Smoothening Functions and the Homomorphism Learning Problem

by

Luis Antonio Ruiz-Lopez

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Combinatorics and Optimization

Waterloo, Ontario, Canada, 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Jintai Ding Professor Department of Mathematical Sciences University of Cincinnati Cincinnati OH 45221-0025 USA
Supervisor	David Jao Professor
Internal Members	Douglas Stebila Professor
	Michele Mosca Professor
Internal-External Member	Sergey Gorbunov Professor Cheriton School of Computer Science

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Luis Ruiz Lopez is the sole author of Chapters 1 and 2, which were written to write narrative and context to the material presented in the subsequent chapters.

The content of Chapters 3 and 4 is based on research performed by Luis Ruiz Lopez and Christopher Leonardi for the University of Waterloo and Isara Corp. The project was led and the draft was written by Luis Ruiz Lopez. Christopher Leonardi contributed with intellectual input for the material in Sections 4.3 and 4.4. As a result, a research paper was published in CFAIL 2019. [LR19].

The content of Chapters 5 and 6 is partially based on research performed by Luis Ruiz Lopez, Filip Pawlega and Elena Bakos-Lang for Isara Corp. and the University of Waterloo. The version that appears in this work is has been partially modified and adapted by Luis Ruiz Lopez to provide a more coherent narrative with the rest of the thesis. The draft was written by Luis Ruiz Lopez, except for parts of Sections 5.5, 5.6, 5.7 and 6.3. All authors contributed intellectually to the obtained results.

The content of Chapter 7 is partially based on the research performed by Luis Ruiz Lopez, Filip Pawlega and Elena Bakos-Lang for Isara Corp. and the University of Waterloo. The version that appears in this work is has been partially modified and adapted by Luis Ruiz Lopez to provide a more coherent narrative with the rest of the thesis. The project was lead and the draft was written by Luis Ruiz Lopez. All authors contributed intellectually to the obtained results.

Abstract

This thesis is an exploration of certain algebraic and geometrical aspects of the Learning With Errors (LWE) problem introduced in [Reg05]. On the algebraic front, we view it as a Learning Homomorphisms with Noise problem, and provide a generic construction of a public-key cryptosystem based on this generalization. On the geometric front, we explore the importance of the Gaussian distribution for the existing relationships between LWE and lattice problems. We prove that their smoothing properties does not make them special, but rather, the fact that it is infinitely divisible and ℓ_2 symmetric are important properties that make the Gaussian unique.

Acknowledgements

I want to start by thanking my supervisor David Jao, who became my friend and collaborator along the way, has always been supportive and has shown trust and patience at every step of this process. Thanks David!

To the member of the evaluation committee: Prof. Jintai Ding, Prof. Sergey Gorbunov, Prof. Michele Mosca and Prof. Douglas Stebila for their feedback and valuable input during the defense.

To my coauthors: David Jao, Chris Leonardi, Jason LeGrow, Elena Bakos Lang and Filip Pawlega. The dream is to do research with one's friends, that dream became reality while doing my PhD. Without you this would have been impossible.

To all the friends I made in Waterloo, specially to Alan, Nanis, Julian, Diana, Naye, Pavel, Randy, Filip, Meagan, Kelvin, Alan W, Ben, Remy, Chris, Jason, Archer, Melissa C., Megan M. and David Jackson.

I would like to thank Isara Corp. for having given me the opportunity of being part of such a great team and for supporting the project for these (almost) three years.

A mi familia: Luis, Martha y Erick. Muchas gracias por siempre mostrar su apoyo y su amor. Siempre están en mi mente y en mi corazón.

A mis eternos amigos Abdón, Héctor, Poke y Daniel, porque siempre los estaré recordando.

Finally, to Joy. Because you pushed, supported and motivated me until the very end. Because of all the chapter we are writing together. Thank you mi amor!

Dedication

To Joy and Echo

Table of Contents

Chapter 1

Introduction

"It is always the case, with mathematics, that a little direct experience of thinking over things on your own can provide a much deeper understanding than merely reading about them."

— Roger Penrose

This thesis is a theoretical exploration of the Learning Problem in a context that makes it suitable for its use for public-key cryptography, as well as the relation that it has with other computational problems. The starting point of this journey is Learning With Errors (LWE), as introduced by Regev in [Reg05]; and, closely related to it, Short Integer Solutions (SIS), studied in [Ajt96, MR07]. These two problems are in the center of countless other works that constitute a very large portion of what is known nowadays as Lattice Based Cryptography. Our goal is to develop a deeper understanding of the Learning Problem and some of the techniques and tools that are commonly used in this area.

Lattice-Based Cryptography, at large, is recognized for its versatility and the several relations that it has with well-established problems in mathematics that are believed to be hard, even for quantum algorithms. These are some of the reasons why cryptographic constructions based on LWE and SIS have grown in popularity in the last decade, beginning with the initial proposals [LPR10, DXL12, Pei14]. However, what makes LWE and SIS particularly fascinating is their seamless combination of basic notions of algebra and geometry. The algebraic component provides them with the versatility that allows for the construction of a great variety of cryptographic primitives. The geometric component adds a new dimension of complexity—when we think about LWE or SIS as problems to be solved, we do not only look for an algebraic solution, we instead look for a solution that satisfies two types of constraints: algebraic and geometric. It is important to notice that this feature is not exclusive to LWE and SIS. Rather, it is a characteristic property that they share with lattice problems, which is what makes possible the connection between these two kinds of problems.

Lattice-based cryptosystems are also known for being "simple", when compared to their elliptic-curve counterparts. This simplicity, however, manifests itself only computationally, so to speak. It is not completely reflected in the description of the protocols—lattice-based cryptosystems are notorious for having a large number of parameters and the relation between them is often intricate or unclear—neither it is true that the simplicity can be seen in the reductions from these constructions to mathematical problems, which are far from simple and often misunderstood.

There are several essential aspects surrounding lattice-based cryptosystems— specifically those based on variants of LWE and SIS—that do not appear to be completely well understood, yet we rely on them to build protocols and set parameters. In fact, a case has been made in [CKMS17] and online forums that some of the existing reductions fail at providing an argument for the effective security of the cyptographic constructions. That is not to say that lattice-based cryptosystems are insecure—there are several reasons to believe that these constructions are safe even against adversaries with access to a full-fledged quantum computer—rather, it gives the community another reason to study other aspects of the problem that have not yet been explored.

In this work we dissect LWE to its core elements, with the goal of analyzing what role each component plays in the construction of cryptographic primitives and in its relation to lattice-problems.

1.1 The Learning Problem

Suppose that a function $f\colon X \to Y$ is defined between two sets X and Y (whose descriptions are available) and, further, suppose that we have access to a collection of input/output pairs $\big(x_i, f(x_i)\big) \in X \times Y$. Our task is to find a description of the function f using only the provided information. Solving this problem is impossible since, in general, there several—in fact, an exponential or even infinite number of—functions that, when evaluated at x_i, result in $f(x_i)$. Thus, without any restriction on f, the given pairs provide very little information about the function.

This problem becomes more interesting when the function f is limited to a well-defined family of functions. This restriction sometimes provides additional information that makes the task of finding the function possible. For example, if we know, in addition, that the function is a quadratic polynomial defined over $\mathbb{R}$, then three input/output pairs are enough to (quickly) find a description of such polynomial.

Learning A Function

If we talk about generic groups then it is natural to wonder *what does it mean to learn a function?*. To make sense of this question consider the following examples. A linear function f over an n-dimensional vector space V is completely described by the set of images $\big\{f(\mathbf{b})\colon \mathbf{b} \in B\big\}$, where B is a given basis for V. Given any list n of input/output pairs $\big(\mathbf{v}_i, f(\mathbf{v}_i)\big)$, assuming that the set $\{\mathbf{v}_i\}$ is linearly independent, provides a characterization of the function f. The same can be said in a more general setting. Suppose that M is a free module over a ring R—in other words, $M \cong \bigoplus_{i=1}^{n} R$. Then a module homomorphism $f\colon M \to M'$ to any R-module M' is completely determined by its value on a basis of $\{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$ of M.

Perhaps against our intuition, outlining the previous linear functions in such manner generally does not provide a "useful" description of the functions. By "useful" we mean that it allows us to efficiently evaluate the function on any element of the domain. In the previous cases, in order to compute $f(\mathbf{v})$ for an arbitrary element $\mathbf{v}$ we need to find an expression of $\mathbf{v}$ as a linear combination of the generators $\mathbf{v}_1, \ldots, \mathbf{v}_n$. For generic groups, there is no known classical algorithm that solves this problem.

We may circumvent this problem in the definition by outsourcing the task of evaluating the function to the solver. In other words, we say that an algorithm $\mathcal{A}$ *learns* the function f if it is able to find its value on any element of the domain. This is the notion we use in this thesis working in the context of generic structures. We explain this idea in more detail in Section 3.2.

This is sometimes equivalent to other conditions. For example, it is well known that the problem of learning a linear function can be performed efficiently when the underlying ring is $\mathbb{Z}_q$, and the module is $\mathbb{Z}_q^n$. Given a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathbb{Z}_q^n$ and an arbitrary element $\mathbf{v} \in \mathbb{Z}_q^n$, we can find an expression of $\mathbf{v}$ as a linear combination of the given basis by solving the associated system of linear equations.

Learning With Errors

Every module homomorphism $f\colon \mathbb{Z}_q^n \to \mathbb{Z}_q$ is given by an inner product. In other words, there exists $\mathbf{s} \in \mathbb{Z}_q^n$ such that for any $\mathbf{v}$, $f(\mathbf{v}) = \langle \mathbf{v}, \mathbf{s} \rangle$. Then, in this case, learning the function f is equivalent to finding the vector $\mathbf{s}$. This can also be formulated as a system of linear equations and thus be efficiently solved.

Intuitively, *Learning With Errors* (LWE) is the problem of solving "noisy linear equations modulo q". Using the previous equivalence, we can describe LWE as the problem of finding $\mathbf{s} \in \mathbb{Z}_q^n$ given a collection of pairs

$$\big(\mathbf{v}_i, \langle \mathbf{v}_i, \mathbf{s} \rangle + e_i\big) \in \mathbb{Z}_q^n \times \mathbb{Z}_q, \qquad i \in \{1, \dots, m\}, \tag{1.1}$$

where e_i is sampled from a known fixed distribution χ over $\mathbb{Z}_q$, and the vectors $\mathbf{v}_1, \dots, \mathbf{v}_m$ are assumed to be uniformly random. The list of pairs outlined in Equation (1.1) describe a probability distribution $A_{\mathbf{s},\chi}$ over $\mathbb{Z}_q^n \times \mathbb{Z}_q$ that depends on $\mathbf{s}$, as well as on the error distribution χ. The *decision* version of LWE is the problem of distinguishing $A_{\mathbf{s},\chi}$ from the uniform distribution over $\mathbb{Z}_q^n \times \mathbb{Z}_q$.

The problem was introduced by Regev in [Reg05], together with a quantum reduction to it from classical lattice problems for which it is believed no efficient quantum solution exists. Along with these results, in the same work the author describes the construction of a public-key encryption scheme whose security is guaranteed by the hardness of LWE, thus also by the quantum hardness the lattice problems this is related to. This attracted the attention of the cryptographic community to evaluate its hardness and improve the efficiency of the construction. To this day there is strong belief that, for certain parameterizations of LWE, this problem remains hard even for quantum algorithms.

In turn, LWE is a generalization of classical problem known as *Learning Parity with Noise* (LPN), a noisy version of a learning problem known as *Parity Learning*. The latter is the problem of learning a function $f\colon \{0,1\}^n \to \{0,1\}$ that computes the parity of the number of bits of a string at some unknown fixed locations. When the correctness of the value can only be guaranteed with probability $\chi \in \big(\frac{1}{2}, 1\big]$, then this is the particular case of LWE where the underlying ring is $\mathbb{Z}_2$.

Solving LWE

Regev's average-case to worst-case reduction to classical lattice problems, such as GapSVP_γ and SIVP_γ, is a strong indication that LWE is a hard problem in the average case. This had been a long desired property for a problem backing a cryptographic construction. However, the given reduction is rarely used to set the parameters of LWE-based constructions. Instead, at the present moment we rely on the best-known solutions and idealized attack models for this problem. These have been surveyed in several works [APS15, Pla18, BLP+13, MP13].

There are three main paths to solve LWE. We may try to find the secret $\mathbf{s}$, which directly solves the problem. Another way is to try to find the error, which would yield a solution of the problem by solving the resulting system of linear equations. Alternatively, given a collection of m samples $\left(\mathbf{a}_i, \langle \mathbf{a}_i, \mathbf{s} \rangle\right)$, we may try to find a short element $\mathbf{x}$ in the kernel of the matrix $\mathbf{A}$ whose rows are the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$. This way yields a method to distinguish this distribution from uniform, since $\langle \mathbf{x}, \mathbf{e} \rangle$ is expected to be small. The problem of finding a small vector in the kernel of a matrix is known as *Short Integer Solutions* SIS.

To follow these strategies, there are two categories of algorithms besides the exhaustive search and meet-in-the-middle kind of solutions. The more common is to transform an LWE into a lattice-type problem and try to find the solution of the latter. The best solutions of this kind are described in [Alb17, ADPS16]. These algorithms involve finding a “good” basis for the lattice corresponding to the given instance. The effectiveness of a solution of this type is thus affected by any improvement in lattice reduction algorithms.

On the other hand, Arora and Ge [AG11] proposed an algebraic solution for the associated SIS problem that runs in subexponential time whenever the error is small enough. This solution requires unlimited access to an oracle that generates samples of the LWE distribution $A_{\mathbf{s},\chi}$, which may not be the case in many practical situations.

In the particular case of LPN, the classic solution is the BKW algorithm proposed by Blum, Kalai and Wasserman [BKW00]. Despite running in $O\left(2^{n/\log n}\right)$ time, its main downside is that it requires $O\left(2^{n/\log n}\right)$ samples. Other modifications of this algorithm have been proposed that require only a polynomial number of samples [Lyu05, Kir11].

Learning Homomorphisms With Noise

In [BFN+11], Baumslag et al. introduced *Learning Homomorphisms with Noise* (LHN), which is a generalization of LWE to any abstract groups given a compact presentation of it. They do so by endowing the group with the geometry that is induced by the word distance—or Cayley distance, as it is called in the cited paper. As the authors mention, this distance is not always easy to compute; moreover, the problem is known to be NP-complete for certain instances [RMUV10]. However, as a particular example the authors propose the use of Burnside groups of exponent 3 (denoted as B_3) to instantiate their construction. In a follow-up paper, Fazio et at. [FIN+15] make a deeper study of the hardness of this problem on B_3, and provide a worst-case to average-case reduction of this problem, by proving that it is random self-reducible.

Using this problem as a basis, Baumslag et al. construct in [BFN+11] a symmetric-key cryptosystem that resembles Regev's encryption scheme. An important difference is that the noise in a ciphertext cannot be "erased" with the secret key, but can only distinguished from an arbitrary element of the group. As a consequence the message space is strictly limited to a single bit. Moreover, noise elements cannot be chosen to commute with the rest of the group. This makes it complicated to use the symmetric-key cryptosystem to obtain public-key encryption as done in [Reg05].

We addressed the problems described abobe in [LR19]. In this paper the authors propose that, by keeping an additional piece of information secret—namely, a normal subgroup of the image group—it is possible construct a public-key encryption scheme without the necessity of a norm defined over the group. The noise is sampled from a secret normal subgroup that is then collapsed in the decryption process.

Other Generalizations of LWE

The learning with errors problem has received special attention, and several efforts have been made to improve its efficiency. As a consequence, cryptosystems based on LWE have particularly enjoyed of a large number of improvements and generalizations. In 2009, *Polynomial Learning With Errors* (PLWE) was introduced by Stehlé et al. [SSTX09] as a way to optimize computation and key-sizes for the constructions based on LWE, apparently without compromising the practical security of these constructions. Shortly after, Lyubashevsky et al. [LPR10] independently proposed *Ring Learning With Errors* (RLWE), which further generalizes PLWE by allowing the objects to belong to the ring of integers of a number field. The recently popular *module*-LWE—first introduced in [BGV12] with the name *general*-LWE—is the generalization of RLWE to a multidimensional module over the same ring, generalizing both LWE and RLWE. Lastly, *Learning With Rounding* (LWR)[BPR12] is a variant of the original LWE problem on which the error is sampled deterministically.

There have been several works outlining generalized versions of LWE in different contexts. Short after Regev's introduction of LWE in 2005 [Reg05], Peikert published a work on the hardness of *error-correction in the exponent* [Pei06], on which he proves that, for suitable parameters on the error, *Bounded Distance Decoding* (BDD) for a black-box cyclic group is at least as hard as the discrete logarithm problem on the same group. This work lead to posterior analysis of the learning with errors problem in the exponent by Demarest et al. [DFR18], which generalizes the original formulation of LWE to the problem of decoding over the group C_p^n, and uses a new technique to provide a generic lower bound on the number of queries necessary to solve the decoding problem in this group. Independently, Dagdelen et al. studied the same problem in [DGG16], where they describe a relation of this to a generalization of the computational Diffie-Hellman problem.

Another approach generalizing the learning problem was proposed by Gama et al. in [GINX16]. They generalize the LWE and SIS problems to finite Abelian groups. The authors show that the more general versions of the problems still enjoy the same worst-case to average-case reductions that the original formulations have, provided that the instance group is large enough.

Another attempt to use non-commutative groups is described in [CZZ16]. In this manuscript, Cheng et al. study the learning problem over the group ring $R[G]$, an algebraic structure which consists of formal sums of element of G with coefficients in R. As a concrete instance they choose $R = \mathbb{Z}$ and $G = D_{2n}$, the dihedral group of order $2n$. Their main motivation is to recreate ring-LWE using a non-commutative (using integer coefficients) instead of cyclic groups (using coefficients in the integers or in a cyclotomic ring), to avoid attacks on principal ideal lattices.

Lastly, a recent work by Bootland et al. [BCSV19] describes a framework in linear algebra that encompasses different problems that have appeared in lattice based cryptography, such as LWE, MLWE and RLWE, as well as in code-based cryptography and the recent constructions modulo Mersenne primes. This framework allows one to obtain a generalization of problems such as LWE and SIS by choosing the environment: a parent ring, a ciphertext, modulus and a rank.

1.2 Constructing a Public Key Encryption Scheme

Regev provided a blueprint for the construction of cryptosystems based on the learning problem. The idea is to first consider a symmetric-key encryption scheme that works in a manner that resembles that of ElGamal encryption scheme. Provided that samples $\big(\mathbf{a}, b = \langle \mathbf{a}, \mathbf{s} \rangle + e\big)$ are indistinguishable from the uniform distribution over $\mathbb{Z}_q^n \times \mathbb{Z}_q$, an encrypting party encodes a bit β as $E(\beta) = \big(\mathbf{a}, b = \langle \mathbf{a}, \mathbf{s} \rangle + e + \beta\tau\big)$, with $\tau = \left\lfloor \frac{q}{2} \right\rceil$. To guarantee that the construction is semantically secure, the pairs of this form must also be indistinguishable from uniform.

The decryption process consists of two steps. A decrypting party uses the secret to compute $b - \langle \mathbf{s}, \mathbf{a} \rangle = \beta\tau + e$. If the resulting element is small we conclude that $\beta = 0$, and $\beta = 1$ if otherwise. This is the strategy used by Regev in the first proposed cryptosystem, and later adapted by Baumslag et al. in [BFN+11] in the construction of a symmetric-key encryption scheme over generic groups. The correctness follows as long as the noise e is bounded to the set $\left\{-\left\lfloor \frac{q}{4} \right\rceil, \ldots, \left\lfloor \frac{q}{4} \right\rceil\right\}$.

Using this mechanism and the bilinearity of the inner product, it is possible to obtain a this into a public-key encryption scheme. The public-key is a collection of encryptions of 0. To encrypt, choose a random subset of that collection. The summation of the elements of this subset results in pair of the form $\big(\mathbf{a}, b = \langle \mathbf{a}, \mathbf{s} \rangle + e\big)$, which follows an LWE distribution with larger error rate. If the accumulated error is still in the set $\left\{-\left\lfloor \frac{q}{4} \right\rceil, \ldots, \left\lfloor \frac{q}{4} \right\rceil\right\}$, we can encrypt and recover the message by using the algorithms described above.

This process can be easily generalized to a protocol that encrypts more than one bit per ciphertext. Say the message space is $\mathbb{Z}_t$, with $t < q$. In this case, the constant τ is chosen to be $\lfloor \frac{q}{t} \rceil$. To encrypt $\beta \in \mathbb{Z}_t$, consider $\beta \in \{0, \ldots, t-1\}$ and encode it as $\mu = \beta\tau \in \mathbb{Z}_q$. After this we proceed to complete the encryption process in the same way as before, by adding μ to the second coordinate of an LWE sample.

The decryption procedure is similar to the one above. The knowledge of the secret $\mathbf{s}$ allows the decrypting party to distinguish samples from the distribution $A_{\mathbf{s},\chi}$. This is done by reverting the action of the key on the second coordinate by computing $b - \langle \mathbf{s}, \mathbf{a} \rangle = \mu + e$. After this, the error is erased by "rounding off" $\mu + e$ to the closest multiple of τ. The correctness of the cryptosystem follows as long as the noise e is bounded to the set $\left\{-\left\lfloor \frac{q}{2t} \right\rceil, \ldots, \left\lfloor \frac{q}{2t} \right\rceil\right\}$.

Transferring This Idea to Generic Groups

The construction of the cryptosystem described above is concerned with homomorphisms $f\colon \mathbb{Z}_q^n \to \mathbb{Z}_q$. In the context of generic groups, an LWE sample is a pair of the form

$$\big(g, \varphi(g)h\big) \in G \times H, \tag{1.2}$$

where G and H are groups, $\varphi\colon G \to H$ is a group homomorphism, g is sampled from the uniform distribution over G and $h \in H$ is a "small" element of H. The public key is a collection of elements $\big(g_i, \varphi(g_i)h_i\big)$ sampled from this distribution. From a group-theoretic point of view, however, there are several important properties of these particular groups which make the construction possible that can be easily overlooked.

Combining Elements. The encryption procedure follows from the creation of a new element of the LWE distribution by means of taking a subset-sum of the elements in the public key. The commutativity of $\mathbb{Z}_q$ allows to accumulate the error in an error term e as

$$\sum_i r_i\big(\langle \mathbf{a}_i, \mathbf{s}\rangle + e_i\big) = \left\langle \sum_i r_i \mathbf{a}_i, \mathbf{s} \right\rangle + \sum_i r_i e_i,$$

where $\sum_i r_i e_i = e$. When the group is non-Abelian, the error terms find themselves interspersed with elements of the form $\varphi(g_i)$—unless, of course, the error is sampled form the center of the group; this is inconvenient for two main reasons. One is that the elements $\varphi(g_i)$ cannot be recombined into an element $\varphi(g)$. The second reason is that the error does not accumulate on one side of the expression; thus it is not obvious what is the effect of the noise terms in the combined sample.

Making the Noise Erasable. Erasing the noise by "rounding it off" to the closest multiple of a public constant is possible, in part, because there is a notion of distance between the elements of $\mathbb{Z}_q$. When speaking of generic groups, there are known notions of distance that can be used, such as the word metric. However, in general it is a hard problem to compute the word distance between two elements of a generic group. As a consequence, the strategy of decoding the noise seems hard to generalize in the context of abstract groups, since it is not clear what the proper notion of rounding is in general. Moreover, because of the manner in which the encryption step is performed, any effort of this kind will likely require the noise elements to be commutative with the rest of the group.

A Purely Algebraic Solution. An alternative is to specify the nature of the noise by algebraic means entirely, instead of using the geometry of the group. The idea is to erase the noise by using a secret homomorphism that maps it to the identity element of a third group K. Effectively this means that the noise elements are drawn from a secret normal subgroup N of H and the error is erased by performing a projection of H onto the quotient group H/N. An adversary without access to a description of the group N is then unable to distinguish error elements from other group elements.

In the quotient H/N, elements of the denominator N become the identity element, which commute with the rest of the elements of the quotient group. After erasing the noise, the terms $\varphi(g_i)$ can now be recombined in a way that relates to the first coordinate. However, the resulting combination is now in the quotient group. This strategy requires reversing the order of the steps in the traditional decryption algorithm for LWE. We elaborate on this idea in Section 4.1.

A Generic Public Key Encryption Scheme

Building on the idea discussed above, we propose an encryption scheme that can be instantiated with generic groups. We explain this construction in detail in Section 4.2. Consider groups G, H and K as part of the setup.

Generating the Keys. Choose two homomorphisms $\varphi\colon G \to H$, $\psi\colon H \to K$. Generate a collection of m pairs $\big(g_i, \varphi(g_i)e_i\big) \in G \times H$ where, for all i, g_i is chosen from a fixed probability distribution over G and e_i is chosen from a probability distribution over H with support in $\mathrm{Ker}(\psi)$. The secret key is the pair of homomorphisms (φ, ψ). The public key is the collection of pairs $\big(g_1, \varphi(g_1)e_1\big), \ldots, \big(g_m, \varphi(g_m)e_m\big)$ along with an element $\tau \in H \setminus \mathrm{Ker}(\psi)$.

Encryption. Similar to Regev's original encryption scheme, public-key encryption is achieved by randomly mixing elements from the public key, generating a new pair $(g, h) \in G \times H$. Contrary to the traditional setting, where mixing noisy elements by summing a random subset of the public key generates an element with the same structure, in the generic setting it is not guaranteed that the structure of the samples is preserved. Nevertheless, the result of mixing public key elements is a pair with specific form

$$(g, h') = \big(g_{i_1} \cdots g_{i_\ell}, \varphi(g_{i_1})e_{i_1} \cdots \varphi(g_{i_\ell})e_{i_\ell}\big).$$

This mixing can be done by forming an arbitrary word with the elements of the public key. To encrypt a message β, encode it by computing $\mu = \tau^\beta$ and output

$$\mathrm{Enc}(\beta) = (g, h'\mu).$$

Decryption. To decrypt, it is necessary to reverse the steps of the traditional LWE encryption scheme. More specifically, let $(g, h) \in G \times H$. If this pair is a well-formed ciphertext, then h is of the form

$$\varphi(g_{i_1})e_{i_1} \cdots \varphi(g_{i_\ell})e_{i_\ell}\tau^\beta,$$

where $g_{i_1} \cdots g_{i_\ell} = g$. Thus, by applying ψ to h we obtain

$$\psi\big(\varphi(g_{i_1})\big) \cdots \psi\big(\varphi(g_{i_\ell})\big)\psi\big(\tau^\beta\big) = \psi\big(\varphi(g)\big)\psi(\tau)^\beta.$$

This step effectively eliminates the noise elements. By multiplying by $\psi\big(\varphi(g)\big)^{-1}$ on the left, the right hand side becomes $\psi(\tau)^\beta$. The message β can be recovered computing the discrete logarithm with base $\psi(\tau)$.

Finding Instances. This is a generic protocol; thus by nature it cannot specify how to choose parameters and other specific details. These must be established once a particular group is chosen to instantiate the protocol. Finding an appropriate group is then the first step that needs to be taken. To instantiate this protocol, the corresponding groups must satisfy the following properties:

(a) The groups G must have enough normal subgroups. Otherwise the key-recovery problem is trivially broken.

(b) The corresponding probability distributions must be efficient to sample from.

We explore these requirements in Section 4.3.

How Safe is This Construction?

The short answer to this question is that we do not know. Nevertheless, we are able to say more than this. The conclusion is that the hardness of the underlying learning problem is related to the algebraic properties of the group. On the one hand, when only given a presentation of the group, the noiseless homomorphism learning problem is a generalization of the *Conjugacy* problem. This problem is known to be undecidable for certain classes of groups. For some others, such as Coxeter groups, a polynomial time solution has been found [Kra94]. For the *polycyclic*, *braid*, *Garside* and other groups, this problem is believed to be hard, yet still decidable.

In the presence of noise that is restricted to a normal group, the hardness of the problem is even more unclear. In the case of non-Abelian groups, we are only able to claim that the problem is at least as hard as the corresponding Conjugacy problem. More work has been done in the case of Abelian groups. In [Pei06], Peikert showed that the underlying Bounded Distance Decoding problem over a (finite) product of finite cyclic groups is at least as hard as the discrete logarithm problem on that group. On the other hand, Leonardi and Ruiz proved in [LR19] this construction is susceptible to quantum attacks when instantiated with any Abelian group. The details of this attack are outlined in Section 4.4.

Key Recovery. Given enough samples $(g_i, h_i) \in G \times H$ it is possible, by using Shor's algorithm, to find a combination of the elements g_i to generate the neutral 0_G element in the group G. By applying this combination to the given pairs themselves we obtain a pair of the form $(0_G, h)$, where h is in the secret normal subgroup of H. Repeating this attack would yield a set of generators of the secret subgroup.

Message Recovery. An immediate way to counter the previous attack is to provide the public key with fewer elements than the rank of G. However, this is far from a good solution. With few elements in the public key, it is easy to recover the linear combination performed in the encryption by using Shor's algorithm, which makes the encryption scheme completely insecure.

Hardness in the Classical Setting. It is important to notice that the previous attacks require a quantum algorithm in a generic setting. The previous attacks depend on the ability of the adversary to compute multidimensional discrete logarithms. An unanswered question at the moment is to what extent that ability is necessary. A positive answer would yield an interesting instance of an LWE-like cryptosystem using isogenies over elliptic curves. This problem is left as an open question for the future.

1.3 The Nature of Noise, Why Gaussians?

Having explored some algebraic aspects of LWE, we turn now to the geometry that is involved in the development of the theory around this problem. *Geometry of numbers* is a vast area of mathematics with a long history that is concerned with the geometry of the lattices themselves. It is in this area where we find the classical hard problems on which lattice-based cryptography is supported. However, when we talk about "small noise", geometry is also present in the group on which the learning problem is given. It is necessary to make use of a geometrical notion to define what a small group element is. Then is it not surprising that the tools that allow us to relate these two different kinds of geometries are very important. One of these tools is the Gaussian distribution.

Gaussians are ubiquitous in the theory of lattice-based cryptography. They are a fundamental tool for the construction of lattice cryptosystems, as well as for proving the security of them. They allow us to make a seamless transition from a finite discrete universe to an infinite continuous one. Thus perhaps it is natural to ask the question what makes Gaussians so special? Is it possible to obtain a similar transition from a different probability distribution?

Smoothing Parameter

Micciancio and Regev's reduction in [MR07] introduced a concept called the "smoothing parameter". This allows them to hide the discrete structure of a lattice in a quantifiable manner. Intuitively speaking, the *smoothing parameter* is the minimum stretching of a function f defined over the Euclidean space $\mathbb{R}^n$ that "hides the discreteness" of a lattice $\mathcal{L} \subset \mathbb{R}^n$ when centered at each $\mathbf{v} \in \mathcal{L}$. More concretely, it is the minimum scaling s of the function f such that the overall weight of the function on a (possibly shifted) lattice,

$$\sum_{\mathbf{x} \in \mathcal{L}+\mathbf{c}} f(s\mathbf{x}),$$

is approximately independent from the shift $\mathbf{c}$. Equivalently, it is the minimum scaling s of f such that the distribution of

$$g(\mathbf{z}) = \sum_{\mathbf{x} \in \mathcal{L}+\mathbf{z}} f(s\mathbf{x}),\ \mathbf{z} \in \mathcal{P}(\mathbf{B})$$

is approximately uniform over the fundamental parallelepiped $\mathcal{P}(\mathbf{B})$ of any basis $\mathbf{B}$ of $\mathcal{L}$. In each interpretation, the smoothing parameter is a function of the size of the lattice $\mathcal{L}$ and a parameter $\varepsilon \in (0, 1]$ which precisely quantifies these approximations.

It is not immediately clear from the description above that a smoothing parameter must exist, and even when it does, that it is reasonably sized. In fact this is not always the case. There exits integrable functions and lattices such that such scaling is not guaranteed to exist. Micciancio and Regev showed in [MR07] that for the Gaussian function $f(\mathbf{x}) = \rho(\mathbf{x}) = e^{-\pi\|\mathbf{x}\|_2^2}$, the smoothing parameter not only exists, but also only grows as $O\left(\sqrt{\log \frac{n}{\varepsilon}}\right)$, where n is the number of dimensions of the lattice $\mathcal{L}$, and ε measures roughly how far the distribution of g is from the uniform distribution over $\mathcal{P}(\mathbf{B})$.

The nature of smoothing parameters, the conditions for their existence, and methods for finding them become more apparent once we leverage the Poisson summation formula,

$$\sum_{\mathbf{x} \in s\mathcal{L}} f(\mathbf{x}+\mathbf{c}) = \frac{\det \mathcal{L}^*}{s} \sum_{\mathbf{y} \in \frac{1}{s}\mathcal{L}^*} \widehat{f}(\mathbf{y}) e^{-2\pi i \langle \mathbf{c}, \mathbf{y} \rangle}, \tag{1.3}$$

where $\widehat{f}$ is the Fourier transform of f, and $\mathcal{L}^*$ is the dual lattice of $\mathcal{L}$. Using this expression, it is possible to prove that under certain conditions, the weight of the shifted lattice becomes approximately independent from the shift—see [Reg05, Claim 3.8]. To see this, observe that on the right hand side of Equation (1.3), the vector $\mathbf{c}$ appears only in the exponent $-2\pi i\langle \mathbf{c}, \mathbf{y}\rangle$. Thus, if $\sum_{\mathbf{y}\in\frac{1}{s}\mathcal{L}^*\setminus\{\mathbf{0}\}} \left|\widehat{f}(\mathbf{y})\right|$ is small, $\sum_{\mathbf{y}\in\frac{1}{s}\mathcal{L}^*\setminus\{\mathbf{0}\}} \widehat{f}(\mathbf{y})e^{-2\pi i\langle \mathbf{c}, \mathbf{y}\rangle}$ is also small since the norm of $e^{-2\pi i\langle \mathbf{c}, \mathbf{y}\rangle}$ is bounded by 1. Moreover, for $\mathbf{y} = \mathbf{0}$, the term $\widehat{f}(\mathbf{y})e^{-2\pi i\langle \mathbf{c}, \mathbf{y}\rangle} = \widehat{f}(\mathbf{0})$. This implies that the overall weight of f over the shifted lattice $\mathcal{L} + \mathbf{c}$ is approximately $\frac{\det \mathcal{L}^*}{s}\widehat{f}(\mathbf{0})$, which is independent from the shift $\mathbf{c}$.

The above discussion allows us to precisely define the smoothing parameter of a function f with respect to a lattice $\mathcal{L}$ and a real $\varepsilon > 0$, as the minimum positive real number η such that for every $s \geq \eta$,

$$\sum_{\mathbf{x}\in s\mathcal{L}^*\setminus\{\mathbf{0}\}} \left|\widehat{f}(\mathbf{x})\right| < \varepsilon. \tag{1.4}$$

Using the newly defined smoothing parameter, the authors of [MR07] also presented a worst-case to average-case reduction from SIVP$_\gamma$ to SIS, by exploiting the ability to sample uniformly from cosets of $\mathcal{P}(\mathbf{B})$. Gentry, Peikert and Vaikuntanathan [GPV08] later presented an algorithm that samples (almost exactly) from a discrete Gaussian distribution over a lattice, which hides the geometry of the choice of basis used. They then showed how to use this algorithm to obtain tighter reductions from SIVP$_\gamma$ to SIS, and to instantiate a trapdoor signature scheme that enjoys worst-case hardness via a reduction from SIS.

The smoothing parameter is also essential to Regev's worst-case to average-case reduction from SIVP$_\gamma$ to LWE [Reg05]. To carry out the reduction, Regev exploited properties of a smoothening function as well as other properties of the Gaussian function—such as the fact that the family of Gaussians is closed under convolution and Fourier transform. Both the SIS and LWE reductions sparked many other derivative results, many of which fundamentally rely on properties of Gaussians and the smoothing parameter.

Currently, the formal study of lattice-related problems involving probability distributions different from the Gaussian has been very limited. As a consequence, there are no functions in the cryptographic literature, other than the Gaussian, for which the smoothing parameter has been quantified. In this paper we make a study of the family of functions that admit a smoothing parameter, and describe a subfamily that has been previously well studied in the mathematical literature. Afterwards we analyze the consequences, limitations and possible applications that these functions may have in the hardness of known problems in lattice cryptography.

Smoothening Functions, Smoothing Parameter and Tail Bounds

After briefly exploring part of the literature in lattice cryptography, it is perhaps not clear what extent it is necessary to use a Gaussian distribution to obtain a worst-case guarantee for a lattice-based cryptosystem. This leads us to wonder what is the largest family of functions for which we have a similar behavior? We explore these questions in Chapter 5.

The left hand side of Equation (1.4) is a summation over the non-zero vectors in some discrete set $\mathcal{L} \subset \mathbb{R}^n$, in other words, over vectors that are "far" from the origin. If the shortest vector of the lattice is "large enough", we can think of this sum as the weight of the tails of f over the lattice. Whenever f decreases fast enough (e.g. if $f(\mathbf{x}) = e^{-\|\mathbf{x}\|^2}$) then by scaling up the lattice, the weight of the function in the tails decreases. This means the overall weight on the lattice $\sum_{\mathbf{x}\in\mathcal{L}} f(\mathbf{x})$ tends towards $f(\mathbf{0})$ or, alternatively, that the ratio $f\big(\mathcal{L} \setminus \{\mathbf{0}\}\big)/f(\mathcal{L})$ becomes negligible.

In [Ban93], Banaszczyk proved that, in the case of the Gaussian distribution, this ratio decreases exponentially in the scaling factor. What is probably more important is to observe that Banaszczyk actually proved an upper bound on this ratio which holds for any lattice. This result allows the establishment of bounds for the smoothing parameter depending only on the density of the lattice, which may be quantified by means of parameters associated to the lattice such as the successive minima $\lambda_1(\mathcal{L})$, $\lambda_n(\mathcal{L})$, among others [MR07, ZZX20].

To develop analogous tools in a more general scenario—for norms other than the Euclidean norm and functions other than the Gaussian distribution—it is essential to understand how the ratio

$$\frac{f(\mathcal{L} \setminus rK)}{f(\mathcal{L})} \tag{1.5}$$

behaves whenever K is the ball of radius 1 of the norm that is being considered and $r > 0$. More importantly, in order to establish a bound for the smoothing parameter that only depends on the density of the lattice, we must find an upper bound $\nu_f(K)$ for Equation (1.5) that holds for every lattice $\mathcal{L}$. We refer to $\nu_f(K)$ as a tail bound for the function f. A tail bound is thus a parameter associated to th function that measures the distribution of its weight with respect to a certain region.

Generic Tail Bounds. In [MSD19], Miller and Stephens-Davidowitz make a study of the tail bounds for a family of strictly positive but rapidly decreasing functions f whose Fourier transform $\widehat{f}$ stays positive. Such is the case, for instance, for the Gaussian distribution itself. The results presented in their paper are thus a generalization of those in [Ban93]. We present a generalization of these results in Section 5.5.

For most applications in cryptography, f must be a non-negative function, as it is used to represent a probability distribution. On the other hand, there is in general no requirement for the characteristic function—the Fourier transform $\widehat{f}$—to be a probability distribution over the dual space. Moreover, minor modifications to a function may dramatically affect the behavior of its Fourier transform. Hence, several functions with potential cryptographic applications are left out if we further require the Fourier transform to be positive.

It is possible, however, to adapt some of these techniques to work in a more general case. To that end it is enough to find a positive and strictly decaying function that accurately describes the rate of eventual decay of the absolute value of $\widehat{f}$. This is, perhaps, the more technically challenging part of the process, since a description for $\widehat{f}$ is rarely available. Once we get past this obstacle, finding a tail bound for $\left|\widehat{f}\right|$ can be, roughly speaking, reduced to find an upper and lower tail bound for the so-called “bounding” function. The details are discussed in Section 5.7.

The procedure described above allows us to compute a tail bound for a function defined over $\mathbb{R}$. The last obstacle to overcome is to use it to compute a tail bound for a function defined over $\mathbb{R}^n$. Unfortunately, at the moment we do not know of a way to adapt this process to work for *any* such function. Nonetheless, if a function is described as a product of unidimensional functions evaluated on each coordinate, then it is possible to transform the unidimensional tail bound into a tail bound for the multidimensional function.

Bounding the Smoothing Parameter. Finally, after a tail bound for a function has been established, it is possible to bound the smoothing parameter. At fist sight, these two notions must be related—after all, the smoothing parameter is found once most of the weight of the characteristic function restricted to the dual lattice is found in the origin. However, when we give a second look, we realize that the notion of a tail bound intrinsically depends on a given geometry which determines whether a point is close or far from the origin. Fortunately it is possible to find a transformation that is somewhat oblivious to this, finding a smoothing parameter that only depends on the successive minima of the lattice with respect of the aforementioned geometry.

Generalized Gaussians and Other Functions. In [MSD19], Miller and Stephens-Davidowitz use the asymptotic geometry of the function $1/(1 + 2\cosh)$ to prove a transference theorem in the ℓ_1-norm by leveraging a transformation that uses the tail bound of the function. As a proof of concept, in Subsection 5.7 we use this number to obtain a bound for the smoothing parameter of this function. This particular example bypasses the necessity of finding a bounding function for the Fourier transform.

In the same work, the authors also bound the tails for the supergaussians $\rho^{[p]} \colon x \mapsto e^{-|x|^p}$, where $p \in (0, 2]$. Finding a smoothing parameter for these functions can be accomplished similar to the previous case. A more challenging endeavor is to compute the smoothing parameter for the supergaussians when $p \in \mathbb{R}_{>2}$ since, in this case, little is known about the characteristic function; moreover, it is possible to prove that they show a behavior that is not compatible with the techniques presented in their work.

We circumvent this problem by making use of the saddle point to find an asymptotic approximation for $\sigma_p = \widehat{\rho}^{[p]}$. Nevertheless, this approximation has several limitations. For instance, the obtained function is not positive, and the error rate—the difference between σ_p and its approximation—is not specifically quantified. As a consequence, the resulting bound is somewhat loose and the proof relies on experimental observations.

Average-Case to Worst-Case Reductions

As mentioned at the beginning of this section, this part of our work was partially motivated by questioning the possibility of obtaining a average-case to worst-case relation between lattice problems and SIS/LWE that makes no use of Gaussians. We focus on the particular case of LWE, following closely the ideas found in [Reg05]. To start investigating that question, in Section 6.1 we identify the steps in the reduction where the Gaussian is used.

Narrow Discrete Gaussians and SIVP. The main idea of proof is to construct an algorithm that outputs vectors according to a narrow discrete Gaussian distribution. Since the Gaussian distribution is ℓ_2 spherical, a significant portion of the mass of the discrete Gaussian lies outside of any proper subspace. Then it is expected—and it can be proven—that after a polynomial number of calls to a discrete Gaussian sampling algorithm we will obtain a set of n short and linearly independent vectors. Thus finding such a sampling algorithm is the focus of the reduction. Notice, however, that this could potentially be achieved by any discrete distribution whose weight is not concentrated on any proper subspace.

From Wide Discrete Gaussian to Narrow Mountains. Sampling from an exponentially wide discrete Gaussian can be done efficiently by leveraging the LLL algorithm. The characteristic function (Fourier transform) of this distribution is, in the Fourier space, a series of narrow Gaussians centered at each point in the dual lattice. Conversely, once one is able to reproduce a series of wider Gaussians centered at each point in the dual lattice, then we are able to simulate a narrower discrete Gaussian in the primal lattice. This wide-narrow correspondence between a discrete distribution and its characteristic function is a property of the Fourier transform itself, and not specific to Gaussian functions.

Building a Gaussian Quantum State and BDD**.** An elegant way to reproduce the characteristic function of a discrete Gaussian in a useful way is to construct a quantum state that simulates its behavior. The idea to construct this state is to have a superposition of every point in the dual lattice in one register and form a narrow Gaussian around each one of these points in a second quantum register. To obtain only a state encoding the value of the function—to make the value from the two registers independent—it is necessary to "erase" values stored in the first register. Since most of the weight of this function is centered closely around (dual) lattice points, this step becomes a Bounded Distance Decoding (BDD) problem over the dual lattice.

From BDD **to** LWE**.** Arguably the most important contribution of [Reg05] is the connection between LWE and BDD. This idea—or slight variations of it—has been used to provide an argument for the classical security of LWE-based constructions such as [ACPS09, Pei08, LPR10, BCD+19]. Moreover, this step of the reduction is completely classical, and it is the only one that makes use of the sampling algorithm that makes use of the LWE oracle.

The idea is the following. We assume we have access to a polynomial number of samples $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ of the discrete Gaussian distribution over the primal lattice whose generator matrix is B. Let $\mathbf{x}$ be a BDD instance in the dual lattice and call $\mathbf{t}$ the difference between $\mathbf{x}$ and the vector that is closest to it in the dual lattice. Then we consider the collection of pairs $\big(B^{-1}\mathbf{v}_i,\ \langle \mathbf{x}, \mathbf{v}_i \rangle\big)$. Each one of these new samples can be expressed as $\big(\mathbf{a}_i, \langle \mathbf{a}_i, \mathbf{s}\rangle + \langle \mathbf{t}, \mathbf{v}_i \rangle\big)$. When reducing modulo q, they now resemble LWE instances, however, there is a caveat. The difference between them and legitimate instances of LWE is that the probability distribution describing $\langle \mathbf{t}, \mathbf{v}_i \rangle$ is not a Gaussian distribution. Furthermore, the distribution itself depends on $\mathbf{t}$ and the given lattice. This issue is later fixed by adding the right amount of Gaussian noise to the second coordinate, obtaining $\big(\mathbf{a}_i, \langle \mathbf{a}_i, \mathbf{s}\rangle + \langle \mathbf{t}, \mathbf{v}_i \rangle + e'\big)$. Finally, is possible to prove that the corrected error $e = \langle \mathbf{t}, \mathbf{v}_i \rangle + e'$ is described by a continuous Gaussian (of unknown width), which makes it (almost) independent of these two parameters.

What Is Possible

We can summarize the impact of the results obtained in chapters 5 and 6 in the following points:

- It is possible to generalize several classical results in lattice cryptography to a context that is independent from the Gaussian distribution.
- There exists an infinite family of functions that admit a smoothing parameter.

However, this is not to say that average-case to worst-case reductions can be proved without using Gaussians. In particular, the analysis of cryptographic constructions based on LWE that use non-Gaussian distributions in the noise is still future work. Some of the remaining roadblocks are covered in the following subsection.

Relating BDD **to a Learning Problem.** One important aspect of the transformation described above, from an instance of BDD to a collection of instances $\big(\mathbf{a}_i, \langle \mathbf{a}_i, \mathbf{s} \rangle + e\big)$ of LWE, is that the distribution of the vectors $\mathbf{a}_i$ is provably close to uniform. This is precisely a consequence of the fact that the distribution of the auxiliary samples $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\} \subset \mathcal{L}$ is wide enough that is probably smoothening for the sublattice $q\mathcal{L}$.

For this reason it is not hard to imagine that, using auxiliary lattice vectors sampled from any distribution that is smoothening for $q\mathcal{L}$, it is possible to transform a BDD instance into a collection of instances of a learning problem. Nonetheless, as mentioned above, the noise defining the problem is not independent of the BDD instance itself; hence it is not possible to guarantee a connection from the worst case of BDD to an average case of this learning problem.

Initializing the Quantum Sampler. The second part of the reduction consists of using an oracle that solves the BDD on the dual lattice to sample vectors in the primal lattice according to their Gaussian weight. This is done by means of a quantum algorithm. To initiate this algorithm it is necessary to create a quantum state representing a (very) wide Gaussian over the elements of the (dual) lattice—which is later used to represent a (very) narrow Gaussian centered at every point in the dual lattice.

Despite the proof of correctness using similar techniques, the construction of the state relies on several properties of the Gaussian distribution, in addition to it being smoothening. Since the state can only be a finite superposition of elements, we must guarantee that most of the information about the function is contained in a compact set—in other words, that the tails of the function are negligible. In addition, the final state represents a slight shift of the function. Thus we also rely in the function being somewhat "stable" around the origin. More precisely, we need that the value of the function remains relatively constant in a neighborhood on the origin containing the fundamental parallelepiped generated by the (reduced) basis. We analyze this process and the requirements on the function in Section 6.2.

An important note is that we have not been able to circumvent the need for the function to be positive. Since this is the characteristic function of a probability distribution, we consider this to be a strong roadblock towards constructing a generic quantum sampler.

SIVP **from** DGS**.** The final step in the LWE–SIVP relation is to use discrete Gaussian sampler to solve the SIVP problem. Intuitively, it should be sufficient to obtain a large enough collection of samples from the algorithm to guarantee that this set contains a short basis for the lattice. The proof technique followed by Regev exploits the fact that the Gaussian distribution is ℓ_2-symmetric to prove that, with very high probability, the algorithm will sample a vector outside any given hyperplane. However, by assuming that the function is smoothening with respect to a proper sublattice, it is possible to argue that the distribution that is obtained on the quotient is close enough to uniform. Thus after a polynomial number of samples we obtain, with high probability, a set of linearly independent vectors.

The next step is to quantify how small the resulting set is. To this end we make use of the tools developed to bound the tails of a function, this time applied to the primal function. These ideas are explored in Section 6.3.

What Seems Impossible

In summary, it is not possible to adapt the current techniques to obtain an average-case to worst case reduction from lattice problems—say BDD—to LWE. The reason for the previous statement is because a key part of the reduction outlined in [Reg05] uses an additional property of the Gaussian, namely that the Gaussian is an *invariant distribution*.

Also, of independent but related interest, we prove that the only probability distribution over $\mathbb{R}^n$ that is described as a joint unidimensional distributions on its coordinates and some other coordinate system must be the Gaussian. In particular, this implies that the any result requiring an ℓ_2-symmetric distribution must use the a Gaussian distribution.

Recovering the Noise Shape. The last step of the BDD to LWE reduction—correcting the noise distribution to obtain a sample that is compatible with the LWE oracle—is the only step we identified that necessarily relies in characteristics that are almost exclusive to Gaussian functions. As mentioned above, the raw error obtained after the transformation is given as the inner product $\langle \mathbf{t}, \mathbf{v} \rangle$. The vector $\mathbf{t}$ is a small vector—the offset from the dual lattice and the BDD instance. However, the vector $\mathbf{v}$ follows a discrete distribution over the lattice. For this reason, the distribution of $\langle \mathbf{f}, \mathbf{v} \rangle$ is intrinsically dependent on the structure of the lattice as well as on the particular offset $\mathbf{t}$.

There are several different approaches in the literature to solve this problem. In [Reg05], Regev opted for correcting the noise by adding to it a "continuous noise", then obtaining a final error described as a continuous Gaussian. The continuity dispenses with the relation between the noise and the lattice. Moreover, the noise is only related to the offset by its magnitude—not the direction—which is dealt with separately.

The given equivalence between the corrected noise and the continuous Gaussian uses a divisibility argument of the distribution to write the added noise e as an inner product $\langle \mathbf{h}, \mathbf{t} \rangle$, where $\mathbf{h}$ follows a Gaussian distribution.

Nearest Planes Discrete Gaussian Sampling. Finally, we address the possibility of constructing lattice trapdoors following the strategy described in [GPV08]. This construction relies on an algorithm to sample from a discrete Gaussian distribution for any given lattice. The algorithm proposed in the same work is a randomized variant of the Babai's Nearest Planes algorithm [Bab86].

This process starts by computing the Gram-Schmidt basis from the given basis of the lattice. Notice that the resulting vectors, despite forming an orthogonal basis, have somewhat arbitrary directions. The algorithm proceeds to sample integer multiples of the vectors in the original basis according to a scaled discrete Gaussian. These scaling factors yield the same proportion in each direction determined by the Gram-Schmidt vectors. As a consequence, the resulting discrete distribution is proportional to a spherical Gaussian.

The correctness of the previous procedure relies on two important facts about Gaussian functions. The first one is that it is spherical, which means that it is symmetrical with respect to the ℓ_2-distance. The second (and, perhaps, most important) is that it factors along *any* orthonormal basis, in other words, the random variable generated by sampling from an n-dimensional Gaussian and projecting over *any* orthonormal basis generates n independently distributed random variables. A consequence of Kac-Bernstein's Theorem [Kac39, Ber41] is that this is a property that characterizes the normal distributions over $\mathbb{R}^n$, among distributions with finite variance. A proof is given in Section 6.3. As a result, any discrete sampling algorithm that works for a non-Gaussian function must necessarily follow a different strategy.

1.4 The Abstract Pieces

Motivated by the understanding acquired while working with abstract functions in the real space we now explore the possibility to extend this understanding beyond $\mathbb{R}^n$. The purpose is to develop a framework that allows a different interpretation of the standard lattice theory that is used for cryptography. In Chapter 7, we argue that several concepts and results that are commonly used in lattice-based cryptography are not entirely geometrical, or not geometrical at all, and can be seen in a more abstract algebraic framework.

What is a Lattice?

Instead of the more traditional expression of a lattice, as a integer linear combination of vectors, we consider its alternative definition as a discrete subgroup of $\mathbb{R}^n$. Discrete subgroups exists in any (Hausdorff) topological group. However, in many cases these subgroups lack a meaningful structure that relates them to the traditional lattices in $\mathbb{R}^n$. Particularly, we are interested in the idea that a lattice tessellates the space into regular bodies of finite size. To replicate this behavior, we impose an additional restriction on the discrete group, which is that the quotient group has a finite well-behaved *volume function*.

Haar Measures and Topological Groups. The idea of "volume" is formalized by the notion of *measure*. In the real space, this is a well understood idea that allows the formal study of integration and probability. Translating the same idea to other groups requires to endow new group with a suitable topology—one that makes the group operations continuous. In such a group we are able to find a special measure that is invariant under the group operation. In other words, as in the real space, moving a (measurable) body via the group operation does not change its volume.

Pontryagin Duality and Fourier Analysis

The development and current understanding of lattice based cryptography strongly relies on the properties of the Fourier transform, the ability of expressing a probability distribution in terms of its characteristic function and so on. Using the Haar measure and the integral functions that are possible to construct based on it, we encounter a complete theory of harmonic analysis that provides a similar relation between a group and its dual. The most natural mathematical objects where all these concepts find natural definitions are locally compact Abelian (LCA) groups. The commutativity allows for the set of all continuous complex representations to be one-dimensional, hence to have a group structure via a correspondence with its group of characters.

Given a finite group, any group homomorphism to (the multiplicative group of) $\mathbb{C}$ has its image included in S^1—the subgroup of elements of norm 1. In general, the set of homomorphisms from an LCA group to $S^1 = \mathbb{R}/\mathbb{Z}$ forms a group, which is better known as the *dual group*. In relation to traditional lattices, the underlying group where lattices live is the real space $\mathbb{R}^n$; moreover, the dual group of $\mathbb{R}^n$ is $\mathbb{R}^n$ as well. More generally, the dual of an LCA group is itself an LCA group. However, the dual group is not always isomorphic to the primal group.

Smoothening Functions over LCA groups

The intuition given in the case of the reals is that a function is smoothening whenever the induced function over the quotient is close to constant. This intuition is usually presented by centering a Gaussian, wider than the covering radius, around every point in the lattice. This idea cannot be immediately translated to a generic LCA group. Nonetheless, the formal description of a smoothening function that we described for real lattices—a description in terms of the characteristic function of the distribution—can be easily interpreted in a more abstract language.

In Section 7.2 we prove that some of the most important theoretical results—namely [Reg05, Claim 3.8] and [MR07, Lemma 4.4]—can be obtained in the generic setting of LCA groups. This opens the possibility of generalizing the theory of lattice cryptography, including some cryptographic constructions, to other LCA groups.

1.5 The Roadmap

This thesis is roughly divided into four main parts. We first give an overview of the general background that is utilized throughout the document. A reader familiar with most or all of the topics is advised to start in the following chapters. The second part consists of Chapters 3 and 4, where we make an exploration of the algebraic structure that makes possible constructing a public-key encryption scheme from the Learning With Errors problem. In Chapter 3 we view LWE as a learning problem and describe a model for learning a homomorphism in the black-box group model. Chapter 4 explores the possible applications of the homomorphism learning problem in public-key cryptography. The contents of Chapters 3 and 4 appear in [LR19].

The third part consists of chapters 5 and 6. This part is almost completely standalone. Chapter 5 is the longest chapter of the thesis. In it we make an exploration of what it means for a wide Gaussian to smoothen a lattice, and show that this property is not exclusive of the Gaussian functions. Once we know this, in Chapter 6 we explore the possibility of completely eliminating Gaussians from the classical reduction in [Reg05].

Finally, the last part, which consists only of Chapter 7, is dedicated to providing a translation of the common concepts from lattice theory to the more abstract world of harmonic analysis over LCA groups. To understand the final goal it is advised to be familiar with the content of Chapter 5. We argue that some geometric ideas have their underpinnings on a more general theory, which opens the door to explore similar ideas in very different objects.

Chapter 2

Background

"Impara a vedere. Renditi conto che ogni cosa è connessa con tutte le altre."

— Leonardo Da Vinci

This chapter is dedicated to providing the theoretical background. Most of the contents of this part of the thesis are only given for completeness and are meant to be for reference exclusively. A reader comfortable with the topic corresponding to each section may skip it completely.

2.1 Notation

We denote $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{C}$ the sets of integer, rational, real and complex numbers, respectively. For a totally ordered set S and $a \in S$ let $S_{>a}$ denote the set of element in S greater than a. We use an analogous notation for the relations $<$, $\geq$, $\leq$, $\neq$. For $n, m \in \mathbb{Z}_{>0}$ and a ring R, let $R^{n\times m}$ denote the set of matrices over R with n rows and m columns. Throughout the thesis, we use boldfaced lowercase letters to denote elements in a module of the form R^n. Let $\mathbf{e}_i \in \mathbb{R}^n$ denote the *ith canonical vector*, that is, the vector whose entries are all 0 except for the ith entry that is equal to 1. For $n \in \mathbb{Z}_{>0}$, let $[n]$ denote the set $\{1, \ldots, n\}$.

In chapters 5 and 6 we use boldfaced to denote multivariate functions—functions defined over $\mathbb{R}^n$.

2.2 Groups and Semigroups

A *semigroup* is a set S together with an associative binary operation $\cdot\colon S \times S \to S$, sometimes called the *semigroup law*. An element $e \in S$ is called *identity* if, for all $s \in S$, $s \cdot e = e \cdot s = s$. The identity element is unique in S. Given $s \in S$, an *inverse* of s is an element $s' \in S$ such that $s \cdot s' = s' \cdot s = e$. It follows that for all $s \in S$, the inverse is unique. The inverse of s is denoted by s^{-1}. A semigroup with an identity element and closed under inverses is called a *group*. A semigroup is *commutative* (or *Abelian*) if, for all $s, s' \in S$, $s \cdot s' = s' \cdot s$. If S is a group, a *subgroup* of S is a subset $H \subseteq S$ closed under the group operation and inverses, and such that $e \in H$. The subgroup relation is denoted as $H \leq S$. The *center* $Z(S)$ of a semigroup S is the set of elements $z \in S$ such that for all s, $zs = sz$. The subgroup *generated* by a collection $\{s_1, \ldots, s_\ell\} \subseteq S$ is the minimum subgroup $\langle s_1, \ldots s_\ell \rangle$ of S containing them. The *order* $O(s)$ of an element $s \in S$ is the cardinality of the group generated by s. A subgroup H of S is *normal* if, for all $s \in S$, $s^{-1}Hs = H$. We denote this relation as $H \trianglelefteq S$. The center is a normal subgroup. The (left) cosets sH of a normal subgroup H of S form a group under the operation $sHs'H = ss'H$. This is called the *quotient group*, and it is denoted as S/H.

Given two semigroups S and S', a mapping $\varphi\colon S \to S'$ is a *semigroup homomorphism* if for all $s, s' \in S$, $\varphi(ss') = \varphi(s)\varphi(s')$. If S and S' are groups, then it follows that $\varphi(e_S) = e_{S'}$ and for all $s \in S$, $\varphi(s^{-1}) = \varphi(s)^{-1}$. In this case φ is called a *group homomorphism.* A bijective homomorphism is called an *isomorphism.* If e' is the identity element of S', the *kernel* of a homomorphism is the set $\operatorname{Ker}(\varphi) \coloneqq \varphi^{-1}(e') \subseteq S$. The kernel of a homomorphism is a normal subgroup of S, moreover, if $\varphi\colon S \to S'$ is a homomorphism, the image of φ is a subgroup of S' isomorphic to $S/\operatorname{Ker}(\varphi)$. Given two homomorphisms $\varphi\colon H \to G$ and $\psi\colon G \mapsto K$, we say that the sequence

$$H \xrightarrow{\varphi} G \xrightarrow{\psi} K$$

is *exact* if $\varphi(H) = \operatorname{Ker}(\psi)$. A *short exact sequence* is a sequence of the form $0 \to H \to G \to K \to 0$, where every two consecutive morphisms form an exact sequence. In particular, in this case, the morphism $H \to G$ is injective and the morphism $G \to K$ is surjective.

For a subset $\Sigma = \{s_1, \ldots, s_m\}$ of a group S, a *word* on Σ of *length* ℓ is an expression of the form $s_{w_1}^{\sigma_1} \cdots s_{w_\ell}^{\sigma_\ell}$, where ℓ is a non-negative integer and for all $i \in \{1, \ldots, \ell\}$, $w_i \in \{1, \ldots, m\}$ and $\sigma_i = \pm 1$. The *empty word* is defined as the unique word of length 0. In this thesis we denote the sequence of indices $w_1, \ldots, w_\ell$ as $\mathbf{w}$, and the word $s_{w_1}^{\sigma_1} \cdots s_{w_\ell}^{\sigma_\ell}$ as $\prod_{\mathbf{w}} s_{w_i}^{\sigma_i}$. A word is *reduced* if it contains no subwords of the form ss^{-1} or $s^{-1}s$.

The *free group* with generating set Σ is the group of words over Σ with the concatenation-and-reduction operation. Every (finitely generated) group G is isomorphic a quotient of a (finitely generated) free group. If Σ is a generating set of such free group and R is a generating set of the kernel, we say that $\langle \Sigma \colon R \rangle$ is a *presentation* of G.

2.3 Metric and Normed Spaces

Let X be a set. A function $d \colon X \times X \to \mathbb{R}$ is called a *metric* if for all $x, y, z \in X$ it satisfies

1. $d(x, y) = d(y, x)$,
2. $d(x, y) = 0$ if and only if $x = y$,
3. $d(x, z) \leq d(x, y) + d(y, z)$.

Given two sets $Y, Z \subseteq X$, define $d(X, Y) = \inf \big\{ d(x, y) \colon x \in X, y \in Y \big\}$.

Let V be a vector space over a field $\mathbb{F} \in \{\mathbb{R}, \mathbb{Q}\}$. A function $|\cdot| \colon V \to \mathbb{R}_{\geq 0}$ is called a *norm* if for every $\mathbf{u}, \mathbf{v} \in V$ and every $a \in \mathbb{F}$, it satisfies

1. $|\mathbf{u} + \mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|$,
2. $|a\mathbf{u}| = |a||\mathbf{u}|$,
3. if $|\mathbf{u}| = 0$, then $\mathbf{u} = \mathbf{0}$.

A subset $S \subseteq \mathbb{R}^n$ is said to be *balanced* if for all $r \in [-1, 1]$, the set $rS \subseteq S$. For a balanced set S containing $\mathbf{0}$ in its interior, the function $\| \cdot \|_S \colon \mathbb{R}^n \to \mathbb{R}$ given by $\|\mathbf{x}\|_S = \inf\{r \in \mathbb{R}_{\geq 0} \colon \mathbf{x} \in rS\}$ defines a positive definite homogeneous function. The function $\| \cdot \|_S$ is a norm if and only if S is convex. Let $\mathcal{K}_n$ be the set of balanced convex and compact sets $K \subset \mathbb{R}^n$.

ℓ_p Norms. Let $p \in \mathbb{R}_{>0}$. For $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ define

$$\|\mathbf{x}\|_p \coloneqq \left(\sum_{i \in [n]} |x_i|^p \right)^{1/p} .$$

We extend this notion to $p = \infty$ by defining $\|\mathbf{x}\|_\infty \coloneqq \max \big\{ |x_i| \colon i \in [n] \big\}$. For $p \in \mathbb{R}_{\geq 1} \cup \{\infty\}$, the quantity $\|\mathbf{x}\|_p$ is called the *ℓ_p-norm* of $\mathbf{x}$. It is perhaps important to notice that, for p in the interval $(0, 1)$, the function $\| \cdot \|_p$ fails to satisfy the triangle inequality, hence it does not define a norm.

If $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_m\}$ is a set of vectors, we write $\|S\|_p = \max_i \|\mathbf{s}_i\|_p$. For $r \geq 0$ and $\mathbf{c} \in \mathbb{R}^n$, the set

$$rB_n^p(\mathbf{c}) \coloneqq \left\{\mathbf{x} \in \mathbb{R}^n \colon \|\mathbf{x} - \mathbf{c}\|_p \leq r\right\}$$

is the ℓ_p *ball* of radius r centered at $\mathbf{c}$. For $r = 1$ and $\mathbf{c} = \mathbf{0}$, we simply denote $1B_n^p(\mathbf{0})$ as B_n^p.

It is true that for $p, q \in \mathbb{R}_{>0}$, if $p \geq q$ then $B_n^p \subseteq B_n^q$. As a consequence, for any $\mathbf{x} \in \mathbb{R}^n$, if $p \geq q$ then $\|\mathbf{x}\|_q \leq \|\mathbf{x}\|_p$. Moreover, under these conditions on p and q,

$$\|\mathbf{x}\|_q \leq \|\mathbf{x}\|_p \leq n^{1/p-1/q}\|\mathbf{x}\|_q. \tag{2.1}$$

This relation is a direct application of Hölder's inequality [MVR97, Theorem 2].

2.4 Probability

Measure Theoretic Definitions

We use several concepts of probability theory in various different scenarios. While Chapters 5 and 6 use standard probability theory over $\mathbb{R}^n$, certain parts in Chapters 3, 4 and 7 require the more general language of measure theory. To avoid repetition, in this section we introduce the relevant concepts in their most general form.

Definition 2.1 (Measurable Spaces and Sets)**.** For a non-empty set S let $\mathcal{M}$ be a subset of the power set $\wp(S)$ of S. The set $\mathcal{M}$ is called a *σ-algebra* in S if

1. $\emptyset \in \mathcal{M}$,
2. For all $E \in \mathcal{M}$, $S \setminus E \in \mathcal{M}$,
3. $\{E_i \colon i \in I\} \subseteq \mathcal{M}$ implies that $\bigcup_{i \in I} E_i \in \mathcal{M}$.

If $\mathcal{M}$ is a σ-algebra of a non-empty set S, then the pair $(S, \mathcal{M})$ is called a *measurable space* and the members of $\mathcal{M}$ are called the *measurable sets* in S. For any collection of subsets $S \subseteq \mathcal{P}$, the smallest σ-algebra $\mathcal{M}$ in S containing S is called the *σ-algebra generated by S*.

In this thesis we mostly consider measures over topological spaces. Possibly the most intuitive way to regard a topological space S as a measurable space is to consider the σ-algebra consisting of arbitrary unions and intersections of open sets of S. Elements in this σ-algebra are called *Borel sets*. In chapters 3–6 we only consider measures over Borel sets.

Definition 2.2 (Measures and Probability Spaces)**.** Let $(S, \mathcal{M})$ be a measurable space. A *measure* over S is a function $\mu\colon \mathcal{M} \to \mathbb{R}_{\geq 0}\{\infty\}$ satisfying the following properties.

(i) $\mu(\emptyset) = 0$.

(ii) If $\{A_i\colon i \in \mathbb{N}\} \subseteq \mathcal{M}$ is a collection of pairwise disjoint measurable sets, then $\sum_{i\in\mathbb{N}} \mu(A_i) = \mu\Big(\bigcup_{i\in\mathbb{N}} A_i\Big)$.

We say that a property holds for *almost every* $x \in S$, if it fails to hold for only elements in a set of measure 0. A measure μ is called a *probability measure* if $\mu(S) = 1$. In this case, the triple $(S, \mathcal{M}, \mu)$ is called a *probability space*.

Definition 2.3 (Measurable Functions and Random Variables)**.** Let $(S, \mathcal{M})$, $(S', \mathcal{M}')$ be measurable spaces. A function $\phi\colon S \to S'$ is said to be *measurable* if for all $A \in \mathcal{M}'$, the preimage $\phi^{-1}(A) \in \mathcal{M}$.

Let $(S, \mathcal{M}, f)$ be a probability space and let $(S', \mathcal{M}')$ be a measurable space. A *random variable* is a measurable function $X\colon S \to S'$.

It follows from the previous definition that, when considering the σ-algebra of Borel sets on topological spaces, every continuous function is measurable.

Probability Over $\mathbb{R}^n$

We now consider $\mathbb{R}^n$ as a measurable space with the σ-algebra of Borel sets. The usual way to endow $\mathbb{R}^n$ with a measure is to consider the *Lebesgue measure* which, intuitively, is given by the volume of a set.

A *probability distribution* over $\mathbb{R}^n$ is a non-negative integrable function $f\colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ such that $\|f\|_1 := \int_{\mathbb{R}^n} f(\mathbf{x})d\mathbf{x} = 1$. A probability distribution induces a probability measure on $\mathbb{R}^n$ given, for every Borel set A, by $\mu(A) = \int_A f(\mathbf{x})d\mathbf{x}$.

Given two probability distributions f_1, f_2 over $\mathbb{R}^n$, the addition of their corresponding random variables is described by the *convolution* defined as

$$(f_1 * f_2)(\mathbf{x}) := \int_{\mathbb{R}^n} f_1(\mathbf{z}) f_2(\mathbf{x} - \mathbf{z}) d\mathbf{z}.$$

The *statistical distance* between f_1 and f_2 is given by

$$\Delta(f_1, f_2) := \frac{1}{2} \int_{\mathbb{R}^n} \big|f_1(\mathbf{x}) - f_2(\mathbf{x})\big| d\mathbf{x}.$$

Given a probability distribution f over $\mathbb{R}$ and $n \in \mathbb{Z}_{>0}$, the *nth moment* of f with respect to $c \in \mathbb{R}$ is given by

$$\int_{\mathbb{R}} (x-c)^n f(x) dx.$$

Gaussian Distributions. For $s \in \mathbb{R}$ with $s \neq 0$ and $\mathbf{c} \in \mathbb{R}^n$, the *spheric Gaussian* of *width* s *centered* at $\mathbf{c}$ is the function $\rho_{s,\mathbf{c}} \colon \mathbb{R}^n \to \mathbb{R}$ given by $\rho_{s,\mathbf{c}}(\mathbf{x}) = \exp\left(-\frac{(\mathbf{x}-\mathbf{c})^2}{2s^2}\right)$. If the center $\mathbf{c} = \mathbf{0}$, we denote $\rho_{s,\mathbf{0}}$ as ρ_s.

Proposition 2.4. *Let $X_1, \ldots, X_n$ be a collection of independent Gaussian random variables of widths $s_1, \ldots, s_n$ and center $c_1, \ldots, c_n$, respectively. Then the random variable obtained from the sum*

$$\sum_{i \in [n]} X_i$$

is a Gaussian random variable of width $\sqrt{\sum_{i \in [n]} s_i^2}$ centered at $(c_1, \ldots, c_n)$.

Proposition 2.4 allows us to prove the divisibility property of the Gaussians in the opposite direction. More precisely, given $s \in \mathbb{R}_{>0}$ and $n \in \mathbb{Z}_{>0}$ it is possible to express ρ_s as a sum of n identical independent (Gaussian) distributions. This property of Gaussian is called *infinite divisibility.*

Given a discrete set $A \subset \mathbb{R}^n$ and $s \in \mathbb{R}_{>0}$, the *discrete Gaussian* is the probability distribution $D_{A,s}$ over A defined as

$$D_{A,s} \colon \mathbf{x} \mapsto \frac{\rho_s(\mathbf{x})}{\rho_s(A)}, \tag{2.2}$$

where $\rho_s(A) \coloneqq \sum_{\mathbf{x} \in A} \rho_s(\mathbf{x})$, whenever this series is convergent.

2.5 Lattices

Lattices are central objects of study in lattice-based cryptography and geometry of numbers. In this section we provide the fundamentals of lattices over $\mathbb{R}^n$. A generalization of this concept is presented in Chapter 7.

Fundamentals

Definition 2.5. Let n be a positive integer and let $\mathbf{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_d\} \in \mathbb{R}^n$ be a set of linearly independent vectors. The *lattice* generated by $\mathbf{B}$ is the set

$$\mathcal{L}(\mathbf{B}) \coloneqq \left\{ \mathbf{v} \in \mathbb{R}^n \colon \mathbf{v} = \sum_{i=1}^{n} a_i \mathbf{b_i},\ a_i \in \mathbb{Z} \right\} \subset \mathbb{R}^n.$$

The number of basis vectors d is called the *rank* of the lattice. We say that a lattice is *full rank* whenever $d = n$. If $\mathcal{L}'$ is a lattice contained in $\mathcal{L}$, we say that $\mathcal{L}'$ is a *sublattice* of $\mathcal{L}$, and $\mathcal{L}$ is a *superlattice* of $\mathcal{L}'$.

Let $\mathcal{L}$ be a lattice generated by a basis $\mathbf{B}$. The *fundamental region* $\mathcal{P}(\mathbf{B})$ generated by $\mathbf{B}$ is the parallelepiped determined by the vectors in $\mathbf{B}$. A matrix B whose rows are the elements of the basis $\mathbf{B}$ is called a *generator matrix.* The *determinant* of a lattice with generator matrix B is given by $\sqrt{\det B^T B}$.

Any set of linearly independent vectors generates a lattice, on the other hand, any lattice has an infinite number of different bases. Moreover, two different bases generate the same lattice if and only if one can be obtained from the other via an unimodular transformation—a linear transformation whose determinant is ± 1. Under the given definition, the determinant of a full rank lattice $\mathcal{L}$ is the absolute value of the determinant of any generator matrix B of $\mathcal{L}$. In this case, the determinant of a lattice is also the volume of any if its fundamental regions.

Definition 2.6. Given a lattice $\mathcal{L} \subset \mathbb{R}^n$, its *dual lattice* is defined as

$$\mathcal{L}^* \coloneqq \big\{ \mathbf{u} \in \mathbb{R}^n \colon \text{for all } \mathbf{v} \in \mathcal{L},\ \langle \mathbf{u}, \mathbf{v} \rangle \in \mathbb{Z} \big\}.$$

Remark 2.7. If B is a generator matrix for a lattice $\mathcal{L}$, then $B^* \coloneqq (B^T)^{-1}$ is a generator matrix for the lattice $\mathcal{L}^*$. It follows that, for $s \in \mathbb{R}_{>0}$, the dual lattice of $s\mathcal{L}$ is given by $\frac{1}{s}\mathcal{L}^*$.

Definition 2.8 (Successive Minima)**.** Let $K \in \mathcal{K}_n$ and let $\mathcal{L} \subset \mathbb{R}^n$ be a lattice. Then, for $i \in \{1, \ldots, n\}$, let

$$\lambda_i(K, \mathcal{L}) \coloneqq \min \big\{ \lambda \in \mathbb{R}_{>0} \colon \dim(\lambda K \cap \mathcal{L}) = i \big\}. \tag{2.3}$$

When $K = B_1^p$, we denote $\lambda_i(B_1^p, \mathcal{L})$ as $\lambda_i^p(\mathcal{L})$. If $p = 2$, then the exponent is omitted, thus $\lambda_i(\mathcal{L}) = \lambda_i^2(\mathcal{L})$.

The successive minima associated to a lattice can be related to their counterparts for the dual lattice. These are known as *transference theorems*. As an example, the following theorem provides an upper and lower bound for $\lambda_i(\mathcal{L})$ in terms of $\lambda_{n-i+1}(\mathcal{L}^*)$.

Theorem 2.9 ([Ban93, Theorem 2.1])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$, and $i \in [1, n]$,*

$$1 \leq \lambda_i(\mathcal{L}) \cdot \lambda_{n-i+1}(\mathcal{L}^*) \leq n. \tag{2.4}$$

Lattice Problems

We now review a few of the lattice problems that are relevant for the discussion in the following chapters. The problems are presented in the most generic form, that is, for any norm $\|\cdot\|_K$, with $K \in \mathcal{K}_n$. As before, if $K = B_n^p$ inducing an ℓ_p norm, then K is substituted for the letter p. If $p = 2$, then this parameter is omitted in the notation.

Definition 2.10 (Shortest Vector Problem (SVP_γ^K))**.** Let $\gamma > 1$ be an approximation function and let $K \in \mathcal{K}_n$. Given a set $\mathbf{B} \subset \mathbb{R}^n$ of linearly independent vectors, find a non-zero vector $\mathbf{v} \in \mathcal{L} = \mathcal{L}(\mathbf{B})$ such that $\|\mathbf{v}\|_K \leq \gamma(n) \cdot \lambda_1(K, \mathcal{L})$.

The following can be seen as a decision version of SVP_γ^K. Notice, however, that not every instance of the problem can be *decided*. In other words, the definition of the problem allows for the existence of lattices that are neither YES or NO instances.

Definition 2.11 (Gap Shortest Vector Problem (GapSVP_γ^K))**.** Let $\gamma > 1$ be an approximation function. Given a set $\mathbf{B} \subset \mathbb{R}^n$ of linearly independent vectors and a number $d \in \mathbb{R}_{>0}$, output

$$\begin{cases} \text{YES} & \text{if } \lambda_1\big(K, \mathcal{L}(\mathbf{B})\big) \leq d, \\ \text{NO} & \text{if } \lambda_1\big(K, \mathcal{L}(\mathbf{B})\big) > \gamma(n) \cdot d. \end{cases}$$

The following problem can be seen as a variant of SVP_γ. In this case, the goal is to find a set of linearly independent lattice vectors where every element is "small".

Definition 2.12 (Shortest Independent Vector Problem (SIVP_γ))**.** Let $\gamma > 1$ be an approximation function and let $K \in \mathcal{K}_n$. Given a set $\mathbf{B} \subset \mathbb{R}^n$ of linearly independent vectors, find a set of linearly independent vectors $\mathbf{S} \subset \mathcal{L} = \mathcal{L}(\mathbf{B})$ such that $\|\mathbf{S}\|_K \leq \gamma(n)\lambda_n(K, \mathcal{L})$.

In definitions 2.13 and 2.14, let dist_K denote the metric function induced by the norm $\|\cdot\|_K$. Then, in particular, for a set $A \subset \mathbb{R}^n$ and a vector $\mathbf{t} \in \mathbb{R}^n$,

$$\mathrm{dist}_K(A, \mathbf{t}) \coloneqq \inf\big\{\|\mathbf{x} - \mathbf{t}\|_K \colon \mathbf{x} \in A\big\}.$$

Definition 2.13 (Closest Vector Problem (CVP_γ))**.** Let $\gamma > 1$ be an approximation function and let $K \in \mathcal{K}_n$. Given a set $\mathbf{B} \subset \mathbb{R}^n$ of linearly independent vectors and a target vector $\mathbf{t} \in \mathbb{R}^n$, find a lattice vector $\mathbf{v} \in \mathcal{L} = \mathcal{L}(\mathbf{B})$ such that $\|\mathbf{v} - \mathbf{t}\|_K < \gamma(n) \cdot \mathrm{dist}_K(\mathcal{L}, \mathbf{t})$.

The Closest Vector Problem imposes no restrictions on the lattice $\mathcal{L}$ and the target vector $\mathbf{t}$. It is known to be NP-Hard for approximation factor $\gamma(n) = n^{1/\log\log n}$ [DKRS03]. The following problem is a variant of CVP, where the target vector is required to be close enough to the lattice.

Definition 2.14 (Bounded Distance Decoding (BDD))**.** Let $d \in \mathbb{R}_{>0}$. Given a set $\mathbf{B} \subset \mathbb{R}^n$ of linearly independent vectors and a target vector $\mathbf{t} \in \mathbb{R}^n$ such that $\mathrm{dist}_p\big(\mathcal{L}(\mathbf{B}), \mathbf{t}\big) < d$, find a lattice vector $\mathbf{v} \in \mathcal{L}(\mathbf{B})$ such that $\|\mathbf{v} - \mathbf{t}\|_K = \mathrm{dist}_K\big(\mathcal{L}(\mathbf{B}), \mathbf{t}\big)$.

Finally, the following is a problem about sampling from a discrete Gaussian distribution, as defined in Equation (2.2). Despite being relatively new, in recent years it has become a standard problem in lattice cryptography literature. The parameter φ is an arbitrary real valued function meant to represent a parameter related to the lattice, for instance, $\lambda_1(\mathcal{L})$.

Definition 2.15 (Discrete Gaussian Sampling (DGS_φ))**.** Let φ be a function associated to the lattice. Given a set $\mathbf{B} \subset \mathbb{R}^n$ of linearly independent vectors and $r \in \mathbb{R}_{>\varphi(\mathcal{L})}$, output a sample from the distribution $D_{\mathcal{L},r}$.

2.6 Learning With Errors and Short Integer Solutions

There are several different versions of the Learning With Errors problem in the literature. Here we review the original definition appearing in [Reg05].

Definition 2.16 (Learning With Errors (LWE) [Reg05])**.** Let $n, q \in \mathbb{Z}_{>0}$ and let $\chi\colon \mathbb{Z}_q \to [0,1]$ be a probability distribution. For $\mathbf{s} \in \mathbb{Z}_q^n$ let $A_{\mathbf{s},\chi}$ be the distribution over $\mathbb{Z}_q^n \times \mathbb{Z}_q$ obtained by sampling $\mathbf{a} \leftarrow U(\mathbb{Z}_q^n)$, $e \leftarrow \chi$, and outputting the pair $\big(\mathbf{a}, \langle \mathbf{a}, \mathbf{s}\rangle + e\big)$. *Search Learning With Errors* is the problem of finding $\mathbf{s} \in \mathbb{Z}_q^n$ given a collection of samples $(\mathbf{a}_i, b_i) \leftarrow A_{\mathbf{s},\chi}$. *Decision Learning With Errors* is the problem of deciding whether there exists $\mathbf{s} \in \mathbb{Z}_q^n$ such that a given collection of pairs $(\mathbf{a}_i, b_i) \in \mathbb{Z}_q^n \times \mathbb{Z}_q$ are sampled according to $A_{\mathbf{s},\chi}$, or they are sampled uniformly.

The *Short Integer Solutions* problem, in its inhomogeneous form, first appeared in [Ajt96]. In the cited paper, Ajtai constructed a one-way function and provided a reduction from the worst case of this problem to the average case of SVP_γ and SIVP_γ. A few years later, in [MR07], Micciancio and Regev improved these results quantitatively and qualitatively, by providing a tighter reduction and improving the approximation factors.

Definition 2.17 (Short Integer Solutions (SIS))**.** Let $n, m, q \in \mathbb{Z}_{>0}$, let $\beta \in \mathbb{R}$ and let $p \in [1, \infty]$. *Short Integer Solutions* over ℓ_p is the problem of finding a solution to a system of linear equations $\mathbf{A}\mathbf{x} \equiv \mathbf{0} \mod q$, with $\mathbf{A} \in \mathbb{Z}^{n\times m}$, such that $\mathbf{x} \in \mathbb{Z}^m \setminus \{\mathbf{0}\}$ and $\|\mathbf{x}\|_p \leq \beta$.

2.7 Elliptic Curves

Let $\mathbb{F}$ be field of characteristic different from 2 or 3. An *elliptic curve* is the set $E(\mathbb{F})$ of solutions to an equation of the form $E\colon y^2 = x^3 + ax + b$ over $\mathbb{F}$ and an additional identity element. An elliptic curve has a natural associative operation such that it becomes an Abelian group. The identity element is usually referred to as *the point at infinity*, and in this thesis is denoted as 0. An *isogeny* over $\mathbb{F}$ is a non-constant map $\phi\colon E_1(\mathbb{F}) \to E_2(\mathbb{F})$ of the form

$$(x, y) \mapsto \left(\frac{f_1(x)}{g_1(x)}, \frac{f_2(x)}{g_2(x)}y\right)$$

that fixes the point at infinity, where f_1, f_2, g_1, g_2 are polynomials in $\mathbb{F}[x]$. In this case, E_1 and E_2 are called *isogenous*. An isogeny induces a group homomorphism from $E_1(\mathbb{F})$ to $E_2(\mathbb{F})$. The *degree* of an isogeny is $\max\left\{f_1(x, y), g_1(x, y)\right\}$. An isogeny is called *separable* if the derivative of $\frac{f_1(x)}{g_1(x)}$ is nonzero.

Not every two curves are isogenous, however, for any prime power q, two elliptic curves E_1, E_2 are isogenous over $\mathbb{F}_q$ if and only if $\left|E_1(\mathbb{F}_q)\right| = \left|E_2(\mathbb{F}_q)\right|$. Furthermore, given a fixed curve $E_1(\mathbb{F}_q)$ and a subgroup $G \leq E_1(\mathbb{F}_q)$ there exist a curve $E_2(\mathbb{F}_q)$ and a separable isogeny $\phi\colon E_1 \to E_2$ over $\mathbb{F}_q$ with kernel G; and, E_2 and ϕ are unique up to $\mathbb{F}_q$-isomorphism. This isogeny can be computed from a set of generators of its kernel by using Velu's formulas [Vél71].

Definition 2.18 (Isogeny Problem)**.** Let E_1, E_2 be two elliptic curves over $\mathbb{F}_q$ such that $\left|E_1(\mathbb{F}_q)\right| = \left|E_2(\mathbb{F}_q)\right|$. The *isogeny problem* is the problem of finding an isogeny $\phi\colon E_1 \to E_2$.

Chapter 3

The Learning Problem

"Meaning lies as much
in the mind of the reader
as in the Haiku."
— Douglas Hofstadter in *Gödel, Escher, Bach:*
An Eternal Golden Braid

Learning, as a human phenomenon, is perhaps easy to understand. After all, humans experience learning on a regular basis. However, it is remarkably complicated to articulate into words what the concept means. It goes without saying that establishing the appropriate models for its formal study from the point of view of computers has been a particularly complex problem. This challenge has motivated several attempts to formulate models for learning that are appropriate for different scenarios [Val84, Val85, AL88, Kea98].

Despite the discussion above, we have an intuitive idea of what is to learn a function. For instance, it is well known that a polynomial function $p(x) = a_0 + a_1x + \ldots + a_nx^n$ of degree n over any field can be uniquely determined from $n+1$ input/output pairs $\big(a, p(a)\big)$. We "determine" this function by computing the coefficients $a_0, a_1, \ldots, a_n$ of $p(x)$. With this information we can efficiently compute the polynomial function at any point in the field, hence we can say we "learned" the function. The concept of "learning" a function can thus be thought as the process of acquiring enough information to efficiently simulate the behavior of the function at any point in the domain.

In [BFN$^+$11], Baumslag et. al present a framework to study the Learning With Errors problem over abstract groups. One of the missing pieces in that work, however, is a precise definition of what "learning a homomorphism" means in the context of groups. Since this is a generalization of Regev's definition in [Reg05], a suitable model for learning must be able to accommodate the existing notions of learning in the context of LWE. In this chapter we argue the reason why different existing models are not suitable for our purposes and present a model that we deem appropriate and discuss its limitations.

Portions of this chapter are based on [LR19]. These portions represent my contribution to the cited paper.

3.1 Learning Models

Blum, Kalai and Wasserman proposed, in [BKW00], an algorithm—widely known nowadays as BKW—that would turn into one of the best known solutions for LWE more than half a decade before the publication of Regev's work. In the mentioned paper, the authors give a subexponential time solution for the *Learning Parity with Noise* problem. In that context, the parity problem is thought as a boolean classifier—it returns 0 if the parity of the input bits is even and 1 otherwise. By doing so, they prove that two models of learning that have been previously proposed are not equivalent.

In [Val84], Valiant introduced the concept of *Probably Approximately Correct* (PAC) learning in the context of boolean classifiers. In short, a family $\mathcal{C}$ of boolean functions defined over $\{0,1\}^n$ is said to be (*efficiently*) PAC-*learnable* if there exists a (polynomial time) algorithm $\mathcal{A}$ such that, given a function $c \in \mathcal{C}$, a probability distribution χ over the domain, parameters $\varepsilon, \delta \in \left(0, \frac{1}{2}\right)$, and a set S_n of $m = \mathit{Poly}\left(\frac{1}{\varepsilon}, \frac{1}{\delta}, n\right)$ input samples x from the distribution χ such that $c(x) = 1$, the algorithm $\mathcal{A}$ outputs a function $h \in \mathcal{C}$ such that

$$\Pr_{S_n \leftarrow \chi^m}\Big(\Pr_{x \leftarrow \chi}\big(h(x) = c(x)\big) \geq 1 - \varepsilon\Big) \geq 1 - \delta.$$

A possible downside of Valiant's model is that it allows a learning algorithm to have access to a distinguishing oracle—an oracle that, for any function $c \in \mathcal{C}$, discerns whether or not a particular input x is such that $c(x) = 1$. This feature is noticeably difficult to reconcile when trying to model learning in the presence of noise.

The problem of establishing a model for learning in the presence of noise is addressed by a number of works [AL88, Lai88, Kea98]. In [AL88], Angluin and Laird extend Valiant's PAC model by allowing the learning algorithm to have access to a possibly faulty verification oracle. Kearns [Kea98] proposed a different approach, where instead of giving access to a faulty sampling oracle, it provides access to an oracle that accurately estimates the probability of the samples obtained from the given distribution χ to be faulty. A consequence of the BKW algorithm is that these two models of learning, [AL88] and [Kea98], are not equivalent.

3.2 Homomorphism Learning

In order to establish the appropriate learning model for our purposes we start with a simple example. Let V be a linear space over a field $\mathbb{F}$ of dimension n and consider $\mathcal{F} = \mathrm{Hom}_{\mathbb{F}}(V, \mathbb{F})$, the set of all linear functions from V to its field of scalars—also called *functionals*. Notice then that, by using the algebraic structure of V, it is possible to learn f given n samples $\big(\mathbf{v}_1, f(\mathbf{v}_1)\big), \ldots, \big(\mathbf{v}_n, f(\mathbf{v}_n)\big)$, provided that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent. This can be done by writing $\mathbf{v}$ in terms of $\mathbf{v}_1, \ldots, \mathbf{v}_n$—as a linear combination $\mathbf{v} = \sum_{i=1}^{n} b_i \mathbf{v}_i$—computing the inverse of the matrix whose columns are the vectors $\mathbf{v}_i$. Using the linearity of f we obtain $f(\mathbf{v}) = \sum_i b_i f(\mathbf{v}_i)$. Moreover, let $\mathbf{e_1}, \ldots, \mathbf{e}_n$ be the canonical basis and let $s_i = f(\mathbf{e}_i)$—this can be computed using Gaussian elimination. Thus, for $\mathbf{v} = (a_1, \ldots a_n)$ we can write $f(\mathbf{v}) = \sum_{i=1}^{n} a_i s_i = \langle \mathbf{s}, \mathbf{v} \rangle$, where $\mathbf{s} = (s_1, \ldots, s_n)$. This means that every $f \in \mathcal{F}$ can be expressed as an inner product by a constant vector $\mathbf{s}$, where $\mathbf{s}$ depends only on f and can be found efficiently.

In the case of morphisms between algebraic objects, the precise notion of "learning a morphism" is intrinsically dependent on the model used for the algebraic structures. For instance, assume that we know (the encodings of) a generating set $g_1, \ldots, g_m$ for a group G, as well as (the encodings of) their corresponding images $\varphi(g_1), \ldots, \varphi(g_m)$ under a morphism $\varphi\colon G \to H$. This information uniquely determines the morphism φ, as the value of $\varphi(g)$ for an element $g \in G$ is given by $\prod_{\mathbf{w}} \varphi(g_{w_i})$, where $g = \prod_{\mathbf{w}} g_{w_i}$. However, computing the word $\mathbf{w}$ may be a hard problem in the group G.

A *black-box semigroup* is a finitely generated semigroup S together with an injective encoding function $\mathrm{enc}\colon S \to \{0,1\}^*$, and an oracle $\mathcal{O}$ that returns the encoding result of operations in a predetermined operation set Π, where Π contains, at least, the group law. We say that an algorithm A has *black-box access* to a finitely generated semigroup $S = \langle s_1, \ldots, s_m \rangle$ if it has access to the list of encodings $\big\{\,\mathrm{enc}(s_1), \ldots, \mathrm{enc}(s_m)\big\}$ and input/output access to the oracle $\mathcal{O}$.

Definition 3.1 (Homomorphism Learning)**.** Let G and H be finitely generated semigroups and let $\varphi\colon G \to H$ be a homomorphism. Let ξ be a probability distribution over G. Suppose that an algorithm A has black-box access to G and H. We say that an algorithm A *learns* the function φ with respect to ξ from m samples $\big(g_i, \varphi(g_i)h_i\big)$, with $h_i \leftarrow \chi$, if given $g \leftarrow \xi\big(\langle g_1, \ldots, g_m \rangle\big)$, the algorithm A outputs $\varphi(g)$ with non-negligible probability, where $\xi(S)$ denotes the probability distribution ξ restricted to $S \leq G$.

Notice that in the case of noiseless samples, the learning problem reduces to the problem of finding an expression for g in terms of $g_1, \ldots, g_m$. This problem, in the case of semigroups, is called the *constructive semigroup membership* problem. In [CI14], Childs and Ivanyos proved that the generic constructive semigroup membership problem has an exponential quantum query lower bound.

Let G and H be groups. Notice that the set $\mathrm{Hom}(G, H)$ of homomorphisms $\varphi\colon G \to H$ is not empty, since the function that maps every element in G to the identity element in H is itself a homomorphism. In general, however, this set may contain several other elements. Let $\varphi \in \mathrm{Hom}(G, H)$ and let $g_1, \ldots, g_m \in G$.

In order to frame this as a computational problem, we shall assume that it is possible to efficiently sample from a probability distribution χ over G. For $\varphi \in \mathrm{Hom}(G, H)$ let $\Gamma^{\xi}_{\varphi,\chi}$ be the probability distribution over $G \times H$ obtained by sampling $g \in G$ according to ξ, $h \in H$ according to χ and outputing $\big(g, \varphi(g)h\big)$. If G is a finite group and ξ is the uniform distribution over G, we will omit ξ and denote $\Gamma^{\xi}_{\varphi,\chi}$ as $\Gamma_{\varphi,\chi}$. The problem of learning φ given samples from $\Gamma^{\xi}_{\varphi,\chi}$ is formally described in the following definition.

Definition 3.2 (Homomorphism Learning Problem)**.** Let G and H be finitely generated groups. Let ξ and χ be probability distributions over G and H, respectively. We say that an algorithm $\mathcal{A}$ solves the *learning homomorphism with noise problem* (LHN) for G, H, ξ and χ if for any $\varphi\colon G \to H$, $\mathcal{A}$ is able to learn φ with respect to ξ given a set of samples from the distribution $\Gamma^{\xi}_{\varphi,\chi}$ with non-negligible probability.

In the previous definition it is not required for the groups G and H to be finite, as this restriction would leave out several basic examples, such as the integers. Ideally, we would like to have the possibility to consider infinite groups for the distinguishing version of LHN. Nonetheless, the distinguishing versions of hard problems are usually about differentiating a particular distribution from uniform. To consider an infinite group, therefore, we need to replace the uniform distribution with a fixed distribution defined on the group, as the uniform distribution is not defined on infinite sets.

Definition 3.3 (Distinguishing Homomorphism Learning Problem)**.** Let G and H be finitely generated groups and fix a probability distribution Ξ over $G \times H$. Let ξ and χ be probability distributions over G and H, respectively. We say that an algorithm $\mathcal{A}$ solves the *distinguishing homomorphism with noise problem* (DHN) for G, H and ξ and χ with respect to Ξ if for any $\varphi\colon G \to H$, $\mathcal{A}$ is able to distinguish the distribution $\Gamma^{\chi}_{\varphi,\chi}$ from the distribution Ξ over $G \times H$.

3.3 Examples, Applications and Limitations

At the beginning of the previous section we argue that, given a functional $f\colon V \to \mathbb{F}$, it is enough to determine its value on the canonical basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ in order to find a vector $\mathbf{s}$ to express f as an inner product. This implies that an algorithm that is able to *learn* f given possibly noisy samples, in the sense of Definition 3.1, is also able to solve the Learning With Errors problem, in the sense of Definition 2.16. Other examples that appear in the literature are described next.

The Conjugacy Problem

Let $\langle s_1, \ldots, s_n \colon r_1, \ldots, r_m \rangle$ be a presentation of a group G. An instance of the *conjugacy (decision) problem* is a pair of elements $g_1, g_2 \in G$. The goal is to decide whether there exists $g \in G$ such that $g_1 = gg_2g^{-1}$. Given a pair of conjugate elements $g_1, g_2 \in G$—that is, a YES instance of the conjugacy problem—the *conjugacy search problem* is the problem of finding g such that $g_1 = gg_2g^{-1}$.

Remark 3.4. Recall that group conjugation is an automorphism. As a consequence, the conjugacy search problem is closely related the homomorphism learning problem. Nonetheless, there might not be a direct relation between them if we consider the model of learning that is given above in Definition 3.1. An algorithm that learns a conjugation, according to this definition, is able to evaluate the function $h \mapsto ghg^{-1}$. However, it is not clear how to use this algorithm to find g.

The hardness of the conjugacy problem naturally depends on the particular group. It is known to be undecidable in its generic form. On the other hand, it is known to have a polynomial time solution for certain families such as Coxeter groups [Kra94]. Nonetheless, several instances of this problem—and variants of this problem—have been used for the construction of Diffie-Hellman type key exchanges and signature schemes. The literature related to non-commutative cryptography is vast. Examples of this include [AAG99], which is one of the earliest construction of a cryptosystem using groups and proposes a key exchange based on the hardness of the conjugacy problem on braid groups; [KCCL02], which proposes a construction of a signature scheme using braid groups; [EK04], where the authors make use of polycyclic groups to construct a key exchange mechanism; and others [KLC+00, Sti05, BFX06].

Learning Isogenies

An isogeny between two elliptic curves $\phi\colon E_1 \to E_2$ is also a group homomorphism. The isogeny problem, as described in Definition 2.18, asks to find an isogeny between to isogenous curves. Naively, that would mean finding the corresponding polynomials that constitute the isogeny. However, in practice this is not the case. Algorithms such as Vélu's [Vél71] allows us to completely describe the isogeny only by finding a set of generators of its kernel. This yields a way to evaluate the isogeny on any point of E_1.

The converse is also true in the context of quantum algorithms; that is, given access to an algorithm that evaluates the isogeny, it is possible to recover a set of generators for its kernel. To see this notice that, since $\phi\colon E_1 \to E_2$ is a group homomorphism, it defines an instance for the Hidden Subgroup Problem. By using Shor's algorithm, it is possible to find a set of generators for the kernel of ϕ.

3.4 Generic Solutions to the Homomorphism Learning Problem

To finalize this chapter, we present two generic approaches to obtain information about instances of the learning problem to solve the Distinguishing Homomorphism Learning Problem (Definition 3.3). It is important to notice that the information obtained by both of these methods, in general, do not yield a solution to the learning problem according to Definition 3.1.

Order Finding Approach

Order Finding is the problem of finding the order of a group element, given oracle access to the group, where the allowed operations are the group law and inverse. This problem, to the best of our knowledge, is hard to solve classically. A quantum algorithm, however, can solve the order finding problem by using phase estimation [Mos99]. The solution for this problem is at the core of Shor's algorithm for factoring and solving discrete logarithm over $\mathbb{Z}_n$.

When trying to solve the Distinguishing Learning problem, the simplest case is when the samples have not been altered by random noise, in other words, where the input of the problem is a collection of samples of the form $\big(g, \varphi(g)\big)$. Let G and H be groups and let

$\varphi\colon G \to H$ be a homomorphism. By definition, $\varphi(e_G) = e_H$. Then, for $g \in G$, the order of g, $O(g)$, is bounded below by the order of $\varphi(g)$. Moreover, $O(g)$ is a multiple of $O\big(\varphi(g)\big)$. Hence, given a collection of samples $\big(g_i, \varphi(g_i)\big) \in G \times H$, an attacker can distinguish this distribution from $U(G \times H)$, by observing that the order of the left coordinate is always a multiple of the order of the right coordinate. Finding the order, in general, requires Shor's algorithm. However, this problem can be solved classically for several particular groups.

Noise in a Known Normal Subgroup

In [BFN+11], the authors remark that, for the distribution $\Gamma_{\varphi,\chi}$ to be indistinguishable from $U(G \times H)$, the support of φ should not be contained in a proper normal subgroup of H, as otherwise an attacker can "factor out" this subgroup, and obtain a noiseless distribution, on which the attacker can perform the order attack previously described to distinguish it from the uniform distribution. In more detail, let $N \trianglelefteq H$ be a normal subgroup of H containing the support of φ. Then the mapping

$$\bar{\varphi}\colon g \mapsto \varphi(g)N$$

is a homomorphism $\bar{\varphi}$ from G to the quotient group H/N. The distribution $\big(g, \bar{\varphi}(g)\big)$ is a noiseless distribution over $G \times H/N$.

Notice that in order to define $\bar{\varphi}$, and to be able to perform operations in the group H/N, it is necessary to know what the group N is. Therefore, performing this attack requires the knowledge of the normal subgroup on which the support of the noise is contained.

Chapter 4

An Algebraic Approach to LHN

> *"Siempre estoy haciendo lo que no puedo hacer, para poder aprender a hacerlo."*
>
> — Pablo Ruiz Picasso

This chapter is a continuation of Chapter 3, where we focus on the possible applications of the Learning Homomorphism with Noise problem to construct a public-key encryption scheme. As in the previous chapter, portions of this chapter are based on [LR19]. These portions represent my contribution to the cited paper.

4.1 Public-key cryptography from LHN

In 2011, Baumslag et al. proposed a generic framework for the study of the problem of learning noisy homomorphisms over abstract groups, using the word norm as their tool to measure noise. From a hard instance of this problem it is easy to derive a symmetric key encryption scheme. The idea is to share a homomorphism $\varphi\colon G \to H$ as the secret key, which allows one to recover $e\tau^{\mu}$ from the pair $\big(g, \varphi(g)e\tau^{\mu}\big)$. If τ is large and the noise is small, it is possible to distinguish whether μ is 0 or 1.

Deriving a public-key cryptosystem, however, is significantly more challenging. Using this problem in a way that is similar to the one described in [Reg05], requires the group to have certain properties. In generic language, the idea of Regev's cryptosystem is to randomly mix samples $\big(g_i, \varphi(g_i)e_i\big) \in G \times H$ from the public key to obtain a new sample

(g, h) whose distribution provides no information about the secret key φ. This allows us to encode a message μ in an element $\tau_\mu \in H$ by "hiding" it in the second coordinate as $(g, h\tau_\mu)$. To recover τ_μ it is enough to compute h from g and the secret key φ. However, h is formed by alternating multiplication of $\varphi(g_{w_i})$ and elements from the error distribution

$$h = \prod_{\mathbf{w}} \varphi(g_{w_i})e_{w_i} = \varphi(g_{w_1})e_{w_1}\varphi(g_{w_2})e_{w_2}\cdots\varphi(g_{w_\ell})e_{w_\ell}, \tag{4.1}$$

while g is only related to $g_{w_1}\cdots g_{w_\ell}$; in other words, the error elements are "in the way" of h.

One way to solve this problem is to use private information to erase the errors first. As a concrete example, let K be a group and let $\psi\colon H \to K$ be a second secret homomorphism, and assume that the error distribution over H efficiently samples elements $e \in \mathrm{Ker}(\psi)$. Hence we can erase the error elements by first applying ψ to h to obtain

$$\begin{aligned}\psi(h) &= \psi\big(\varphi(g_{w_1})\big)\psi(e_{w_1})\psi\big(\varphi(g_{w_2})\big)\psi(e_{w_2})\cdots\psi\big(\varphi(g_{w_\ell})\big)\psi(e_{w_\ell})\\ &= \psi\big(\varphi(g_{w_1})\big)\psi\big(\varphi(g_{w_2})\big)\cdots\psi\big(\varphi(g_{w_\ell})\big).\end{aligned}$$

Since φ and ψ are group homomorphisms, we may now recover the relation of the second coordinate with g by computing $\psi \circ \psi(g)$. This motivates the following definition.

Definition 4.1. Let G, H and K be groups and let $\varphi\colon G \to H$, $\psi\colon H \to K$ be group homomorphisms. Let χ be a probability distribution over H whose support is a subset of $\mathrm{Ker}(\psi)$. We say that an algorithm $\mathcal{A}$ solves the *normal-Learning Homomorphism with Noise* problem (normal-LHN) if $\mathcal{A}$ is able to learn φ from a set of samples from the distribution $\Gamma_{\varphi,\chi}$.

Notice that if the group H is Abelian—or, more generally, if the errors are sampled from the center of H—Equation (4.1) can be rewritten as

$$h = \prod_{\mathbf{w}} \varphi(g_{w_i}) \prod_{\mathbf{w}} e_{w_i}.$$

Nevertheless, this may lead to weaknesses in the construction. If the center $Z(H)$ of H is a proper subgroup, and the projection $H \mapsto H/Z(H)$ is efficiently computable, we may use the generic solution described in Section 3.4 when the noise is restricted to a known normal subgroup. This procedure does not provide additional information to an attacker when H is an Abelian group, since the projection onto the quotient yields a trivial distribution $(g, 1)$. However, in Section 4.4 we describe a more effective way to solve normal-LHN in this case.

4.2 A Public Key Cryptosystem based on Normal-LHN

In the previous section we argued the possible difficulties when using LHN to obtain cryptographic primitives, and motivated the definition of normal-LHN based on this discussion, with the possibility of arriving to a general procedure to construct a public-key cryptosystem form a generic group. In this section we describe this procedure. As with constructions based on LWE, we start by describing a symmetric encryption scheme that is later transformed into a public-key encryption scheme using the algebraic properties inherent to LHN. In Section 4.3 we describe two constructions using different algebraic objects: polynomial rings and elliptic curves. However, in Section 4.4, we argue why these constructions are insecure in the quantum setting.

Start by recalling that a subgroup $N \leq H$ is normal if and only if it is the kernel of a homomorphism from H. Consider three finitely generated groups G, H and K, and let ξ and χ be probability distributions over G and H respectively such that both distributions can be sampled efficiently.

A Symmetric-Key Construction

KeyGen(1^λ)**:** Given the security parameter λ, choose efficiently computable homomorphisms $\varphi\colon G \to H$ and $\psi\colon H \to K$, such that it is efficient to sample from χ restricted to $\mathrm{Ker}(\psi) \leq H$. Let $\tau \in H \setminus \mathrm{Ker}(\psi)$. The shared key is a description of φ and ψ, together with the group element τ.

Enc(β)**:** Given a message $\beta \in \{0,1\}$, sample an element g from G according to ξ and h from $\mathrm{Ker}(\psi) \leq H$ according to χ. The encryption of β is $\big(g, \varphi(g)h\tau^\beta\big)$

Dec(g, h')**:** Given a pair $(g,h') \in G \times H$, compute $\nu = \psi\big(\varphi(g)\big)^{-1} \cdot \psi(h')$ and output

$$\beta' = \begin{cases} 0 & \text{if } \nu = 1_K, \\ 1 & \text{if } \nu \neq 1_K. \end{cases}$$

Correctness. Suppose that (g,h') is a correctly formed encryption of $\beta \in \{0,1\}$. Then the intermediate step of the decryption algorithm computes

$$\begin{array}{rcl}\nu & = & \psi\big(\varphi(g)\big)^{-1}\cdot\psi(h') \\ & = & \psi\big(\varphi(g)\big)^{-1}\cdot\psi\big(\varphi(g)h\tau^{\beta}\big) \\ & = & \psi\big(\varphi(g)\big)^{-1}\cdot\psi\big(\varphi(g)\big)\cdot\psi(h)\cdot\psi(\tau)^{\beta} \\ & = & \psi(\tau)^{\beta}.\end{array}$$

The correctness then follows since τ is not in the kernel of ψ. □

A public-key construction

KeyGen(1^{λ})**:** Given the security parameter λ, choose efficiently computable homomorphisms $\varphi\colon G\to H$ and $\psi\colon H\to K$. For $i\in\{1,\dots,m\}$ compute

$$\big(g_i,\varphi(g_i)h_i\big)\in G\times H,$$

where g_i is sampled from ξ and h_i is sampled from $\mathrm{Ker}(\psi)\leq H$ according to χ. The private key is a description of φ and ψ. The public key is the set

$$\Big\{\big(g_i,\varphi(g_i)h_i\big)\colon i=1,\dots,m\Big\}\subseteq G\times H,$$

together with a public element $\tau\in H\setminus\mathrm{Ker}(\psi)$.

Enc(β)**:** Given a message $\beta\in\{0,1\}$, sample a word $\omega=w_1\cdots w_\ell$ over the indices $\{1,\dots,m\}$ of length ℓ and compute

$$(g,h')=\left(\prod_{i=1}^{\ell}g_{w_i},\prod_{i=1}^{\ell}\varphi(g_{w_i})h_{w_i}\right).$$

Then output $(g,h'\tau^{\beta})$.

Dec(g,h)**:** Run the decryption procedure described in Subsection 4.2.

Correctness. Suppose that (g,h) is a correctly formed encryption of $\beta\in\{0,1\}$. Then the intermediate step of the decryption algorithm computes

$$\begin{array}{rcl}\nu & = & \psi\big(\varphi(g)\big)^{-1}\cdot\psi(h) \\ & = & \psi\left(\varphi\left(\prod_{i=1}^{\ell}g_{w_i}\right)\right)^{-1}\cdot\psi\left(\left(\prod_{i=1}^{\ell}\varphi(g_{w_i})h_{w_i}\right)\cdot\tau^{\beta}\right) \\ & = & \psi\left(\varphi\left(\prod_{i=1}^{\ell}g_{w_i}\right)\right)^{-1}\cdot\left(\prod_{i=1}^{\ell}\psi\big(\varphi(g_{w_i})\big)\psi(h_{w_i})\right)\cdot\psi(\tau)^{\beta} \\ & = & \psi\left(\varphi\left(\prod_{i=1}^{\ell}g_{w_i}\right)\right)^{-1}\cdot\psi\left(\varphi\left(\prod_{i=1}^{\ell}g_{w_i}\right)\right)\cdot\psi(\tau)^{\beta} \\ & = & \psi(\tau)^{\beta}.\end{array}$$

The correctness then follows since τ is not in the kernel of ψ. $\square$

Properties

Despite being inspired by the traditional LWE cryptosystem, there are several differences between this and the construction described in the previous subsection that may yield different useful properties, as well as different lines of cryptanalysis.

Noise Accumulation and Decryption Errors. Due to the geometric nature of LWE, it is necessary to be careful when handling the noise. Large noise yields decryption errors, which in turn give way to key recovery attacks. Noise may accumulate during encryption, making decryption errors difficult to mitigate—unless an error correcting code is implemented alongside. Moreover, noise accumulation has been the main obstacle for the design of effective homomorphic cryptosystems based on lattices, making necessary the use of bootstrapping to achieve unbounded depth fully-homomorphic encryption.

A cryptosystem built as in the previous subsection does not suffer from noise accumulation or decryption errors. Elements sampled from the noise distribution χ are all contained in the kernel of the secret homomorphism ψ.

Unbounded Homomorphic. Suppose that H is a group with non-trivial center Z, and assume that τ is a non-trivial central element of H of order 2 in the set $H \setminus \mathrm{Ker}(\psi)$. Then $\psi(\tau)$ is also a non-trivial central element in the image of ψ. Let β, β' be two messages and (g, h), (g', h') their corresponding encryptions. Then

$$\begin{aligned} hh' &= \big(\textstyle\prod_{\mathbf{w}} \varphi(g_{w_i})h_{w_i}\big)\tau^{\beta} \big(\textstyle\prod_{\mathbf{w}} \varphi(g_{w'_i})h_{w'_i}\big)\, \tau^{\beta'} \\ &= \big(\textstyle\prod_{\mathbf{w}} \varphi(g_{w_i})h_{w_i}\big) \cdot \big(\textstyle\prod_{\mathbf{w}} \varphi(g_{w'_i})h_{w'_i}\big) \cdot \tau^{\beta} \cdot \tau^{\beta'} \\ &= \Big(\big(\textstyle\prod_{\mathbf{w}} \varphi(g_{w_i})h_{w_i}\big) \cdot \big(\textstyle\prod_{\mathbf{w}} \varphi(g_{w'_i})h_{w'_i}\big)\Big) \cdot \tau^{\beta'+\beta'}. \end{aligned}$$

It follows that the coordinate-wise product $(g, h) \cdot (g', h') = (gg', hh')$ is a valid encryption of $\beta + \beta'$.

It is worth noticing that this property itself does not imply that the construction is unbounded *fully homomorphic*. In other words, the construction is only able to evaluate a single operation, since it is based on a group. We conjecture that this is not a sufficient condition to evaluate any complete set of gates.

(Potentially) Small Keys. The encryption mechanism used in traditional LWE mixes elements of the public key by taking a random linear combination of them, where the coefficients are in $\{0, 1\}$. Such a restriction is necessary in order to keep the noise small. This is, however, not necessary in this case since noise accumulation does not induce decryption errors. In particular, the number of possible linear combinations of elements $g_1, \dots, g_m$ of an Abelian group increases according to their order. In the case of non-Abelian groups, however, the number of combinations obtained—words in the set $S = \{g_1, \dots g_m\}$—is strictly greater, and depends on the relations that hold for the set S.

(Potentially) Large Message Space. Suppose that a central element $\tau \in Z(G)$ is such that the discrete logarithm can be solved efficiently in the group generated by $\psi(\tau)$. Then there is a way to modify the decryption procedure in 4.2 to increase the size of the message space. In particular this is true whenever the discrete logarithm is solvable in K. This allows for the message space to be of size $O\big(\psi(\tau)\big)$. Notice, however, that this depends on ψ, which is part of the secret key.

4.3 Obtaining instances

In the previous section we described a way to obtain public-key encryption from the normal-LHN problem over a generic group. However, the feasibility of the construction, as well as the security of it, depend on the specific group that is chosen to instantiate it. In this case, the chosen groups G, H and K, the homomorphisms $\varphi\colon G \to H$, $\psi\colon H \to K$, and the corresponding probability distributions must have certain desired properties.

Large key space. The groups $\mathrm{Hom}(G, H)$ and $\mathrm{Hom}(H, K)$ must be of exponential size in the security parameter.

Feasibility. There is an efficient algorithm to sample from the distribution χ. Since the support of χ (the set of elements where χ is non-zero) must be contained in the kernel of ψ, there must be an efficient algorithm to sample from $\mathrm{Ker}(\psi)$.

One way to ensure that the first condition is satisfied is to choose a group G with a large number of normal subgroups, which holds trivially for Abelian groups. In this section we present two instances of the construction described in Section 4.1 using Abelian groups. We remark that both constructions are vulnerable to the attacks described in Section 4.4, moreover, the attack to the first example, the instance using polynomials, does not require the use of a quantum algorithm, rendering the scheme completely insecure, as proved in Section 4.4. The second condition is slightly more difficult to guarantee since the difficulty of finding the kernel of a homomorphism depends on the way that the homomorphism is described, and this, in general, might be a difficult task. In the following constructions this problem is addressed by describing the homomorphisms through the description of their corresponding kernels.

A polynomial ring instance

Let $\mathbb{F}$ be a finite field and let $f(x) \in \mathbb{F}[x]$ be a polynomial of degree n. For $g(x) \in \mathbb{F}[x]$ let $[g(x)]$ denote the coset in $R = \mathbb{F}[x]/f(x)$ containing $g(x)$, and let $\overline{g}(x)$ denote the residue of $g(x)$ divided by $f(x)$. Notice that $\overline{g}(x)$ is the unique polynomial of degree less than n in the coset $\big[g(x)\big]$. We have that for every $\alpha \in \mathbb{F}$, the function

$$\psi\colon [g(x)] \mapsto \overline{g}(\alpha)$$

is a group homomorphism from the additive group of $R = \mathbb{F}[x]/f(x)$ to the additive group of $\mathbb{F}$. Notice that this is not a ring homomorphism. The kernel of this homomorphism can be described by the set of polynomials in $\mathbb{F}[x]$ of degree less than n that have α as a root:

$$\begin{aligned} \mathrm{Ker}(\psi) &= \big\{[g(x)]\colon g(\alpha) = 0, \deg(g) < n\big\} \\ &= \big\{[(x-\alpha)p(x)]\colon \deg(p) < n-1\big\}. \end{aligned}$$

If $\mathbb{F}$ is a finite field, the previous description yields an efficient procedure to sample from the uniform distribution over $\mathrm{Ker}(\psi)$, by sampling uniformly a polynomial $p(x)$ of degree less than $n-1$ and returning $(x-\alpha)p(x)$.

KeyGen(1^λ)**:** Pick a polynomial $f(x) \in \mathbb{F}[x]$. Choose $\alpha, s_0, \dots, s_{n-1}$ from the uniform distribution over $\mathbb{F}$ and let $s(x) = \sum_{j=0}^{n-1} s_j x^j$. For $i \in \{1, \dots, m\}$ choose $a_i(x)$ uniformly from R and $p(x)$ uniformly from the set of polynomials in $\mathbb{F}[x]$ of degree less than $n-1$. Compute

$$b_i(x) = a_i(x)s(x) + p_i(x)(x-\alpha).$$

The private key is the pair $\big(s(x), \alpha\big)$. The public key is the set of pairs $\big(a_i(x), b_i(x)\big)$.

Enc(μ)**:** Given a message μ, encode it as an element of the field $\mathbb{F}$. Choose a random subset $J \subseteq \{1, \dots, m\}$ and compute the ciphertext

$$\big(a(x), b(x)\big) = \left(\sum_{i\in J} a_i(x), \sum_{i\in J} b_i(x) + \mu\right) \in \mathbb{F}[x] \times \mathbb{F}[x].$$

Dec $\big(a(x), b(x)\big)$**:** Compute $d(x) = b(x) - a(x)s(x)$ and output $\mu' = d(\alpha)$.

Isogeny LWE

Keeping the kernel of a homomorphism secret is the main idea behind isogeny-based cryptography. The isogeny problem is the problem of computing an isogeny between two curves E_1, E_2 just by knowing the equations that describe the curves, provided that this isogeny exists (that the curves are *isogenous*). In constructions such as SIKE [JAC+17], it is assumed that this problem remains hard even after giving away the image of two points in the curve (specifically the generators of the 2 or 3 torsion subgroup of E_1).

Let p be a prime number and $\mathbb{F}_{p^2}$ be the field with p^2 elements.

KeyGen(1^λ)**:** For simplicity we divide this section into the isogeny generation and the point generation.

- **Isogenies:** Choose $k_1, k_2 \in \mathbb{Z}_3^n$ uniformly at random. Set $G_0 = [k_1]R_0 + [k_2]S_0 \in E_0[3^n]$, and find a point $H_0 \in E_0[3^n]$ which is independent from G_0. Use G_0 to compute the isogeny $\phi : E_0 \to E_1$ with $\mathrm{Ker}(\phi) = \langle G_0 \rangle$, along with $\phi(H_0)$.

  Next, compute a basis R_1, S_1 for $E_1[3^n]$. Choose $k_3, k_4 \in \mathbb{Z}_3^n$ uniformly at random. Set $G_1 = [k_1]R_1 + [k_2]S_1 \in E_1[3^n]$, and test that G_1 is independent from $\phi(H_0)$. Otherwise choose k_3 and k_4 again and repeat the previous line. Once G_1 and $\phi(H_0)$ are independent, compute the isogeny use G_1 to compute the isogeny $\psi : E_1 \to E_2$ with $\ker(\phi) = \langle G_1 \rangle$.

- **Points:** Construct points P_1, $Q_1 \in E_1[2^n]$ such that $\langle P_1, Q_1 \rangle = E_1[2^n]$. Choose $2m$ points at random:

  $$X_1, \ldots, X_m \in_R E_0(\mathbb{F}_{p^2}),$$

  $$Y_1, \ldots, Y_m \in_R \mathrm{Ker}(\psi) \subseteq E_1[3^n].$$

  For each $i \in \{1, \ldots m\}$ compute the image of $X_1, \ldots, X_m$ under ϕ. The public key is P_1, Q_1 and the tuples $\big(X_i,\ \phi(X_i) + Y_i\big)$, for $i \in \{1, \ldots m\}$. The private key is $k_1, k_2, k_3, k_4 \in \mathbb{Z}_3^n$ and $\psi(P_1)$, $\psi(Q_1)$.

Enc(μ)**:** Encode the message μ into $(M_1,\ M_2) \in (\mathbb{Z}_2^m)^2$, where not both M_1 and M_2 are divisible by 2. Choose a random subset $J \subseteq \{1, \ldots, t\}$ and compute the ciphertext:

$$(X, Y) = \left(\sum_{i \in J} X_i, \left(\sum_{i \in J} \phi(X_i) + Y_i \right) + [M_1]P_1 + [M_2]Q_1 \right) \in E_0(\mathbb{F}_{p^2}) \times E_1(\mathbb{F}_{p^2}).$$

Dec(X, Y)**:** Given a ciphertext $(X, Y) \in E_0(\mathbb{F}_{p^2}) \times E_1(\mathbb{F}_{p^2})$, compute

$$Z = \psi\big(Y - \phi(X)\big).$$

Using the knowledge of $\psi(P_1),\ \psi(Q_1)$ solve the two dimensional elliptic curve discrete logarithm problem:

$$Z = [M_1']\psi(P_1) + [M_2']\psi(Q_1)$$

and recover the message M from $(M_1',\ M_2')$.

4.4 Solving Normal-LHN for Abelian Groups

In this section we prove the impossibility of constructing a quantum-resistant cryptosystem based on the hardness of normal-LHN for Abelian groups. As a warm-up, we start by recalling the standard way to reduce LWE to SIS. Suppose that we are given m LWE samples $\big(\mathbf{a}_i, b_i = \langle \mathbf{a}_i, \mathbf{s}\rangle + e_i\big)$. Finding the secret $\mathbf{s}$ is equivalent to solving the equation

$$A\mathbf{s} + \mathbf{e} = \mathbf{b},$$

where the matrix A and the vectors $\mathbf{e}$ and $\mathbf{b}$ are formed with the entries of the samples. To solve this one may try to "get rid of the action of $\mathbf{s}$" by computing a vector $\mathbf{t}$ in the null-space of A^T and multiplying $\mathbf{b}$ by $\mathbf{t}^T$. This way we obtain

$$\mathbf{t}^T\mathbf{b} = \mathbf{t}^T A\mathbf{s} + \mathbf{t}^T\mathbf{e} = \mathbf{t}^T\mathbf{e}.$$

When $\mathbf{t}$ is a small vector, the product $\mathbf{t}^T\mathbf{e}$ is also small. Hence it is possible to solve the decisional version of LWE.

The previous idea can also be used to solve LHN in the case of Abelian groups. When the number of samples in the public key exceeds the rank of the group it is possible to mount a key-recovery attack from the public key. In the other case, when the number of samples that constitute the public-key is less than or equal to the rank of the group, it is possible to recover a message from any encryption of it. Observe that any group homomorphism is constant on the cosets of its kernel; hence a group homomorphism is a hiding function of its kernel.

Secret key recovery

Let G, H, K be Abelian groups (denoted additively) and let $\varphi\colon G \to H$ and $\psi\colon H \to K$ be two secret homomorphisms. Let ℓ be the rank of G and suppose that we are given $m > \ell$ samples of the form

$$\big(g_i, \varphi(g_i) + h_i\big) \in G \times H$$

with $h_i \in \mathrm{Ker}(\psi)$. Now consider the map $f\colon \mathbb{Z}^m \to G$ given by

$$f\colon (a_1, \dots, a_m) \mapsto \sum_{i=1}^{m} a_i g_i \in G.$$

This map is a group homomorphism. Using Shor's algorithm it is possible to find a generating set for the kernel of f. If $(a_1, \dots, a_m) \in \mathrm{Ker}(f)$, we have that

$$\sum_{i=1}^{m} a_i\big(g_i, \varphi(g_i) + h_i\big) = \left(0, \sum_{i=1}^{m} a_i h_i\right) \in \{0\} \times \ker(\psi),$$

obtaining a random element in $\ker\psi$. By repeating this process we can obtain a generating set of $\ker\psi$.

Message recovery

Suppose that we have the same setup as before, but this time $m \leq \ell$. Let $\Big\{\big(g_i, \varphi(g_i) + h_i\big)\colon i = 1, \dots, m\Big\}$ be the public key and let $(g, h) = \sum_{i=1}^{m} r_i\big(g_i, \varphi(g_i) + h_i\big) + (0, \beta\tau)$ be an encryption of β. Consider the function $f\colon \mathbb{Z}^{m+1} \to G$ given by

$$f\colon (a_1, \dots, a_m, a_{m+1}) \mapsto -a_{m+1}g + \sum_{i=1}^{m} a_i g_i.$$

As before, this is a group homomorphism. Using Shor's algorithm it is possible to find a generating set for the kernel of f. Moreover, this has rank one and is generated by the tuple $(r_1, \dots, r_m, 1)$. Using these recovered coefficients and the public key, it is possible to recover $\beta\tau$ from the given ciphertext.

4.5 Conclusion

In this chapter we proposed a generic adaptation of the construction proposed by Regev in [Reg05] of a public-key cryptosystem that uses the hardness of solving a noisy learning problem. Extending previous works in the area such as [BBFR08, BFN$^+$11, BFX06], here we proposed the first construction of this kind that is able to make use of non-commutative groups, as long as these satisfy certain properties. We showed, however, that instantiating this construction using Abelian groups results in an insecure cryptosystem, when the attacker has access to a quantum computer.

There are several questions that remain unanswered. Finding an non-Abelian instantiation of this construction is a work in progress, as well as the case of evaluating the security of the Abelian instantiation in the classical setting—if no attack is found this would yield a unbounded (additively) homomorphic construction.

Chapter 5

A Generalized Notion of the Smoothing Parameter

"(. . .) un auteur ne nuit jamais tant à ses lecteurs que quand il dissimule une difficulté."

— Évariste Galois

We now turn our attention to certain geometric aspects of LWE and the lattice problems associated to it. Specifically, in this chapter we make an exploration of the smoothing parameter from a mathematical point of view. The smoothing parameter has been a fundamental concept in the study of lattice theory and the development of lattice based cryptography. Roughly speaking, among other things, it allows one to conceal the instance of the lattice problem that is encoded in the instance of an LWE problem. Hence it is a component that unlocks the possibility of an average-case to worst-case reduction from LWE to lattice problems.

Intuitively speaking, the smoothing parameter can be thought of as the minimum stretching of the Gaussian $\rho_{s,\mathbf{c}}$ over $\mathbb{R}^n$ that hides the discreteness of a lattice $\mathcal{L} \subset \mathbb{R}^n$ when centered at each $\mathbf{v} \in \mathcal{L}$. Under this understanding, the smoothing parameter is then a quantity associated to the lattice. Nonetheless, it is perhaps expected that, for some other functions, it is also possible to find an adequate stretching that accomplishes the same goal. Thus a natural question is "is it necessary to consider a Gaussian noise for LWE to obtain an average-case to worst-case reduction, or can we use other noise distributions?".

The main result of this chapter is the following.

Theorem 5.1 (Main (informal))**.** *There exists an infinite family of functions that admit polynomially large smoothing parameters for every lattice.*

The analysis of the possible applications, such as its use in LWE and BDD, is left for Chapter 6.

5.1 Analytic Background

Notation

Throughout this and the next chapter, we use boldface letters to denote multivariate functions. The reason for this is to differentiate between single and multivariate functions. Several properties that are given for functions over $\mathbb{R}$ do not hold—or are difficult to prove—for functions over $\mathbb{R}^n$.

Given a function $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}$, a vector $\mathbf{c} \in \mathbb{R}^n$ and a scalar $s \in \mathbb{R}$ let $\mathbf{f}_{s,\mathbf{c}}$ denote the function

$$\mathbf{f}_{s,\mathbf{c}}\colon \mathbf{x} \mapsto \mathbf{f}\left(\frac{1}{s}(\mathbf{x}-\mathbf{c})\right).$$

When the shift $\mathbf{c} = \mathbf{0}$, we denote $\mathbf{f}_{s,\mathbf{0}}$ as $\mathbf{f}_s$.

Smooth and Integrable Spaces

Given $k \in \mathbb{Z}_{\geq 0}$, let C^k denote the set of functions $f\colon \mathbb{R} \to \mathbb{R}$ such that the all the derivatives $f = f^{(0)}, f^{(1)}, \ldots, f^{(k)}$ exists and are continuous. In particular, C^0 denotes the set of real valued continuous functions over $\mathbb{R}$.

For $p \in \mathbb{R}_{>0}$, let $L^p(\mathbb{R}^n)$ be the set of functions $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}$ such that the value

$$\|\mathbf{f}\|_p := \left(\int_{\mathbb{R}^n} \left|\mathbf{f}(\mathbf{x})\right|^p \, d\mathbf{x}\right)^{1/p}$$

exists and is finite. A function $\mathbf{f} \in L^1(\mathbb{R}^n)$ is said to be *integrable* over $\mathbb{R}^n$. We include the case $p = \infty$ by defining the ∞-*norm* as

$$\|\mathbf{f}\|_\infty := \inf\left\{b \in \mathbb{R}\colon \text{for almost all } \mathbf{x} \in \mathbb{R}^n, \mathbf{f}(\mathbf{x}) \leq b\right\}.^{1}$$

[1] The quantifiers *for almost all* and *almost everywhere* are formally defined as "for all except for a set of measure zero".

The L^∞ space thus consists of the functions that are bounded almost everywhere. In the literature, these functions are also known as *essentially bounded.*

Fourier Transform

The Fourier transform is central in the development of lattice theory. In particular, it is a fundamental tool for the analysis of probability distributions over $\mathbb{R}^n$—or any measurable space.

Definition 5.2. Let $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{C}$ be an integrable function. The *Fourier transform* of $\mathbf{f}$ is the function $\widehat{\mathbf{f}}\colon \mathbb{R}^n \to \mathbb{C}$ defined as

$$\widehat{\mathbf{f}}(\mathbf{y}) \coloneqq \int_{\mathbb{R}^n} \mathbf{f}(\mathbf{x}) e^{-2\pi i \langle \mathbf{x}, \mathbf{y} \rangle} d\mathbf{x}. \tag{5.1}$$

For $\mathbf{f} \in L^\infty(\mathbb{R}^n)$ let $\check{\mathbf{f}}$ denote the *inverse Fourier transform* of $\mathbf{f}$. This is given by

$$\check{\mathbf{f}}(\mathbf{y}) \coloneqq \int_{\mathbb{R}^n} \mathbf{f}(\mathbf{x}) e^{2\pi i \langle \mathbf{x}, \mathbf{y} \rangle} d\mathbf{x}. \tag{5.2}$$

It is possible to obtain a relation between the 1-norm of a function and the ∞-norm of its Fourier transform, which is given by the Hausdorff-Young inequality. This results in the following relation.

$$\int_{\mathbb{R}^n} \mathbf{f}(\mathbf{x}) d\mathbf{x} = \widehat{\mathbf{f}}(\mathbf{0}) \leq \|\widehat{\mathbf{f}}\|_\infty \leq \|\mathbf{f}\|_1 = \int_{\mathbb{R}^n} \big|\mathbf{f}(\mathbf{x})\big| d\mathbf{x}.$$

In particular we obtain an equality whenever $\mathbf{f}$ is a non-negative function.

Proposition 5.3. *Let $\mathbf{f} \in L^1(\mathbb{R}^n)$ and let $s \in \mathbb{R}_{\neq 0}$ and $\mathbf{c} \in \mathbb{R}^n$. Then the Fourier transform of $\mathbf{f}_{s,\mathbf{c}}$ is given by*

$$\widehat{\mathbf{f}_{s,\mathbf{c}}}(\mathbf{y}) = s^n e^{-2\pi i \langle \mathbf{c}, \mathbf{y} \rangle} \widehat{\mathbf{f}}\left(s\mathbf{y}\right).$$

Proof. Let $s \in \mathbb{R}_{\neq 0}$ and $\mathbf{c} \in \mathbb{R}^n$. The Fourier transform of $\mathbf{f}_{s,\mathbf{c}}$ is given by

$$\begin{aligned}
\widehat{\mathbf{f}_{s,\mathbf{c}}}(\mathbf{y}) &= \int_{\mathbb{R}^n} \mathbf{f}_{s,\mathbf{c}}(\mathbf{x}) e^{-2\pi i \langle \mathbf{x},\mathbf{y} \rangle} d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \mathbf{f}\left(\frac{1}{s}(\mathbf{x}-\mathbf{c})\right) e^{-2\pi i \langle \mathbf{x},\mathbf{y} \rangle} d\mathbf{x} \\
&= \int_{\mathbb{R}^n} s^n \mathbf{f}(\mathbf{z}) \exp\Big(-2\pi i \big\langle s\mathbf{z}+\mathbf{c}, \mathbf{y} \big\rangle\Big) d\mathbf{z} \\
&= e^{-2\pi i \langle \mathbf{c},\mathbf{y} \rangle} s^n \int_{\mathbb{R}^n} \mathbf{f}(\mathbf{z}) e^{-2\pi i \langle \mathbf{z}, s\mathbf{y} \rangle} d\mathbf{z} \\
&= s^n e^{-2\pi i \langle \mathbf{c},\mathbf{y} \rangle} \widehat{\mathbf{f}}\left(s\mathbf{y}\right).
\end{aligned}$$

Notice that the change of variables in the third equality, $\mathbf{z} = \frac{1}{s}(\mathbf{x}-\mathbf{c})$, yields the differential $d\mathbf{x} = s^n d\mathbf{z}$. This observation finishes the proof. □

The Schwartz Space

For $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geq 0}$ let ∂^α denote the operator $\partial^{\alpha_1} \cdots \partial^{\alpha_n}$, and given $\mathbf{x} \in \mathbb{R}^n$ let $\mathbf{x}^\alpha$ be the polynomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. Let $C^\infty(\mathbb{R}^n)$ denote the set of functions $\mathbf{f} \colon \mathbb{R}^n \to \mathbb{R}$ such that for all $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}_{\geq 0}$, the derivative $\partial^\alpha \mathbf{f}$ exists and is continuous. For $\mathbf{f} \in C^\infty(\mathbb{R}^n)$ and $\alpha, \beta \in \mathbb{Z}_{\geq 0}$ let

$$\|\mathbf{f}\|_{\alpha,\beta} = \sup_{\mathbf{x} \in \mathbb{R}^n} \big|\mathbf{x}^\alpha \partial^\beta \mathbf{f}(\mathbf{x})\big|. \tag{5.3}$$

Definition 5.4 (Schwartz Function)**.** A function $\mathbf{f} \colon \mathbb{R}^n \to \mathbb{C}$ is a *Schwartz function* if for all $\alpha, \beta \in \mathbb{Z}^n_{\geq 0}$, $\|\mathbf{f}\|_{\alpha,\beta}$ is finite. The set of Schwartz functions defined over $\mathbb{R}^n$ is denoted by $\mathcal{S}(\mathbb{R}^n)$.

Any Schwartz function $\mathbf{f}$ is eventually bounded by the inverse of any polynomial. As a consequence, any Schwartz function is integrable; thus the Fourier transform of any element in $\mathcal{S}(\mathbb{R}^n)$ is defined. Moreover, the set of Schwartz functions over $\mathbb{R}^n$ has the following properties.

Proposition 5.5. *Let* $\mathbf{f}, \mathbf{g} \in \mathcal{S}(\mathbb{R}^n)$*. Then the functions*

- $\mathbf{f} + \mathbf{g} \colon \mathbf{x} \mapsto \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})$,
- $\mathbf{f} \cdot \mathbf{g} \colon \mathbf{x} \mapsto \mathbf{f}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x})$,
- $\widehat{\mathbf{f}} \colon \mathbf{y} \mapsto \int_{\mathbb{R}^n} \mathbf{f}(\mathbf{x}) e^{-2\pi i \langle \mathbf{x}, \mathbf{y} \rangle} d\mathbf{x}$ *and*
- $\mathbf{f} * \mathbf{g} \colon \mathbf{x} \mapsto \int_{\mathbb{R}^n} \mathbf{f}(\mathbf{z}) \mathbf{g}(\mathbf{x} - \mathbf{z}) d\mathbf{z}$

all belong to $\mathcal{S}(\mathbb{R}^n)$*.*

For this reason $\mathcal{S}(\mathbb{R}^n)$ is commonly referred as the *Schwartz space* over $\mathbb{R}^n$. The Gaussian function $\rho(\mathbf{x}) = e^{-x^2}$ is a common example of a Schwartz function. Other common examples are the functions $\mathbf{f} \in C^\infty(\mathbb{R}^n)$ with compact support, which are better known as *bump functions*.

The Poisson Summation Formula

Another property satisfied by any Schwartz function is the following relation between it and its dual with respect to any given lattice.

Lemma 5.6 (Poisson Summation Formula)**.** *Let* $\mathbf{f} \colon \mathbb{R}^n \to \mathbb{R}$ *be a Schwartz function. Then for any lattice* $\mathcal{L} \subset \mathbb{R}^n$*,*

$$\mathbf{f}(\mathcal{L}) = \det(\mathcal{L}^*) \widehat{\mathbf{f}}(\mathcal{L}^*). \tag{5.4}$$

It is known that this result can be extended to a larger family of functions. In this and the following Chapter we denote by $\mathcal{D}_n \subset L^1(\mathbb{R}^n)$ the set of functions such that, for every lattice $\mathcal{L} \subset \mathbb{R}^n$, Equation (5.4) holds and both, $\mathbf{f}(\mathcal{L})$ and $\widehat{\mathbf{f}}(\mathcal{L}^*)$ are absolutely convergent. By [MSD19, Theorem A.1], $\mathcal{D}_n$ contains every continuous function $\mathbf{f}$ such that

1. there exists $\delta \in \mathbb{R}_{>0}$ such that $\mathbf{f}(\mathbf{x}) = O\left(1 + \|\mathbf{x}\|_2^{-(n+\delta)}\right)$ and
2. for every lattice $\mathcal{L} \subset \mathbb{R}^n$, the series $\sum_{\mathbf{x} \in \mathcal{L}} \left|\widehat{\mathbf{f}}(\mathbf{x})\right|$ converges.

In turn, this is a relaxation of a well known set of conditions that appear in the literature. See, for instance, [SW71, Chapter VII, Corollary 2.6].[2]

Combining Equation (5.4) with Proposition 5.3 we obtain the slightly more general expression,

$$\sum_{\mathbf{x} \in k\mathcal{L}} \mathbf{f}(\mathbf{x} + \mathbf{c}) = \frac{\det \mathcal{L}^*}{k} \sum_{\mathbf{x} \in \frac{1}{k}\mathcal{L}^*} \widehat{\mathbf{f}}(\mathbf{x}) e^{-2\pi i \langle \mathbf{c}, \mathbf{x} \rangle}. \tag{5.5}$$

[2] Several other works have presented other generalizations of the Poisson Summation Formula—generalizing to other families and generalizing Equation (5.4) itself. See, for example [BZ97, NU15].

A consequence of the previous lemma is that the weight of a function over a lattice coset $\mathcal{L}+\mathbf{c}$ is maximized at $\mathbf{c}=\mathbf{0}$, whenever the Fourier transform of the function is non-negative over the dual lattice.

Proposition 5.7. *Let $\mathbf{f} \in \mathcal{S}(\mathbb{R}^n)$ and let $\mathcal{L} \subset \mathbb{R}^n$ be a lattice. Suppose that for every $\mathbf{y} \in \mathcal{L}^*$, $\widehat{\mathbf{f}}(\mathbf{y}) \geq 0$. Then for any $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{f}(\mathcal{L}-\mathbf{c}) \leq \mathbf{f}(\mathcal{L})$.*

Proof. Let $\mathbf{c} \in \mathbb{R}^n$. Then, by Proposition 5.3 and Lemma 5.6,

$$\begin{aligned}
\mathbf{f}(\mathcal{L}-\mathbf{c}) &= \mathbf{f}_{\mathbf{c}}(\mathcal{L}) \\
&= \sum_{\mathbf{y}\in\mathcal{L}^*} \widehat{\mathbf{f}_{\mathbf{c}}}(\mathbf{y}) \\
&= \sum_{\mathbf{y}\in\mathcal{L}^*} e^{2\pi i\langle \mathbf{c},\mathbf{y}\rangle}\widehat{\mathbf{f}}(\mathbf{y}) \\
&\leq \sum_{\mathbf{y}\in\mathcal{L}^*} \left|e^{2\pi i\langle \mathbf{c},\mathbf{y}\rangle}\right|\widehat{\mathbf{f}}(\mathbf{y}) \\
&\leq \sum_{\mathbf{y}\in\mathcal{L}^*} \widehat{\mathbf{f}}(\mathbf{y}) \\
&= \mathbf{f}(\mathcal{L}),
\end{aligned}$$

as required. □

Normalization and Standard Form

Definition 5.8 (Normalization and Standard Form)**.** We say that a function $\mathbf{f} \in L^1(\mathbb{R}^n)$ is *normalized* if $\int_{\mathbb{R}^n} \mathbf{f}(\mathbf{x})d\mathbf{x} = \widehat{\mathbf{f}}(\mathbf{0}) = 1$. The function $\mathbf{f}$ is said to be in *standard form* if $\mathbf{f}(\mathbf{0}) = \widehat{\mathbf{f}}(\mathbf{0}) = 1$.

Given a function $\mathbf{f} \in L^1(\mathbb{R}^n)$, we denote

$$D_{\mathbf{f}} \coloneqq \frac{1}{\widehat{\mathbf{f}}(\mathbf{0})}\mathbf{f}.$$

For $s \in \mathbb{R}_{>0}$ and $\mathbf{c} \in \mathbb{R}^n$, the normalized scaled and translated function is denoted by

$$D_{\mathbf{f},s,\mathbf{c}} \coloneqq D_{\mathbf{f}_{s,\mathbf{c}}} : \mathbf{x} \mapsto \frac{1}{\widehat{\mathbf{f}}(\mathbf{0})}\mathbf{f}\left(\frac{1}{s}(\mathbf{x}-\mathbf{c})\right).$$

Similarly, the *discrete probability distribution* over a discrete set A is the mapping defined as

$$D_{A,\mathbf{f},s,\mathbf{c}} \colon (\mathbf{x}) \mapsto \mathbf{f}_{s,\mathbf{c}}(\mathbf{x})/\mathbf{f}_{s,\mathbf{c}}(A).$$

When $s = 1$ or $\mathbf{c} = 0$, the corresponding parameter is omitted in the notation. Notice that the Fourier transform of $D_{\mathbf{f}}$ is given by

$$\widehat{D_{\mathbf{f}}} = \widehat{\left(\frac{1}{\widehat{\mathbf{f}}(\mathbf{0})}\mathbf{f}\right)} = \frac{1}{\widehat{\mathbf{f}}(\mathbf{0})}\widehat{\mathbf{f}}.$$

5.2 Intuition of the (Gaussian) Smoothing Parameter

There are several different ways to think about the current notion of the smoothing parameter that appear in the literature. Perhaps the most intuitive way is to think about it as the minimum stretching of a Gaussian function that is above the covering radius of a lattice. This way, by centering a Gaussian on every point of the lattice and adding them all up, we obtain a function over $\mathbb{R}^n$ that is very close to constant.

Equivalently it can be thought as the largest scaling k of the space such that the overall weight of any translation of the (scaled) lattice,

$$\sum_{\mathbf{v} \in k\mathcal{L}} \rho(\mathbf{v} + \mathbf{c}), \tag{5.6}$$

is approximately the same, in the sense that its value is approximately independent from the shift $\mathbf{c}$.

According to this intuition, we define a smoothing parameter below with respect to a lattice $\mathcal{L} \subset \mathbb{R}^n$, and a value $\varepsilon > 0$ that bounds how much the value of Equation (5.6) can vary for different shift vectors $\mathbf{c} \in \mathbb{R}^n$. The nature of smoothing parameters, under what conditions they exist, and how to find them becomes more approachable by leveraging Lemma 5.6. Combining this result with Proposition 5.3

$$\sum_{\mathbf{v} \in k\mathcal{L}} \mathbf{f}(\mathbf{v} + \mathbf{c}) = \frac{\det \mathcal{L}^*}{k} \sum_{\mathbf{v} \in \frac{1}{k}\mathcal{L}^*} \widehat{\mathbf{f}}(\mathbf{v}) e^{-2\pi i \langle \mathbf{c}, \mathbf{v} \rangle}, \tag{5.7}$$

where $\widehat{\mathbf{f}}$ is the Fourier transform of $\mathbf{f}$, and $\mathcal{L}^*$ is the dual lattice of $\mathcal{L}$. Using this expression, it is possible to prove that under certain conditions, the weight of the shifted lattice becomes approximately independent from the shift—for the formal proof see Lemma 5.14 or [Reg05, Claim 3.8]. To see this intuitively, observe that on the right hand side of Equation (5.7), the vector $\mathbf{c}$ appears only in the exponent $-2\pi i\langle\mathbf{c},\mathbf{v}\rangle$; moreover, for $\mathbf{v} = \mathbf{0}$, the term $\widehat{\mathbf{f}}(\mathbf{v})e^{-2\pi i\langle\mathbf{c},\mathbf{v}\rangle} = \widehat{\mathbf{f}}(\mathbf{0})$. Thus, if $\sum_{\mathbf{v}\in\frac{1}{k}\mathcal{L}^*\setminus\{\mathbf{0}\}}\widehat{\mathbf{f}}(\mathbf{v})e^{-2\pi i\langle\mathbf{c},\mathbf{v}\rangle}$ is sufficiently small for every $\mathbf{c}$, the overall weight of $\mathbf{f_c}$ over the lattice is approximately $\frac{\det\mathcal{L}^*}{k}\widehat{\mathbf{f}}(\mathbf{0})$. An important observation is that this condition can be guaranteed if $\sum_{\mathbf{v}\in\frac{1}{k}\mathcal{L}^*\setminus\{\mathbf{0}\}}\left|\widehat{\mathbf{f}}(\mathbf{v})\right|$ is small, as $\left|e^{-2\pi i\langle\mathbf{c},\mathbf{v}\rangle}\right| \leq 1$. This observation is the motivation for the following definition.

Definition 5.9 (Smoothing Parameter (Gaussian))**.** Given a lattice $\mathcal{L}\subset\mathbb{R}^n$, the *smoothing parameter* $\eta_\varepsilon(\mathcal{L})$ of $\mathcal{L}$, is the minimum positive real s such that for every $s' \geq s$,

$$\widehat{\tfrac{1}{s'}\rho_{s'}}\big(\mathcal{L}^*\setminus\{\mathbf{0}\}\big) = \sum_{\mathbf{v}\in\mathcal{L}^*\setminus\{\mathbf{0}\}}\rho_{\frac{1}{s'}}(\mathbf{v}) < \varepsilon. \tag{5.8}$$

This description of the smoothing parameter is a generalization of the original definition in [MR07] by Micciancio and Regev, where it is only defined for the Gaussian distribution. There are a few details about the previous definition that must be taken into account. Since $\mathbf{f}$ is an integrable function, the Fourier transform $\widehat{\mathbf{f}}$ exists; however, integrability is not enough to guarantee that the weight of the function over every lattice is finite. Moreover, if $\widehat{\mathbf{f}}$ is a decreasing function, then for any real number s larger than the smoothing parameter $\eta_{\mathbf{f},\varepsilon}$, we have that

$$\sum_{\mathbf{v}\in s\mathcal{L}^*\setminus\{\mathbf{0}\}}\left|\widehat{\mathbf{f}}(\mathbf{v})\right| \leq \sum_{\mathbf{v}\in\eta_{\mathbf{f},\varepsilon}(\mathcal{L})\cdot\mathcal{L}^*\setminus\{\mathbf{0}\}}\left|\widehat{\mathbf{f}}(\mathbf{v})\right| < \varepsilon.$$

However, it is impossible to guarantee that $\widehat{\mathbf{f}}$ is a decreasing function. The Fourier transform of a probability distribution on $\mathbb{R}^n$ is commonly an oscillating function. This might go against our intuition, since we might expect that if a particular scaling of a function hides the discrete structure of a lattice, every wider scaling will do so as well, but in general this is not true. To permit an additional degree of freedom in our language, we say that a function $\mathbf{f}$ is ε-*smoothening* with respect to a lattice $\mathcal{L}\subset\mathbb{R}^n$ if $|\widehat{\mathbf{f}}|\big(\mathcal{L}\setminus\{\mathbf{0}\}\big) < \varepsilon$, and note that the smoothing parameter is the smallest scaling of the function $\mathbf{f}$ such that *every* wider scaling of the function is an ε-smoothening function.

5.3 Smoothening Functions

We dedicate this section to introducing two different definitions that generalize the concept of smoothing parameter that appear in the literature for Gaussian functions. The first definition is of an analytic nature, and captures only the necessary requirements for a function to itself hide a given lattice. The second is of geometric nature, which is perhaps a more natural generalization of the traditional concept, and we expect it to better capture the intuition of a reader who is already familiar with the literature.

Definition 5.10 (Smoothening Function)**.** Let $\mathcal{L} \subset \mathbb{R}^n$ and let $\varepsilon \in \mathbb{R}_{>0}$. A function $\mathbf{f} \in L^1(\mathbb{R}^n)$ is said to be *ε-smoothening for $\mathcal{L}$* if

$$\left|\widehat{D_{\mathbf{f}}}\right|\left(\mathcal{L}^* \setminus \{\mathbf{0}\}\right) = \frac{1}{\widehat{\mathbf{f}}(\mathbf{0})} \sum_{\mathbf{y}\in\mathcal{L}^*\setminus\{\mathbf{0}\}} \left|\widehat{\mathbf{f}}(\mathbf{y})\right| < \varepsilon. \tag{5.9}$$

If a function $\mathbf{f}$ is ε-smoothening for a lattice $\mathcal{L}$, intuitively we may expect that this property is preserved by translating the domain of the function. Moreover, scaling the domain or the image of the function should only affect the parameter ε that quantifies how smoothening the function is. In fact, this intuition can be verified and quantified, but only in certain cases, as in the following claim.

Claim 5.11. Let $\varepsilon \in \mathbb{R}_{>0}$ and let $\mathbf{f} \in L^1(\mathbb{R}^n)$ be an ε-smoothening function for a lattice $\mathcal{L}$. Consider $s \in \mathbb{R}_{>0}$ and $\mathbf{c} \in \mathbb{R}^n$. Then the following hold.

1. The function $\mathbf{f}_{s,\mathbf{c}}$ is ε-smoothening for the lattice $s\mathcal{L}$.

2. If $s \in \mathbb{R}_{\geq 1}$ and $\widehat{\mathbf{f}}$ is decreasing on rays, then $\mathbf{f}_{s,\mathbf{c}}$ is ε-smoothening for $\mathcal{L}$.

3. If $s \in \mathbb{Z}_{\geq 1}$, then $\mathbf{f}_{s,\mathbf{c}}$ is ε-smoothening for $\mathcal{L}$.

Proof. Let $\mathbf{c} \in \mathbb{R}^n$, and $r, t \in \mathbb{R}_{\geq 1}$. Then, using Proposition 5.3, we have the following observation.

$$\begin{aligned}
\sum_{\mathbf{y}\in(r\mathcal{L})^*\setminus\{\mathbf{0}\}} \left|\widehat{D_{\mathbf{f},t,\mathbf{c}}}(\mathbf{y})\right| &= \frac{1}{\widehat{\mathbf{f}_{t,\mathbf{c}}}(\mathbf{0})} \sum_{\mathbf{y}\in\frac{1}{r}\mathcal{L}^*\setminus\{\mathbf{0}\}} \left|\widehat{\mathbf{f}_{t,\mathbf{c}}}(\mathbf{y})\right| \\
&= \frac{1}{t^n\widehat{\mathbf{f}}(\mathbf{0})} t^n \sum_{\mathbf{y}\in\frac{1}{r}\mathcal{L}^*\setminus\{\mathbf{0}\}} \left|e^{2\pi i\langle\mathbf{c},\mathbf{y}\rangle}\widehat{\mathbf{f}}\left(t\mathbf{y}\right)\right| \\
&\leq \frac{1}{\widehat{\mathbf{f}}(\mathbf{0})} \sum_{\mathbf{y}\in\frac{t}{r}\mathcal{L}^*\setminus\{\mathbf{0}\}} \left|\widehat{\mathbf{f}}\left(\mathbf{y}\right)\right| \\
&= \sum_{\mathbf{y}\in\frac{t}{r}\mathcal{L}^*\setminus\{\mathbf{0}\}} \left|\widehat{D_{\mathbf{f}}}(\mathbf{y})\right|.
\end{aligned}$$

By making $r = t = s$, Part 1 follows since $\sum_{\mathbf{y}\in\mathcal{L}^*\setminus\{\mathbf{0}\}} \left|\widehat{D_{\mathbf{f}}}(\mathbf{y})\right| < \varepsilon$. For parts 2 and 3, consider $r = 1$, $t = s$. If $\widehat{\mathbf{f}}$ is decreasing on rays, since $s \in \mathbb{R}_{\geq 1}$, we have that for every $\mathbf{y} \in \mathbb{R}^n$, $\widehat{D_{\mathbf{f}}}(\mathbf{y}) \leq \widehat{D_{\mathbf{f}}}(s\mathbf{y})$. Therefore

$$\sum_{\mathbf{y}\in s\mathcal{L}^*\setminus\{\mathbf{0}\}} \left|\widehat{D_{\mathbf{f}}}(\mathbf{y})\right| \leq \sum_{\mathbf{y}\in\mathcal{L}^*\setminus\{\mathbf{0}\}} \left|\widehat{D_{\mathbf{f}}}(\mathbf{y})\right| < \varepsilon. \tag{5.10}$$

Part 2 follows. For Part 3, notice that for $s \in \mathbb{Z}_{\geq 1}$, $s\mathcal{L}^*$ is a sublattice of $\mathcal{L}^*$. Thus we have that Equation (5.10) holds, and the proof follows. □

Definition 5.12 (Smoothing Parameter)**.** Given a lattice $\mathcal{L} \subset \mathbb{R}^n$, a function $\mathbf{f} \in L^1(\mathbb{R}^n)$ and $\varepsilon > 0$, the *smoothing parameter* $\eta_{\mathbf{f},\varepsilon}(\mathcal{L})$ of $\mathbf{f}$ over the lattice $\mathcal{L}$, if it exists, is the minimum positive real s such that for every $s' \in \mathbb{R}_{>s}$, the function $\mathbf{f}_{s'}$ is ε-smoothening.

It is important to remark that definitions 5.10 and 5.12 refer to different concepts. Though they are very similar in nature, it is important to understand the subtle differences that they entail. In short, not every smoothening function possesses a smoothing parameter. Traditionally, we think of the smoothing parameter as the minimum stretching of the Gaussian distribution that induces a distribution on the quotient that is ε-close to uniform. Since the Gaussian is strictly decreasing—in addition to being its own Fourier transform—we have that every stretching larger than the smoothing parameter makes it also a smoothening function. This is not true in general, since the Fourier transform of arbitrary probability distributions is not strictly decreasing, in general. Thus the weight of the Fourier transform over increasing scalings of the dual lattice may fluctuate.

Example 5.13. For instance, consider the function f which is the characteristic function of the interval $\left[-\frac{1}{2}, \frac{1}{2}\right] \subset \mathbb{R}$, that is

$$f(x) = \begin{cases} 1 & \text{if } x \in \left[-\frac{1}{2}, \frac{1}{2}\right], \\ 0 & \text{otherwise.} \end{cases}$$

Intuitively, every non-integer stretching of this function will assign more weight to some elements than to others in the quotient $\mathbb{R}/\mathbb{Z}$. This intuition is reflected in the Poisson summation formula. Since f is normalized, then $D_f = f$. The Fourier transform of D_f is given by

$$\widehat{f}(y) = \operatorname{sinc}(y) := \begin{cases} \frac{\sin(\pi y)}{\pi y} & \text{if } y \neq 0, \\ 1 & \text{if } y = 0. \end{cases}$$

Thus the series $\sum_{y\in\mathbb{Z}\setminus\{0\}} \left|\widehat{f}(y)\right| = 0$. This indicates that the probability distribution induced by f on the quotient is (perfectly) uniform. On the other hand, for any positive integer k,

$$\begin{aligned}
\sum_{y\in\frac{2k+1}{2}\mathbb{Z}\setminus\{0\}} \left|\widehat{f}(y)\right| &= \sum_{y\in\mathbb{Z}\setminus\{0\}} \left|\frac{2\sin\big((2k+1)y\pi/2\big)}{(2k+1)y}\right| \\
&= \frac{1}{2k+1}\sum_{y\in\mathbb{Z}\setminus\{0\}} \left|\frac{2\sin\big(y\pi/2\big)}{y}\right| \\
&= \frac{1}{2k+1}\sum_{y\in 2\mathbb{Z}+1} \left|\frac{2\sin\big(y\pi/2\big)}{y}\right| + \sum_{y\in 2\mathbb{Z}\setminus\{0\}} \left|\frac{2\sin\big(y\pi/2\big)}{y}\right| \\
&= \frac{1}{2k+1}\sum_{y\in 2\mathbb{Z}+1} \frac{2}{|y|},
\end{aligned}$$

which is not convergent. As a result, there exists arbitrarily large $s \in \mathbb{R}\setminus\mathbb{Z}$ such that the series $\sum_{y\in s\mathbb{Z}\setminus\{0\}} \left|\widehat{f}(y)\right|$ is not convergent. Thus the function f cannot have a smoothing parameter for the lattice $\mathbb{Z}$, despite the fact that, for every ε, f itself is ε-smoothening for $\mathbb{Z}$.

By now it is probably clear that the existence of smoothing parameters for an arbitrary function is not obvious from the definition and, in fact, this may be a rare property in real-valued functions. Indeed, we start the following section by giving an example of an integrable, infinitely differentiable function for which a smoothing parameter does not exist (with respect to $\mathbb{Z}$). Moreover, to the best of our knowledge, the concept of smoothing parameter has been studied and its existence proved only in the case of the Gaussian distribution.

Properties of Smoothening Functions

There are several results in the literature that are proved for Gaussians that are wider than the smoothing parameter. In this subsection we argue that many of the important and useful results can be proved for smoothening functions. Moreover, we argue that the results given in this subsection can all be seen as a direct consequence of a generalized version of [Reg05, Claim 3.8].

Lemma 5.14. *Let $\mathcal{L} \subset \mathbb{R}^n$ be a lattice and let $\varepsilon \in \mathbb{R}_{>0}$. Consider an ε-smoothening function $\mathbf{f} \in L^1(\mathbb{R}^n)$ for the lattice $\mathcal{L}$. Then for any $\mathbf{c} \in \mathbb{R}^n$,*

$$\mathbf{f}(\mathcal{L}+\mathbf{c}) \in \det(\mathcal{L}^*)\widehat{\mathbf{f}}(\mathbf{0})\,(1-\varepsilon, 1+\varepsilon)\,.$$

Proof. Consider the function $\mathbf{f}_{-\mathbf{c}}$ given by $\mathbf{f}_{-\mathbf{c}} \colon \mathbf{x} \mapsto \mathbf{f}(\mathbf{x}+\mathbf{c})$. Then the weight of $\mathbf{f}$ over the shifter lattice is expressed as

$$\begin{aligned}\mathbf{f}(\mathcal{L}+\mathbf{c}) &= \sum_{\mathbf{x}\in\mathcal{L}+\mathbf{c}} \mathbf{f}(\mathbf{x}) \\ &= \sum_{\mathbf{x}\in\mathcal{L}} \mathbf{f}_{-\mathbf{c}}(\mathbf{x}).\end{aligned}$$

By the Poisson summation formula (Lemma 5.6) and Proposition 5.3,

$$\begin{aligned}\sum_{\mathbf{x}\in\mathcal{L}} \mathbf{f}_{-\mathbf{c}}(\mathbf{x}) &= \det(\mathcal{L}^*) \sum_{\mathbf{y}\in\mathcal{L}^*} \widehat{\mathbf{f}_{-\mathbf{c}}}(\mathbf{y}) \\ &= \det(\mathcal{L}^*) \sum_{\mathbf{y}\in\mathcal{L}^*} \widehat{\mathbf{f}}(\mathbf{y}) e^{2\pi i\langle \mathbf{y},\mathbf{c}\rangle} \\ &= \det(\mathcal{L}^*) \left(\widehat{\mathbf{f}}(\mathbf{0}) + \sum_{\mathbf{y}\in\mathcal{L}^*\setminus\{\mathbf{0}\}} \widehat{\mathbf{f}}(\mathbf{y}) e^{2\pi i\langle \mathbf{y},\mathbf{c}\rangle} \right), \\ &= \det(\mathcal{L}^*)\widehat{\mathbf{f}}(\mathbf{0}) \left(1 + \sum_{\mathbf{y}\in\mathcal{L}^*\setminus\{\mathbf{0}\}} \widehat{D_{\mathbf{f}}}(\mathbf{y}) e^{2\pi i\langle \mathbf{y},\mathbf{c}\rangle} \right),\end{aligned}$$

where the second equality follows from Proposition 5.3. Now, recall that $\mathbf{f}$ is a real-valued function, and $\mathrm{Re}\left(e^{2\pi i\langle \mathbf{y},\mathbf{c}\rangle}\right) = \cos\left(2\pi\langle \mathbf{y},\mathbf{c}\rangle\right) \in [-1,1]$. Therefore, from Equation (5.9) we have that $\sum_{\mathbf{y}\in\mathcal{L}^*\setminus\{\mathbf{0}\}} \widehat{D_{\mathbf{f}}}(\mathbf{y}) e^{2\pi i\langle \mathbf{y},\mathbf{c}\rangle} \in (-\varepsilon,\varepsilon)$, as required. □

Corollary 5.15. *Let $\mathbf{f}$ an ε-smoothening function with respect to a lattice $\mathcal{L} \subset \mathbb{R}^n$. Then for any $\mathbf{c} \in \mathbb{R}^n$,*

$$\mathbf{f}(\mathcal{L}+\mathbf{c}) \in \det(\mathcal{L}^*) \int_{\mathbb{R}^n} \mathbf{f}(\mathbf{x}) d\mathbf{x}\,(1-\varepsilon,1+\varepsilon)\,.$$

Proof. If follows from $\widehat{\mathbf{f}}(\mathbf{0}) = \int_{\mathbb{R}^n} \mathbf{f}(\mathbf{x}) e^{2\pi i\langle \mathbf{x},\mathbf{0}\rangle}\, d\mathbf{x} = \int_{\mathbb{R}^n} \mathbf{f}(\mathbf{x})\, d\mathbf{x}$. □

This proposition demonstrates that the total measure of a ε-smoothening function cannot deviate very much (in terms of ε) under arbitrary shifts $\mathbf{c}$ of the measure—or, equivalently, shifts of the lattice. Note that since Lemma 5.14 and Corollary 5.15 hold for any $\mathbf{c} \in \mathbb{R}^n$, they can be alternatively stated in terms of cosets of a lattice, that is, $\mathbf{f}_s(\mathcal{L}+\mathbf{c})$ via the change of variables $\mathbf{f}_s(\mathcal{L}+\mathbf{c}) - \mathbf{c} = \mathbf{f}_{s,-\mathbf{c}}(\mathcal{L})$.

Proposition 5.16. *Let $\varepsilon \in \mathbb{R}_{>0}$, and let $\mathbf{f}$ be a ε-smoothening function for a lattice $\mathcal{L}$. Then for any $\mathbf{c}_1, \mathbf{c}_2 \in \mathbb{R}^n$, we have*

$$\mathbf{f}_{\mathbf{c}_1}(\mathcal{L}) \in \left(\frac{1-\varepsilon}{1+\varepsilon}, \frac{1+\varepsilon}{1-\varepsilon}\right) \cdot \mathbf{f}_{\mathbf{c}_2}(\mathcal{L}).$$

Proof. Let $\mathbf{c}_1, \mathbf{c}_2 \in \mathbb{R}^n$. By Lemma 5.14,

$$\widehat{\mathbf{f}}(\mathbf{0}) \det(\mathcal{L}^*)\,(1-\varepsilon) \le \mathbf{f}_{\mathbf{c}_1}(\mathcal{L}) \le \widehat{\mathbf{f}}(\mathbf{0}) \det(\mathcal{L}^*)\,(1+\varepsilon)\,.$$

Similarly for $\mathbf{c}_2$. Thus, combining these two inequalities we obtain

$$\frac{1-\varepsilon}{1+\varepsilon} = \frac{\widehat{\mathbf{f}}(\mathbf{0}) \det(\mathcal{L}^*)\,(1-\varepsilon)}{\widehat{\mathbf{f}}(\mathbf{0}) \det(\mathcal{L}^*)\,(1+\varepsilon)} \le \frac{\mathbf{f}_{\mathbf{c}_1}(\mathcal{L})}{\mathbf{f}_{\mathbf{c}_2}(\mathcal{L})} \le \frac{\widehat{\mathbf{f}}(\mathbf{0}) \det(\mathcal{L}^*)\,(1+\varepsilon)}{\widehat{\mathbf{f}}(\mathbf{0}) \det(\mathcal{L}^*)\,(1-\varepsilon)} = \frac{1+\varepsilon}{1-\varepsilon}.$$

The result follows. □

Lemma 5.17. *Let $\varepsilon \in \mathbb{R}_{>0}$, and let $\mathbf{f}$ be an ε-smoothening function for a lattice $\mathcal{L} = \mathcal{L}(\mathbf{B})$. Then the distribution over the fundamental domain $\mathcal{P}(\mathbf{B})$ given by*

$$D_{\mathbf{f}} \bmod \mathcal{P}(\mathbf{B})$$

is within statistical distance $\varepsilon/2$ from the uniform distribution over $\mathcal{P}(\mathbf{B})$.

Proof. Write $\psi = D_{\mathbf{f}} \mod \mathcal{P}(\mathbf{B})$. Then, for $\mathbf{x} \in \mathcal{P}(\mathbf{B})$, $\psi(\mathbf{x})$ is given by

$$\psi(\mathbf{x}) = \frac{1}{\widehat{\mathbf{f}}(\mathbf{0})} \sum_{\mathbf{y} \in \mathcal{L}} \mathbf{f}(\mathbf{x}+\mathbf{y}) = \frac{1}{\widehat{\mathbf{f}}(\mathbf{0})} \mathbf{f}(\mathcal{L}+\mathbf{x}).$$

Therefore by Lemma 5.14,

$$\psi(\mathbf{x}) = \frac{1}{\widehat{\mathbf{f}}(\mathbf{0})} \mathbf{f}(\mathcal{L}+\mathbf{x}) = \det \mathcal{L}^* \big(1 + \delta(\mathbf{x})\big),$$

where $\delta(\mathbf{x}) \in (-\varepsilon, \varepsilon)$. On the other hand, the density function of the uniform distribution over $\mathcal{P}(\mathbf{B})$ is given by

$$U(\mathbf{x}) = 1/\operatorname{vol}\big(\mathcal{P}(\mathbf{B})\big) = \det\big(\mathcal{L}^*\big).$$

Hence, the statistical distance $\Delta(U, \psi)$ between ψ and U is

$$\begin{aligned}
2\Delta(U, \psi) &= \int_{\mathbf{x}\in\mathcal{P}(\mathbf{B})} \big|\psi(\mathbf{x}) - U(\mathbf{x})\big| d\mathbf{x} \\
&\leq \operatorname{vol}\big(\mathcal{P}(\mathbf{B})\big) \max_{\mathbf{x}\in\mathcal{P}(\mathbf{B})} \big|\psi(\mathbf{x}) - \det\mathcal{L}^*\big| \\
&= \det\mathcal{L} \max_{\mathbf{x}\in\mathcal{P}(\mathbf{B})} \Big|\det\mathcal{L}^*\big(1+\delta(\mathbf{x})\big) - \det\mathcal{L}^*\Big| \\
&= \det\mathcal{L}\det\mathcal{L}^* \max_{\mathbf{x}\in\mathcal{P}(\mathbf{B})} \Big|1+\delta(\mathbf{x}) - 1\Big| \\
&< \varepsilon,
\end{aligned}$$

which completes the proof. □

Lemma 5.18. *Let $\mathcal{L} \subset \mathbb{R}^n$ be a lattice and let $\varepsilon \in \mathbb{R}_{>0}$. Consider an ε-smoothening function $\mathbf{f}$ for $\mathcal{L}$ such that $\|\mathbf{f}\|_\infty = \mathbf{f}(\mathbf{0}) = 1$. Then for any $s \in \mathbb{Z}_{\geq 2}$,*

$$D_{\mathcal{L},\mathbf{f},s,\mathbf{c}}(\mathbf{x}) \leq s^{-n}\left(\tfrac{1+\varepsilon}{1-\varepsilon}\right).$$

Proof. First, note that by Corollary 5.15,

$$s^n \det(\mathcal{L}^*) \cdot (1-\varepsilon) \leq \mathbf{f}_{s,\mathbf{c}}(\mathcal{L}) \leq s^n \det(\mathcal{L}^*) \cdot (1+\varepsilon).$$

Now, since $\|\mathbf{f}\|_\infty = \mathbf{f}(\mathbf{0}) = 1$, it follows that

$$\begin{aligned}
D_{\mathcal{L},\mathbf{f},s,\mathbf{c}}(\mathbf{x}) &= \frac{\mathbf{f}_{s,\mathbf{c}}(\mathbf{x})}{\mathbf{f}_{s,\mathbf{c}}(\mathcal{L})} \\
&\leq \frac{1}{s^n\widehat{\mathbf{f}}(\mathbf{0})\det(\mathcal{L}^*)\cdot(1-\varepsilon)} \\
&\leq \frac{f(\mathcal{L})}{s^n\widehat{\mathbf{f}}(\mathbf{0})\det(\mathcal{L}^*)\cdot(1-\varepsilon)} \\
&\leq \frac{\widehat{\mathbf{f}}(\mathbf{0})\det(\mathcal{L}^*)}{s^n\widehat{\mathbf{f}}(\mathbf{0})\det(\mathcal{L}^*)} \cdot \frac{(1+\varepsilon)}{(1-\varepsilon)} \\
&\leq s^{-n}\left(\tfrac{1+\varepsilon}{1-\varepsilon}\right),
\end{aligned}$$

as required. □

5.4 Families of Smoothening Functions

Intuitively speaking, a function is smoothening whenever one is able to confine the majority of the weight of its Fourier transform to a bounded set. This intuition may lead us to think that a function is smoothening whenever its Fourier transform is integrable. This intuition, however, ignores the fact that the smoothening property is defined with respect to a discrete set. We elaborate on this argument more in the following example.

Example 5.19. For each $n \in \mathbb{Z}$ consider the function

$$b_n(x) = \begin{cases} \exp\left(\dfrac{1}{1 - \left(2^{n+1}(x-n)\right)^2}\right) & \text{if } x \in \left(n - \frac{1}{2^{n+1}}, n + \frac{1}{2^{n+1}}\right), \\ 0 & \text{otherwise.} \end{cases}$$

This is a bump function supported on a set centered around n with diameter $\frac{1}{2^{|n|}}$. Its maximum value is 1, which is reached at $x = n$. Since the function is positive and bounded by 1, the integral of b_n over $\mathbb{R}$ is bounded by $\frac{1}{2^{|n|}}$. Using these functions as building blocks, construct the function $b = \sum_{n \in \mathbb{Z}} b_n$. It is clear that b is bounded and has finite integral (moreover, it is infinitely differentiable). Therefore, its inverse Fourier transform is an integrable function f.

Consider the lattice $\mathcal{L} = \mathbb{Z} \subset \mathbb{R}$. Notice that for any non-zero rational $s \in \mathbb{Q}$, the lattice $\frac{1}{s}\mathcal{L}^* = \frac{1}{s}\mathbb{Z}$ has an infinite number of integers. Thus the series $\sum_{x \in \frac{1}{s}\mathcal{L}^*} b(x)$ does not converge for any non-zero rational s. As a consequence, no positive rational s can be a smoothing parameter for $\check{f}$ with respect to $\mathbb{Z}$.

A consequence of this example is that not every integrable function has a smoothing parameter. This is stated more precisely with the following claim.

Claim 5.20. For every $n \in \mathbb{Z}_{>0}$ there exist a function $\mathbf{f} \in L^1(\mathbb{R}^n)$ and a lattice $\mathcal{L} \subset \mathbb{R}$ such that for all $s \in \mathbb{R}$, s is not a smoothing parameter of $\mathbf{f}$ with respect to $\mathcal{L}$.

Proof. Let $n \in \mathbb{Z}_{>0}$ and let g be the function described in Example 5.19. Consider the function $\mathbf{g}\colon (x_1, \ldots, x_n) \mapsto \prod_{i \in [n]} g(x_i)$. Then, for any non-zero rational $s \in \mathbb{Q}$, the lattice $\frac{1}{s}\mathbb{Z}^N$ has an infinite number elements in $\mathbb{Z}^n$. The result follows. □

Remark 5.21. Recall that $\mathcal{D}_n$ denotes the set of real-valued functions over $\mathbb{R}^n$ that admit the Poisson summation formula over any given lattice. In particular, for any $f \in \mathcal{D}_n$ and for any lattice $\mathcal{L} \subset \mathbb{R}^n$, the series $\sum_{\mathbf{x} \in \mathcal{L}^*} f(\mathbf{x})$ is absolutely convergent. It follows, by definition, that for $\varepsilon = \sum_{\mathbf{x} \in \mathcal{L}^* \setminus \{\mathbf{0}\}} \left|\widehat{f}(\mathbf{x})\right|$, the function f is ε-smoothening for $\mathcal{L}$.

In the rest of this section we present different conditions for a function to have smoothening properties. As a result we have the following theorem.

Theorem 5.22. *For each $n \in \mathbb{Z}_{>0}$ there exist families of functions $\mathcal{F}' \subset \mathcal{F} \subset \mathcal{D}_n$, which are infinitely large, such that the following holds.*

- *For every $\mathbf{f} \in \mathcal{F}$, lattice $\mathcal{L} \subset \mathbb{Q}^n$ and $\varepsilon > 0$, the smoothing parameter $\eta_{\mathbf{f},\varepsilon}(\mathcal{L})$ is finite.*
- *For every $\mathbf{f} \in \mathcal{F}'$, lattice $\mathcal{L} \subset \mathbb{Q}^n$ and $\varepsilon > 0$, $\eta_{\mathbf{f},\varepsilon}(\mathcal{L}) \in O\Big(\operatorname{poly}\Big(n, \frac{1}{\log \varepsilon}, \lambda_n(\mathcal{L})\Big)\Big)$.*

Absolutely Convergent Functions

Our first example is the family of functions that are absolutely convergent over every lattice. This is a standard pre-requisite for the application of the Poisson summation formula. Moreover, following Remark 5.21 we have that for every function $\mathbf{f} \in \mathcal{D}_n$ there exists $\varepsilon \in \mathbb{R}_{>0}$ such that $\mathbf{f}$ is ε-smoothening.

Consider a rational lattice $\mathcal{L} \in \mathbb{Q}^n$ generated by a basis $\mathbf{b}_1, \dots, \mathbf{b}_n$ and let D be the set of denominators of all the entries of all the basis vectors. Notice that, by taking m to be the lowest common multiple of the members of D, we have that for every $i \in [n]$, the vector $m\mathbf{b}_i$ has integer entries. Therefore $m\mathcal{L} \subset \mathbb{Z}^n$. If $d = 1/m$, then $\mathcal{L} \subset d\mathbb{Z}^n$.

Proposition 5.23. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a function and consider the mapping defined by $\mathbf{f}\colon (x_1, \dots, x_n) \mapsto \prod_{i\in[n]} f(x_i)$. Suppose that for every $\mathcal{L}_1 \subset \mathbb{Q}$, the series $\sum_{x\in\mathcal{L}_1} f(\mathbf{x})$ converges absolutely. Then for every $\mathcal{L} \subset \mathbb{Q}^n$, the series $\sum_{\mathbf{x}\in\mathcal{L}} \mathbf{f}(\mathbf{x})$ converges absolutely.*

Proof. Let $\mathcal{L} \subset \mathbb{Q}^n$. Following the discussion above, there exists $d \in \mathbb{Q}$ such that $\mathcal{L} \subset d\mathbb{Z}^n$. Then the summation of the absolute values of $\mathbf{f}(\mathbf{x})$ is bounded by

$$\sum_{\mathbf{x}\in\mathcal{L}} \big|\mathbf{f}(\mathbf{x})\big| \le \sum_{(x_1,\dots,x_n)\in d\mathbb{Z}^n} \prod_{i\in[n]} \big|f(x_i)\big| = \prod_{i\in[n]} \sum_{x_i\in d\mathbb{Z}} \big|f(x_i)\big| < \infty.$$

The proof follows. □

Claim 5.24. Let $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}$ be a function and let $\mathcal{L} \subset \mathbb{R}^n$ be a lattice. Consider a balanced set K such that $\mathbf{0}$ is contained in its interior. If the series $\sum_{\mathbf{x}\in\mathcal{L}} \mathbf{f}(\mathbf{x})$ converges absolutely, then for every $\varepsilon \in \mathbb{R}_{>0}$ there exists $\eta \in \mathbb{R}$ such that $\sum_{\mathbf{x}\in\mathcal{L}\setminus\eta K} \mathbf{f}(\mathbf{x}) < \varepsilon$.

Proof. Consider the function $W\colon r \mapsto \sum_{\mathbf{x}\in\mathcal{L}\setminus rK} \mathbf{f}(\mathbf{x})$. For every r, $W(r)$ is finite. Moreover, since K is balanced and contains the vector $\mathbf{0}$ in its interior, the function W is non-increasing and for every $\mathbf{x} \in \mathbb{R}^n$ there exists r_0 such that for every $r \geq r_0$, $\mathbf{x} \in rK$. Therefore $\lim_{r\to\infty} W(r) = 0$. The result follows. □

Proposition 5.25. *Let $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}$ be a function and let $\mathcal{L} \subset \mathbb{R}^n$ be a lattice. If the series $\sum_{\mathbf{x}\in\mathcal{L}} \mathbf{f}(\mathbf{x})$ converges absolutely, then for every $\varepsilon \in \mathbb{R}_{>0}$ there exists $\eta \in \mathbb{R}$ such that $\sum_{\mathbf{x}\in\eta\mathcal{L}\setminus\{\mathbf{0}\}} \mathbf{f}(\mathbf{x}) < \varepsilon$.*

Proof. Let $\varepsilon > 0$. Consider the ball $B_n^2(\mathbf{0})$. By Claim 5.24, there exists $r_0 \in \mathbb{R}_{>0}$ such that $\sum_{\mathbf{x}\mathcal{L}\setminus rB_n^2(\mathbf{0})} \mathbf{f}(\mathbf{x}) < \varepsilon$. Consider $\eta = \lceil r_0/\lambda_1(\mathcal{L})\rceil \in \mathbb{Z}_{>0}$. Then $\lambda_1(\eta\mathcal{L}) \geq r_0$, which implies that $\eta\mathcal{L} \cap r_0 B_n^2(\mathbf{0}) = \{\mathbf{0}\}$. Moreover, since η is an integer, $\eta\mathcal{L} \subseteq \mathcal{L}$. As a consequence

$$\sum_{\mathbf{x}\in\eta\mathcal{L}\setminus\{0\}} \left|\mathbf{f}(\mathbf{x})\right| = \sum_{\mathbf{x}\in\eta\mathcal{L}\setminus r_0 B_n^2(\mathbf{0})} \left|\mathbf{f}(\mathbf{x})\right| \leq \sum_{\mathbf{x}\in\mathcal{L}\setminus r_0 B_n^2(\mathbf{0})} \left|\mathbf{f}(\mathbf{x})\right| < \varepsilon,$$

as desired. □

A consequence of the previous proposition is that, given a lattice $\mathcal{L}$ and $\varepsilon \in \mathbb{R}_{>0}$, every function in $\mathcal{D}_n$ admits an integral dilation of the domain that makes it ε-smoothening for $\mathcal{L}$.

Proposition 5.26. *Let k be a positive integer and consider a function $f \in L^1(\mathbb{R})$ such that for every $\ell \leq k$, $f^{(\ell)}$ exists and belongs to $L^1(\mathbb{R})$. Then $\left|\widehat{f}(y)\right| \in O\big(y^{-k}\big)$.*

Proof. Since $f \in L^1(\mathbb{R})$, $\lim_{x\to\pm\infty} f(x) = 0$. Thus, expressing $\widehat{f}$ as an integral by parts, we have that

$$\begin{aligned}
\widehat{f}(y) &= \int_{\mathbb{R}} f(x)e^{-2\pi ixy}\,dx \\
&= -\frac{f(x)}{2\pi iy}e^{-2\pi ixy}\Big|_{-\infty}^{\infty} + \frac{1}{2\pi iy}\int_{\mathbb{R}} f'(x)e^{-2\pi ixy}\,dx \\
&= \frac{1}{2\pi iy}\widehat{f}'(y).
\end{aligned}$$

Inductively, using the same approach we obtain $\widehat{f}(y) = \left(\frac{1}{2\pi iy}\right)^k \widehat{f^{(k)}}(y)$. Since $\widehat{f^{(k)}} \in L^\infty$, its absolute value is bounded by a positive constant $\vartheta_{f,k} \in \mathbb{R}$. Therefore

$$\left|\widehat{f}(y)\right| \leq \vartheta_{f,k}\left(\frac{1}{2\pi y}\right)^k,$$

as required. □

Notice that the constant $\vartheta_{f,k}$ in the last proposition is the ∞-norm of the function $\widehat{f^{(k)}}$, which is equal to $\left\|f^{(k)}\right\|_1 = \int_{\mathbb{R}} f^{(k)}$.

Corollary 5.27. *Let f be a function in $L^1(\mathbb{R})$ such that its first and second derivative both exist and belong to $L^1(\mathbb{R})$. Then for every lattice $\mathcal{L} \in \mathbb{R}$, the series $\sum_{y \in \mathcal{L}^*} \widehat{f}(y)$ converges absolutely.*

Proof. Let $\mathcal{L} \subset \mathbb{R}$ be a lattice. Then, there exists $a \in \mathbb{R}$ such that the dual lattice $\mathcal{L}^*$ is equal to $a\mathbb{Z}$. By Proposition 5.26,

$$\begin{aligned}\sum_{y \in \mathcal{L}^*} \left|\widehat{f}(y)\right| &= \sum_{y \in \mathbb{Z}} \left|\widehat{f}(ay)\right| \\ &\leq \vartheta_{f,2} \left(\frac{1}{2\pi a}\right)^2 \sum_{y \in \mathbb{Z}} \frac{1}{y^2} \\ &= \frac{\vartheta_{f,2}}{6} \left(\frac{\pi}{2\pi a}\right)^2,\end{aligned}$$

as required. □

Schwartz and Rapidly Decreasing Functions

Integrability is a property that only bounds the volume of the function. The function b described in the last example diverges over a lattice because the integrability property still allows it to "misbehave" on an infinite number of points. Nonetheless, Proposition 5.26 provides us with sufficient analytic conditions on the function that dictate its asymptotic behavior. We combine this result with the following lemma by Betke et.al. to prove that rapidly decreasing functions are absolutely convergent over any lattice.

Lemma 5.28 ([BHW93], Theorem 2.1)**.** *Let $K \in \mathcal{K}_n$ and let $\mathcal{L} \subset \mathbb{R}^n$ be a lattice. Then*

$$|K \cap \mathcal{L}| \leq \left\lfloor \frac{2}{\lambda_1(K, \mathcal{L})} + 1 \right\rfloor^n .$$

Proposition 5.29. *Let $K \in \mathcal{K}_n$ and let $\mathbf{f} \colon \mathbb{R}^n \to \mathbb{R}$. Suppose there exists $\delta \in \mathbb{R}_{>1}$ such that $\left|\mathbf{f}(\mathbf{x})\right| = O\big((1 + \|\mathbf{x}\|_K)^{-n-\delta}\big)$. Then for every lattice $\Lambda \subset \mathbb{R}^n$, the series $\mathbf{f}(\Lambda)$ is uniformly convergent.*

Proof. The idea is to divide $\mathbb{R}^n$ into shells $S_1, S_2, \ldots$, and bound the number of lattice points in every shell $\mathcal{L} \cup S_i$ by a polynomial of degree n. Thus, by the hypothesis on $\mathbf{f}$, the sum of the function over the lattice points in every shell, $\sum_{\mathbf{v}\in\mathcal{L}\cup S_i} |\mathbf{f}(\mathbf{v})|$, is bounded by the inverse of a superlinear function. By considering all the shells, the series $\sum_{\mathbf{v}\in\mathcal{L}} \mathbf{f}(\mathbf{v})$ converges absolutely.

Formally, let $S_1 = K$ and for each integer $r > 1$, let $S_r = rK \setminus (r-1)K$. By definition, for any $r > 0$, $\lambda_1(rK, \mathcal{L}) = \frac{1}{r}\lambda_1(K, \mathcal{L})$. Thus, following Lemma 5.28, we have that for any $r > 0$,

$$|rK \cap \mathcal{L}| \leq \left\lfloor \frac{2r}{\lambda_1(K, \mathcal{L})} + 1 \right\rceil^n .$$

Hence, since $S_r \subseteq rK$,

$$|S_r \cap \mathcal{L}| \leq \left\lfloor \frac{2r}{\lambda_1(K, \mathcal{L})} + 1 \right\rceil^n . \tag{5.11}$$

On the other hand, by definition, there exists a constant $c \in \mathbb{R}_{>0}$ and $N \in \mathbb{Z}_{>1}$ such that for all $\mathbf{x} \in \mathbb{R}^n$ such that $\|\mathbf{x}\|_K \in \mathbb{R}_{>N}$, $\big|\mathbf{f}(\mathbf{x})\big| \leq c\left(\frac{1}{r}\right)^{n+\delta}$. Therefore, for any $r \in \mathbb{Z}_{>N}$,

$$\begin{aligned}\sum_{\mathbf{v}\in S_r\cap\mathcal{L}} \big|\mathbf{f}(\mathbf{x})\big| &\leq c\left(\frac{2}{\lambda_1(K,\mathcal{L})} + \frac{1}{r}\right)^n \left(\frac{1}{r}\right)^\delta \\ &\leq c\left(\frac{2}{\lambda_1(K,\mathcal{L})} + 1\right)^n \left(\frac{1}{r}\right)^\delta .\end{aligned}$$

By adding up for all $r \geq 1$,

$$\begin{aligned}\sum_{\mathbf{x}\in\mathcal{L}} \big|\mathbf{f}(\mathbf{x})\big| &= \mathbf{f}\big((N-1)K \cap \mathcal{L}\big) + \sum_{r\in\mathbb{Z}_{\geq N}} \sum_{\mathbf{x}\in S_r\cap\mathcal{L}} \big|\mathbf{f}(\mathbf{x})\big| \\ &\leq \mathbf{f}\big((N-1)K \cap \mathcal{L}\big) + \sum_{r\in\mathbb{Z}_{\geq N}} c\left(\frac{2}{\lambda_1(K,\mathcal{L})} + 1\right)^n \left(\frac{1}{r}\right)^\delta \\ &= \mathbf{f}\big((N-1)K \cap \mathcal{L}\big) + c\left(\frac{1}{\lambda_1(K,\mathcal{L})} + \frac{1}{2}\right)^n \sum_{r\in\mathbb{Z}_{\geq N}} \left(\frac{1}{r}\right)^\delta \\ &\leq \mathbf{f}\big((N-1)K \cap \mathcal{L}\big) + c\left(\frac{1}{\lambda_1(K,\mathcal{L})} + \frac{1}{2}\right)^n \left(\left(\frac{1}{N}\right)^\delta + \left(\frac{1}{N}\right)^{\delta-1}\right) \\ &< \infty\end{aligned}$$

This completes the proof. □

Product of Smoothing Functions Over Rational Lattices

In the following claim we construct a partition of a rational lattice $\mathcal{L}$ into an infinite collection of parallel copies of an $(n-1)$-dimensional sublattice of $\mathcal{L}$, all of them orthogonal to a given vector.

Claim 5.30. Let $\mathcal{L} \subset \mathbb{Q}^n$ be a full-rank lattice. For every $\mathbf{e} \in \mathbb{Q}^n$ such that $\|\mathbf{e}\|_2 = 1$, there exists a collection of subsets $\{A_i \colon i \in \mathbb{Z}\}$, such that the following properties are satisfied.

1. The lattice $\mathcal{L}$ is the disjoint union
$$\mathcal{L} = \bigcup_{i \in \mathbb{Z}} A_i.$$
2. $A_0 = \big\{\mathbf{x} \in \mathcal{L} \colon \langle \mathbf{x}, \mathbf{e} \rangle = 0\big\}$ is an $n-1$ dimensional lattice.
3. There exists a vector $\mathbf{v} \in \mathcal{L} \setminus A_0$ such that for each $i \in \mathbb{Z}$, $A_i - i\mathbf{v} = A_0$.

Proof. Without loss of generality, it is enough to show the claim for $\mathbf{e} = \mathbf{e}_1 = (1, 0, \ldots, 0)$. Clearly, A_0 contains the vector $\mathbf{0}$ and is closed under addition; thus A_0 is a lattice. We show that there exists $\mathbf{v} \neq \mathbf{0}$ also contained in A_0. Note that, in this case, a vector $\mathbf{v} = (v_1, \ldots, v_n)$ is in A_0 if and only if $v_1 = 0$.

Let $\mathbf{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for $\mathcal{L}$. If there is a vector in $\mathbf{B}$ contained in A_0, then the statement above is clear. Otherwise, $\mathbf{b}_1, \mathbf{b}_2 \notin A_0$. (in other words, their component perpendicular to A_0 is non-zero). Since $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{Q}^n$, there exist integers x and y such that $x\langle \mathbf{b}_1, \mathbf{e} \rangle + y\langle \mathbf{b}_2, \mathbf{e} \rangle = 0$. Hence $\langle x\mathbf{b}_1 + y\mathbf{b}_2, \mathbf{e} \rangle = 0$, which implies that the integer combination $x\mathbf{b}_1 + y\mathbf{b}_2 \in A_0$. Moreover, $x\mathbf{b}_1 + y\mathbf{b}_2 \neq \mathbf{0}$ as $\mathbf{b}_1$ and $\mathbf{b}_2$ are linearly independent. As a consequence, A_0 is a non-trivial lattice.

To show that A_0 has exactly $n-1$ dimensions, assume, without loss of generality, that the first coordinate of $\mathbf{b}_1 \in \mathbf{B}$ is different from 0. By the process above, for each $i \in \{2, \ldots, n\}$, we can find a linear combination $\mathbf{a}_i = x_i\mathbf{b}_1 + y_i\mathbf{b}_i \in A_0$, with $x_i, y_i \in \mathbb{Z}$. Additionally, as the first coordinate of $\mathbf{b}_1$ is non-zero, we have that both x_i and y_i are different from 0. Hence, since all $\mathbf{b}_i$ are linearly independent, we have that $\mathbf{a}_2, \ldots, \mathbf{a}_n$ are $n-1$ linearly independent vectors, and thus the lattice contained in A_0 has dimension $n-1$.

Finally, let $\mathbf{v} \in \mathcal{L} \setminus A_0$ be a vector that is closest to the hyperplane $\mathbf{e}^{\perp} = \big\{\mathbf{x} \in \mathbb{R}^n \colon \langle \mathbf{x}, \mathbf{e} \rangle\big\} \supset A_0$ and, for each $i \in \mathbb{Z}$, let $A_i = i\mathbf{v} + A_0$. Notice that for $i \neq j$, the sets A_i and A_j are disjoint—$\mathbf{a}_1 + j\mathbf{v} = \mathbf{a}_2 + i\mathbf{v}$ with $\mathbf{a}_1, \mathbf{a}_2 \in A_0$, implies that $j\mathbf{v} - i\mathbf{v} \in A_0$. Since

$\mathbf{0}$ is the only multiple of $\mathbf{v}$ contained in A_0, it follows that $j = i$. Now, consider $\mathbf{u} \in \mathcal{L}$, and let $i \in \mathbb{Z}$ be such that the distance between the affine space K_i containing A_i and $\mathbf{u}$ is minimal. Take a vector $\mathbf{w} \in A_i$; then $\mathbf{u} - \mathbf{w} \in \mathcal{L}$. Moreover, the distance between $\mathbf{u} - \mathbf{w}$ and $\mathbf{e}^\perp = K_i - \mathbf{w}$ is minimal for vectors in $\mathcal{L} \setminus \mathbf{e}^\perp$. Therefore $\mathbf{u} - \mathbf{w} \notin A_0$, which is a contradiction for the choice of $\mathbf{v}$. Thus the collection $\{A_i \colon i \in \mathbb{Z}\}$ is a partition of $\mathcal{L}$. The result follows. □

The previous claim is a very useful tool to describe some properties of functions that factor over an orthogonal basis. Consider a function $\mathbf{f}(x_1, \ldots, x_n) = \prod_{i\in[n]} f(x_i)$. With the notation used in the claim, let $v = \langle \mathbf{v}, \mathbf{e}_n \rangle$ and $\mathbf{u} = \big(\mathbf{v} - \langle \mathbf{v}, \mathbf{e}_n \rangle \mathbf{e}_n\big) \in \mathbf{e}_n^\perp$. Then we can express the value of $\mathbf{f}$ over a rational lattice $\mathcal{L}$ as

$$\begin{aligned}\sum_{\mathbf{x}\in\mathcal{L}} \mathbf{f}(\mathbf{x}) &= \sum_{i\in\mathbb{Z}} \sum_{\mathbf{x}\in A_0 + i\mathbf{v}} \mathbf{f}(\mathbf{x}) \\ &= \sum_{i\in\mathbb{Z}} \sum_{\mathbf{x}\in A_0 + i\mathbf{v}} f(x_1)\mathbf{g}\big((x_1, \ldots, x_{n-1})\big) \\ &= \sum_{i\in\mathbb{Z}} f(iv) \sum_{\mathbf{x}\in \overline{A}_0 + i\mathbf{u}} \mathbf{g}\big((x_1, \ldots, x_{n-1})\big).\end{aligned} \tag{5.12}$$

The previous expression then allows us, in certain cases, to use induction for generalizing properties that hold for functions over $\mathbb{R}$ to a family of functions over $\mathbb{R}^n$. This can be seen in the following results.

Claim 5.31. Let $f \colon \mathbb{R}^n \to \mathbb{R}$ and consider the product $\mathbf{f} \colon (x_1, \ldots, x_n) \mapsto \prod_{i\in[n]} f(x_i)$. Suppose that there exists $\kappa \in \mathbb{R}$ such that for all $\mathcal{L}_1 \subset \mathbb{Q}$ and all $c \in \mathbb{Q}$, $f(\mathcal{L}_1 + c) \leq \kappa f(\mathcal{L}_1)$. Then, there exists $\kappa' \in \mathbb{R}$ such that for all $\mathcal{L} \subset \mathbb{Q}^n$ and all $\mathbf{c} \in \mathbb{Q}^n$, $\mathbf{f}(\mathcal{L} + \mathbf{c}) \leq \kappa' \mathbf{f}(\mathcal{L})$.

Proof. We proceed by induction on n. For $n = 1$, this is given by hypothesis. Now let $n \in \mathbb{Z}_{>2}$.

Let $\mathcal{L} \subset \mathbb{Q}^n$ be a lattice and let $\mathbf{c} \in \mathbb{Q}^n$. Consider the decomposition $\mathcal{L} = \bigcup_{i\in\mathbb{Z}} A_0 + i\mathbf{v}$ with respect to $\mathbf{e}_n$ described in Claim 5.30. Recall that A_0 is a lattice contained in $\mathbf{e}_n^\perp = \big\{\mathbf{x} \in \mathbb{R}^n \colon \langle \mathbf{x}, \mathbf{e}_n \rangle = 0\big\}$—the hyperplane orthogonal to $\mathbf{e}_n$. Let $\overline{A}_0$ be the lattice $\mathbb{R}^{n-1}$ formed by the vectors $(x_1, \ldots, x_{n-1})$ such that $(x_1, \ldots, x_{n-1}) \in A_0$, and let $\mathbf{g}$ denote the mapping $\mathbf{g}\colon (x_1, \dots, x_{n-1}) \mapsto \prod_{[n-1]} g(x_i)$.

$$\begin{aligned}
\sum_{\mathbf{x}\in\mathcal{L}} \mathbf{f}(\mathbf{x}+\mathbf{c}) &= \sum_{i\in\mathbb{Z}} \sum_{\mathbf{x}\in A_0+i\mathbf{v}} \mathbf{f}(\mathbf{x}+\mathbf{c}) \\
&= \sum_{i\in\mathbb{Z}} f(iv+c_n) \sum_{\mathbf{x}\in\overline{A}_0+i\mathbf{u}} \mathbf{g}\big(\mathbf{x}+(c_1,\dots,c_{n-1})\big) \\
&\le \left(\kappa_1 \sum_{\mathbf{x}\in\overline{A}_0} \mathbf{g}(\mathbf{x})\right) \sum_{i\in\mathbb{Z}} f(iv+c_n) \\
&\le \left(\kappa_1 \sum_{\mathbf{x}\in\overline{A}_0} \mathbf{g}(\mathbf{x})\right) \kappa_2 \sum_{i\in\mathbb{Z}} f(iv) \\
&\le \kappa_1\kappa_2 \sum_{i\in\mathbb{Z}} f(iv) \left(\kappa_3 \sum_{\mathbf{x}\in\overline{A}_0+i\mathbf{u}} \mathbf{g}(\mathbf{x})\right) \\
&= \kappa_1\kappa_2\kappa_3 \sum_{\mathbf{x}\in\mathcal{L}} \mathbf{f}(\mathbf{x}),
\end{aligned} \tag{5.13}$$

as desired. □

Proposition 5.32. *Let $f \in \mathcal{F}_1(\mathbb{Q})$ be a positive function in standard form. Suppose that there exists $\kappa \in \mathbb{R}$ such that for every lattice $A_0 \subset \mathbb{Q}$ and for every shift $\mathbf{c} \in \mathbb{Q}$, $f(A_0+\mathbf{c}) \le \kappa f(A_0)$. Then the mapping given by the product $\mathbf{f}\colon (x_1, \dots, x_n) \mapsto \prod_{i\in[n]} f(x_i)$ is an element of $\mathcal{F}_n(\mathbb{Q}^n)$.*

Proof. We proceed with the proof using induction on n. The case $n = 1$ is then given by our hypothesis on f. Let $n \in \mathbb{Z}_{\ge 2}$.

Let $\mathcal{L} \subset \mathbb{Q}^n$ be a lattice and let $\varepsilon \in \mathbb{R}_{>0}$. Notice that the dual $\mathcal{L}^*$ is also a rational lattice. Consider the decomposition of $\mathcal{L}^* = \bigcup_{i\in\mathbb{Z}} A_0 + i\mathbf{v}$ with respect to $\mathbf{e}_n$ described in Claim 5.30. Let $\overline{A}_0$ be the lattice $\mathbb{Q}^{n-1}$ formed by the vectors $(y_1, \dots, y_{n-1})$ such that $(y_1, \dots, y_{n-1}, 0) \in A_0$, and let $\mathbf{g}$ denote the mapping $\mathbf{g}\colon (x_1, \dots, x_{n-1}) \mapsto \prod_{[n-1]} g(x_i)$. Recall that $\widehat{\mathbf{g}}(y_1, \dots, y_{n-1}) = \prod_{i\in[n]} \widehat{f}(y_i)$. Then, for any $s \in \mathbb{R}_{>0}$,

$$\begin{aligned}
\sum_{\mathbf{y}\in\mathcal{L}^*\setminus\{\mathbf{0}\}} \big|\widehat{\mathbf{f}}_s(\mathbf{y})\big| &= \sum_{i\in\mathbb{Z}\setminus\{0\}} \sum_{\mathbf{y}\in A_0+i\mathbf{v}} \big|\widehat{\mathbf{f}}_s(\mathbf{y})\big| + \sum_{\mathbf{y}\in A_0\setminus\{\mathbf{0}\}} \big|\widehat{\mathbf{f}}_s(\mathbf{y})\big| \\
&= \sum_{i\in\mathbb{Z}\setminus\{\mathbf{0}\}} \big|\widehat{f}_s(iv)\big| \left(\sum_{\mathbf{y}\in\overline{A}_0+i\mathbf{u}} \big|\widehat{\mathbf{g}}_s(\mathbf{y})\big|\right) + f(0) \sum_{\mathbf{y}\in\overline{A}_0\setminus\{\mathbf{0}\}} \big|\widehat{\mathbf{g}}_s(\mathbf{y})\big|.
\end{aligned} \tag{5.14}$$

Where $v = \langle \mathbf{v}, \mathbf{e}_n \rangle$ and $\mathbf{u} = \big(\mathbf{v} - \langle \mathbf{v}, \mathbf{e}_n \rangle \mathbf{e}_n\big) \in \mathbf{e}_n^{\perp}$. Now, it is necessary to bound $\big|\widehat{\mathbf{g}_s}\,\big|(\overline{A}_0 + \mathbf{u}_i)$. To do so, recall that by Proposition 5.3,

$$\sum_{\mathbf{y} \in \overline{A}_0 + \mathbf{u}_i} \big|\widehat{\mathbf{g}_s}(\mathbf{y})\big| = s^{n-1} \sum_{\mathbf{y} \in \overline{A}_0 + \mathbf{u}_i} \big|\widehat{\mathbf{g}}(s\mathbf{y})\big| \leq \kappa^{n-1} s^{n-1} \sum_{\mathbf{y} \in \overline{A}_0} \big|\widehat{\mathbf{g}}(s\mathbf{y})\big| = \kappa^{n-1} \sum_{\mathbf{y} \in \overline{A}_0} \big|\widehat{\mathbf{g}_s}(\mathbf{y})\big|.$$

Then, by Equation (5.14), $\big|\widehat{\mathbf{f}_s}\big|\big(\mathcal{L} \setminus \{\mathbf{0}\}\big)$ is bounded by

$$\begin{aligned} \sum_{\mathbf{y} \in \mathcal{L}^* \setminus \{\mathbf{0}\}} \big|\widehat{\mathbf{f}_s}(\mathbf{y})\big| &\leq \kappa^{n-1} \left(\sum_{\mathbf{y} \in \overline{A}_0} \big|\widehat{\mathbf{g}_s}(\mathbf{y})\big| \right) \sum_{i \in \mathbb{Z} \setminus \{0\}} \big|\widehat{f_s}(v_i)\big| + \sum_{\mathbf{y} \in \overline{A}_0 \setminus \{\mathbf{0}\}} \big|\widehat{\mathbf{g}_s}(\mathbf{y})\big|. \\ &= \kappa^{n-1} \left(1 + \sum_{\mathbf{y} \in \overline{A}_0 \setminus \{\mathbf{0}\}} \big|\widehat{\mathbf{g}_s}(\mathbf{y})\big| \right) \sum_{i \in \mathbb{Z} \setminus \{0\}} \big|\widehat{f_s}(v_i)\big| + \sum_{\mathbf{y} \in \overline{A}_0 \setminus \{\mathbf{0}\}} \big|\widehat{\mathbf{g}_s}(\mathbf{y})\big|. \end{aligned} \tag{5.15}$$

Notice that the functions $\mathbf{f}$ and $\mathbf{g}$ are also in standard form. By induction the hypothesis, the smoothing parameter $\eta_{\mathbf{g},1}$ is finite. Then, in particular, for $s \geq \eta_{\mathbf{g},1}(A_0)$, the weight $\big|\widehat{\mathbf{g}_s}\big|(\overline{A}_0) \leq 2$.

Now, let $\varepsilon_1 = \varepsilon / \left(4\kappa^{n-1}\right)$ and $\varepsilon_2 = \min\{\varepsilon/2, 1\}$, and consider

$$\eta = \max\Big\{ \eta_{\mathbf{g},\varepsilon_1}\big(\langle \mathbf{v}, \mathbf{e}_n \rangle \mathbb{Z}\big), \eta_{f,\varepsilon_2}(A_0) \Big\}.$$

By induction the hypothesis, η is a finite real number since both smoothing parameters in the set are finite. Then, by Equation (5.15) we have that, for $s \geq \eta$,

$$\sum_{\mathbf{y} \in \mathcal{L}^* \setminus \{\mathbf{0}\}} \big|\widehat{\mathbf{f}_s}(\mathbf{y})\big| < \Big(\kappa^{n-1} \min\{2, 1 + \varepsilon_2\} \varepsilon_1 + \varepsilon_2 \Big) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Therefore, the smoothing parameter of $\mathbf{f}$ with respect to $\mathcal{L}$ is $\eta_{\mathbf{f},\varepsilon}(\mathcal{L}) \leq \eta < \infty$, as required. □

5.5 Tail Bounds

The traditional method to find (or bound) the smoothing parameter for Gaussian functions is to bound the proportion of the weight of the function over a lattice that is outside a certain region. This is better known in the literature as the tail bound of the function. The following definition formally captures this intuition.

Definition 5.33 (Tail Bound Functions)**.** Let $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}$ be a continuous function and let $\mathcal{A}$ be a family of discrete sets $A \subset \mathbb{R}^n$. A function $\nu_{\mathbf{f}}$ is said to be a *tail bound* for $\mathbf{f}$ if for every $K \subset \mathbb{R}^n$ and every $A \in \mathcal{A}$,

$$\nu_{\mathbf{f}}(K, A) := \frac{\mathbf{f}(A \setminus K)}{\mathbf{f}(A)}. \tag{5.16}$$

A tail bound for $\mathbf{f}$ and the family $\mathcal{A}$ is defined as

$$\nu_{\mathbf{f},\mathcal{A}}(K) := \sup_{A \in \mathcal{A}} \nu_{\mathbf{f}}(K, A), \tag{5.17}$$

whenever the supremum on the right hand side of (5.17) is well defined.

Notice that the quantity $\nu_{\mathbf{f}}(K, \mathcal{L} + \mathbf{v})$ denotes the fraction of weight of the function $f_{\mathbf{v}}$ over $\mathcal{L}$ outside of the set K. It follows from the definition above that for every arbitrary subset K of $\mathbb{R}^n$ and every lattice $\mathcal{L} \subset \mathbb{R}^n$,

$$\mathbf{f}(\mathcal{L} + \mathbf{v} \setminus K) \leq \nu_{\mathbf{f}}(K)\mathbf{f}(\mathcal{L}).$$

Obtaining A Generic Tail Bound

In [MSD19], Miller and Stephens-Davidowitz provide a construction of a tail bound $\nu_f(K)$ for a class of functions that admit the Poisson summation formula. We now show how to derive bounds on the smoothing parameter in terms of various lattice quantities using similar ideas. We begin by generalizing the main result from [MSD19], by allowing the Fourier transform of the function to present a non-monotonous behavior. To do so we introduce a new parameter that allows to bound the potential growth of the function.

Lemma 5.34 (Tail Bounds for Eventually Decreasing functions)**.** *Let $\mathbf{b}\colon \mathbb{R}^n \to \mathbb{R}_{>0}$ be a continuous real-valued positive function for which the Poisson summation formula holds. Further, assume that its Fourier transform $\widehat{\mathbf{b}}$ is positive and define $\beta\colon \mathbb{R}_{\geq 1} \to \mathbb{R}_{>0}$ as*

$$\beta(t) := \sup_{\mathbf{y} \in \mathbb{R}^n} \frac{\widehat{\mathbf{b}}(t\mathbf{y})}{\widehat{\mathbf{b}}(\mathbf{y})}.$$

Then, for any $\mathbf{v} \in \mathbb{R}^n$ and $K, \mathcal{L} \subset \mathbb{R}^n$, where $\mathcal{L}$ is a lattice and K is any set, we have that

$$\sum_{\substack{\mathbf{x} \in \mathcal{L} \\ \mathbf{x}+\mathbf{v} \notin K}} \mathbf{b}(\mathbf{x} + \mathbf{v}) \geq \nu_{\mathbf{b}}(K) \sum_{\mathbf{x} \in \mathcal{L}} \mathbf{b}(\mathbf{x}),$$

where $\nu_b(K)$ is given by

$$\nu_{\mathbf{b}}(K) = \inf_{u\in(0,1]} \sup_{\mathbf{z}\in\mathbb{R}^n\setminus K} \frac{\beta\big(u^{-1}\big)}{u^n}\frac{\mathbf{b}(\mathbf{z})}{\mathbf{b}(u\mathbf{z})}.$$

Proof. Start by noticing that, by definition of β, we have that for all $\mathbf{y}\in\mathbb{R}^n$ and $t\in\mathbb{R}_{\geq 1}$, $\widehat{\mathbf{b}}(t\mathbf{y}) \leq \beta(t)\widehat{\mathbf{b}}(\mathbf{x})$. Let $\mathcal{L}\subset\mathbb{R}^n$ be a lattice, and let $u\in(0,1]$ and $\mathbf{v}\in\mathbb{R}^n$. By the Poisson summation formula we have that the summation of the shifted and scaled lattice is bounded above by

$$\sum_{\mathbf{x}\in\mathcal{L}} \mathbf{b}\big(u(\mathbf{x}+\mathbf{v})\big) \leq \frac{1}{u^n\det\mathcal{L}}\sum_{\mathbf{y}\in\mathcal{L}^*}\widehat{\mathbf{b}}\big(u^{-1}\mathbf{y}\big) \leq \frac{1}{u^n\det\mathcal{L}}\sum_{\mathbf{y}\in\mathcal{L}^*}\beta\big(u^{-1}\big)\widehat{\mathbf{b}}(\mathbf{y}) = \frac{\beta\big(u^{-1}\big)}{u^n}\sum_{\mathbf{x}\in\mathcal{L}}\mathbf{b}(\mathbf{x}).$$

On the other hand, this same summation is bounded below by

$$\sum_{\mathbf{x}\in\mathcal{L}} \mathbf{b}\big(u(\mathbf{x}+\mathbf{v})\big) \geq \sum_{\substack{\mathbf{x}\in\mathcal{L}\\ \mathbf{x}+\mathbf{v}\notin K}} \mathbf{b}\big(u(\mathbf{x}+\mathbf{v})\big) \geq \inf_{\mathbf{z}\in\mathbb{R}^n\setminus K}\frac{\mathbf{b}(u\mathbf{z})}{\mathbf{b}(\mathbf{z})}\sum_{\substack{\mathbf{x}\in\mathcal{L}\\ \mathbf{x}+\mathbf{v}\notin K}}\mathbf{b}(\mathbf{x}+\mathbf{v})$$

These two inequalities combined imply that for all $u\in(0,1)$,

$$\sum_{\substack{\mathbf{x}\in\mathcal{L}\\ \mathbf{x}+\mathbf{v}\notin K}}\mathbf{b}(\mathbf{x}+\mathbf{v}) \leq \frac{\beta\big(u^{-1}\big)}{u^n}\sup_{\mathbf{z}\in\mathbb{R}^n\setminus K}\frac{\mathbf{b}(\mathbf{z})}{\mathbf{b}(u\mathbf{z})}\sum_{\mathbf{x}\in\mathcal{L}} b(\mathbf{x}).$$

The result follows. □

The parameter function β accounts for the maximum rate of eventual growth of $\widehat{\mathbf{b}}$. When $\widehat{b}$ is monotonically decreasing on rays, β is the constant function 1. In fact, the proof itself does not restrict β to have any particular behavior. However, the utility of the bound obtained in Lemma 5.34 is closely dependent on the behavior of β. Thus, of course, large upper bounds on β would negatively impact the usefulness of the result.

Increasing the Dimension

We now describe a general procedure to transform a tail bound for a function g over $\mathbb{R}$ to the function $\mathbf{g}\colon (x_1,\ldots,x_n)\mapsto\prod_{i\in[n]} g(x_i)$. In the following results consider the following sets. For $r\in\mathbb{R}_{\geq 0}$ and $i\in\{1,\ldots,n\}$ let $rQ_n^{(i)} \coloneqq \big\{\mathbf{x}=(x_1,\ldots,x_n)\in\mathbb{R}^n\colon |x_i|<r\big\}$ and $rB_n^\infty = \big\{\mathbf{x}\in\mathbb{R}^n\colon \|\mathbf{x}\|_\infty\leq r\big\}$.

Lemma 5.35 (From $rQ_n^{(i)}$ to rB_n^∞)**.** *Let $g\colon \mathbb{R} \to \mathbb{R}_{>0}$ be a non-negative function and consider $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}$ defined by $\mathbf{g}(\mathbf{y}) = \prod_{i=1}^n g(y_i)$. Suppose that for all $i \in [n]$, $\nu_g(rQ_n^i)$ is known. Then*

$$\nu_{\mathbf{g}}(rB_n^\infty) \le n \cdot \max_{i\in[n]} \nu_g(rQ_n^i).$$

Proof. We have that

$$\mathcal{L} \setminus rB_n^\infty = \bigcup_{i\in[n]} \left(\mathcal{L} \setminus rQ_n^{(i)}\right).$$

Therefore, since $\mathbf{g}$ is non-negative,

$$\begin{aligned}\mathbf{g}(\mathcal{L} \setminus rB_n^\infty) &= \mathbf{g}\left(\bigcup_{i\in[n]} \left(\mathcal{L} \setminus rQ_n^{(i)}\right)\right)\\ &\le \sum_{i\in[n]} \mathbf{g}\left(\mathcal{L} \setminus rQ_n^{(i)}\right)\\ &\le \sum_{i\in[n]} \nu_g\left(rQ_n^{(i)}\right)\mathbf{g}\left(\mathcal{L}\right)\\ &= n \max_{i\in[n]} \nu_g\left(rQ_n^{(i)}\right)\mathbf{g}\left(\mathcal{L}\right),\end{aligned}$$

as desired. □

In the following theorem we make use of the lattice decomposition for rational lattices described in Claim 5.30 in a different context. This is perhaps not a coincidence, as the following theorem can be thought as the analog result to Proposition 5.32 for tail bound functions. The requirement that the weight of the function has bounded growth when shifted—in other words, that for all $\mathbf{c}$, $\mathbf{f}(\mathcal{L} + \mathbf{c}) \le \kappa\mathbf{f}(\mathcal{L})$—is also present as a hypothesis for the function. Notice, however, that when the function and the dual are both positive, this is automatically satisfied from Proposition 5.7, by taking $\kappa = 1$.

Theorem 5.36. *Let $g\colon \mathbb{R} \to \mathbb{R}_{>0}$ be a non-increasing function with $g(0) = 1$ and consider $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}$ defined by $\mathbf{g}(\mathbf{x}) = \prod_{i=1}^n g(x_i)$. Suppose that there exists a constant $\kappa \in \mathbb{R}_{\ge 1}$ such that for every $\mathbf{c} \in \mathbb{R}^n$ and every lattice $\mathcal{L} \subset \mathbb{Q}^n$, $\mathbf{g}(\mathcal{L} + \mathbf{c}) \le \kappa \cdot \mathbf{g}(\mathcal{L})$. Given a tail bound function ν_g for g we have that, for any n-dimensional lattice $\mathcal{L} \subset \mathbb{Q}^n$,*

$$\nu_{\mathbf{g}}\left(rQ_n^{(i)}, \mathcal{L}\right) = \kappa \cdot g\left((-r, r) \cap \mathcal{L}_1^{(i)}\right) \cdot \nu_g\left((-r, r), \mathcal{L}_1^{(i)}\right),$$

where $\mathcal{L}_1^{(i)}$ denotes the projection of $\mathcal{L}$ to the subspace generated by $\mathbf{e}_i$.

Proof. Let $\mathcal{L} \subset \mathbb{Q}^n$ be a lattice. Without loss of generality, we fix $i = 1$. Consider the lattice decomposition for $\mathcal{L}$ with respect to $\mathbf{e}_1$ described in Claim 5.30,

$$\mathcal{L} = \bigcup_{j \in \mathbb{Z}} A_j,$$

where $A_j = A_0 + j\mathbf{w}$. Let $\mathbf{v} = \langle \mathbf{w}, \mathbf{e}_1 \rangle \mathbf{e}_1$ be the projection of $\mathbf{w}$ onto $\mathbf{e}_1$. Thus $\mathbf{v}$ is a generator for $\mathcal{L}_1 = \mathcal{L}_1^{(1)}$. Let $v = \langle \mathbf{w}, \mathbf{e}_1 \rangle = \|\mathbf{v}\|_2$ and, for $r \in \mathbb{Z}$, let $A_{<r}$ denote the union of the of sets A_j such that $j \in I_r = (-v/r, v/r) \cap \mathbb{Z}$. Then

$$\begin{aligned}
\mathbf{g}(\mathcal{L} \setminus A_{<r}) &= \sum_{j \in I_r} \mathbf{g}(A_j) \\
&= \sum_{j \in I_r} \mathbf{g}(A_0 + j\mathbf{w}) \\
&= \sum_{j \in I_r} \mathbf{g}\big(A_0 + j(\mathbf{w} - \mathbf{v}) + j\mathbf{v}\big) \\
&= \sum_{j \in I_r} \mathbf{g}\big(A_0 + j(\mathbf{w} - \mathbf{v})\big) g(j\mathbf{v}),
\end{aligned}$$

where the last equality is given by the definition of $\mathbf{g}$, since $j(\mathbf{w} - \mathbf{v}) \in \mathbf{e}_1^\perp$ and $j\mathbf{v}$ is a multiple of $\mathbf{e}_1$. Now we bound $\mathbf{g}\big(A_0 + j(\mathbf{w} - \mathbf{v})\big)$. To that end, observe that for every vector $\mathbf{c} \in \mathbf{e}_1^\perp$ we have, by hypothesis,

$$\begin{aligned}
\kappa \mathbf{g}(A_0) \sum_{j \in \mathbb{Z}} \mathbf{g}(j\mathbf{w}) &= \kappa \mathbf{g}\left(\bigcup_{j \in \mathbb{Z}} A_j\right) \\
&= \kappa \mathbf{g}(\mathcal{L}) \\
&\geq \mathbf{g}\big(\mathcal{L} + \mathbf{c}\big) \\
&= \sum_{j \in \mathbb{Z}} \mathbf{g}(A_j + \mathbf{c}) \\
&= \mathbf{g}(A_0 + \mathbf{c}) \sum_{j \in \mathbb{Z}} \mathbf{g}(j\mathbf{w})
\end{aligned}$$

As a consequence, we can bound the value of $\mathbf{g}$ over a shift of K_0 by the same constant κ

times $\mathbf{g}(A_0)$. In particular, $\mathbf{g}\big(A_0 + j(\mathbf{w} - \mathbf{v})\big) \le \kappa \mathbf{g}(A_0)$. Hence,

$$\begin{aligned}
\mathbf{g}(\mathcal{L} \setminus A_{<r}) &= \sum_{j \in I_r} \mathbf{g}\big(A_0 + j(\mathbf{w} - \mathbf{v})\big) g(j\mathbf{v}) \\
&\le \sum_{j \in I_r} \kappa \mathbf{g}(A_0) g(j\mathbf{v}) \\
&= \kappa \mathbf{g}(A_0) \sum_{j \in I_r} g(j\mathbf{v}) \\
&\le \kappa \mathbf{g}(A_0) \cdot \frac{\nu_g\big((-r,r), \mathcal{L}_1\big)}{1 - \nu_g\big((-r,r), \mathcal{L}_1\big)} \cdot g\big((-r,r) \cap \mathcal{L}_1\big) \\
&\le \kappa \mathbf{g}(A_{<r}) \cdot \frac{\nu_g\big((-r,r), \mathcal{L}_1\big)}{1 - \nu_g\big((-r,r), \mathcal{L}_1\big)} \cdot g\big((-r,r) \cap \mathcal{L}_1\big),
\end{aligned} \tag{5.18}$$

where the second to last inequality is given since $A_0 \subseteq A_{<r}$ and

$$\begin{aligned}
\sum_{j \in I_r} g(j\mathbf{v}) &\le \nu_g\big((-r,r), \mathcal{L}_1\big) \sum_{j \in I_r} g(j\mathbf{v}) \\
&= \nu_g\big((-r,r), \mathcal{L}_1\big) \left(g\big((-r,r) \cap \mathcal{L}_1\big) + \sum_{j \in I_r} g(j\mathbf{v}) \right).
\end{aligned}$$

For ease of notation, let R denote the fraction

$$R = \frac{\nu_g\big((-r,r), \mathcal{L}_1\big)}{1 - \nu_g\big((-r,r), \mathcal{L}_1\big)}.$$

With this, the bound for $\mathbf{g}(\mathcal{L} \setminus A_{<r})$ given in (5.18) is given as

$$\begin{aligned}
\mathbf{g}(\mathcal{L} \setminus A_{<r}) &\le \kappa \cdot R \cdot g\big((-r,r) \cap \mathcal{L}_1\big) \cdot \mathbf{g}(A_{<r}) \\
&= \kappa \cdot R \cdot g\big((-r,r) \cap \mathcal{L}_1\big) \big[\mathbf{g}(\mathcal{L}) - \mathbf{g}(\mathcal{L} \setminus A_{<r})\big],
\end{aligned}$$

which yields the following alternate bound,

$$\begin{aligned}
\mathbf{g}(\mathcal{L} \setminus A_{<r}) &\le \kappa \cdot \frac{R}{1 + \kappa \cdot R \cdot g\big((-r,r) \cap \mathcal{L}_1\big)} \cdot g\big((-r,r) \cap \mathcal{L}_1\big) \cdot \mathbf{g}(\mathcal{L}) \\
&\le \kappa \cdot \frac{R}{1 + R} \cdot g\big((-r,r) \cap \mathcal{L}_1\big) \cdot \mathbf{g}(\mathcal{L}) \\
&\le \kappa \cdot \nu_g\big((-r,r), \mathcal{L}_1\big) \cdot g\big((-r,r) \cap \mathcal{L}_1\big) \cdot \mathbf{g}(\mathcal{L}).
\end{aligned}$$

where

$$\frac{R}{1+\kappa\cdot R\cdot g\left((-r,r)\cap\mathcal{L}_1\right)} \le \frac{R}{1+R} = \frac{\frac{\nu_g\big((-r,r),\mathcal{L}_1\big)}{1-\nu_g\big((-r,r),\mathcal{L}_1\big)}}{1+\frac{\nu_g\big((-r,r),\mathcal{L}_1\big)}{1-\nu_g\big((-r,r),\mathcal{L}_1\big)}} = \nu_g\big((-r,r),\mathcal{L}_1\big),$$

as $\kappa \ge 1$ and $g\big((-r,r)\cap\mathcal{L}_1\big) \ge g(0) = 1$. Finally, note that $\mathcal{L}\cap A_{<r} = \mathcal{L}\cap rQ_n^{(1)}$ as both consist of the set of vectors $(x_1,\dots,x_n)\in\mathcal{L}$ with $x_1 < r$. Hence,

$$\mathbf{g}\left(\mathcal{L}\setminus rQ_n^{(1)}\right) = \mathbf{g}(\mathcal{L}\setminus A_{<r}) \le \kappa\cdot g\big((-r,r)\cap\mathcal{L}_1\big)\cdot\nu_g\big((-r,r),\mathcal{L}_1\big)\cdot\mathbf{g}(\mathcal{L}).$$

This finishes the proof. □

Corollary 5.37. *Suppose the function g satisfies all the conditions of Theorem 5.36. Then*

$$\nu_{g,\mathcal{L}}(rB_n^\infty) = n\cdot\kappa\cdot\max_{i\in[n]} g\left((-r,r)\cap\mathcal{L}_1^{(i)}\right)\cdot\nu_{g,\mathcal{L}_1^{(i)}}(-r,r).$$

If the function additionally satisfies the following properties:

1. *For all $\mathbf{y}\in\mathbb{R}^n$, $\widehat{\mathbf{g}}(\mathbf{y})\ge 0$.*
2. *$r \le \lambda_1^\infty(\mathcal{L})$.*

Then

$$\nu_{\mathbf{g},\mathcal{L}}(rQ_n^{(i)}) = \nu_{g,\mathcal{L}_1^{(i)}}(-r,r)$$

and

$$\nu_{g,\mathcal{L}}(rB_n^\infty) = n\cdot\max_{i\in[n]}\nu_{g,\mathcal{L}_1^{(i)}}(-r,r).$$

Proof. Composing Lemma 5.35 and Theorem 5.36 immediately gives

$$\nu_{g,\mathcal{L}}(rB_n^\infty) = n\cdot\max_{i\in[n]}\nu_{\mathbf{g},\mathcal{L}}\left(rQ_n^{(i)}\right) \le n\cdot\kappa\cdot\max_{i\in[n]} g\left((-r,r)\cap\mathcal{L}_1^{(i)}\right)\cdot\nu_{g,\mathcal{L}_1^{(i)}}(-r,r).$$

Additionally, $\widehat{\mathbf{g}}(\mathbf{y})\ge 0$ for all $y\in\mathbb{R}^n$ implies that $\kappa = 1$ (by Proposition 5.7), and $r < \lambda_1^\infty(\mathcal{L})$ implies that $(-r,r)\cap\mathcal{L}_1^{(i)} = \{0\}$ for all i. Hence

$$\max_{i\in[n]} g\left((-r,r)\cap\mathcal{L}_1^{(i)}\right) = g(0) = 1$$

which implies

$$\nu_{\mathbf{g},\mathcal{L}}\left(rQ_n^{(i)}\right) = \nu_{g,\mathcal{L}_1^{(i)}}(-r,r)$$

and

$$\nu_{g,\mathcal{L}}(rB_n^{\infty}) = n \cdot \max_{i\in[n]} \nu_{g,\mathcal{L}_1^{(i)}}(-r,r)$$

if the additional conditions are satisfied. □

Remark 5.38. If lattice invariant tail bounds are known, the corollary gives us

$$\nu_{\mathbf{g},\mathcal{L}}\left(rQ_n^{(i)}\right) = \nu_{\mathbf{g}}\big((-r,r)\big)$$

and

$$\nu_{g,\mathcal{L}}(rB_n^{\infty}) = n \cdot \nu_{\mathbf{g}}\big((-r,r)\big).$$

5.6 Bounding the Smoothing Parameter

The first application of the tools developed in the previous sections is to bound the smoothing parameter of a function.

Lemma 5.39 (Tail Bounds from Tail Bounds on Bounding Functions)**.** *Let $K \subset \mathbb{R}^n$ be an arbitrary subset. Let $\mathcal{J}$ be a family of lattices in $\mathbb{R}^n$ and consider functions $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{b} \in L^1(\mathbb{R}^n)$ satisfying the following properties.*

1. *For all $\mathbf{y} \in \mathbb{R}^n$, $\big|\mathbf{f}(\mathbf{y})\big| \leq \mathbf{b}(\mathbf{y})$.*
2. *There exists a constant C such that for every lattice $\mathcal{L} \in \mathcal{J}$, $\mathbf{b}(\mathcal{L}) \leq C \cdot |\mathbf{f}|(\mathcal{L})$.*

Then for all $\mathcal{L} \in \mathcal{J}$ and for all $\mathbf{c} \in \mathbb{R}^n$,

$$\nu_{|\mathbf{f}|}(K, \mathcal{L} + \mathbf{c}) \leq C \cdot \nu_{\mathbf{b}}(K, \mathcal{L} + \mathbf{c}).$$

Proof. Let $\mathcal{L} \in \mathcal{J}$ and let $\mathbf{c} \in \mathbb{R}^n$. By the hypotheses of the lemma we have that

$$\begin{aligned} |\mathbf{f}|(\mathcal{L} + \mathbf{c} \setminus K) &\leq \mathbf{b}(\mathcal{L} + \mathbf{c} \setminus K) \\ &\leq \nu_{\mathbf{b}}(K)\mathbf{b}(\mathcal{L}) \\ &= \left(\nu_{\mathbf{b}}(K)\frac{\mathbf{b}(\mathcal{L})}{|\mathbf{f}|(\mathcal{L})}\right)|\mathbf{f}|(\mathcal{L}) \\ &\leq C \cdot \nu_b(K) \cdot |\mathbf{f}|(\mathcal{L}). \end{aligned}$$

The result follows. □

Lemma 5.40. *Let $\mathcal{L}, K \subset \mathbb{R}^n$, where $\mathcal{L}$ is a lattice and K is an arbitrary set. Consider a function $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}$ such that $\mathbf{g}(\mathbf{0}) = 1$ and let $\nu_{\mathbf{g}}$ be a tail bound function for $\mathbf{g}$. Let $\varepsilon \in (0,1)$ and assume that there exists $r \in \mathbb{R}_{>0}$ such that*

$$\nu_{\mathbf{g}}(rK, \mathcal{L}) \leq \frac{\varepsilon}{1+\varepsilon}.$$

Then,

$$\mathbf{g}\left(\left(\frac{r}{\lambda_1(K,\mathcal{L})}\right)\mathcal{L} \setminus \{\mathbf{0}\}\right) < \varepsilon. \tag{5.19}$$

Proof. Let t denote $\lambda_1(K, \mathcal{L})$. Since $tK \cap \mathcal{L} = \{\mathbf{0}\}$, it follows that for all $s \in \mathbb{R}_{>0}$,

$$\mathbf{g}_{1/s}\big(\mathcal{L} \setminus \{\mathbf{0}\}\big) = \mathbf{g}\big(s\mathcal{L} \setminus \{\mathbf{0}\}\big) = \mathbf{g}(s\mathcal{L} \setminus stK).$$

Consider $s = \frac{r}{t}$.

$$\begin{aligned}\mathbf{g}_{1/s}\big(\mathcal{L} \setminus \{\mathbf{0}\}\big) &= \mathbf{g}(s\mathcal{L} \setminus stK)\\ &\leq \nu_{\mathbf{g}}\big(stK, \mathcal{L}\big)\mathbf{g}(s\mathcal{L})\\ &= \nu_{\mathbf{g}}\big(rK, \mathcal{L}\big)\Big(\mathbf{g}(\mathbf{0}) + \mathbf{g}\big(s\mathcal{L} \setminus \{\mathbf{0}\}\big)\Big).\end{aligned}$$

As $\mathbf{g}(\mathbf{0}) = 1$, using the previous inequality we have that

$$\mathbf{g}\left(\left(\frac{r}{\lambda_1(K,\mathcal{L})}\right)\mathcal{L} \setminus \{\mathbf{0}\}\right) = \mathbf{g}_{1/s}\big(\mathcal{L} \setminus \{\mathbf{0}\}\big) \leq \frac{\nu_{\mathbf{g}}\big(rK, \mathcal{L}\big)}{1 - \nu_{\mathbf{g}}\big(rK, \mathcal{L}\big)} < \frac{\frac{\varepsilon}{1+\varepsilon}}{1 - \frac{\varepsilon}{1+\varepsilon}} = \varepsilon, \tag{5.20}$$

as required. □

Corollary 5.41. *Let $\mathcal{L} \subset \mathbb{R}^n$, where $\mathcal{L}$ is a lattice and K is an arbitrary set. Let $\mathbf{f} \in \mathcal{D}_n$ and let $\mathbf{g} = \left|\widehat{D_{\mathbf{f}}}\right|$. Consider a tail bound function $\nu_{\mathbf{g}}$ for $\mathbf{g}$. Suppose that for $\varepsilon \in (0,1)$ there exists $r \in \mathbb{R}_{>0}$ such that*

$$\nu_{\mathbf{g}}(rK, \mathcal{L}) \leq \frac{\varepsilon}{1+\varepsilon}.$$

Then for $s = r/\lambda_1(K, \mathcal{L})$, the function f_s is ε-smoothening for the lattice $\mathcal{L}$. Moreover, if the mapping $s' \mapsto \nu_{\mathbf{g}}(s'K, \mathcal{L})$ is monotonically decreasing, then the smoothing parameter $\eta_{\mathbf{f},\varepsilon} < s$.

Proof. By Proposition 5.3, Equation (5.19) is equivalent to

$$\left|\widehat{D_{\mathbf{f},s}}\right|\big(\mathcal{L} \setminus \{\mathbf{0}\}\big) = \mathbf{g}_{1/s}\big(\mathcal{L} \setminus \{\mathbf{0}\}\big) < \varepsilon.$$

Thus $\mathbf{f}$ is ε-smoothening. Suppose now that the mapping $s' \mapsto \nu_{\mathbf{g}}(s'K, \mathcal{L})$. Then, by Equation (5.20), for $s' > s$

$$\mathbf{g}_{1/s'}\big(\mathcal{L} \setminus \{\mathbf{0}\}\big) \leq \frac{\nu_{\mathbf{g}}\big(s'tK, \mathcal{L}\big)}{1 - \nu_{\mathbf{g}}\big(s'tK, \mathcal{L}\big)} \leq \frac{\nu_{\mathbf{g}}\big(stK, \mathcal{L}\big)}{1 - \nu_{\mathbf{g}}\big(stK, \mathcal{L}\big)} < \varepsilon,$$

where $t = \lambda_1(K, \mathcal{L})$. The proof follows. □

5.7 Smoothing Parameter for Non-Gaussian Functions

In this section we quantify the smoothing parameter for an infinite set of non-Gaussian functions by using the tools developed in the previous sections. We start by giving an overview of the general procedure starting with an arbitrary function in $\mathcal{D}_1$—the set of real valued functions over $\mathbb{R}$ which are suitable for the application of the Poisson Summation Formula. Next we provide a short example by computing the smoothing parameter of a function for which a tail bound has been previously computed. We finalize this section by computing the smoothing parameter for the family of generalized Gaussian distributions.

An Overview of the Process

We now provide an overview of the general methodology for finding smoothing parameters This description is used as a reference in the forthcoming sections of this chapter. Every step of the sequence essentially subsumes all the previous ones; that is, depending on the available information, it is possible to start the process at any given step. For instance, finding a bounding function (Step 1) is not necessary if we know an appropriate representation for the Fourier transform. Similarly, the availability of $\nu_{\widehat{\mathbf{f}}}(rB_n^\infty)$ bypasses the need to obtain $\nu_{\widehat{\mathbf{f}}}\left(rQ_n^{(i)}\right)$. The evolution of the information is then given by the following flow chart.

$$b(y) \xrightarrow{5.34} \nu_b\big((-r, r)\big) \xrightarrow{5.39} \nu_{\widehat{f}}\big((-r, r)\big) \xrightarrow[*]{5.37} \nu_{\widehat{\mathbf{f}}}\left(rQ_n^{(i)}, \mathcal{L}\right) \xrightarrow{5.35} \nu_{\widehat{\mathbf{f}}}(rB_n^\infty, \mathcal{L}) \xrightarrow{5.41} \eta_{\mathbf{f},\varepsilon}(\mathcal{L}).$$

The arrow marked with * marks the transition from 1-dimensional function to an n-dimensional one.

In the following steps, let $f \in \mathcal{D}_1$ be a positive normalized function and consider $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}$ given by $\mathbf{f}(\mathbf{x}) = \prod_{i\in[n]} f(x_i)$. Let K denote the set $(-r, r) \subset \mathbb{R}$, where $r \in \mathbb{R}_{>0}$ is fixed in Step 4.

Step 1 We start by finding a suitable bounding function for $\left|\widehat{f}\right|$, that is, a function b satisfying the conditions of Theorem 5.34 (in other words, the function is Poisson admissible and has a positive Fourier transform) and such that for all $y \in \mathbb{R}$, $b(y) \geq \left|\widehat{f}(y)\right|$. This may require manipulation and piece-wise completion such that we can evaluate the expression

$$\nu_{\mathbf{b}}(K) = \inf_{0<u\leq 1} \sup_{\mathbf{y}\in\mathbb{R}^n\setminus K} \beta(u^{-1})\frac{\mathbf{b}(\mathbf{y})}{u^n\mathbf{b}(u\mathbf{y})}, \quad \text{where } \beta(u^{-1}) = \sup_{y\in\mathbb{R}} \frac{\widehat{b}(u^{-1}y)}{b(y)}.$$

Step 2 Obtain an upper-bound on $\nu_{\widehat{f}}(K) \leq C \cdot \nu_b(K)$. Start by finding $C_1, C_2 \in \mathbb{R}^n$ such that

$$C_1 \cdot b(\mathcal{L}) \leq b(\mathcal{L} \cap K) \leq C_2 \cdot |\widehat{f}|(\mathcal{L} \cap K).$$

From the expression above we obtain $b(\mathcal{L}) \leq \frac{C_2}{C_1}|\widehat{f}|(\mathcal{L})$ and the bound follows from Lemma 5.39.

Remark 5.42. Notice that $b(\mathcal{L}\cap K)/b(\mathcal{L}) = 1-\nu_b(K,\mathcal{L})$. Thus finding C_1 is equivalent to finding an upper bound on $\nu_b(K,\mathcal{L})$.

Step 3 Use Corollary 5.37 to lift the tail bound $\nu_{|\widehat{f}|}\big((-r,r)\big)$ to the function $\eta_{|\widehat{\mathbf{f}}|}\left(rB_n^\infty, \mathcal{L}\right)$ (passing through $\eta_{|\widehat{\mathbf{f}}|}\left(rQ_n^{(i)}, \mathcal{L}\right)$). Note that the choice of r in Step 4 always satisfies the extra condition of Corollary 5.37.

Step 4 Use the tail bound $\nu_{|\widehat{\mathbf{f}}|}(rB_n^\infty, \mathcal{L})$ to find a scaling of the lattice s such that $s\mathcal{L}^* \cap rB_n^\infty = \mathbf{0}$. Then use the tail bound from Step 3 (as $r \leq \lambda_1^\infty(s\mathcal{L}^*)$ by construction) to obtain a smoothing parameter for $\mathbf{f}$. This is formalized in Corollary 5.41.

Note 5.43. Notice that Step 3 implicitly requires that the function $\mathbf{f}$ can be written as $\mathbf{f}(\mathbf{x}) = \prod_{i\in[n]} f(x_i)$. This property of $\mathbf{f}$, however, is not reflected in Step 4, since Corollary 5.41 may be applied to an arbitrary function.

Smoothing Parameter for $1/(1+2\cosh)$

Throughout this subsection, let $\mathbf{f}$ denote the function

$$\mathbf{f}\colon (x_1,\ldots,x_n) \mapsto \prod_{i\in[n]} \frac{1}{1+2\cosh\left(2\pi x_i/\sqrt{3}\right)}, \tag{5.21}$$

where $\cosh(x) = \frac{1}{2}(e^x + e^{-x})$ is the hyperbolic cosine. This function is used in [MSD19] to obtain a transference theorem for the ℓ_1 norm—this is since $\mathbf{f}$ is approximately ℓ_1 symmetrical.

For $C^* \approx 0.425$ and $\alpha \geq \frac{\sqrt{3}}{2\pi}$, let

$$K_\alpha \coloneqq \left\{\mathbf{x} \in \mathbb{R}^n \colon \|\mathbf{x}\|_1 \leq (1+C^*)\alpha n\right\}$$

be the ℓ_1 ball of radius $(1+C^*)\alpha$. By [MSD19, Lemma 3.7],

$$\nu_{\mathbf{f}}(K_\alpha) \leq \left(\frac{2\pi\alpha}{\sqrt{3}}\right)^n e^{-\left(\frac{2\pi\alpha}{\sqrt{3}}-1\right)n} = \theta(\alpha)^n,$$

with $\theta(\alpha) = \left(\frac{2\pi\alpha}{\sqrt{3}} e^{-\left(\frac{2\pi\alpha}{\sqrt{3}}-1\right)}\right)$. By definition, this means that

$$\mathbf{f}(\mathcal{L} \setminus K_\alpha) \leq \nu_{\mathbf{f}}(K_\alpha)\mathbf{f}(\mathcal{L}). \tag{5.22}$$

One of the more important properties of $\mathbf{f}$ is the fact that it is self-reciprocal. In other words, the function

$$f\colon x \mapsto \frac{1}{1+2\cosh\left(2\pi x/\sqrt{3}\right)}$$

is its own Fourier transform [Tit48, pages 262,263]. In the following proposition, we use this fact to convert the expression in Equation (5.22) for a tail bound of $\mathbf{f}$ into a smoothing parameter for $\mathbf{f}$, for a particular choice of ε.

Proposition 5.44. *Let $\mathcal{L} \subset \mathbb{R}^n$ be a lattice and let $\varepsilon = 2^{-n}$. Then the smoothing parameter of $\mathbf{f}$ with respect to $\mathcal{L}$ is bounded as*

$$\eta_{\mathbf{f},\varepsilon}(\mathcal{L}) \leq \frac{(1+C^*)n}{\lambda_1^{(1)}(\mathcal{L}^*)}.$$

Proof. Let $\varepsilon = 2^{-n}$. Note that $K_\alpha = (1+C^*)\alpha n B_n^1$ by definition, and consider the equation

$$\nu_{\mathbf{f}}\big((1+C^*)nB_n^1\big) = \nu_{\mathbf{f}}(K_\alpha) \leq \left(\frac{2\pi\alpha}{\sqrt{3}}\right)^n e^{-\left(\frac{2\pi\alpha}{\sqrt{3}}-1\right)n} \leq \frac{\varepsilon}{1+\varepsilon} = \frac{2^{-n}}{1+2^{-n}}.$$

Note that for $\alpha = 1$, we have that

$$\left(\frac{2\pi\alpha}{\sqrt{3}}\right)^n e^{-\left(\frac{2\pi\alpha}{\sqrt{3}}-1\right)n} = \left(\frac{2\pi}{\sqrt{3}} e^{-\left(\frac{2\pi}{\sqrt{3}}-1\right)}\right)^n < \left(\frac{1}{3}\right)^n < \frac{2^{-n}}{1+2^{-n}},$$

so the equation holds for any $\alpha \geq 1$. Thus, by Corollary 5.41, as $K_{\alpha=1} = (1+C^*)nB_n^1$, we have that

$$\eta_{\mathbf{f},\varepsilon}(\mathcal{L}) \leq \frac{(1+C^*)n}{\lambda_1^{(1)}(\mathcal{L}^*)},$$

as required. □

In the following proposition we provide an alternative bound for the smoothing parameter of $\mathbf{f}$ that works for any $\varepsilon \in \mathbb{R}_{>0}$. To express this bound we use the product-logarithm function W.[3] This is the inverse of the mapping $z \mapsto ze^z$. Over the reals, the W function is divided into two branches. Given the nature of the relevant quantities, we are only interested on the branch -1. For a comprehensive study of this function see [CGH+96].

Proposition 5.45. *Let $\mathcal{L} \subset \mathbb{R}^n$ and let $\varepsilon \in \mathbb{R}_{>0}$. Then the smoothing parameter of $\mathbf{f}$ with respect to $\mathcal{L}$ is bounded as*

$$\eta_{\mathbf{f},\varepsilon}(\mathcal{L}) \leq \frac{-\sqrt{3}(1+C^*)}{2\pi\lambda_1^\infty(\mathcal{L})} W_{-1}\left(-\frac{1}{e}\left(\frac{\varepsilon}{n(1+\varepsilon)}\right)\right).$$

Proof. The function $\mathbf{f}$ is written as the product of f evaluated on each coordinate; thus it is enough to find a tail bound for f. As before, by [MSD19, Lemma 3.7], a one-dimensional tail bound for f is given by

$$\nu_f\big((1+C^*)\alpha B_1^1\big) \leq \left(\frac{2\pi\alpha}{\sqrt{3}}\right) e^{-\left(\frac{2\pi\alpha}{\sqrt{3}}-1\right)}.$$

Then by Corollary 5.37, an n-dimensional tail bound is expressed as

$$\nu_{\mathbf{f}}\big((1+C^*)\alpha B_n^\infty\big) \leq n\left(\frac{2\pi\alpha}{\sqrt{3}}\right) e^{-\left(\frac{2\pi\alpha}{\sqrt{3}}-1\right)}.$$

[3] In the literature, the function W is also known as the *Lambert W function.*

We now use this tail bound function to find a bound for the smoothing parameter. Bounding the right hand side of the previous inequality by $\frac{\varepsilon}{1+\varepsilon}$ yields

$$-\left(\frac{2\pi\alpha}{\sqrt{3}}\right)e^{-\left(\frac{2\pi\alpha}{\sqrt{3}}\right)} > -\left(\frac{\varepsilon}{ne(1+\varepsilon)}\right).$$

By applying the (branch -1 of the) product logarithm function W_{-1} we have

$$\alpha > \frac{-\sqrt{3}}{2\pi}W_{-1}\left(-\frac{1}{e}\left(\frac{\varepsilon}{n(1+\varepsilon)}\right)\right).$$

Notice that when evaluating both sides on W_{-1} the direction of the inequality is inverted since this branch is a strictly decreasing function. Finally, applying Corollary 5.41 we obtain that the smoothing parameter is bounded by

$$\eta_{\mathbf{f},\varepsilon}(\mathcal{L}) \leq \frac{-\sqrt{3}(1+C^*)}{2\pi\lambda_1^{\infty}(\mathcal{L})}W_{-1}\left(-\frac{1}{e}\left(\frac{\varepsilon}{n(1+\varepsilon)}\right)\right),$$

as required. □

Smoothing Parameter for Generalized Gaussians

In [MSD19], Miller and Stephens-Davidowitz successfully compute a tail bound for a set of p-supergaussians—a set of functions that generalize the Gaussian distribution. Their work strongly depends on the assumption that the Fourier transform is a positive function. In this section we consider the p-supergaussians for $p \in 2\mathbb{Z}_{>0}$, where we do not enjoy such a guarantee. As a result, we are able to compute the smoothing parameter for an infinite family of functions, each of which is symmetric over an ℓ_p norm.

Definition 5.46 (p-supergaussian)**.** Fix $p \in \mathbb{R}_{>0}$. For $s \in \mathbb{R}_{>0}$ and $\mathbf{c} = (c_1, \ldots, c_n) \in \mathbb{R}^n$, the *$p$-supergaussian* (or *supergaussian of degree p*) centered at $\mathbf{c}$ of width s is the function $\rho^{[p]}_{\mathbf{c},s} \colon \mathbb{R}^n \to \mathbb{R}$ defined for $\mathbf{x} = (x_1, \ldots, x_n)$ as

$$\rho^{[p]}_{\mathbf{c},s}(\mathbf{x}) \coloneqq \exp\left(-\frac{1}{s^p}\sum_{i=1}^{n}|x_i - c_i|^p\right). \tag{5.23}$$

For the sake of simplification, in the rest of this work the function $\rho^{[p]}_{\mathbf{0},1}$ is denoted as $\rho^{[p]}$.

In the literature, (the normalized versions of) these functions are also commonly known as the *generalized normal distributions*. We use the name "supergaussians" to be relatively consistent with the nomenclature in [MSD19]. The traditional Gaussian distribution is obtained by setting $p = 2$. Its 1-norm is given by $\int_{\mathbb{R}} f_s^{[p]}(x)dx = \frac{2s\Gamma(1/p)}{p}$. Similar to the Gaussian function, the functions in this family show a nice interplay with the ℓ_p norms.

Proposition 5.47. *The p-supergaussian function $\rho^{[p]}$ is symmetric with respect to the homogeneous function $\|\cdot\|_p$.*

Proof. It follows from $\rho^{[p]}(\mathbf{x}) = \exp\left(-\sum_{i\in[n]}|x_i|^p\right) = \exp\big(-\|\mathbf{x}\|_p^p\big)$. □

Proposition 5.48. *The p-supergaussian function $\rho^{[p]}$ factors over the canonical basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$.*

Proof. Let $\mathbf{x} = \sum_{i\in[n]} x_i\mathbf{e}_i$. Then we have that

$$\rho^{[p]}(\mathbf{x}) = \exp\left(-\sum_{i\in[n]}|x_i|^p\right) = \prod_{i\in[n]} \exp\big(-|x_i|^p\big) = \prod_{i\in[n]} \rho^{[p]}(x_i),$$

as desired. □

An inductive argument is enough to prove that for any $p \in 2\mathbb{Z}_{\geq 1}$, the p-supergaussian is a Schwartz function. For the rest of this document we refer to a function of this form as an *even degree supergaussian*.

As has been argued in this chapter, the Fourier transform of a function is fundamental for study the smoothing properties of a function. For the p-supergaussian, in this thesis we denote the Fourier transform of $\rho^{[p]}$ as

$$\sigma_p(y) \coloneqq \widehat{\rho^{[p]}}(y) = \int_{\mathbb{R}} e^{-x^p}e^{-2\pi ixy}dx. \tag{5.24}$$

Proposition 5.49. *Let $p \in 2\mathbb{Z}_{>0}$. Then the Fourier transform of $\rho^{[p]}$ is asymptotically approximated by the function*

$$\sigma_p(y) \approx \sqrt{\frac{(p-1)}{y^{(p+2)/(p-1)}}\left(\frac{p}{2\pi}\right)^{3/(p-1)}}$$
$$\sum_{k=0}^{(p-2)/2} \exp\left(i\pi\left(\frac{p-2-4k}{4(p-1)}\right) - (p-1)\left(\frac{2\pi y}{p}\right)^{p/(p-1)} e^{i\pi\frac{p-2-4k}{2(p-1)}}\right).$$

We compute the previous approximation in Appendix A by means of an asymptotic approximation technique. Specifically, we make use of the saddle point method. Please refer to the appendix for a full proof of the proposition, along with an overview of the technique.

Step 1—Finding a Bound for σ_p. We begin by finding a bounding function for $|\sigma_p|$ and computing its tail bounds. Using a saddle-point method described in Appendix A, we can show that, for all $p \geq 4$ and for all $y \neq 0$,

$$|\widehat{f^{[p]}}|(y) = |\sigma_p(y)| \approx \xi(p)|y|^{-\alpha(p)}e^{-\vartheta(p)|y|^{\tau(p)}} \tag{5.25}$$

where

$$\alpha(p) = \frac{(p+2)}{2(p-1)}, \quad \vartheta(p) = (p-1)\cos\left(\pi\frac{p-2}{2(p-1)}\right)\left(\frac{2\pi}{p}\right)^{\frac{p}{p-1}}, \quad \tau(p) = \frac{p}{p-1}$$

and

$$\xi(p) = \frac{p-2}{2}\sqrt{(p-1)\left(\frac{p}{2\pi}\right)^{3/(p-1)}}$$

To ease the notation, when p is clear from context, we sometimes use α, ϑ, τ to denote $\alpha(p)$, $\vartheta(p)$, $\tau(p)$, respectively.

Note 5.50. There are several limitations with the approximation of σ_p obtained in Appendix A. As such, the function described in Equation (5.25) is only an asymptotic approximation, that is not itself a function that bounds $|\sigma_p|$. The distance between the obtained function and σ_p is asymptotically decreasing; however, the proportion between the two grows with $|y|$. Nonetheless, we have experimental evidence that this growth is, at most, linear in $|y|$, which leads us to conjecture that this is the case. Thus, assuming that this conjecture is indeed true, we obtain a bounding function by a slight adjustment in the parameter $\alpha(p)$.

As discussed at the beginning of this section, we define the bounding function as the following continuous piecewise completion

$$b_p(y) := \xi(p) \cdot \begin{cases} b_1(y) := e^{-\vartheta|y|^\tau} & |y| \leq 1 \\ b_2(y) := |y|^{-\alpha}e^{-\vartheta|y|^\tau} & |y| \geq 1 \end{cases}$$

where $\alpha, \vartheta, \tau \in \mathbb{R}_{>0}$ are defined as above.

We now proceed to compute the tail bounds for the bounding function $b^{[p]}$. By Lemma 5.34, a tail bound for the function b is given by

$$\nu_{b_p}\big([-r,r]\big) = \inf_{0<u\leq 1}\sup_{|y|\geq r}\beta(u^{-1})\frac{b_p(y)}{ub_p(uy)}. \tag{5.26}$$

Note 5.51. Computing the exact value of $\beta(u^{-1})$ would require the exact computation of $\widehat{b}$. This is, naturally, a big obstacle to obtaining an exact expression for ν_{b_p}. However, we can make some observations about this function. Since $\widehat{b}$ is positive and eventually decreasing, we expect β to be decreasing. Moreover, based on our experiments we conjecture that its value close to 1 is bounded. Thus, in our computations, we regard β as some fixed constant, and use $\beta = 1$ when evaluating expressions. However, note that the final value of any evaluated expression needs to be adjusted since $\beta > 1$.

Evaluating Equation (5.26), we find that

$$\nu_{b_p}\big([-r,r]\big) \leq \beta\left(\frac{1}{r}\right)^{\alpha-1}\exp\big(-\vartheta(r^{\tau}-1)\big). \tag{5.27}$$

The computations are summarized in Appendix B. Recall that $\overline{\sigma_p}$ denotes the Fourier transform of the standard form of the supergaussian (see the notation part of Section 5.1). It follows that $\overline{\sigma_p}$ is bounded by $\overline{b_p}(y) := \frac{1}{a_p}b_p\left(\frac{y}{a_p}\right)$. Using Equation (5.27), a tail bound for $\overline{b_p}$ is given by

$$\nu_{\overline{b_p}}\big([-r,r]\big) = \frac{1}{a_p}\cdot\nu_b\left(\left[-\frac{r}{a_p},\frac{r}{a_p}\right]\right) = \frac{\beta}{a_p}\left(\frac{a_p}{r}\right)^{\alpha-1}\cdot\exp\left(-\vartheta\left(\frac{r}{a_p}\right)^{\tau}+\vartheta\right). \tag{5.28}$$

Step 2—A Tail Bound Function for σ_p. The next step is to convert a tail bound of b_p to a tail bound of $|\sigma_p|$. As described at the beginning of this section, the idea is to use Lemma 5.39. Thus the goal is to find a constant $C \in \mathbb{R}_{>0}$ such that for every lattice $\mathcal{L}_1 \subset \mathbb{R}$ for which $\lambda_1(\mathcal{L}_1) > r$,

$$\overline{b_p}(\mathcal{L}) \leq C\cdot|\overline{\sigma_p}|(\mathcal{L}). \tag{5.29}$$

Once C is found we obtain the bound

$$\nu_{|\overline{\sigma_p}|}\big([-r,r]\big) \leq C\cdot\nu_{\overline{b_p}}\big([-r,r]\big) = C\frac{\beta}{a_p}\left(\frac{a_p}{r}\right)^{\alpha-1}\cdot\exp\left(-\vartheta\left(\frac{r}{a_p}\right)^{\tau}+\vartheta\right). \tag{5.30}$$

Remark 5.52. Notice that the defining property of C, which is Equation (5.29), must hold for every lattice $\mathcal{L}$. Thus, after scaling both sides of the inequality, as well as every lattice by a_p, it is equivalent to consider

$$b_p(\mathcal{L}) \leq C \cdot |\sigma_p|(\mathcal{L}). \tag{5.31}$$

Note 5.53. We now follow the methodology described at the beginning of this section. Following a naive method for finding $C = C_2/C_1$, the more technically challenging part becomes finding an appropriate value for C_1. In the process below, we start by setting a value for C_1. We use this and the bound described in Equation (5.27) to get a condition for the set $K = (-r, r) \subset \mathbb{R}$ for which we can apply Lemma 5.39. We note that only by forcing the bound for ν_{b_p} in Equation (5.27) to be non-trivial (that is, less than 1) can we establish a similar condition on K. As a consequence, we are only able to provide a statement valid for a fixed family of lattices with a certain property.

Fix $\delta \in (0, 1)$. Bounding the right hand side of Equation (5.27) by δ, we obtain

$$\vartheta + \ln(\beta\delta^{-1}) < \vartheta r^\tau + (\alpha - 1)\ln r. \tag{5.32}$$

It is difficult to obtain an expression for r in terms of p from the above inequality. To obtain an alternative inequality, first observe that for all $r \in \mathbb{R}_{>0}$, $\ln r < r^\tau$. Moreover, for $r > e^{1/\tau} > 1$, the fraction $r^\tau/\ln r$ is strictly increasing and approaches infinity as r increases. Consider $r_1 \in \mathbb{R}_{>1}$ such that for all $r \geq r_1$,

$$\frac{r^\tau}{\ln r} > \frac{-2(\alpha - 1)}{\vartheta}. \tag{5.33}$$

Notice that $\alpha < 1$, thus the right hand side is positive. Under this restriction, we then have that for all $r > r_1$,

$$\frac{-\vartheta}{2} r^\tau < (\alpha - 1)\ln r. \tag{5.34}$$

On the other hand, let

$$r_2 = \left(\frac{2\big(\vartheta + \ln(\beta\delta^{-1})\big)}{\vartheta} \right)^{1/\tau}. \tag{5.35}$$

Since r^τ is an increasing function of r, we have that for all $r > r_2$,

$$\vartheta + \ln(\beta\delta^{-1}) < \frac{\vartheta}{2} r^\tau. \tag{5.36}$$

Putting equations (5.34) and (5.36) together, we have that for all $r > r_0 \coloneqq \max\{r_1, r_2\}$,

$$\vartheta + \ln(\beta\delta^{-1}) < \frac{\vartheta}{2} r^\tau = \vartheta r^\tau - \frac{\vartheta}{2} r^\tau < \vartheta r^\tau + (\alpha - 1)\ln r, \tag{5.37}$$

thus satisfying Equation (5.32).

Notice that r_0 is a function of p. Since $\vartheta(p) \to 0$ as $p \to \infty$, the behavior of r_0 is dominated by $\left(\frac{1}{\vartheta}\right)^{1/\tau}$ as p grows. Now observe that for $p \in \mathbb{Z}_{>1}$,

$$\begin{aligned}
\vartheta(p) &= (p-1)\cos\left(\pi\frac{p-2}{2(p-1)}\right)\left(\frac{2\pi}{p}\right)^{\frac{p}{p-1}} \\
&= (p-1)\sin\frac{\pi}{2(p-1)}\left(\frac{2\pi}{p}\right)^{\frac{p}{p-1}} \\
&\sim (p-1)\left(\frac{\pi}{2(p-1)}\right)\left(\frac{2\pi}{p}\right)^{\frac{p}{p-1}} \\
&= \Theta\left(p^{-\frac{p}{p-1}}\right).
\end{aligned}$$

Hence,

$$r_0 = O\left(p^{\frac{p}{p-1}}\right)^{\frac{1}{\tau(p)}} = O\left(p^{\frac{p}{p-1}}\right)^{\frac{p-1}{p}} = O(p).$$

We conclude that, by letting $K = [-r_0, r_0]$, we have $C_1 = \frac{1}{2}$. To find C_2, recall that $K \cap \mathcal{L}_1 = \{0\}$ in the context of finding a smoothing parameter. Hence, we have

$$C_2 \le \frac{b_p(0)}{\sigma_p(0)} \le \frac{\xi(p)\cdot 1}{1} = \xi(p).$$

Note that a slightly more complex argument allows us to find a C_2 that is lattice invariant, leading to lattice invariant tail bounds.

Substituting the above estimate for r_0, C_1, and the definition of $b_p(0)$, we get

$$\begin{aligned}
C &= \frac{C_2}{C_1} \\
&\le 2\xi(p) \\
&= 2\cdot\frac{(p-2)}{2}\sqrt{(p-1)\left(\frac{p}{2\pi}\right)^{3/(p-1)}} \\
&= O\left((p-2)\sqrt{2(p-1)}\right) \\
&= O\left(p^{1.5}\right).
\end{aligned}$$

Thus, for any lattice $\mathcal{L}_1 \subset \mathbb{R}$ with $\lambda_1(\mathcal{L}_1) > r$,

$$\nu_{|\sigma_p|}\big((-r,r),\mathcal{L}_1\big) \leq C \cdot \nu_{\overline{b_p}}\big((-r,r)\big) = O\left(p^{1.5}\right) \cdot \frac{a_p^{\alpha-2}\beta}{r} \cdot \exp\left(-\vartheta\left(\frac{r}{a_p}\right)^{\tau} + \vartheta\right).$$

Step 3—n Dimensional Tail Bound. Note that, in Step 4, the lattice that is considered is such that $\lambda_1^{\infty}(\mathcal{L}) > r$, for a given $r > r_0$. Thus it is enough to apply Corollary 5.37 to get

$$\nu_{\overline{|\sigma^{[p]}|}}\big(rB_n^{\infty},\mathcal{L}\big) \leq n \cdot \nu_{|\sigma_p|}\big((-r,r),\mathcal{L}_1\big) \leq nO\left(p^{1.5}\right) \cdot \frac{a_p^{\alpha-2}\beta}{r} \cdot \exp\left(-\vartheta\left(\frac{r}{a_p}\right)^{\tau} + \vartheta\right). \quad (5.38)$$

Note 5.54. In the above application of Corollary 5.37 we assumed that the constant κ—which bounds the weight of a coset in terms of the weight of the lattice—has value 1. In general this is guaranteed if the Fourier transform of a function is positive by Proposition 5.7. However, at the moment we do not have concrete evidence that the Fourier transform of $|\sigma_p|$ is a positive function. We conjecture, however, that the weight of the coset is indeed maximized by the lattice itself, which is the reason for out assumption.

Step 4—Bounding the Smoothing Parameter. Finally, we find the smoothing parameter for $f^{[p]}$ using Corollary 5.41. Let $\varepsilon \in (0,1)$. To find the smoothing parameter $\eta_{\overline{\mathbf{f}^{[p]}},\varepsilon}$, we must find r such that

$$\nu_{\overline{|\sigma^{[p]}|}}(rB_n^{\infty},\mathcal{L}) \leq \frac{\varepsilon}{1+\varepsilon}.$$

To satisfy the above inequality, we use Equation (5.38) by bounding its right hand side by $\varepsilon/(1+\varepsilon)$, obtaining

$$r \geq a_p \left(\frac{\ln\left(O\left(p^{1.5}\right)\beta n e^{\vartheta} a_p^{\alpha-2}\left(1+\frac{1}{\varepsilon}\right)\right)+5}{\vartheta}\right)^{\frac{1}{\tau}}.$$

Thus, by Corollary5.41, the smoothing parameter is given by

$$
\begin{aligned}
s &= \frac{a_p}{\lambda_1(\mathcal{L}^*)} \left(\frac{\ln\Big(O\left(p^{2.5}\right) \beta n e^{\vartheta} a_p^{\alpha-2} \big(1+\frac{1}{\varepsilon}\big)\Big) + 5}{\vartheta} \right)^{\frac{1}{\tau}} \\
&= \frac{a_p}{\lambda_1(\mathcal{L}^*)} \left(\frac{\ln\Big(O\left(p^{2.5}\right) n \left(1+\frac{1}{\varepsilon}\right)\Big)}{\vartheta} \right)^{\frac{1}{\tau}} \\
&= \frac{2\Gamma\left(\frac{p}{p+1}\right)}{\lambda_1(\mathcal{L}^*)} \left(\ln\left(O\left(p^{2.5}\right) n \left(1+\frac{1}{\varepsilon}\right)\right) \cdot O\left(p^{\frac{p}{p-1}}\right) \right)^{\frac{p-1}{p}} \\
&= \frac{O(p)}{\lambda_1(\mathcal{L}^*)} \left(\ln\left(O\left(p^{2.5}\right) n \left(1+\frac{1}{\varepsilon}\right)\right) \right)^{\frac{p-1}{p}},
\end{aligned}
$$

where the second equality holds as

$$
a_p^{\alpha-2} = \left(2\Gamma\left(\frac{p}{p+1}\right)\right)^{\frac{(p+2)}{2(p-1)}-2} \leq 1.
$$

for $f^{[p]}$ for the lattice $\mathcal{L}$. Thus, for all even $p \geq 2$,

$$
\eta_{\mathbf{f}^{[p]},\varepsilon}(\mathcal{L}) \leq \frac{O\left(p\right)}{\lambda_1(\mathcal{L}^*)} \left(\ln\left(O\left(p^{1.5}\right) n \left(1+\frac{1}{\varepsilon}\right)\right) \right)^{\frac{p-1}{p}}.
$$

5.8 Conclusion

In this chapter we studied the smoothening property that is present in the Gaussian distributions, and we extended this notion to any function for which the Poisson summation formula can be applied, which we defined as *smoothening functions*. In addition, we defined what the smoothing parameter of a lattice with respect to a function is. We clarified that a function being smoothening with respect to a lattice does not imply that the lattice has a smoothing parameter with respect to the function. We identified the necessity of studying the behavior of the Fourier transform of a function, in particular, the behavior of the tails of the function. To do so, we extended the tools presented in [MSD19] to bound the tails of a function, by allowing its Fourier transform to oscillate around 0; we provided a way to transform tail bounds for one-dimensional functions to n-dimensional functions; and we provided a direct connection between how smoothening a function is to the tail bound of its Fourier transform.

In addition, we proved that several classical results can be proved generically for smoothening functions. This proves the following points:

- Current proof techniques can often be adapted and extended to a generic setting.
- Several parts of the theory of lattice based cryptography can be generalized to a wider family of functions.

Finally, we described an infinite family of smoothening functions, and an infinite subfamily of functions for which the smoothing parameter exists and is sufficiently small. Based on our observations and examples—the fact that the Gaussian seems to be the function that converges "faster" to the uniform distribution in the quotient $\mathbb{R}^n/\mathcal{L}$, independently of what the lattice $\mathcal{L}$ is—we conjecture that, even though the Gaussian is not the only distribution that can be used to obtain certain results, the Gaussian is often the optimal choice of distribution, in regards to the parameters that are necessary to obtain the generalized result.

Chapter 6

Average-Case to Worst-Case Reductions Without Gaussians

> *"A technical argument by a trusted author, which is hard to check and looks similar to arguments known to be correct, is hardly ever checked in detail."*
>
> — Vladimir Voevodsky

This work was partially motivated by the questions "is it possible to use a distribution different from Gaussians for LWE while still enjoying an average-case to worst-case reduction to lattice problems?", "What is the underlying structure that makes the arguments work?". Before starting to explore this question it is necessary to understand and differentiate the multiple roles that the Gaussian distribution has in the reductions that exist in the literature.

For this we first turn our focus on Regev's work [Reg05], since this is the blueprint of subsequent related works such as [Pei08, LPR10, PRS17]. We start by making a detailed analysis of the main parts of the process and identify the main steps. We then describe what are the different properties of the Gaussian that are used in each step, and explore the possibility of using different distributions in each one.

One of the main steps of the reduction is the relation between the Discrete Gaussian Sampling (DGS) and Shortest Independent Vectors problems. In addition, the ability to sample from a discrete Gaussian has been used as an intermediate step for several other reductions and security proofs [MR07, Pei08, LPR10, ADS15], as well as many constructions [GPV08, MP12]. For this reason, we pay special attention to the Discrete Sampling problem and some of its different uses in lattice cryptography.

6.1 An Overview of [Reg05]

The main result in [Reg05] is a reduction from LWE to two lattice problems, namely the approximation version of GapSVP and SIVP within polynomial factors over a full-rank lattice $\mathcal{L}$. The road map of this reduction, however, is far from straightforward. The solutions for GapSVP and SIVP are found by solving the Discrete Gaussian Sampling problem over $\mathcal{L}$

The General Idea. Let W be an oracle that solves the (search) LWE problem for n, q and a Gaussian distribution χ. The main idea of the reduction is to use W to construct a quantum algorithm that efficiently samples vectors from a discrete Gaussian distribution of small enough radius. The algorithm starts by sampling a collection of vectors in $\mathcal{L}$ from a discrete Gaussian distribution of a large width. If the width is large enough, say exponentially large on the dimension, then this step can be efficiently done by means of a lattice reduction algorithm such as LLL.

The intention now is to progressively shorten the radius of the vectors by a constant factor. It does so by iterating an algorithm $\mathcal{A}$ that takes as input a collection of vectors V from the discrete Gaussian distribution over $\mathcal{L}$ of radius r, and outputs a collection of vectors U from the discrete Gaussian distribution over $\mathcal{L}$ of radius $r/2$ (or less). By iterating this algorithm over its own output eventually we obtain a collection of vectors of a sufficiently small radius.

The Iterative Step. For a given radius r—which is sufficiently greater than the smoothing parameter ρ with respect to $\mathcal{L}$—the algorithm constructs a quantum state representing the Fourier transform of $D_{\mathcal{L},\rho_{r/2}}$. This is a series of thin Gaussians centered at every vector in the dual lattice $\mathcal{L}^*$. The desired sample vector is obtained after applying the quantum Fourier transform to the constructed state and collapsing the result.[1] For this construction, however, it is necessary to call an oracle that solves a BDD instance on $\mathcal{L}^*$ that arises during the process. Thus the BDD oracle is the piece to be constructed next.

[1] The Fourier transform of $D_{\mathcal{L},\rho_{r/2}}$ is a continuous function; therefore there is naturally some loss of fidelity when computing this state.

From LWE **to** BDD. This is arguably the most important part of the reduction, to such an extent that it has subsequently been used to show worst-case to average-case reductions from LWE to itself under weaker conditions [ACPS09, BCD$^+$19, GMPW20].

The relationship between solving LWE on a "random" lattice and solving BDD on *any* lattice is given by a simple transformation of instances of the latter problem to instances of the former. This transformation, however, in addition assumes access to a number of samples from a discrete distribution over the dual lattice. This part itself consists of several pieces that are implied by more general results, some of which we have covered in the previous chapter.

At this step we have a collection $V = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ of samples from $D_{\mathcal{L},\rho_r}$—which are obtained either from a previous iteration or is the starting collection of exponentially long vectors. Suppose that $\mathcal{L}$ is given by the matrix B. Given an instance $\mathbf{x}$ of the BDD problem over $\mathcal{L}^*$, we compute

$$\big(\mathbf{a}_i \coloneqq B^{-1}\mathbf{v}_i,\ \langle \mathbf{x}, \mathbf{v}_i\rangle\big) \mod q, \quad \text{for } i \in \{1, \ldots, k\}. \tag{6.1}$$

This expression is very close to an LWE sample. To see this, let $\mathbf{u}$ be the closest lattice vector to $\mathbf{x}$—the solution of the BDD problem—and consider the difference $\mathbf{t} = \mathbf{x} - \mathbf{u} = (t_1, \ldots, t_n)$. Then we have

$$\langle \mathbf{x}, \mathbf{v}_i\rangle \equiv \langle \mathbf{u}, \mathbf{v}_i\rangle + \langle \mathbf{t}, \mathbf{v}_i\rangle \mod q. \tag{6.2}$$

Now, since $(B^*)^{-1} = B^T$, the first term can be written as

$$\langle \mathbf{u}, \mathbf{v}_i\rangle = \mathbf{u}^T\mathbf{v}_i = \mathbf{u}^T BB^{-1}\mathbf{v}_i = \left(B^T\mathbf{u}\right)^T B^{-1}\mathbf{v}_i = \left\langle (B^*)^{-1}\mathbf{u},\ B^{-1}\mathbf{v}\right\rangle. \tag{6.3}$$

By defining $\mathbf{s} \coloneqq (B^*)^{-1}\mathbf{u}$, we obtain the more familiar expression

$$\langle \mathbf{x}, \mathbf{v}_i\rangle \equiv \langle \mathbf{u}, \mathbf{v}_i\rangle + \langle \mathbf{t}, \mathbf{v}_i\rangle = \langle \mathbf{a}_i, \mathbf{s}\rangle + \langle \mathbf{t}, \mathbf{v}_i\rangle \mod q. \tag{6.4}$$

Thus the BDD problem is solved once $\mathbf{s}$ is recovered. Notice that $\langle \mathbf{t}, \mathbf{v}_i\rangle$ is small, since $\mathbf{t}$ is bounded by a small real parameter, and $\mathbf{v}_i$ is a sample from $D_{\mathcal{L},\rho_r}$. Nevertheless, the above expression is not precisely a sample from the traditional instantiation of the LWE problem with Gaussian noise.

Noise Analysis. Write $\mathbf{v}_i = (v_{i,1}, \ldots, v_{i,n})$. Then error term is given by

$$\langle \mathbf{t}, \mathbf{v}_i\rangle = \sum_{j\in[n]} t_j v_{i,j}. \tag{6.5}$$

The vector $\mathbf{v}_i$ is distributed according to a discrete Gaussian distribution over $\mathcal{L}$. However, the distribution of its coordinates $v_{i,1}, \ldots, v_{i,n}$ over $\mathbb{R}$ is significantly harder to describe. Moreover, our assumption is that W solves the LWE problem where the noise is sampled from a particular Gaussian distribution χ. To guarantee that the noise in our sample is described by such a distribution, we add a small Gaussian error e. The distribution of the resulting noise is statistically close to a Gaussian distribution.

In fact, the distribution of e can be expressed as a linear combination of independent Gaussian random variables. This property of the Gaussian distribution is called *infinite divisibility*. Thus e can be written as $\langle \mathbf{t}, \mathbf{h} \rangle$, where $\mathbf{h} = (h_1, \ldots, h_n)$ is described by a Gaussian distribution over $\mathbb{R}^n$. Thus the final noise is described by

$$e' = \langle \mathbf{t}, \mathbf{v}_i \rangle + e = \langle \mathbf{t}, \mathbf{h} + \mathbf{v}_i \rangle = \sum_{j \in [n]} t_j (h_j + v_{i,j}). \tag{6.6}$$

Finally, the distribution of $\mathbf{h} + \mathbf{v}_i$—or equivalently, in this case, the distribution of each coordinate $v_{i,j}$ over $\mathbb{R}$—which is an addition of a discrete and a continuous random variable, can be proved to be statistically close to a Gaussian. Therefore, after carefully selecting e, the distribution of e' is statistically close to χ.

6.2 The Critical Steps

In the analysis of [Reg05] given in the previous section we can observe that the Gaussian distribution is used as a tool in several different parts of the process, and for several different purposes. In this section we isolate every instance where a Gaussian distribution is used, and study the possibility of using a different distribution for the corresponding goal.

Initialization of the Quantum Sampler.

The quantum discrete Gaussian sampler is initialized with a quantum state representing an n-dimensional wide Gaussian distribution. It uses an algorithm by Grover and Rudolph [GR02] to create the 1-dimensional components of the state, and the smoothening property of the Gaussian, as well as the Gaussian's own tail bounds, to argue that the resulting state—after a careful manipulation—is sufficiently close to the desired state.

In the following propositions we show sufficient conditions on a function to construct the quantum state used to initialize the discrete sampler. We do so utilizing the language and tools related to smoothening functions and tail bound functions developed in Chapter 5. It is worth noticing, however, that this algorithm is later used to simulate a continuous distribution over the dual space.

In the setting for [Reg05, Lemma 3.12], the relevant function is a Gaussian that is wide with respect to the lattice. This width allows us to guarantee that for two vectors that are not far from the origin and not far from each other, the corresponding weight assigned to each by the Gaussian is very similar. Consequently the function is also smoothening for the lattice.

Lemma 6.1. *Let $\varepsilon \in \mathbb{R}_{>0}$ and let $\mathbf{f}$ be a non-negative ε-smoothening function for a lattice $\mathcal{L} \subset \mathbb{R}^n$. Consider a compact set $K \subset \mathbb{R}^n$ and let ν be a tail bound function for $\mathbf{f}$. Let $\mathbf{c} \in \mathbb{R}^n$ and let $\gamma_{\mathbf{c}} \colon \mathbf{x} \mapsto \mathbf{f}(\mathbf{x})/\mathbf{f}(\mathbf{x}-\mathbf{c})$. Let $\mathcal{L}$ be a lattice and assume that there exists a constant κ such that for every $\mathbf{x} \in \mathcal{L}$, $\gamma_{\mathbf{c}}(\mathbf{x}) < \kappa$. If for all $\mathbf{x} \in K \cap \mathcal{L}$, $\left|1 - \gamma_{\mathbf{c}}(\mathbf{x})\right| < \varepsilon'$, then the statistical distance between $D_{\mathcal{L},\mathbf{f}}$ and $D_{\mathcal{L},\mathbf{f},\mathbf{c}}$ is bounded by*

$$\frac{2\varepsilon(1+\varepsilon')}{1-\varepsilon} + \left(\kappa\left(1 + \frac{2\varepsilon}{1-\varepsilon}\right) + 1\right)\nu_{\mathbf{f}}(K, \mathcal{L}).$$

Proof. Following the corresponding definitions, the statistical distance Δ between $D_{\mathcal{L},\mathbf{f}}$ and $D_{\mathcal{L},\mathbf{f},\mathbf{c}}$ is given by

$$\Delta = \sum_{\mathbf{x}\in\mathcal{L}} \left|D_{\mathcal{L},\mathbf{f}}(\mathbf{x}) - D_{\mathcal{L},\mathbf{f},\mathbf{c}}(\mathbf{x})\right| = \sum_{\mathbf{x}\in\mathcal{L}} \left|\frac{\mathbf{f}(\mathbf{x})}{\mathbf{f}(\mathcal{L})} - \frac{\mathbf{f}(\mathbf{x}-\mathbf{c})}{\mathbf{f}(\mathcal{L}-\mathbf{c})}\right|.$$

On the other hand, since $\mathbf{f}$ is ε smoothening we have that, by Proposition 5.16, $\mathbf{f}(\mathcal{L}+\mathbf{c}) = \delta(\mathbf{c})\mathbf{f}(\mathcal{L})$, with $\delta(\mathbf{c}) \in \left(\frac{1-\varepsilon}{1+\varepsilon}, \frac{1+\varepsilon}{1-\varepsilon}\right) = \left(1 - \frac{2\varepsilon}{1+\varepsilon}, 1 + \frac{2\varepsilon}{1-\varepsilon}\right)$. Thus we can rewrite the above expression as

$$\begin{aligned}
\Delta &= \frac{1}{\mathbf{f}(\mathcal{L})} \sum_{\mathbf{x}\in\mathcal{L}} \left|\mathbf{f}(\mathbf{x}) - \frac{\mathbf{f}(\mathbf{x}-\mathbf{c})}{\delta(\mathbf{c})}\right| \\
&= \frac{1}{\mathbf{f}(\mathcal{L})} \sum_{\mathbf{x}\in\mathcal{L}} \left|\mathbf{f}(\mathbf{x}) - \frac{\gamma_{\mathbf{c}}(\mathbf{x})}{\delta(\mathbf{c})}\mathbf{f}(\mathbf{x})\right| \\
&= \frac{1}{\mathbf{f}(\mathcal{L})} \sum_{\mathbf{x}\in\mathcal{L}} \left|1 - \frac{\gamma_{\mathbf{c}}(\mathbf{x})}{\delta(\mathbf{c})}\right| \mathbf{f}(\mathbf{x}) \\
&= \frac{1}{\mathbf{f}(\mathcal{L})} \left(\sum_{\mathbf{x}\in\mathcal{L}\cap K} \left|1 - \frac{\gamma_{\mathbf{c}}(\mathbf{x})}{\delta(\mathbf{c})}\right| \mathbf{f}(\mathbf{x}) + \sum_{\mathbf{x}\in\mathcal{L}\setminus K} \left|1 - \frac{\gamma_{\mathbf{c}}(\mathbf{x})}{\delta(\mathbf{c})}\right| \mathbf{f}(\mathbf{x})\right).
\end{aligned} \tag{6.7}$$

The behaviour of $\gamma_{\mathbf{c}}$ provided by the hypotheses is only guaranteed for $x \in K$—exploiting this information is the main reason for splitting the above summation into two parts. Consider $\mathbf{x}_0 \in \mathcal{L} \cap K$ such that $\left|1 - \gamma_{\mathbf{c}}(\mathbf{x})/\delta(\mathbf{c})\right|$ is maximized, and let $\gamma_0 = \gamma(\mathbf{x}_0, \mathbf{c})$. Additionally, since $\gamma_{\mathbf{c}}$ is bounded, both parts of the summation can be bounded as in the following expression.

$$\begin{aligned}
\Delta &\leq \frac{1}{\mathbf{f}(\mathcal{L})} \left(\left|1 - \frac{\gamma_0}{\delta(\mathbf{c})}\right| \sum_{\mathbf{x} \in \mathcal{L} \cap K} \mathbf{f}(\mathbf{x}) + \left(1 + \left|\frac{\kappa}{\delta(\mathbf{c})}\right|\right) \sum_{\mathbf{x} \in \mathcal{L} \setminus K} \mathbf{f}(\mathbf{x}) \right) \\
&\leq \frac{1}{\mathbf{f}(\mathcal{L})} \left(\left|1 - \frac{\gamma_0}{\delta(\mathbf{c})}\right| \cdot \mathbf{f}(\mathcal{L}) + \left(1 + \left|\frac{\kappa}{\delta(\mathbf{c})}\right|\right) \nu_{\mathbf{f}}(K, \mathcal{L}) \cdot \mathbf{f}(\mathcal{L}) \right) \qquad (6.8) \\
&= \left|1 - \frac{\gamma_0}{\delta(\mathbf{c})}\right| + \left(1 + \left|\frac{\kappa}{\delta(\mathbf{c})}\right|\right) \nu_{\mathbf{f}}(K, \mathcal{L}).
\end{aligned}$$

Finally, notice that $1/\delta(\mathbf{c})$ is also a number in the interval $\left(1 - \frac{2\varepsilon}{1+\varepsilon}, 1 + \frac{2\varepsilon}{1-\varepsilon}\right)$. Thus

$$\left|1 - \frac{\gamma_0}{\delta(\mathbf{c})}\right| \in \left[0, \frac{2\varepsilon}{1-\varepsilon}(\varepsilon' + 1)\right) \quad \text{and} \quad \left|\frac{\kappa}{\delta(\mathbf{c})}\right| \in \left[0, \kappa\left(1 + \frac{2\varepsilon}{1-\varepsilon}\right)\right).$$

The result follows. □

Notice that the third equality in the series of equations (6.7) depends on the function being positive. When this is not the case, the last expression in (6.8) is multiplied by a factor of $|\mathbf{f}|(\mathcal{L})/\mathbf{f}(\mathcal{L})$, which is difficult to bound.

Moving on to the next part of the analysis, the construction of the quantum state given in the following proposition relies on an algorithm proposed by Grover and Rudolph. In [GR02], they describe a procedure to obtain the superposition

$$\sum_{i \in I} \sqrt{f(i)}\ |i\rangle$$

representing a probability distribution f over a set $I \subset \mathbb{R}$, whenever the function is *efficiently integrable*, which means that there exists an algorithm such that, for every $i, j \in I$, efficiently computes the sum $\sum_{i \leq k \leq j} f(k)$. The authors of the aforementioned paper make mention of enough conditions for the function to have the above property, namely, whenever the function is log-concave it is possible to compute the sum using Monte-Carlo integration.

Remark 6.2. Notice that for any $a, b \in \mathbb{R}_{>0}$ we have $(a-b)^2 < a^2 - b^2$. As a consequence, given real valued positive functions f and g defined over a discrete set $A \subset \mathbb{R}^n$, we may bound the ℓ_2-distance between the states

$$\sum_{\mathbf{x}\in A} \mathbf{f}(\mathbf{x})\,|\mathbf{x}\rangle \quad \text{and} \quad \sum_{\mathbf{x}\in A} \mathbf{g}(\mathbf{x})\,|\mathbf{x}\rangle$$

by computing—or bounding—the statistical distance between the probability distributions they induce over A.

Proposition 6.3. *Let $f \in D_1$ be a positive function which is efficiently integrable over B_1 and consider the function $\mathbf{f}\colon (x_1, \ldots, x_n) \mapsto \prod_{i\in[n]} f(x_i)$. Consider as well a tail bound function ν_f for f and for $\mathbf{c} \in \mathbb{R}^n$ consider the mapping $\gamma_c\colon \mathbf{x} \mapsto \mathbf{f}(\mathbf{x})/\mathbf{f}(\mathbf{x}-\mathbf{c})$. Furthermore, suppose that $\mathbf{f}$ satisfies the following properties.*

1. *There exists a constant κ such that for every $\mathbf{x}, \mathbf{c} \in \mathbb{Z}^n$, $\gamma_{\mathbf{c}}(\mathbf{x}) < \kappa$*
2. *For $r \in \mathbb{R}_0$ there exists $\varepsilon' \in \mathbb{R}_{>0}$ such that for all $\mathbf{x} \in \mathbb{Z}^n \cap B_n^\infty$ and $\mathbf{c} \in \mathbb{Z}^n$, $\gamma_{\mathbf{c}}(\mathbf{x}) < \varepsilon$.*

Then there exists an efficient quantum algorithm that, on input a set $\mathbf{B} \subset \mathbb{Z}^n$ of n linearly independent vectors, outputs a quantum state $|\psi\rangle$ such that

$$\left\| |\psi\rangle - \sum_{\mathbf{x}\in\mathcal{L}(\mathbf{B})} \sqrt{|\mathbf{f}(\mathbf{x})|}\;|\mathbf{x}\rangle \right\|_2 < \frac{2\varepsilon(1+\varepsilon')}{1-\varepsilon} + n\kappa'\nu_f(rB_n^\infty) + \sqrt{n\nu_f(rB_n^\infty)\mathbf{f}(\mathbb{Z}^n)},$$

where $\varepsilon = \sum_{\mathbf{x}\in\mathcal{L}(\mathbf{B})^*\setminus\{\mathbf{0}\}} |\widehat{f}(\mathbf{x})|$.

Proof. For $i \in [n]$, use the algorithm by Grover and Rudolph given in [GR02] to compute the superposition

$$|\phi_i\rangle = \sum_{x_i\in\mathbb{Z}\cap rB_1^\infty} f(x_i)\;|x_i\rangle.$$

By taking the tensor product $|\phi_1\rangle \otimes \ldots \otimes |\phi_n\rangle$ of the states obtained above, we obtain a superposition of integer vectors

$$|\phi\rangle = \sum_{\mathbf{x}\in\mathbb{Z}^n\cap rB_n^\infty} \mathbf{f}(\mathbf{x})\;|\mathbf{x}\rangle.$$

This state is a finite superposition of vectors in the integer lattice. Nonetheless, if the function f has negligible tails, this information should be enough to simulate a superposition over the entirety of $\mathbb{Z}^n$. Using the given tail bound for f, the ℓ_2 distance is thus bounded by

$$\left\| |\phi\rangle - \sum_{\mathbf{x}\in\mathbb{Z}^n} \sqrt{\mathbf{f}(\mathbf{x})}\, |\mathbf{x}\rangle \right\|_2 = \left(\sum_{\mathbf{x}\in\mathbb{Z}^n \setminus rB_n^\infty} \mathbf{f}(\mathbf{x}) \right)^{1/2} < \sqrt{n\nu_f(rB_n^\infty)\mathbf{f}(\mathbb{Z}^n)}.$$

On the other hand, let $\mathbf{B}'$ be the result of applying the LLL basis reduction to $\mathbf{B}$, and let $\mathcal{P}(\mathbf{B}')$ be the parallelepiped generated by $\mathbf{B}'$. Using an ancillary quantum register, compute the reduction of $\mathbf{x}$ modulo $\mathcal{P}(\mathbf{B}')$, thus obtaining the state

$$\sum_{\mathbf{x}\in\mathbb{Z}^n} \sqrt{\mathbf{f}(\mathbf{x})}\, \big|\mathbf{x},\ \mathbf{x} \mod \mathcal{P}(\mathbf{B}')\big\rangle.$$

When the second register is collapsed to an element $\mathbf{c} \in \mathcal{P}(\mathbf{B}')$, the resulting state, contained in the first register, is the superposition over all vectors $\mathbf{x} \in \mathbb{Z}^n$ such that, $\mathbf{x} \equiv \mathbf{c} \mod \mathcal{P}(\mathbf{B})$. Thus it can be written as

$$\sum_{\mathbf{x}\in\mathcal{L}+\mathbf{c}} \sqrt{\mathbf{f}(\mathbf{x})}\, |\mathbf{x}\rangle.$$

After subtracting $\mathbf{c}$ we obtain the superposition over lattice points

$$|\psi_2\rangle := \sum_{\mathbf{x}\in\mathcal{L}+\mathbf{c}} \sqrt{\mathbf{f}(\mathbf{x})}\, |\mathbf{x}-\mathbf{c}\rangle = \sum_{\mathbf{x}\in\mathcal{L}} \sqrt{\mathbf{f}(\mathbf{x}+\mathbf{c})}\, |\mathbf{x}\rangle.$$

Finally, to compare $|\psi_2\rangle$ with the desired state $|\psi_1\rangle = \sum_{\mathbf{x}\in\mathcal{L}(\mathbf{B})} \sqrt{\mathbf{f}(\mathbf{x})}\, |\mathbf{x}\rangle$, we compare the probability distributions they induce on the lattice by computing the statistical distance between them. For $\mathbf{x} \in \mathcal{L}$, the amplitude square of the ket $|\mathbf{x}\rangle$ in $|\psi_1\rangle$ is $\mathbf{f}(\mathbf{x})$, whereas the amplitude square of the same ket in $|\psi_2\rangle$ is $\mathbf{f}(\mathbf{x}+\mathbf{c})$. Thus the corresponding distributions are $D_{\mathcal{L},\mathbf{f}}$ and $D_{\mathcal{L},\mathbf{f},\mathbf{c}}$, respectively. By Lemma 6.1, the statistical distance between $D_{\mathcal{L},\mathbf{f}}$ and $D_{\mathcal{L},\mathbf{f},\mathbf{c}}$ is negligible. The proof follows by Remark 6.2. □

From Discrete Sampling to A Learning Problem

As described by equations (6.1)-(6.4), the association between BDD and LWE is facilitated by the interplay between the lattice $\mathcal{L}$ on which the BDD problem is given, and its dual $\mathcal{L}$. The solution $\mathbf{u}$ is written in terms of the primal lattice as $\mathbf{s} = (B^*)^{-1}\mathbf{u} \in \mathcal{L}$, whose residue modulo q is then the secret that is to be recovered by an LWE solver. It is important to keep in mind that this reduction, as well as every other variant of it that appears in the literature, requires a list of samples $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ from a discrete distribution $D_{\mathcal{L},\mathbf{f}}$ over $\mathcal{L}$.

Uniform Samples of a Learning Problem. Let $q \in \mathbb{Z}_{>2}$ and let $\mathbf{f}$ be a real valued function over $\mathbb{R}^n$. If $\mathbf{f}$ is ε-smoothening for the lattice $q\mathcal{L}$, the smoothening property guarantees that the function induces a distribution over $\mathbb{R}^n$. Then, intuitively, the discrete distribution $D_{q\mathcal{L},\mathbf{f}}$ induces a uniform distribution over the cosets of $q\mathcal{L}$ modulo $\mathcal{L}$

Recall that, in the case of Gaussian functions, the smoothing parameter is in direct proportion with the covering radius of the lattice. Since the covering radius of a superlattice is no larger than that of the lattice, the smoothing parameter is also smaller. Thus it is natural to expect that, given a smoothening function for a lattice, this function is also smoothening for every superlattice. We show this in the following claim.

Claim 6.4. Let $\mathcal{L} \subseteq \mathcal{L}' \subset \mathbb{R}^n$ be two lattices. If a function $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}$ is ε-smoothening for $\mathcal{L}$ then $\mathbf{f}$ is ε-smoothening for $\mathcal{L}'$.

Proof. Since $\mathcal{L}$ is a sublattice of $\mathcal{L}'$, it follows from the definition that $(\mathcal{L}')^* \subseteq \mathcal{L}^* \subset \mathbb{R}$. Therefore

$$\left|\widehat{D_{\mathbf{f}}}\right|\left((\mathcal{L}')^* \setminus \{\mathbf{0}\}\right) \le \left|\widehat{D_{\mathbf{f}}}\right|\left(\mathcal{L}^* \setminus \{\mathbf{0}\}\right) < \varepsilon,$$

as required. □

The following corollary is a discrete analogue of Lemma 5.17, which shows that samples from an ε-smoothening distribution over a lattice $\mathcal{L}$ are themselves close to uniform modulo any sub-lattice $\mathcal{L}'$. This result is implicit in the proof of [Reg05, Claim 3.11], but was later better quantified in [GPV08].

Corollary 6.5. *Let $\mathcal{L}' \subseteq \mathcal{L} \subset \mathbb{R}^n$ be two lattices. For $\varepsilon \in \left(0, \frac{1}{2}\right)$, consider a ε-smoothening function $\mathbf{f}$ for $\mathcal{L}'$. Then for any $\mathbf{c} \in \mathbb{R}^n$, the distribution of $D_{\mathcal{L},\mathbf{f},\mathbf{c}} \bmod \mathcal{L}'$ is within statistical distance at most 2ε of uniform over $\mathcal{L} \bmod \mathcal{L}'$.*

Proof. Let $\mathbf{x} \in \mathcal{L}$ and let $\mathbf{c} \in \mathbb{R}^n$. An implication of Claim 6.4 is that $\mathbf{f}$ is ε-smoothening for $\mathcal{L}$. By Proposition 5.16, it follows that

$$D_{\mathcal{L},\mathbf{f},\mathbf{c}}(\mathbf{x} + \mathcal{L}') = \sum_{\mathbf{v} \in \mathbf{x} + \mathcal{L}'} D_{\mathcal{L},\mathbf{f},\mathbf{c}}(\mathbf{v}) = \frac{1}{\mathbf{f_c}(\mathcal{L})} \sum_{\mathbf{v} \in \mathbf{x} + \mathcal{L}'} \mathbf{f_c}(\mathbf{v}) = \frac{\mathbf{f_{c-x}}(\mathcal{L}')}{\mathbf{f_c}(\mathcal{L})} = \delta(\mathbf{x}) \frac{\mathbf{f_c}(\mathcal{L}')}{\mathbf{f_c}(\mathcal{L})},$$

where $\delta(\mathbf{x}) \in \left(\frac{1-\varepsilon}{1+\varepsilon}, \frac{1+\varepsilon}{1-\varepsilon}\right)$. For $\varepsilon \in \left(0, \frac{1}{2}\right)$, $\varepsilon(1 \pm 2\varepsilon) \ge 0$. As a consequence $1 + \varepsilon \le (1-\varepsilon)(1+4\varepsilon)$ and $(1-4\varepsilon)(1+\varepsilon) \le 1 - \varepsilon$. Since $\frac{\mathbf{f_c}(\mathcal{L}')}{\mathbf{f_c}(\mathcal{L})} \le 1$, putting it all together we have that for all $\mathbf{x} \in \mathcal{L}$,

$$D_{\mathcal{L},\mathbf{f},\mathbf{c}}(\mathbf{x} + \mathcal{L}') = \delta(\mathbf{x}) \frac{\mathbf{f_c}(\mathcal{L}')}{\mathbf{f_c}(\mathcal{L})} \in (1 - 4\varepsilon, 1 + 4\varepsilon) \frac{\mathbf{f_c}(\mathcal{L}')}{\mathbf{f_c}(\mathcal{L})} \subseteq \left(\frac{\mathbf{f_c}(\mathcal{L}')}{\mathbf{f_c}(\mathcal{L})} - 4\varepsilon, \frac{\mathbf{f_c}(\mathcal{L}')}{\mathbf{f_c}(\mathcal{L})} + 4\varepsilon\right).$$

The result follows. □

Noise Correction

An important component of the BDD to LWE reduction part is the analysis of the noise constituting the LWE instance. To this end, Regev uses a re-shaping technique by adding a small error e, as described in Equation (6.6). When the added error e follows a Gaussian distribution, it can be expressed as a linear combination of independent Gaussian random variables. This is implied by the following remark, which itself follows from Proposition 2.4.

Remark 6.6. Given $X_1, \dots, X_n$ be a collection of independent identically distributed Gaussian random variables of width s and a vector $\mathbf{t} = (t_1, \dots, t_n) \in \mathbb{R}^n$, the random variable obtained from the linear combination $\sum_{i \in [n]} t_i X_i$ is a Gaussian random variable of width $s\|\mathbf{t}\|_2$.

In particular, given $\mathbf{t}$ and any desired width for the distribution of e, it is straightforward to obtain s. As a consequence, e can be written as $\langle \mathbf{t}, \mathbf{h} \rangle$, where $\mathbf{h} = (h_1, \dots, h_n)$ is described by a Gaussian distribution over $\mathbb{R}^n$. Performing this step in a more general setting may require our desired distribution of e to satisfy a property that is analogous to that given by Remark 6.6.

After the addition of e, the final noise is described by

$$e' = \langle \mathbf{t}, \mathbf{v}_i \rangle + e = \langle \mathbf{t}, \mathbf{h} + \mathbf{v}_i \rangle = \sum_{j \in [n]} t_j (h_j + v_{i,j}). \tag{6.9}$$

Here $\mathbf{z} = \mathbf{h} + \mathbf{v}_i$ is the result of an addition of a discrete and a continuous random variable over $\mathbb{R}^n$ (or over $\mathbb{R}$ if we consider $z_j = h_j + v_{i,j}$). This can be proved to be related to the addition of the two continuous random variables, independently of their shape, whenever certain conditions about the functions with respect to the lattice are satisfied.

Proposition 6.7. *Let $\mathcal{L} \subset \mathbb{R}^n$ be a lattice. Consider functions $\mathbf{f} \in \mathcal{D}_n$ and $\mathbf{g} \in L^1(\mathbb{R}^n)$ and suppose that for $\mathbf{x} \in \mathbb{R}^n$, the function $\mathbf{h}_\mathbf{x}(\mathbf{y}) := \mathbf{f}(\mathbf{y})\mathbf{g}(\mathbf{x} - \mathbf{y})$ is ε-smoothening for $\mathcal{L}$. Then*

$$\big(D_{\mathcal{L},\mathbf{f}} * \mathbf{g}\big)(\mathbf{x}) \in \frac{\det \mathcal{L}^*}{\mathbf{f}(\mathcal{L})} \big(\mathbf{f} * \mathbf{g}\big)(\mathbf{x}) + (-\varepsilon, \varepsilon).$$

Proof. By the Poisson summation formula (Lemma 5.6),

$$\begin{aligned}
\big(D_{\mathcal{L},\mathbf{f}} * \mathbf{g}\big)(\mathbf{x}) &= \sum_{\mathbf{y}\in\mathcal{L}} D_{\mathcal{L},\mathbf{f}}(\mathbf{y})\mathbf{g}(\mathbf{x}-\mathbf{y}) \\
&= \frac{1}{\mathbf{f}(\mathcal{L})} \sum_{\mathbf{y}\in\mathcal{L}} \mathbf{h}_{\mathbf{x}}(\mathbf{y}) \\
&= \frac{\det \mathcal{L}^*}{\mathbf{f}(\mathcal{L})} \sum_{\mathbf{z}\in\mathcal{L}^*} \widehat{\mathbf{h}_{\mathbf{x}}}(\mathbf{z}) \\
&= \frac{\det \mathcal{L}^*}{\mathbf{f}(\mathcal{L})} \left(\widehat{\mathbf{h}_{\mathbf{x}}}(\mathbf{0}) + \sum_{\mathbf{z}\in\mathcal{L}^*\setminus\{\mathbf{0}\}} \widehat{\mathbf{h}_{\mathbf{x}}}(\mathbf{z}) \right).
\end{aligned}$$

Since $\mathbf{h}_{\mathbf{x}}$ is ε-smoothening, $\sum_{\mathbf{z}\in\mathcal{L}^*\setminus\{\mathbf{0}\}} \widehat{\mathbf{h}_{\mathbf{x}}}(\mathbf{z}) \in (-\varepsilon,\varepsilon)$. As for $\widehat{\mathbf{h}_{\mathbf{x}}}(\mathbf{0})$ we have that

$$\begin{aligned}
\widehat{\mathbf{h}_{\mathbf{x}}}(\mathbf{0}) &= \int_{\mathbb{R}^n} \mathbf{h}_{\mathbf{x}}(\mathbf{y}) e^{2\pi i\langle \mathbf{0},\mathbf{y}\rangle} d\mathbf{y} \\
&= \int_{\mathbb{R}^n} \mathbf{h}_{\mathbf{x}}(\mathbf{y}) d\mathbf{y} \\
&= \int_{\mathbb{R}^n} \mathbf{f}(\mathbf{y})\mathbf{g}(\mathbf{x}-\mathbf{y}) d\mathbf{y} \\
&= \big(\mathbf{f} * \mathbf{g}\big)(\mathbf{x}),
\end{aligned}$$

which completes the proof. □

It is worth noticing, however, that this portion of the proof in the case of the Gaussians can be completed without the necessity of continuous distributions thanks to the recent work by Genise et al. [GMPW20]. In the cited paper the authors analyze the behavior of convolutions of discrete Gaussians.

The final piece is to describe the inner product $\langle \mathbf{t}, \mathbf{z}\rangle$, where $\mathbf{t}$ is a bounded vector and $\mathbf{z}$ is distributed according to some known function ψ. At this point we make our final assumption. We assume that ψ can be written as a product $\psi(\mathbf{x}) = \prod_{i=1}^n \varphi_i(x_i)$. This implies that the random variable $\mathbf{Z}$ that describes $\mathbf{z}$ can be written as a sum of independent random variables $\mathbf{e}_1 Z_1, \ldots, \mathbf{e}_n Z_n$, where for each i, Z_i is a random variable over $\mathbb{R}$. Under this assumption,

$$\langle \mathbf{t}, \mathbf{Z}\rangle = \sum_{i=1}^{n} t_i Z_i. \tag{6.10}$$

If the distribution φ_i describing each Z_i has finite variance then, by the Central Limit Theorem, the expression in (6.10) is "close" to a Gaussian distribution (of unknown width). The distance is determined by the distance of each φ_i to a Gaussian. The Berry-Essen inequality provides a polynomial bound on the ∞-distance between (6.10) and a Gaussian. A polynomial bound on the ℓ_1 norm—which is the statistical distance between the distributions—is given by Goldstein in [Gol10]. This limits the possibilities for χ to distributions that are polynomially close to a Gaussian. When χ is a Gaussian, the width problem is cleanly dealt with by leveraging 6.6 to add additional noise. While the possibility of being polynomially far from a Gaussian could yield a useful relaxation for practical hardness with more effort, this approach to the reduction seems to necessarily produce some form of "bell" shaped distribution very different from the choice of smoothening function.

The Necessity of Correcting the Noise. The noise given prior to the correction step

$$e' = \langle \mathbf{t}, \mathbf{v} \rangle = \sum_{j \in [n]} t_j v_j. \tag{6.11}$$

is an arbitrary yet bounded linear combination of the entries of the vector $\mathbf{v}$. Since $\mathbf{v}$ is sampled according to a discrete distribution with support on a lattice $\mathcal{L}$, each v_j follows a discrete distribution supported on the projection of $\mathcal{L}$ onto the subspace generated by $\mathbf{e}_j$. The projection of $\mathcal{L}$ onto $\mathbf{e}_j$ is also a lattice generated by a vector $a_j\mathbf{e}_j$. Thus, for a fixed instance $\mathcal{L}$ and $\mathbf{t}$ of BDD, the error is follows a distribution over $\mathbb{R}$ with support

$$\mathcal{L}_1 = t_1 a_1 \mathbb{Z} + \ldots + t_n a_n \mathbb{Z} \subset \mathbb{R}.$$

It is clear that $\mathcal{L}_1$ is a lattice, whose structure is inherently dependent on the given BDD instance.

The noise correction technique described at the beginning of this subsection aims to eliminate this dependency by making the noise continuous. After adding a small amount of noise, the resulting distribution is virtually independent from a lattice in the BDD.

If the noise correction step is omitted, the element e' is then distributed according to a discrete distribution over $\mathcal{L}_1$.

6.3 Discrete Sampling

In the final part of this chapter we turn our attention to the Discrete Sampling problem. The hardness of sampling from a discrete Gaussian varies significantly, depending on the width of the desired distribution. When the width is small, sampling from a discrete Gaussian is hard, since a vector obtained from this distribution is a small vector of the lattice (with high probability). When the width is (exponentially) large, however, sampling form this distribution can be performed in polynomial time after pre-processing the input basis using LLL.

Sampling Algorithms.

Many existing lattice samplers use ℓ_2 spherical symmetry as a convenient shorthand to argue independence of the output distribution from the private information (in other words, the secret basis) used to produce short elements in the lattice. Sampling statistically close to any distribution over a lattice which is independent from a particular choice of basis is also similar. For example, we could choose some other "spherical" distribution for a different choice of ℓ_p.

The framework of [MP12, LW15] allows sampling from a variety of distributions for certain specially constructed lattices. However, the components of samplers which work for arbitrary lattices (given a short basis) rely on spherical symmetry and additional properties of Gaussians.

The "randomized nearest planes" samplers of [Kle00, GPV08] exploit the spherical symmetry of the Gaussian in their analysis to make use of samples from $D_{\mathbb{Z},s}$, independently of the direction of the Gram-Schmidt vectors. In particular, in the analysis of [GPV08] it is shown that

$$\prod_i \rho_s \left(c_i \cdot |\tilde{\mathbf{b}_i}|_2 \right) = \rho_s \left(\sum_i c_i \cdot \tilde{\mathbf{b}}_i \right).$$

The Gaussian factors not only along the axis, but through any inner-product norm, which gives the result through orthogonality of the Gram-Schmidt vectors. In the following subsection we prove that the Gaussian is the only distribution with finite variance which can be factored simultaneously along different orthonormal bases as independent distributions from the same family. This is a fundamental barrier to generalizing any sampling algorithm following the [GPV08] strategy, which analyzes the joint distribution of many samples drawn from scaled/re-centered versions of the same one-dimensional distribution. However, the same algorithm and analysis can still be adapted directly if we restrict to lattices with axes-aligned Gram-Schmidt vectors. This is sufficient for sampling statistically close to a distribution over $\mathbb{Z}^m$, as well as reductions where we are allowed to rotate the lattice to impose this condition a priori—but not for cryptographic constructions relying on uniformly random lattices.

We can try to get around this limitation in two ways. As noted in [GPV08], smoothening functions are also amendable to rejection sampling by quantizing samples from $D_{\mathbf{f},s}$ (or $D_{\mathbb{Z}^m,\mathbf{f},s}$) by "rounding" to the lattice using a short basis. Smoothening ensures every point in the lattice with non-negligible measure has a preimage. However, ensuring efficiency requires $\mathbf{f}$ to be sufficiently "flat" to lower the probability of rejection. A bound on the number of draws before a sample is accepted can be quantified by β satisfying

$$\mathbb{P}_{\mathbf{x}\leftarrow D_{\mathcal{L},\mathbf{f},s}(\mathbf{x})}\left[D_{\mathcal{L},\mathbf{f},s}(\mathbf{x})/D_{\mathcal{L},\mathbf{f},s,\mathbf{d}}(\mathbf{x}) < \beta\right] = 1 - \mathrm{negl}(n).$$

where $\mathbf{d}$ arises from the shift induced by rounding to the lattice. The shift is bounded by the decoding radius of the Nearest Planes Algorithm [Bab86]. In the Gaussian case this requires taking s proportional to the *diameter* of the Gram-Schmidt basis $\widetilde{\mathbf{B}}$, rather than simply $|\widetilde{\mathbf{B}}|$. The other component which quantifies the probability of rejection depends on the rate of decay of $\mathbf{f}$, which can also be quite unfavourable for functions which decay appreciably faster than the Gaussian. Another possible direction is to apply rejection sampling to samples from a distribution over a suitably rotated lattice. This strategy also requires increasing s when the distribution is not ℓ_2 symmetric, but of a choice of distribution which has other useful geometric properties—this could conceivably require smaller s than required when bounding the effect of large shifts.

Later approaches to sampling over lattices such as [Pei10] do not directly require spherical symmetry. However, these constructions rely on *decompositions* of Gaussians being well defined (as elliptical Gaussians). A possible direction of future research is to see if we can instantiate such samplers with decomposable smoothening functions, but this necessarily rules out smoothening functions which (loosely speaking) cannot be decomposed as a convolution of two density functions. In a separate direction, since we know the Schwartz space is closed under convolution, an interesting question is whether we can precisely analyze iterative samplers in a way which produces distributions that are independent from private information, while allowing the one-dimensional and final smoothening functions to differ.

A Characterization of Gaussian Functions

We now provide a characterization of the Gaussian distribution in terms of a property of it that is commonly used in proof techniques appearing in the lattice cryptography literature. Namely, a spherical Gaussian distribution can be factored over any given orthogonal basis. A direct application of Theorem 6.9 is that the spherical Gaussian is the only distribution with this property.

Definition 6.8. Consider a function $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}$ and let $\mathbf{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a set of n linearly independent vectors. We say that $\mathbf{f}$ *factors over* $\mathbf{B}$ if there exist functions $f_1, \dots, f_n\colon \mathbb{R} \to \mathbb{R}$ such that for all $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{f}(\mathbf{x}) = \prod_{i\in[n]} f_i\big(\langle \mathbf{b}_i, \mathbf{x}\rangle\big)$.

Recall that a set of real random variables $\{Z_1, \dots, Z_n\}$ is pairwise independent if and only if for all $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, the probability distribution function $\mathbf{f}$ of the combined random variable $(Z_1, \dots, Z_n)$ can be written as

$$\mathbf{f}(\mathbf{x}) = \prod f_i(x_i).$$

As a consequence, $\mathbf{f}$ factors over a basis $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ if and only if the random variables

$$Z_i = \langle \mathbf{b}_i, \mathbf{X}\rangle$$

are pairwise independent. The following result, better known as the Kac-Bernstein's Theorem, is a characterization of the Gaussian distribution—originally proved by Bernstein [Ber41] and Kac [Kac39] independently—and later generalized by Lukacs and King in [LK54].

Theorem 6.9 ([LK54])**.** *Let $X_1, \dots, X_n$ be independently distributed random variables. For each suppose that the nth moment of X_i exists and let s_i denote its variance. Given two vectors $(a_1, \dots, a_n), (b_1, \dots, b_n) \in \mathbb{R}^n$ consider the random variables given by the linear combinations $Z_1 = \sum_{i\in[n]} a_i X_i$ and $Z_2 = \sum_{i\in[n]} b_i X_i$. Then Z_1 and Z_2 are independent random variables if and only if*

1. *for each $i \in [n]$, if $a_i b_i \neq 0$, then X_i is Gaussian and*

2. *$\sum_{i\in[n]} a_i b_i s_i^2 = 0$.*

Corollary 6.10. *Let $\mathbf{f}$ be a probability distribution over $\mathbb{R}^n$ which factors over two different orthonormal bases $\mathbf{B}_1$ and $\mathbf{B}_2$. If $\mathbf{B}_1 \cup \mathbf{B}_2$ has no two collinear vectors, then $\mathbf{f}$ is a Gaussian distribution. Moreover, if $\mathbf{B}_2$ contains a vector whose coordinates are all non-zero with respect to $\mathbf{B}_1$, then $\mathbf{f}$ is ℓ_2-symmetric.*

Proof. Let $\mathbf{X}$ denote the random variable associated to $\mathbf{f}$. Without loss of generality, $\mathbf{f}$ factors over the bases $\mathbf{B}_1$ and $\mathbf{B}_2$, where $\mathbf{B}_1 = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is the canonical basis. Then, for every $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, $\mathbf{f}(\mathbf{x}) = \prod_{i\in[n]} f_i(x_i)$. Thus the random variables $X_1, \dots, X_n$ describing each coordinate of $\mathbf{X}$ are pairwise independent.

To show that each coordinate of $\mathbf{X}$ is a Gaussian random variable, fix and index $i \in [n]$. Notice that $\mathbf{B}_2$ contains, at least, two vectors $\mathbf{u} = (a_1, \dots, a_n), \mathbf{v} = (b_1, \dots, b_n)$ such that

$a_i b_i \neq 0$—as otherwise, since $\mathbf{B}_2$ is an orthonormal basis, the uniqueness of such vector in $\mathbf{B}_2$ implies that the corresponding entry is ± 1, therefore we conclude that $\pm\mathbf{e}_i \in \mathbf{B}_2$, contradicting the hypotheses. Let $Z_1 = \langle \mathbf{u}, \mathbf{X} \rangle$, $Z_2 = \langle \mathbf{v}, \mathbf{X} \rangle$. Since $\mathbf{f}$ factors over $\mathbf{B}_2$, the random variables Z_1 and Z_2 are independently distributed. Thus, by Theorem 6.9, X_i is a Gaussian distribution.

Finally, we prove that for all $i \in [n]$, the variance s_i is equal to a constant. Let $\mathbf{B}_2 = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ and for each $i \in [n]$ write $\mathbf{v}_i = (v_{i,1}, \dots, v_{i,n})$. Without loss of generality, assume that every coordinate of $\mathbf{v}_1$ is non-zero. Consider the matrix

$$M = \left(\begin{array}{c|c|c|c} v_{1,1} \cdot \begin{bmatrix} v_{2,1} \\ \vdots \\ v_{n,1} \end{bmatrix} & v_{1,2} \cdot \begin{bmatrix} v_{2,2} \\ \vdots \\ v_{n,2} \end{bmatrix} & \cdots & v_{1,n} \cdot \begin{bmatrix} v_{2,n} \\ \vdots \\ v_{n,n} \end{bmatrix} \end{array} \right).$$

Since $v_{1,1}, \dots, v_{1,n}$ are all non-zero, M is equivalent to a matrix whose rows are the vectors $\mathbf{v}_2, \dots, \mathbf{v}_n$. Thus it is clear that the rank of M is $n-1$. Let $V \subset \mathbb{R}^n$ be the space spanned by the rows of M. Notice that, for every row of M, the sum of its entries is equal to 0. It follows that $V^\perp$ is the subspace generated by the vector $(1, \dots, 1)$, since this sum is the inner product of the corresponding $\mathbf{v}_i$ with $\mathbf{v}_1$. Since the vector $(s_1^2, \dots, s_n^2) \in V^\perp$, the result follows. □

Notice that the last condition is necessary to prove the ℓ_2-symmetry of $\mathbf{f}$ in the previous result.

From DGS to SIVP

In Chapter 5, we showed how to find smoothening functions and smoothing parameters from tail bounds of the function $\widehat{\mathbf{f}}$ living in the Fourier space. Beyond smoothening, we often require that samples from a discrete distribution have bounded norm with high probability. To this end we may make use of the tools developed for tail bounds on functions in the primal space. Recall that when $\mathbf{f}$ is a probability measure over the lattice, we can also interpret tail bounds as probabilities

$$\begin{aligned} \Pr_{\mathbf{x} \sim D_{\mathcal{L},\mathbf{f}}} \left[\|\mathbf{x}\|_p > r \right] &= \Pr_{\mathbf{x} \sim D_{\mathcal{L},\mathbf{f}}} [\mathbf{x} \in \mathcal{L} \setminus rB_n^p] \\ &= \frac{\mathbf{f}\left(\mathcal{L} \setminus rB_n^p\right)}{\mathbf{f}\left(\mathcal{L}\right)} \\ &\leq \nu_{\mathbf{f}}(rB_n^p). \end{aligned}$$

It follows that the probability of a sample having ℓ_p norm bounded by r can be expressed as

$$\Pr_{\mathbf{x}\sim D_{\mathcal{L},\mathbf{f}}}\big[\|\mathbf{x}\|_p \le r\big] \ge 1 - \nu_{\mathbf{f}}(rB_n^p).$$

Hence, by taking a suitably large r, we can ensure a large amount of the measure is assigned to lattice points with a proportionally bounded norm. Although this approach is very general, the exact statements we can make necessarily depend on what tail bounds we can demonstrate for a particular function—which may be challenging even given a closed form expression. Similar bounds on elements sampled from normalized scalings/shifts of the function also follow analogously. In the latter case, this is achieved by leveraging a tail bound that holds for all cosets. Hence, a weaker tail bound which holds only for the family of lattices in $\mathbb{R}^n$ may be insufficient for this purpose. However when $\mathbf{f}$ is ε-smoothening, the effect of the shift can also be bounded in terms of ε, in which case a limited tail bound may be extensible.[2] These intuitions are formally expressed in the next lemma.

Lemma 6.11. *Let $\mathcal{L} \subset \mathbb{R}^n$ be an n-dimensional lattice, and let $\mathbf{f} \in \mathcal{F}$. For $p \in \mathbb{R}_{>1}$ consider the ℓ_p norm over $\mathbb{R}^n$. Then for any $\varepsilon \in (0,1)$, $s \ge \eta_{\mathbf{f},\varepsilon}(\mathcal{L})$, $r \in \mathbb{R}_{>0}$, and vector $\mathbf{c} \in \mathbb{R}^n$ we have*

$$\Pr_{\mathbf{x}\sim D_{\mathcal{L},\mathbf{f},s,\mathbf{c}}}\big[\|\mathbf{x}-\mathbf{v}\|_p > sr\big] \le \frac{1+\varepsilon}{1-\varepsilon}\cdot \nu_{\mathbf{f}}(rB_n^p). \tag{6.12}$$

Additionally, for any vector $\mathbf{c} \in \mathcal{L}$ (e.g. $\mathbf{c} = \mathbf{0}$),

$$\Pr_{\mathbf{x}\sim D_{\mathcal{L},\mathbf{f},s,\mathbf{c}}}\big[\|\mathbf{x}-\mathbf{v}\|_p > sr\big] \le \eta_{\mathbf{f}}(rB_n^p). \tag{6.13}$$

Proof. We can write the probability in Equation (6.12) as

$$\frac{\mathbf{f}_s\big((\mathcal{L}-\mathbf{c}) \setminus rB_n^p)\big)}{\mathbf{f}_{s,\mathbf{c}}(\mathcal{L})}.$$

First, note that the numerator is upper bounded as

$$\begin{aligned}\mathbf{f}_s\big((\mathcal{L}-\mathbf{c}) \setminus srB_n^q\big) &\le \nu_{\mathbf{f}}\big((1/s)\cdot srB_n^q\big)\cdot \mathbf{f}_s(\mathcal{L})\\ &= \nu_{\mathbf{f}}(rB_n^q)\cdot \mathbf{f}_s(\mathcal{L}).\end{aligned}$$

[2] Note that in the case of $\mathbf{c} \in \mathcal{L}$ as in the second case of Lemma 6.11, it is also sufficient to take $s \in \mathbb{R}_{\ge 1}$ and any $\mathbf{f} \in D_n$.

which is well-defined for any r such that $\eta_{\mathbf{f}}(rB_n^q) \in (0,1]$. Next, if $\mathbf{f}$ is ε-smoothening, by Proposition 5.16 it follows that

$$\frac{\mathbf{f}_s(\mathcal{L})}{\mathbf{f}_{s,\mathbf{c}}(\mathcal{L})} \cdot \nu_{\mathbf{f}}(rB_n^p) \leq \frac{1+\varepsilon}{1-\varepsilon} \cdot \eta_{\mathbf{f}}(rB_n^p).$$

Conversely, for any $\mathbf{c} \in \mathcal{L}$, we have $\mathbf{f}_{s,\mathbf{c}}(\mathcal{L}) = \mathbf{f}_s(\mathcal{L})$, which yields the second result. □

Lemma 6.12. *Let $\mathcal{L} = \mathcal{L}(\mathbf{B}) \subset \mathbb{R}^n$ be a full-rank lattice and let $\varepsilon \in (0, \frac{1}{2})$. Consider a normalized positive function $f\colon \mathbb{R}^n \to \mathbb{R}$ that is ε-smoothening for $\mathcal{L}$. Then after n^2 samples from $D_{\mathcal{L},f}$ we obtain a set of n linearly independent vectors from $\mathcal{L}$ with probability, at least, $1 - \left(\frac{1}{2} + 2\varepsilon\right)^n$.*

Proof. We consider the following event. Sample a set $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{L}$ according to $D_{\mathcal{L},f}$ and consider the matrix $\mathbf{A}$ whose columns are these vectors. Let $\mathbf{C} = \mathbf{B}^{-1}\mathbf{A} \bmod 2 \in \mathbb{Z}_2^{n\times n}$. By hypothesis, the function f is ε-smoothening for $2\mathcal{L}$; therefore, by Corollary 6.5, the distribution of the column vectors of $\mathbf{C}$ within 2ε in statistical distance from the uniform distribution over $\mathbb{Z}_2^n$. Since applying a function does not increase the statistical distance, the mapping $\mathbf{C} \mapsto \det \mathbf{C}$ induces a probability distribution over $\mathbb{Z}_2$ that is within 2ε from uniform. Thus, $\det \mathbf{C} = 0$ with probability, at most, $\frac{1}{2} + 2\varepsilon$.

By repeating this experiment n times, we obtain a matrix $\mathbf{C}$ with determinant 1 with probability at least $1-\left(\frac{1}{2} + 2\varepsilon\right)^n$. Since reducing modulo 2 (or any modulus) maps matrices of determinant 0 to matrices of determinant 0; we can conclude that the matrix $\mathbf{B}^{-1}\mathbf{A}$ has non-zero determinant, thus it is invertible over $\mathbb{R}$. As a consequence, the column vectors of $\mathbf{A}$ are linearly independent. The result follows. □

We now show a generalization of the SIVP to DGS reduction. The core approach of the reduction is the same as [Reg05, Lemma 3.17], and gives a result that is very similar to the natural generalization of the original, with two slight modifications. First, while the original reduction only found solutions to SIVP_γ in the ℓ_2 norm, our solution allows us to find solutions in any ℓ_p norm directly (as opposed to passing through norm inequalities), based on the tail bounds of the function f. This allows us to exploit symmetries in the function f, such as factoring through the ℓ_p norm. A second change is that the reduction works even when a lower bound on the smoothing parameter is not known, with the caveat that we can only guarantee solutions up to the best known upper bound h for $\eta_{\mathbf{f},\varepsilon}(\mathcal{L})$.

Lemma 6.13 (SIVP to $DS_{\mathbf{f},\varphi}$)**.** *Let $\varepsilon \in (0, \frac{1}{2})$, and consider any ℓ_p norm. Let $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}$ be a function in $\mathcal{F}_n'$ and suppose there exists $k \geq 1$ such that $1-\nu_{\mathbf{f}}(kB_n^p)$ is non-negligible in n. Then, for every choice of full-rank lattice $\mathcal{L} \subset \mathbb{R}^n$, any efficiently computable function*

$h(n) = \Omega\left(\frac{1}{\mathrm{poly}(n)}\right)$ such that $h(n) \cdot \lambda_n^p(\mathcal{L}) > \eta_{\mathbf{f}}(\mathcal{L})$, and any function $\varphi(\mathcal{L}) \geq h(n) \cdot \lambda_n^p(\mathcal{L})$ there exists a polynomial time reduction from $\mathrm{SIVP}_{2k\varphi}^p$ to $\mathrm{DS}_{\mathbf{f},\varphi}$.[3] Additionally, given any efficiently computable function $\pi(n) = \frac{1}{\mathrm{poly}(n)}$ such that $\eta_{\mathbf{f}}(\mathcal{L}) \geq \pi(n) \cdot \lambda_n^p(\mathcal{L})$, this holds for any $\varphi(\mathcal{L}) \geq \eta_{\mathbf{f}}(\mathcal{L})$.

Proof. To show this, we follow the reduction of [Reg05, Lemma 3.17]. Given a lattice $\mathcal{L}$, we first apply the LLL algorithm to obtain n linearly independent vectors of length at most $2^n\lambda_n \leq 2^{2n}\lambda_n^p$ (by norm inequalities), which we label S, and denote $\|S\|_p$ as $\widetilde{\lambda_n^p}$.

If $\varphi(\mathcal{L}) > \widetilde{\lambda_n^p}$, then S is already shorter than $2k\varphi(\mathcal{L})$. Otherwise, assume $\varphi(\mathcal{L}) \leq \widetilde{\lambda_n^p}$. For each $i \in \left[\left\lceil \log_2 \frac{2^{2n}\lambda_n^p}{h(\mathcal{L})\lambda_n^p} \right\rceil\right]$, call the $\mathrm{DS}_{\mathbf{f}}$ oracle $O(n^2)$ times with the pair $\left(\mathcal{L}, r_i = \widetilde{\lambda_n^p} 2^{-i}\right)$, and let S_i be the resulting set of vectors. Then, look for a set of n linearly independent vectors in each of $S, S_0, \ldots, S_{2n}$ and output the shortest set (in the ℓ_p norm) found. To prove correctness, note that $\varphi(\mathcal{L}) \leq \widetilde{\lambda_n}$ implies there must exist an $i \in \{1, \ldots, 2n\}$ with $\varphi(\mathcal{L}) < r_i \leq 2\varphi(\mathcal{L})$. Note that since $\varphi(\mathcal{L}) \geq h(n) \cdot \lambda_n^p(\mathcal{L}) = \Omega\left(\frac{1}{\mathrm{poly}(n)}\right) \cdot \lambda_n^p(\mathcal{L})$, the number of i's we need to consider is

$$\left\lceil \log_2 \frac{2^{2n}\lambda_n^p(\mathcal{L})}{\varphi(\mathcal{L})} \right\rceil \leq \left\lceil \log_2 \frac{2^{2n}\lambda_n^p(\mathcal{L})}{h\left(n\right) \cdot \lambda_n^p(\mathcal{L})} \right\rceil < \log_2 \frac{2^{2n}\lambda_n^p(\mathcal{L})}{\frac{1}{\mathrm{poly(n)}} \cdot \lambda_n^p(\mathcal{L})} < \mathrm{poly}(n).$$

The case where $\varphi(\mathcal{L}) \geq \eta_{\mathbf{f}}(\mathcal{L}) \geq \pi(n) \cdot \lambda_n^p(\mathcal{L})$ for some $\pi(n) = \frac{1}{\mathrm{poly}(n)}$ is similar. By Lemma 6.12, S_i contains n linearly independent vectors with very high probability. Finally, by Lemma 6.11, all vectors output by the $\mathrm{DS}_{\mathbf{f}}$ oracle, which come from the distribution of $\mathbf{f}_{r_i}$ over the lattice $\mathcal{L}$, are in the set $kr_iB_n^p$ and satisfy $\|\mathbf{x}_i\|_p \leq kr_i \leq 2k\varphi(\mathcal{L})$ with probability at least

$$1 - \nu_{\mathbf{f}}(kB_n^p),$$

which is non-negligible by the primal tail bound assumption. This completes the proof. □

Remark 6.14. Note that any bound on $\eta_{\mathbf{f}}(\mathcal{L})$ given in terms of $1/\lambda_1^\infty(\mathcal{L}^*)$ (e.g. those given by Corollary 5.41) can be expressed in terms of $\lambda_n^p(\mathcal{L})$ using transference theorems (e.g. Lemma 2.9) and norm inequalities.

[3] The condition $f \in \mathcal{F}_n'$ restricts this reduction to *interesting* instantiations—that is, instantiations that have polynomially sized smoothing parameters, and hence (potentially) polynomially sized approximation factors for SIVP_γ. The lower bound requirement for $h(n)$ or $\varphi(\mathcal{L})$ is needed to guarantee a polynomial time reduction and is the only limitation we include, although we note that choosing an exponentially large $h(n)$ or $\varphi(\mathcal{L})$ would yield similarly sized approximation factors.

6.4 Conclusion

In this chapter we answer the question: is it sufficient to consider the smoothening property of the Gaussian for the purposes of lattice cryptography, in particular LWE? More precisely, we explore the possibility of using a generic smoothening function to reconstruct an average-case to worst-case reduction from lattice problems to LWE. Our conclusion can be summarized in the following points:

- In the case of LWE, the techniques used in [Reg05] cannot directly be adapted to use a generic smoothening function. Nonetheless, we believe that the remaining roadblocks can be circumvented using alternative paths, as is the case for the proof of Lemma 6.13.
- There are results in the literature whose proofs can only work for Gaussian functions. As an example, the sampling algorithm using Nearest Planes, as described in [GPV08], can only work for Gaussian distributions.

Chapter 7

A Geometric Approach to LHN

"Every explorer is therefore, by necessity, a revolutionary, and every successful revolutionary is a peacemaker."
— Jordan B Peterson in *Maps of Meaning: The Architecture of Belief*

We dedicate this chapter to exploring what is beyond the boundaries of the real space with regards to the underlying theory behind lattice cryptography. In Chapter 5, we developed a general theory of the smoothing parameter that extends to an infinite family of functions. By doing so we proved that several properties that were only previously known for the Gaussian distribution can be found in an infinite family of functions. This opens the possibility of instantiating certain lattice problems and cryptosystems with non-Gaussian errors. In Chapter 6 we studied the possibility of obtaining an average-case to worst-case reduction that is independent from Gaussians. We do so by making an analysis and identifications of the specific properties of the Gaussian that are used in classical reductions such as [Reg05, LPR10].

Continuing with that philosophy, in this chapter we intend to extend the theory of smoothening functions to non-real spaces. We start this study by compiling what is known and relevant to the theory in the field of locally compact Abelian groups, and proving that, from an abstract point of view, being smoothening can be seen as a structural property of the function and the group.

It is worth mentioning that several results and techniques shown in this chapter are similar to those appearing in recent works of probability in abstract structures. As an example, Proposition 7.15 is a generalization of [App17, Theorem 6.1].

7.1 Locally-Compact Abelian Groups

The algebraic structure present in the set of real numbers is intuitively different from that of a generic group. The structure of the operations in it is far from arbitrary, in the sense that both addition and multiplication preserve closeness between elements. This idea is traditionally abstracted formally in the following definition.

Definition 7.1. A *topological group* is a group G endowed with a topology such that the functions

$$G \times G \to G \colon (g, h) \mapsto gh \quad \text{and} \quad G \to G \colon g \mapsto g^{-1}$$

are both continuous, where the topology considered for $G \times G$ is the product topology.

Another important property of $\mathbb{R}$ as a topological group is that it shows a certain degree of regularity. Two points can always be separated, and every point can be seen as contained in a compact space. This important property is known as *local compactness* in topology. We are particularly interested in topological groups with such a property.

Definition 7.2 (Locally Compact Group)**.** A topological space is said to be *locally compact* if every element of it has a compact neighborhood. A *locally compact Abelian* (LCA) group is a topological Abelian group whose underlying topological space is locally compact and Hausdorff.

Measures on Locally Compact Groups

The structure of a topological space allow us to regard it as a measurable space—by considering the σ-algebra of its Borel sets (see Section 2.4). When this space is a group, it is natural to wonder whether this structure is compatible with the group operation.

Definition 7.3 (Invariant Measure)**.** Let G be a locally compact group. A measure μ over G is called *left invariant* (respectively *right invariant*) if for every $g \in G$ and every measurable set $E \in \mathcal{M}$, $\mu(E) = \mu(gE)$ (respectively, $\mu(E) = \mu(Eg)$). The measurable space is called *left invariant* (respectively *right invariant*) if every element in $\mathcal{M}$ is left invariant (respectively right invariant). A measure is called *invariant* if it is both left and right invariant.

An invariant measure then allows us to say that *volume* of a set does not change under translations. This is an important feature of the Euclidean space $\mathbb{R}^n$. In particular, for a lattice $\mathcal{L} \subset \mathbb{R}^n$, any translation of a fundamental region has the same volume.

Note 7.4. Recall the definition measure and Borel sets given in Section 2.4 from Chapter 2. To provide a topological space with a measure, a slight variation of this is to consider *Baire sets*, which are the elements of the σ-algebra generated by the G_σ sets—the compact sets that can be expressed as a countable intersection of open sets. If X is second countable (in other words, it its topology has a countable basis), locally compact and Hausdorff, these two concepts are equivalent. In the rest of this chapter, we will only consider second countable spaces; thus measurable sets, Borel sets and Baire sets refer to the same concept in the context of a topological space. This distinction is important as it allows us to give a simplified version of Haar's Theorem. In general this version only holds for the σ-algebra of Baire sets.

Theorem 7.5 (Haar's Theorem). *Let G be a second countable locally compact topological group. Then there exists a left invariant measure over G that is finite on every compact set. Moreover, any two left invariant measures over G are equal up to multiplication by a constant.*

A (*left*) *Haar measure* is then defined as a (left) invariant measure over an LCA group. It follows then that if G is an LCA group, then there exist a unique—up to scalar multiplication—Haar measure for G. Theorem 7.5 result can be extended to topological groups that are not necessarily second countable, by requiring the Haar measures to satisfy inner and outer regularity properties. However, in this Chapter we only consider spaces that are second countable.

Integrals. The theory of Lebesgue integration allows us to define an integral over a measurable space by leveraging the σ-algebra of Borel sets. Thus a function is integrable if and only if it is measurable and the value of the integral over the group is finite. The set of integrable functions defined over a locally compact group G is denoted as $L^1(G)$.

Over a locally compact group with a Haar measure, this integral function is better known as the *Haar integral.* In this case, the change of variables $x \mapsto cx$ does not imply a change in the differential. A general change of variables, however, implies a change in the differential.

Definition 7.6. Let $(X, \mathcal{M})$, $(X', \mathcal{M}')$ be two measurable spaces and let $\mu\colon \mathcal{M} \to [0, \infty]$ be a measure over X. For a surjective measurable function $f\colon X \to X'$, the *pushforward* *measure of μ along f* is the measure over X' defined for $A \in \mathcal{M}'$ as

$$f_*\mu(A) \coloneqq \mu\big(f^{-1}(A)\big).$$

It is not immediate from the previous definition that a pushforward measure must be invariant. In the following proposition we use the surjectivity of the function to prove that $\varphi * \mu$ is indeed a Haar measure over H.

Proposition 7.7. *Let G, H be locally compact groups and let $\varphi\colon G \to H$ be a measurable surjective homomorphism. Then the pushforward of the Haar measure in G is a Haar measure in H.*

First proof of Proposition 7.7. Let $x \in G$ and let A be a measurable set in H. Then we have that

$$\begin{aligned}
\varphi_*\mu(xA) &= \mu\big(\varphi^{-1}(xA)\big)\\
&= \mu\big(x^{-1} \cdot \varphi^{-1}(xA)\big)\\
&\leq \mu\Big(\varphi^{-1}\big(\varphi\big(x^{-1}\varphi^{-1}(xA)\big)\big)\Big)\\
&= \mu\Big(\varphi^{-1}\big(\varphi(x^{-1}) \cdot \varphi\big(\varphi^{-1}(xA)\big)\big)\Big)\\
&= \mu\Big(\varphi^{-1}\big(x^{-1} \cdot (xA)\big)\Big)\\
&= \mu\big(\varphi^{-1}(A)\big)\\
&= \varphi_*(\mu)(A).
\end{aligned}$$

Therefore

$$\varphi_*\mu(A) = \varphi_*\mu\big(x^{-1} \cdot (xA)\big) \leq \varphi_*\mu(xA) \leq \varphi_*\mu(A),$$

which implies that $\varphi_*\mu(x + A) = \varphi_*\mu(A)$. This proves that $\varphi_*\mu$ is left-invariant, thus by Haar's Theorem (Theorem 7.5), $\varphi_*\mu$ is a Haar measure over H, as desired. □

Harmonic Analysis on Locally Compact Groups

The theory that has been presented up to this point in the chapter is independent of whether the corresponding group is Abelian. This is somewhat intentional. We leave the restriction to Abelian groups until this subsection to emphasize that this the point where the commutative structure of the group becomes important. Although harmonic analysis has been studied on non-Abelian groups, it is unclear whether is possible to develop a similar theory as the one presented here. Thus, for the rest of this chapter, we focus our attention on locally compact Abelian groups, and use the additive notation to denote the group operation.

Pontryagin Duality. In the theory of lattices on the Euclidean space it is common to study or quantify properties of a lattice in terms of properties of the dual lattice. Classic examples include the smoothing parameter, successive minima, transference theorems etc. In this particular case, from an algebraic point of view, this duality allows us to regard vectors as functions defined over the primal space as $\mathbf{y}\colon \mathbf{x} \mapsto e^{-2\pi i\langle \mathbf{x},\mathbf{y}\rangle}$. This idea in traditionally encapsulated and transported to a more general framework through the following definition.

Definition 7.8 (Dual Group)**.** For a topological Abelian group G, the *dual group* $\widehat{G} = \mathrm{Hom}(G, \mathbb{T})$ is the set of continuous homomorphisms $y\colon G \to \mathbb{T} = \mathbb{R}/\mathbb{Z}$.

In the case of LCA groups, the set $\widehat{G}$ is also known as the *Pontryagin dual* of G. This set has a group structure with the operation given by pointwise multiplication, that is, $\big(\sigma \cdot \tau\big)(g) := \sigma(g)\tau(g)$. Moreover, $\widehat{G}$ can be endowed with a particular topology which makes $\widehat{G}$ a topological group. Furthermore, if G is a LCA group, then $\widehat{G}$ (with the aforementioned topology) is also LCA. For completeness we describe this topology in the following definition.

Definition 7.9 (Compact-Open Topology)**.** Consider a LCA group G and let $S^1 \cong \mathbb{R}/\mathbb{Z}$ be the unit circle in $\mathbb{C}$. Let $C(G, S^1)$ denote the set of continuous functions $G \to S^1$. For a compact subset $K \subseteq G$ and an open set $U \subseteq S^1$. Let $V(K, U)$ be the subset of $C(G, S^1)$ consisting of the functions f that map K into U; in other words, $V(K, U) = \big\{f \in C(G, S^1)\colon f(K) \subseteq U\big\}$. The *compact-open* topology on $C(G, S^1)$ is the minimal topology such that for every compact set $K \subseteq G$ and open set $U \subseteq S^1$, the set $V(K, U)$ is open. Thus the set $C(G, S^1)$, endowed with the compact-open topology, is a topological group under pointwise multiplication.

By definition of the group operation on $\widehat{G}$, $(x, y) \mapsto y(x)$ defines a bilinear form from $G \times \widehat{G}$ to $\mathbb{R}$. For this reason, given $x \in G$ and $y \in \widehat{G}$ we denote $\langle x, y\rangle := y(x)$. In the case of the real space, the dual group of $\mathbb{R}^n$ is isomorphic to the same group. As mentioned before, the elements of the dual group can, more precisely, be described as the functions $\mathbf{x} \mapsto e^{-2\pi i\langle \mathbf{x},\mathbf{y}\rangle}$. Pointwise multiplication of two of these functions, $e^{-2\pi i\langle \mathbf{x},\mathbf{y}\rangle}$ and $e^{-2\pi i\langle \mathbf{x},\mathbf{y}\rangle}$, agrees with the addition of the corresponding defining vectors $\mathbf{y}$, $\mathbf{y}'$.

It is, perhaps, expected that not every group is isomorphic to its dual. Consider the group $\mathbb{Z}^n$. Then any homomorphism from $\mathbb{Z}^n$ to S^1 is given by $\mathbf{x} \mapsto e^{2\pi i\langle \mathbf{x},\mathbf{y}\rangle}$, for some $\mathbf{y} \in \mathbb{R}^n$. However, for $\mathbf{a} \in \mathbb{Z}^n$ we have that $e^{2\pi i\langle \mathbf{x},\mathbf{y}\rangle} = e^{2\pi i\langle \mathbf{x},\mathbf{y}+\mathbf{a}\rangle}$. As a consequence, the dual group of $\mathbb{Z}^n$ is isomorphic to $\mathbb{R}^n/\mathbb{Z}^n$.

A group that is isomorphic to its dual group is called *self-dual*. By the previous example, not every group is self-dual. Nonetheless, it is possible to draw a relationship between an LCA group and its double dual via the following theorem.

Theorem 7.10 (Pontryagin Duality)**.** *Let G be an* LCA *group. Then the mapping*

$$\mathrm{ev}\colon x \mapsto \big(y \mapsto \langle x, y\rangle\big)$$

is an isomorphism between G and its double dual $\widehat{(\widehat{G})}$.

Finally we state an important consequence of Theorem 7.10. Recall that a short exact sequence is a sequence of homomorphisms $0 \to H \to G \to K \to 0$ where the image of one homomorphism equals the kernel of the following. Given a short exact sequence of LCA groups, their duals are also in a sequence.

Proposition 7.11 ([Rud62, Section 2.1])**.** *Let G, H, K be* LCA *groups and suppose that*

$$0 \to H \to G \to K \to 0$$

is a short exact sequence. Then their corresponding dual groups form the short exact sequence

$$0 \to \widehat{K} \to \widehat{G} \to \widehat{H} \to 0.$$

Quotient Measure. A sequence LCA of groups $0 \to H \to G \to K \to 0$ is a *short exact sequence of* LCA *groups* if H is closed in G and H and K are equipped with the subset and quotient topologies, respectively. Given such a sequence, then there exists a measure μ_K over K such that for any integrable function $f\colon G \to \mathbb{C}$, we have that the following Fubini-type formula—which is also known as the Weil formula—holds.

$$\int_G f(g)d\mu_G = \int_K \int_H f(h+k)d\mu_H d\mu_K, \tag{7.1}$$

where $\int_H f(h+k)d\mu_H$ denotes $\int_H f(h+g)d\mu_H$, where $g \in G$ is any element that maps to $k \in K$. (See [RRS00, p. 87-88].)

Definition 7.12 (Quotient Measure)**.** Given an LCA group G and a subgroup $H \leq G$, the *quotient measure* of G/H is the measure obtained from Equation (7.1) after considering the sequence $0 \to H \to G \to G/H \to 0$.

The Fourier Transform. Real-valued integrable functions over $\mathbb{R}$ can, traditionally, be related to their counterparts over the dual space via the Fourier transform. It is a fundamental tool for the content of chapters 5 and 6. For LCA groups, this concept can be derived from their Haar measure.

Definition 7.13 (Fourier Transform over LCA groups)**.** Let G be an LCA group and let μ be a Haar measure over G. Given a $f \in L^1(G)$, the *Fourier transform* of f is the function $\widehat{f}\colon \widehat{G} \to \mathbb{R}$ defined as

$$\widehat{f}\colon y \mapsto \int_G f(x)\langle x, y\rangle d\mu(x). \tag{7.2}$$

Stretching and Translating Maps. In an LCA group G, let $x, c \in G$. For $s \in \mathbb{Z}$ let

$$s \cdot x \coloneqq \operatorname{sign}(s) \sum_{i=1}^{|s|} x = \begin{cases} \sum_{i=1}^{|s|} x & \text{if } s \in \mathbb{Z}_{\geq 0}, \\ -\sum_{i=1}^{|s|} x & \text{if } s \in \mathbb{Z}_{<0}. \end{cases}$$

Since group operations are continuous, this is a continuous function from G to itself. Moreover, since G is Abelian, $s \cdot (x + y) = \operatorname{sign}(s) \sum_{i=1}^{n}(x + y) = \operatorname{sign}(s) \sum_{i=1}^{n} x + \operatorname{sign}(s) \sum_{i=1}^{n} y$. Thus this mapping is a group homomorphism. Moreover, it is an injective homomorphism whenever the s-torsion subgroup of G is trivial.

Consider an injective homomorphism $\phi\colon G \to H$ between two groups. We say that ϕ is a *section* if there exists a homomorphism $\psi\colon H \to G$ such that $\psi \circ \phi = \mathrm{Id}_G$. Whenever $x \mapsto s \cdot x$ is a section, its retraction is a continuous surjective homomorphism. We denote this function as $x \mapsto s^{-1} \cdot x$. Furthermore, for $c \in G$ denote

$$\tau_{s,c}(x) \coloneqq s^{-1} \cdot (x - c).$$

Corollary 7.14. *Let G be a locally compact Abelian group and let $s \in \mathbb{Z}$ be such that the map $x \mapsto s \cdot x$ is a section. Let μ be a Haar measure on G and let $c \in G$. Then there exists a constant κ_s so that the pushforward measure of μ along $\tau_{s,c}$ is κ_s times μ.*

Proof. Since $x \mapsto s \cdot x$ is a section, $\tau_{s,c}$ is a measurable surjective homomorphism. Thus, by Proposition 7.7, $\tau_{s,c} * \mu$ is a Haar measure. The result follows from Theorem 7.5. □

Proposition 7.15. *Let G be a* LCA *group written additively and fix a Haar measure μ over G. Consider $f \in L^1(G)$ and let $\widehat{f}$ denote its Fourier transform. For $c \in G$ and $s \in \mathbb{Z}$ let*

$$f_{s,c}(x) \coloneqq f \circ \tau_{s,c}(x) = f\big(s \cdot (x + c)\big).$$

Then, for $y \in \widehat{G}$, we have that

$$\widehat{f_{s,c}}(y) = \frac{\langle c, y\rangle}{\kappa_s}\widehat{f}(s^{-1}y),$$

where $\kappa_s \in \mathbb{C}$ is a constant that depends on s.

Proof. Let $c \in G$, $s \in \mathbb{Z}$ and $y \in \widehat{G}$.

$$\begin{aligned}\widehat{f_{s,c}}(y) &= \int_G f_{s,c}(x)\overline{y}(x) \cdot d\mu(x) \\ &= \int_G f\big(\tau_{s,c}(x)\big)\overline{y}(x) d\mu(x)\end{aligned}$$

By Corollary 7.14, the pushforward measure of μ along $\tau_{s,c}$ is $\kappa_s\mu$. Making $z = \tau_{s,c}(x) = s \cdot (x + c)$ we have

$$\begin{aligned}\int_G f\big(\tau_{s,c}(x)\big)\overline{y}(x) d\mu(x) &= \int_G f(z)\overline{y}\left(s^{-1}z - c\right) \cdot d(\tau_{s,c} * \mu)(z) \\ &= \frac{1}{\kappa_s}\int_G f(z)\overline{y}\left(s^{-1}z\right)\overline{y}(-c) \cdot d\mu(z) \\ &= \frac{y(c)}{\kappa_s}\int_G f(z)\left((\overline{s^{-1}y})(z)\right) d\mu(z) \\ &= \frac{y(c)}{\kappa_s}\widehat{f}\left(s^{-1}\overline{y}\right),\end{aligned}$$

as required. □

Lattices Over LCA **groups.** Up to this point, the theory of locally compact Abelian groups is, perhaps, still quite distant from the traditional theory of lattices over $\mathbb{R}^n$. An important missing element is the concept of lattice itself. Recall that lattices over the Euclidean space are characterized as the discrete subgroups of $\mathbb{R}^n$. In general, a subgroup A of a topological group G is said to be *discrete* if its induced topology as a subspace of G is discrete.

Definition 7.16 (Lattice)**.** Let G be an LCA group and consider a Haar measure on G. A discrete subgroup $\mathcal{L} \le G$ is called a *lattice* if the quotient $G/\mathcal{L}$ is compact.

An equivalent condition for an LCA group to be compact is that its Haar measure is finite. Notice that, under this notion, a lattice $\mathcal{L} \subset \mathbb{R}^n$ is also a lattice in the sense of LCA groups if and only if the rank of $\mathcal{L}$ is n.

Another important missing element is the dual of a lattice $\mathcal{L} \leq G$. A priori, a reader might be tempted to take it as the Pontryagin dual $\widehat{\mathcal{L}}$ of $\mathcal{L}$. However, it follows from Proposition 7.11 $\widehat{\mathcal{L}}$ is not necessarily (isomorphic to) a subgroup of $\widehat{G}$. In fact, it rarely is. To construct the "dual of a lattice", we first look at the following proposition.

Proposition 7.17 ([Fol94, Proposition 4.4])**.** *Let G be an* LCA *group. Then G is compact if and only if $\widehat{G}$ is discrete.*

Remark 7.18. It follows from Theorem 7.10 that the converse also follows. In other words, an LCA group G is discrete if and only if its dual $\widehat{G}$ is compact.

As we remark above, the Pontryagin dual $\widehat{\mathcal{L}}$ of $\mathcal{L}$ is rarely a subgroup of $\widehat{G}$. On the other hand, if follows from Proposition 7.11 that

$$0 \to \widehat{G/\mathcal{L}} \to \widehat{G} \to \widehat{\mathcal{L}} \to 0 \tag{7.3}$$

is a short exact sequence. Therefore $\widehat{G/\mathcal{L}}$ is—or, more precisely, can be seen as—a subgroup of $\widehat{G}$. Moreover, since $G/\mathcal{L}$ is compact, it follows from Proposition 7.17, that $\widehat{G/\mathcal{L}}$ is a discrete group. Furthermore, by Remark 7.18, $\widehat{\mathcal{L}}$ is compact. Hence, since (7.3) is a short exact sequence, the group

$$\widehat{\mathcal{L}} \cong \widehat{G}\Big/\widehat{G/\mathcal{L}}$$

is compact. Thus $\widehat{G/\mathcal{L}}$ is a lattice. This motivates the following definition.

Definition 7.19 (Orthogonal Lattice)**.** Let G be an LCA group and let $\mathcal{L} \leq G$ be a lattice. The *orthogonal* lattice of $\mathcal{L}$ is defined as the group $\mathcal{L}^{\perp} := \widehat{G/\mathcal{L}}$.

Poisson Summation Formula. The Poisson Summation Formula presented in Chapter 5 for lattices and functions over $\mathbb{R}^n$ has its equivalent counterpart over LCA groups. In this context, the result is well-known to hold for functions in the *Schwartz-Bruhat* space, which is the natural generalization of the Schwartz space over $\mathbb{R}^n$. However, recently it was proved to hold for functions in the *Feichtinger algebra* $\mathcal{S}_0(G)$ (see [Jak18, Theorem 5.7]).

Lemma 7.20 (Poisson Summation Formula)**.** *Suppose we have a group G, a subgroup $A \leq G$, and $C = G/A$. If we equip these sets with the Haar measures μ_A, μ_G, μ_C respectively satisfying Equation (7.1), then for any function $f \in \mathcal{S}_0(G)$, we have*

$$\int_A f(a) d\mu_A(a) = \int_{\widehat{C}} \widehat{f}(\hat{c}) d\mu_{\widehat{C}}(\hat{c}). \tag{7.4}$$

Remark 7.21. If A is a lattice, then G/A is a compact LCA groups. Therefore, by Proposition 7.17, $\widehat{G/A}$ is also discrete, and Equation (7.4) becomes

$$\sum_{x\in A} f(x) = \frac{1}{\mu(G/A)} \sum_{y\in\widehat{G/A}} \widehat{f}(y).$$

In particular, if $G = \mathbb{R}^n$ and $A = \mathcal{L}$ is a real lattice, we recover the familiar expression

$$f(\mathcal{L}) = \frac{1}{\det \mathcal{L}} \widehat{f}(\mathcal{L}^*) = \det(\mathcal{L}^*)\widehat{f}(\mathcal{L}^*).$$

7.2 Smoothening Functions Over LCA Groups

We dedicate this section to reconstructing the fundamentals of smoothening functions that appear in Section 5.3 for locally compact Abelian groups. In particular, we show that important results such as [Reg05, Claim 3.8] (Lemma 5.14) and [MR07, Lemma 4.4] (Proposition 5.16) can be seen as results that are independent from the specific structure of $\mathbb{R}^n$.

In this section, let G be an LCA group and fix a Haar measure μ over G. This choice allows us to consider a unique integral over G. In addition, similar to chapters 5 and 6, for a function $f \in L^1(G)$ let D_f denote the mapping

$$D_f \colon x \mapsto \frac{1}{\widehat{f}(0)} f(x),$$

which represents the probability distribution induced by f over G.

Smoothening Functions Over LCA groups

We start with the definition of smoothening function. This is a direct generalization of Definition 5.10 for real lattices.

Definition 7.22 (Smoothening function)**.** Let $\mathcal{L} \leq G$ be a lattice. Given $\varepsilon \in \mathbb{R}_{>0}$, we say that a function $f \in L^1(G)$ is *ε-smoothening for $\mathcal{L}$* if

$$\left|D_{\widehat{f}}\right| \left(\mathcal{L}^\perp \setminus \{0\}\right) = \frac{1}{\widehat{f}(0)} \sum_{x\in\mathcal{L}^\perp\setminus\{0\}} \left|f(x)\right| < \varepsilon. \tag{7.5}$$

Claim 7.23. Let $\varepsilon \in \mathbb{R}_{>0}$ and let $f \in L^1(G)$ be an ε-smoothening function for a lattice $\mathcal{L} \leq G$. Then for every $s \in \mathbb{Z}_{>0}$ and $c \in G$, the function $f_{s,c}$ is ε-smoothening for $\mathcal{L}$.

Proof. The proof is a generalization of that given for Claim 5.11. Let $c \in G$, and $s \in \mathbb{Z}_{>1}$. Then, using Proposition 7.15, we have the following observation.

$$\begin{aligned}
\sum_{y \in \mathcal{L}^\perp \setminus \{0\}} \left| \widehat{D_{f,s,c}}(y) \right| &= \frac{1}{\widehat{f_{s,c}}(0)} \sum_{y \in \mathcal{L}^\perp \setminus \{0\}} \left| \widehat{f_{s,c}}(y) \right| \\
&= \frac{1}{\kappa_s \widehat{f}(0)} \kappa_s \sum_{y \in \mathcal{L}^\perp \setminus \{0\}} \left| \langle c, y \rangle \widehat{f}\,(ty) \right| \\
&\leq \frac{1}{\widehat{f}(0)} \sum_{y \in s\mathcal{L}^\perp \setminus \{0\}} \left| \widehat{f}\,(y) \right| \\
&= \sum_{y \in s\mathcal{L}^\perp \setminus \{0\}} \left| \widehat{D_f}(y) \right| \\
&\leq \sum_{y \in \mathcal{L}^\perp \setminus \{0\}} \left| \widehat{D_f}(y) \right| < \varepsilon,
\end{aligned}$$

where the second to last inequality holds since for $s \in \mathbb{Z}_{>0}$, $s\mathcal{L}^\perp$ is a subgroup of $\mathcal{L}^\perp$. This finishes the proof. □

Smoothing Parameters on LCA **groups.** In Section 5.3 we argue the difference between smoothening functions and functions for which the smoothing parameter exists. In short, a smoothening function $\mathbf{f}$ may not have a smoothing parameter (for the same lattice).

Given an arbitrary dilation $\mathbf{f}_s$, this may yield a Fourier transform whose weight outside the origin is arbitrarily large. There are two main reasons for this. The more evident one is that the Fourier transform may not be decreasing; thus its weight outside the origin may increase. The second (and, perhaps, less evident reason) is that the definition of smoothing parameter requires that any *real* dilation is smoothening. This requirement follows since $\mathbb{R}^n$, the domain of the function, is an $\mathbb{R}$-module.

Nonetheless, in general this situation is different because a generic LCA group is only guaranteed to be a $\mathbb{Z}$ module. By Claim 7.23, any integer dilation of an ε-smoothening function remains ε-smoothening.

Definition 7.24 (Discrete Smoothing Parameter)**.** Let G be an LCA group and let $f \in L^1(G)$. For a lattice $\mathcal{L} \leq G$ and a number $\varepsilon \in \mathbb{R}_{>0}$, define the *discrete smoothing parameter* for f with respect to $\mathcal{L}$, if exists, to be the minimum $s \in \mathbb{Z}_{>0}$ such that the function f_s is ε-smoothening.

Smoothening Functions and Lattices

The following result is a direct generalization of Lemma 5.14—which, in turn, generalizes [Reg05, Claim 3.8]. As we argue in Chapter 5, this result is a cornerstone for a large part of the theory and tools developed related to lattice cryptography.

Lemma 7.25. *Let G be an* LCA *group and let $\mathcal{L} \leq G$ be a discrete subgroup. Let $\varepsilon > 0$ and let f be an ε-smoothening function over G for $\mathcal{L}$. Then for any $z \in G$,*

$$f(\mathcal{L}+z) \in \frac{\widehat{f}(0)}{\mu(G/\mathcal{L})}\left(\varepsilon, -\varepsilon\right).$$

Proof. Let $z \in G$ and let $h(x) \coloneqq f(x+z)$. Then the weight of the function on the shifted lattice can be expressed as

$$\begin{aligned} f(\mathcal{L}+z) &= \sum_{x \in \mathcal{L}+z} f(x) \\ &= \sum_{x \in \mathcal{L}} h(x). \end{aligned}$$

By the Poisson summation formula (Lemma 7.20)

$$\begin{aligned} \sum_{x \in \mathcal{L}} h(x) &= \frac{1}{\mu(G/\mathcal{L})} \sum_{y \in \mathcal{L}^\perp} \widehat{h}(y) \\ &= \frac{1}{\mu(G/\mathcal{L})} \sum_{y \in \mathcal{L}^\perp} \widehat{f}(y)\langle y, z\rangle \\ &= \frac{\widehat{f}(0)}{\mu(G/\mathcal{L})}\left(1 + \sum_{y \in \mathcal{L}^\perp \setminus \{0\}} \widehat{D_f}(y)\langle y, z\rangle\right). \end{aligned}$$

Recall that f is a real-valued function and the real part of $\langle y, z\rangle \in [-1,1]$. Therefore, since f is ε-smoothening we have that $\sum_{y \in \mathcal{L}^\perp \setminus \{0\}} \widehat{D_f}(y)\langle y, z\rangle \in (-\varepsilon, \varepsilon)$, as required. □

Lastly, we state some of the results given in Section 5.3 in their most generic form, which are directly derived from Lemma 7.25. Since the proofs are both identical to their real counterparts, we omit their corresponding proofs.

Corollary 7.26. *Let f be an ε-smoothening function with respect to a discrete subgroup $\mathcal{L} \leq G$. Then for any $z \in G$,*

$$f(\mathcal{L}+z) \in \frac{1}{\mu(G/\mathcal{L})}\left(\int_G f \ d\mu + (-\varepsilon, \varepsilon)\right).$$

Corollary 7.27 ([MR07, Lemma 4.4])**.** *Let $\varepsilon \in \mathbb{R}_{>0}$ and let $f \colon G \to \mathbb{R}$ be an ε-smoothening function for a lattice $\mathcal{L} \leq G$. Then for any $x, y \in G$,*

$$f(\mathcal{L}+x) \in \left(\frac{1+\varepsilon}{1-\varepsilon}\right) f(\mathcal{L}+y).$$

7.3 A Case for Other Groups

In this last section we explore the possibility of using smoothening functions to obtain an average-case to worst-case relation among lattice problems over LCA groups. In addition, we propose a few directions for future research.

The Longest Vector Problem Over Local Fields. Another example of an LCA group is the p-adic reals. The field of *p-adic numbers* is defined as the set of formal power series

$$\mathbb{Q}_p := \left\{ \sum_{i \in \mathbb{Z}_{\geq j}} a_i p^i \colon a_i \in \{0, \dots, p-1\}, j \in \mathbb{Z} \right\}.$$

We may then consider the group $G = \mathbb{Q}_p^n$.

In [DLX18], Luo et. al. proposed several computational problems over the p-adic fields that are analogs to SVP and CVP. In the cited manuscript, the authors provide an argument for why, in the case of the p-adic fields, finding the shortest vector in a lattice is not a well-defined problem. Nonetheless, every lattice over $\mathbb{Q}_p$ has a longest non-trivial vector. The nature of the computational problems that can be defined in a certain group inherently depends on the geometry that the group is endowed with.

Groups With No Geometry. At the beginning of this thesis we mentioned that one of the most appealing characteristics of LWE and SIS is their seamless combination of algebra and geometry. More precisely, LWE is a combination of the algebraic and geometric and geometric aspects of a particular example, the group $\mathbb{R}^n$ and the lattices over it. However, in this chapter we argue that most of the relevant ideas can be materialized with topological and algebraic concepts.

This leads us to formulate the following questions:

- What is the topological nature of lattice problems over LCA groups?
- Is it possible to find an LCA group where the average case of LWE is connected to the worst case of a lattice problem?

7.4 Conclusion

The development of lattice cryptography accelerated in the last one and a half decades. This was greatly motivated by landmark publications such as [Reg05] and [MR07]. Since then, many aspects of the results and constructions that appear in these papers have changed, mainly for the sake of practicality. However, the amount of choices that have been made along the way has made us wonder if any of those are arbitrary, and to what extent some of them are indeed necessary.

In this chapter we explored the possibility of extending these ideas to the most abstract conceivable scenario, while still keeping part of the geometric and algebraic components that characterize the area. We do so with the hope of eventually finding other instantiations of LWE and SIS that yield cryptographic constructions that have different properties, are more efficient or depend on a different kind of mathematical problem.

References

[AAG99] Iris Anshel, Michael Anshel, and Dorian Goldfeld. An algebraic method for public-key cryptography. *Mathematical Research Letters*, 6(3):287–291, 1999.

[ACPS09] Benny Applebaum, David Cash, Chris Peikert, and Amit Sahai. Fast cryptographic primitives and circular-secure encryption based on hard learning problems. In *Proceedings of the 29th Annual International Cryptology Conference on Advances in Cryptology*, CRYPTO '09, pages 595–618. Springer-Verlag, 2009.

[ADPS16] Erdem Alkim, Léo Ducas, Thomas Pöppelmann, and Peter Schwabe. Post-quantum key exchange - A new hope. In *25th USENIX Security Symposium, USENIX Security 16, Austin, TX, USA, August 10-12, 2016.*, pages 327–343, 2016.

[ADS15] D. Aggarwal, D. Dadush, and N. Stephens-Davidowitz. Solving the closest vector problem in 2^n time – the discrete Gaussian strikes again! In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 563–582, 2015.

[AG11] Sanjeev Arora and Rong Ge. New algorithms for learning in presence of errors. In Luca Aceto, Monika Henzinger, and Jiří Sgall, editors, *Automata, Languages and Programming*, pages 403–415. Springer Berlin Heidelberg, 2011.

[Ajt96] M. Ajtai. Generating hard instances of lattice problems (extended abstract). In *Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing*, STOC '96, pages 99–108. ACM, 1996.

[AL88] Dana Angluin and Philip Laird. Learning from noisy examples. *Machine Learning*, 2:343–370, april 1988.

[Alb17] Martin R. Albrecht. On dual lattice attacks against small-secret LWE and parameter choices in HElib and SEAL. In Jean-Sébastien Coron and Jesper Buus Nielsen, editors, *Advances in Cryptology – EUROCRYPT 2017*, pages 103–129. Springer International Publishing, 2017.

[App17] David Applebaum. Probabilistic trace and poisson summation formulae on locally compact abelian groups. *Forum Math*, 29:501–17, 2017.

[APS15] Martin Albrecht, Rachel Player, and Samuel Scott. On the concrete hardness of learning with errors. *Journal of Mathematical Cryptology*, 9(3):169–203, 10 2015.

[Bab86] László Babai. On Lovász' lattice reduction and the nearest lattice point problem. *Combinatorica*, 6:1–13, March 1986.

[Ban93] W. Banaszczyk. New bounds in some transference theorems in the geometry of numbers. *Mathematische Annalen*, 296(1):625–635, Dec 1993.

[BBFR08] Gilbert Baumslag, Yegor Bryukhov, Benjamin Fine, and Gerhard Rosenberger. Some cryptoprimitives in noncommutative algebraic crytography. In *Aspects of infinite groups*, pages 26–44. World Scientific, 2008.

[BCD$^+$19] Joppe Bos, Craig Costello, Léo Ducas, Ilya Mironov, Michael Naehrig, Valeria Nikolaenko, Chris Peikert, Ananth Raghunathan, and Douglas Stebila. Frodo: Take off the ring! practical, quantum-secure key exchange from LWE. NIST PQC Standardization Process Round 2, 2019. https://frodokem.org/.

[BCSV19] Carl Bootland, Wouter Castryck, Alan Szepieniec, and Frederik Vercauteren. A framework for cryptographic problems from linear algebra. *Journal of Mathematical Cryptology*, 2019.

[Ber41] S Bernstein. Sur une propriete caracteristique de la loi de gauss. *Trans. Leningrad Polytechn. Inst*, 3:21–22, 1941.

[BFN$^+$11] Gilbert Baumslag, Nelly Fazio, Antonio R. Nicolosi, Vladimir Shpilrain, and William E. Skeith. Generalized learning problems and applications to noncommutative cryptography. In Xavier Boyen and Xiaofeng Chen, editors, *Provable Security*, pages 324–339. Springer Berlin Heidelberg, 2011.

[BFX06] Gilbert Baumslag, Benjamin Fine, and Xiaowei Xu. A proposed public key cryptosystem using the modular group. *Contemporary Mathematics*, 421:35, 2006.

[BGV12] Zvika Brakerski, Craig Gentry, and Vinod Vaikuntanathan. (Leveled) fully homomorphic encryption without bootstrapping. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12, pages 309–325. ACM, 2012.

[BHW93] U. Betke, M. Henk, and J. M. Wills. Successive-minima-type inequalities. *Discrete & Computational Geometry*, 9(2):165–175, Feb 1993.

[BKW00] Avrim Blum, Adam Kalai, and Hal Wasserman. Noise-tolerant learning, the parity problem, and the statistical query model. *J. ACM*, 50:2003, 2000.

[BLP$^+$13] Zvika Brakerski, Adeline Langlois, Chris Peikert, Oded Regev, and Damien Stehlé. Classical hardness of learning with errors. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, STOC '13, pages 575–584. ACM, 2013.

[Boy14] John P. Boyd. The Fourier transform of the quartic Gaussian $\exp(-a \times 4)$: Hypergeometric functions, power series, steepest descent asymptotics and hyperasymptotics and extensions to $\exp(-a \times 2n)$. *Applied Mathematics and Computation*, 241:75 – 87, 2014.

[BPR12] Abhishek Banerjee, Chris Peikert, and Alon Rosen. Pseudorandom functions and lattices. In David Pointcheval and Thomas Johansson, editors, *Advances in Cryptology – EUROCRYPT 2012*, pages 719–737. Springer Berlin Heidelberg, 2012.

[BZ97] John J. Benedetto and Georg Zimmermann. Sampling multipliers and the poisson summation formula. *Journal of Fourier Analysis and Applications*, 3:505–523, Sept 1997.

[CGH$^+$96] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the lambert w function. *Advances in Computational Mathematics*, 5:329–359, Dec 1996.

[CI14] Andrew M. Childs and Gábor Ivanyos. Quantum computation of discrete logarithms in semigroups. *Journal of Mathematical Cryptology*, 8:405–416, 2014.

[CKMS17] Sanjit Chatterjee, Neal Koblitz, Alfred Menezes, and Palash Sarkar. Another look at tightness ii: Practical issues in cryptography. In Raphaël C.-W. Phan and Moti Yung, editors, *Paradigms in Cryptology – Mycrypt 2016. Malicious* *and Exploratory Cryptology*, pages 21–55. Springer International Publishing, 2017.

[CZZ16] Qi Cheng, Jun Zhang, and Jincheng Zhuang. LWE from non-commutative group rings. Cryptology ePrint Archive, Report 2016/1169, 2016. https://eprint.iacr.org/2016/1169.

[DBPS18] Alex Dytso, Ronit Bustin, H. Vincent Poor, and Shlomo Shamai. Analytical properties of generalized Gaussian distributions. *Journal of Statistical Distributions and Applications*, 5:6, Dec 2018.

[DFR18] Luke Demarest, Benjamin Fuller, and Alexander Russell. Handling correlated errors: Hardness of LWE in the exponent. Cryptology ePrint Archive, Report 2018/1005, 2018. https://eprint.iacr.org/2018/1005.

[DGG16] Özgür Dagdelen, Sebastian Gajek, and Florian Göpfert. Learning with errors in the exponent. In Soonhak Kwon and Aaram Yun, editors, *Information Security and Cryptology - ICISC 2015*, pages 69–84. Springer International Publishing, 2016.

[DKRS03] I. Dinur, G. Kindler, R. Raz, and S. Safra. Approximating CVP to within almost-polynomial factors is NP-Hard. *Combinatorica*, 23:205–243, Apr 2003.

[DLX18] Yingpu Deng, Lixia Luo, and Guanju Xiao. On some computational problems in local fields. Cryptology ePrint Archive, Report 2018/1229, 2018. https://eprint.iacr.org/2018/1229.

[DXL12] Jintai Ding, Xiang Xie, and Xiaodong Lin. A simple provably secure key exchange scheme based on the learning with errors problem. Cryptology ePrint Archive, Report 2012/688, 2012. https://eprint.iacr.org/2012/688.

[EK04] Bettina Eick and Delaram Kahrobaei. Polycyclic groups: A new platform for cryptology? arXiv, 2004.

[FIN$^+$15] Nelly Fazio, Kevin Iga, Antonio R. Nicolosi, Ludovic Perret, and William E. Skeith. Hardness of learning problems over burnside groups of exponent 3. *Designs, Codes and Cryptography*, 75(1):59–70, Apr 2015.

[Fol94] G.B. Folland. *A Course in Abstract Harmonic Analysis*. Studies in Advanced Mathematics. Taylor & Francis, 1994.

[GINX16] Nicolas Gama, Malika Izabachène, Phong Q. Nguyen, and Xiang Xie. Structural lattice reduction: Generalized worst-case to average-case reductions and homomorphic cryptosystems. In *Proceedings, Part II, of the 35th Annual International Conference on Advances in Cryptology — EUROCRYPT 2016 - Volume 9666*, pages 528–558. Springer-Verlag New York, Inc., 2016.

[GMPW20] Nicholas Genise, Daniele Micciancio, Chris Peikert, and Michael Walter. Improved discrete Gaussian and subgaussian analysis for lattice cryptography. In Aggelos Kiayias, Markulf Kohlweiss, Petros Wallden, and Vassilis Zikas, editors, *Public-Key Cryptography - PKC 2020*, volume 12110 of *Lecture Notes in Computer Science*, pages 623–651. Springer, 2020.

[Gol10] Larry Goldstein. Bounds on the constant in the mean central limit theorem. *Ann. Probab*, 38(4):1672–1689, 2010.

[GPV08] Craig Gentry, Chris Peikert, and Vinod Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, pages 197–206. ACM, 2008.

[GR02] Lov Grover and Terry Rudolph. Creating superpositions that correspond to efficiently integrable probability distributions. arXiv, 2002.

[JAC$^+$17] David Jao, Reza Azarderakhsh, Matthew Campagna, Craig Costello, Luca De Feo, Basil Hess, Amir Jalali, Brian Koziel, Brian LaMacchia, Patrick Longa, Michael Naehrig, Joost Renes, Vladimir Soukharev, and David Urbanik. Supersingular isogeny key encapsulation. Supporting documentation for the NIST PQC project submission, 2017. https://www.cs.ru.nl/~jrenes/publications/sike.pdf.

[Jak18] Mads S. Jakobsen. On a (no longer) new segal algebra: A review of the feichtinger algebra. *Journal of Fourier Analysis and Applications*, 24:1579–1660, Dec 2018.

[Kac39] M. Kac. On a characterization of the normal distribution. *American Journal of Mathematics*, 61(3):726–728, 1939.

[KCCL02] Ki Hyoung Ko, Doo Ho Choi, Mi Sung Cho, and Jang Won Lee. New signature scheme using conjugacy problem. Cryptology ePrint Archive, Report 2002/168, 2002.

[Kea98] Michael Kearns. Efficient noise-tolerant learning from statistical queries. In *JOURNAL OF THE ACM*, pages 392–401. ACM Press, 1998.

[Kir11] Paul Kirchner. Improved generalized birthday attack. Cryptology ePrint Archive, Report 2011/377, 2011. https://eprint.iacr.org/2011/377.

[KLC+00] Ki Hyoung Ko, Sang Jin Lee, Jung Hee Cheon, Jae Woo Han, Ju-sung Kang, and Choonsik Park. New public-key cryptosystem using braid groups. In Mihir Bellare, editor, *Advances in Cryptology — CRYPTO 2000*, pages 166–183. Springer Berlin Heidelberg, 2000.

[Kle00] Philip Klein. Finding the closest lattice vector when it's unusually close. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '00, page 937–941. Society for Industrial and Applied Mathematics, 2000.

[Kra94] Daan Krammer. *The Conjugacy Problem for Coxeter Groups*. PhD thesis, University of Warwick, 1994. https://homepages.warwick.ac.uk/~masbal/index_files/me.pdf.

[Lai88] Philip D. Laird. *Learning from Good and Bad Data*. Kluwer Academic Publishers, 1988.

[LK54] Eugene Lukacs and Edgar P. King. A property of the normal distribution. *The Annals of Mathematical Statistics*, 25(2):389–394, 1954.

[LPR10] Vadim Lyubashevsky, Chris Peikert, and Oded Regev. On ideal lattices and learning with errors over rings. In *Proceedings of the 29th Annual International Conference on Theory and Applications of Cryptographic Techniques*, EUROCRYPT'10, pages 1–23. Springer-Verlag, 2010.

[LR19] Christopher Leonardi and Luis Ruiz. Homomorphism learning problems and its applications to public-key cryptography. CFail 2019, May 2019. https://eb6d8399-e0bd-4258-90d3-01dbd1dd7ad6.filesusr.com/ugd/6a3e24_e30f7aeab3de44ebbcab0b16b3369675.pdf.

[LW15] Vadim Lyubashevsky and Daniel Wichs. Simple lattice trapdoor sampling from a broad class of distributions. In Jonathan Katz, editor, *Public-Key Cryptography – PKC 2015*, pages 716–730. Springer Berlin Heidelberg, 2015.

[Lyu05] Vadim Lyubashevsky. The parity problem in the presence of noise, decoding random linear codes, and the subset sum problem. In Chandra Chekuri, Klaus Jansen, José D. P. Rolim, and Luca Trevisan, editors, *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 378–389. Springer Berlin Heidelberg, 2005.

[Mos99] Michele Mosca. *Quantum Computer Algorithms*. PhD thesis, Oxford University, 1999. http://cacr.uwaterloo.ca/ mmosca/moscathesis.ps.

[MP12] Daniele Micciancio and Chris Peikert. Trapdoors for lattices: Simpler, tighter, faster, smaller. In David Pointcheval and Thomas Johansson, editors, *Advances in Cryptology – EUROCRYPT 2012*, pages 700–718. Springer Berlin Heidelberg, 2012.

[MP13] Daniele Micciancio and Chris Peikert. Hardness of SIS and LWE with small parameters. In Ran Canetti and Juan A. Garay, editors, *Advances in Cryptology – CRYPTO 2013*, pages 21–39. Springer Berlin Heidelberg, 2013.

[MR07] Daniele Micciancio and Oded Regev. Worst-case to average-case reductions based on Gaussian measures. *SIAM J. Comput.*, 37(1):267–302, apr 2007.

[MSD19] Stephen Miller and Noah Stephens-Davidowitz. Kissing numbers and transference theorems from generalized tail bounds. *SIAM Journal on Discrete Mathematics*, 33(3):1313–1325, jul 2019.

[Mur12] J.D. Murray. *Asymptotic Analysis*. Applied Mathematical Sciences. Springer New York, 2012.

[MVR97] R. Meise, D. Vogt, and M.S. Ramanujan. *Introduction to Functional Analysis*. Clarendon Press, 1997.

[NU15] H.Q. Nguyen and M. Unser. Generalized poisson summation formula for tempered distributions. *Proceedings of the Eleventh International Workshop on Sampling Theory and Applications (SampTA'15)*, (Washington DC, USA):1–5, 2015.

[Pei06] Chris Peikert. On error correction in the exponent. In Tal Halevi, Shaicand Rabin, editor, *Theory of Cryptography*, pages 167–183. Springer Berlin Heidelberg, 2006.

[Pei08] Chris Peikert. Public-key cryptosystems from the worst-case shortest vector problem. *Electronic Colloquium on Computational Complexity (ECCC)*, 2008.

[Pei10] Chris Peikert. An efficient and parallel Gaussian sampler for lattices. In Tal Rabin, editor, *Advances in Cryptology – CRYPTO 2010*, pages 80–97. Springer Berlin Heidelberg, 2010.

[Pei14] Chris Peikert. Lattice cryptography for the internet. In Michele Mosca, editor, *Post-Quantum Cryptography*, pages 197–219. Springer International Publishing, 2014.

[Pla18] Rachel Player. *Parameter selection in lattice-based cryptography.* PhD thesis, Royal Holloway, University of London, 2018.

[PRS17] Chris Peikert, Oded Regev, and Noah Stephens-Davidowitz. Pseudorandomness of Ring-LWE for any ring and modulus. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 461–473, 2017.

[Reg05] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. In *Proceedings of the Thirty-seventh Annual ACM Symposium on Theory of Computing*, STOC '05, pages 84–93. ACM, 2005.

[RMUV10] Vitaly Roman'kov, Alexei Miasnikov, Alexander Ushakov, and Anatoly Vershik. The word and geodesic problems in free solvable groups. *Transactions of the American Mathematical Society*, 362:4655–4682, 01 2010.

[RRS00] H. Reiter, P.M.H. Reiter, and J.D. Stegeman. *Classical Harmonic Analysis and Locally Compact Groups.* London Mathematical Society monographs. Clarendon Press, 2000.

[Rud62] Walter Rudin. *Fourier analysis on groups.* Interscience tracts in pure and applied mathematics ; no. 12. 1962.

[SSTX09] Damien Stehlé, Ron Steinfeld, Keisuke Tanaka, and Keita Xagawa. Efficient public key encryption based on ideal lattices. In Mitsuru Matsui, editor, *Advances in Cryptology – ASIACRYPT 2009*, pages 617–635. Springer Berlin Heidelberg, 2009.

[Sti05] Eberhard Stickel. A new method for exchanging secret keys. In *Third International Conference on Information Technology and Applications (ICITA'05)*, volume 2, pages 426–430. IEEE, 2005.

[SW71] Elias M. Stein and Guido Weiss. *Introduction to Fourier Analysis on Euclidean Spaces (PMS-32)*. Princeton University Press, 1971.

[Tit48] Edward C Titchmarsh. *Introduction to the theory of Fourier integrals*. Clarendon Press, 1948.

[Val84] L. G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, nov 1984.

[Val85] L. G. Valiant. Deductive learning. In *Proc. of a Discussion Meeting of the Royal Society of London on Mathematical Logic and Programming Languages*, pages 107–112. Prentice-Hall, Inc., 1985.

[Vél71] Jacques Vélu. Isogénies entre courbes elliptiques. *Comptes rendus de l'Académie des Sciences, Paris, Serie A*, 273:238–241, 1971.

[ZZX20] Zhongxiang Zheng, Chunhuan Zhao, and Guangwu Xu. Discrete Gaussian measures and new bounds of the smoothing parameter for lattices. *Applicable Algebra in Engineering, Communication and Computing*, Feb 2020.

Appendix A

Asymptotic Behavior of σ_p

We dedicate this appendix to giving a formal proof of Proposition 5.49. Explicitly, we use the steepest descent method to find an asymptotic approximation of the Fourier transform of the p-supergaussian

$$\sigma_p(y) := \int_{\mathbb{R}} e^{-x^p} e^{-2\pi i x y} dx, \tag{A.1}$$

where p is a positive even integer. There are several reasons to use this method. On one hand, approximations given by Taylor/Laurent series, by definition, are only well-behaved locally, but they provide no information about the asymptotical behavior of the function. On the other hand, other approaches to find asymptotic approximations—such as Laplace's method or method of stationary phase—have different requirements on the behavior of the integrand in Equation (A.1), which are not satisfied in our case.

We remark that there have been previous attempts to use the steepest descent method to find an approximation of σ_p in the literature, for instance, [Boy14, DBPS18]. However, we should mention that, in [Boy14], the author backs up the claim that the found function is an asymptotic approximation, by observing that the difference between the obtained result and a numerical approximation of the real function does indeed decrease for $p = 4$. However, there is a fatal flaw in the provided argument which makes the result invalid for $p > 4$. [1] This flaw becomes evident once we provide a more rigorous proof of the correctness of the process.

[1] Precisely, the summation (A.7) must range over all the critical points whose level curves approximate the real line in a "closed" manner. In our case, this is the collection of critical points existing in the lower half of the complex plane.

The Steepest Descent Method, An Overview. Suppose a function $I\colon \mathbb{R} \to \mathbb{C}$ is given by

$$I(\kappa) = \int_{\mathcal{C}} f(t) \exp\big(\kappa\psi(t)\big) dt, \tag{A.2}$$

where f and ψ are analytic functions in some open set $U \subset \mathbb{C}$ that contains the contour $\mathcal{C}$. The exponential part of the integrand can be expressed as

$$\exp\big(\kappa\psi(t)\big) = \exp\Big(\kappa \operatorname{Re}\big(\psi(t)\big)\Big) \exp\Big(i\kappa \operatorname{Im}\big(\psi(t)\big)\Big).$$

The first factor provides the magnitude of the function, and the second provides the direction. Let Σ be the set of critical points of ψ. The idea behind this process—which is similar to all saddle-point approximations—is to deform $\mathcal{C}$ to a new contour $\mathcal{C}'$ containing a critical point $\omega \in \Sigma$, leaving fixed the endpoints of the contour and avoiding poles throughout the deformation. Thus, by Cauchy's Integral Formula, the value of the integral is the same when it is taken over $\mathcal{C}'$. However, when integrating over $\mathcal{C}'$, for large values of κ, most of the weight of the integral is centered around ω and (possibly also) around the endpoints. If, furthermore, the value of the function is negligible at the endpoints, the weight of the function on the tails can be disregarded. Finally, by obtaining the Taylor expansion of ψ around ω, we obtain a sufficiently good approximation of the exponential in this neighborhood.

The formalization of this process can be found in [Mur12, Chapter 3]. The following proposition gives an approximation of the integral over the steepest path of $\operatorname{Re}\big(\psi(t)\big)$.

Proposition A.1 ([Mur12, Equation 3.29]). *Let $f, \psi\colon \mathbb{C} \to \mathbb{C}$ be analytic in some open set $U \subset \mathbb{C}$. Let $I(\kappa) = \int_{\mathcal{C}} f(t) \exp\big(\kappa\psi(t)\big)dt$, where $\mathcal{C} \subset U$ is a steepest contour for the function $\operatorname{Re}\big(\psi(t)\big)$. Assume that $\mathcal{C}$ contains a single saddle point $\omega \in \Sigma$. Then*

$$I(\kappa) = f(\omega) \exp\big(\kappa\psi(\omega)\big) \sqrt{\frac{-2\pi}{\kappa\psi''(\omega)}} + O\left(\frac{\exp\big(\kappa\psi(\omega)\big)}{\kappa}\right). \tag{A.3}$$

Notice, however, that there are two possible values to choose for $\sqrt{\frac{-2\pi}{\kappa\psi''(\omega)}}$. The chosen value determines the direction of integration. To understand this relation, it is useful to analyze part of the derivation of Proposition A.1. In [Mur12], Equation (A.3) is obtained after parameterizing the path of the steepest descent $\mathcal{C}$ of $\operatorname{Re}(\psi)$ following the relation

$$\psi(t) - \psi(\omega) = -s^2.$$

Since ψ is a holomorphic function, by the Cauchy-Riemann equations, a path of steepest descent for $\mathrm{Re}(\psi)$ is a level curve of the function $\mathrm{Im}(\psi)$. As a consequence, $\psi(t) - \psi(\omega)$ is a real negative number; thus s is real. Expanding the Taylor series of ψ around ω gives

$$\frac{1}{2}(t-\omega)^2\psi''(\omega) + O\Big((t-\omega)^3\Big) = -s^2$$

which implies that

$$t - \omega = s\sqrt{\frac{-2}{\psi''(\omega)}} + O\left(s^2\right). \tag{A.4}$$

This approximation is only useful in a neighborhood of ω. Nonetheless, by our previous discussion, this is enough for our purposes. From Equation (A.4) we have that for points t over $\mathcal{C}$ close to ω, $t-\omega$ is close to a scalar multiple of $\sqrt{\frac{-2}{\psi''(\omega)}}$. Thus the direction of integration is given by the derivative of the curve at the point ω.

A Preliminary Transformation. This is a good point to remember that our goal is to find an approximation for the expression in Equation (A.1). We first start by modifying this equation to obtain an expression as in Equation (A.2). [2] Let $x = \left(\frac{2\pi y}{p}\right)^{1/(p-1)} t$. Then σ_p is expressed as follows.

$$\begin{aligned}
\sigma_p(y) &= \int_{\mathbb{R}} e^{-x^p} e^{-2\pi i x y} dx \\
&= \int_{\mathbb{R}} \exp\left(-x^p - 2\pi i x y\right) dx \\
&= \left(\frac{p}{2\pi y}\right)^{1/(p-1)} \int_{\mathbb{R}} \exp\left(-\left(\frac{2\pi y}{p}\right)^{p/(p-1)} t^p - 2\pi i y \left(\frac{2\pi y}{p}\right)^{1/(p-1)} t\right) dt \\
&= \left(\frac{p}{2\pi y}\right)^{1/(p-1)} \int_{\mathbb{R}} \exp\left((2\pi y)^{p/(p-1)} \left(\frac{1}{p}\right)^{1/(p-1)} \left(\frac{-1}{p} t^p - it\right)\right) dt \\
&= \left(\frac{p}{2\pi y}\right)^{1/(p-1)} \int_{\mathbb{R}} \exp\big(\kappa(y)\psi(t)\big) dt,
\end{aligned}$$

where

$$\kappa(y) \coloneqq \left(\frac{1}{p}\right)^{1/(p-1)} (2\pi y)^{p/(p-1)} \in \Omega(y) \qquad \text{and} \qquad \psi(t) \coloneqq \frac{-1}{p} t^p - it \tag{A.5}$$

[2] Notice that the naive attempt of taking $\psi(t) = -2\pi i y t$ yields no critical points for ψ. Thus it is not suitable for the application of any saddle-point method.

The newly defined ψ is a holomorphic function. Its derivatives are given by $\psi'(t) = -t^{p-1} - i$ and $\psi''(t) = \frac{-t^{p-2}}{p-1}$. Thus the set of critical points for ψ is described by $\Sigma = \{\omega \colon \omega^{p-1} = -i\}$.

Deforming the Integration Path. For each $\omega \in \Sigma$, consider the level set given by the equation

$$\mathcal{Z}_\omega \colon \operatorname{Im}\big(\psi(t)\big) = \operatorname{Im}\big(\psi(\omega)\big). \tag{A.6}$$

It is then clear that the path $\mathcal{C}_\omega$ of steepest descent of the function $\operatorname{Re}\big(\psi(t)\big)$ that passes through ω is contained in the set $\mathcal{Z}_\omega$. However, these two sets, $\mathcal{Z}_\omega$ and $\mathcal{C}_\omega$ are not the same in general. In our particular case, $\mathcal{Z}_\omega$ consists of several connected components, one of which is $\mathcal{C}_\omega$.

As discussed before, the idea is to deform the integration path, which in this case is $\mathbb{R}$, into a steepest path $\mathcal{C}_\omega$. In this case, however, such deformation cannot be performed in a straightforward manner. The reason is because none of the level curves $\mathcal{C}_\omega$ share the same "endpoints" with $\mathbb{R}$; in other words, none of these curves approximate $\mathbb{R}$ in both tails.

Here we give the intuitive idea for the solution of this problem. We start by noticing that the curves $\mathcal{C}_\omega$ divide the lower half of complex plane into several regions. Moreover, two of the curves in this half have tails approximating the positive and the negative tails of $\mathbb{R}$. The remaining curves in the lower half approximate their neighboring curves in the tails. Thus, intuitively, when integrating along each one of them, the tails of two neighboring curves "cancel out", except for the tails that approximate $\mathbb{R}$. This intuition is made formal with the following claims.

Claim A.2. Let C_{p-1} be the the group of $p-1$ roots of unity. Then $\Sigma = -i^{p+1}C_{p-1}$.

Proof. Let $z \in C_{p-1}$. Then we have

$$\big(i^{p+1} \cdot z\big)^{p-1} = i^{p^2-1} \cdot z^{p-1} = i^{-1} \cdot 1 = -i,$$

which completes the proof. □

As a consequence of Claim *A.2*, it is possible to write any element $\omega \in \Sigma$ as

$$\omega = \exp\left(-2\pi i \left(\frac{k}{p-1} + \frac{1}{4(p-1)}\right)\right),$$

with $k \in \{0, \ldots, p-1\}$. In particular, $\exp\left(\frac{i\pi}{2(p-1)}\right)$ is an element of Σ. Let Σ_- be the subset of elements in Σ whose imaginary part is negative. A consequence of Claims A.3 and A.4 is that the lower half of the complex plane is roughly divided by the curves $\mathcal{C}_\omega$ into $p/2$ regions, each of which contains one element of Σ_-.

Claim A.3. The set Σ_- has exactly $p/2$ elements.

Proof. To see this, observe that, by the pigeonhole principle, it follows that if $p \equiv 0 \mod 4$, there are exactly $p/2$ elements in C_{p-1} whose real part is negative. Similarly, if $p \equiv 2 \mod 4$, C_{p-1} has exactly $p/2$ elements whose real part is positive. □

Claim A.4. For $k \in \{0, \ldots, p-1\}$ let R_k be the ray $e^{-2\pi ik/p}\mathbb{R}_{>0}$ and if $0 \leq k < p/2$ let $\omega_k = \exp\left(-2\pi i\left(\frac{k}{p-1} + \frac{1}{4(p-1)}\right)\right) \in \Sigma_-$. Then the curve $\mathcal{C}_{\omega_k}$ is asymptotically close to the rays R_k and R_{k+1}.

Sketch of the proof. Consider the function $f(t) = \operatorname{Re}\big(\Psi(t)\big)$. For large values of $|t|$, the function $\psi(t) = \frac{-1}{p}t^p - it$ behaves like $\frac{-1}{p}t^p$, whose real part has a valley exactly on the rays R_k for $k \in \{0, \ldots, p-1\}$. Therefore f has a valley whose center approaches each one of these rays—to prove this formally, we would need to prove that as r approaches infinity, the minimum of f restricted to the elements of norm r is reached in elements every time closer to the rays.

Notice that the R_k with $k \in \{0, \ldots, p/2\}$ divide the lower half of the complex plane into $p/2$ regions, and each one of these regions intersects Σ_- in exactly one element. As a consequence, $\omega_k = \exp\left(-2\pi i\left(\frac{k}{p-1} + \frac{1}{4(p-1)}\right)\right) \in \Sigma_-$, is contained in the region between the rays R_k and R_{k+1}. Thus the curve $\mathcal{C}_{\omega_k}$ asymptotically follows the centers of the valleys of $f(t)$ closer to ω_k, which in turn asymptotically follows the rays R_k and R_{k+1}. □

Proposition A.5. *For any $\kappa > 0$,*

$$\int_{\mathbb{R}} \exp\big(\kappa\psi(t)\big)dt = \sum_{\omega\in\Sigma_-} \int_{\mathcal{C}_\omega} \exp\big(\kappa\psi(t)\big)dt, \tag{A.7}$$

where every curve $\mathcal{C}_\omega$ is parameterized counterclockwise.

Proof. The idea is to consider a bounded version of the curves $\mathcal{C}_\omega$ and use Cauchy's integral theorem to approximate $\int_{\mathbb{R}} \exp\big(\kappa\psi(t)\big)dt$. Formally, let $r \in \mathbb{R}_{>1}$ and let $\mathcal{C}_{\omega,r} = \mathcal{C}_\omega \cup B_r$. Consider the straight line segments $\mathcal{T}_{\omega,\omega',r}$ that are the closest to the ends of the curves $\mathcal{C}_{\omega,r}$ and $\mathcal{C}_{\omega' r}$ whenever these are in adjacent areas. Finally let $\mathcal{T}_{0,r}$ and $\mathcal{T}_{1,r}$ be the straight line segments joining the two remaining unmatched ends with $-r, r \in \mathbb{R} \cap B_r$.

It is clear that the union of the previously defined curves

$$\mathcal{C}_r = T_{0,r} \cup T_{1,r} \cup \left(\bigcup_{\omega} \mathcal{C}_{\omega,r}\right) \cup \left(\bigcup_{\omega,\omega'} \mathcal{T}_{\omega,\omega',r}\right)$$

forms a closed curve. Thus by Cauchy's Integral Theorem, $\oint_{\mathcal{C}_r \cup [-r,r]} f(z)dz = 0$, which implies that

$$\int_{[-r,r]} \exp\big(\kappa\psi(t)\big)dt = \int_{\mathcal{C}_r} \exp\big(\kappa\psi(t)\big)dt,$$

where the parameterization of $\mathcal{C}_r$ starts from (its point closest to) $-r$.

To finalize, observe that Claim A.4 implies that the length of the segments $\mathcal{C}_{\omega,r}$, $\mathcal{T}_{0,r}$, $\mathcal{T}_{1,r}$ decreases to zero as r goes to infinity. However, since the real part of the exponent is negative, the value of the function on these curves is bounded in absolute value. Therefore

$$\lim_{r\to\infty} \int_{\mathcal{C}_{\omega,r}} \exp\big(\kappa\psi(t)\big)dt = 0 \qquad \text{and} \qquad \lim_{r\to\infty} \int_{\mathcal{T}_{\beta,r}} \exp\big(\kappa\psi(t)\big)dt = 0, \tag{A.8}$$

for $\beta \in \{0, 1\}$. Since $\int_{\mathcal{C}_\omega} \exp\big(\kappa\psi(t)\big)dt = \lim_{r\to\infty} \int_{\mathcal{C}_{\omega,r}} \exp\big(\kappa\psi(t)\big)dt$, the result follows. □

Computing An Approximation of σ_p. We now combine propositions A.1 and A.5. This allows us to obtain the following approximation of σ_p:

$$\sigma_p(y) = \left(\frac{p}{2\pi y}\right)^{1/(p-1)} \sum_{\omega\in\Sigma_-} \exp\left(\kappa(y)\left(\frac{-1}{p}\omega^p - i\omega\right)\right)\sqrt{\frac{-2\pi}{\kappa(y)\frac{-\omega^{p-2}}{p-1}}} + \mathcal{E}(y,\omega), \tag{A.9}$$

where $\mathcal{E}$ is an error term which appears in each element of the summation, given by

$$\mathcal{E}(y,\omega) = O\left(\frac{\exp\big(\kappa(y)\psi(\omega)\big)}{\kappa(y)}\right).$$

Observe that for every $\omega \in \Sigma_-$, the real part of $\psi(\omega)$ is negative. Consequently, since the function $\kappa(y)$ is in $\Omega(y)$, the error function $\mathcal{E}$ rapidly approximates 0 as y tends to infinity.

As a result, σ_p is approximated by the following expression.

$$\begin{aligned}
\sigma_p(y) &\approx \left(\frac{p}{2\pi y}\right)^{1/(p-1)} \sum_{\omega\in\Sigma_-} \exp\left(\kappa(y)\left(\frac{-1}{p}(-i\omega) - i\omega\right)\right)\sqrt{\frac{-2\pi}{\kappa(y)\frac{i}{(p-1)\omega}}} \\
&= \left(\frac{p}{2\pi y}\right)^{1/(p-1)} \sum_{\omega\in\Sigma_-} \exp\left(\kappa(y)\left(\frac{-1}{p}(-i\omega) - i\omega\right)\right)\sqrt{\frac{2\pi(p-1)i\omega}{\kappa(y)}} \\
&= \left(\frac{p}{2\pi y}\right)^{1/(p-1)} \sum_{\omega\in\Sigma_-} \exp\left(-i\omega\kappa(y)\left(\frac{-1}{p} + 1\right)\right)\sqrt{\frac{2\pi(p-1)i\omega}{\kappa(y)}} \qquad \text{(A.10)} \\
&= \left(\frac{p}{2\pi y}\right)^{1/(p-1)} \sum_{\omega\in\Sigma_-} \exp\left(-i\omega\kappa(y)\left(\frac{p-1}{p}\right)\right)\sqrt{\frac{2\pi(p-1)i\omega}{\kappa(y)}} \\
&= \left(\frac{p}{2\pi y}\right)^{1/(p-1)} \sqrt{\frac{2\pi(p-1)}{\kappa(y)}} \sum_{\omega\in\Sigma_-} \sqrt{i\omega}\exp\left(-i\omega\kappa(y)\left(\frac{p-1}{p}\right)\right).
\end{aligned}$$

The expression above is well defined up to the choice of $\sqrt{i\omega}$. As discussed previously, the particular value represents the direction of integration at each $\mathcal{C}_\omega$. Moreover, the chosen value of $\sqrt{i\omega}$ approximates the derivative of the curve at ω. Since we are integrating counterclockwise, the corresponding value of $\sqrt{i\omega}$ is that which has a positive real part. For $\omega \in \Sigma_-$, write

$$\begin{aligned}
i\omega &= \exp\left(i\pi\left(\frac{1}{2} - \frac{1}{2(p-1)} - \frac{2k}{(p-1)}\right)\right) \\
&= \exp\left(i\pi\left(\frac{p-1-1-4k}{2(p-1)}\right)\right) \\
&= \exp\left(i\pi\left(\frac{p-2-4k}{2(p-1)}\right)\right),
\end{aligned}$$

where $k \in \big\{0,\ldots,(p-2)/2\big\}$. Then we have that $\sqrt{i\omega} = \exp\left(i\pi\left(\frac{p-2-4k}{4(p-1)} + \beta\right)\right)$ with $\beta \in \{0,1\}$. Thus the real part is given by $\operatorname{Re}\sqrt{i\omega} = \cos\left(\pi\left(\frac{p-2-4k}{4(p-1)} + \beta\right)\right)$. Since this value must be positive, we have that for all ω, the corresponding $\beta = 0$. Hence we simplify the expression in (A.10) as follows. The factor outside the summation can be expressed as

$$\begin{aligned}
\left(\frac{p}{2\pi y}\right)^{1/(p-1)} \sqrt{\frac{2\pi(p-1)}{\kappa(y)}} &= \left(\frac{p}{2\pi y}\right)^{1/(p-1)} \sqrt{\frac{2\pi(p-1)}{\left(\frac{1}{p}\right)^{1/(p-1)} (2\pi y)^{p/(p-1)}}} \\
&= \left(\frac{p}{2\pi y}\right)^{1/(p-1)} \sqrt{\frac{(p-1)p^{1/(p-1)}}{y^{p/(p-1)}(2\pi)^{1/(p-1)}}} \\
&= \left(\frac{p}{2\pi y}\right)^{1/(p-1)} \sqrt{\frac{(p-1)}{y^{p/(p-1)}} \left(\frac{p}{2\pi}\right)^{1/(p-1)}} \\
&= \sqrt{\frac{(p-1)}{y^{(p+2)/(p-1)}} \left(\frac{p}{2\pi}\right)^{3/(p-1)}},
\end{aligned}$$

and the summation is given by

$$\begin{aligned}
&\sum_{\omega\in\Sigma_-} \sqrt{i\omega} \exp\left(-i\omega\kappa(y)\left(\frac{p-1}{p}\right)\right) \\
&= \sum_{\omega\in\Sigma_-} \sqrt{i\omega} \exp\left(-i\omega\left(\frac{1}{p}\right)^{1/(p-1)} (2\pi y)^{p/(p-1)} \left(\frac{p-1}{p}\right)\right) \\
&= \sum_{k=0}^{(p-2)/2} \exp\left(i\pi\left(\frac{p-2-4k}{4(p-1)}\right)\right) \exp\left(-e^{i\pi\frac{p-2-4k}{2(p-1)}} \left(\frac{1}{p}\right)^{1/(p-1)} (2\pi y)^{p/(p-1)} \left(\frac{p-1}{p}\right)\right) \\
&= \sum_{k=0}^{(p-2)/2} \exp\left(i\pi\left(\frac{p-2-4k}{4(p-1)}\right) - e^{i\pi\frac{p-2-4k}{2(p-1)}} \left(\frac{2\pi y}{p}\right)^{p/(p-1)} (p-1)\right).
\end{aligned}$$

Thus, σ_p is asymptotically approximated by the expression

$$\begin{aligned}
\sigma_p(y) \approx &\sqrt{\frac{(p-1)}{y^{(p+2)/(p-1)}} \left(\frac{p}{2\pi}\right)^{3/(p-1)}} \\
&\sum_{k=0}^{(p-2)/2} \exp\left(i\pi\left(\frac{p-2-4k}{4(p-1)}\right) - (p-1)\left(\frac{2\pi y}{p}\right)^{p/(p-1)} e^{i\pi\frac{p-2-4k}{2(p-1)}}\right).
\end{aligned} \tag{A.11}$$

A Positive Approximation for σ_p. We now use the result obtained above to construct a positive approximation for σ_p. The main idea is to smooth out the oscillations from that function to obtain a smooth bound. Note that σ_p is a real valued function—as the Fourier transform of any real function is a real function—so we only need to consider the real part of the approximation. Consider one term inside the sum of Equation (A.11). For $k \in \left\{1, \ldots, \frac{p-2}{2}\right\}$ we have

$$\begin{aligned}
&\exp\left(i\pi\left(\frac{p-2-4k}{4(p-1)}\right) - (p-1)\left(\frac{2\pi y}{p}\right)^{\frac{p}{p-1}} e^{\left(i\pi\left(\frac{p-2-4k}{2(p-1)}\right)\right)}\right) \\
&= \exp\left(i\pi\left(\frac{p-2-4k}{4(p-1)}\right)\right. \\
&\qquad\left. -(p-1)\left(\frac{2\pi y}{p}\right)^{\frac{p}{p-1}}\left(\cos\pi\left(\frac{p-2-4k}{2(p-1)}\right) + i\sin\pi\left(\frac{p-2-4k}{2(p-1)}\right)\right)\right) \\
&= \exp\left(-(p-1)\left(\frac{2\pi y}{p}\right)^{\frac{p}{p-1}}\cos\pi\left(\frac{p-2-4k}{2(p-1)}\right)\right) \\
&\qquad \exp\left(i\pi\left(\frac{p-2-4k}{4(p-1)}\right) - (p-1)\left(\frac{2\pi y}{p}\right)^{\frac{p}{p-1}}\sin\pi\left(\frac{p-2-4k}{2(p-1)}\right)\right)
\end{aligned}$$

Since for all $\theta \in \mathbb{R}$, $\left|\mathrm{Re}(e^{i\theta})\right| \leq 1$, the real part of each term is upper bounded by

$$\exp\left(-(p-1)\left(\frac{2\pi y}{p}\right)^{\frac{p}{p-1}}\cos\left(\pi\frac{p-2-4k}{2(p-1)}\right)\right),$$

as the only imaginary term comes from the second exponent in the last expression. Note that $0 \leq k \leq \frac{p-2}{2}$; thus the argument of the cosine in the above expression is in the range $(-\pi/2, \pi/2)$. (See the discussion above about the simplification of Equation (A.10)). Thus we have that for every term,

$$\cos\left(\pi\frac{p-2-4k}{2(p-1)}\right) > 0.$$

Thus,

$$\begin{aligned}
&\exp\left(-(p-1)\left(\frac{2\pi y}{p}\right)^{\frac{p}{p-1}}\cos\left(\pi\frac{p-2-4k}{2(p-1)}\right)\right) \\
&\qquad \leq \exp\left(-(p-1)\left(\frac{2\pi y}{p}\right)^{\frac{p}{p-1}}\cos\left(\pi\frac{p-2}{2(p-1)}\right)\right) \\
&\qquad = \exp\left(-\vartheta(p)y^{\frac{p}{p-1}}\right)
\end{aligned}$$

where $\vartheta(p) \coloneqq (p-1)\left(\frac{2\pi}{p}\right)^{\frac{p}{p-1}} \cos\left(\pi \frac{p-2}{2(p-1)}\right)$. Thus, combining all the terms of the sum, we get that

$$\begin{aligned}
\sigma_p(y) &\approx \sqrt{(p-1)\left(\frac{p}{2\pi}\right)^{3/(p-1)}} y^{-(p+2)/2(p-1)} \\
&\quad \sum_{k=0}^{(p-2)/2} \exp\left(i\pi\left(\frac{p-2-4k}{4(p-1)}\right) - (p-1)\left(\frac{2\pi y}{p}\right)^{p/(p-1)} e^{\left(i\pi\left(\frac{p-2-4k}{2(p-1)}\right)\right)}\right) \\
&\leq \sqrt{(p-1)\left(\frac{p}{2\pi}\right)^{3/(p-1)}} y^{-(p+2)/2(p-1)} \sum_{k=0}^{(p-2)/2} \exp\left(-\vartheta(p) y^{\frac{p}{p-1}}\right) \\
&= \frac{p-2}{2}\sqrt{(p-1)\left(\frac{p}{2\pi}\right)^{3/(p-1)}} y^{-(p+2)/2(p-1)} \exp\left(-\vartheta(p) y^{\frac{p}{p-1}}\right) \\
&= \xi(p) y^{-(p+2)/2(p-1)} \exp\left(-\vartheta(p) y^{\frac{p}{p-1}}\right)
\end{aligned} \tag{A.12}$$

where $\xi(p) \coloneqq \frac{p-2}{2}\sqrt{(p-1)\left(\frac{p}{2\pi}\right)^{3/(p-1)}}$.

Appendix B

Tail Bound for σ_p

In this appendix show how to compute the tail bounds of the bounding function for the Fourier transform of supergaussians. For the rest of the section, we only consider this function restricted to $\mathbb{R}_{\geq 0}$, as the same techniques apply to $\mathbb{R}_{\leq 0}$ by symmetry since $f^{[p]}$ and $\left|\widehat{f^{[p]}}\right|$ are both even functions. Recall that the bounding function is given by

$$b(y) := \xi(p) \cdot \begin{cases} b_1(y) := e^{-\vartheta y^{\tau}} & 0 \leq y \leq 1 \\ b_2(y) := y^{-\alpha} e^{-\vartheta y^{\tau}} & y \geq 1 \end{cases}$$

with $\alpha = \frac{(p+2)}{2(p-1)}$, $\vartheta = (p-1)\cos\left(\pi\frac{p-2}{2(p-1)}\right)\left(\frac{2\pi}{p}\right)^{\frac{p}{p-1}}$ and $\tau = \frac{p}{p-1}$. Note that for the computation of a tail bound, we can assume $\xi(p) = 1$, as both $b(y)$ and $b(uy)$ are scaled by the same factor $\xi(p)$.

Consider a β that satisfies the conditions of Lemma 5.34. As pointed in Note 5.51, we assume that β is a constant. Let $r \in \mathbb{R}_{\geq 1}$.[1] There are two cases to consider, if $b(uy) = b_1(uy)$ and if $b(uy) = b_2(uy)$. The first case happens if $y \leq \frac{1}{u}$. This case is dependent on the chosen completion of the bounding function.

The second case happens if $y \geq \frac{1}{u}$. Note that this case is independent of the chosen completion of the bounding function. In that case,

$$\frac{b(y)}{ub(uy)} = \frac{y^{-\alpha}e^{-\vartheta y^{\tau}}}{u(uy)^{-\alpha}e^{-\vartheta(uy)^{\tau}}} = u^{\alpha-1}e^{-\vartheta y^{\tau}(1-u^{\tau})}.$$

[1] This will only allow us find tail bounds for sets of a given minimum size (namely, at lest $[-1,1]$), which will be sufficient for all of our applications. Should tail bounds for smaller sets be needed, they can be computed using the same method.

Thus,

$$\sup_{|y|\geq r} \frac{b(y)}{ub(uy)} = \sup_{|y|\geq r} u^{\alpha-1} e^{-\vartheta y^{\tau}(1-u^{\tau})} = u^{\alpha-1} e^{-\vartheta r^{\tau}(1-u^{\tau})},$$

as the function is monotonically decreasing for $y \geq 0$.

Since $y \geq r$, we get case 2 for any $u \geq \frac{1}{r}$ (as $uy \geq ur \geq \frac{1}{r}r = 1$). Moreover, given that $r \geq 1$, the interval $I = \left[\frac{1}{r}, 1\right] \subset (0,1]$ is non-empty, and we get case 2 for all y for any $u \in I$. Note that as long as $I \cap (0,1] = I$ is non-empty, we have that

$$\inf_{0<u\leq 1} \sup_{|y|\geq r} \frac{b(y)}{ub(uy)} \leq \inf_{u\in I} \sup_{|y|\geq r} \frac{b(y)}{ub_2(uy)} = \inf_{u\in I} u^{\alpha-1} e^{-\vartheta r^{\tau}(1-u^{\tau})}.$$

By taking the second derivative of the left-hand side for u over $I \cap (0,1] = I = \left[\frac{1}{r}, 1\right]$, we have that the infimum for the last equation occurs at $u = 1/r$. As a consequence,

$$\inf_{0<u\leq 1} u^{\alpha-1} e^{-\vartheta r^{\tau}(1-u^{\tau})} = \left(\frac{1}{r}\right)^{\alpha-1} e^{-\vartheta r^{\tau}(1-\frac{1}{r^{\tau}})} = \left(\frac{1}{r}\right)^{\alpha-1} e^{-\vartheta(r^{\tau}-1)}.$$

Thus,

$$\begin{aligned} \nu_{b^{[p]}}\big([r,r]\big) &\leq \beta \inf_{0<u\leq 1} \sup_{|y|\geq r} \frac{b(y)}{ub(uy)} \\ &\leq \beta \left(\frac{1}{r}\right)^{\alpha-1} e^{-\vartheta(r^{\tau}-1)}. \end{aligned}$$

Index