的困难性并没有直接的帮助。幸运的是，对于特定参数的离散高斯分布，仍然可以证明其具有“加和”性质。特别地，有引理2成立。

引理2　如果标准差$s>\omega(\mathrm{lb}\kappa)$，那么离散高斯分布$D_{Z,s}$对于特定的参数具有“加和”性质。特别地，对于任意$s_1>\omega(\mathrm{lb}\kappa), s_2>\sqrt{{s_1}^2+\omega(\mathrm{lb}\kappa)^2}$，存在一个多项式时间的算法$S(\cdot,\cdot)$使得分布

$$\{x_1+x_2\in Z|x_1\leftarrow D_{Z,s_1}, x_2\leftarrow S(s_1,s_2)\} \tag{8}$$

和分布D_{Z,s_2}的统计距离关于安全参数κ是可忽略的。

为了证明引理2，需要用到蕴含在文献[5,10]中的两个引理。

引理3　对于任意实数$s_1,s_2>\omega(\mathrm{lb}\kappa)$，式(15)的分布

$$\{x_1+x_2|x_1\leftarrow D_{Z,s_1}, x_2\leftarrow D_{s_2}\} \tag{9}$$

和连续高斯分布D_s的统计距离关于安全参数κ是可忽略的，其中$s=\sqrt{{s_1}^2+{s_2}^2}$。

引理4　对于任意实数$s_1>0$和$s_2>\omega(\mathrm{lb}\kappa)$，式(16)的分布

$$\{x_1+x_2|x_1\leftarrow D_{s_1}, x_2\leftarrow D_{Z-x_1,s_2}\} \tag{10}$$

和离散高斯分布$D_{Z,s}$的统计距离关于安全参数κ是可忽略的，其中$s=\sqrt{{s_1}^2+{s_2}^2}$。

直观上，引理3 的意思是对于特定的参数，一个离散高斯分布加上一个连续高斯分布可以得到一个连续高斯分布，而引理4则可以理解为对一个连续高斯分布进行随机高斯取整后可得到一个离散的高斯分布。

引理2的证明　给定参数$s_1,s_2>\omega(\mathrm{lb}\,\kappa)$，算法$S(s_1,s_2)$进行如下运算：

(1) 计算$t=\sqrt{{s_2}^2-{s_1}^2}$；

(2) 抽样并输出x_2。

由引理4可知，x_2的分布统计接近于对于连续高斯分布进行随机高斯取整后得到的分布

$$\left\{x_{2,1}+x_{2,2}|x_{2,1}\leftarrow D_{t/\sqrt{2}}, x_{2,2}\leftarrow D_{Z-x_{2,1},t/\sqrt{2}}\right\} \tag{11}$$

因此，对于任意$x_1\leftarrow D_{Z,s_1}$，有x_1+x_2的分布统计接近于分布

$$\begin{aligned}&\left\{x_1+x_{2,1}+x_{2,2}|x_{2,1}\leftarrow D_{t/\sqrt{2}}, x_{2,2}\leftarrow D_{Z-x_{2,1},t/\sqrt{2}}\right\} \\ &\quad=\left\{x_1+x_{2,1}+x_{2,2}|x_{2,1}\leftarrow D_{t/\sqrt{2}},\right. \\ &\quad\left. x_{2,2}\leftarrow D_{Z-x_1-x_{2,1},t/\sqrt{2}}\right\}\end{aligned} \tag{12}$$

等式(12)成立是因为x_1是整数。进一步，由引理3可知，有x_1+x_2的分布统计接近于分布

$$\left\{x_3+x_{2,2}|x_3\leftarrow D_{t_1}, x_{2,2}\leftarrow D_{Z-x_3,t/\sqrt{2}}\right\} \tag{13}$$

其中，$t_1=\sqrt{{s_1}^2+t^2/2}$。再次使用引理4可知，x_1+x_2的分布统计接近于分布离散高斯分布D_{Z,s_2}，其中$s_2=\sqrt{{t_1}^2+t^2/2}$。由此可知引理2得证。

6　结束语

本文研究了非对称含错学习问题和标准学习问题之间困难关系，证明了对于具有“加和”性质的错误分布，非对称含错学习问题和标准学习问题是等价的。特别地，本文还证明了特定参数下的离散高斯分布和二项分布均满足“加和”的性质，从而为基于离散高斯分布或二项分布的非对称含错学习问题设计安全的格密码方案奠定了理论基础。

参 考 文 献

[1] SHOR P W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer[J]. *SIAM Journal on Computing*, 1997, 26(5): 1484–1509. doi: 10.1137/S0097539795293172.

[2] NSA. National Security Agency. Cryptography today[EB/OL]. https://www.nsa.gov/ia/programs/suiteb_cryptography/, 2015.

[3] NIST. Post-quantum cryptography standardization [EB/OL]. http://csrc.nist.gov/groups/ST/post-quantum-crypto/submission-requirements/index.html, 2016.

[4] 中国科学技术学会. 科普时报: 中国科协发布12个领域60大科技难题[EB/OL]. http://www.cast.org.cn/art/2018/6/22/art_90_77662.html, 2018.

[5] 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.

[6] AJTAI M. Generating hard instances of lattice problems (extended abstract)[C]. The 28th Annual ACM Symposium on Theory of Computing, Philadelphia, USA,1996: 99–108.

[7] ZHANG Jiang, YU Yu, FAN Shuqin, *et al*. Tweaking the asymmetry of asymmetric-key cryptography on lattices: KEMs and signatures of smaller sizes[R]. Cryptology ePrint Archive 2019/510, 2019.

[8] APPLEBAUM B, CASH D, PEIKERT C, *et al*. Fast cryptographic primitives and circular-secure encryption based on hard learning problems[C]. The 29th Annual International Cryptology Conference on Advances in Cryptology, Santa Barbara, USA, 2009: 595–618.

[9] MICCIANCIO D and REGEV O. Worst-case to average-case reductions based on Gaussian measures[C]. The 45th Annual IEEE Symposium on Foundations of Computer Science, Rome, Italy, 2004: 372–381.

[10] PEIKERT C. An efficient and parallel Gaussian sampler for lattices[C]. The 30th Annual Conference on Advances in Cryptology, Santa Barbara, USA, 2010: 80–97.

张 江：男，1986年生，副研究员，主要研究方向为基于格的密码协议及其可证明安全.

范淑琴：女，1978年生，教授，主要研究方向为基于格的密码分析.