SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Rényi Divergence on Learning with Errors

Yang TAO[1,2], Han WANG[1,2*] & Rui ZHANG[1,2]

[1]*State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing* 100093*, China;*
[2]*School of Cyber Security, University of Chinese Academy of Sciences, Beijing* 100049*, China*

Appendix A Average Argument in Proof of Theorem 1

Theorem 1. For decision problems P and P', assume that $D_0(\cdot)$ and $D_1(\cdot)$ satisfy the adaptively public sampling property. Then, given a T-time PPT distinguisher $\mathcal{A}$ for Problem P with advantage ϵ, we can construct a PPT distinguisher $\mathcal{A}'$ for Problem P' with advantage bounded by $\frac{\epsilon}{8\cdot R_a(\Phi\|\Phi')}\cdot(\frac{\epsilon}{8})^{\frac{a}{a-1}}-\frac{1}{2}\epsilon_1$ for any $a\in(1,+\infty])$ with running time at most $O(\frac{1}{\epsilon^2}\log(\frac{R_a(\Phi\|\Phi')}{\epsilon^{\frac{a}{a-1}+1}})(T_S+T))$ where T_S is the upper bound of running time of S_0 and S_1.

Recall that $\mathcal{S}_1'=\{r|p_1(r)-p_0(r)\geqslant\frac{\epsilon}{2}+3\epsilon_1\}$, $\mathcal{S}_2'=\{r|-\epsilon_1\leqslant p_1(r)-p_0(r)<\frac{\epsilon}{2}+3\epsilon_1\}$ and $\mathcal{S}_3'=\{r|p_1(r)-p_0(r)<-\epsilon_1\}$. Let $a_i=\frac{1}{2}\Pr[\mathcal{A}(x)=1|x\leftarrow D_1(r)]+\frac{1}{2}\Pr[\mathcal{A}(x)=0|x\leftarrow D_0(r)]=\frac{1}{2}(1+p_1(r)-p_0(r))$ when randomness r is in $\mathcal{S}_i'$ for $i=1,2,3$. Due to the definition of $\mathcal{S}_i'$ for $i=1,2,3$, we have

$$\frac{1}{2}(1+\frac{\epsilon}{2}+3\epsilon_1)\leqslant a_1\leqslant 1, \tag{A1}$$

$$\frac{1}{2}(1-\epsilon_1)\leqslant a_2<\frac{1}{2}(1+\frac{\epsilon}{2}+3\epsilon_1), \tag{A2}$$

$$0\leqslant a_3<\frac{1}{2}(1-\epsilon_1). \tag{A3}$$

Since the advantage of $\mathcal{A}$ is $Adv(\mathcal{A})=\epsilon$, we have

$$\epsilon=Adv(\mathcal{A})=\sum_r\{\Phi(r)(\Pr[\mathcal{A}(x)=1|x\leftarrow D_1(r)]-\Pr[\mathcal{A}(x)=1|x\leftarrow D_0(r)])\} \tag{A4}$$

$$=\sum_{i=1}^{3}\Phi(\mathcal{S}_i')(2a_i-1). \tag{A5}$$

Hence, the probability

$$\frac{1+\epsilon}{2}=\Phi(\mathcal{S}_1')a_1+\Phi(\mathcal{S}_2')a_2+\Phi(\mathcal{S}_3')a_3\leqslant\Phi(\mathcal{S}_1')a_1+(\Phi(\mathcal{S}_2')+\Phi(\mathcal{S}_3'))a_2 \tag{A6}$$

$$\leqslant\Phi(\mathcal{S}_1')+(1-\Phi(\mathcal{S}_1'))a_2\leqslant\Phi(\mathcal{S}_1')+a_2 \tag{A7}$$

$$\leqslant\Phi(\mathcal{S}_1')+\frac{1}{2}(1+\frac{\epsilon}{2}+3\epsilon_1), \tag{A8}$$

infers $\Phi(\mathcal{S}_1')\geqslant\frac{\epsilon}{4}-\frac{3}{2}\epsilon_1$. Since ϵ_1 is a negligible function of n, we assert $\epsilon_1<\frac{\epsilon}{12}$ as the security parameter n increases. Hence, it follows $\Phi(\mathcal{S}_1')\geqslant\frac{\epsilon}{8}$.

Appendix B Proof of Theorem 2

Lemma 1. Let X,Y be two random variables taking values in a common set A. For any (possibly randomized) function f with domain A, the statistical distance between $f(X)$ and $f(Y)$ satisfies $\Delta(f(X),f(Y))\leqslant\Delta(X,Y)$.

Lemma 2 ([2], Lemma 7)**.** Let $\mathcal{R}=\mathbb{Z}[x]/(x^n+1)$, $\mathcal{R}_q=\mathbb{Z}_q[x]/(x^n+1)$ for $n\geqslant 4$ a power of 2 and $q=3^k$ a power of 3. Let $m\geqslant 2\lceil\log q\rceil+2$ and $\alpha\geqslant\omega(\sqrt{\ln nm})$. With overwhelming probability over the choice of $\mathbf{a}\leftarrow\mathcal{R}_q^{1\times m}$, if $\mathbf{x}\leftarrow D_{\mathcal{R},\alpha}^m$, then $\mathbf{ax}$ is within negligible statistical distance from the uniform distribution over $\mathcal{R}$.

* Corresponding author (email: wanghan@iie.ac.cn)

Theorem 2. Let χ' and $\chi = D_{\mathcal{R},\alpha}$ be two error distributions over $\mathcal{R}$ with $\mathsf{Supp}(\chi') \subseteq \mathsf{Supp}(\chi)$. Let $\mathbf{a} \leftarrow \mathcal{R}_q^{1\times m}$, $\mathbf{s} \leftarrow \mathcal{R}_q^{m\times 1}$, $m \geqslant 2\lceil \log_2 q\rceil + 2$ and $\alpha \geqslant \omega(\sqrt{\ln nm})$. Then, if there is a PPT distinguisher $\mathcal{A}$ against $\mathrm{MLWE}_{n,m,q}(\chi')$ with advantage ε, there exists a PPT distinguisher $\mathcal{A}'$ against the $\mathrm{MLWE}_{n,m,q}(\chi)$ with the access to oracle $\mathcal{O}_x$ with advantage $\Omega(\frac{\varepsilon^{1+a/(a-1)}}{R_a(\chi'\|\chi)})$ and running time $O(\frac{1}{\epsilon^2}\log(\frac{R_a(\Phi\|\Phi')}{\epsilon^{\frac{a}{a-1}+1}})(T_S+T))$ where T_S is the upper bound of running time of S_0 and S_1, for any $a \in (1,+\infty]$.

Proof. By Theorem 1, it suffices to verify the distributions satisfying the adaptively public sampling property. Set the error term as the randomness from χ. Define the distribution $D_0(\mathbf{e}) = (\mathbf{a}, \mathbf{b} = \mathbf{as}+\mathbf{e})$ with $\mathbf{a} \leftarrow \mathcal{R}_q^{1\times m}, \mathbf{s} \leftarrow \mathcal{R}_q^m$, $D_1(\mathbf{e}) = (\mathbf{a},\mathbf{u})$ with $\mathbf{a} \leftarrow \mathcal{R}_q^{1\times m}$ and $\mathbf{u} \leftarrow \mathcal{R}_q$. Since the distribution $D_1(\mathbf{e})$ is independent from the randomness $\mathbf{e}$, we consider instances of D_1 can be corresponding to any randomness.

In details, for any instance $x = (\mathbf{a},\mathbf{b}) \in \mathcal{R}_q^{1\times m} \times \mathcal{R}_q$ from the distribution $D_b(\mathbf{e})$ for $b \in \{0,1\}$, we define the adaptively public sampling algorithms as follows. Define $\widetilde{D}_0 = D_0, \widetilde{D}_1 = D_1, U_1 = U_2 = U_4$ is the uniform distribution over $\mathcal{R}_q^{1\times m}\times\mathcal{R}_q$ and U_3 is a distribution statistically close to the uniform distribution, which is defined later.

- Algorithm S_0 with $x = (\mathbf{a},\mathbf{b})$:
- $S_0(0,x)$ outputs $(\mathbf{a},\mathbf{b}+\mathbf{at})$ with $\mathbf{t} \leftarrow \mathcal{R}_q^m$.
- $S_0(1,x)$ outputs $(\mathbf{a},\mathbf{u})$ with $\mathbf{u} \leftarrow \mathcal{R}_q$.
- Algorithm S_1 with $x = (\mathbf{a},\mathbf{b})$:
- $S_1(0,x)$ first samples a random $\mathbf{e}$, and then gets additional $m-1$ samples $\mathbf{b}'_i \in \mathcal{R}_q$ by accessing the oracle $\mathcal{O}_x$ for $i \in \{1,\cdots,m-1\}$. Set $\mathbf{b}' = [\mathbf{b}'_1|\mathbf{b}'_2|\cdots|\mathbf{b}'_{m-1}]^t \in \mathcal{R}_q^{m-1}$, $\mathbf{b}^* = \begin{bmatrix}\mathbf{b}\\ \mathbf{b}'\end{bmatrix} \in \mathcal{R}_q^m$ and output $(\mathbf{a},\mathbf{ab}^*+\mathbf{e})$.
- $S_1(1,x)$ outputs $(\mathbf{a},\mathbf{u})$ with $\mathbf{u} \leftarrow \mathcal{R}_q$.

Now we claim the output of algorithm S_0 and S_1 satisfies the properties. First, for the algorithm S_0,
(1) when $x \leftarrow D_0(\mathbf{e})$, i.e. $x = (\mathbf{a},\mathbf{b} = \mathbf{as}+\mathbf{e})$,

- $S_0(0,x)$ outputs $(\mathbf{a},\mathbf{b}+\mathbf{at})$ with $\mathbf{t} \leftarrow \mathcal{R}_q^m$, which is a fresh sample from $D_0(\mathbf{e})$ with secret $\mathbf{s}+\mathbf{t}$.
- $S_0(1,x)$ outputs $(\mathbf{a},\mathbf{u})$ with $\mathbf{u} \leftarrow \mathcal{R}_q$, which is a fresh sample from $D_1(\mathbf{e})$.

(2) when $x \leftarrow D_1(\mathbf{e})$, i.e. $x = (\mathbf{a},\mathbf{b})$ with randomly uniform $\mathbf{b}$,

- $S_0(0,x)$ outputs $(\mathbf{a},\mathbf{b}+\mathbf{at})$ with $\mathbf{t} \leftarrow \mathcal{R}_q^m$. Since $\mathbf{b}$ is randomly uniform and independent from $\mathbf{a}$ and $\mathbf{t}$, it follows $\mathbf{b}+\mathbf{at}$ is uniform, which is a fresh sample from U_1.
- $S_0(1,x)$ outputs $(\mathbf{a},\mathbf{u})$ with $\mathbf{u} \leftarrow \mathcal{R}_q$, which is a fresh sample from U_2.

Then, for the algorithm S_1,
(1) when $x \leftarrow D_0(\mathbf{e})$, i.e. $x = (\mathbf{a},\mathbf{b} = \mathbf{as}+\mathbf{e})$. The oracle $\mathcal{O}_x$ outputs LWE instances with the same secret $\mathbf{s}$ with $\mathbf{b}$.

- $S_1(0,x)$ chooses $\mathbf{e}'$ from the error distribution χ and outputs $(\mathbf{a},\mathbf{ab}^*+\mathbf{e}')$. Since $\mathbf{b} = \mathbf{as}+\mathbf{e}$ and $\mathbf{b}' = \mathbf{A}'\mathbf{s}+\bar{\mathbf{e}}$, where $\mathbf{A}' \leftarrow \mathcal{R}_q^{(m-1)\times m}$, $\mathbf{b}^* = \begin{bmatrix}\mathbf{b}\\ \mathbf{b}'\end{bmatrix} = \mathbf{A}_1\mathbf{s}+\mathbf{e}_1$ with $\mathbf{A}_1 = \begin{bmatrix}\mathbf{a}\\ \mathbf{A}'\end{bmatrix} \in \mathcal{R}_q^{m\times m}$ and $\mathbf{e}_1 = \begin{bmatrix}\mathbf{e}\\ \bar{\mathbf{e}}\end{bmatrix} \in \mathcal{R}_q^m$. Hence, $\mathbf{ab}^*+\mathbf{e}' = \mathbf{a}(\mathbf{A}_1\mathbf{s}+\mathbf{e}_1)+\mathbf{e}' = \mathbf{aA}_1\mathbf{s}+\mathbf{ae}_1+\mathbf{e}'$.

Now, we claim the distribution of $(\mathbf{a},\mathbf{ab}^*+\mathbf{e}')$ is statistically close to the uniform distribution over $\mathcal{R}_q^{1\times m}\times\mathcal{R}_q$. Define $f(\mathbf{a},\mathbf{u}^*) = \mathbf{aA}_1\mathbf{s}+\mathbf{u}^*+\mathbf{e}'$ conditioned on any prescribed secret $\mathbf{s}$. By Lemma 1 and Lemma 2, $\Delta(f(\mathbf{a},\mathbf{ae}_1), f(\mathbf{a},\mathbf{u})) \leqslant \Delta((\mathbf{a},\mathbf{ae}_1),(\mathbf{a},\mathbf{u})) = negl(n)$, where $\mathbf{u}$ is a uniform vector of $\mathcal{R}_q$. The latter $f(\mathbf{a},\mathbf{u})$ is a uniform vector since $\mathbf{u}$ is independent from $\mathbf{aA}_1\mathbf{s}$ and $\mathbf{e}'$. Thus, the output distribution U_3 of $S_1(0,x)$ is statistically close to the uniform distribution.

- $S_1(1,x)$ outputs $(\mathbf{a},\mathbf{u})$ with $\mathbf{u} \leftarrow \mathcal{R}_q$, which is a fresh sample from U_4.

(2) when $x \leftarrow D_1(\mathbf{e})$, i.e. $x = (\mathbf{a},\mathbf{b})$ where $\mathbf{b}$ is a uniform vector on $\mathcal{R}_q$.

- $S_1(0,x)$ outputs a fresh sample from $D_0(\mathbf{e})$ with the secret $\mathbf{b}^*$, since $\mathcal{O}_x$ is a uniform distribution.
- $S_1(1,x)$ outputs $(\mathbf{a},\mathbf{u})$ with $\mathbf{u} \leftarrow \mathcal{R}_q$, which is a fresh sample from $D_1(\mathbf{e})$.

In conclusion, $D_0(\mathbf{e})$ and $D_1(\mathbf{e})$ satisfy the adaptively public sampling property and the proof is completed by Theorem 1.

Appendix C Proof of Corollary

Lemma 3 (Adapted from [1], Lemma 5.2)**.** Let α,β be real numbers with $\beta \geqslant \alpha$. Let U_β be uniform distribution over $[-\beta,\beta]$ and $D_{\mathbb{Z},\alpha}$ be a discrete Gaussian distribution. Define distribution $\psi = D_{\mathbb{Z},\alpha}+U_\beta$. We have

$$R_2(U_\beta\|\psi) = \frac{1}{C}(1+\frac{1}{1-\exp^{-\pi\beta^2/\alpha^2}}\frac{\alpha}{\beta}) < \frac{1}{C}(1+1.05\frac{\alpha}{\beta}),$$

where $C = \rho_\alpha(\mathbb{Z})$ is a constant.

Corollary 1. Let $m \geqslant 2\lceil\log_2 q\rceil+2$, $\alpha \geqslant \omega(\sqrt{\ln nm})$ and $\alpha,\beta > 0$ be real numbers with $\beta = \Omega(n\alpha/\log n)$ for positive integers n. Then there is a polynomial-time reduction from $\mathrm{MLWE}_{n,m,q}(D_{\mathcal{R},\alpha})$ to $\mathrm{MLWE}_{n,m,q}(\bar{U}_\beta)$, where $\bar{U}_\beta = \frac{1}{q}\lfloor qU_\beta\rceil$ and U_β is a continuous uniform distribution over $[-\beta,\beta]$.

Proof. Let U_β denote the uniform distribution over $[-\beta,\beta]$ and $\psi = D_{\mathbb{Z},\alpha}+U_\beta$ denote the convolution of $D_{\mathbb{Z},\alpha}$ and U_β. Our reduction contains three steps.

- First, we claim there is a reduction from $\mathrm{MLWE}_{n,m,q}(D_{\mathcal{R},\alpha})$ to $\mathrm{MLWE}_{n,m,q}(\psi)$.
- Second, we prove a reduction from $\mathrm{MLWE}_{n,m,q}(\psi)$ to $\mathrm{MLWE}_{n,m,q}(U_\beta)$.
- At last, we reduce $\mathrm{MLWE}_{n,m,q}(U_\beta)$ to $\mathrm{MLWE}_{n,m,q}(\bar{U}_\beta)$ by discretization.

Step 1: Given an instance $(\mathbf{a},\mathbf{b})$ from $\mathrm{MLWE}_{n,m,q}(D_{\mathcal{R},\alpha})$ problem. We choose independent samples $\mathbf{b}'_i$ from U_β as the coefficients of element $\mathbf{b}' \in \mathcal{R}$ and transform $(\mathbf{a},\mathbf{b})$ into $(\mathbf{a},\mathbf{b}+\mathbf{b}')$. If $(\mathbf{a},\mathbf{b})$ is from uniform distribution, $(\mathbf{a},\mathbf{b}+\mathbf{b}')$ is uniform. Otherwise, if $(\mathbf{a},\mathbf{b})$ is from MLWE distribution with error from $D_{\mathcal{R},\alpha}$, then $(\mathbf{a},\mathbf{b}+\mathbf{b}')$ is from MLWE distribution with each error coefficient from ψ.

Step 2: It suffices to argue the $R_2(U_\beta^n\|\psi^n)$ is polynomial bounded, where the error term from ψ_α (resp. U_β^n) contains n independent coefficients from ψ (resp. U_β). By the multiplicative property of RD and Lemma 3, we have $R_2(U_\beta^n\|\psi^n) \leqslant R_2(U_\beta\|\psi)^n < (1+1.05\frac{\alpha}{\beta})^n \leqslant n^{O(1)}$ due to $\beta = \Omega(\frac{n\alpha}{\log n})$. Therefore, a distinguisher of $\mathrm{MLWE}_{n,m,q}(U_\beta)$ problem can be converted to a distinguisher of $\mathrm{MLWE}_{n,m,q}(\psi)$ problem.

Step 3: Given an instance $(\mathbf{a},\mathbf{b})$ from $\mathrm{MLWE}_{n,m,q}(U_\beta)$ problem. We round each coefficient b_i of $\mathbf{b}$ to $\frac{1}{q}\lfloor qb_i\rceil$. If $(\mathbf{a},\mathbf{b})$ is from uniform distribution, it is uniform. Otherwise, it is a sample from $\mathrm{MLWE}_{n,m,q}(\bar{U}_\beta)$.

In conclusion, there is a polynomial-time reduction from $\mathrm{MLWE}_{n,m,q}(D_\alpha)$ to $\mathrm{MLWE}_{n,m,q}(\bar{U}_\beta)$.

References

1 Bai S, Langlois A, Lepoint T, et al. Improved security proofs in lattice-based cryptography: using the rényi divergence rather than the statistical distance. In: Iwata T, Cheon J H, eds. Advances in Cryptology - ASIACRYPT 2015, Auckland, 2015. 3-24

2 Ducas L, Micciancio D. Improved short lattice signatures in the standard model. In: Garay J A, Gennaro R, eds. Advances in Cryptology - CRYPTO 2014, Santa Barbara, 2014. 335-352