J Cryptol (2018) 31:610–640
https://doi.org/10.1007/s00145-017-9265-9

Journal of
CRYPTOLOGY

Improved Security Proofs in Lattice-Based Cryptography: Using the Rényi Divergence Rather than the Statistical Distance

Shi Bai

Department of Mathematical Sciences, Florida Atlantic University, Boca Raton, FL, USA
sbai@fau.edu
http://cosweb1.fau.edu/~sbai/

Tancrède Lepoint

SRI International, New York, NY, USA
tancrede.lepoint@sri.com
https://tlepoint.github.io/

Adeline Roux-Langlois

CNRS/IRISA, Rennes, France
adeline.roux-langlois@irisa.fr
http://people.irisa.fr/Adeline.Roux-Langlois/

Amin Sakzad

Faculty of Information Technology, Monash University, Clayton, VIC, Australia
amin.sakzad@monash.edu
http://monash.edu/research/explore/en/persons/amin-sakzad(dd21c248-0e1f-489d-8df0-d1f5250ea5df).html/

Damien Stehlé

ENS de Lyon, Laboratoire LIP (U. Lyon, CNRS, ENSL, INRIA, UCBL), Lyon, France
damien.stehle@ens-lyon.fr
http://perso.ens-lyon.fr/damien.stehle/

Ron Steinfeld

Faculty of Information Technology, Monash University, Clayton, VIC, Australia
ron.steinfeld@monash.edu
http://users.monash.edu.au/~rste/

Communicated by Kenneth G. Paterson.

Received 19 May 2016 / Revised 26 July 2017
Online publication 28 August 2017

Abstract. The Rényi divergence is a measure of closeness of two probability distributions. We show that it can often be used as an alternative to the statistical distance in security proofs for lattice-based cryptography. Using the Rényi divergence is particularly suited for security proofs of primitives in which the attacker is required to solve a search problem (e.g., forging a signature). We show that it may also be used in the case of distinguishing problems (e.g., semantic security of encryption schemes), when they enjoy a public sampleability property. The techniques lead to security proofs for schemes with smaller parameters, and sometimes to simpler security proofs than the existing ones.

Keywords. Lattice-based cryptography, Rényi divergence, Statistical distance, Security proofs.

1. Introduction

Let D_1 and D_2 be two non-vanishing probability distributions over a common measurable support X. Let $a \in (1, +\infty)$. The *Rényi divergence* [33,35] (RD for short) $R_a(D_1 \| D_2)$ of order a between D_1 and D_2 is defined as the (($a-1$)th root of the) expected value of $(D_1(x)/D_2(x))^{a-1}$ over the randomness of x sampled from D_1. For notational convenience, our definition of the RD is the exponential of the classical definition [35]. The RD is an alternative to the statistical distance (SD for short) $\Delta(D_1, D_2) = \frac{1}{2}\sum_{x \in X} |D_1(x) - D_2(x)|$ as measure of distribution closeness, where we replace the difference in SD, by the ratio in RD. RD enjoys several properties that are analogous of those enjoyed by SD, where the additive property of SD is replaced by a multiplicative property in RD (see Sect. 2.3).

SD is ubiquitous in cryptographic security proofs. One of its most useful properties is the so-called *probability preservation property*: For any measurable event $E \subseteq X$, we have $D_2(E) \geq D_1(E) - \Delta(D_1, D_2)$. RD enjoys the analogous property $D_2(E) \geq D_1(E)^{\frac{a}{a-1}}/R_a(D_1 \| D_2)$. If the event E occurs with significant probability under D_1, and if the SD (resp. RD) is small, then the event E also occurs with significant probability under D_2. These properties are particularly handy when the success of an attacker against a given scheme can be described as an event whose probability should be non-negligible, e.g., the attacker outputs a new valid message-signature pair for a signature scheme. If the attacker succeeds with good probability in the real scheme based on distribution D_1, then it also succeeds with good probability in the simulated scheme (of the security proof) based on distribution D_2.

To make the SD probability preservation property useful, it must be ensured that the SD $\Delta(D_1, D_2)$ is smaller than any $D_1(E)$ that the security proof must handle. Typically, the quantity $D_1(E)$ is assumed to be greater than some success probability lower bound ε, which is of the order of $1/\mathrm{poly}(\lambda)$ where λ refers to the security parameter, or even $2^{-o(\lambda)}$ if the proof handles attackers whose success probabilities can be sub-exponentially small (which we believe better reflects practical objectives). As a result, the SD $\Delta(D_1, D_2)$ must be $< \varepsilon$ for the SD probability preservation property to be relevant. In contrast, the RD probability preservation property is non-vacuous when the RD $R_a(D_1 \| D_2)$ is $\leq \mathrm{poly}(1/\varepsilon)$. In many cases, the latter seems less demanding than the former: in all our applications, the RD between D_1 and D_2 is small enough for the RD probability preservation property while their SD is too large for the SD probability preservation to be applicable (see Sect. 2.3). This explains the superiority of the RD in several of our applications.

Although RD seems more amenable than SD for search problems, it seems less so for distinguishing problems. A typical cryptographic example is semantic security of an encryption scheme. Semantic security requires an adversary $\mathcal{A}$ to distinguish between the encryption distributions of two plaintext messages of its choosing: the distinguishing advantage $\mathrm{Adv}_{\mathcal{A}}(D_1, D_2)$, defined as the difference of probabilities that $\mathcal{A}$ outputs 1 using D_1 or D_2, should be sufficiently large. In security proofs, algorithm $\mathcal{A}$ is often called on distributions D_1' and D_2' that are close to D_1 and D_2 (respectively). If the SDs between D_1 and D_1' and D_2 and D_2' are both bounded from above by ε, then, by the SD probability preservation property (used twice), we have $\mathrm{Adv}_{\mathcal{A}}(D_1', D_2') \geq \mathrm{Adv}_{\mathcal{A}}(D_1, D_2) - 2\varepsilon$. As a result, SD can be used for distinguishing problems in a similar fashion as for search problems. The multiplicativity of the RD probability preservation property seems to prevent RD from being applicable to distinguishing problems.

We replace the statistical distance by the Rényi divergence in several security proofs for lattice-based cryptographic primitives. *Lattice-based cryptography* is a relatively recent cryptographic paradigm in which cryptographic primitives are shown at least as secure as it is hard to solve standard problems over lattices (see the surveys [26,29]). Security proofs in lattice-based cryptography involve different types of distributions, often over infinite sets, such as continuous Gaussian distributions and Gaussian distributions with lattice supports. The RD seems particularly well suited to quantify the closeness of Gaussian distributions. Consider for example two continuous Gaussian distributions over the reals, both with standard deviation 1, but one with center 0 and the other one with center c. Their SD is linear in c, so that c must remain extremely small for the SD probability preservation property to be useful. On the other hand, their RD of order $a = 2$ is bounded as $\exp(O(c^2))$ so that the RD preservation property remains useful even for slightly growing c.

RD was first used in lattice-based cryptography in Lyubashevsky et al. [19], in the decision to search reduction for the Ring Learning With Errors problem (which serves as a security foundation for many asymptotically fast primitives). It was then exploited in Langlois et al. [21] to decrease the parameters of the [14] (approximation to) cryptographic multilinear maps. In the present work, we present a more extensive study of the power of RD in lattice-based cryptography, by showing several independent applications of RD. In some cases, it leads to security proofs allowing to take smaller parameters in the cryptographic schemes, hence leading to efficiency improvements. In other cases, this leads to alternative security proofs that are conceptually simpler.

Our applications of RD also include distinguishing problems. To circumvent the aforementioned a priori limitation of the RD probability preservation property for distinguishing problems, we propose an alternative approach that handles a class of distinguishing problems enjoying a special property that we call *public sampleability*. This public sampleability allows to estimate success probabilities via Hoeffding's bound.

The applications we show in lattice-based cryptography are as follows:

- Smaller storage requirement for the Fiat-Shamir BLISS signature scheme [11,13,27].
- Smaller parameters in the dual-Regev encryption scheme from Gentry et al. [16].
- Alternative proof that the Learning With Errors (LWE) problem with noise chosen uniformly in an interval is no easier than the Learning With Errors problem with Gaussian noise [12]. Our reduction does not require the latter problem to be hard, and it is hence marginally more general as it also applies to distributions with smaller noises.

Further, our reduction preserves the LWE dimension n, and is hence tighter than the one from Refs. [12] (the latter degrades the LWE dimension by a constant factor).[1]

- Alternative proof that the Learning With Rounding (LWR) problem [7] is no easier than LWE. Our reduction is the first which preserves the dimension n without resorting to noise flooding (which significantly degrades the noise rate): the reductions from Refs. [3,4] do not preserve the dimension, and the one from Banerjee et al. [7] preserves the dimension but makes use of noise flooding. Alwen et al. [3], the authors can get close to preserve the dimension up to a constant but at a price of larger polynomial modulos. Denoting by $\mathbb{Z}_p$ the ring in which we perform rounding, our new reduction also gains extra factors of $p\sqrt{\log n}$ and $pn\sqrt{\log n}$ in the number of LWR samples handled, compared with Bogdanov et al. [4] and Alwen et al. [3], respectively.

We think RD is likely to have further applications in lattice-based cryptography, for both search and distinguishing problems.

Related Work The framework for using RD in distinguishing problems was used in Ling et al. [20], in the context of the k-LWE problem (a variant of LWE in which the attacker is given extra information). Pöppelmann et al. [27] used the Kullback–Leibler divergence (which is the RD of order $a = 1$) to lower the storage requirement of BLISS scheme [11]. Asymptotically, using the Kullback–Leibler divergence rather than SD only leads to a constant factor improvement. Our approach allows bigger savings in the case where the number of signature queries is limited, as explained in Sect. 3.

Recently, Bogdanov et al. [4] adapted parts of (an earlier version of) our RD-based hardness proof for LWE with noise uniform in a small interval, to the LWR problem. In particular, they obtained a substantial improvement over the hardness results of Refs. [3, 7]. In this revised and extended version of our earlier conference paper [5], we show an alternative LWR hardness proof that improves on that of Bogdanov et al. [4], exploiting the equivalence of LWR to LWE with noise uniform in an interval; an equivalence was also established in Bogdanov et al. [4] but not used there to relate the hardness of LWE to that of LWR.

After the publication of earlier versions of this article, some of our results have been improved [34] and used in Libert et al. [18] in the context of dynamic group signatures and in Alkim et al. [1] to replace the LWE error distribution by a more efficiently samplable distribution.

Road-map In Sect. 2, we provide necessary background on lattice-based cryptography, and on the Rényi divergence. In Sect. 3, we use RD to improve lattice-based signature scheme parameters via more efficient Gaussian sampling. Section 4 contains the description of the framework in which we can use RD for distinguishing problems, which we apply to improve the parameters of the dual-Regev encryption scheme. In Sect. 5, we describe an alternative hardness proof for LWE with noise uniformly chosen in an interval. Section 6 shows an application of the previous section to give a new hardness proof for the LWR problem. Finally, Sect. 7 concludes with open problems.

[1] Note that LWE with uniform noise in a small interval is also investigated in Alwen et al. [24], with a focus on the number of LWE samples. The reduction from Micciancio et al. [24] does not preserve the LWE dimension either.

Notation If x is a real number, we let $\lfloor x \rceil$ denote a closest integer to x. The notation ln refers to the natural logarithm and the notation log refers to the base 2 logarithm. We define $\mathbb{T} = ([0, 1], +)$, where the addition operation is just modulo 1 operation. For an integer q, we let $\mathbb{Z}_q$ denote the ring of integers modulo q. We let $\mathbb{T}_q$ denote the group $\mathbb{T}_q = \{i/q \bmod 1 : i \in \mathbb{Z}\} \subseteq \mathbb{T}$. Vectors are denoted in bold. If $\boldsymbol{b}$ is a vector in $\mathbb{R}^d$, we let $\|\boldsymbol{b}\|$ denote its Euclidean norm. By default, all our vectors are column vectors.

If D is a probability distribution, we let $\mathrm{Supp}(D) = \{x : D(x) \neq 0\}$ denote its support. For a set X of finite weight, we let $U(X)$ denote the uniform distribution on X. To ease notation, we let U_β denote the distribution $U([-\beta, \beta])$ for a positive real β. The statistical distance between two distributions D_1 and D_2 over a countable support X is $\Delta(D_1, D_2) = \frac{1}{2}\sum_{x\in X} |D_1(x) - D_2(x)|$. This definition is extended in the natural way to continuous distributions. If $f : X \to \mathbb{R}$ takes non-negative values, then for all countable $Y \subseteq X$, we define $f(Y) = \sum_{y\in Y} f(y) \in [0, +\infty]$. For any vector $\boldsymbol{c} \in \mathbb{R}^n$ and any real $s > 0$, the (spherical) Gaussian function with standard deviation parameter s and center $\boldsymbol{c}$ is defined as follows: $\forall \boldsymbol{x} \in \mathbb{R}^n, \rho_{s,\boldsymbol{c}}(\boldsymbol{x}) = \exp(-\pi\|\boldsymbol{x} - \boldsymbol{c}\|^2/s^2)$. The Gaussian distribution is $D_{s,\boldsymbol{c}} = \rho_{s,\boldsymbol{c}}/s^n$. When $\boldsymbol{c} = \boldsymbol{0}$, we may omit the subscript $\boldsymbol{c}$.

We use the usual Landau notations. A function $f(\lambda)$ is said negligible if it is $\lambda^{-\omega(1)}$. A probability $p(\lambda)$ is said overwhelming if it is $1 - \lambda^{-\omega(1)}$.

The distinguishing advantage of an algorithm $\mathcal{A}$ between two distributions D_0 and D_1 is defined as $\mathrm{Adv}_{\mathcal{A}}(D_0, D_1) = |\Pr_{x\hookleftarrow D_0}[\mathcal{A}(x) = 1] - \Pr_{x\hookleftarrow D_1}[\mathcal{A}(x) = 1]|$, where the probabilities are taken over the randomness of the input x and the internal randomness of $\mathcal{A}$. Algorithm $\mathcal{A}$ is said to be an (ε, T)-distinguisher if it runs in time $\leq T$ and if $\mathrm{Adv}_{\mathcal{A}}(D_0, D_1) \geq \varepsilon$.

We say a distribution χ is B-bounded, for some positive real B, if its support be in the interval $[-B, B]$. In the case where χ is over $\mathbb{Z}_q$, we assume that $B \leq (q - 1)/2$. A B-bounded distribution χ is said to be balanced if $\Pr[\chi \leq 0] \geq 1/2$ and $\Pr[\chi \geq 0] \geq 1/2$.

2. Preliminaries

We assume the reader is familiar with standard cryptographic notions, as well as with lattices and lattice-based cryptography. We refer to Refs. [26,31] for introductions on the latter topic.

2.1. *Lattices*

A (full-rank) n-dimensional *Euclidean lattice* $\Lambda \subseteq \mathbb{R}^n$ is the set of all integer linear combinations $\sum_{i=1}^n x_i\boldsymbol{b}_i$ of some $\mathbb{R}$-basis $(\boldsymbol{b}_i)_{1\leq i\leq n}$ of $\mathbb{R}^n$. In this setup, the tuple $(\boldsymbol{b}_i)_i$ is said to form a $\mathbb{Z}$-basis of Λ. For a lattice Λ and any $i \leq n$, the ith successive minimum $\lambda_i(\Lambda)$ is the smallest radius r such that Λ contains i linearly independent vectors of norm at most r. The dual Λ^* of a lattice Λ is defined as $\Lambda^* = \{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{y}^t\Lambda \subseteq \mathbb{Z}^n\}$.

The (spherical) *discrete Gaussian distribution over a lattice* $\Lambda \subseteq \mathbb{R}^n$, with standard deviation parameter $s > 0$ and center $\boldsymbol{c}$ is defined as:

$$\forall \boldsymbol{x} \in \Lambda,\ D_{\Lambda,s,\boldsymbol{c}} = \frac{\rho_{s,\boldsymbol{c}}(\boldsymbol{x})}{\rho_{s,\boldsymbol{c}}(\Lambda)}.$$

When the center is $\mathbf{0}$, we omit the subscript $\boldsymbol{c}$.

The *smoothing parameter* [25] of an n-dimensional lattice Λ with respect to $\varepsilon > 0$, denoted by $\eta_\varepsilon(\Lambda)$, is the smallest $s > 0$ such that $\rho_{1/s}(\Lambda^* \setminus \{0\}) \leq \varepsilon$. We use the following properties.

Lemma 2.1. *([25, Lemma 3.3]) Let Λ be an n-dimensional lattice and $\varepsilon > 0$. Then*

$$\eta_\varepsilon(\Lambda) \leq \sqrt{\frac{\ln(2n(1+1/\varepsilon))}{\pi}} \cdot \lambda_n(\Lambda).$$

Lemma 2.2. *(Adapted from [16, Lemma 5.3]) Let $m, n \geq 1$ and q a prime integer, with $m \geq 2n \ln q$. For $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ we define $\Lambda_{\mathbf{A}}^{\perp}$ as the lattice $\{\boldsymbol{x} \in \mathbb{Z}^m : \mathbf{A}\boldsymbol{x} = \mathbf{0} \bmod q\}$. Then*

$$\forall \varepsilon < 1/2 : \Pr_{\mathbf{A} \hookleftarrow U(\mathbb{Z}_q^{n\times m})}\left[\eta_\varepsilon(\Lambda_{\mathbf{A}}^{\perp}) \geq 4\sqrt{\frac{\ln(4m/\varepsilon)}{\pi}}\right] \leq q^{-n}.$$

Lemma 2.3. *(Adapted from [16, Cor. 2.8]) Let Λ, Λ' be n-dimensional lattices with $\Lambda' \subseteq \Lambda$ and $\varepsilon \in (0, 1/2)$. Then for any $\boldsymbol{c} \in \mathbb{R}^n$ and $s \geq \eta_\varepsilon(\Lambda')$ and any $x \in \Lambda/\Lambda'$ we have*

$$(D_{\Lambda,s,\boldsymbol{c}} \bmod \Lambda')(x) \in \left[\frac{1-\varepsilon}{1+\varepsilon}, \frac{1+\varepsilon}{1-\varepsilon}\right] \cdot \frac{\det(\Lambda)}{\det(\Lambda')}.$$

2.2. *The SIS, LWE, and LWR Problems*

The Small Integer Solution (SIS) problem was introduced by Ajtai [2]. It serves as a security foundation for numerous cryptographic primitives, including, among many others, hash functions [2] and signatures [11,16].

Definition 2.4. Let $m \geq n \geq 1$ and $q \geq 2$ be integers, and β a positive real. The $\mathrm{SIS}_{n,m,q,\beta}$ problem is as follows: given $\mathbf{A} \hookleftarrow U(\mathbb{Z}_q^{n\times m})$, the goal is to find $\boldsymbol{x} \in \mathbb{Z}^m$ such that $\mathbf{A}\boldsymbol{x} = \mathbf{0} \bmod q$ and $0 < \|\boldsymbol{x}\| \leq \beta$.

The SIS problem was proven by Ajtai [2] to be at least as hard as some standard worst-case problems over Euclidean lattices, under specific parameter constraints. We refer to Gentry et al. [16] for an improved (and simplified) reduction.

The Learning With Errors (LWE) problem was introduced in 2005 by Regev [30,32]. LWE is also extensively used as a security foundation, for encryption schemes [16,32], fully homomorphic encryption schemes [8], and pseudorandom functions [3,7], among many others. Its definition involves the following distribution. Let χ be a distribution over $\mathbb{T}$, $q \geq 2$, $n \geq 1$ and $\boldsymbol{s} \in \mathbb{Z}_q^n$. A sample from $A_{\boldsymbol{s},\chi}$ is of the form $(\boldsymbol{a}, b) \in \mathbb{Z}_q^n \times \mathbb{T}$, with $\boldsymbol{a} \hookleftarrow U(\mathbb{Z}_q^n)$, $b = \frac{1}{q}\langle \boldsymbol{a}, \boldsymbol{s}\rangle + e$ and $e \hookleftarrow \chi$.

Definition 2.5. Let χ be a distribution over $\mathbb{T}$, $q \geq 2$, and $m \geq n \geq 1$. The search variant $\mathrm{sLWE}_{n,q,\chi,m}$ of the LWE problem is as follows: given m samples from $A_{\boldsymbol{s},\chi}$ for some $\boldsymbol{s} \in \mathbb{Z}_q^n$, the goal is to find $\boldsymbol{s}$. The decision variant $\mathrm{LWE}_{n,q,\chi,m}$ consists in distinguishing between the distributions $(A_{\boldsymbol{s},\chi})^m$ and $U(\mathbb{Z}_q^n \times \mathbb{T})^m$, where $\boldsymbol{s} \hookleftarrow U(\mathbb{Z}_q^n)$.

Definition 2.6. The $\mathrm{sbinLWE}_{n,q,\chi,m}$ (resp. $\mathrm{binLWE}_{n,q,\chi,m}$) for any error distribution χ denotes the $\mathrm{sLWE}_{n,q,\chi,m}$ problem (resp. $\mathrm{LWE}_{n,q,\chi,m}$ problem) when the vector $\boldsymbol{s}$ is uniformly sampled in $\{0, 1\}^n$.

Bogdanov et al. [4], the secret $\boldsymbol{s}$ can be drawn from any distribution over $\{0, 1\}^n$ similar to what we defined above. It would be more consistent with the definition of sLWE to let the secret $\boldsymbol{s}$ be arbitrary, but it does not seem possible to prove equivalence via the random self reducibility property of LWE. A less direct reduction from worst-case sbinLWE to uniform-secret sbinLWE is as follows: worst-case sbinLWE reduces to LWE, then Goldwasser et al. [15] and Brakerski et al. [6, Theorem 4.1] provide reductions from LWE to binLWE, and finally [4] contains a reduction from binLWE to uniform-secret sbinLWE. In any case, we will only use uniform-secret sbinLWE so we stick to this variant in the present article. In some cases, it is convenient to use an error distribution χ whose support is $\mathbb{T}_q$. In these cases, the definition of LWE is adapted such that $U(\mathbb{Z}_q^n \times \mathbb{T})$ is replaced by $U(\mathbb{Z}_q^n \times \mathbb{T}_q)$. Note also that for a fixed number of samples m, we can represent the LWE samples using matrices. The $\boldsymbol{a}_i$'s form the rows of a matrix $\mathbf{A}$ uniform in $\mathbb{Z}_q^{m \times n}$, and the scalar product is represented by the product between $\mathbf{A}$ and $\boldsymbol{s}$.

Regev [32] gave a quantum reduction from standard worst-case problems over Euclidean lattices to sLWE and LWE, under specific parameter constraints. Classical (but weaker) reductions have later been obtained (see [6,28]). We will use the following sample-preserving search to decision reduction for LWE.

Theorem 2.7. *(Adapted from [23, Proposition 4.10] If $q \leq \mathrm{poly}(m, n)$ is prime and the error distribution χ has support in $\mathbb{T}_q$, then there exists a reduction from $\mathrm{sLWE}_{n,q,\chi,m}$ to $\mathrm{LWE}_{n,q,\chi,m}$ that is polynomial in n and m.*

For integers $p, q \geq 2$, the rounding function from $\mathbb{Z}_q$ to $\mathbb{Z}_p$ is defined by

$$\lfloor x \rceil_p = \lfloor (p/q)\bar{x} \rfloor \pmod{p},$$

where $\bar{x} \in \mathbb{Z}$ is any integer congruent to x modulo q. This can also be extended componentwise to vectors and matrices.

For a secret vector $\boldsymbol{s} \in \mathbb{Z}_q^n$, a sample $(\boldsymbol{a}, b)$ from the LWR distribution $B_{\boldsymbol{s}}$ over $\mathbb{Z}_q^n \times \mathbb{Z}_p$ is obtained by choosing a vector $\boldsymbol{a} \hookleftarrow U\left(\mathbb{Z}_q^n\right)$ and setting $b = \lfloor \langle \boldsymbol{a}, \boldsymbol{s} \rangle \rceil_p$.

Definition 2.8. The decision variant $\mathrm{LWR}_{n,q,p,m}$ of LWR consists in distinguishing between the distributions $(B_{\boldsymbol{s}})^m$ and $U(\mathbb{Z}_q^n \times \mathbb{Z}_p)^m$, where $\boldsymbol{s} \hookleftarrow U(\mathbb{Z}_q^n)$.

The LWR problem was introduced in Banerjee et al. [7] and used there and in subsequent works to construct pseudorandom functions (PRFs) based on the hardness of LWE.

2.3. *The Rényi Divergence*

For any two discrete probability distributions P and Q such that $\mathrm{Supp}(P) \subseteq \mathrm{Supp}(Q)$ and $a \in (1, +\infty)$, we define the Rényi divergence of order a by

$$R_a(P\|Q) = \left(\sum_{x \in \mathrm{Supp}(P)} \frac{P(x)^a}{Q(x)^{a-1}} \right)^{\frac{1}{a-1}} .$$

We omit the a subscript when $a = 2$. We define the Rényi divergences of orders 1 and $+\infty$ by

$$R_1(P\|Q) = \exp\left(\sum_{x \in \mathrm{Supp}(P)} P(x) \log \frac{P(x)}{Q(x)} \right) \text{ and } R_\infty(P\|Q) = \max_{x \in \mathrm{Supp}(P)} \frac{P(x)}{Q(x)}.$$

The definitions are extended in the natural way to continuous distributions. The divergence R_1 is the (exponential of) the Kullback–Leibler divergence.

For any fixed P, Q, the function $a \mapsto R_a(P\|Q) \in (0, +\infty]$ is non-decreasing, continuous over $(1, +\infty)$, tends to $R_\infty(P\|Q)$ when a grows to infinity, and if $R_a(P\|Q)$ is finite for some a, then $R_a(P\|Q)$ tends to $R_1(P\|Q)$ when a tends to 1 (we refer to Van Erven et al. [35] for proofs). A direct consequence is that if $P(x)/Q(x) \leq c$ for all $x \in \mathrm{Supp}(P)$ and for some constant c, then $R_a(P\|Q) \leq R_\infty(P\|Q) \leq c$. In the same setup, we have $\Delta(P, Q) \leq c/2$.

The following properties can be considered the multiplicative analogues of those of the SD. We refer to Refs. [21,35] for proofs.

Lemma 2.9. *Let $a \in [1, +\infty]$. Let P and Q denote distributions with $\mathrm{Supp}(P) \subseteq \mathrm{Supp}(Q)$. Then the following properties hold:*

- ***Log. Positivity*** $R_a(P\|Q) \geq R_a(P\|P) = 1$.
- ***Data Processing Inequality*** $R_a(P^f\|Q^f) \leq R_a(P\|Q)$ *for any function f, where P^f (respectively, Q^f) denotes the distribution of $f(y)$ induced by sampling $y \hookleftarrow P$ (respectively, $y \hookleftarrow Q$).*
- ***Multiplicativity*** *Assume P and Q are two distributions of a pair of random variables (Y_1, Y_2). For $i \in \{1, 2\}$, let P_i (resp. Q_i) denote the marginal distribution of Y_i under P (resp. Q), and let $P_{2|1}(\cdot|y_1)$ (resp. $Q_{2|1}(\cdot|y_1)$) denote the conditional distribution of Y_2 given that $Y_1 = y_1$. Then we have:*
  - · $R_a(P\|Q) = R_a(P_1\|Q_1) \cdot R_a(P_2\|Q_2)$ *if Y_1 and Y_2 are independent for $a \in [1, \infty]$.*
  - · $R_a(P\|Q) \leq R_\infty(P_1\|Q_1) \cdot \max_{y_1 \in X} R_a(P_{2|1}(\cdot|y_1)\|Q_{2|1}(\cdot|y_1))$.
- ***Probability Preservation*** *Let $E \subseteq \mathrm{Supp}(Q)$ be an arbitrary event. If $a \in (1, +\infty)$, then $Q(E) \geq P(E)^{\frac{a}{a-1}}/R_a(P\|Q)$. Further, we have*

$$Q(E) \geq P(E)/R_\infty(P\|Q).$$

Let P_1, P_2, P_3 *be three distributions with* $\mathrm{Supp}(P_1) \subseteq \mathrm{Supp}(P_2) \subseteq \mathrm{Supp}(P_3)$. *Then we have:*

- ***Weak Triangle Inequality***

$$R_a(P_1 \| P_3) \leq \begin{cases} R_a(P_1 \| P_2) \cdot R_\infty(P_2 \| P_3), \\ R_\infty(P_1 \| P_2)^{\frac{a}{a-1}} \cdot R_a(P_2 \| P_3) & \textit{if } a \in (1, +\infty). \end{cases}$$

Getting back to the setup in which $P(x)/Q(x) \leq c$ for all $x \in \mathrm{Supp}(P)$ and for some constant c, the RD probability preservation property above is relevant even for large c, whereas the analogous SD probability preservation property starts making sense only when $c < 2$.

Pinsker's inequality is the analogue of the probability preservation property for $a = 1$: for an arbitrary event $E \subseteq \mathrm{Supp}(Q)$, we have $Q(E) \geq P(E) - \sqrt{\ln R_1(P\|Q)/2}$ (see [27, Lemma 1] for a proof). Analogously to the statistical distance, this probability preservation property is useful for unlikely events E only if $\ln R_1(P\|Q)$ is very small. We refer to Sect. 3 for additional comments on this property.

2.4. *Some RD Bounds*

As we have already seen, if two distributions are close in a uniform sense, then their RD is small. We observe the following immediate consequence of Lemma 2.3, that allows replacing the SD with the RD in the context of smoothing arguments, in order to save on the required parameter s. In applications of Lemma 2.3, it is customary to use $s \geq \eta_\varepsilon(\Lambda')$ with $\varepsilon \leq 2^{-\lambda}$, in order to make the distribution $D_{\Lambda/\Lambda',s,\boldsymbol{c}} = D_{\Lambda,s,\boldsymbol{c}} \bmod \Lambda'$ within SD $2^{-\lambda}$ of the uniform distribution $U(\Lambda/\Lambda')$. This translates via Lemma 2.1 to use $s = \Omega(\sqrt{\lambda + \log n} \cdot \lambda_n(\Lambda'))$. If using an RD bound, the fact that $R_\infty(D_{\Lambda/\Lambda',s,\boldsymbol{c}} \| U_{\Lambda/\Lambda'}) = O(1)$ suffices for the application: one can take $\varepsilon = O(1)$ in the corollary below, which translates to just $s = \Omega(\sqrt{\log n} \cdot \lambda_n(\Lambda'))$, saving a factor $\Theta(\sqrt{\lambda})$.

Lemma 2.10. *Let* Λ, Λ' *be* n*-dimensional lattices with* $\Lambda' \subseteq \Lambda$ *and* $\varepsilon \in (0, 1/2)$. *Let* $D_{\Lambda/\Lambda',s,\boldsymbol{c}}$ *for any* $\boldsymbol{c} \in \mathbb{R}^n$ *denote the distribution on* Λ/Λ' *induced by sampling from* $D_{\Lambda,s,\boldsymbol{c}}$ *and reducing modulo* Λ', *and let* $U_{\Lambda/\Lambda'}$ *denote the uniform distribution on* Λ/Λ'. *Then for* $s \geq \eta_\varepsilon(\Lambda')$, *we have*

$$R_\infty(D_{\Lambda/\Lambda',s,\boldsymbol{c}} \| U_{\Lambda/\Lambda'}) \leq \frac{1+\varepsilon}{1-\varepsilon}.$$

In our hardness analysis of the LWR problem, the following Gaussian tail-cut lemma is used. It bounds the RD of order ∞ between a continuous Gaussian D_α and the same Gaussian with its tail cut to be B-bounded, that we denote by $D'_{\alpha,B}$. This allows, via an application of the RD probability preservation property, to conclude that any algorithm with success probability ε for m-sample search LWE with noise coordinates sampled from the tail-cut Gaussian $D'_{\alpha,B}$, is also an algorithm for LWE with noise coordinates sampled from the true Gaussian D_α with success probability $\geq \varepsilon/O(1)$, as long as $B = \Omega(\alpha \cdot \sqrt{\log m})$. This improves upon the bound $B = \Omega(\alpha \cdot \sqrt{\log(m \cdot \varepsilon^{-1})})$ that one obtains with an application of the SD to get the same conclusion.

Lemma 2.11. *Let $D'_{\alpha,B}$ denote the continuous distribution on $\mathbb{R}$ obtained from D_α by cutting its tail (by rejection sampling) to be B-bounded. Then we have*

$$R_\infty(D'_{\alpha,B} \| D_\alpha) \leq \frac{1}{1 - \exp(-\pi B^2/\alpha^2)}.$$

Furthermore, for m independent samples, we have $R_\infty((D'_{\alpha,B})^m \| (D_\alpha)^m) \leq \exp(1)$ if $B \geq \alpha \cdot \sqrt{\ln(2m)/\pi}$.

Proof. For $x \in \mathbb{R}$, we have $D'_{\alpha,B}(x) = c \cdot D_\alpha(x)$ for $|x| < B$ and $D'_{\alpha,B}(x) = 0$ otherwise, where c is a normalization constant such that $\int_{-\infty}^{\infty} D'_{\alpha,B}(x)dx = 1$. It follows that $c = \frac{1}{1-2Q_\alpha(B)}$, where $Q_\alpha(B) = \int_B^\infty D_\alpha(x)dx$ is the tail probability $\Pr_{z \hookleftarrow D_\alpha}[z \geq B]$. By a standard Gaussian tail bound, we have $Q_\alpha(B) \leq \frac{1}{2} \cdot \exp(-\pi B^2/\alpha^2)$, and hence $c \leq \frac{1}{1-\exp(-\pi B^2/\alpha^2)}$. The first part of the lemma now follows from the observation that $R_\infty(D'_{\alpha,B} \| D_\alpha) = \max_x \frac{D'_{\alpha,B}(x)}{D_\alpha(x)} = c$. For the second part of the lemma, observe that $c \leq \exp(4Q_\alpha(B))$ if $2Q_\alpha(B) \leq 1/2$ using the inequality $1-x \geq \exp(-2x)$ for $0 < x \leq 1/2$. It follows by the multiplicativity property of RD that $R_\infty((D'_{\alpha,B})^m \| (D_\alpha)^m) \leq \exp(4mQ_\alpha(B)) \leq \exp(1)$ if $2Q_\alpha(B) \leq \frac{1}{2m}$. The latter condition is satisfied by the above tail bound on $Q_\alpha(B)$ if $B \geq \alpha \cdot \sqrt{\ln(2m)/\pi}$. □

3. Application to Lattice-Based Signature Schemes

In this section, we use the RD to improve the security proofs of the BLISS signature scheme [11], allowing to take smaller parameters for any fixed security level.

More precisely, we show that the use of RD in place of SD leads to significant savings in the required precision of integers sampled according to a discrete Gaussian distribution in the security analysis of lattice-based signature schemes. These savings consequently lower the precomputed table storage for sampling discrete Gaussians with the method described in Refs. [11,27]. In Tables 1 and 2, we provide a numerical comparison of RD and SD based on an instantiations of BLISS-IV and BLISS-I.

Discrete Gaussian Sampling In the BLISS signature scheme [11] (and similarly in earlier variants [22]), each signature requires the signing algorithm to sample $O(n)$ independent integers from the 1-dimensional discrete Gaussian distribution $D_{\mathbb{Z},s}$, where $s = O(m)$ is the deviation parameter (here the variable m denotes a parameter related to the underlying lattice dimension, and is typically in the order of several hundreds).[2]

Ducas et al. [11], a particularly efficient sampling algorithm for $D_{\mathbb{Z},s}$ is presented. To produce a sample from $D_{\mathbb{Z},s}$, this algorithm samples about $\ell = \lfloor \log((k-1) \cdot (k-1+2k \cdot \tau\sigma_2) \rfloor + 1$ Bernoulli random variables of the form $B_{\exp(-\pi 2^i/s^2)}$ for $0 \leq i \leq \ell - 1$. Here, $\sigma_2 = \frac{1}{\sqrt{2\ln(2)}}$ is the standard deviation of a 'small width' Gaussian (sampled by Algorithm 10 in Ducas et al. [11]), $k = \frac{s}{\sigma_2 \cdot \sqrt{2\pi}}$ is the standard deviation 'amplification

[2] Note that [11,22] consider the unnormalized Gaussian function $\rho'_{\sigma,\boldsymbol{c}}(\boldsymbol{x}) = \exp(-\|\boldsymbol{x}-\boldsymbol{c}\|/(2\sigma^2))$ instead of $\rho_{s,\boldsymbol{c}}$. We have $\rho_{s,\boldsymbol{c}} = \rho'_{\sigma,\boldsymbol{c}}$ when $\sigma = s/\sqrt{2\pi}$.

factor' (in Algorithm 11 in Ducas et al. [11]), and τ is the tail-cut factor for the 'small width' Gaussian samples (i.e., those samples are cut by rejection sampling to be less than $\tau \cdot \sigma_2$). To sample the required Bernoulli random variables $B_{\exp(-\pi 2^i/s^2)}$, the authors of [11] use a precomputed table of the probabilities $c_i = \exp(-\pi 2^i/s^2)$, for $0 \leq i \leq \ell - 1$. Since these probabilities are real numbers, they must be truncated to some bit precision p in the precomputed table, so that truncated values $\tilde{c}_i = c_i + \varepsilon_i$ are stored, where $|\varepsilon_i| \leq 2^{-p} c_i$ are the truncation errors.

In previous works, the precision was determined by an analysis either based on the statistical distance (SD) [11] or the Kullback–Leibler divergence (KLD) [27]. In this section, we review and complete these methods, and we propose an RD-based analysis that in some cases leads to bigger savings, asymptotically and in practice, in particular for larger security levels and or smaller number of sign queries, when the number of attack sign queries is significantly less than $2^{\lambda/2}$ for security level λ (see Tables 1, 2). More precisely, we give sufficient lower bounds on the precision p in terms of the number of signing queries q_s and security parameter λ to ensure security level λ for the scheme implemented with truncated values against adversaries making $\leq q_s$ signing queries in time T, assuming that the scheme implemented with *untruncated* (exact) values has security level $\lambda + 1$ (i.e., our truncated scheme loses at most 1 bit of security with respect to the untruncated scheme).

Here, and in the following analysis, we say that a scheme has security level λ against (T, q_s, ε) forging adversaries running in time T, making q_s sign queries (where each sign query involves $\ell \cdot m$ Bernoulli samples), and succeeding with probability ε, if $T/\varepsilon \geq 2^\lambda$ for all adversaries with $T \geq Q = q_s \cdot \ell \cdot m$ and $q_s \geq 1$ (we count each Bernoulli sampling in signing queries as a unit time operation, so that $T \geq Q$, where Q is the total number of Bernoulli samples over all signing queries).

For any adversary, the distributions Φ' and Φ denote the signatures in the view of the adversary in the untruncated (resp. truncated) cases.

SD-based Analysis [11] Any forging adversary $\mathcal{A}$ with success probability $\geq \varepsilon$ in time T on the scheme implemented with truncated Gaussian has a success probability $\varepsilon' \geq \varepsilon - \Delta(\Phi, \Phi')$ against the scheme implemented with perfect Gaussian sampling in time T'. We guarantee a security level λ for truncated scheme if $T/\varepsilon < 2^\lambda$. This means that an adversary $\mathcal{A}'$ against the untruncated scheme has $T'/\varepsilon' \leq (2T)/\varepsilon$ if $\varepsilon' \geq \varepsilon/2$. Therefore, we select parameters to handle adversaries with success probabilities $\geq \varepsilon/2$ against the untruncated scheme; we can set the required precision p so that $\Delta(\Phi, \Phi') \leq \varepsilon/2$. Each signature requires $\ell \cdot m$ samples from the Bernoulli random variables $(B_{\tilde{c}_i})_i$. To ensure security against q_s signing queries, each of the truncated Bernoulli random variables $B_{\tilde{c}_i}$ should be within SD $\Delta(\Phi, \Phi')/(\ell \cdot m \cdot q_s)$ of the desired B_{c_i} (by the union bound). Using $\Delta(B_{\tilde{c}_i}, B_{c_i}) = |\varepsilon_i| \leq 2^{-p} c_i \leq 2^{-p-1}$ leads to a precision requirement

$$p \geq \log(\ell \cdot m \cdot q_s / \Delta(\Phi, \Phi')) \geq \log\left(\frac{\ell \cdot m \cdot q_s}{\varepsilon}\right).$$

Letting $Q = \ell \cdot m \cdot q_s$ it is sufficient to take $p \geq \log(\frac{Q}{\varepsilon})$. For each $\ell \cdot m \leq Q \leq 2^\lambda$, the maximum value of $\frac{Q}{\varepsilon}$ under the constraint $\frac{T}{\varepsilon} \leq 2^\lambda$ is $\frac{Q}{T} \cdot 2^\lambda$ which in turn has maximum value 2^λ using $T \geq Q$. Therefore, the SD-based precision requirement for truncated scheme security level λ is

$$p \geq \lambda. \tag{1}$$

The overall precomputed table is hence of bit size $L_{\text{SD}} = p \cdot \ell \geq \log(\ell \cdot m \cdot q_s/\varepsilon) \cdot \ell$.

One may also set the precision p_i depending on i for $0 \leq i \leq \ell - 1$. It is sufficient to set

$$Q \cdot 2^{-p_i} c_i \leq \varepsilon/2.$$

Hence, since the maximum of Q/ε is $T/\varepsilon \leq 2^\lambda$, the precision p_i is

$$p_i \geq \lambda + 1 + \log\left(\min\left(c_i, 1 - c_i\right)\right), \; 0 \leq i \leq \ell - 1. \tag{2}$$

The bit size of the overall precomputed table can be computed as a sum of the above p_i's. The min in the precision estimate above exploits the symmetry of the Bernoulli variable to decrease the bit size of the precomputed table (i.e., we may sample $B_{1-\tilde{c}_i}$ and flip the sampled bit to get a bit with distribution $B_{\tilde{c}_i}$).

KLD-Based Analysis [27] Pöppelman et al. [27] replace the SD-based analysis by a KLD-based analysis (i.e., using the RD of order $a = 1$) to reduce the precision p needed in the precomputed table. They show that any forging adversary $\mathcal{A}$ with success probability ε on the scheme implemented with truncated Gaussian has a success probability $\varepsilon' \geq \varepsilon - \sqrt{\ln R_1(\Phi \| \Phi')/2}$ on the scheme implemented with perfect Gaussian (see remark at the end of Sect. 2.3). By the multiplicative property of the RD over the $Q = \ell \cdot m \cdot q_s$ independent Bernoulli samples needed for signing q_s times, we get that $R_1(\Phi \| \Phi') \leq (\max_{1 \leq i \leq \ell} R_1(B_{\tilde{c}_i} \| B_{c_i}))^{\ell \cdot m \cdot q_s}$. Now, we have:

$$\begin{aligned} \ln R_1(B_{\tilde{c}_i} \| B_{c_i}) &= (1 - c_i - \varepsilon_i) \ln \frac{1 - c_i - \varepsilon_i}{1 - c_i} + (c_i + \varepsilon_i) \ln \frac{c_i + \varepsilon_i}{c_i} \\ &\leq -(1 - c_i - \varepsilon_i) \frac{\varepsilon_i}{1 - c_i} + (c_i + \varepsilon_i) \frac{\varepsilon_i}{c_i} = \frac{\varepsilon_i^2}{(1 - c_i)c_i}. \end{aligned}$$

Exploiting the symmetry of the distribution, $|\varepsilon_i| \leq 2^{-p} \min(c_i, 1 - c_i)$, we obtain $\ln R_1(B_{\tilde{c}_i} \| B_{c_i}) = 2^{-2p} \min(\frac{c_i}{1-c_i}, \frac{1-c_i}{c_i}) \leq 2^{-2p}$. Therefore, we get $\varepsilon' \geq \varepsilon - \sqrt{Q \cdot 2^{-2p-1}}$. We can select parameters such that $\sqrt{Q \cdot 2^{-2p-1}} \leq \varepsilon/2$. This leads to a precision requirement

$$p \geq \frac{1}{2} \log\left(\frac{Q}{\varepsilon^2}\right) + \frac{1}{2}. \tag{3}$$

To minimize the required precision, if the attacker has run-time $T < 2^\lambda$, makes $Q \leq T$ queries, and has success probability $\epsilon \geq T/2^\lambda$, we assume, as in Pöppelman et al. [27], that the attacker is first converted, by re-running it $\approx 2^\lambda/T$ times with independent public keys and random coins and returning the forgery from any successful run, to an attacker with run-time $\widehat{T} = (2^\lambda/T) \cdot T = 2^\lambda$, making $\widehat{Q} = 2^\lambda \cdot (Q/T)$ queries, and having success probability $\widehat{\varepsilon} \geq 1 - (1 - \varepsilon)^{2^\lambda/T} \geq 1 - \exp\left(-2^\lambda/(T/\varepsilon)\right) \geq 1 - \exp(-1) \geq 0.63$. We remark that this new attacker works in a *multi-key* model, in which an attacker gets as input $2^\lambda/T$ keys, and outputs a forgery for any one of them. Then, since $\frac{\widehat{Q}}{\varepsilon^2} \leq (2^\lambda \cdot Q/T)/0.63^2 \leq 2^\lambda/0.63^2$ using $Q \leq T$, the required precision is

$$p \geq \frac{1}{2}\log\left(\frac{2^\lambda}{0.63^2}\right) + \frac{1}{2} \approx \frac{\lambda}{2} + 1.2. \tag{4}$$

The overall precomputed table is hence of bit size $L_{\mathrm{KLD}} \geq (\lambda/2 + 1.2) \cdot \ell$.

One may also set the precision p_i depending on i. It is sufficient to set

$$\sqrt{\widehat{Q} \cdot \frac{\left(2^{-p_i}\min(c_i, 1-c_i)\right)^2}{2(1-c_i)c_i}} \leq \frac{\widehat{\varepsilon}}{2}.$$

Hence, since as above the maximum of $\frac{\widehat{Q}}{\varepsilon^2}$ is $\leq 2^\lambda/0.63^2$ using $Q \leq T$, the precision p_i is

$$p_i \geq \frac{\lambda}{2} + 1.2 + \frac{1}{2}\log\left(\min\left(\frac{c_i}{1-c_i}, \frac{1-c_i}{c_i}\right)\right),\ 0 \leq i \leq \ell - 1. \tag{5}$$

R_∞-based analysis. The probability preservation property of the Rényi divergence from Lemma 2.9 is multiplicative for $a > 1$ (rather than additive for $a = 1$). Here we use the order $a = \infty$. This property gives that any forging adversary $\mathcal{A}$ having success probability ε on the scheme implemented with truncated Gaussian sampling has a success probability $\varepsilon' \geq \varepsilon/R_\infty(\Phi\|\Phi')$ on the scheme implemented with perfect Gaussian. If $R = R_\infty(\Phi\|\Phi') \leq O(1)$, then $\varepsilon' = \Omega(\varepsilon)$. By the multiplicative property of the RD over the $Q = \ell \cdot m \cdot q_s$ samples needed for signing q_s times, we have $R_\infty(\Phi\|\Phi') \leq \prod_{i \leq Q} R_\infty(B_{\tilde{c}_i}\|B_{c_i})$. By our assumption that $c_i \leq 1/2$, we have $R_\infty(B_{\tilde{c}_i}\|B_{c_i}) = 1 + |\varepsilon_i|/c_i \leq 1 + 2^{-p}$. Therefore, we get $R_\infty(\Phi\|\Phi') \leq (1 + 2^{-p})^Q$ and hence $\varepsilon' \geq \varepsilon/(1 + 2^{-p})^Q$. We select parameters to get adversaries with success probabilities $\geq \varepsilon/2$ against the untruncated scheme and hence set the precision so that $(1 + 2^{-p})^Q \leq 2$. Using the inequality $1 + x \leq \exp(x)$, this yields a sufficient precision requirement

$$p \geq \log(Q) + \log(1/\ln(2)) \approx \lambda_Q + 0.16, \tag{6}$$

where $\lambda_Q = \log Q$. Overall, we get a precomputed table of bit size $L_{\mathrm{RD}} = \lambda_Q \cdot \ell$. In terms of the security parameter λ, the precision requirement (6) for R_∞ is lower than the requirement (4) for R_1 if the number of on-line queries Q is smaller than $2^{\lambda/2}$. In practice this condition may be satisfied, especially for larger security parameters λ (see numberical examples below).

R_a-based analysis. We may also consider R_a-based analysis for general $a > 1$. It should be noted that the reductions here are not tight: for R_a-based analysis with $a > 1$, the probability preservation shows $\varepsilon' > \varepsilon^{a/(a-1)}/R_a(\Phi\|\Phi')$. The Rényi divergence can be computed, as follows

$$
\begin{aligned}
(R_a(B_{\tilde{c}_i}\|B_{c_i}))^{a-1} &= \frac{(1-c_i-\varepsilon_i)^a}{(1-c_i)^{a-1}} + \frac{(c_i+\varepsilon_i)^a}{c_i^{a-1}} \\
&= (1-c_i-\varepsilon_i)\left(1-\frac{\varepsilon_i}{1-c_i}\right)^{a-1} + (c_i+\varepsilon_i)\left(1+\frac{\varepsilon_i}{c_i}\right)^{a-1}.
\end{aligned}
$$

If a is much smaller than 2^p, we obtain

$$
\begin{aligned}
(R_a(B_{\tilde{c}_i}\|B_{c_i}))^{a-1} &\approx (1-c_i-\varepsilon_i)\left(1-\frac{(a-1)\varepsilon_i}{1-c_i}+\frac{(a-1)(a-2)}{2}\cdot\frac{\varepsilon_i^2}{(1-c_i)^2}\right) \\
&\quad + (c_i+\varepsilon_i)\left(1+\frac{(a-1)\varepsilon_i}{c_i}+\frac{(a-1)(a-2)}{2}\cdot\frac{\varepsilon_i^2}{c_i^2}\right) \\
&\approx 1+\frac{a(a-1)}{2}\cdot\frac{\varepsilon_i^2}{c_i(1-c_i)} \le 1+\frac{a(a-1)}{2}\cdot 2^{-2p}.
\end{aligned}
$$

For instance, if we take $a = 2$, we have $R_2(B_{\tilde{c}_i}\|B_{c_i}) \le 1 + 2^{-2p}$ and hence $\varepsilon' \ge \varepsilon^2/R_2(B_{\tilde{c}_i}\|B_{c_i})$. On the other hand, if a is much larger than 2^p, then we have

$$
\begin{aligned}
(R_a(B_{\tilde{c}_i}\|B_{c_i}))^{a-1} &= (1-c_i-\varepsilon_i)\left(1-\frac{\varepsilon_i}{1-c_i}\right)^{a-1} + (c_i+\varepsilon_i)\left(1+\frac{\varepsilon_i}{c_i}\right)^{a-1} \\
&\approx (c_i+\varepsilon_i)\exp\left(\frac{(a-1)\varepsilon_i}{c_i}\right).
\end{aligned}
$$

Hence the Rényi divergence satisfies

$$
R_a(B_{\tilde{c}_i}\|B_{c_i}) \approx (c_i+\varepsilon_i)^{1/(a-1)}\exp\left(\frac{\varepsilon_i}{c_i}\right) \approx 1+\frac{\varepsilon_i}{c_i}.
$$

As $a \to \infty$, we have $R_a(B_{\tilde{c}_i}\|B_{c_i}) \to 1 + 2^{-p}$.

Thus if the tightness of the reduction is not a concern, using R_a with small a reduces the precision requirement. Subsequent work [34] shows that by choosing an adequate a, tightness can be reached ($\varepsilon' \approx \varepsilon$) for the same number of queries. This may however lead to a slightly larger precision (compared to the case of using a tiny Rényi order a).

Numerical Examples

In Tables 1 and 2, we consider a numerical example which gives the lower bound on the precision p and table bit size for Gaussian sampling in the schemes BLISS-IV ($\lambda = 192$) and BLISS-I ($\lambda = 128$), and three settings for the number of sign queries $q_s = (2^{42}, 2^{50}, 2^{64})$ allowed for the adversary. In all cases, we assume that the 'small deviation' (positive) Gaussian samples of standard deviation $\sigma_2 = \frac{1}{\sqrt{2\ln(2)}}$ (sampled in Algorithm 11 of Ducas et al. [11]) are tail cut to $\tau\sigma_2$, where we set the tail-cut factor $\tau = \sqrt{2\ln(2mq_s)}$ by applying Lemma 2.11, to make the R_∞ bound $\le \exp(1)$ between the cut and uncut distributions, over all mq_s 'small deviation' Gaussian samples.

For the BLISS-IV parameters, we use $\lambda = 192$, $m = 1024$, $k = \lceil 271/\sigma_2\rceil = 320$, $\tau = \sqrt{2\ln(2mq_s)} \approx (7.3, 7.8, 8.7)$, $\ell = \lfloor \log((k-1)\cdot(k-1+2k\cdot\tau\sigma_2)\rfloor + 1 = 21$, $s =$

Table 1. Comparison of the precision needed to obtain 2^λ security for the finite precision BLISS-IV scheme against adversaries with off-line run-time $T \leq 2^\lambda$ and making less than q_s sign queries (resulting in $Q = \ell \cdot m \cdot q_s$ Bernoulli samples over all sign queries), assuming $\approx 2^{\lambda+1}$ security of the infinite precision scheme.

Method	Precision p	Example p	Example table bit sizes
SD [Eq. (1)]	λ	192, 192, 192	4032, 4032, 4032
SD [Eq. (2)]	$\lambda + 1 + \log c_i$	–	3882, 3882, 3882
KLD [Eq. (4)]	$\lambda/2 + 1.2$	97, 97, 97	2037, 2037, 2037
KLD [Eq. (5)]	$\lambda/2 + 1.2 + \log(\sqrt{\frac{c_i}{1-c_i}})$	–	1957, 1957, 1957
R_∞ [Eq. (6)]	$\lambda_Q + 0.16$	57, 65, 79	1197, 1365, 1659

The example numerical values of precision p and table size are for off-line security parameter $\lambda = 192$ and (in order) three cases $q_s = (2^{42}, 2^{50}, 2^{64})$ for the number of sign queries. Our R_∞ parameters are on the last line. The Bernoulli probabilities are $c_i = \exp(-\pi 2^i/s^2)$ for $i = 0, \ldots, \ell - 1$. For the BLISS-IV parameters, we use $m = 1024$, $\ell = 21$, $s = 682$ and $\tau \approx (7.3, 7.8, 8.7)$

Table 2. Comparison of the precision needed to obtain 2^λ security for the finite precision BLISS-I scheme against adversaries with off-line run-time $T \leq 2^\lambda$ and making less than q_s sign queries (resulting in $Q = \ell \cdot m \cdot q_s$ Bernoulli samples over all sign queries), assuming $\approx 2^{\lambda+1}$ security of the infinite precision scheme.

Method	Precision p	Example p	Example table bit sizes
SD [Eq. (1)]	λ	128, 128, 128	2560, 2560, 2560
SD [Eq. (2)]	$\lambda + 1 + \log c_i$	–	2429, 2429, 2429
KLD [Eq. (4)]	$\lambda/2 + 1.2$	65, 65, 65	1300, 1300, 1300
KLD [Eq. (5)]	$\lambda/2 + 1.2 + \log(\sqrt{\frac{c_i}{1-c_i}})$	–	1222, 1222, 1222
R_∞ [Eq. (6)]	$\lambda_Q + 0.16$	57, 65, 79	1140, 1300, 1580

The example numerical values of precision p and table size are for off-line security parameter $\lambda = 128$ and (in order) three cases $q_s = (2^{42}, 2^{50}, 2^{64})$ for the number of sign queries. Our R_∞ parameters are on the last line. The Bernoulli probabilities are $c_i = \exp(-\pi 2^i/s^2)$ for $i = 0, \ldots, \ell - 1$. For the BLISS-I parameters, we use $m = 1024$, $\ell = 20$, $s = 541$ and $\tau \approx (7.3, 7.8, 8.7)$

$\lceil\sqrt{2\pi} \cdot k \cdot \sigma_2\rceil = 682$ and $Q = \ell \cdot m \cdot q_s \approx (2^{56}, 2^{64}, 2^{78})$. For the BLISS-I parameters, we use $\lambda = 128$, $m = 1024$, $k = \lceil 215/\sigma_2 \rceil = 254$, $\tau = \sqrt{2\ln(2mq_s)} \approx (7.3, 7.8, 8.7)$, $\ell = \lfloor \log((k-1) \cdot (k - 1 + 2k \cdot \tau\sigma_2) \rfloor + 1 = 20$, $s = \lceil\sqrt{2\pi} \cdot k \cdot \sigma_2\rceil = 541$ and $Q = \ell \cdot m \cdot q_s \approx (2^{56}, 2^{64}, 2^{78})$. In all cases, we assume that the underlying BLISS scheme with perfect (infinite precision) Bernoulli sampling has security level $2^{\lambda+1}$.

Note that we assume, as is common in practice, that the allowed 'off-line' attack run-time $T = 2^\lambda$ is much bigger than the allowed 'on-line' number of sign queries q_s. This assumption may be satisfied in practice since in many applications the number of issued signatures q_s is limited by computation, communication and/or policy restrictions of the attacked user's application running the signing algorithm, whereas the 'off-line' run-time T depends only on the attacker's resources and may be much larger. For example, even for the scenario with the smallest allowed number of signatures $q_s = 2^{42}$ considered in the Tables, if the attacked user's signing algorithm runs on a single Intel Core i7 CPU at 3.4 GHz, it would take the attacker more than 17 years to collect all q_s signatures, even if the signer was continuously signing messages throughout this time.

4. Rényi Divergence and Distinguishing Problems

In this section, we prove Theorem 4.2 which allows to use the RD for distinguishing problems, and we show how to apply it to the dual-Regev encryption scheme.

4.1. *Problems with Public Sampleability*

A general setting one comes across in analyzing the security of cryptographic schemes has the following form. Let P denote a decision problem that asks to distinguish whether a given x was sampled from distribution X_0 or X_1, defined as follows:

$$X_0 = \{x : r \hookleftarrow \Phi, x \hookleftarrow D_0(r)\}, \quad X_1 = \{x : r \hookleftarrow \Phi, x \hookleftarrow D_1(r)\}.$$

Here r is some parameter that is sampled from the same distribution Φ in both X_0 and X_1. The parameter r then determines the conditional distributions $D_0(r)$ and $D_1(r)$ from which x is sampled in X_0 and X_1, respectively, given r. Now, let P' denote another decision problem that is defined similarly to P, except that in P' the parameter r is sampled from a different distribution Φ' (rather than Φ). Given r, the conditional distributions $D_0(r)$ and $D_1(r)$ are the same in P' as in P. Let X'_0 (resp. X'_1) denote the resulting marginal distributions of x in problem P'. Now, in the applications we have in mind, the distributions Φ' and Φ are "close" in some sense, and we wish to show that this implies an efficient reduction between problems P' and P, in the usual sense that every distinguisher with efficient run-time T and non-negligible advantage ε against P implies a distinguisher for P' with efficient run-time T' and non-negligible advantage ε'. In the classical situation, if the SD $\Delta(\Phi, \Phi')$ between Φ' and Φ is negligible, then the reduction is immediate. Indeed, for $b \in \{0, 1\}$, if p_b (resp. p'_b) denotes the probability that a distinguisher algorithm $\mathcal{A}$ outputs 1 on input distribution X_b (resp. X'_b), then we have, from the SD probability preservation property, that $|p'_b - p_b| \leq \Delta(\Phi, \Phi')$. As a result, the advantage $\varepsilon' = |p'_1 - p'_0|$ of $\mathcal{A}$ against P' is bounded from below by $\varepsilon - 2\Delta(\Phi, \Phi')$ which is non-negligible (here $\varepsilon = |p_1 - p_0|$ is the assumed non-negligible advantage of $\mathcal{A}$ against P).

Unfortunately, for general decision problems P, P' of the above form, it seems difficult to obtain an RD-based analogue of the above SD-based argument, in the weaker setting when the SD $\Delta(\Phi, \Phi')$ is non-negligible, but the RD $R = R(\Phi \| \Phi')$ is small. Indeed, the probability preservation property of the RD in Lemma 2.9 does not seem immediately useful in the case of general decision problems P, P'. With the above notations, it can be used to conclude that $p'_b \geq p_b^2/R$ but this does not allow us to usefully relate the advantages $|p'_1 - p'_0|$ and $|p_1 - p_0|$.

Nevertheless, we now make explicit a special class of "publicly sampleable" problems P, P' for which such a reduction can be made. In such problems, it is possible to efficiently sample from both distributions $D_0(r)$ (resp. $D_1(r)$) given the single sample x from the unknown $D_b(r)$. This technique is implicit in the application of RD in the reductions of Lyubashevsky et al. [19]: we abstract it and make it explicit in the following.

Before going ahead to state one of the main results of this paper, we recall Hoeffding's bound [17]:

Lemma 4.1. *Let $X_1, \ldots, X_N$ be independent random variables for which $a_i \leq X_i \leq b_i$. Let $\overline{X}$ denotes $\frac{X_1+\cdots+X_n}{N}$, then*

$$\mathbb{P}\left(\left|\overline{X} - \mathrm{E}\left[\overline{X}\right]\right| \geq t\right) \leq 2\exp\left(-\frac{2N^2t^2}{\sum_{i=1}^{N}(b_i - a_i)^2}\right),$$

is valid for all positive t and E denotes the expected value.

Theorem 4.2. *Let Φ, Φ' denote two distributions with* $\mathrm{Supp}(\Phi) \subseteq \mathrm{Supp}(\Phi')$*, and $D_0(r)$ and $D_1(r)$ denote two distributions determined by some parameter $r \in \mathrm{Supp}(\Phi')$. Let P, P' be two decision problems defined as follows:*

- *Problem P: distinguish whether input x is sampled from distribution X_0 or X_1, where*

$$X_0 = \{x : r \hookleftarrow \Phi, x \hookleftarrow D_0(r)\}, \quad X_1 = \{x : r \hookleftarrow \Phi, x \hookleftarrow D_1(r)\}.$$

- *Problem P': distinguish whether input x is sampled from distribution X'_0 or X'_1, where*

$$X'_0 = \{x : r \hookleftarrow \Phi', x \hookleftarrow D_0(r)\}, \quad X'_1 = \{x : r \hookleftarrow \Phi', x \hookleftarrow D_1(r)\}.$$

Assume that $D_0(\cdot)$ and $D_1(\cdot)$ satisfy the following public sampleability *property: there exists a sampling algorithm S with run-time T_S such that for all (r, b), given any sample x from $D_b(r)$:*

- $\mathsf{S}(0, x)$ *outputs a fresh sample distributed as $D_0(r)$ over the randomness of* S,
- $\mathsf{S}(1, x)$ *outputs a fresh sample distributed as $D_1(r)$ over the randomness of* S.

Then, given a T-time distinguisher $\mathcal{A}$ for problem P with advantage ε, we can construct a distinguisher $\mathcal{A}'$ for problem P' with run-time and distinguishing advantage, respectively, bounded from above and below by (for any $a \in (1, +\infty]$):

$$\frac{64}{\varepsilon^2}\log\left(\frac{8R_a(\Phi\|\Phi')}{\varepsilon^{a/(a-1)+1}}\right) \cdot (T_S + T) \text{ and } \frac{\varepsilon}{4 \cdot R_a(\Phi\|\Phi')} \cdot \left(\frac{\varepsilon}{2}\right)^{\frac{a}{a-1}}.$$

Proof. For each $\hat{r} \in \mathrm{Supp}(\Phi)$, and $b \in \{0, 1\}$, we let $p_b(\hat{r}) = \Pr_{x \hookleftarrow D_b(\hat{r})}(\mathcal{A}(x) = 1)$ and $p_b = \sum_{\hat{r} \in \mathrm{Supp}(\Phi)} p_b(\hat{r})\Phi(\hat{r})$. The advantage of $\mathcal{A}$ is defined as $|p_0 - p_1|$, which we assume is bigger than ε. Without loss of generality, we may assume that $p_0 > p_1$. Distinguisher $\mathcal{A}'$ is given an input x sampled from $D_b(r)$ for some r sampled from Φ' and some unknown $b \in \{0, 1\}$. For an ε' to be determined later, it runs distinguisher $\mathcal{A}$ on $N \geq 32\varepsilon^{-2}\log(4/\varepsilon')$ independent inputs sampled from $D_0(r)$ and $D_1(r)$ calling algorithm S on $(0, x)$ and $(1, x)$ to obtain estimates $\hat{p}_0$ and $\hat{p}_1$ for the acceptance probabilities $p_0(r)$ and $p_1(r)$ of $\mathcal{A}$ given as inputs samples from $D_0(r)$ and $D_1(r)$ (with the r fixed to the value used to sample the input x of $\mathcal{A}'$). By letting $t = \varepsilon/8$ and $N = 32\varepsilon^{-2}\log(4/\varepsilon')$ for X_i's being Bernoulli with probability $p_b(r)$ over $[a_i, b_i] = [0, 1]$, for $1 \leq i \leq N$, the Hoeffding's bound implies that, the estimation errors $|\hat{p}_0 - p_0|$ and $|\hat{p}_1 - p_1|$ are $< \varepsilon/8$

except with probability $< 2\exp(-2Nt^2) = \varepsilon'/2$ over the randomness of S. Then, if $\hat{p}_1 - \hat{p}_0 > \varepsilon/4$, distinguisher $\mathcal{A}'$ runs $\mathcal{A}$ on input x and returns whatever $\mathcal{A}$ returns, else distinguisher $\mathcal{A}'$ returns a uniformly random bit. This completes the description of distinguisher $\mathcal{A}'$.

Let $\mathcal{S}_1$ denote the set of r's such that $p_1(r) - p_0(r) \geq \varepsilon/2$, $\mathcal{S}_2$ denote the set of r's that are not in $\mathcal{S}_1$ and such that $p_1(r) - p_0(r) \geq 0$, and $\mathcal{S}_3$ denote all the remaining r's. Then:

- If $r \in \mathcal{S}_1$, then except with probability $< \varepsilon'$ over the randomness of S, we will have $\hat{p}_1 - \hat{p}_0 > \varepsilon/4$ and thus $\mathcal{A}'$ will output $\mathcal{A}(x)$. Thus, in the case $b = 1$, we have $\Pr[\mathcal{A}'(x) = 1 | r \in \mathcal{S}_1] \geq p_1(r) - \varepsilon'$ and in the case $b = 0$, we have $\Pr[\mathcal{A}'(x) = 1 | r \in \mathcal{S}_1] \leq p_0(r) + \varepsilon'$.
- Assume that $r \in \mathcal{S}_2$. Let $u(r)$ be the probability over the randomness of S that $\hat{p}_1 - \hat{p}_0 > \varepsilon/4$. Then $\mathcal{A}'$ will output $\mathcal{A}(x)$ with probability $u(r)$ and a uniform bit with probability $1 - u(r)$. Thus, in the case $b = 1$, we have $\Pr[\mathcal{A}'(x) = 1 | r \in \mathcal{S}_2] = u(r) \cdot p_1(r) + (1 - u(r))/2$, and in the case $b = 0$, we have $\Pr[\mathcal{A}'(x) = 1 | r \in \mathcal{S}_2] = u(r) \cdot p_0(r) + (1 - u(r))/2$.
- If $r \in \mathcal{S}_3$, except with probability $< \varepsilon'$ over the randomness of S, we have $\hat{p}_1 - \hat{p}_0 < \varepsilon/4$ and $\mathcal{A}'$ will output a uniform bit. Thus, in the case $b = 1$, we have $\Pr[\mathcal{A}'(x) = 1 | r \in \mathcal{S}_3] \geq 1/2 - \varepsilon'$, and in the case $b = 0$, we have $\Pr[\mathcal{A}'(x) = 1 | r \in \mathcal{S}_3] \leq 1/2 + \varepsilon'$.

Overall, the advantage of $\mathcal{A}'$ is bounded from below by:

$$\sum_{r \in \mathcal{S}_1} \Phi'(r) \left(p_1(r) - p_0(r) - 2\varepsilon'\right) + \sum_{r \in \mathcal{S}_2} \Phi'(r) u(r) \left(p_1(r) - p_0(r)\right) - \sum_{r \in \mathcal{S}_3} \Phi'(r) 2\varepsilon' \geq \Phi'(\mathcal{S}_1) \cdot \frac{\varepsilon}{2} - 2\varepsilon'.$$

By an averaging argument, the set $\mathcal{S}_1$ has probability $\Phi(\mathcal{S}_1) \geq \varepsilon/2$ under distribution Φ. Hence, by the RD probability preservation property (see Lemma 2.9), we have $\Phi'(\mathcal{S}_1) \geq (\varepsilon/2)^{\frac{a}{a-1}} / R_a(\Phi \| \Phi')$. The proof may be completed by setting $\varepsilon' = (\varepsilon/4) \cdot (\varepsilon/2)^{\frac{a}{a-1}} / R_a(\Phi \| \Phi')$. □

4.2. *Application to Dual-Regev Encryption*

Let m, n, q, χ be as in Definition 2.5 and Φ denote a distribution over $\mathbb{Z}_q^{m \times n}$. We define the LWE variant $\mathrm{LWE}_{n,q,\chi,m}(\Phi)$ as follows: Sample $\mathbf{A} \hookleftarrow \Phi$, $\boldsymbol{s} \hookleftarrow U(\mathbb{Z}_q^n)$, $\boldsymbol{e} \hookleftarrow \chi^m$ and $\boldsymbol{u} \hookleftarrow U(\mathbb{T}^m)$; The goal is to distinguish between the distributions $\left(\mathbf{A}, \frac{1}{q}\mathbf{A}\boldsymbol{s} + \boldsymbol{e}\right)$ and $(\mathbf{A}, \boldsymbol{u})$ over $\mathbb{Z}_q^{m \times n} \times \mathbb{T}^m$. Note that standard LWE is obtained by taking $\Phi' = U(\mathbb{Z}_q^{m \times n})$.

As an application to Theorem 4.2, we show that LWE with non-uniform and possibly statistically correlated $\boldsymbol{a}_i$'s of the samples $(\boldsymbol{a}_i, b_i)$'s (with b_i either independently sampled from $U(\mathbb{T})$ or close to $\langle \boldsymbol{a}_i, \boldsymbol{s} \rangle$ for a secret vector $\boldsymbol{s}$) remains at least as hard as standard LWE, as long as the RD $R(\Phi \| U)$ remains small, where Φ is the joint distribution of the given $\boldsymbol{a}_i$'s and U denotes the uniform distribution.

To show this result, we first prove in Corollary 4.3 that there is a reduction from $\mathrm{LWE}_{n,q,\chi,m}(\Phi')$ to $\mathrm{LWE}_{n,q,\chi,m}(\Phi)$ using Theorem 4.2 if $R_a(\Phi\|\Phi')$ is small enough. We then describe in Corollary 4.4 how to use this first reduction to obtain smaller parameters for the dual-Regev encryption. This allows us to save an $\Omega(\sqrt{\lambda/\log\lambda})$ factor in the Gaussian deviation parameter r used for secret key generation in the dual-Regev encryption scheme [16], where λ refers to the security parameter.

Corollary 4.3. *Let Φ and Φ' be two distributions over $\mathbb{Z}_q^{m\times n}$ with $\mathrm{Supp}(\Phi)\subseteq\mathrm{Supp}(\Phi')$. If there exists a distinguisher $\mathcal{A}$ against the $\mathrm{LWE}_{n,q,\chi,m}(\Phi)$ with run-time T and advantage $\varepsilon=o(1)$, then there exists a distinguisher $\mathcal{A}'$ against the $\mathrm{LWE}_{n,q,\chi,m}(\Phi')$ with run-time $T'=O(\varepsilon^{-2}\log\frac{R_a(\Phi\|\Phi')}{\varepsilon^{a/(a-1)}}\cdot(T+\mathrm{poly}(m,\log q)))$ and advantage*

$$\Omega\left(\frac{\varepsilon^{1+a/(a-1)}}{R_a(\Phi\|\Phi')}\right),$$

for any $a\in(1,+\infty]$.

Proof. Apply Theorem 4.2 with $r=\mathbf{A}\in\mathbb{Z}_q^{m\times n}$, $x=(\mathbf{A},\boldsymbol{b})\in\mathbb{Z}_q^{m\times n}\times\mathbb{T}^m$, $D_0(r)=(\mathbf{A},\mathbf{A}\cdot\boldsymbol{s}+\boldsymbol{e})$ with $\boldsymbol{s}\hookleftarrow U(\mathbb{Z}_q^n)$ and $\boldsymbol{e}\hookleftarrow\chi^m$, and $D_1(r)=(\mathbf{A},\boldsymbol{u})$ with $\boldsymbol{u}\hookleftarrow U(\mathbb{Z}_q^m)$. The sampling algorithm S is such that $\mathsf{S}(0,x)$ outputs $(\mathbf{A},\mathbf{A}\cdot\boldsymbol{s}'+\boldsymbol{e}')$ for $\boldsymbol{s}'\hookleftarrow U(\mathbb{Z}_q^n)$ and $\boldsymbol{e}'\hookleftarrow\chi^m$, while $\mathsf{S}(1,x)$ outputs $(\mathbf{A},\boldsymbol{u}')$ with $\boldsymbol{u}'\hookleftarrow U(\mathbb{Z}_q^m)$. □

We recall that the dual-Regev encryption scheme has a general public parameter $\mathbf{A}\in\mathbb{Z}_q^{m\times n}$, a secret key of the form $\mathrm{sk}=\boldsymbol{x}$ with $\boldsymbol{x}\hookleftarrow D_{\mathbb{Z}^m,r}$ and a public key of the form $\boldsymbol{u}=\mathbf{A}^t\boldsymbol{x}\bmod q$. A ciphertext for a message $M\in\{0,1\}$ is obtained as follows: Sample $\boldsymbol{s}\hookleftarrow U(\mathbb{Z}_q^n)$, $\boldsymbol{e}_1\hookleftarrow\chi^m$ and $e_2\hookleftarrow\chi$; return ciphertext

$$(\boldsymbol{c}_1,c_2)=\left(\frac{1}{q}\mathbf{A}\boldsymbol{s}+\boldsymbol{e}_1,\frac{1}{q}\langle\boldsymbol{u},\boldsymbol{s}\rangle+e_2+\frac{M}{2}\right)\in\mathbb{T}^m\times\mathbb{T}.$$

Corollary 4.4. *Suppose that q is prime, $m\geq 2n\log q$ and $r\geq 4\sqrt{\log(12m)/\pi}$. If there exists an adversary against the IND-CPA security of the dual-Regev encryption scheme with run-time T and advantage ε, then there exists a distinguishing algorithm for $\mathrm{LWE}_{n,q,\chi,m+1}$ with run-time $O((\varepsilon')^{-2}\log(\varepsilon')^{-1}\cdot(T+\mathrm{poly}(m)))$ and advantage $\Omega((\varepsilon')^2)$, where $\varepsilon'=\varepsilon-2q^{-n}$.*

Proof. Breaking the security of the dual-Regev encryption scheme as described above is at least as hard as $\mathrm{LWE}_{n,q,\chi,m+1}(\Phi)$ where Φ is obtained by sampling $\mathbf{A}\hookleftarrow U(\mathbb{Z}_q^{m\times n})$, $\boldsymbol{u}\hookleftarrow\mathbf{A}^t\cdot D_{\mathbb{Z}^m,r}\bmod q$ and returning the $(m+1)\times n$ matrix obtained by appending $\boldsymbol{u}^t$ at the bottom of $\mathbf{A}$. We apply Corollary 4.3 with $\Phi'=U(\mathbb{Z}_q^{(m+1)\times n})$.

Since q is prime, if $\mathbf{A}$ is full rank, then the multiplication by $\mathbf{A}^t$ induces an isomorphism between the quotient group $\mathbb{Z}^m/\Lambda_{\mathbf{A}}^{\perp}$ and $\mathbb{Z}_q^n$, where $\Lambda_{\mathbf{A}}^{\perp}=\{\boldsymbol{x}\in\mathbb{Z}^m:\mathbf{A}^t\cdot\boldsymbol{x}=\mathbf{0}\bmod q\}$. By Lemma 2.2, we have $\eta_{1/3}\left(\Lambda_{\mathbf{A}}^{\perp}\right)\leq 4\sqrt{\log(12m)/\pi}\leq r$, except for a fraction $\leq q^{-n}$

of the $\mathbf{A}$'s. Let Bad denote the union of such bad $\mathbf{A}$'s and the $\mathbf{A}$'s that are not full rank. We have $\Pr[\mathsf{Bad}] \leq 2q^{-n}$.

By the multiplicativity property of Lemma 2.9, we have:

$$R_\infty(\Phi \| \Phi') \leq \max_{\mathbf{A} \notin \mathsf{Bad}} R_\infty \left(D_{\mathbb{Z}^m, r} \bmod \Lambda_{\mathbf{A}}^{\perp} \| U_{\mathbb{Z}^m / \Lambda_{\mathbf{A}}^{\perp}} \right).$$

Thanks to Lemma 2.10, we know that the latter is ≤ 2. The result now follows from Corollary 4.3. □

In all applications, we are aware of, the parameters satisfy $m \leq \mathrm{poly}(\lambda)$ and $q^{-n} \leq 2^{-\lambda}$, where λ refers to the security parameter. The $r = \Omega(\sqrt{\log \lambda})$ bound of our Corollary 4.4, that results from using $\delta = 1/3$ in the condition $r \geq \eta_\delta \left(\Lambda_{\mathbf{A}}^{\perp}\right)$ in the RD-based smoothing argument of the proof above, improves on the corresponding bound $r = \Omega(\sqrt{\lambda})$ that results from the requirement to use $\delta = O(2^{-\lambda})$ in the condition $r \geq \eta_\delta \left(\Lambda_{\mathbf{A}}^{\perp}\right)$ in the SD-based smoothing argument of the proof of [16, Theorem 7.1], in order to handle adversaries with advantage $\varepsilon = 2^{-o(\lambda)}$ in both cases. Thus our RD-based analysis saves a factor $\Omega\left(\sqrt{\lambda / \log \lambda}\right)$ in the choice of r, and consequently of a^{-1} and q. (The authors of [16] specify a choice of $r = \omega(\sqrt{\log \lambda})$ for their scheme because they use in their analysis the classical "no polynomial attacks" security requirement, corresponding to assuming attacks with advantage $\varepsilon = \lambda^{-O(1)}$, rather than the stronger $\varepsilon = \omega(2^{-\lambda})$ but more realistic setting we take.)

5. Application to LWE with Uniform Noise

The LWE problem with noise uniform in a small interval was introduced first in Döttling et al. [12]. In that article, the authors exhibit a reduction from LWE with Gaussian noise, which relies on a new tool called *lossy codes*. The main proof ingredients are the construction of lossy codes for LWE (which are lossy for the uniform distribution in a small interval), and the fact that lossy codes are pseudorandom.

We note that the reduction from Döttling et al. [12] needs the number of LWE samples to be bounded by $\mathrm{poly}(n)$ and that it degrades the LWE dimension by a constant factor. The parameter β (when the interval of the noise is $[-\beta, \beta]$) should be at least $mn^\sigma \alpha$ where α is the LWE Gaussian noise parameter and $\sigma \in (0, 1)$ is an arbitrarily small constant.

Another hardness result for LWE with uniform noise can be obtained by composing the hardness result for Learning With Rounding (LWR) from Bogdanov et al. [4] (based on RD-based techniques inspired by an earlier version of our paper, see Theorem 6.1 in Sec. 6 and the discussion there) with the reduction of Chow [10] (see Theorem 6 in Bogdanov et al. [4]) from LWR to LWE with uniform noise. The resulting reduction maps the $\mathrm{LWE}_{n', q, D_\alpha, m}$ problem to the $\mathrm{LWE}_{n, q, U([-\beta, \beta]), m}$ problem with $n' = n / \log q$ and $\beta = \Omega(m\alpha / \sqrt{\log n})$, and hence, like the reduction of D ́ottling[12], it also degrades the LWE dimension.

We now provide an alternative reduction from the $\mathrm{LWE}_{n, q, D_\alpha, m}$ distinguishing problem to the $\mathrm{LWE}_{n, q, U([-\beta, \beta]), m}$ distinguishing problem, and analyze it using RD. Our reduction preserves the LWE dimension n, and is hence tighter in terms of dimension than the reductions from Döttling et al. [12] and Bogdanov et al. [4] discussed above. In terms of noise, our reduction requires that $\beta = \Omega(m\alpha/\log n)$ (so in this respect is slightly less tight than the reduction of Bogdanov et al. [4] by a factor $\sqrt{\log n}$).

We remark that the search-decision equivalence idea in the proof of Theorem 5.1 could be extended to show the hardness of the decision LWE problem with any noise distribution ψ, with respect to the hardness of LWE with Gaussian noise D_α if either ψ is 'close' to D_α in the sense of RD (i.e., $R(\psi \| D_\alpha)$ is 'small'), or (as below) if ψ is sufficiently 'wider' than a D_α so that $R(\psi \| \psi + D_\alpha)$ is 'small'. The first generalization could be applied to prove the IND-CPA security of LWE-based encryption schemes (such as Regev [30] and Dual-Regev [16]) schemes with low-precision Gaussian sampling, as used for signature schemes in Sect. 3.

Theorem 5.1. *Let $\alpha, \beta > 0$ be real numbers with $\beta = \Omega(m\alpha/\log n)$ for positive integers m and n. Let $m > \frac{n \log q}{\log(\alpha+\beta)^{-1}} \geq 1$ with $q \leq \mathrm{poly}(m, n)$ prime. Then there is a polynomial-time reduction from $\mathrm{LWE}_{n,q,D_\alpha,m}$ to $\mathrm{LWE}_{n,q,\phi,m}$, with $\phi = \frac{1}{q}\lfloor qU_\beta \rceil$.*

Proof. Our reduction relies on five steps:

- A reduction from $\mathrm{LWE}_{n,q,D_\alpha,m}$ to $\mathrm{LWE}_{n,q,\psi,m}$ with $\psi = D_\alpha + U_\beta$,
- A reduction from $\mathrm{LWE}_{n,q,\psi,m}$ to $\mathrm{sLWE}_{n,q,\psi,m}$,
- A reduction from $\mathrm{sLWE}_{n,q,\psi,m}$ to $\mathrm{sLWE}_{n,q,U_\beta,m}$,
- A reduction from $\mathrm{sLWE}_{n,q,U_\beta,m}$ to $\mathrm{sLWE}_{n,q,\phi,m}$, with $\phi = \frac{1}{q}\lfloor qU_\beta \rceil$,
- A reduction from $\mathrm{sLWE}_{n,q,\phi,m}$ to $\mathrm{LWE}_{n,q,\phi,m}$.

First Step The reduction is given m elements $(\boldsymbol{a}_i, b_i) \in \mathbb{Z}_q^n \times \mathbb{T}$, all drawn from $A_{\boldsymbol{s},D_\alpha}$ (for some $\boldsymbol{s}$), or all drawn from $U(\mathbb{Z}_q^n \times \mathbb{T})$. The reduction consists in adding independent samples from U_β to each b_i. The reduction maps the uniform distribution to itself, and $A_{\boldsymbol{s},D_\alpha}$ to $A_{\boldsymbol{s},\psi}$.

Second Step Reducing the distinguishing variant of LWE to its search variant is direct. In particular, suppose that there exists a solver, which finds the secret for $\mathrm{sLWE}_{n,q,\psi,m}$ with success probability ε. We use this solver to construct a distinguisher for $\mathrm{LWE}_{n,q,\psi,m}$. Let $(\mathbf{A}, \boldsymbol{b})$ be the input to the distinguisher, which comes from either the LWE distribution or the uniform distribution. Let $\boldsymbol{s}$ be the output of the solver on input $(\mathbf{A}, \boldsymbol{b})$. Given $\boldsymbol{s}$, $\mathbf{A}$, and $\boldsymbol{b}$, we compute $\|\boldsymbol{b} - \frac{1}{q}\mathbf{A}\boldsymbol{s}\|_\infty$. If this quantity is smaller than

$$t_0 = \beta + 2\pi\alpha\sqrt{\log\left(4\varepsilon^{-1}\right)}, \tag{7}$$

then the distinguisher outputs 1; otherwise, it outputs 0. We now analyze the advantage $\varepsilon_{\mathrm{adv}}$ of such a distinguisher. On the one hand, if the input to the distinguisher comes from the LWE distribution, the probability that the distinguisher outputs 1 is bounded from below by

$$\varepsilon - \Pr_{\boldsymbol{e} \hookleftarrow \psi}\left(\|\boldsymbol{e}\|_\infty \geq t_0\right).$$

On the other hand, when the input comes from the uniform distribution, the probability of having 1 as the output of the constructed distinguisher is bounded from above by

$$\Pr_{\boldsymbol{b} \hookleftarrow U(\mathbb{Z}_q^n)}\left[\exists \boldsymbol{s} \in \mathbb{Z}_q^n : \left\| \boldsymbol{b} - \frac{1}{q}\mathbf{A}\boldsymbol{s} \right\|_\infty \leq t_0\right].$$

Hence the overall distinguishing advantage satisfies

$$\varepsilon_{\text{adv}} \geq \left(\varepsilon - \Pr_{\boldsymbol{e} \hookleftarrow \psi}\left[\|\boldsymbol{e}\|_\infty \geq t_0\right]\right) - \Pr_{\boldsymbol{b} \hookleftarrow U(\mathbb{Z}_q^n)}\left[\exists \boldsymbol{s} \in \mathbb{Z}_q^n : \left\| \boldsymbol{b} - \frac{1}{q}\mathbf{A}\boldsymbol{s} \right\|_\infty \leq t_0\right]. \quad (8)$$

Since

$$\Pr_{\boldsymbol{e} \hookleftarrow \psi}\left[\|\boldsymbol{e}\|_\infty \geq t_0\right] \leq \Pr_{\boldsymbol{e} \hookleftarrow D_\alpha}\left[\|\boldsymbol{e}\|_\infty \geq t\right],$$

where t is defined to be $2\pi\alpha\sqrt{\log\left(4\varepsilon^{-1}\right)}$, the lower bound on ε_{adv} given in (8) can be re-written as

$$\varepsilon - \Pr_{\boldsymbol{e} \hookleftarrow D_\alpha}[\|\boldsymbol{e}\|_\infty \geq t] - \Pr_{\boldsymbol{b} \hookleftarrow U(\mathbb{Z}_q^n)}\left[\exists \boldsymbol{s} \in \mathbb{Z}_q^n | \left\| \boldsymbol{b} - \frac{1}{q}\mathbf{A}\boldsymbol{s} \right\|_\infty \leq t_0\right]. \quad (9)$$

If both the above probabilities are $\leq \varepsilon/4$, then ε_{adv} is at least $\varepsilon/2$. We now give an upper bound to each probability and enforce the parameters to satisfy these bounds. For the first probability, since $\boldsymbol{e} \hookleftarrow D_\alpha$, a standard Gaussian tail bound follows

$$\Pr_{\boldsymbol{e} \hookleftarrow D_\alpha}[\|\boldsymbol{e}\|_\infty \geq t] \leq \exp\left(-\left(\frac{2\pi\alpha}{t}\right)^2\right).$$

To ensure that the latter is less than $\varepsilon/4$, we need $t \geq 2\pi\alpha\sqrt{\log\left(4\varepsilon^{-1}\right)}$. Our t_0 defined in (7) satisfies the latter condition. For the second probability, by using a union bound argument, we have that

$$\Pr_{\boldsymbol{b} \hookleftarrow U(\mathbb{Z}_q^n)}[\exists \boldsymbol{s} \in \mathbb{Z}_q^n : \|\boldsymbol{b} - \mathbf{A}\boldsymbol{s}\|_\infty \leq t_0] \leq q^n \left(\frac{2\lfloor t_0 q \rfloor + 1}{q}\right)^m \leq \left(q^{n/m}(4t_0)\right)^m.$$

To ensure that the right hand side of the above inequality is less than $\varepsilon/4$, we require to satisfy two conditions. First, we impose that $q^{n/m}(4t_0) < 1/2$, which is equivalent to $(n \log q)/m \leq \log\left(\frac{1}{8t_0}\right)$. Now that $q^{n/m}(4t_0) < 1/2$, we impose $(1/2)^m \leq \varepsilon/4$ to enforce the second constraint, that is $m \geq \log\left(4\varepsilon^{-1}\right)$. Combining the above two conditions, it suffices to have

$$m \geq \max\left(\frac{n \log q}{\log\left(\frac{1}{8t_0}\right)}, \log\left(4\varepsilon^{-1}\right)\right).$$

By replacing t_0 from (7) and inserting $\varepsilon^{-1} = O(\text{poly}(n))$, we get

$$m \geq \frac{n \log q}{\log{(\alpha+\beta)}^{-1}},$$

if $(\alpha+\beta)^{-1} = 2^{o\left(\frac{n \log q}{\log n}\right)}$.

Third Step The reduction from $\text{sLWE}_{n,q,\psi,m}$ to $\text{sLWE}_{n,q,U_\beta,m}$ is vacuous: by using the RD (and in particular the probability preservation property of Lemma 2.9), we show that an oracle solving $\text{sLWE}_{n,q,U_\beta,m}$ also solves $\text{sLWE}_{n,q,\psi,m}$.

Lemma 5.2. *Let α, β be real numbers with $\alpha \in (0, 1/e)$ and $\beta \geq \alpha$. Let $\psi = D_\alpha + U_\beta$. Then*

$$R_2(U_\beta \| \psi) = 1 + \frac{1}{1 - e^{-\pi\beta^2/\alpha^2}} \frac{\alpha}{\beta} < \ 1 + 1.05 \cdot \frac{\alpha}{\beta}.$$

Proof. The density function of ψ is the convolution of the density functions of D_α and U_β:

$$f_\psi(x) = \frac{1}{2\alpha\beta} \int_{-\beta}^{\beta} e^{\frac{-\pi(x-y)^2}{\alpha^2}} \mathrm{d}y.$$

Using Rényi of order 2, we have:

$$R_2(U_\beta \| \psi) = \int_{-\beta}^{\beta} \frac{\frac{1}{(2\beta)^2}}{\frac{1}{2\alpha\beta} \int_{-\beta}^{\beta} e^{\frac{-\pi(x-y)^2}{\alpha^2}} \mathrm{d}y} \mathrm{d}x = \frac{\alpha}{\beta} \int_0^{\beta} \frac{1}{\int_{-\beta}^{\beta} e^{\frac{-\pi(x-y)^2}{\alpha^2}} \mathrm{d}y} \mathrm{d}x.$$

The denominator in the integrand is a function for $x \in [0, \beta]$.

$$\phi(x) = \alpha - \int_{\beta+x}^{\infty} \exp\left(\frac{-\pi y^2}{\alpha^2}\right) \mathrm{d}y - \int_{\beta-x}^{\infty} \exp\left(\frac{-\pi y^2}{\alpha^2}\right) \mathrm{d}y.$$

For standard Gaussian, we use the following tail bound [9]:

$$\frac{1}{\sqrt{2\pi}} \int_z^{\infty} \mathrm{e}^{-x^2/2} \mathrm{d}x \leq \frac{1}{2} \mathrm{e}^{-z^2/2}.$$

Then we have

$$\phi(x) \geq \alpha \left(1 - \frac{1}{2}\exp\left(\frac{-\pi(\beta+x)^2}{\alpha^2}\right) - \frac{1}{2}\exp\left(\frac{-\pi(\beta-x)^2}{\alpha^2}\right)\right).$$

Taking the reciprocal of above, we use the first-order Taylor expansion. Note here

$$t(x) = \frac{1}{2}\exp\left(\frac{-\pi(\beta+x)^2}{\alpha^2}\right) + \frac{1}{2}\exp\left(\frac{-\pi(\beta-x)^2}{\alpha^2}\right). \tag{10}$$

We want to bound the function $t(x)$ by a constant $c \in (0, 1)$. Here $t(x)$ is not monotonic. We take the maximum of the first term and the maximum of the second term of $t(x)$ in (10). Let $\sigma_{\alpha,\beta}$ denote $\frac{1}{2}e^{-\pi\beta^2/\alpha^2}$, then an upper bound ($\beta \geq \alpha$) is:

$$t(x) \leq \frac{1}{2}e^{-\pi\beta^2/\alpha^2} + \frac{1}{2} = \sigma_{\alpha,\beta} + \frac{1}{2} < 1.$$

We then use the fact that $\frac{1}{1-t(x)} = 1 + \frac{1}{1-t(x)}t(x) \leq 1 + \frac{1}{1-2\sigma_{\alpha,\beta}}t(x)$ to bound the Rényi divergence of order 2.

$$\begin{aligned}
R_2(U_\beta\|\psi) &= \frac{\alpha}{\beta}\int_0^\beta \frac{1}{\phi(x)}\mathrm{d}x \\
&\leq \frac{1}{\beta}\int_0^\beta \frac{1}{1-\frac{1}{2}\exp\left(\frac{-\pi(\beta+x)^2}{\alpha^2}\right) - \frac{1}{2}\exp\left(\frac{-\pi(\beta-x)^2}{\alpha^2}\right)}\mathrm{d}x \\
&\leq \frac{1}{\beta}\int_0^\beta \left(1 + \frac{1}{1-2\sigma_{\alpha,\beta}}\exp\left(\frac{-\pi(\beta+x)^2}{\alpha^2}\right)\right. \\
&\quad \left. + \frac{1}{1-2\sigma_{\alpha,\beta}}\exp\left(\frac{-\pi(\beta-x)^2}{\alpha^2}\right)\right)\mathrm{d}x \\
&= 1 + \frac{1}{(1-2\sigma_{\alpha,\beta})\beta}\int_0^{2\beta}\exp\left(\frac{-\pi x^2}{\alpha^2}\right)\mathrm{d}x \\
&= 1 + \frac{1}{2(1-2\sigma_{\alpha,\beta})\beta}\int_{-2\beta}^{2\beta}\exp\left(\frac{-\pi x^2}{\alpha^2}\right)\mathrm{d}x \\
&= 1 + \frac{\alpha}{(1-2\sigma_{\alpha,\beta})\beta}(1-2D_\alpha(2\beta)) \leq 1 + \frac{1}{1-2\sigma_{\alpha,\beta}}\frac{\alpha}{\beta}.
\end{aligned}$$

Hence we have the bound

$$R_2(U_\beta\|\psi) \leq 1 + \frac{1}{1-e^{-\pi\beta^2/\alpha^2}}\frac{\alpha}{\beta}.$$

The second bound in the lemma statement follows from the fact that

$$\frac{1}{1-\mathrm{e}^{-\pi\beta^2/\alpha^2}} < 1.05,$$

for $\beta \geq \alpha$. □

The RD multiplicativity property (see Lemma 2.9) implies that for m independent samples, we have $R_2(U_\beta^m \| \psi^m) \leq R_2(U_\beta \| \psi)^m$. To ensure that the latter mth power is polynomial in n, we use Lemma 5.2 with $\beta = \Omega(m\alpha / \log n)$; with this choice, we have $R_2(U_\beta \| \psi) = 1 + O(\frac{\alpha}{\beta}) \leq \exp(O(\frac{\alpha}{\beta}))$ and $R_2(U_\beta \| \psi)^m = n^{O(1)}$. The RD probability preservation and data processing properties (see Lemma 2.9) now imply that if an oracle solves $\mathrm{sLWE}_{n,q,U_\beta,m}$ with probability ε, then it also solves $\mathrm{sLWE}_{n,q,\psi,m}$ with probability $\varepsilon' \geq \varepsilon^2 / R_2(U_\beta \| \phi)^m \geq \varepsilon^2 / n^{O(1)}$.
Fourth step. We reduce $\mathrm{sLWE}_{n,q,U_\beta,m}$ with continuous noise U_β to $\mathrm{sLWE}_{n,q,\phi,m}$ with discrete noise $\phi = \frac{1}{q}\lfloor qU_\beta \rceil$ with support contained in $\mathbb{T}_q$, by rounding to the nearest multiple of $\frac{1}{q}$ any provided b_i (for $i \leq m$).
Fifth step. We reduce $\mathrm{sLWE}_{n,q,\phi,m}$ to $\mathrm{LWE}_{n,q,\phi,m}$ by invoking Theorem 2.7. □

6. Application to Learning with Rounding (LWR)

In this section, we first review (in Theorem 6.1) and combine with other results (in Theorem 6.2) the recent hardness result of Bogdanov et al. [4] for the Learning With Rounding (LWR) problem introduced in Banerjee et al. [7], based on the hardness of the standard LWE problem. This result of Bogdanov et al. [4] makes use of RD (inspired by an earlier version of our work) within a proof that can be seen as a variant of the Micciancio-Mol search to decision reduction for LWE [23]. Then, we show (in Theorem 6.4) a new dimension-preserving hardness result for LWR, obtained by composing our RD-based hardness result for LWE with uniform noise from the previous section with another reduction from Bogdanov et al. [4] (which we rephrase in Theorem 6.3) that reduces LWE with uniform noise to LWR. Interestingly, our new reduction for LWR also makes use of the Micciancio-Mol reduction [23], but unlike the LWR reduction in Theorem 6.1, ours uses [23] as a black box within the reduction of Theorem 5.1.

6.1. *Adapted Results from [4]*

We first recall the main hardness result on LWR from Bogdanov et al. [4].

Theorem 6.1. *([4, Theorem 3]) For every $\varepsilon > 0$, n, m, $q > 2pB$, and algorithm* Dist *such that*

$$\left|\Pr_{\mathbf{A},\boldsymbol{s}}\left[\mathrm{Dist}\left(\mathbf{A}, \lfloor \mathbf{A}\boldsymbol{s} \rceil_p\right) = 1\right] - \Pr_{\boldsymbol{u}}\left[\mathrm{Dist}\left(\mathbf{A}, \lfloor \boldsymbol{u} \rceil_p\right) = 1\right]\right| \geq \varepsilon$$

where $\mathbf{A} \hookleftarrow U\left(\mathbb{Z}_q^{m\times n}\right)$, $\boldsymbol{s} \hookleftarrow U\left(\{0,1\}^n\right)$ and $\boldsymbol{u} \hookleftarrow U\left(\mathbb{Z}_q^m\right)$ there exists an algorithm Learn *that runs in time polynomial in n, m, the number of divisors of q, and the running time of* Dist *such that*

$$\Pr_{\mathbf{A},\boldsymbol{s}}\left[\mathrm{Learn}\left(\mathbf{A}, \mathbf{A}\boldsymbol{s} + \boldsymbol{e}\right) = \boldsymbol{s}\right] \geq \left(\frac{\varepsilon}{4qm} - \frac{2^n}{p^m}\right)^2 \cdot \frac{1}{\left(1 + \frac{2Bp}{q}\right)^m}, \tag{11}$$

for any noise distribution $\boldsymbol{e}$ *that is* B*-bounded and* B*-balanced in each coordinate.*

We now combine Theorem 6.1 with other results to state it as a reduction from the standard LWE problem, so that it would be comparable with our alternative reduction.

Theorem 6.2. *Let* $qm = O(\text{poly}(n))$*, and* $n \leq m \leq O\left(\frac{\sqrt{\log n}}{p\alpha}\right)$*. Then there is a polynomial-time reduction from* $\mathrm{LWE}_{n/\log q,q,D_\alpha,m}$ *to* $\mathrm{LWR}_{n,q,p,m}$*.*

Proof. The reduction can be obtained in the following five steps:

- A reduction from $\mathrm{LWE}_{n/\log q,q,D_\alpha,m}$ to $\mathrm{binLWE}_{n,q,D_\alpha,m}$,
- A trivial reduction from $\mathrm{binLWE}_{n,q,D_\alpha,m}$ to $\mathrm{sbinLWE}_{n,q,D_\alpha,m}$,
- A reduction from $\mathrm{sbinLWE}_{n,q,D_\alpha,m}$ to $\mathrm{sbinLWE}_{n,q,D'_{\alpha,B'},m}$, with $D'_{\alpha,B'}$ the distribution D_α truncated (by rejection) to the interval $[-B', B']$,
- A reduction from $\mathrm{sbinLWE}_{n,q,D'_{\alpha,B'},m}$ to $\mathrm{sbinLWE}_{n,q,\phi,m}$, with $\phi = \frac{1}{q}\lfloor qD'_{\alpha,B'}\rceil$,
- A reduction from $\mathrm{sbinLWE}_{n,q,\phi,m}$ to $\mathrm{LWR}_{n,q,p,m}$ via Theorem 6.1.

The first reduction is taken from Brakerski et al. [6]. The second one is just the trivial decision to search reduction for binary secret LWE. Note that we provided such a reduction (see the second step of proof of Theorem 5.1) for a more general setting. In fact, there we had binary secret LWE with $\psi = D_\alpha + U_\beta$ as the error distribution while we have non-binary secret LWE and Gaussian noise D_α here. If we simplify the constraints appeared there, we simply get $m \geq n/\log\left(\alpha^{-1}\right)$, which can be further relaxed to $m \geq n$. The third reduction is vacuous and consists in applying the R_∞ probability preservation property from Lemma 2.9 and the m-sample Gaussian tail-cut Lemma 2.11 that ensures that this reduction preserves success probability up to a constant factor by setting $B' = \alpha q\sqrt{\ln(2m)/\pi}$. The fourth reduction consists of applying $\frac{1}{q}\lfloor q(\cdot)\rceil$ to all samples. With this, we have only changed the noise distribution from Gaussian D_α with standard deviation αq to its quantized version ϕ. This only adds a rounding error of magnitude $\leq 1/2$. The last step is exactly Theorem 6.2 mentioned above. The last step reduction holds if (i) the distribution ϕ be B-bounded, and, to ensure the reduction is probabilistic polynomial-time, we need that (ii) the right hand side of (11) is at least $\varepsilon^{O(1)}/n^{O(1)}$. The obtained distribution ϕ in (i) is both B-bounded and B-balanced with $B = \alpha q\sqrt{\ln(2m)/\pi} + 1/2$. For the second condition (ii), we note that there are two terms in the right hand side of (11). We first claim that

$$\frac{\varepsilon}{8qm} > \frac{2^n}{p^m},$$

for $q = \text{poly}(n)$ and $\varepsilon^{-1} = 2^{o(n)}$. To prove this claim, first note that

$$\begin{aligned} \frac{\varepsilon}{8qm} > \frac{2^n}{p^m} &\Leftrightarrow n - m\log p < \log\left(\frac{\varepsilon}{8qm}\right) \\ &\Leftrightarrow m > \frac{n + \log\left(8qm\varepsilon^{-1}\right)}{\log p}. \end{aligned}$$

Now, $qm = \mathrm{poly}(n)$ and $\varepsilon^{-1} = n^{O(1)}$ and the assumption $m \geq n \geq 2n/\log(p)$ imply the above condition for sufficiently large n. Hence, for the first term we get

$$\left(\frac{\varepsilon}{4qm} - \frac{2^n}{p^m}\right)^2 > \left(\frac{\varepsilon}{8qm}\right)^2,$$

which is $\geq \varepsilon^{O(1)}/n^{O(1)}$ using $qm = O\left(\mathrm{poly}(n)\right)$. For the second term, we get

$$\left(1 + \frac{2Bp}{q}\right)^m \leq \exp\left(\frac{2Bpm}{q}\right), \tag{12}$$

since for positive x and y, we have $(1+x)^y \leq \exp(xy)$. The right hand side of (12) is less than $n^{O(1)}$ if $2Bpm/q \leq O\left(\log n\right)$. Replacing B by the value derived from condition in (i), and using that $m \geq n$, we get that some $m = O\left(\sqrt{\log n}/(p\alpha)\right)$ suffices. □

Below, we will give a tighter reduction than above from LWE to LWR. We will make use of the theorem below.

Theorem 6.3. *(Adapted from [4, Theorem 13]) Let p and q be two integers such that p divides q and let $\beta = q/(2p)$. If we have a T-time distinguisher for* $\mathrm{LWR}_{n,q,p,m}$ *with advantage ε, then we can construct a $T' = O\left(T + m'n \cdot \mathrm{poly}(\log q)\right)$ time distinguisher for* $\mathrm{LWE}_{n,q,U_\beta,m'}$ *with $m' = m \cdot q/p$ and advantage $\varepsilon' \geq \varepsilon/2$.*

Proof. The proof follows the steps of the proof of Theorem 13 in Bogdanov et al. [4]. Suppose that we have access to a T-time distinguisher which runs over m samples $(\boldsymbol{a}, b) = (\boldsymbol{a}, \langle \boldsymbol{a}, \boldsymbol{s}\rangle + e)$ for $\boldsymbol{a} \hookleftarrow U(\mathbb{Z}_q^n)$, and

$$e \hookleftarrow \left[-\frac{q}{2p}, \ldots, \frac{q}{2p}\right) \subseteq \mathbb{Z}_q.$$

The authors of [4] run the LWE oracle until they hit a ‘good’ sample $(\boldsymbol{a}, b)$ with $b \in (q/p)\mathbb{Z}_p$ and output the LWR sample $(\boldsymbol{a}, (p/q)b) \in \mathbb{Z}_q^n \times \mathbb{Z}_p$. Since the LWE error e is distributed uniformly in U_β, each sample output by the LWE oracle is ‘good’ with probability p/q, and the *expected* number of LWE samples needed by this reduction to produce m LWR samples is therefore $m' = m \cdot q/p$. Instead, here we modify the reduction to work with a fixed number $m' = m \cdot q/p$ of LWE samples. Namely, if the $m' = m \cdot q/p$ given LWE samples contain at least m ‘good’ samples (which we call event Good), the modified reduction uses them to compute m LWR samples and runs the LWR distinguisher on them, outputting whatever it outputs, as in Bogdanov et al. [4]. Else, if the m' given LWE samples contain $< m$ ‘good’ samples, the LWE distinguisher outputs 0. The proof of Theorem 13 in [4] shows that conditioned on event Good, the input samples to the LWR distinguisher come from the LWR distribution (resp. uniform distribution) if the LWE oracle generates samples from the LWE distribution (resp. uniform distribution). It follows that the advantage of our LWE distinguisher is $\geq \Pr[\mathsf{Good}] \cdot \varepsilon \geq \varepsilon/2$, where we have used the fact that $\Pr[\mathsf{Good}] \geq 1/2$, since the number of ‘good’ samples is binomially distributed with parameters $(m', p/q)$ and has median $m' \cdot p/q = m$. □

6.2. *New Results*

One can compose the reduction in Theorem 6.3 with ours from LWE with Gaussian noise to LWE with uniform noise (Theorem 5.1) to get a new reduction from LWE to LWR. Hence, this combination can be summarized as:

Theorem 6.4. *Let p divide q, $m' = m \cdot q/p$ with $m = O\left(\log n/\alpha\right)$ for $m' \geq m \geq n \geq 1$. There is a polynomial-time reduction from* $\mathrm{LWE}_{n,q,D_\alpha,m'}$ *to* $\mathrm{LWR}_{n,q,p,m}$.

Proof. Let $\beta = q/(2p)$. The reduction has two steps:

- A reduction from $\mathrm{LWE}_{n,q,D_\alpha,m'}$ to $\mathrm{LWE}_{n,q,U_\beta,m'}$,
- A reduction from $\mathrm{LWE}_{n,q,U_\beta,m'}$ to $\mathrm{LWR}_{n,q,p,m}$.

On the one hand, $\mathrm{LWE}_{n,q,U_\beta,m'}$ is at least as hard as $\mathrm{LWE}_{n,q,D_\alpha,m'}$ where $\beta = \Omega\left(m'\alpha/\log n\right)$ (see Theorem 5.1). On the other hand, the second phase of the reduction follows from Theorem 6.3; namely we have a reduction from $\mathrm{LWE}_{n,q,U_\beta,m'}$ to $\mathrm{LWR}_{n,q,p,m}$ subject to the condition that p divides q and $m' = m \cdot q/p$. Combining these two reductions completes the proof. Note that, by putting all the conditions together, it turns out that

$$\beta = \Omega\left(\frac{m'\alpha}{\log n}\right) \Leftrightarrow \frac{q}{2p} \geq \frac{\frac{mq}{p}\alpha}{\log n} \Leftrightarrow m = O\left(\frac{\log n}{\alpha}\right),$$

where the first equivalence is derived by replacing β and m', by $q/(2p)$ and mq/p. □

Table 3 compares the parameters of Theorems 6.2 and 6.4, and a reduction from [3]. The reduction in Theorem 6.2 loses a $\log q$ factor in dimension, while our uniform noise reduction preserves the dimension, which is the first of its kind without resorting to the noise-flooding technique (as [7]). On the downside, our reduction does not preserve the number of samples.

Note that setting $\gamma = 1$ gives n' equal to that of Theorem 6.2, while it loses an extra factor n in the denominator of m. On the other hand, setting $\gamma = q$ allows for approximately $n = n'$, however for an expense of much smaller m. The reduction in

Table 3. Comparing the main parameters of different reductions from $\mathrm{LWE}_{n',q,D_\alpha,m'}$ to $\mathrm{LWR}_{n,q,p,m}$ for a fixed n and another flexible parameter $\gamma \geq 1$.

Param.	[3, Theorem 4.1]	Theorem 6.2 ([4])	Theorem 6.4
n'	$O\left(\frac{n\log(2\gamma)}{\log q}\right)$	$\frac{n}{\log q}$	n
m	$O\left(\frac{\sqrt{\log n}}{\gamma n p \alpha}\right)$	$O\left(\frac{\sqrt{\log n}}{p\alpha}\right)$	$O\left(\frac{\log n}{\alpha}\right)$
m'	m	m	$m \cdot \frac{q}{p}$

Theorem 6.2 also restricts the number of LWR samples m by a further $O\left(p\sqrt{\log n}\right)$ factor in comparison with our results. This factor is equal to $O\left(\gamma pn\sqrt{\log n}\right)$ if we compare our result with that of Theorem 4.1 from [3].

7. Open Problems

Our results show the utility of the Rényi divergence in several areas of lattice-based cryptography. A natural question is to find further new applications of RD to improve the efficiency of cryptosystems. Our results suggest some natural open problems, whose resolution could open up further applications. In particular, can we extend the applicability of RD to more general distinguishing problems than those satisfying our 'public sampleability' requirement? This may extend our results further. For instance, can we use RD-based arguments to prove the hardness of LWE with uniform noise without using the search to decision reduction of Micciancio and Mol [23]? This may allow the proof to apply also to Ring-LWE with uniform noise and Ring-LWR. Another open problem will be discussed in the next section.

7.1. *GPV Signature Scheme*

The RD can also be used to reduce the parameters obtained via the SD-based analysis of the GPV signature scheme in Gentry et al. [16].

In summary, the signature and the security proof from Gentry et al. [16] work as follows. The signature public key is a matrix $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ with n linear in the security parameter λ, $q = \text{poly}(n)$, and $m = O(n \log q)$. The private signing key is a short basis matrix $\mathbf{T}$ for the lattice $\Lambda_{\mathbf{A}}^{\perp} = \{\boldsymbol{x} \in \mathbb{Z}^m : \mathbf{A}\cdot\boldsymbol{x} = \mathbf{0} \bmod q\}$, whose last successive minimum satisfies $\lambda_m\left(\Lambda_{\mathbf{A}}^{\perp}\right) \leq O(1)$ when $m = \Omega(n \log q)$ (see [16]). A signature $(\boldsymbol{\sigma}, s)$ on a message M is a short vector $\boldsymbol{\sigma} \in \mathbb{Z}^m$ and a random salt $s \in \{0,1\}^{\lambda}$, such that $\mathbf{A}\cdot\boldsymbol{\sigma} = H(M,s) \bmod q$, where H is a random oracle hashing into $\mathbb{Z}_q^n$. The short vector $\boldsymbol{\sigma}$ is sampled by computing an arbitrary vector $\boldsymbol{t}$ satisfying $\mathbf{A}\cdot\boldsymbol{t} = H(M,s) \bmod q$ and using $\mathbf{T}$ along with a Gaussian sampling algorithm (see [6,16]) to produce a sample from $\boldsymbol{t} + D_{\Lambda_{\mathbf{A}}^{\perp},r,-\boldsymbol{t}}$.

The main idea in the security proof from the SIS problem [16] is based on simulating signatures without $\mathbf{T}$, by sampling $\boldsymbol{\sigma}$ from $D_{\mathbb{Z}^m,r}$ and then programming the random oracle H at (M,s) according to $H(M,s) = \mathbf{A}\cdot\boldsymbol{\sigma} \bmod q$. As shown in Gentry et al. [16, Lemma 5.2], the conditional distribution of $\boldsymbol{\sigma}$ given $\mathbf{A}\cdot\boldsymbol{\sigma} \bmod q$ is exactly the same in the simulation and in the real scheme. Therefore, the SD between the simulated signatures and the real signatures is bounded by the SD between the marginal distribution D_1 of $\mathbf{A}\cdot\boldsymbol{\sigma} \bmod q$ for $\boldsymbol{\sigma} \hookleftarrow D_{\mathbb{Z}^m,r}$ and $U(\mathbb{Z}_q^m)$. This SD for one signature is bounded by ε if $r \geq \eta_{\varepsilon}\left(\Lambda_{\mathbf{A}}^{\perp}\right)$. This leads, over the q_s sign queries of the attacker, in the SD-based analysis of Gentry et al. [16], to take $\varepsilon = O(2^{-\lambda}q_s^{-1})$ and thus $r = \Omega(\sqrt{\lambda + \log q_s})$ (using Lemma 2.2), in order to handle attackers with success probability $2^{-o(\lambda)}$.

Now, by Lemma 2.10, we have that the RD $R_{\infty}(D_1\|U)$ is bounded by $1 + c\cdot\varepsilon$ for one signature, for some constant c. By the multiplicativity property of Lemma 2.9, over q_s queries, it is bounded by $(1+c\varepsilon)^{q_s}$. By taking $\varepsilon = O(q_s^{-1})$, we obtain overall an RD bounded as $O(1)$ between the view of the attacker in the real attack and simulation, leading to a security proof with respect to SIS but with a smaller $r = \Omega(\sqrt{\log(nq_s)}) = \Omega(\sqrt{\log \lambda + \log q_s})$. When the number of sign queries q_s allowed to the adversary is much smaller than 2^λ, this leads to significant parameter savings, because SIS's parameter β is reduced and hence n, m, q may be set smaller for the same security parameter λ.

The above analysis indeed reduces the smoothing condition in the security proof from $r = \Omega(\sqrt{\lambda})$ to $r = \Omega(\sqrt{\log \lambda})$. But to make Gaussian sampling on $\Lambda_{\mathbf{A}}^{\perp}$ efficient in signature generation, we also need r lower bounded by the Euclidean norm of the trapdoor basis for $\Lambda_{\mathbf{A}}^{\perp}$. The latter is lower bounded by $\lambda_1\left(\Lambda_{\mathbf{A}}^{\perp}\right)$, which is $\Omega(\sqrt{m}) \geq \Omega(\sqrt{\lambda})$ with high probability. That is actually similar to (or even larger than) the old SD-based smoothing condition. Overall, we relaxed the smoothing condition while the sampling condition remained unchanged. Hence, relaxing both conditions together is left as an open problem.

Acknowledgements

We thank Léo Ducas, Vadim Lyubashevsky and Fabrice Mouhartem for useful discussions. We thank Katsuyuki Takashima and Atsushi Takayasu for informing us about an error in the conference version of this work, and another one in the computations of the R_a-based analysis of Sect. 3. This work has been supported in part by ERC Starting Grant ERC-2013-StG-335086-LATTAC, an Australian Research Fellowship (ARF) from the Australian Research Council (ARC), ARC Discovery Grants DP0987734, DP110100628 and DP150100285, and the European Unions H2020 Programme under Grant Agreement Number ICT-644209.

References

[1] E. Alkim, L. Ducas, T. Pöppelmann, P. Schwabe, Post-quantum key exchange—a new hope, in *25th USENIX Security Symposium (USENIX Security 16)* (USENIX Association, Austin, 2016), pp. 327–343

[2] M. Ajtai, Generating hard instances of lattice problems (extended abstract), in *Proceedings of STOC* (ACM, 1996), pp. 99–108

[3] J. Alwen, S. Krenn, K. Pietrzak, D. Wichs, Learning with rounding, revisited—new reduction, properties and applications, in *Proceedings of CRYPTO*. LNCS, vol. 8042 (Springer, 2013), pp. 57–74

[4] A. Bogdanov, S. Guo, D. Masny, S. Richelson, A. Rosen, On the hardness of learning with rounding over small modulus, in *Proceedings of TCC A*. LNCS, vol. 9562 (Springer, 2016), pp. 209–224

[5] S. Bai, A. Langlois, T. Lepoint, D. Stehlé, R. Steinfeld, Improved security proofs in lattice-based cryptography: using the Rényi divergence rather than the statistical distance, in *Proceedings of ASIACRYPT, Part I*. LNCS, vol. 9452 (Springer, 2015), pp. 3–24

[6] Z. Brakerski, A. Langlois, C. Peikert, O. Regev, D. Stehlé, Classical hardness of learning with errors, in *Proceedings of STOC* (ACM, 2013), pp. 575–584

[7] A. Banerjee, C. Peikert, A. Rosen, Pseudorandom functions and lattices, in *Proceedings of EUROCRYPT*. LNCS, vol. 7237 (Springer, 2012), pp. 719–737

[8] Z. Brakerski, V. Vaikuntanathan, Efficient fully homomorphic encryption from (standard) LWE, in *Proceedings of FOCS* (IEEE Computer Society Press, 2011), pp. 97–106

[9] M. Chiani, D. Dardari, M.K. Simon, New exponential bounds and approximations for the computation of error probability in fading channels. *IEEE Trans. Wireless. Commun.***2**(4):840–845 (2003)

[10] C.-W. Chow, *On Algorithmic Aspects of the Learning with Errors Problem and Its Variants*, Masters thesis, The Chinese University of Hong Kong (2003)

[11] L. Ducas, A. Durmus, T. Lepoint, V. Lyubashevsky, Lattice signatures and bimodal Gaussians, in *Proceedings of CRYPTO*. LNCS, vol. 8042 (Springer, 2013), pp. 40–56

[12] N. Döttling, J. Müller-Quade, Lossy codes and a new variant of the learning-with-errors problem, in *Proceedings of EUROCRYPT*. LNCS, (Springer, 2013), pp. 18–34

[13] L. Ducas, *Accelerating Bliss: The Geometry of Ternary Polynomials*. Cryptology ePrint Archive, Report 2014/874 (2014). http://eprint.iacr.org/

[14] S. Garg, C. Gentry, S. Halevi, Candidate multilinear maps from ideal lattices, in *Proceedings of EUROCRYPT*. LNCS, vol. 7881 (Springer, 2013), pp. 1–17

[15] S. Goldwasser, Y.T. Kalai, C. Peikert, V. Vaikuntanathan, Robustness of the learning with errors assumption, in *Proceedings of ICS* (Tsinghua University Press, 2010), pp. 230–240

[16] C. Gentry, C. Peikert, V. Vaikuntanathan, Trapdoors for hard lattices and new cryptographic constructions, in *Proceedings of STOC* (ACM, 2008), pp. 197–206

[17] W. Hoeffding, Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **58**(301):13–30 (1963)

[18] B. Libert, S. Ling, F. Mouhartem, K. Nguyen, H. Wang, *Signature Schemes with Efficient Protocols and Dynamic Group Signatures from Lattice Assumptions*. Cryptology ePrint Archive, Report 2016/101 (2016). http://eprint.iacr.org/

[19] V. Lyubashevsky, C. Peikert, O. Regev, On ideal lattices and learning with errors over rings. *J. ACM* **60**(6):43 (2013)

[20] S. Ling, D.H. Phan, D. Stehlé, R. Steinfeld, Hardness of k-LWE and applications in traitor tracing, in *Proceedings of CRYPTO, Part I*. LNCS, vol. 8616 (Springer, 2014), pp. 315–334. Full version available at http://eprint.iacr.org/2014/494

[21] A. Langlois, D. Stehlé, R. Steinfeld, GGHLite: more efficient multilinear maps from ideal lattices, in *Proceedings of EUROCRYPT*. LNCS (Springer, 2014), pp. 239–256. Full version available at http://eprint.iacr.org/2014/487

[22] V. Lyubashevsky, Lattice signatures without trapdoors, in *Proceedings of EUROCRYPT*. LNCS, vol. 7237, ed. By D. Pointcheval, T. Johansson (Springer, 2012), pp. 738–755

[23] D. Micciancio, P. Mol. Pseudorandom knapsacks and the sample complexity of LWE search-to-decision reductions, in *Proceeding of CRYPTO*. LNCS, vol. 6841 (Springer, 2011), pp. 465–484

[24] D. Micciancio, C. Peikert. Hardness of SIS and LWE with small parameters, in *Proceeding of CRYPTO*. LNCS, vol. 8042 (Springer, 2013) pp. 21–39

[25] D. Micciancio, O. Regev, Worst-case to average-case reductions based on Gaussian measures. *SIAM J. Comput.* **37**(1):267–302 (2007)

[26] D. Micciancio, O. Regev, Lattice-based cryptography, in *Post-Quantum Cryptography*, ed By D.J. Bernstein, J. Buchmann, E. Dahmen (Springer, 2009), pp. 147–191

[27] T. Pöppelmann, L. Ducas, T. Güneysu, Enhanced lattice-based signatures on reconfigurable hardware, in *Proceeding of CHES* (2014), pp. 353–370

[28] C. Peikert, Public-key cryptosystems from the worst-case shortest vector problem, in *Proceeding of STOC* (ACM, 2009), pp. 333–342

[29] C. Peikert, *A Decade of Lattice Cryptography*. Cryptology ePrint Archive, Report 2015/939 (2015). http://eprint.iacr.org/

[30] O. Regev, On lattices, learning with errors, random linear codes, and cryptography, in *Proceeding of STOC* (2005), pp. 84–93

[31] O. Regev, *Lecture Notes of Lattices in Computer Science*, Computer Science Tel Aviv University (2009). Available at http://www.cims.nyu.edu/~regev

[32] O. Regev, On lattices, learning with errors, random linear codes, and cryptography. *J. ACM*, **56**(6) (2009)

[33] A. Rényi, On measures of entropy and information, in *Proceeding of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1 (1961), pp. 547–561

[34] K. Takashima, A. Takayasu, Tighter security for efficient lattice cryptography via the rényi divergence of optimized orders, in *Proceeding of ProvSec*. LNCS (Springer, 2015), pp. 412–431

[35] T. van Erven, P. Harremoes, Rényi divergence and Kullback–Leibler divergence. *IEEE Trans. Inf. Theory* **60**(7):3797–3820 (2014)