On the Gaussian Measure Over Lattices

by

Noah Stephens-Davidowitz

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Computer Science

New York University

September 2017

Oded Regev

Acknowledgements

I am grateful to my advisors, Oded Regev and Yevgeniy Dodis, for showing me how to be a computer scientist. Yevgeniy introduced me to my first real research questions. He also introduced me to many of my friends, colleagues, and co-authors in the computer science community. In particular, Yevgeniy brought me to the 2015 Simons Institute cryptography summer program, where I got to know some of the wonderful cast of characters in the cryptographic community.

It would be difficult to overstate Oded's influence on me. He introduced me to the topics of this thesis, taught me how to think about them, and showed me their beauty. More generally, Oded (patiently) taught me how to do research, and how to present it to others. I am simply awestruck by Oded, and the best aspects of my work are inherited directly from him. I hope to continue to steal as much of his insight as I can.

I thank my co-authors for the innumerable hours of fun and confusion and the occasional epiphanies that we had together—Divesh Aggarwal, Navid Alamati, Huck Bennett, Daniel Dadush, Yevgeniy Dodis, Sasha Golovnev, Shai Halevi, Tzipi Halevi, Ilya Mironov, Chris Peikert, Oded Regev, Adi Shamir, Victor Shoup, and Daniel Wichs. I'm particularly thankful for the co-authors on the works included in this thesis, Divesh Aggarwal, Daniel Dadush, and Oded Regev.

I thank my friends and co-conspirators at NYU for the laughs, games, and adventures that we shared; for the advice that they gave me; and for putting up with me—Azam Asl, Huck Bennett, Sandro Coretti, Laura Florescu, Chaya Ganesh, Sasha Golovnev, Siyao Guo, Shravas Rao, Igor Shinkar, Deva Thiruvenkatachari, and Omri Weinstein.

I am indebted to Daniel Dadush and Divesh Aggarwal, two postdocs who were at NYU

when I arrived. Daniel's three-person class (co-taught with Oded) is what originally got me hooked on lattices, and Divesh sat next to me for two years and responded to my countless stupid questions with thorough and seemingly judgment-free answers.

I thank Ilya Mironov for having me as his intern at Microsoft Research in the summer of 2014, and Chris Peikert for inviting me to visit him in Michigan in the summer of 2016. Each of them gave me far more of their time than I deserved and introduced me to new ways of thinking.

I thank the professors at Brown who originally introduced me to the beauty and elegance of abstract mathematics and computer science, Anna Lysyanskaya, Stephen Lichtenbaum, Alf van der Poorten, and Michael Rosen.

I thank my friends and colleagues at IBM, Fabrice Benhamouda, Craig Gentry, Shai Halevi, Tzipi Halevi, Justin Holmgren, Charanjit Jutla, Hugo Krawczyk, Antigoni Polychroniadou, and Tal Rabin.

I thank my family, Mitch, Esther, Seth, and Lauren, for always encouraging me—and also for gently discouraging me during my strange detour into the poker world. I offer mad props to my nephew Jonah, who is cooler at two-going-on-three than I'll ever be, and I'm looking forward to seeing how quickly my seven-week-old niece Sasha upstages me. And, of course, my brother-in-law Mark deserves roughly half the credit for this precocious coolness.

Lastly, I thank my committee, Daniel Dadush, Yevgeniy Dodis, Chris Peikert, Oded Regev, and Victor Shoup.

Abstract

We study the *Gaussian mass* of a lattice coset

$$\rho_s(\mathcal{L} - \boldsymbol{t}) := \sum_{\boldsymbol{y} \in \mathcal{L}} \exp(-\pi \|\boldsymbol{y} - \boldsymbol{t}\|^2 / s^2) \ ,$$

where $\mathcal{L} \subset \mathbb{R}^n$ is a lattice and $\boldsymbol{t} \in \mathbb{R}^n$ is a vector describing a shift of the lattice. In particular, we use bounds on this Gaussian mass to obtain a partial converse to Minkowski's celebrated theorem bounding the number of lattice points in a ball.

We also consider the *discrete Gaussian distribution* $D_{\mathcal{L}-\boldsymbol{t},s}$ induced by the Gaussian measure over $\mathcal{L} - \boldsymbol{t}$, and we use procedures for sampling from this distribution to construct the current fastest known algorithms for the two most important computational problems over lattices, the Shortest Vector Problem (SVP) and the Closest Vector Problem (CVP).

Finally, we study $\rho_s(\mathcal{L} - \boldsymbol{t})$ and $D_{\mathcal{L}-\boldsymbol{t},s}$ as interesting computational and mathematical objects in their own right. In particular, we show that the computational problem of sampling from $D_{\mathcal{L}-\boldsymbol{t},s}$ is equivalent to CVP in a very strong sense (and that sampling from $D_{\mathcal{L},s}$ is no harder than SVP). We also prove a number of bounds on the moments of $D_{\mathcal{L}-\boldsymbol{t},s}$ and various monotonicity properties of $\rho_s(\mathcal{L} - \boldsymbol{t})$.

Contents

List of Figures

List of Tables

What's this?

Before I begin to motivate and summarize the contents of this thesis, I should make clear that it contains no substantial original work. This is instead a new presentation (necessary for the completion of my doctorate) of results from joint work with my co-authors from five papers. These papers are all available on the arXiv and roughly as accessible as this thesis [ADRS15, ADS15, Ste16a, RS17a, RS17b]. Indeed, much of this work is taken verbatim from the original papers, so it would be perfectly reasonable for the reader to choose to read those instead.

The most substantial differences between this thesis and the papers from which it inherits its content come in Chapter 4. First, this chapter is a merger of two papers, one on SVP [ADRS15] and one on CVP [ADS15]. The two original papers overlap quite a bit, so this might prove to be beneficial.[1] Second, I have modified the $2^{n/2+o(n)}$-time algorithm in Section 4.5. Any readers looking to improve this algorithm to work below the smoothing parameter might prefer the version in this thesis, which is a bit simpler and which gets around one of the obstructions to doing so. (My co-authors and I have long felt that we were very close to improving the algorithm in Section 4.5. I resisted the urge to include in this thesis a long treatise on why we feel this way and what seemingly minor obstacles remain.)

Additional differences include (1) the omission of some of the more technical and/or less salient proofs and results; (2) condensed and merged preliminaries in Chapter 1—including some extra love and attention devoted to the Gaussian measure over lattices—(3) a global introduction (below); (4) notation that is more consistent across papers; (5) extremely minor

[1] On the other hand, the original papers have their own benefits. The first paper [ADRS15] is a bit easier to read because it deals only with a nice special case—the *centered* discrete Gaussian, as opposed to an arbitrary discrete Gaussian—and the second paper [ADS15] is written more concisely, under the assumption that readers interested in the smallest details will have read [ADRS15] first.

improvements to some secondary results; (6) some reworded paragraphs; etc. In fact, this exciting opportunity to unilaterally make whatever changes I saw fit to joint work with my co-authors was remarkably fruitless—a testament to the relative skill of my co-authors, and a bit of a bummer. I'm very much at risk of having worked hard only to have made things worse, and any errors introduced are of course my own fault and not the fault of my co-authors.

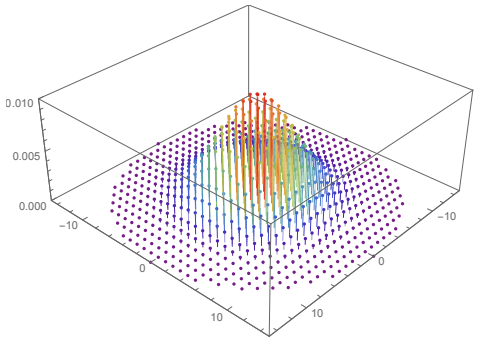
Introduction

A lattice $\mathcal{L} \subset \mathbb{R}^n$ is the set of integer linear combinations of linearly independent basis vectors $\mathbf{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n) \in \mathbb{R}^{n \times n}$,

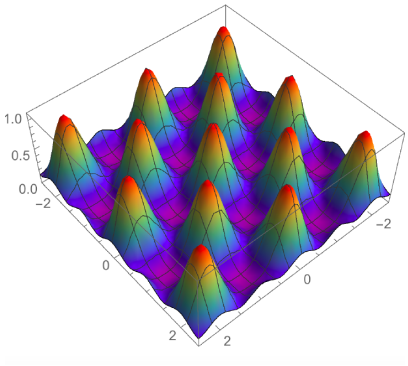
$$\mathcal{L}(\mathbf{B}) := \left\{ \sum a_i \boldsymbol{b}_i \ : \ a_i \in \mathbb{Z} \right\} .$$

Lattices are studied classically as natural geometric objects with connections to number theory, convex geometry, and many other fields (see, e.g., [GL87, CS98]). More recently, computer scientists have studied computational problems on lattices because of their applications in integer programming [Len83, Kan87], factoring polynomials over the rationals [LLL82], cryptanalysis [Sha84, Bri85, LO85], etc.

And, more recently still, lattice-based cryptographic constructions have revolutionized cryptography [Ajt04, Reg09, Gen09, BV11]. (See [Pei16] for a survey.) In particular, nearly all lattice-based cryptographic constructions come with the advantage that their security can be based on the hardness of approximating computational lattice problems *in the worst case* [Ajt04, Reg09, Pei09, BLP+13, PRS17]. And, many powerful cryptographic primitives are only known via lattice-based constructions, such as fully homomorphic encryption [Gen09, BV11, BV14]. Furthermore, by far the most well-studied public-key cryptographic constructions that are thought to be secure against quantum computers are based on lattices, and these schemes are now nearing widespread deployment in anticipation of future developments in quantum computing [NIS16, ADPS16, BCD+16].



(a) The discrete Gaussian distribution on $\mathbb{Z}^2$ with parameter $s = 10$.



(b) The periodic Gaussian function on the lattice spanned by $(1, 1)$ and $(1, -1)$ with parameter $s = 3/4$.

Figure 1

The *Gaussian measure*,

$$\rho_s(\boldsymbol{x}) := \exp(-\pi\|\boldsymbol{x}\|^2/s^2)$$

for a vector $\boldsymbol{x} \in \mathbb{R}^n$ and parameter $s > 0$, has become an essential tool in the study of lattices. In particular, we are interested in the four closely related mathematical objects introduced below.

The *Gaussian mass* of $\mathcal{L} - \boldsymbol{t}$,

$$\rho_s(\mathcal{L} - \boldsymbol{t}) := \sum_{\boldsymbol{y} \in \mathcal{L}} \rho_s(\boldsymbol{y} - \boldsymbol{t}) \ ,$$

for a lattice $\mathcal{L} \subset \mathbb{R}^n$ and shift $\boldsymbol{t} \in \mathbb{R}^n$ can be viewed as a smooth analogue of the lattice point-counting function $|\mathcal{L} \cap (rB_2^n + \boldsymbol{t})|$, which counts the number of lattice points in a ball of radius r around a vector $\boldsymbol{t}$. In particular, we have the trivial inequality $|\mathcal{L} \cap (rB_2^n + \boldsymbol{t})| \leq \exp(\pi r^2/s^2)\rho_s(\mathcal{L} - \boldsymbol{t})$, which is often quite tight for a suitably chosen parameter $s > 0$.[2] More

[2]For example, Mazo and Odlyzko showed that this inequality gives a very accurate estimate of $|\mathbb{Z}^n \cap (rB_2^n + \boldsymbol{t})|$ when $r := \Theta(\sqrt{n})$ and s is chosen appropriately [MO90]. In Section 2.8, we show weaker bounds in a more general setting.

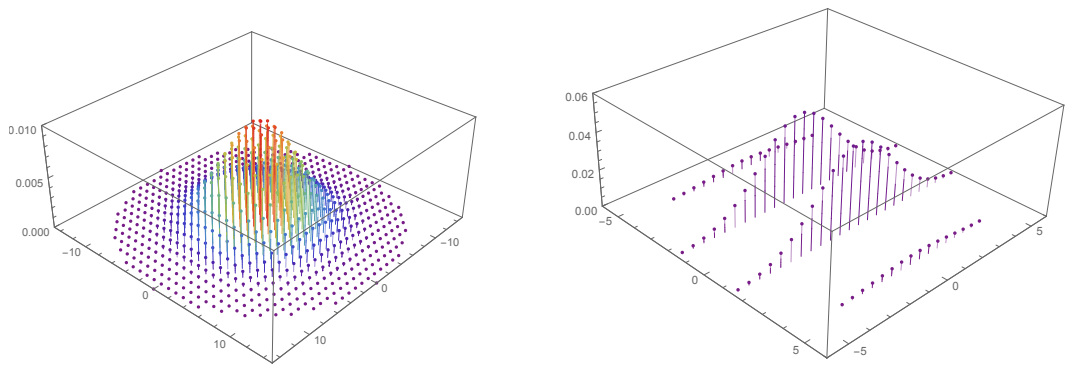


Figure 2: Two rather different discrete Gaussian distributions in two dimensions. On the left is $D_{\mathbb{Z}^2,10}$. On the right is $D_{\mathcal{L}-\boldsymbol{t},5}$, where $\mathcal{L}$ is spanned by $3\boldsymbol{e}_1$ and $\boldsymbol{e}_2/2$, and $\boldsymbol{t} = 3\boldsymbol{e}_1/2 + \boldsymbol{e}_2/4$ is a "deep hole."

generally, studying $\rho_s(\mathcal{L} - \boldsymbol{t})$ allows us to use analytic tools to understand the geometry of $\mathcal{L}$ (see, e.g., [Ban93], or our treatment in Section 1.3.1). The function $s \mapsto \rho_s(\mathcal{L})$ also arises naturally in many applications in number theory, where it is typically parametrized differently and referred to as the lattice *theta function* (e.g., [Jac28, Rie57, BPY01, Mum07]).

The *discrete Gaussian distribution* $D_{\mathcal{L}-\boldsymbol{t},s}$ is the distribution over $\mathcal{L} - \boldsymbol{t}$ induced by the Gaussian measure,

$$\Pr_{\boldsymbol{X} \sim D_{\mathcal{L}-\boldsymbol{t},s}}[\boldsymbol{X} = \boldsymbol{y} - \boldsymbol{t}] := \frac{\rho_s(\boldsymbol{y} - \boldsymbol{t})}{\rho_s(\mathcal{L} - \boldsymbol{t})} ,$$

for $\boldsymbol{y} \in \mathcal{L}$. In particular, this distribution allows us to sample relatively short vectors $\boldsymbol{x} \in \mathcal{L} - \boldsymbol{t}$, which makes it extremely useful in cryptographic constructions [Reg09, GPV08], algorithms for lattice problems [ADRS15, ADS15], and in worst-case to average-case reductions [MR07, Reg09, Pei09, BLP+13, PRS17]. Indeed, there are by now many different algorithms for sampling from $D_{\mathcal{L}-\boldsymbol{t},s}$, as well as reductions to other lattice problems. We summarize what is known in Table 1.

The *periodic Gaussian function*

$$f_{\mathcal{L},s}(\boldsymbol{t}) := \frac{\rho_s(\mathcal{L} - \boldsymbol{t})}{\rho_s(\mathcal{L})}$$

	Shift	Parameter	Time	Notes
	Any $\boldsymbol{t}$	$s \geq \gamma\sqrt{\log n} \cdot \lambda_n$	–	Reduces to γ-SVP or γ-SIVP [GPV08, BLP+13].
	Any $\boldsymbol{t}$	$s \geq 2^{\frac{n \log n}{\log \log n}} \cdot \lambda_n$	$\mathrm{poly}(n)$	[AKS01, GPV08]
	Any $\boldsymbol{t}$	$s \gg \sqrt{n} \cdot \eta_{n^{-\omega(1)}}$	–	Quantum reduces to BDD or LWE [Reg09].
*	Any $\boldsymbol{t}$	$s \geq \sqrt{2} \cdot \eta_{1/2}$	$2^{n/2+o(n)}$	Outputs $2^{n/2}$ samples [ADRS15].
*	Any $\boldsymbol{t}$	$s > 2^{-\frac{n}{\log n}} \operatorname{dist}(\boldsymbol{t}, \mathcal{L})$	$2^{n+o(n)}$	Outputs many samples [ADS15].
*	Any $\boldsymbol{t}$	Any s	–	Equivalent to CVP [Ste16a].
*	Any $\boldsymbol{t}$	Any s	$2^{n+o(n)}$	Follows from equivalence and [ADS15].
*	$\boldsymbol{t} = \boldsymbol{0}$	Any s	$2^{n+o(n)}$	Outputs $2^{n/2}$ samples [ADRS15].
*	$\boldsymbol{t} = \boldsymbol{0}$	Any s	–	Reduces to SVP [Ste16a].

Table 1: Known results concerning the problem of sampling from $D_{\mathcal{L}-\boldsymbol{t},s}$. Lines marked with a * are presented in this thesis. We have left out some constants and $\omega(1)$ factors for readability.

is the probability density function of a continuous Gaussian distribution with parameter $s > 0$ modulo the lattice $\mathcal{L}$ (up to scaling). Equivalently, it is the heat kernel on the flat torus $\mathbb{R}^n/\mathcal{L}$. It can also be thought of as a smooth approximation of the function $\boldsymbol{t} \mapsto \exp(-\pi \operatorname{dist}(\boldsymbol{t}, \mathcal{L})^2/s^2)$, and it has therefore found applications in algorithms for approximating $\operatorname{dist}(\boldsymbol{t}, \mathcal{L})$ and for finding a close lattice vector to $\boldsymbol{t}$ [AR05, LLM06, DRS14]. It is intimately related to the discrete Gaussian distribution, as its Fourier series is given by the discrete Gaussian over the dual lattice (see Eq. (1.3)).

The *smoothing parameter*,

$$\eta_\varepsilon(\mathcal{L}) := \min\{s \ : \ \rho_{1/s}(\mathcal{L}^*) \leq 1 + \varepsilon\} \ ,$$

where

$$\mathcal{L}^* := \{\boldsymbol{w} \in \mathbb{R}^n \ : \ \forall \boldsymbol{y} \in \mathcal{L}, \ \langle \boldsymbol{w}, \boldsymbol{y} \rangle \in \mathbb{Z}\}$$

is the dual lattice. We are typically interested in the case when $\varepsilon \in (0, 1)$, in which case $\eta_\varepsilon(\mathcal{L})$ intuitively represents "the scale at which the discrete structure of $\mathcal{L}$ is no longer visible", such that for parameters $s \geq \eta_\varepsilon(\mathcal{L})$ the distribution $D_{\mathcal{L}-\boldsymbol{t},s}$ "behaves like a continuous Gaussian

distribution up to error ε." The smoothing parameter, which was introduced by Micciancio and Regev [MR07], takes its name from the fact that $f_{\mathcal{L},s}(\boldsymbol{t})$ is nearly constant (or "smooth") for $s \geq \eta_\varepsilon(\mathcal{L})$. (See Eq. (1.4).) Both of these facts are extremely useful in applications, and nearly all algorithmic applications of $D_{\mathcal{L}-\boldsymbol{t},s}$ only work above the smoothing parameter (the key exceptions being [ADRS15, ADS15]).

Our Contributions

In this thesis, we describe new geometric and computational applications of these objects, and we study them as interesting mathematical objects in their own right. Below, we summarize the contents of this thesis, which describe results that originally appeared in joint works with Regev [RS17a, RS17b]; Aggarwal, Dadush, and Regev [ADRS15]; Aggarwal and Dadush [ADS15]; and in a single-author paper [Ste16a].

A Reverse Minkowski Theorem

In Chapter 2, we show a partial converse to Minkowski's Theorem (sometimes called "Minkowski's First Theorem"), which is the foundational result in the geometry of numbers [Min10]. Minkowski's Theorem guarantees that a centered ball rB_2^n of a certain radius $r > 0$ must have at least one non-zero lattice point. Below, we present a slight generalization due to Blichfeldt and van der Corput,[3]

Theorem 1 ([vdC36])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ with $\det(\mathcal{L}) \leq 1$ and radius $r > 0$,*

$$|\mathcal{L} \cap rB_2^n| \geq 2^{-n} \cdot \mathrm{vol}(rB_2^n) = \frac{1}{\sqrt{\pi n}} \left(\frac{\pi e r^2}{2n}\right)^{n/2} (1 + o(1)) \; .$$

[3]They actually showed the slightly stronger bound $|\mathcal{L} \cap rB_2^n| \geq 2\lfloor 2^{-n} \cdot \mathrm{vol}(rB_2^n)\rfloor + 1$. And, like Minkowski, they considered arbitrary norms, not just ℓ_2. (See, e.g., [GL87, Thm. 1 of Ch. 2, Sec. 7].).

Here, the *determinant* of the lattice is the inverse "global density" of the lattice

$$\det(\mathcal{L}) := \lim_{r\to\infty} \frac{\mathrm{vol}(B_2^n)}{|\mathcal{L} \cap rB_2^n|} ,$$

which can be computed as the absolute value of the determinant of any lattice basis. So, geometrically, Minkowski's Theorem says that "a globally dense lattice must also be locally dense." In terms of the Gaussian mass (which we think of as a smoothed version of the point-counting function $|\mathcal{L} \cap rB_2^n|$), Minkowski's theorem tells us that, e.g.,

$$\rho_3(\mathcal{L}) \geq \frac{3}{2} \tag{1}$$

for any lattice $\mathcal{L} \subset \mathbb{R}^n$ with $\det(\mathcal{L}) \leq 1$. I.e., $\eta_{1/2}(\mathcal{L}^*) \geq 1/3$. (There is a much easier proof of Eq. (1) that achieves a better constant. See Eq. (1.1).)

It is natural to ask whether a converse of Minkowski's Theorem holds. I.e., must a lattice with many points inside a relatively small ball have small determinant? If a lattice has high Gaussian mass, does it have low determinant?

Unfortunately, the answer to these (rather naive) questions is no. If a lattice $\mathcal{L} \subset \mathbb{R}^n$ has a low-determinant *sublattice* $\mathcal{L}' \subset \mathcal{L}$, then it will have many points in a small ball and therefore large Gaussian mass, though $\det(\mathcal{L})$ can be arbitrarily large. For example, take $\mathcal{L} \subset \mathbb{R}^2$ to be the lattice generated by $(\boldsymbol{e}_1/t, t^2\boldsymbol{e}_2)$ for some large $t > 0$ and $\mathcal{L}' \subset \mathcal{L}$ to be the sublattice generated by $\boldsymbol{e}_1/t$. Then, $\det(\mathcal{L}) = t$, but $|\mathcal{L} \cap rB_2^2| \geq |\mathcal{L}' \cap rB_2^2| \geq rt$, and it follows that $\rho_s(\mathcal{L}) \geq \rho_s(\mathcal{L}') \geq e^{-\pi}st$. So, we can make the determinant and the Gaussian mass arbitrarily large simultaneously.

Dadush conjectured that lattices like the one described in the previous paragraph are essentially the only counterexample. I.e., he conjectured that any lattice with large Gaussian mass must have a low-determinant sublattice [Dad12a, DR16]. In joint work with Oded

Regev, we proved this conjecture (written here in the contrapositive).

Theorem 2 ([RS17b])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ with $\det(\mathcal{L}') \geq 1$ for all sublattices $\mathcal{L}' \subseteq \mathcal{L}$,*

$$\rho_{1/t}(\mathcal{L}) \leq \frac{3}{2} ,$$

where $t := 10(\log n + 2)$. Equivalently, $\eta_{1/2}(\mathcal{L}^*) \leq t$.

Recall from Eq.(1) that $\eta_{1/2}(\mathcal{L}^*) \geq 1/3$ for any lattice with a sublattice of determinant less than one. (Here, we use the trivial fact that $\eta_{1/2}(\mathcal{L}^*) \geq \eta_{1/2}((\mathcal{L}')^*)$ for any sublattice $\mathcal{L}' \subseteq \mathcal{L}$.) So, (after scaling appropriately), Theorem 2 characterizes the smoothing parameter of any lattice up to a factor of $O(\log n)$. Furthermore, we must have $t \geq \eta_{1/2}(\mathbb{Z}^n) = \sqrt{\log n/\pi} + o(1)$, so that the theorem is the best possible up to a factor of $O(\sqrt{\log n})$ in t.

Theorem 2 has many applications. In particular, we derive from it the following novel point-counting bounds.

Theorem 3 ([RS17b, Corollary 1.4])**.** *For every lattice $\mathcal{L} \subset \mathbb{R}^n$ with $\det(\mathcal{L}') \geq 1$ for all sublattices $\mathcal{L}' \subseteq \mathcal{L}$, and every shift vector $\boldsymbol{u} \in \mathbb{R}^n$,*

1. *for any $r \geq 1$, $|\mathcal{L} \cap (rB_2^n + \boldsymbol{u})| \leq 3e^{\pi t^2 r^2}/2$;*
2. *for any $\sqrt{n/(2\pi)} \cdot t^{-1} \leq r \leq \sqrt{n/(2\pi)} \cdot t$, $|\mathcal{L} \cap (rB_2^n + \boldsymbol{u})| \leq (Ctr/\sqrt{n})^{n/2}$ for some universal constant $C > 0$; and*
3. *for any $r \geq \sqrt{n/(2\pi)} \cdot t$, $|\mathcal{L} \cap (rB_2^n + \boldsymbol{u})| \leq 2(2\pi e r^2/n)^{n/2}$,*

where $t := 10(\log n + 2)$.

In addition, Dadush and Regev presented many applications of Theorem 2 to a wide range of areas, from complexity theory to Brownian motion on tori [DR16]. (They presented these applications when Theorem 2 was still an unproven conjecture.)

This chapter is primarily based on joint work with Oded Regev, which appeared in the Symposium on the Theory of Computing (STOC), 2017 [RS17b], and some passages have been taken verbatim from this source.

A Rotation Identity and Related Inequalities

Chapter 3 shows a "rotation" identity concerning the Gaussian mass of lattice cosets that is closely related to Riemann's quartic theta identities (see, e.g., [Mum07]). (The $2^{n+o(n)}$-time algorithm for discrete Gaussian sampling discussed in the next section is perhaps best understood in terms of this identity.) From this identity, we derive the following "rotation inequality" relating the periodic Gaussian function evaluated at different points.

Theorem 4 ([RS17a])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$, parameter $s > 0$, and vectors $\boldsymbol{t}, \boldsymbol{u} \in \mathbb{R}^n$,*

$$f_{\mathcal{L},s}(\boldsymbol{t})^2 f_{\mathcal{L},s}(\boldsymbol{u})^2 \leq f_{\mathcal{L},s}(\boldsymbol{t}+\boldsymbol{u}) f_{\mathcal{L},s}(\boldsymbol{t}-\boldsymbol{u}) \ .$$

To understand Theorem 4, we note the easy identity

$$\rho_s(\boldsymbol{t})^2 \rho_s(\boldsymbol{u})^2 = \rho_s(\boldsymbol{t}+\boldsymbol{u}) \rho_s(\boldsymbol{t}-\boldsymbol{u}) \ . \tag{2}$$

So, Theorem 4 can be viewed as a relaxation of Eq. (2) that holds for the periodic Gaussian. Indeed, notice that Eq. (2) is in fact a rotation identity, as it follows from the basic fact that the vector $(\boldsymbol{t}+\boldsymbol{u}, \boldsymbol{t}-\boldsymbol{u})/\sqrt{2} \in \mathbb{R}^{2n}$ is a rotation of $(\boldsymbol{t}, \boldsymbol{u}) \in \mathbb{R}^{2n}$. (This is essentially the reason that we call Theorem 4 a "rotation inequality.")

From the (perhaps rather opaque) inequality in Theorem 4, we derive a surprising number of interesting corollaries about the discrete Gaussian and the periodic Gaussian function. For example, we show that $f_{\mathcal{L},s}(\boldsymbol{t})$ is monotonic in the parameter s (answering a question asked by Price [Pri14b]; see Proposition 3.4.1), and that the centered discrete Gaussian $D_{\mathcal{L},s}$ has

minimal covariance (answering a question asked by Dadush [Dad12a]; see Corollary 3.3.2).

This chapter is primarily based on joint work with Oded Regev that appeared in the SIAM Journal of Discrete Mathematics (SIDMA), 31(2) 2017 [RS17a], and passages have been taken verbatim from this source.

Algorithms for DGS (and CVP and SVP)

In Chapter 4, we show an algorithm that samples from the discrete Gaussian distribution and use this to obtain the fastest known algorithms for the two most important computational problems on lattices. As we discussed a bit above, algorithms for sampling from the discrete Gaussian $D_{\mathcal{L}-\boldsymbol{t},s}$ have played a central role in cryptographic constructions [Reg09, GPV08] and worst-case to average-case reductions [MR07, Reg09, Pei09, BLP+13] for over a decade. However, such algorithms previously only worked for parameters $s \gg \eta_\varepsilon(\mathcal{L})$. (See Table 1.) For such parameters, $D_{\mathcal{L}-\boldsymbol{t},s}$ is rather well-behaved in that it "looks like a continuous Gaussian," which makes it significantly easier to work with in practice. Even with this restriction, these algorithms often required access to powerful oracles, trapdoors, and/or quantum computers.

In joint work with Aggarwal, Dadush, and Regev [ADRS15] and follow-up work with Aggarwal and Dadush [ADS15], we show the first algorithm for sampling from the discrete Gaussian $D_{\mathcal{L}-\boldsymbol{t},s}$ for any parameter $s > \operatorname{dist}(\boldsymbol{t},\mathcal{L})/2^{o(n/\log n)}$. (In [Ste16a], discussed in Chapter 5, we show how to extend this to any parameter $s > 0$. See Theorem 5.3.6 and Corollary 9.)

Theorem 5 ([ADRS15, ADS15])**.** *There is a $2^{n+o(n)}$-time algorithm that takes as input a lattice $\mathcal{L} \subset \mathbb{R}^n$, a shift vector $\boldsymbol{t} \in \mathbb{R}^n$, and a parameter $s > \operatorname{dist}(\boldsymbol{t},\mathcal{L})/2^{o(n/\log n)}$ and outputs at least one sample from $D_{\mathcal{L}-\boldsymbol{t},s}$.*

Furthermore, for the special case when $\boldsymbol{t} = \boldsymbol{0}$, the algorithm outputs $2^{n/2}$ independent samples from $D_{\mathcal{L},s}$ (in the same running time), and for any $\boldsymbol{t} \in \mathbb{R}^n$, the algorithm outputs at

least

$$\frac{\rho_s(\mathcal{L}-\boldsymbol{t})}{\max_{\boldsymbol{c}\in\mathcal{L}/(2\mathcal{L})}\rho_s(2\mathcal{L}-\boldsymbol{c}-\boldsymbol{t})} \geq 1$$

independent samples from $D_{\mathcal{L}-\boldsymbol{t},s}$.

Applying Theorem 5 with $\boldsymbol{t}=\boldsymbol{0}$ more-or-less immediately yields a $2^{n+o(n)}$-time algorithm for the most important lattice problem, the Shortest Vector Problem (SVP, which asks for a shortest non-zero vector in a lattice), which is the current fastest known algorithm. With quite a bit more work, in [ADS15] we are able to use Theorem 5 for arbitrary $\boldsymbol{t}\in\mathbb{R}^n$ to obtain a $2^{n+o(n)}$-time algorithm for a harder problem, the Closest Vector Problem (CVP, which asks for the closest lattice vector to some target vector $\boldsymbol{t}$).

Theorem 6 ([ADRS15, ADS15])**.** *There is a* $2^{n+o(n)}$*-time algorithm for SVP, and a* $2^{n+o(n)}$*-time algorithm for CVP.*

We also show an algorithm for sampling from $D_{\mathcal{L}-\boldsymbol{t},s}$ that runs in $2^{n/2}$ time but only works above the smoothing parameter. This algorithm has a number of applications as well, which are discussed in [ADRS15].

Theorem 7 ([ADRS15])**.** *There is a* $2^{n/2+o(n)}$*-time algorithm that takes as input a lattice* $\mathcal{L}\subset\mathbb{R}^n$*, a shift vectors* $\boldsymbol{t}\in\mathbb{R}^n$*, and a parameter* $s>\sqrt{2}\eta_{1/2}(\mathcal{L})$ *and outputs* $2^{n/2}$ *independent samples from* $D_{\mathcal{L}-\boldsymbol{t},s}$.

We are primarily interested in Theorem 7 because of the potential to extend it to arbitrary parameters $s>0$ (at least for one sample, at least in the $\boldsymbol{t}=\boldsymbol{0}$ case). This would immediately yield a faster algorithm for SVP. (We do not expect to obtain faster algorithms for CVP via the methods of [ADS15], and recent work with Bennett and Golovnev [BGS17] shows some evidence that no faster algorithms for CVP exist at all.)

This chapter is primarily based on joint work with Divesh Aggarwal, Daniel Dadush, and Oded Regev, which appeared in the Symposium on the Theory of Computing (STOC),

2015 [ADRS15], and joint work with Divesh Aggarwal and Daniel Dadush, which appeared in the Symposium on the Foundations of Computer Science (FOCS), 2015 [ADS15]. Some passages have been taken verbatim from these sources.

Reduction(s) from DGS to CVP (and SVP)

The above shows the fastest known algorithm for CVP, which follows from an algorithm for discrete Gaussian sampling. Indeed, there is an easy efficient, dimension-preserving reduction from CVP to discrete Gaussian sampling. (See Corollary 1.3.11 and the discussion above it.) So, any algorithm that allows us to sample from $D_{\mathcal{L}-\boldsymbol{t},s}$ for arbitrary $\mathcal{L} \subset \mathbb{R}^n$, $\boldsymbol{t} \in \mathbb{R}^n$, and $s > 0$ immediately implies an algorithm for CVP with essentially the same running time.[4] It is therefore natural to ask whether there is a reduction in the other direction. In [Ste16a] and Chapter 5, we show that there is.

Theorem 8 ([Ste16a])**.** *CVP is equivalent to sampling from the discrete Gaussian distribution under dimension-preserving polynomial-time reductions.*

In particular, this theorem suggests that our discrete-Gaussian-based techniques for solving CVP are optimal in a certain (rather weak) sense. At the very least, it shows that we are not solving an unnecessarily difficult problem. Together with Theorem 6, this also implies the first algorithm for discrete Gaussian sampling with no restrictions at all on the shift $\boldsymbol{t}$ or the parameter $s > 0$, as follows.

Corollary 9 ([Ste16a])**.** *There is a $2^{n+o(n)}$-time algorithm that takes as input a lattice $\mathcal{L} \subset \mathbb{R}^n$, a shift vectors $\boldsymbol{t} \in \mathbb{R}^n$, and* any *parameter $s > 0$ and outputs a sample from $D_{\mathcal{L}-\boldsymbol{t},s}$.*

In contrast, the relationship between SVP and discrete Gaussian sampling is not yet settled. Here, the right notion of discrete Gaussian sampling seems to be sampling from the

[4] This reduction requires a discrete Gaussian sampling oracle that works for arbitrarily small parameters $s > 0$. In [ADS15] and Chapter 4, we have to work quite a bit harder because Theorem 5 works only for $s > \operatorname{dist}(\boldsymbol{t}, \mathcal{L})/2^{-o(n/\log n)}$.

centered distribution $D_{\mathcal{L},s}$, in which the shift vector $\boldsymbol{t}$ is zero. Indeed, we show a reduction from centered discrete Gaussian sampling to SVP.

Theorem 10 ([Ste16a])**.** *There is a dimension-preserving polynomial-time reduction from the problem of sampling from the* centered *discrete Gaussian $D_{\mathcal{L},s}$ to SVP.*

However, there is no known efficient reduction from SVP to sampling from $D_{\mathcal{L},s}$, even one that does not preserve the dimension.[5] So, the complexity of sampling from the *centered* discrete Gaussian $D_{\mathcal{L},s}$ is still poorly understood. In particular, we do not know whether it is NP-hard or whether it can be placed in some complexity class that is unlikely to contain NP-hard problems.

The techniques used to prove Theorems 8 and 10 are interesting in their own right. In particular, we show powerful generalizations of Khot's sparsification technique [Kho05] (see also Dadush and Kun [DK13] and [DRS14]). Indeed, since their original publication in [Ste16a], these techniques have found additional applications (e.g., [Ste16b, BSW16]).

This chapter is primarily based on work that appeared in the Symposium on Discrete Algorithms (SODA), 2016 [Ste16a], and passages have been taken verbatim from this source.

[5] In [ADRS15], we use exponentially many samples from $D_{\mathcal{L},s}$ to solve SVP.

Chapter 1

Preliminaries

We use c, C, C', C_1, C_2 to denote arbitrary positive universal constants, whose value might change from one occurrence to the next. Logarithms are base e unless otherwise specified, and we write $\exp(x) := e^x$. Vectors $\boldsymbol{x} \in \mathbb{R}^n$ are column vectors. We write $\|\boldsymbol{x}\|$ to represent the Euclidean norm of $\boldsymbol{x}$, and we write I_n for the identity matrix in n dimensions. For a matrix $A \in \mathbb{R}^{n \times n}$, we write A^T for the transpose of A. We write $B_2^n := \{\boldsymbol{x} \in \mathbb{R}^n \ : \ \|\boldsymbol{x}\| \leq 1\}$ for the Euclidean ball in $\mathbb{R}^n$. We write $\pi_S(\boldsymbol{x})$ for the orthogonal projection of $\boldsymbol{x}$ onto $\mathrm{span}(S)$ for some $S \subseteq \mathbb{R}^n$. (E.g., $\pi_{\boldsymbol{y}}(\boldsymbol{x}) = \langle \boldsymbol{y}, \boldsymbol{x} \rangle \boldsymbol{y} / \|\boldsymbol{y}\|^2$.) We write $S^\perp$ for the subspace of vectors orthogonal to S. For two additive subgroups $S_1 \subseteq \mathbb{R}^n$ and $S_2 \subseteq \mathbb{R}^m$, their direct sum $S_1 \oplus S_2 \subseteq \mathbb{R}^{n+m}$ is $\{(\boldsymbol{x}, \boldsymbol{y}) \ : \boldsymbol{x} \in S_1, \boldsymbol{y} \in S_2\}$.

When we discuss computational problems and algorithms, we strongly prefer to avoid getting bogged down in the nuances of how real numbers are represented computationally. In general, we are not particularly interested in the bit length of the input, but instead only on the dimension n. All of our results hold unambiguously if we assume that the input always consists of rational numbers with bit length that is polynomial in the dimension n.

1.1 Lattice basics

A *lattice* $\mathcal{L} \subset \mathbb{R}^n$ is the set

$$\mathcal{L} := \left\{ \sum z_i \boldsymbol{b}_i \ : \ z_i \in \mathbb{Z} \right\}$$

of integer linear combinations of linearly independent basis vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d$. The matrix $\mathbf{B} := (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$ is a *basis* of the lattice, and we sometimes write $\mathcal{L}(\mathbf{B})$ to denote the lattice generated by $\mathbf{B}$. We call d the *rank* of the lattice and write $\text{rank}(\mathcal{L}) := d$. By associating the span of the lattice with $\mathbb{R}^d$, we may always assume without loss of generality that $d = n$, though it is still sometimes convenient to talk about the rank of a sublattice. (Many of the definitions that follow only make sense for full-rank lattices.)

We write

$$\lambda_1(\mathcal{L}) := \min_{\boldsymbol{y} \in \mathcal{L} \setminus \{\mathbf{0}\}} \|\boldsymbol{y}\|$$

for the length of the shortest non-zero vector in the lattice. More generally, for $1 \leq i \leq n$, we define the *ith successive minimum*,

$$\lambda_i(\mathcal{L}) := \min\{r \ : \ \dim(\text{span}(\mathcal{L} \cap rB_2^n)) = i\} \ .$$

For $\boldsymbol{t} \in \mathbb{R}^n$, we write

$$\text{dist}(\boldsymbol{t}, \mathcal{L}) := \min_{\boldsymbol{y} \in \mathcal{L}} \|\boldsymbol{y} - \boldsymbol{t}\|$$

for the distance between $\boldsymbol{t}$ and the lattice. The *covering radius* is then

$$\mu(\mathcal{L}) := \max_{\boldsymbol{t} \in \mathbb{R}^n} \text{dist}(\boldsymbol{t}, \mathcal{L}) \ ,$$

the farthest distance from the lattice.

The *dual lattice* is

$$\mathcal{L}^* := \{\boldsymbol{w} \in \mathbb{R}^n \ : \ \forall \boldsymbol{y} \in \mathcal{L},\ \langle \boldsymbol{w}, \boldsymbol{y} \rangle \in \mathbb{Z}\} \ .$$

One can check that $\mathcal{L}^*$ is itself a lattice with basis $\mathbf{B}^{-T}$, and therefore that $(\mathcal{L}^*)^* = \mathcal{L}$.

The *determinant* of the lattice is given by $\det(\mathcal{L}(\mathbf{B})) := |\det(\mathbf{B})|$. One can show that the determinant is well defined (i.e., it does not depend on the choice of basis $\mathbf{B}$). It follows that, if $\mathcal{L} \subset \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$ is non-singular, then $\det(A\mathcal{L}) = |\det(A)| \det(\mathcal{L})$, and that $\det(\mathcal{L}^*) = 1/\det(\mathcal{L})$.

A *sublattice* $\mathcal{L}' \subseteq \mathcal{L}$ is an additive subgroup of $\mathcal{L}$. We say that $\mathcal{L}'$ is *primitive* if $\mathcal{L}' = \mathcal{L} \cap \mathrm{span}(\mathcal{L}')$, and we call $\mathrm{span}(\mathcal{L}')$ a *lattice subspace*. For a primitive sublattice $\mathcal{L}' \subseteq \mathcal{L}$, we define the quotient lattice $\mathcal{L}/\mathcal{L}' := \pi_{\mathcal{L}'^{\perp}}(\mathcal{L})$ to be the projection of $\mathcal{L}$ onto the space orthogonal to $\mathcal{L}'$. In particular, $\mathcal{L}/\mathcal{L}'$ is a lattice, and we have the identities $(\mathcal{L}/\mathcal{L}')^* = \mathcal{L}^* \cap \mathrm{span}(\mathcal{L}')^{\perp}$ and $\det(\mathcal{L}/\mathcal{L}') = \det(\mathcal{L})/\det(\mathcal{L}')$. If $\mathcal{L}' \subset \mathcal{L}$ is a full-rank sublattice (and therefore is not primitive), we write

$$\mathcal{L}/\mathcal{L}' := \{\boldsymbol{y} \bmod \mathcal{L}' \ : \ \boldsymbol{y} \in \mathcal{L}\}$$

for the set of cosets of $\mathcal{L}$ over $\mathcal{L}'$. This slightly overloaded notation should not cause any confusion, as we use the two different notions in very different contexts. And, for $\boldsymbol{c} \in \mathcal{L}/\mathcal{L}'$ for some full-rank sublattice $\mathcal{L}'$, we always write $\mathcal{L}' + \boldsymbol{c}$ instead of just $\boldsymbol{c}$ to make it clear that the coset is a set (as opposed to a single vector).

Given a basis, $\mathbf{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$, we define its *Gram-Schmidt orthogonalization* $(\widetilde{\boldsymbol{b}}_1, \ldots, \widetilde{\boldsymbol{b}}_n)$ by

$$\widetilde{\boldsymbol{b}}_i = \pi_{\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{i-1}\}^{\perp}}(\boldsymbol{b}_i) \ .$$

1.2 Computational Problems

Definition 1.2.1. *For any approximation factor $\gamma = \gamma(n) \geq 1$, the γ-approximate Shortest Vector Problem (γ-SVP) is defined as follows. The input is (a basis for) a lattice $\mathcal{L} \subset \mathbb{R}^n$, and the goal is to output a non-zero lattice vector $\boldsymbol{y} \in \mathcal{L} \setminus \{\mathbf{0}\}$ with $\|\boldsymbol{y}\| \leq \gamma(n) \cdot \lambda_1(\mathcal{L})$.*

Definition 1.2.2. *For any approximation factor $\gamma = \gamma(n) \geq 1$, the γ-approximate Closest Vector Problem (γ-CVP) is defined as follows. The input is (a basis for) a lattice $\mathcal{L} \subset \mathbb{R}^n$ and a target vector $\boldsymbol{t} \in \mathbb{R}^n$, and the goal is to output a lattice vector $\boldsymbol{y} \in \mathcal{L}$ with $\|\boldsymbol{y} - \boldsymbol{t}\| \leq \gamma(n) \cdot \mathrm{dist}(\boldsymbol{t}, \mathcal{L})$.*

When $\gamma = 1$, we often omit it and simply write SVP or CVP.

The following theorem was proven by Ajtai, Kumar, and Sivakumar [AKS01], building on work of Schnorr [Sch87].

Theorem 1.2.3 ([Sch87, AKS01])**.** *There is an algorithm that takes as input a lattice $\mathcal{L} \subset \mathbb{R}^n$ and $u \geq 2$ and outputs an $u^{n/y}$-reduced basis of $\mathcal{L}$ in time $\exp(O(u)) \cdot \mathrm{poly}(n)$, where we say that a basis $\mathbf{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ of a lattice $\mathcal{L}$ is γ-reduced for some $\gamma \geq 1$ if*

1. *$\|\boldsymbol{b}_1\| \leq \gamma \cdot \lambda_1(\mathcal{L})$; and*
2. *$\pi_{\{\boldsymbol{b}_1\}^\perp}(\boldsymbol{b}_2), \ldots, \pi_{\{\boldsymbol{b}_1\}^\perp}(\boldsymbol{b}_n)$ is a γ-reduced basis of $\pi_{\{\boldsymbol{b}_1\}^\perp}(\mathcal{L})$.*

We will also need the following celebrated result due to Babai [Bab86].

Theorem 1.2.4 ([Bab86])**.** *There is an efficient algorithm that solves $2^{n/2}$-CVP.*

1.3 Gaussian measure on lattices

For a vector $\boldsymbol{x} \in \mathbb{R}^n$ and a parameter $s > 0$, we write $\rho_s(\boldsymbol{x}) := \exp(-\pi\|\boldsymbol{x}\|^2/s^2)$ for the *Gaussian mass of $\boldsymbol{x}$ with parameter s*. Then, for any discrete set $A \subset \mathbb{R}^n$, we can extend this

notion in the natural way:

$$\rho_s(A) := \sum_{\boldsymbol{y} \in A} \rho_s(\boldsymbol{y}) \ .$$

We are particularly interested in $\rho_s(\mathcal{L} - \boldsymbol{t})$ for a lattice $\mathcal{L} \subset \mathbb{R}^n$ and shift vector $\boldsymbol{t} \in \mathbb{R}^n$. By the Poisson Summation Formula, we have

$$\rho_s(\mathcal{L}) := \frac{s^n}{\det(\mathcal{L})} \cdot \rho_{1/s}(\mathcal{L}^*) \ . \tag{1.1}$$

More generally,

$$\rho_s(\mathcal{L} - \boldsymbol{t}) := \frac{s^n}{\det(\mathcal{L})} \cdot \sum_{\boldsymbol{w} \in \mathcal{L}^*} \rho_{1/s}(\boldsymbol{w}) \cos(2\pi \langle \boldsymbol{w}, \boldsymbol{t} \rangle) \ . \tag{1.2}$$

In particular, $\rho_s(\mathcal{L} - \boldsymbol{t}) \leq \rho_s(\mathcal{L})$ for any $\boldsymbol{t} \in \mathbb{R}^n$ with equality if and only if $\boldsymbol{t} \in \mathcal{L}$.

The *discrete Gaussian distribution over* $\mathcal{L} - \boldsymbol{t}$ *with parameter* $s > 0$ is the probability distribution $D_{\mathcal{L}-\boldsymbol{t},s}$ induced by this measure. I.e., for any $\boldsymbol{y} \in \mathcal{L}$,

$$\Pr_{\boldsymbol{X} \sim D_{\mathcal{L}-\boldsymbol{t},s}} [\boldsymbol{X} = \boldsymbol{y} - \boldsymbol{t}] = \rho_s(\boldsymbol{y} - \boldsymbol{t}) / \rho_s(\mathcal{L} - \boldsymbol{t}) \ .$$

The *periodic Gaussian function with parameter* $s > 0$ is defined as

$$f_{\mathcal{L},s}(\boldsymbol{t}) := \frac{\rho_s(\mathcal{L} - \boldsymbol{t})}{\rho_s(\mathcal{L})} \ .$$

The Poisson Summation Formula (Eq. (1.2)) shows that the discrete Gaussian and the periodic Gaussian are in some sense duals of each other:

$$f_{\mathcal{L},s}(\boldsymbol{t}) := \mathop{\mathbb{E}}_{\boldsymbol{w} \sim D_{\mathcal{L}^*}} [\cos(2\pi \langle \boldsymbol{w}, \boldsymbol{t} \rangle)] \ . \tag{1.3}$$

For $\varepsilon > 0$, the *lattice smoothing parameter* is defined as

$$\eta_\varepsilon(\mathcal{L}) := \min\{s \ : \ \rho_{1/s}(\mathcal{L}^*) \le 1 + \varepsilon\} \ .$$

The smoothing parameter was introduced by Micciancio and Regev [MR07], and it takes its name from Eq. (1.3), which in particular shows that for $s \ge \eta_\varepsilon(\mathcal{L})$,

$$\frac{1-\varepsilon}{1+\varepsilon} \le f_{\mathcal{L},s}(\boldsymbol{t}) \le 1 \ , \tag{1.4}$$

so that for small ε, $f_{\mathcal{L},s}(\boldsymbol{t})$ is nearly constant or "smooth."

Claim 1.3.1. *For any lattice $\mathcal{L} \subset \mathbb{R}^n$, primitive sublattice $\mathcal{L}' \subseteq \mathcal{L}$, and parameter $s > 0$,*

$$\rho_s(\mathcal{L}) \le \rho_s(\mathcal{L}')\rho_s(\mathcal{L}/\mathcal{L}')$$

with equality if and only if $\mathcal{L} \cong \mathcal{L}' \oplus \mathcal{L}/\mathcal{L}'$.

Proof. Let $\widehat{\mathcal{L}} \subset \mathbb{R}^n$ be a lattice such that every vector $\boldsymbol{y} \in \mathcal{L}$ can be written uniquely as $\widehat{\boldsymbol{y}} + \boldsymbol{y}'$ with $\widehat{\boldsymbol{y}} \in \widehat{\mathcal{L}}$ and $\boldsymbol{y}' \in \mathcal{L}'$. (It is easy to see that such a lattice exists if and only if $\mathcal{L}'$ is primitive.) Let $\pi := \pi_{\mathcal{L}'}$ and $\pi^\perp := \pi_{(\mathcal{L}')^\perp}$. We have

$$\begin{aligned}
\rho_s(\mathcal{L}) &= \sum_{\boldsymbol{y} \in \mathcal{L}} \rho_s(\boldsymbol{y}) \\
&= \sum_{\widehat{\boldsymbol{y}} \in \widehat{\mathcal{L}}} \sum_{\boldsymbol{y}' \in \mathcal{L}'} \rho_s(\pi(\widehat{\boldsymbol{y}}) + \pi(\boldsymbol{y}'))\rho_s(\pi^\perp(\widehat{\boldsymbol{y}})) \\
&= \sum_{\widehat{\boldsymbol{y}} \in \widehat{\mathcal{L}}} \rho_s(\mathcal{L}' + \pi(\widehat{\boldsymbol{y}}))\rho_s(\pi^\perp(\widehat{\boldsymbol{y}})) \\
&\le \rho_s(\mathcal{L}') \sum_{\widehat{\boldsymbol{y}} \in \widehat{\mathcal{L}}} \rho_s(\pi^\perp(\widehat{\boldsymbol{y}})) \\
&= \rho_s(\mathcal{L}')\rho_s(\mathcal{L}/\mathcal{L}') \ ,
\end{aligned}$$

as needed. Here, the inequality follows from Eq. (1.2) (and in particular the observation afterwards), as does the fact that equality holds if and only if $\pi(\widehat{\boldsymbol{y}}) \in \mathcal{L}'$ for all $\widehat{\boldsymbol{y}} \in \widehat{\mathcal{L}}$, i.e., if and only if $\mathcal{L} \cong \mathcal{L}' \oplus \mathcal{L}/\mathcal{L}'$. □

Chapters 4 and 5 will be primarily focused on the following computational problem, which asks us to sample from $D_{\mathcal{L}-\boldsymbol{t},s}$.

Definition 1.3.2. *For any $\varepsilon = \varepsilon(n) \in (0,1)$ and functions $\sigma(\mathcal{L}-\boldsymbol{t}), m(\mathcal{L}-\boldsymbol{t}) \geq 0$ over lattice cosets $\mathcal{L}-\boldsymbol{t}$, the Discrete Gaussian Sampling problem with error ε, parameter σ, and output size m (ε-DGS$_\sigma^m$) is defined as follows. The input is a (basis for a) lattice $\mathcal{L} \subset \mathbb{R}^n$, shift vector $\boldsymbol{t} \in \mathbb{R}^n$, and parameter $s > \sigma(\mathcal{L}-\boldsymbol{t})$. The goal is to output $\widehat{m} \geq m(\mathcal{L}-\boldsymbol{t})$ vectors in $\mathcal{L}-\boldsymbol{t}$ whose joint distribution is within statistical distance ε of independent samples from $D_{\mathcal{L}-\boldsymbol{t},s}$.*

This rather technical definition might be easier to understand with a few examples. A natural value for $\sigma(\mathcal{L}-\boldsymbol{t})$ is simply $\sigma(\mathcal{L}-\boldsymbol{t}) = \eta_\varepsilon(\mathcal{L})$ for some $\varepsilon > 0$, which corresponds to sampling "above the smoothing parameter." One could also imagine, say, $\sigma(\mathcal{L}-\boldsymbol{t}) := \min\{s : \rho_s(\mathcal{L}-\boldsymbol{t}) \geq (1-\varepsilon)\rho_s(\mathcal{L})\}$. In our primary example, we will have $\sigma(\mathcal{L}-\boldsymbol{t}) = f(n)\operatorname{dist}(\boldsymbol{t},\mathcal{L})$ for some tiny function $f(n) \approx 2^{-n/\log n}$. (It is much less natural to take m to be a function of $\mathcal{L}-\boldsymbol{t}$, but our algorithm in Chapter 4 happens to output a different number of vectors depending on $\mathcal{L}-\boldsymbol{t}$.)

1.3.1 Banaszczyk's theorem and some consequences

Here, we present Banaszczyk's celebrated theorem (Theorem 1.3.4) [Ban93] and many of its immediate consequences. Since these results are so important to the study of the Gaussian measure on lattices, we are a bit pedantic, and we prove more than just what we will need in the sequel.

Lemma 1.3.3 ([Ban93]). *For any lattice $\mathcal{L} \subset \mathbb{R}^n$, parameter $s \geq 1$, and shift $\boldsymbol{t} \in \mathbb{R}^n$, $\rho_s(\mathcal{L} - \boldsymbol{t}) \leq s^n \rho(\mathcal{L})$.*

Proof. By Eq. (1.2), we have

$$\rho_s(\mathcal{L} - \boldsymbol{t}) = \frac{s^n}{\det(\mathcal{L})} \sum_{\boldsymbol{w} \in \mathcal{L}^*} \rho_{1/s}(\boldsymbol{w}) \cos(2\pi \langle \boldsymbol{w}, \boldsymbol{t} \rangle) \leq \frac{s^n}{\det(\mathcal{L})} \sum_{\boldsymbol{w} \in \mathcal{L}^*} \rho(\boldsymbol{w}) = s^n \rho(\mathcal{L}) \ ,$$

where the inequality follows from the fact that $\cos(x) \leq 1$ and the fact that the Gaussian measure is monotonically increasing in s. □

With this, we can prove Banaszczyk's main theorem.

Theorem 1.3.4 ([Ban93]). *For any lattice $\mathcal{L} \subset \mathbb{R}^n$, parameter $s > 0$, shift $\boldsymbol{t} \in \mathcal{L}$, and $u \geq 1/\sqrt{2\pi}$,*

$$\rho_s((\mathcal{L} - \boldsymbol{t}) \setminus u\sqrt{n} s B_2^n) \leq (2\pi e u^2)^{n/2} \cdot \exp(-\pi u^2 n) \cdot \rho_s(\mathcal{L})$$

Proof. We may assume without loss of generality that $s = 1$. For any $\sigma \geq 1$, we have by Lemma 1.3.3 that $\rho_\sigma(\mathcal{L} - \boldsymbol{t}) \leq \sigma^n \rho(\mathcal{L})$. On the other hand,

$$\begin{aligned}
\rho_\sigma(\mathcal{L} - \boldsymbol{t}) &\geq \sum_{\boldsymbol{y} \in \mathcal{L} \setminus u\sqrt{n} B_2^n} \rho_\sigma(\boldsymbol{y} - \boldsymbol{t}) \\
&\geq \exp(\pi u^2 (1 - 1/\sigma^2)) \sum_{\boldsymbol{y} \in \mathcal{L} \setminus u\sqrt{n} B_2^n} \rho(\boldsymbol{y} - \boldsymbol{t}) \\
&= \exp(\pi u^2 n - \pi u^2 n/\sigma^2) \cdot \rho((\mathcal{L} - \boldsymbol{t}) \setminus u\sqrt{n} B_2^n) \ .
\end{aligned}$$

We therefore have

$$\rho((\mathcal{L} - \boldsymbol{t}) \setminus u\sqrt{n} B_2^n) \leq (\sigma e^{\pi u^2/\sigma^2})^n \exp(-\pi u^2 n) \ .$$

The result follows by setting $\sigma := \sqrt{2\pi} u \geq 1$. □

The following slightly weaker version of Theorem 1.3.4 is a bit more convenient and often sufficient.

Corollary 1.3.5. *For any lattice $\mathcal{L} \subset \mathbb{R}^n$, parameter $s > 0$, shift $\boldsymbol{t} \in \mathbb{R}^n$, and radius $r > \sqrt{n/(2\pi)} \cdot s$,*

$$\rho_s((\mathcal{L} - \boldsymbol{t}) \setminus rB_2^n) < \exp(-\pi x^2)\rho_s(\mathcal{L}) \ ,$$

where $x := r/s - \sqrt{n/(2\pi)}$.

Proof. Let

$$u := \frac{r}{s\sqrt{n}} = \frac{1}{\sqrt{2\pi}} + \frac{x}{\sqrt{n}} \ .$$

We have

$$\begin{aligned} \frac{\rho_s((\mathcal{L} - \boldsymbol{t}) \setminus rB_2^n)}{\rho_s(\mathcal{L})} &\le (2\pi e u^2)^{n/2} \cdot \exp(-\pi u^2 n) \\ &= (1 + \sqrt{2\pi/n} \cdot x)^n \exp(-\sqrt{2\pi n}x - \pi x^2) \\ &< \exp(-\pi x^2) \ . \end{aligned}$$

□

Banaszczyk's theorem in particular implies a bound on the smoothing parameter of the following form.

Corollary 1.3.6. *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and $\varepsilon \in (0, 1)$,*

$$\sqrt{\frac{\log(2/\varepsilon)}{\pi}} < \eta_\varepsilon(\mathcal{L})\lambda_1(\mathcal{L}^*) < \sqrt{\frac{n}{2\pi}} + \sqrt{\frac{\log(2/\varepsilon)}{\pi}} \ .$$

Proof. The lower bound is trivial. In particular, for any $s \le \sqrt{\log(2/\varepsilon)/\pi}/\lambda_1(\mathcal{L}^*)$,

$$\rho_{1/s}(\mathcal{L}^*) > 1 + 2\exp(-\pi s^2 \lambda_1(\mathcal{L}^*)) \ge 1 + \varepsilon \ .$$

For the upper bound, we assume without loss of generality that $\lambda_1(\mathcal{L}^*) = 1$ and notice that

for any $s > 0$,

$$\rho_{1/s}(\mathcal{L}^*) - 1 = \rho_{1/s}(\mathcal{L}^* \setminus B_2^n) \ .$$

In particular, if we take $r = 1$ and set

$$s := \sqrt{\frac{n}{2\pi}} + \sqrt{\frac{\log(2/\varepsilon)}{\pi}} \ .$$

then by Corollary 1.3.5, we have $\rho_{1/s}(\mathcal{L}^*) - 1 < \varepsilon \rho_{1/s}(\mathcal{L}^*)/2$. The result follows. □

We get the following bound with roughly the same proof.

Corollary 1.3.7. *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and $\varepsilon \in (0, 1)$,*

$$\frac{\mu(\mathcal{L})}{\eta_\varepsilon(\mathcal{L})} < \sqrt{\frac{n}{2\pi}} + \sqrt{\frac{\log((1+\varepsilon)/(1-\varepsilon))}{\pi}} \ .$$

Proof. By scaling the lattice, we may assume without loss of generality that $\eta_\varepsilon(\mathcal{L}) = 1$. Let

$$r := \sqrt{\frac{n}{2\pi}} + \sqrt{\frac{\log((1+\varepsilon)/(1-\varepsilon))}{\pi}} \ .$$

By Corollary 1.3.5, for any $\boldsymbol{t} \in \mathbb{R}^n$, we have

$$\rho((\mathcal{L} - \boldsymbol{t}) \setminus rB_2^n) < \frac{1-\varepsilon}{1+\varepsilon} \cdot \rho(\mathcal{L}) \ .$$

On the other hand, by Eq. (1.2),

$$\rho(\mathcal{L} - \boldsymbol{t}) \geq \frac{1-\varepsilon}{1+\varepsilon} \cdot \rho(\mathcal{L}) > \rho((\mathcal{L} - \boldsymbol{t}) \setminus u\sqrt{n}B_2^n) \ ,$$

so that $(\mathcal{L} - \boldsymbol{t}) \cap u\sqrt{n}B_2^n$ is nonempty. Since this holds for arbitrary $\boldsymbol{t}$, we have $\mu(\mathcal{L}) < r$, as needed. □

Corollary 1.3.8. *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and $\varepsilon \in (0,1)$,*

$$\frac{\lambda_n(\mathcal{L})}{\eta_\varepsilon(\mathcal{L})} < \sqrt{\frac{2n}{\pi}} + 2\sqrt{\frac{\log((1+\varepsilon)/(1-\varepsilon))}{\pi}} \,.$$

Proof. It is easy to see that $\lambda_n(\mathcal{L}) \le 2\mu(\mathcal{L})$. The result then follows immediately from Corollary 1.3.7. □

Corollary 1.3.9. *For any lattice $\mathcal{L} \subset \mathbb{R}^n$,*

$$\mu(\mathcal{L})\lambda_1(\mathcal{L}^*) < \frac{n + 10\sqrt{n}}{2\pi} \,.$$

Proof. Let $\varepsilon := 1/2$. Then, we have

$$\begin{aligned}
\mu(\mathcal{L})\lambda_1(\mathcal{L}^*) &< \left(\sqrt{\frac{n}{2\pi}} + \sqrt{\frac{\log(3)}{\pi}}\right) \cdot \eta_\varepsilon(\mathcal{L})\lambda_1(\mathcal{L}^*) \\
&< \left(\sqrt{\frac{n}{2\pi}} + \sqrt{\frac{\log(3)}{\pi}}\right) \cdot \left(\sqrt{\frac{n}{2\pi}} + \sqrt{\frac{\log(4)}{\pi}}\right) \\
&< \frac{n + 10\sqrt{n}}{2\pi} \,.
\end{aligned}$$

□

Banaszczyk also proved the following lemma.

Lemma 1.3.10 ([Ban93])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$, any parameter $s > 0$, and any shift $\boldsymbol{t} \in \mathbb{R}^n$,*

$$\exp(-\pi \operatorname{dist}(\boldsymbol{t}, \mathcal{L})^2/s^2)\rho_s(\mathcal{L}) \le \rho_s(\mathcal{L} - \boldsymbol{t}) \le \rho_s(\mathcal{L}) \,.$$

Proof. The upper bound follows immediately from Eq. (1.2). For the lower bound, since the function $\boldsymbol{t} \mapsto \rho_s(\mathcal{L}-\boldsymbol{t})$ is periodic over the lattice, we may assume without loss of generality that $\operatorname{dist}(\boldsymbol{t}, \mathcal{L}) = \|\boldsymbol{t}\|$ (i.e., that $\mathbf{0}$ is a closest lattice vector to $\boldsymbol{t}$) so that $\exp(-\pi \operatorname{dist}(\boldsymbol{t}, \mathcal{L})^2/s^2) =$

$\rho_s(\boldsymbol{t})$. Then, we have

$$\begin{aligned}\rho_s(\mathcal{L}-\boldsymbol{t}) &= \sum_{\boldsymbol{y}\in\mathcal{L}} \rho_s(\boldsymbol{y}-\boldsymbol{t}) \\ &= \frac{1}{2}\sum_{\boldsymbol{y}\in\mathcal{L}} (\rho_s(\boldsymbol{y}-\boldsymbol{t}) + \rho_s(-\boldsymbol{y}-\boldsymbol{t})) \\ &= \rho_s(\boldsymbol{t}) \sum_{\boldsymbol{y}\in\mathcal{L}} \rho_s(\boldsymbol{y}) \cosh(2\pi\langle \boldsymbol{y}, \boldsymbol{t}\rangle/s^2) \\ &\geq \rho_s(\boldsymbol{t})\rho_s(\mathcal{L}) \ . \end{aligned}$$

□

From this and Theorem 1.3.4, we derive the following convenient result, which shows that samples from $D_{\mathcal{L}-\boldsymbol{t},s}$ yield good approximate solutions for CVP. (We make little attempt to optimize the parameters here.) Furthermore, notice that a single sample from $D_{\mathcal{L}-\boldsymbol{t},s}$ with an arbitrarily small parameter is sufficient to solve *exact* CVP for, say, rational lattices $\mathcal{L} \subset \mathbb{Q}^n$. (See [Ste16a]. One can make this statement more general, with the size of the necessary parameter $s > 0$ depending on the specifics of the input format of the lattice. We therefore view Corollary 1.3.11 as a simple, efficient reduction from CVP to discrete Gaussian sampling with arbitrarily small parameters $s > 0$, without worrying about what conditions on the input format are sufficient to make the bit length of the required parameter s polynomial in the input length.)

Corollary 1.3.11. *For any lattice $\mathcal{L} \subset \mathbb{R}^n$, parameter $s > 0$, shift $\boldsymbol{t} \in \mathbb{R}^n$, and radius $r > \sqrt{n/(2\pi)} \cdot s$, with $r > \operatorname{dist}(\boldsymbol{t}, \mathcal{L})$ and*

$$r^2 > \operatorname{dist}(\boldsymbol{t}, \mathcal{L})^2 + \frac{ns^2}{\pi} \cdot \log(2\pi \operatorname{dist}(\boldsymbol{t}, \mathcal{L})^2/(ns^2)) \ ,$$

we have

$$\Pr_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t},s}} [\|\boldsymbol{X}\| > r] < (2e)^{n/2+1} \exp(-\pi y^2/2) \ ,$$

where $y := \sqrt{r^2 - \operatorname{dist}(\boldsymbol{t}, \mathcal{L})^2}/s$.

Proof. Let $d := \operatorname{dist}(\boldsymbol{t}, \mathcal{L})/s$, and

$$u := \frac{r}{s\sqrt{n}} = \sqrt{(y^2 + d^2)/n} \; .$$

By Theorem 1.3.4, we have

$$\begin{aligned} \frac{\rho_s((\mathcal{L} - \boldsymbol{t}) \setminus rB_2^n)}{\rho_s(\mathcal{L})} &\leq (2\pi e u^2)^{n/2} \cdot \exp(-\pi u^2 n) \\ &= (2\pi e(y^2 + d^2)/n)^{n/2} \cdot \exp(-\pi(y^2 + d^2)) \; . \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\rho_s((\mathcal{L} - \boldsymbol{t}) \setminus rB_2^n)}{\rho_s(\mathcal{L} - \boldsymbol{t})} &\leq (2\pi e(y^2 + d^2)/n)^{n/2} \cdot \exp(-\pi(y^2 + d^2)) \cdot \frac{\rho_s(\mathcal{L})}{\rho_s(\mathcal{L} - \boldsymbol{t})} \\ &\leq (2\pi e(y^2 + d^2)/n)^{n/2} \cdot \exp(-\pi y^2) \; , \end{aligned}$$

where the last line follows from Lemma 1.3.10.

Now, we consider two cases. If $d^2 \geq n/\pi$, then the function $y \mapsto (2\pi e(d^2 + y^2)/n)^{n/2} \cdot \exp(-\pi y^2/2)$ is decreasing in y. Since $y^2 = r^2/s^2 - d^2 > n \log(2\pi d^2/n)/\pi$, it follows that

$$(2\pi e(y^2 + d^2)/n)^{n/2} \cdot \exp(-\pi y^2) < 5 \exp(-\pi y^2/2) \; .$$

On the other hand, if $d^2 < n/\pi$, then we note that

$$(2\pi e(y^2 + d^2)/n)^{n/2} \cdot \exp(-\pi y^2/2) \leq 2^{n/2} \cdot \exp(\pi d^2/2) < (2e)^{n/2} \; .$$

The result follows. □

1.3.2 Algorithms for one-dimensional Gaussians

Brakerski, Langlois, Peikert, Regev, and Stehlé show how to efficiently sample from the one-dimensional discrete Gaussian $D_{\mathbb{Z}+c,s}$ for any $c \in \mathbb{R}$ and $s > 0$ [BLP+13]. For completeness, we describe a slightly modified version of their algorithm to sample from $D_{\mathbb{Z}\setminus\{0\},s}$ (which we will need in Chapter 5).

Lemma 1.3.12. *There is an algorithm that samples from $D_{\mathbb{Z}\setminus\{0\},s}$ for any $s > 0$ in (expected) polynomial time.*

Proof. We describe an algorithm that samples from $D_{\mathbb{Z}^+,s}$, which is clearly sufficient. Let $Z := e^{-\pi/s^2} + \int_1^\infty e^{-\pi x^2/s^2} \mathrm{d}x$. The algorithm outputs 1 with probability $e^{-\pi/s^2}/Z$. Otherwise, it samples x from the one-dimensional continuous Gaussian with parameter s restricted to the interval $(1, \infty)$. Let $y := \lceil x \rceil$. With probability $e^{-\pi(y^2-x^2)/s^2}$, the algorithm outputs y. Otherwise, it repeats.

On a single run of the algorithm, for any integer $z \geq 2$, the probability that the algorithm outputs z is

$$\frac{1}{Z} \cdot \int_{z-1}^{z} e^{-\pi x^2/s^2} \cdot e^{-\pi(z^2-x^2)/s^2} \mathrm{d}x = \frac{e^{-\pi z^2/s^2}}{Z} .$$

And, the probability that the algorithm outputs 1 is of course $e^{-\pi/s^2}/Z$. So, the algorithm outputs the correct distribution.

It remains to bound the expected running time. After a single run, the algorithm outputs an integer with probability

$$\frac{\rho_s(\mathbb{Z}^+)}{Z} = \frac{\rho_s(\mathbb{Z}^+)}{e^{-\pi/s^2} + \int_1^\infty e^{-\pi x^2/s^2} \mathrm{d}x} \geq \frac{1}{2} .$$

It follows that it runs in expected polynomial time. □

Brakerski et al. also noted a simple algorithm to compute $\rho_s(\mathbb{Z})$ for arbitrary $s > 0$ [BLP+13]. (We do our best to avoid the question of what it means to "efficiently compute"

a real number. Here, we simply note a very rapidly convergent series of elementary functions, which is more than sufficient for our purposes.)

Claim 1.3.13. *There is an efficient algorithm that computes $\rho_s(\mathbb{Z} \setminus \{0\})$ for any $s > 0$.*

Proof. If $s \leq 1$, we simply write

$$\rho_s(\mathbb{Z} \setminus \{0\}) = 2 \sum_{z \geq 1} \exp(-\pi z^2/s^2) \ .$$

Notice that this summation converges extremely rapidly. In particular, $O(\sqrt{m})$ terms are sufficient to obtain m bits of accuracy. For $s > 1$, we apply the Poisson Summation Formula (Eq. (1.1)) to obtain

$$\rho_s(\mathbb{Z} \setminus \{0\}) = s\rho_{1/s}(\mathbb{Z}) - 1 = s - 1 + 2s \sum_{z \geq 1} \exp(-\pi s^2 z^2) \ ,$$

which again converges extremely rapidly. □

1.3.3 Algorithms for arbitrary Gaussians with large parameters

For sampling from $D_{\mathcal{L}-\boldsymbol{t},s}$ in high dimensions for large parameters $s \gg \eta_{1/2}(\mathcal{L})$, we can use the celebrated algorithm introduced by Klein and further analyzed by Gentry, Peikert, and Vaikuntanathan suffices [Kle00, GPV08]. For convenience, we use the following strengthening of this result due to Brakerski et al., which provides exact samples and gives slightly better bounds on the parameter s.

Theorem 1.3.14 ([BLP+13, Lemma 2.3])**.** *There is a probabilistic polynomial-time algorithm that takes as input a basis $\mathbf{B}$ for a lattice $\mathcal{L} \subset \mathbb{R}^n$, a shift $\boldsymbol{t} \in \mathbb{R}^n$, and $\hat{s} > C\sqrt{\log n} \cdot \|\widetilde{\mathbf{B}}\|$ and outputs a vector that is distributed exactly as $D_{\mathcal{L}-\boldsymbol{t},\hat{s}}$, where $\|\widetilde{\mathbf{B}}\| := \max\|\widetilde{\boldsymbol{b}}_i\|$.*

(In [GPV08], they show that $C\sqrt{\log n} \cdot \|\widetilde{\mathbf{B}}\| \geq \eta_{1/2}(\mathcal{L})$ for any lattice $\mathcal{L} \subset \mathbb{R}^n$ with basis $\mathbf{B}$, so that this algorithm really does only allow sampling above the smoothing parameter.)

When instantiated with a γ-reduced basis, Theorem 1.3.14 allows us to sample with parameter $\hat{s} = \gamma \cdot \mathrm{poly}(n) \cdot \lambda_n(\mathcal{L})$. After running our combiner $o(n/\log n)$ times, this will allow us to sample with any parameter $s = \gamma \cdot \lambda_n(\mathcal{L})/2^{o(n/\log n)}$.

Corollary 1.3.15. *There is an algorithm that takes as input a basis for a lattice $\mathcal{L} \subset \mathbb{R}^n$, parameters $u \geq 2$ and $\hat{s} > C\sqrt{\log n} \cdot u^{n/u}\lambda_n(\mathcal{L})$, and a positive integer M, and outputs M vectors that are distributed exactly as M independent samples from $D_{\mathcal{L},\hat{s}}$ in time* $\mathrm{poly}(n) \cdot M + \mathrm{poly}(n) \cdot 2^{O(u)}$.

Proof. The algorithm first runs the procedure from Theorem 1.2.3 to obtain a $u^{n/u}$-reduced basis $\mathbf{B}$ for $\mathcal{L}$. It is easy to see that such a basis satisfies $\|\widetilde{\mathbf{B}}\| \geq u^{n/u}\lambda_n(\mathcal{L})$. We then run the procedure from Theorem 1.3.14 M times and output the result. □

Corollary 1.3.16. *There is an algorithm that takes as input a basis $\mathbf{B}$ for a lattice $\mathcal{L} \subset \mathbb{R}^n$, parameters $u \geq 2$ and $\hat{s} > C\sqrt{n\log n} \cdot u^{n/u}\eta_{1/2}(\mathcal{L})$, and a positive integer M, and outputs M vectors that are distributed exactly as M independent samples from $D_{\mathcal{L},\hat{s}}$ in time* $\mathrm{poly}(n) \cdot M + \mathrm{poly}(n) \cdot 2^{O(u)}$.

Proof. Simply combine the above corollary with Corollary 1.3.8. □

1.4 Miscellany

We will need the following (weak variant of the) effective form of Stirling's approximation due to Robbins [Rob55], from which we derive bounds on the binomial coefficient.

Theorem 1.4.1 ([Rob55])**.** *For any integer $n \geq 1$,*

$$\sqrt{2\pi n}(n/e)^n \leq n! \leq \exp(1/(12n)) \cdot \sqrt{2\pi n}(n/e)^n \ .$$

Corollary 1.4.2. *For any integers $n \geq 1$ and $1 \leq k \leq n/2$,*

$$\frac{1}{\sqrt{2\pi e^{1/6} k}} \cdot (e^{1-k/n} n/k)^k \leq \binom{n}{k} \leq \frac{\exp(k/n)}{\sqrt{2\pi k}} \cdot (en/k)^k \ .$$

Proof. For the upper bound, we have

$$\begin{aligned}
\binom{n}{k} &\leq \exp(1/(12n)) \cdot \sqrt{\frac{n}{2\pi k(n-k)}} \cdot \frac{n^n}{k^k (n-k)^{n-k}} \\
&= \frac{\exp(1/(12n)}{\sqrt{2\pi k}} \cdot (n/k)^k \cdot (1-k/n)^{k-n-1/2} \\
&\leq \frac{\exp(k/(2n) + 1/(12n))}{\sqrt{2\pi k}} \cdot (en/k)^k \\
&\leq \frac{\exp(k/n)}{\sqrt{2\pi k}} \cdot (en/k)^k \ .
\end{aligned}$$

For the lower bound, we have

$$\begin{aligned}
\binom{n}{k} &\geq \exp(-1/(12k) - 1/(12(n-k))) \cdot \sqrt{\frac{n}{2\pi k(n-k)}} \cdot \frac{n^n}{k^k (n-k)^{n-k}} \\
&\geq \frac{\exp(-1/12 - 1/(6n))}{\sqrt{2\pi k}} \cdot (n/k)^k \cdot (1-k/n)^{k-n-1/2} \\
&\geq \frac{\exp(-1/(6n))}{\sqrt{2\pi e^{1/6} k}} \cdot (n/k)^k \cdot \exp(-k/n)^{k-n-1/2} \\
&\geq \frac{1}{\sqrt{2\pi e^{1/6} k}} \cdot (e^{1-k/n} n/k)^k \ .
\end{aligned}$$

□

We will also need the Chernoff-Hoeffding bound [Hoe63].

Lemma 1.4.3 (Chernoff-Hoeffding bound)**.** *Let $X_1, \ldots, X_N$ be independent and identically distributed random variables with $0 \leq X_i \leq 1$ and $\overline{X} := \mathbb{E}[X_i]$. Then, for $s > 0$*

$$\Pr\left[N\overline{X} - \sum X_i \geq s\right] \leq \exp(-s^2/N) \ ,$$

and

$$\Pr\left[\sum X_i - N\overline{X} \geq s\right] \leq \exp(-s^2/N) \ .$$

Chapter 2

A Reverse Minkowski Theorem[1]

2.1 Introduction

Recall that the determinant of a lattice $\mathcal{L} \subset \mathbb{R}^n$ with basis $\mathbf{B}$, $\det(\mathcal{L}) = |\det(\mathbf{B})|$, is a measure of its global density in the sense that

$$\det(\mathcal{L}) = \lim_{r\to\infty} \frac{\mathrm{vol}(rB_2^n)}{|\mathcal{L} \cap rB_2^n|} ,$$

where rB_2^n denotes the closed Euclidean ball of radius $r > 0$, whose volume is equal to $(\pi n)^{-1/2}(2\pi e r^2/n)^{n/2}(1 + o(1))$. Minkowski's celebrated theorem shows that a lattice with small determinant must have short non-zero vectors [Min10]. This is the foundational results in the study of lattices and the geometry of numbers, and it has innumerable applications. We consider the following point-counting form of this theorem due to Blichfeldt and van der

[1]This chapter is primarily based on joint work with Oded Regev, which appeared in the Symposium on the Theory of Computing (STOC), 2017 [RS17b], and some passages have been taken verbatim from this source. This work was supported by the National Science Foundation (NSF) under Grant No. CCF-1320188, and the Defense Advanced Research Projects Agency (DARPA) and Army Research Office (ARO) under Contract No. W911NF-15-C-0236. Part of this work was done while visiting Chris Peikert at the University of Michigan and while interning at IBM.

Corput,[2] which says that a lattice with small determinant must have *many* short points, or informally, that "global density implies local density."

Theorem 2.1.1 ([vdC36])**.** *For any lattice* $\mathcal{L} \subset \mathbb{R}^n$ *with* $\det(\mathcal{L}) \leq 1$ *and* $r > 0$*,*

$$|\mathcal{L} \cap rB_2^n| \geq 2^{-n} \cdot \mathrm{vol}(rB_2^n) = \frac{1}{\sqrt{\pi n}} \left(\frac{\pi e r^2}{2n}\right)^{n/2} (1 + o(1)) \,.$$

It is quite natural to ask whether a converse of Theorem 2.1.1 holds. In particular, if a lattice has sufficiently many short points, does it necessarily have small determinant? Does local density imply global density?

It is easy to see that the answer is actually no. Consider, for example, the lattice generated by the vectors $(1/t, 0)$ and $(0, t^2)$ for some arbitrarily large t. This lattice has at least $2\lfloor tr \rfloor + 1$ points of norm at most r, but it has arbitrarily large determinant t. Notice, however, that this lattice contains a *sublattice* generated by $(1/t, 0)$ that does have small determinant. This leads us to a more refined question:

> If a lattice has sufficiently many short points, does it necessarily have a small-determinant *sublattice*? Does local density imply global density over a subspace?

Equivalently, in the contrapositive, the question asks for an upper bound on the number of lattice points in a ball given that there is no sublattice of small determinant.

Dadush conjectured a suitably precise answer to these questions [Dad12a]. Dadash and Regev studied this conjecture in detail [DR16]. They showed a wide range of applications (from computational complexity of lattice problems to Brownian motion on flat tori) and gave some evidence for it. We refer the reader to [DR16] for a full list of their results.

Our main result is a proof of the conjecture of Dadush, which in particular implies the applications mentioned above.

[2] They actually showed the slightly stronger bound $|\mathcal{L} \cap rB_2^n| \geq 2\lfloor 2^{-n} \cdot \mathrm{vol}(rB_2^n) \rfloor + 1$ and considered arbitrary norms, not just ℓ_2. (See, e.g., [GL87, Thm. 1 of Ch. 2, Sec. 7].)

Theorem 2.1.2 (Reverse Minkowski Theorem, [RS17b])**.** *For any lattice* $\mathcal{L} \subset \mathbb{R}^n$ *with* $\det(\mathcal{L}') \geq 1$ *for all sublattices* $\mathcal{L}' \subseteq \mathcal{L}$*,*

$$\rho_{1/t}(\mathcal{L}) \leq \frac{3}{2} ,$$

where $t := 10(\log n + 2)$*. In other words, for any lattice* $\mathcal{L} \subset \mathbb{R}^n$

$$3\eta_{\det}(\mathcal{L})/2 \leq \eta_{1/2}(\mathcal{L}) \leq t\eta_{\det}(\mathcal{L}) ,$$

where

$$\eta_{\det}(\mathcal{L}) := \max_{\mathcal{M} \subseteq \mathcal{L}^*} \det(\mathcal{M})^{-1/\mathrm{rank}(\mathcal{M})} .$$

In Section 2.5, we extend Theorem 2.1.2 to obtain a bound on the Gaussian mass for all parameters, as follows.

Theorem 2.1.3 ([RS17b])**.** *For any lattice* $\mathcal{L} \subset \mathbb{R}^n$ *with* $\det(\mathcal{L}') \geq 1$ *for all sublattices* $\mathcal{L}' \subseteq \mathcal{L}$*,*

1. $\rho_s(\mathcal{L}) \leq 1 + e^{-\pi(1/s^2 - t^2)}/2$ *for any* $s \leq 1/t$*;*
2. $\rho_s(\mathcal{L}) \leq (Cst)^{n/2}$ *for any* $1/t < s < t$ *and some universal constant* $C > 1$*; and*
3. $\rho_s(\mathcal{L}) \leq 2s^n$ *for any* $s \geq t$*,*

where $t := 10(\log n + 2)$*.*

Theorem 2.1.3 implies the following point-counting bounds. (See Section 2.5 for the proof.)

Corollary 2.1.4. *For every lattice* $\mathcal{L} \subset \mathbb{R}^n$ *with* $\det(\mathcal{L}') \geq 1$ *for all sublattices* $\mathcal{L}' \subseteq \mathcal{L}$*, and every shift vector* $\boldsymbol{u} \in \mathbb{R}^n$*,*

1. *for any* $r \geq 1$*,* $|\mathcal{L} \cap (rB_2^n + \boldsymbol{u})| \leq 3e^{\pi t^2 r^2}/2$*;*

2. *for any* $\sqrt{n/(2\pi)} \cdot t^{-1} \le r \le \sqrt{n/(2\pi)} \cdot t$, $|\mathcal{L} \cap (rB_2^n + \boldsymbol{u})| \le (Ctr/\sqrt{n})^{n/2}$ *for some universal constant* $C > 0$*; and*

3. *for any* $r \ge \sqrt{n/(2\pi)} \cdot t$, $|\mathcal{L} \cap (rB_2^n + \boldsymbol{u})| \le 2(2\pi e r^2/n)^{n/2}$,

where $t := 10(\log n + 2)$.

In Section 2.8, we discuss the tightness of Theorem 2.1.3 and Corollary 2.1.4.

2.1.1 Approximation to the covering radius

Notice that the covering radius $\mu(\mathcal{L})$ of a lattice $\mathcal{L}$ must be at least the radius of a ball of volume $\det(\mathcal{L})$, which is at least $\sqrt{n/(2\pi e)}\det(\mathcal{L})^{1/n}$. By considering projections, Kannan and Lovász [KL88] improved this lower bound, as follows. Let $\pi_{W^\perp}(\mathcal{L})$ be the projection of the lattice onto the space $W^\perp$ orthogonal to some *lattice subspace* $W \subset \mathbb{R}^n$—a subspace spanned by $k < n$ linearly independent lattice vectors.[3] Then clearly $\mu(\mathcal{L}) \ge \mu(\pi_{W^\perp}(\mathcal{L}))$, and the latter is at least $(\dim(W^\perp)/(2\pi e))^{1/2} \cdot \det(\pi_{W^\perp}(\mathcal{L}))^{1/\dim(W^\perp)}$. So, we obtain the lower bound

$$\mu(\mathcal{L}) \ge \frac{1}{\sqrt{2\pi e}} \cdot \mu_{\det}(\mathcal{L}) \ ,$$

where

$$\begin{aligned} \mu_{\det}(\mathcal{L}) &:= \max_{W \subset \mathbb{R}^n} \sqrt{\dim(W^\perp)} \cdot \det(\pi_{W^\perp}(\mathcal{L}))^{\frac{1}{\dim(W^\perp)}} \\ &= \max_{\mathcal{M} \subseteq \mathcal{L}^*} \sqrt{\operatorname{rank}(\mathcal{M})} \cdot \det(\mathcal{M})^{-\frac{1}{\operatorname{rank}(\mathcal{M})}} \ , \end{aligned}$$

with the first maximum taken over lattice subspaces $W \subset \mathbb{R}^n$. Kannan and Lovász also observed the upper bound

$$\mu(\mathcal{L}) \le C\sqrt{n} \cdot \mu_{\det}(\mathcal{L})$$

[3] The projection $\pi_{W^\perp}(\mathcal{L})$ is a lattice if and only if W is a lattice subspace.

(see [DR16, Theorem 11.1] for a proof), and asked whether a better upper bound could be found.[4] In Section 2.6, we use Theorem 2.1.2 to derive the following improved bound.

Theorem 2.1.5 (Covering-radius approximation, [RS17b])**.** *For any lattice* $\mathcal{L} \subset \mathbb{R}^n$*,*

$$\frac{1}{\sqrt{2\pi e}} \cdot \mu_{\det}(\mathcal{L}) \leq \mu(\mathcal{L}) \leq 10(\log n + 10)^{3/2} \cdot \mu_{\det}(\mathcal{L}) \; . \tag{2.1}$$

We emphasize that Dadush and Regev [DR16] already proved that Theorem 2.1.5 (with slightly weaker parameters) would follow from a proof of Theorem 2.1.2. Although our proof is shorter and achieves slightly better parameters, it is conceptually similar to the one in [DR16].

We note that the specific polylogarithmic factor that we obtain is likely not optimal. In fact, in Theorem 2.6.7 we prove a bound similar to that in Eq. (2.1) that replaces the factor $10(\log n + 10)^{3/2}$ by $C\sqrt{\log n}$, assuming the celebrated Slicing Conjecture [Bou91, Kla06]. However, it is not difficult to show that this factor cannot be smaller than $\sqrt{\log n/(4e)} + o(1)$.[5]

Covering radius of stable lattices and Minkowski's Conjecture. We say that a lattice $\mathcal{L} \subset \mathbb{R}^n$ is *stable* if $\det(\mathcal{L}) = 1$ and $\det(\mathcal{L}') \geq 1$ for all sublattices $\mathcal{L}' \subseteq \mathcal{L}$. Stable lattices arise in a number of contexts [HN75, Stu76, Gra84] and they play an important role in the rest of this chapter. Shapira and Weiss showed that a tight bound of $\mu(\mathcal{L}) \leq \mu(\mathbb{Z}^n) = \sqrt{n}/2$ on the covering radius of stable lattices would imply a well-known conjecture attributed to Minkowski [SW16]. (See also [Sol16].) We do not manage to prove such a tight bound, but en route to proving Theorem 2.1.5 we do show that $\mu(\mathcal{L}) \leq 4\sqrt{n}(\log n + 10)$ for all stable

[4] They also proved similar bounds for arbitrary norms [KL88, Corollary 3.11].

[5] Consider the lattice $\mathcal{L}$ generated by $(\boldsymbol{e}_1, \boldsymbol{e}_2/2, 2\boldsymbol{e}_3/3^{3/2}, \ldots, (n-1)^{(n-1)/2}\boldsymbol{e}_n/n^{n/2})$. It is not difficult to verify that $\mu_{\det}(\mathcal{L}) = 1$, but

$$\mu(\mathcal{L})^2 = 1/4 + \sum_{k=2}^{n} \frac{(k-1)^{k-1}}{4k^k} = 1/4 + \sum_{k=2}^{n} \frac{(1-1/k)^k}{4(k-1)} = \sum_{k=2}^{n} \frac{1}{4e(k-1)} + O(1) = \frac{\log n}{4e} + O(1) \; .$$

Therefore, $\mu(\mathcal{L}) = \sqrt{\log n/(4e)} + o(1)$.

lattices. (See Theorem 2.6.1.) We also observe that a very strong resolution to the Slicing Conjecture and a better bound between two lattice parameters would yield the desired tight bound. (See Theorem 2.6.6 and the discussion afterwards.)

2.1.2 An optimal bound on the Gaussian mass for "extreme" parameters

It is natural to ask whether $\rho_s(\mathcal{L}) \leq \rho_s(\mathbb{Z}^n)$ for any lattice $\mathcal{L} \subset \mathbb{R}^n$ such that $\det(\mathcal{L}') \geq 1$ for all sublattices $\mathcal{L}' \subseteq \mathcal{L}$ and any parameter $s > 0$. This would be a strict strengthening of Theorem 2.1.3 and would obviously be the strongest possible bound with this form. The next theorem shows that indeed $\rho_s(\mathcal{L}) \leq \rho_s(\mathbb{Z}^n)$ for such lattices, but only for "extremely low" or "extremely high" parameters s. (See Section 2.7 for the proof.)

Theorem 2.1.6 ([RS17b])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ such that $\det(\mathcal{L}') \geq 1$ for all sublattices $\mathcal{L}' \subseteq \mathcal{L}$ and parameter $s > 0$ such that either $s \leq \sqrt{2\pi/(n+2)}$ or $s \geq \sqrt{(n+2)/(2\pi)}$, we have $\rho_s(\mathcal{L}) \leq \rho_s(\mathbb{Z}^n)$.*

We hope that the proof of Theorem 2.1.6 might provide some hints as to how to extend it to all parameters s. We also note that Theorem 2.1.6 implies the full result in low dimensions, as follows.

Corollary 2.1.7. *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ with $n \leq 4$ such that $\det(\mathcal{L}') \geq 1$ for all sublattices $\mathcal{L}' \subseteq \mathcal{L}$, $\rho_s(\mathcal{L}) \leq \rho_s(\mathbb{Z}^n)$ for and $s > 0$.*

2.1.3 Proof overview

In this section, we give a high-level overview of the proof of Theorem 2.1.2.

Bounding the mass of stable lattices. Recall that a lattice $\mathcal{L}$ is *stable* if $\det(\mathcal{L}) = 1$ and $\det(\mathcal{L}') \geq 1$ for all sublattices $\mathcal{L}' \subseteq \mathcal{L}$. I.e., stable lattices are determinant-one lattices

that satisfy the assumption in Theorem 2.1.2. In this proof overview, we focus on bounding the Gaussian mass $\rho_s(\mathcal{L})$ of stable lattices $\mathcal{L}$. The general case follows immediately from such a bound.

Crucially, the stable lattices form a compact subset of the set of determinant-one lattices, so that the continuous function $\rho_s(\mathcal{L})$ must attain a global maximum over the set of stable lattices. We may therefore restrict our attention to a lattice that corresponds to this global maximum. If this lattice is on the *boundary* of the set of stable lattices, then it has a strict sublattice $\mathcal{L}'$ with determinant one. We can then "split the lattice" at $\mathcal{L}'$. Namely, we can replace the original lattice $\mathcal{L}$ by the direct sum $\mathcal{L}' \oplus \mathcal{L}/\mathcal{L}'$. It is not difficult to prove that

$$\rho_s(\mathcal{L}) \leq \rho_s(\mathcal{L}' \oplus \mathcal{L}/\mathcal{L}') = \rho_s(\mathcal{L}')\rho_s(\mathcal{L}/\mathcal{L}')$$

and that $\mathcal{L}'$ and $\mathcal{L}/\mathcal{L}'$ are stable. So, we have reduced the question to a lower-dimensional one. Therefore, if we could show that for any dimension, the global maximizer is on the boundary, then we could use induction to show that the global maximizer of the Gaussian mass is simply the integer lattice $\mathbb{Z}^n = \mathbb{Z} \oplus \cdots \oplus \mathbb{Z}$.

Indeed, this is how we prove Theorem 2.1.6 (in Section 2.7), which shows that $\mathbb{Z}^n$ has maximal Gaussian mass for certain "extreme" parameters s. For such parameters, by taking the second derivative, we show that a stable lattice cannot be a local maximum over the set of determinant-one lattices. Therefore, the global maximizer of $\rho_s(\mathcal{L})$ over the compact subset of stable lattices must be on the boundary, and we can perform the "splitting" procedure described above to show by induction that $\rho_s(\mathcal{L}) \leq \rho_s(\mathbb{Z}^n)$.

However, we do not know if $\rho_s(\mathcal{L})$ can have such stable local maxima for other parameters. As a potential way around this issue, we could use a natural and very elegant idea due to Shapira and Weiss [SW16]—We could try to directly bound the value of $\rho_s(\mathcal{L})$ at any hypothetical local maximum. Then, either the global maximum of $\rho_s(\mathcal{L})$ over the set of stable

lattices is one of these local maxima, in which case we can apply this bound; or it is on the boundary, in which case we can "split the lattice" as above. (Shapira and Weiss suggested using this approach to bound the covering radius of stable lattices to resolve Minkowski's Conjecture [SW16]. Interestingly, local maxima of the covering radius do exist [DSV12].)

Enter the Voronoi cell. Unfortunately, even bounding the value of $\rho_s(\mathcal{L})$ at local maxima seems to be beyond our grasp. So, instead of working with $\rho_s(\mathcal{L})$ directly, we work with a proxy for it: the Gaussian mass of the Voronoi cell of the lattice

$$\gamma_s(\mathcal{V}(\mathcal{L})) := \int_{\mathcal{V}(\mathcal{L})/s} e^{-\pi\|\boldsymbol{x}\|^2} \mathrm{d}\boldsymbol{x} \ ,$$

where the Voronoi cell is the set of all points that are closer to the origin than to any other lattice vector

$$\mathcal{V}(\mathcal{L}) := \{\boldsymbol{x} \in \mathbb{R}^n \ : \ \forall \boldsymbol{y} \in \mathcal{L}, \ \|\boldsymbol{x}\| \leq \|\boldsymbol{y} - \boldsymbol{x}\|\} \ .$$

An elegant proof due to Chung, Dadush, Liu, and Peikert [CDLP13] shows that $\rho_s(\mathcal{L})$ is at most $1/\gamma_s(\mathcal{V}(\mathcal{L}))$. (See Lemma 2.4.1.) So, in order to prove an upper bound on $\rho_s(\mathcal{L})$, it suffices to prove a lower bound on $\gamma_s(\mathcal{V}(\mathcal{L}))$.

We accomplish this via the approach described above. I.e., we reduce the problem to bounding the value of $\gamma_s(\mathcal{V}(\mathcal{L}))$ at local minima. (Here too, we do not know whether these local minima exist.) By comparing gradients, we then show (in Section 2.3) that any lattice corresponding to a local minimum must have a Voronoi cell $\mathcal{V}(\mathcal{L})$ such that the function $A \mapsto \gamma_s(A\mathcal{V}(\mathcal{L}))$ has a critical point at $A = I_n$, where $A \in \mathrm{SL}_n(\mathbb{R})$ ranges over all determinant-one matrices. Using a result due to Bobkov [Bob11], which itself follows from a deep theorem due to Cordero-Erausquin, Fradelizi, and Maurey [CFM04],[6] we can show that any such

[6]We note in passing that one can prove Theorem 2.1.2 (at least up to constants) without using this rather heavy hammer by considering local maxima of the *ℓ-norm* of the Voronoi cell instead of local minima of the Gaussian mass of the Voronoi cell. (We still need the $\ell\ell^*$ theorem, though.)

critical point must actually be a global *maximum* of the function $A \mapsto \gamma_s(A\mathcal{V}(\mathcal{L}))$. I.e., in the language of convex geometry, the Voronoi cell is in a position that maximizes the Gaussian mass. (Note the rather surprising jump from a presumed local minimum over the set of determinant-one lattices to a global *maximum* over the set of positions of the Voronoi cell.)

Finally, we complete the proof by applying the celebrated $\ell\ell^*$ theorem [FT79, Lew79, Pis82]. In particular, this theorem tells us that every convex body K with $\mathrm{vol}(K) = 1$ has a position $A \in \mathrm{SL}_n(\mathbb{R})$ such that $\gamma_{1/t}(AK) \geq 2/3$, with $t := 10(\log n + 2)$ as in Theorem 2.1.2. (See Theorem 2.4.6.) Since the Voronoi cell is already in a position that maximizes the mass, we must have $\gamma_{1/t}(\mathcal{V}(\mathcal{L})) \geq \gamma_{1/t}(A\mathcal{V}(\mathcal{L})) \geq 2/3$. We then obtain the desired bound on $\rho_{1/t}(\mathcal{L})$ by applying the result of [CDLP13].

2.1.4 Related work

Our main theorem was originally conjectured by Dadush [Dad12a], and Dadush and Regev described several applications of the conjecture [DR16]. In particular, they showed the connection between this conjecture and the Kannan-Lovász-style covering-radius approximation given in Theorem 2.1.5. They also used a result from convex geometry (specifically the Milman-Pisier Theorem) as evidence for the conjecture. That theorem is related to the $\ell\ell^*$ theorem that we use in our proof.

The high-level outline of our proof (in which we obtain a bound on a lattice parameter by reducing the question to stable local extrema) is due to Shapira and Weiss [SW16]. They showed that an important conjecture attributed to Minkowski would follow if we could prove that $\mathbb{Z}^n$ has maximal covering radius amongst all stable lattices (i.e., that the covering radius of an n-dimensional stable lattice is at most $\sqrt{n}/2$). They then observed that it would suffice to bound the covering radius of the lattices corresponding to local maxima of the covering radius function over the set of determinant-one lattices.

Stable lattices were introduced (in a more general context) by Harder and Narasimhan [HN75] and by Stuhler [Stu76]. Our presentation more-or-less follows that of Grayson [Gra84].

2.1.5 Directions for future work

The most obvious direction for future work is to try to obtain a better value for t in Theorem 2.1.2. As far as we know, the correct value could be as small as $t = \eta_{1/2}(\mathbb{Z}^n) = \sqrt{\log(n)/\pi} + o(1)$. Our proof seems to be loose in two places: (1) Theorem 2.4.6, which bounds the maximal Gaussian mass of convex bodies; and (2) the induction argument in the proof of Proposition 2.4.10. So, perhaps a different proof technique (such as the one discussed in the next paragraph) would be preferable for this.

A more ambitious goal would be to prove that $\mathbb{Z}^n$ is the exact maximizer of $\rho_s(\mathcal{L})$ for all parameters $s > 0$ over the set of stable lattices $\mathcal{L}$. One might try to prove this by showing that $\rho_s(\mathcal{L})$ has no local maxima over the set of determinant-one (stable) lattices for any parameter $s > 0$. Alternatively, one can try using the technique of "characterizing the local extrema" that we use to prove Theorem 2.1.2. For this, we note that any local maximum of $\rho_s(\mathcal{L})$ must correspond to an "isotropic" lattice $\mathcal{L}$ in the sense that

$$\sum_{\boldsymbol{y} \in \mathcal{L}} \rho_s(\boldsymbol{y}) \boldsymbol{y}\boldsymbol{y}^T = \alpha \cdot I_n$$

for some scalar $\alpha > 0$. So, it would suffice to show that $\rho_s(\mathcal{L}) \leq \rho_s(\mathbb{Z}^n)$ for (stable) "isotropic" lattices. Unfortunately, we do not know how to make use of this.

Theorem 2.1.2 gives quite a tight approximation to the smoothing parameter $\eta_{1/2}(\mathcal{L})$. However, an analogous tightness result does not hold for Theorem 2.1.3 and Corollary 2.1.4. Dadush and Regev therefore suggested a potential refinement that depends on "the full spectrum of dense sublattices," $\min_{\mathcal{L}' \subseteq \mathcal{L},\ \mathrm{rank}(\mathcal{L}')=k} \det(\mathcal{L}')^{1/k}$ for $k = 1, \ldots, n$, rather than just $\min_{\mathcal{L}' \subseteq \mathcal{L}} \det(\mathcal{L}')^{1/\mathrm{rank}(\mathcal{L}')}$ [DR16, Section 9]. This could potentially give a tight characterization

of $|\mathcal{L} \cap rB_2^n|$ for all radii r and all lattices $\mathcal{L} \subset \mathbb{R}^n$.

One can also consider generalizations of Theorems 2.1.5 and 2.1.2 to arbitrary norms, as discussed in [KL88] and [DR16, Section 9] respectively. Extending Theorem 2.1.5 to arbitrary norms could potentially yield faster algorithms for Integer Programming [Dad12b]. Unfortunately, a natural generalization of Theorem 2.1.2 actually fails. (See [DR16, Section 9].)

2.2 Preliminaries

A *convex body* $K \subset \mathbb{R}^n$ is a convex compact subset of $\mathbb{R}^n$ with non-empty interior. It is *symmetric* if $-K = K$. A *position* of a convex body is simply AK for a determinant-one matrix A.

2.2.1 Stable lattices

We say that a lattice $\mathcal{L} \subset \mathbb{R}^n$ is *stable* if $\det(\mathcal{L}) = 1$ and $\det(\mathcal{L}') \geq 1$ for all sublattices $\mathcal{L}' \subseteq \mathcal{L}$. (Some authors call such lattices "semistable.") Note the obvious relationship between this notion and Theorem 2.1.2. Here, we describe the properties of stable lattices that we will need in the sequel, and include proofs for completeness. This theory was developed by [HN75, Stu76, Gra84]. See, e.g., [Gra84, Cas04] for a more thorough treatment.

We can in some sense "decompose" any lattice into stable lattices. To see this, we consider the two-dimensional scatter plot with points

$$\{(\operatorname{rank}(\mathcal{L}'),\ \log\det(\mathcal{L}'))\ :\ \mathcal{L}' \subseteq \mathcal{L}\}\ ,$$

for some lattice $\mathcal{L} \subset \mathbb{R}^n$, where we explicitly include the trivial sublattice $\{\mathbf{0}\}$ and define $\log\det(\{\mathbf{0}\}) := 0$. We call this the *canonical plot* of $\mathcal{L}$. Note that these points are bounded

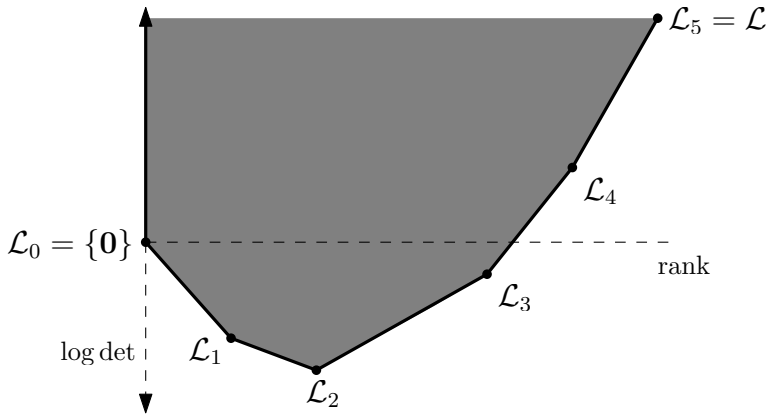


Figure 2.1: The canonical polygon of a (hypothetical) lattice $\mathcal{L}$.

from below and that the minimum log det for each fixed rank is achieved. The convex hull of these points is therefore a degenerate polygon (bounded from below, but unbounded from above), called the *canonical polygon* of $\mathcal{L}$. See Figure 2.1.

We are interested in the vertices of this polygon (i.e., the extremal points), which correspond to certain primitive sublattices of $\mathcal{L}$ with low determinants. (E.g., $\mathcal{L}_0, \dots, \mathcal{L}_5$ in Figure 2.1.) Each vertex corresponds to a unique sublattice, and a lattice $\mathcal{L}_1$ corresponding to a low-rank extremal point is a sublattice of any lattice $\mathcal{L}_2$ corresponding to a higher-rank extremal point, $\mathcal{L}_1 \subset \mathcal{L}_2$. Therefore, the extremal points define a *canonical filtration* of $\mathcal{L}$,

$$\{\mathbf{0}\} = \mathcal{L}_0 \subset \mathcal{L}_1 \subset \cdots \subset \mathcal{L}_k = \mathcal{L} \,.$$

(E.g., the canonical filtration of $\mathbb{Z}^n$ is trivial: $\{\mathbf{0}\} = \mathcal{L}_0 \subset \mathcal{L}_1 = \mathbb{Z}^n$. Note in particular we only include lattices that correspond to *vertices* in the canonical filtration, not any lattice on the boundary.) All of the quotients $\mathcal{L}_i/\mathcal{L}_{i-1}$ of adjacent sublattices in the canonical filtration are scalings of stable lattices. This is what we mean when we say that we can "decompose" a lattice into a sequence of stable lattices. Following [Gra84, Cas04], we make these facts (and more) precise in Proposition 2.2.2, which lists basic properties of the canonical filtration and

stable lattices. We first need the following lemma, due to Stuhler [Stu76].

Lemma 2.2.1. *For any $\mathcal{L} \subset \mathbb{R}^n$ and any two primitive sublattices $\mathcal{L}_1, \mathcal{L}_2 \subseteq \mathcal{L}$,*

$$\operatorname{rank}(\mathcal{L}_1) + \operatorname{rank}(\mathcal{L}_2) = \operatorname{rank}(\mathcal{L}_1 \cap \mathcal{L}_2) + \operatorname{rank}(\mathcal{L}_1 + \mathcal{L}_2) \ ,$$

and

$$\det(\mathcal{L}_1 \cap \mathcal{L}_2) \det(\mathcal{L}_1 + \mathcal{L}_2) \leq \det(\mathcal{L}_1) \det(\mathcal{L}_2) \ ,$$

where we define $\det(\{\mathbf{0}\}) = 1$.

Proof. The equality of ranks follows by considering the dimensions of the subspaces spanned by $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{L}_1 \cap \mathcal{L}_2$, and $\mathcal{L}_1 + \mathcal{L}_2$. For the inequality, suppose that $\mathcal{M}_1, \mathcal{M}_2 \subseteq \mathcal{M}$ are sublattices such that $\mathcal{M}_1 \cap \mathcal{M}_2 = \{\mathbf{0}\}$ and $\mathcal{M}_1 + \mathcal{M}_2 = \mathcal{M}$. Then, we have

$$\det(\mathcal{M}) = \det(\mathcal{M}_1) \cdot \det(\pi_{\operatorname{span}(\mathcal{M}_1)^\perp}(\mathcal{M}_2)) \leq \det(\mathcal{M}_1) \det(\mathcal{M}_2) \ .$$

Plugging in $\mathcal{M} := (\mathcal{L}_1 + \mathcal{L}_2)/(\mathcal{L}_1 \cap \mathcal{L}_2)$, $\mathcal{M}_1 := \mathcal{L}_1/(\mathcal{L}_1 \cap \mathcal{L}_2)$ and $\mathcal{M}_2 := \mathcal{L}_2/(\mathcal{L}_1 \cap \mathcal{L}_2)$ gives

$$\begin{aligned} \det(\mathcal{L}_1 + \mathcal{L}_2)/\det(\mathcal{L}_1 \cap \mathcal{L}_2) &= \det((\mathcal{L}_1 + \mathcal{L}_2)/(\mathcal{L}_1 \cap \mathcal{L}_2)) \\ &\leq \det(\mathcal{L}_1/(\mathcal{L}_1 \cap \mathcal{L}_2)) \det(\mathcal{L}_2/(\mathcal{L}_1 \cap \mathcal{L}_2)) \\ &= \det(\mathcal{L}_1) \det(\mathcal{L}_2)/\det(\mathcal{L}_1 \cap \mathcal{L}_2)^2 \ . \end{aligned}$$

The result follows by rearranging. □

Proposition 2.2.2. *For any lattice $\mathcal{L} \subset \mathbb{R}^n$, let $\{\mathbf{0}\} = \mathcal{L}_0, \mathcal{L}_1, \ldots, \mathcal{L}_k = \mathcal{L}$ be all sublattices corresponding to vertices of the canonical polytope, ordered by their rank. (See Figure 2.1.) Then,*

1. *the $\mathcal{L}_i$ define a filtration $\mathcal{L}_0 \subset \mathcal{L}_1 \subset \cdots \subset \mathcal{L}_k$;*

2. *the quotient lattice* $\mathcal{L}_i/\mathcal{L}_{i-1}$ *is a scaling of a stable lattice for* $1 \le i \le k$ *(i.e.,* $\det(\mathcal{L}_i/\mathcal{L}_{i-1})^{-1/\mathrm{rank}(\mathcal{L}_i/\mathcal{L}_{i-1})} \cdot \mathcal{L}_i/\mathcal{L}_{i-1}$ *is stable); and*

3. *for all* $1 \le i \le k-1$, $\det(\mathcal{L}_i/\mathcal{L}_{i-1})^{1/\mathrm{rank}(\mathcal{L}_i/\mathcal{L}_{i-1})} < \det(\mathcal{L}_{i+1}/\mathcal{L}_i)^{1/\mathrm{rank}(\mathcal{L}_{i+1}/\mathcal{L}_i)}$.

Furthermore,

(i) the dual of a stable lattice is stable;

(ii) the set of all stable lattices is compact;

(iii) the direct sum of stable lattices is stable; and

(iv) a lattice $\mathcal{L} \subset \mathbb{R}^n$ *is on the boundary of the set of stable lattices if and only if* $\mathcal{L}$ *is stable and there is a primitive sublattice* $\mathcal{L}' \subset \mathcal{L}$ *with* $0 < \mathrm{rank}(\mathcal{L}') < n$ *such that* $\mathcal{L}'$ *and* $\mathcal{L}/\mathcal{L}'$ *are both stable.*

Proof. To prove Item 1, we first note that for any two indices $i \le j$, we can interpret Lemma 2.2.1 in terms of the canonical plot as follows. Consider the parallelogram with the three vertices $(\mathrm{rank}(\mathcal{L}_i), \log\det(\mathcal{L}_i))$, $(\mathrm{rank}(\mathcal{L}_i + \mathcal{L}_j), \log\det(\mathcal{L}_i + \mathcal{L}_j))$, and $(\mathrm{rank}(\mathcal{L}_i \cap \mathcal{L}_j), \log\det(\mathcal{L}_i \cap \mathcal{L}_j))$. Lemma 2.2.1 tells us that the point $(\mathrm{rank}(\mathcal{L}_j), \log\det(\mathcal{L}_j))$ lies on or above the fourth point in this parallelogram. This contradicts the assumption that $\mathcal{L}_i$ and $\mathcal{L}_j$ are extremal points of the convex hull of the canonical plot *unless* the parallelogram is degenerate—i.e., unless $\mathcal{L}_i + \mathcal{L}_j = \mathcal{L}_j$ or $\mathcal{L}_i \cap \mathcal{L}_j = \mathcal{L}_i$. This happens if and only if $\mathcal{L}_i \subseteq \mathcal{L}_j$, as needed.

To prove Item 2, let $\mathcal{L}' \subseteq \mathcal{L}_i/\mathcal{L}_{i-1}$. Let $\widehat{\mathcal{L}} \subseteq \mathcal{L}_i$ be a "lift" of $\mathcal{L}'$ so that $\mathcal{L}_{i-1} \subseteq \widehat{\mathcal{L}}$ and $\mathcal{L}' = \widehat{\mathcal{L}}/\mathcal{L}_{i-1}$. Since $\mathcal{L}_{i-1}$ and $\mathcal{L}_i$ are vertices of the canonical polygon, the point $(\mathrm{rank}(\widehat{\mathcal{L}}), \log\det(\widehat{\mathcal{L}}))$ must lie on or above the line between $(\mathrm{rank}(\mathcal{L}_{i-1}), \log\det(\mathcal{L}_{i-1}))$ and

$(\mathrm{rank}(\mathcal{L}_i), \log\det(\mathcal{L}_i))$. Therefore,

$$\begin{aligned}\det(\mathcal{L}') &= \det(\widehat{\mathcal{L}})/\det(\mathcal{L}_{i-1}) \\ &\geq \Big(\frac{\det(\mathcal{L}_i)}{\det(\mathcal{L}_{i-1})}\Big)^{\frac{\mathrm{rank}(\widehat{\mathcal{L}})-\mathrm{rank}(\mathcal{L}_{i-1})}{\mathrm{rank}(\mathcal{L}_i)-\mathrm{rank}(\mathcal{L}_{i-1})}} \\ &= \det(\mathcal{L}_i/\mathcal{L}_{i-1})^{\frac{\mathrm{rank}(\mathcal{L}')}{\mathrm{rank}(\mathcal{L}_i/\mathcal{L}_{i-1})}} \ .\end{aligned}$$

I.e., if we set $\alpha_i := \det(\mathcal{L}_i/\mathcal{L}_{i-1})^{-1/\mathrm{rank}(\mathcal{L}_i/\mathcal{L}_{i-1})}$, then $\det(\alpha_i\mathcal{L}') \geq 1$. It follows that $\alpha_i\mathcal{L}_i/\mathcal{L}_{i-1}$ is stable, as claimed.

Item 3 simply says that the slopes of the lines between vertices on the canonical polytope are strictly increasing. This is essentially just the definition of a vertex. (See Figure 2.1.)

To prove Item (i), let $\mathcal{M} \subset \mathbb{R}^n$ be a stable lattice and let $\mathcal{M}' \subseteq \mathcal{M}^*$ be a primitive sublattice of the dual. We have

$$\det(\mathcal{M}') = \frac{1}{\det(\mathcal{M}^*/\mathcal{M}')} = \det((\mathcal{M}^*/\mathcal{M}')^*) = \det(\mathcal{M} \cap \mathrm{span}(\mathcal{M}')) \geq 1 \ .$$

Therefore, $\mathcal{M}^*$ is stable.

To prove Item (ii), it suffices to find a bounded set in $\mathbb{R}^{n\times n}$ that contains a basis for every stable lattice. Indeed, for any stable lattice $\mathcal{M} \subset \mathbb{R}^n$, by Item (i), we know that its dual $\mathcal{M}^*$ is also stable. Therefore, $\lambda_1(\mathcal{M}^*) \geq 1$. It then follows from [LLS90] that there exists a basis $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ of $\mathcal{M}$ with $1 \leq \|\boldsymbol{b}_i\| \leq Cn^{2.5}$ for all i, as needed.

To prove Item (iii), let $\mathcal{M}_1, \mathcal{M}_2$ be two stable lattices, and let $\mathcal{M}' \subset \mathcal{M}_1 \oplus \mathcal{M}_2$ be a sublattice. Then, applying Lemma 2.2.1, we have

$$\det(\mathcal{M}') \geq \frac{\det(\mathcal{M}' \cap \mathcal{M}_1)\det(\mathcal{M}' + \mathcal{M}_1)}{\det(\mathcal{M}_1)} = \det(\mathcal{M}' \cap \mathcal{M}_1)\det(\mathcal{M}' + \mathcal{M}_1) \ .$$

Note that $\mathcal{M}' \cap \mathcal{M}_1$ is a sublattice of $\mathcal{M}_1$, so that $\det(\mathcal{M}' \cap \mathcal{M}_1) \geq 1$. And $\mathcal{M}' + \mathcal{M}_1 =$

$\mathcal{M}_1 \oplus \pi_{\mathrm{span}(\mathcal{M}_2)}(\mathcal{M}')$ is the direct sum of $\mathcal{M}_1$ with a sublattice of $\mathcal{M}_2$, so that $\det(\mathcal{M}' + \mathcal{M}_1) = \det(\pi_{\mathrm{span}(\mathcal{M}_2)}(\mathcal{M}')) \geq 1$ as well. The result follows.

Finally, Item (iv) follows by first noting that a stable lattice $\mathcal{M}$ is on the boundary if and only if there is some strict primitive non-zero sublattice $\mathcal{M}' \subset \mathcal{M}$ with $\det(\mathcal{M}') = 1$. Clearly, $\mathcal{M}'$ is stable, since it has determinant one and all of its sublattices are also sublattices of $\mathcal{M}$, so that they must have determinant at least one. The proof that $\mathcal{M}/\mathcal{M}'$ is stable is essentially identical to the proof of Item 2. □

2.2.2 The Voronoi cell and fundamental bodies

The *Voronoi cell* of a lattice $\mathcal{L} \subset \mathbb{R}^n$,

$$\mathcal{V}(\mathcal{L}) := \{\boldsymbol{x} \in \mathbb{R}^n \ : \ \langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq \|\boldsymbol{y}\|^2/2, \ \forall \boldsymbol{y} \in \mathcal{L}\} \ ,$$

is the set of vectors in $\mathbb{R}^n$ that are closer to $\mathbf{0}$ than to any other lattice vector. In fact, it is a symmetric polytope.

A *fundamental body* of a lattice $\mathcal{L} \subset \mathbb{R}^n$ is any convex body $K \subset \mathbb{R}^n$ such that $K + \mathcal{L}$ is a tiling of space. Equivalently, $\mathrm{vol}(K) = \det(\mathcal{L})$ and $\mathrm{Int}(K) \cap (K + \boldsymbol{y}) = \emptyset$ for any non-zero lattice point $\boldsymbol{y} \in \mathcal{L} \setminus \{\mathbf{0}\}$. In particular, the Voronoi cell is a fundamental body.

Claim 2.2.3. *For any lattice $\mathcal{L} \subset \mathbb{R}^n$, primitive sublattice $\mathcal{L}' \subset \mathbb{R}^n$, fundamental body K_1 of $\mathcal{L}'$, and fundamental body K_2 of $\mathcal{L}/\mathcal{L}'$, $K := K_1 \times K_2$ is a fundamental body of $\mathcal{L}$. In particular, if $\{\mathbf{0}\} = \mathcal{L}_0 \subset \mathcal{L}_1 \subset \cdots \subset \mathcal{L}_k$ is a filtration of primitive sublattices, then*

$$\mathcal{V}\Big(\bigoplus_i \mathcal{L}_i/\mathcal{L}_{i-1}\Big) = \mathcal{V}(\mathcal{L}_1/\mathcal{L}_0) \times \cdots \times \mathcal{V}(\mathcal{L}_k/\mathcal{L}_{k-1})$$

is a fundamental body of $\mathcal{L}$.

Proof. Notice that

$$\mathrm{vol}(K) = \mathrm{vol}(K_1) \cdot \mathrm{vol}(K_2) = \det(\mathcal{L}') \cdot \det(\mathcal{L}/\mathcal{L}') = \det(\mathcal{L}) \ .$$

It therefore suffices to show that $\mathrm{Int}(K) \cap (K + \boldsymbol{y}) = \emptyset$ for any $\boldsymbol{y} \in \mathcal{L} \setminus \{\boldsymbol{0}\}$. So suppose $\boldsymbol{y} \in \mathcal{L}$ such that $\mathrm{Int}(K) \cap (K + \boldsymbol{y}) \neq \emptyset$. Then, by projecting orthogonally to $\mathcal{L}'$, we see that $\mathrm{Int}(K_2) \cap (K_2 + \pi_{\mathrm{span}(\mathcal{L}')^\perp}(\boldsymbol{y})) \neq \emptyset$. Since K_2 is a fundamental body of $\mathcal{L}/\mathcal{L}'$ and $\pi_{\mathrm{span}(\mathcal{L}')^\perp}(\boldsymbol{y}) \in \mathcal{L}/\mathcal{L}'$, it follows that $\pi_{\mathrm{span}(\mathcal{L}')^\perp}(\boldsymbol{y}) = \boldsymbol{0}$, i.e., $\boldsymbol{y} \in \mathcal{L}'$. Intersecting with $\mathrm{span}(\mathcal{L}')$, this implies that $\mathrm{Int}(K_1) \cap (K_1 + \boldsymbol{y}) \neq \emptyset$. Since $\boldsymbol{y} \in \mathcal{L}'$ and K_1 is a fundamental body of $\mathcal{L}'$, we obtain that $\boldsymbol{y} = \boldsymbol{0}$. The result follows. □

The next lemma and its corollary show that the Voronoi cell is in some sense the "optimal fundamental body." They are very similar to some results due to Dadush [Dad12b, Lemma 6.3.6, Corollary 6.3.7].

Lemma 2.2.4. *For any lattice $\mathcal{L} \subset \mathbb{R}^n$, there is a map $\psi_{\mathcal{L}} : \mathbb{R}^n \to \mathcal{V}(\mathcal{L})$ such that $\|\psi_{\mathcal{L}}(\boldsymbol{x})\| \leq \|\boldsymbol{x}\|$, and for every fundamental body K of $\mathcal{L}$, $\psi_{\mathcal{L}}$ restricted to $\mathrm{Int}(K)$ is injective and volume-preserving.*

Proof. The function $\psi_{\mathcal{L}}$ just maps $\boldsymbol{x}$ to the unique representative of $\boldsymbol{x} \bmod \mathcal{L}$ that is in the Voronoi cell. Specifically, let $\mathsf{CVP}_{\mathcal{L}}(\boldsymbol{x}) := \arg\min_{\boldsymbol{y} \in \mathcal{L}} \|\boldsymbol{y} - \boldsymbol{x}\|$ be the closest lattice vector to $\boldsymbol{x}$, and let $\psi_{\mathcal{L}}(\boldsymbol{x}) := \boldsymbol{x} - \mathsf{CVP}_{\mathcal{L}}(\boldsymbol{x})$. By the definition of CVP, it is immediate that $\|\psi_{\mathcal{L}}(\boldsymbol{x})\| = \min_{\boldsymbol{y} \in \mathcal{L}} \|\boldsymbol{y} - \boldsymbol{x}\| \leq \|\boldsymbol{x}\|$.

Suppose $\psi_{\mathcal{L}}(\boldsymbol{x}) = \psi_{\mathcal{L}}(\boldsymbol{x}')$ for some $\boldsymbol{x}, \boldsymbol{x}' \in \mathrm{Int}(K)$. I.e., $\boldsymbol{x} - \mathsf{CVP}_{\mathcal{L}}(\boldsymbol{x}) = \boldsymbol{x}' - \mathsf{CVP}_{\mathcal{L}}(\boldsymbol{x}')$. Rearranging, we see that $\boldsymbol{y} := \boldsymbol{x} - \boldsymbol{x}' = \mathsf{CVP}_{\mathcal{L}}(\boldsymbol{x}) - \mathsf{CVP}_{\mathcal{L}}(\boldsymbol{x}')$ is a lattice point. But, $\boldsymbol{x} \in \mathrm{Int}(K) \cap (K + \boldsymbol{y})$. Since K is a fundamental body, it follows that $\boldsymbol{y} = \boldsymbol{0}$. I.e., $\boldsymbol{x} = \boldsymbol{x}'$, and $\psi_{\mathcal{L}}$ is injective over $\mathrm{Int}(K)$.

The fact that $\psi_{\mathcal{L}}$ is volume-preserving over $\mathrm{Int}(K)$ follows from the fact that it is an

injective piecewise combination of translations. □

Corollary 2.2.5. *For any non-decreasing measurable function $f : \mathbb{R} \to \mathbb{R}$, lattice $\mathcal{L} \subset \mathbb{R}^n$, and fundamental body K of $\mathcal{L}$,*

$$\int_{\mathcal{V}(\mathcal{L})} f(\|\boldsymbol{x}\|)\mathrm{d}\boldsymbol{x} \leq \int_K f(\|\boldsymbol{x}\|)\mathrm{d}\boldsymbol{x} \ .$$

Proof.

$$\begin{aligned}
\int_K f(\|\boldsymbol{x}\|)\mathrm{d}\boldsymbol{x} &= \int_{\mathrm{Int}(K)} f(\|\boldsymbol{x}\|)\mathrm{d}\boldsymbol{x} \\
&\geq \int_{\mathrm{Int}(K)} f(\|\psi_{\mathcal{L}}(\boldsymbol{x})\|)\mathrm{d}\boldsymbol{x} \\
&= \int_{\psi_{\mathcal{L}}(\mathrm{Int}(K))} f(\|\boldsymbol{x}\|)\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathcal{V}(\mathcal{L})} f(\|\boldsymbol{x}\|)\mathrm{d}\boldsymbol{x} \ ,
\end{aligned}$$

where the last equality follows from the fact that $\psi_{\mathcal{L}}$ preserves volume and $\mathrm{vol}(\mathrm{Int}(K)) = \mathrm{vol}(\mathcal{V}(\mathcal{L}))$, so it must be the case that $\psi_{\mathcal{L}}(\mathrm{Int}(K)) \subset \mathcal{V}(\mathcal{L})$ differs from $\mathcal{V}(\mathcal{L})$ on a set of measure zero. □

2.2.3 Matrix calculus

For a function $g : \mathbb{R}^{n\times n} \to \mathbb{R}$, if for some $Q \in \mathbb{R}^{n\times n}$ there exists an $B \in \mathbb{R}^{n\times n}$ such that

$$\lim_{M\to 0} \frac{g(Q+M) - g(Q) - \mathrm{Tr}(B^T M)}{\|M\|} = 0 \ ,$$

then we say that g is *differentiable* at Q, and we call B the *gradient* of g at Q,

$$\nabla_A g(A)|_{A=Q} := B \ .$$

(Some authors prefer to define $\nabla g(A)|_{A=Q}$ as B^T.)

2.3 Gradients over lattices and over positions of the Voronoi cell

The purpose of this section is to prove the following theorem. (Note that the gradient is actually symmetric, so that the transpose in the definition of the function g is simply a matter of convention.) The proof we give here is elementary but somewhat lengthy. In [RS17b], we include a much shorter proof that works for monotonic f under the assumption that one already knows that the functions g and h are differentiable.

Theorem 2.3.1 ([RS17b, Theorem 3.1])**.** *For any continuously differentiable function $f : \mathbb{R}^{\geq 0} \to \mathbb{R}$ and lattice $\mathcal{L} \subset \mathbb{R}^n$, let*

$$g(A) := \frac{1}{|\det(A)|} \cdot \int_{\mathcal{V}(A^T\mathcal{L})} f(\|\boldsymbol{x}\|^2)\mathrm{d}\boldsymbol{x}, \quad \text{and} \quad h(A) := \frac{1}{|\det(A)|} \cdot \int_{A\mathcal{V}(\mathcal{L})} f(\|\boldsymbol{x}\|^2)\mathrm{d}\boldsymbol{x} \ ,$$

where $A \in \mathrm{GL}_n(\mathbb{R})$ ranges over the set of all non-singular matrices. Then, g and h are differentiable at $A = I_n$, with

$$\nabla_A g(A)|_{A=I_n} = \nabla_A h(A)|_{A=I_n} = 2\int_{\mathcal{V}(\mathcal{L})} f'(\|\boldsymbol{x}\|^2)\boldsymbol{x}\boldsymbol{x}^T\mathrm{d}\boldsymbol{x} \ ,$$

where $f'(x) := \frac{\mathrm{d}}{\mathrm{d}x}f(x)$.

We first compute the gradient of h, which is straightforward.

Claim 2.3.2. *For any continuously differentiable function $f : \mathbb{R}^{\geq 0} \to \mathbb{R}$ and measurable set U, let*

$$h(A) := \frac{1}{|\det(A)|} \cdot \int_{AU} f(\|\boldsymbol{x}\|^2)\mathrm{d}\boldsymbol{x} \ ,$$

where $A \in \mathrm{GL}_n(\mathbb{R})$ ranges over the set of all non-singular matrices. Then, h is differentiable with

$$\nabla_A h(A)|_{A=I_n} = 2\int_U f'(\|\boldsymbol{x}\|^2)\boldsymbol{x}\boldsymbol{x}^T \mathrm{d}\boldsymbol{x} \ ,$$

where $f'(x) := \frac{\mathrm{d}}{\mathrm{d}x} f(x)$, provided that this integral and $h(I_n)$ are well-defined and finite.

Proof. By a change of variables, we have

$$h(A) = \int_U f(\|A\boldsymbol{x}\|^2)\mathrm{d}\boldsymbol{x} \ .$$

Then, by applying an appropriate high-dimensional form of Leibniz's integral rule (see, e.g., [Kam16]), we may swap the gradient and the integral and write

$$\begin{aligned}\nabla_A h(A) &= \int_U \big(\nabla_A f(\|A\boldsymbol{x}\|^2)\big)\mathrm{d}\boldsymbol{x} \\ &= 2\int_U f'(\|A\boldsymbol{x}\|^2)A\boldsymbol{x}\boldsymbol{x}^T \mathrm{d}\boldsymbol{x} \qquad \text{(Chain rule)} \ .\end{aligned}$$

□

2.3.1 Polytopes and "protected cones"

We define

$$H_{\boldsymbol{w}} := \{\boldsymbol{x} \in \mathbb{R}^n \ : \ \langle \boldsymbol{w}, \boldsymbol{x}\rangle \leq 1\} \ .$$

Any convex body with $\mathbf{0}$ in its interior can be written as

$$K(W) := \bigcap_{\boldsymbol{w}\in W} H_{\boldsymbol{w}}$$

for some (possibly infinite) set $W \subset \mathbb{R}^n$. We call $K(W)$ a *polytope* if the set W can be taken to be finite. I.e., a polytope is a bounded finite intersection of half-spaces.

The *facets* of a polytope $K(W)$ are the points in the polytope for which at least one

inequality is tight,

$$\mathcal{F}_{W,\boldsymbol{w}} := \{\boldsymbol{x} \in K(W) \ : \ \langle \boldsymbol{w}, \boldsymbol{x} \rangle = 1\} \ ,$$

for $\boldsymbol{w} \in W$.

A polytope has *normally symmetric facets* if for every $\boldsymbol{w} \in W \setminus \{\mathbf{0}\}$, $\boldsymbol{x} \in \mathcal{F}_{W,\boldsymbol{w}}$ if and only if the "reflection of $\boldsymbol{x}$ through span($\boldsymbol{w}$)," $2\pi_{\boldsymbol{w}}(\boldsymbol{x}) - \boldsymbol{x}$, is in $\mathcal{F}_{W,\boldsymbol{w}}$. In other words, a polytope has normally symmetric facets if each facet is symmetric about the line normal to the facet. Equivalently, if we define

$$R_{W,\boldsymbol{w}} := \{\boldsymbol{x} \in \mathbb{R}^n \ : \ \text{for all } \boldsymbol{w}' \in W, \langle \boldsymbol{w}', \boldsymbol{x} \rangle \leq \langle \boldsymbol{w}, \boldsymbol{x} \rangle\} \tag{2.2}$$

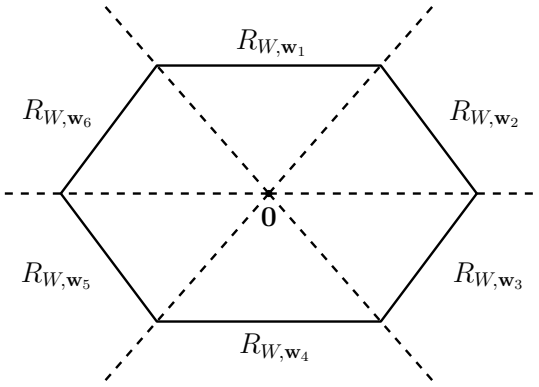
to be the minimal cone containing $\mathcal{F}_{W,\boldsymbol{w}}$ (or $\{\mathbf{0}\}$ if the facet is empty), then a polytope has normally symmetric facets if and only if $\phi_{\boldsymbol{w}}(R_{W,\boldsymbol{w}}) = R_{W,\boldsymbol{w}}$ for all $\boldsymbol{w} \in W \setminus \{\mathbf{0}\}$, where

$$\phi_{\boldsymbol{w}}(\boldsymbol{x}) := 2\pi_{\boldsymbol{w}}(\boldsymbol{x}) - \boldsymbol{x} = 2\langle \boldsymbol{w}, \boldsymbol{x} \rangle \boldsymbol{w}/\|\boldsymbol{w}\|^2 - \boldsymbol{x} \ . \tag{2.3}$$

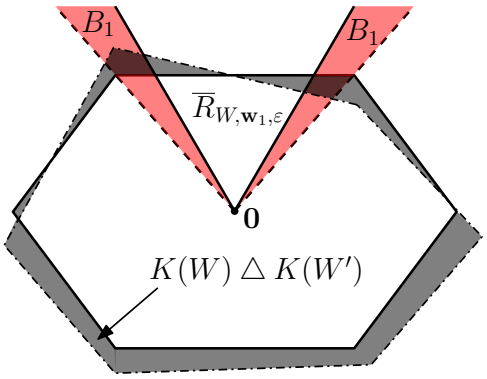
We will be interested in perturbations of polytopes. When we analyze these objects, we will have some trouble with points $\boldsymbol{x}$ that "change cones $R_{W,\boldsymbol{w}}$." The next lemma shows how to find slightly smaller "protected cones" $\overline{R}_{W,\boldsymbol{w}_i,\varepsilon}$ inside the $R_{W,\boldsymbol{w}_i}$ so that the vectors inside these "protected cones" do not leave the larger cone $R_{W,\boldsymbol{w}_i}$ after a small perturbation $W \to W'$. See Figure 2.2.

Lemma 2.3.3. *For any finite set $W := \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k\} \subset \mathbb{R}^n \setminus \{\mathbf{0}\}$ of distinct vectors such that $K(W)$ is a polytope and sufficiently small $\varepsilon > 0$, there exist "protected cones" $\overline{R}_{W,\boldsymbol{w}_1,\varepsilon}, \ldots, \overline{R}_{W,\boldsymbol{w}_k,\varepsilon}$ such that for any $W' := \{\boldsymbol{w}'_1, \ldots, \boldsymbol{w}'_k\} \subset \mathbb{R}^n$ with $\|\boldsymbol{w}'_i - \boldsymbol{w}_i\| \leq \varepsilon$ for all i, we have*

1. *$\overline{R}_{W,\boldsymbol{w}_i,\varepsilon} \subseteq R_{W',\boldsymbol{w}'_i}$ for all i (i.e., vectors in the protected cones "keep the same relevant vector" after any ε perturbation of the $\boldsymbol{w}_i$);*



(a) A polytope $K(W) \subset \mathbb{R}^2$ with the corresponding cones $R_{W,\boldsymbol{w}_1}, \ldots, R_{W,\boldsymbol{w}_6}$.



(b) The "protected cone" $\overline{R}_{W,\boldsymbol{w}_1,\varepsilon}$, "bad" set B_1, and the symmetric difference $K(W)\Delta K(W')$ between the original polytope $K(W)$ and its perturbation $K(W')$.

Figure 2.2: An illustration of Lemma 2.3.3.

2. *for all i, the "bad" set $B_i := \{\boldsymbol{x} \in \mathbb{R}^n \ : \ \boldsymbol{x} \in R_{W,\boldsymbol{w}_i} \setminus \overline{R}_{W,\boldsymbol{w}_i,\varepsilon}\}$ of points not in the protected cone satisfies $\mathrm{vol}(B_i \cap (K(W)\Delta K(W')) \leq O(\varepsilon^2)$, where the $O(\varepsilon^2)$ term hides dependence on W (i.e., the dark red region in Figure 2.2b has volume at most $O(\varepsilon^2)$); and*

3. *if $\phi_{\boldsymbol{w}_i}(R_{W,\boldsymbol{w}_i}) = R_{W,\boldsymbol{w}_i}$, then $\phi_{\boldsymbol{w}_i}(\overline{R}_{W,\boldsymbol{w}_i,\varepsilon}) = \overline{R}_{W,\boldsymbol{w}_i,\varepsilon}$.*

Proof. Let

$$\alpha_\varepsilon := \max_{i \neq j} \frac{2\varepsilon}{\|\boldsymbol{w}_i - \boldsymbol{w}_j\|} \ .$$

We take

$$\overline{R}_{W,\boldsymbol{w}_i,\varepsilon} := \{\boldsymbol{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\} \ : \ \boldsymbol{x} + \alpha_\varepsilon \|\boldsymbol{x}\| B_2^n \subseteq R_{W,\boldsymbol{w}_i}\} \ .$$

Item 3 then follows from the fact that $\phi_{\boldsymbol{w}_i}$ is an isometry. In particular, $\phi_{\boldsymbol{w}_i}(\boldsymbol{x} + \alpha_\varepsilon \|\boldsymbol{x}\| B_2^n) = \phi_{\boldsymbol{w}_i}(\boldsymbol{x}) + \alpha_\varepsilon \|\phi_{\boldsymbol{w}_i}(\boldsymbol{x})\| B_2^n$.

Turning to Item 1, suppose that $\boldsymbol{x} \in \overline{R}_{W,\boldsymbol{w}_i,\varepsilon}$, and let $j \neq i$. Let $\widehat{\boldsymbol{w}} := (\boldsymbol{w}_i - \boldsymbol{w}_j)/\|\boldsymbol{w}_i - \boldsymbol{w}_j\|$.

By the definition of the protected cone, we have

$$\langle \boldsymbol{w}_i, \boldsymbol{x} - \alpha_\varepsilon \|\boldsymbol{x}\| \widehat{\boldsymbol{w}} \rangle \geq \langle \boldsymbol{w}_j, \boldsymbol{x} - \alpha_\varepsilon \|\boldsymbol{x}\| \widehat{\boldsymbol{w}} \rangle \ .$$

Rearranging, we see that

$$\langle \boldsymbol{w}_i, \boldsymbol{x} \rangle \geq \langle \boldsymbol{w}_j, \boldsymbol{x} \rangle + \alpha_\varepsilon \|\boldsymbol{x}\| \langle \boldsymbol{w}_i - \boldsymbol{w}_j, \widehat{\boldsymbol{w}} \rangle = \langle \boldsymbol{w}_j, \boldsymbol{x} \rangle + \alpha_\varepsilon \|\boldsymbol{x}\| \|\boldsymbol{w}_i - \boldsymbol{w}_j\| \geq \langle \boldsymbol{w}_j, \boldsymbol{x} \rangle + 2\varepsilon \|\boldsymbol{x}\| \ .$$

Therefore, by Cauchy-Schwarz, we have

$$\langle \boldsymbol{w}'_i, \boldsymbol{x} \rangle \geq \langle \boldsymbol{w}_i, \boldsymbol{x} \rangle - \varepsilon \|\boldsymbol{x}\| \geq \langle \boldsymbol{w}_j, \boldsymbol{x} \rangle + \varepsilon \|\boldsymbol{x}\| \geq \langle \boldsymbol{w}'_j, \boldsymbol{x} \rangle \ .$$

I.e., $\boldsymbol{x} \in R_{W', \boldsymbol{w}'_i}$, as needed.

Finally, suppose that $\boldsymbol{x} \in B_i \cap (K(W) \Delta K(W'))$. We assume that $\boldsymbol{x} \in K(W) \setminus K(W')$, since the case where $\boldsymbol{x} \in K(W') \setminus K(W)$ is nearly identical. Since $\boldsymbol{x} \in K(W)$, we have

$$\langle \boldsymbol{w}_i, \boldsymbol{x} \rangle \leq 1 \ .$$

Since $\boldsymbol{x} \notin K(W')$, there exists a j such that

$$1 < \langle \boldsymbol{w}'_j, \boldsymbol{x} \rangle = \langle \boldsymbol{w}_j, \boldsymbol{x} \rangle + \langle \boldsymbol{w}'_j - \boldsymbol{w}_j, \boldsymbol{x} \rangle \leq \langle \boldsymbol{w}_i, \boldsymbol{x} \rangle + \varepsilon \|\boldsymbol{x}\| \ ,$$

where we used that $\boldsymbol{x} \in R_{W, \boldsymbol{w}_i}$. It follows that

$$|\langle \boldsymbol{w}_i, \boldsymbol{x} \rangle - 1| \leq \varepsilon \|\boldsymbol{x}\| = O(\varepsilon) \ . \tag{2.4}$$

And, since $\boldsymbol{x} \in B_i$, there must also be some $j \neq i$ such that

$$|\langle \boldsymbol{w}_i - \boldsymbol{w}_j, \boldsymbol{x} \rangle| \leq O(\varepsilon)\|\boldsymbol{x}\| \leq O(\varepsilon) \; . \tag{2.5}$$

In other words, $\boldsymbol{x}$ lies in one of finitely many intersections between a slab of width $O(\varepsilon)$ bounded away from $\mathbf{0}$ (defined by Eq. (2.4)), a slab of width $O(\varepsilon)$ around $\mathbf{0}$ (defined by Eq. (2.5)), and the bounded set $K(W) \cup K(W')$. Item 2 then follows from the fact that any such set has volume $O(\varepsilon^2)$. □

The next rather technical and specific corollary shows that these protected cones in some sense "do not distinguish between perturbations to $\boldsymbol{w}_i'$ and perturbations to $\phi_{\boldsymbol{w}_i}(\boldsymbol{w}_i')$," when $K(W)$ has normally symmetric facets.

Corollary 2.3.4. *For any finite set $W := \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k\} \subset \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that $K(W)$ is a polytope with normally symmetric facets, sufficiently small $\varepsilon > 0$, and $W' := \{\boldsymbol{w}_1', \ldots, \boldsymbol{w}_k'\} \subset \mathbb{R}^n$ with $\|\boldsymbol{w}_i' - \boldsymbol{w}_i\| \leq \varepsilon$ for all i, let $W'' := \{\phi_{\boldsymbol{w}_1}(\boldsymbol{w}_1'), \ldots, \phi_{\boldsymbol{w}_k}(\boldsymbol{w}_k')\}$. Then,*

$$\overline{R}_{W,\boldsymbol{w}_i,\varepsilon} \cap K(W') = \phi_{\boldsymbol{w}_i}(\overline{R}_{W,\boldsymbol{w}_i,\varepsilon} \cap K(W''))$$

for all i, where $\overline{R}_{W,\boldsymbol{w}_i,\varepsilon}$ is the "protected cone" from Lemma 2.3.3.

Proof. By Item 1 of Lemma 2.3.3, we have that $\overline{R}_{W,\boldsymbol{w}_i,\varepsilon} \subseteq R_{W',\boldsymbol{w}_i'} \cap R_{W'',\phi_{\boldsymbol{w}_i}(\boldsymbol{w}_i')}$. It follows that $\overline{R}_{W,\boldsymbol{w}_i,\varepsilon} \cap K(W') = \overline{R}_{W,\boldsymbol{w}_i,\varepsilon} \cap H_{\boldsymbol{w}_i'}$, and similarly $\overline{R}_{W,\boldsymbol{w}_i,\varepsilon} \cap K(W'') = \overline{R}_{W,\boldsymbol{w}_i,\varepsilon} \cap H_{\phi_{\boldsymbol{w}_i}(\boldsymbol{w}_i')}$. Noting that for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $\langle \phi_{\boldsymbol{w}_i}(\boldsymbol{y}), \boldsymbol{x} \rangle = \langle \boldsymbol{y}, \phi_{\boldsymbol{w}_i}(\boldsymbol{x}) \rangle = \langle \boldsymbol{y}, \phi_{\boldsymbol{w}_i}^{-1}(\boldsymbol{x}) \rangle$, we see that for any $\boldsymbol{y} \in \mathbb{R}^n$,

$$H_{\phi_{\boldsymbol{w}_i}(\boldsymbol{y})} = \{\boldsymbol{x} \in \mathbb{R}^n : \langle \phi_{\boldsymbol{w}_i}(\boldsymbol{y}), \boldsymbol{x} \rangle \leq 1\} = \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{y}, \phi_{\boldsymbol{w}_i}^{-1}(\boldsymbol{x}) \rangle \leq 1\} = \phi_{\boldsymbol{w}_i}(H_{\boldsymbol{y}}) \; .$$

We therefore have

$$\phi_{\boldsymbol{w}_i}(\overline{R}_{W,\boldsymbol{w}_i,\varepsilon} \cap H_{\phi_{\boldsymbol{w}_i}(\boldsymbol{w}_i')}) = \phi_{\boldsymbol{w}_i}(\overline{R}_{W,\boldsymbol{w}_i,\varepsilon} \cap \phi_{\boldsymbol{w}_i}(H_{\boldsymbol{w}_i'})) = \phi_{\boldsymbol{w}_i}(\overline{R}_{W,\boldsymbol{w}_i,\varepsilon}) \cap H_{\boldsymbol{w}_i'} \ ,$$

where the last equality follows from the fact that $\phi_{\boldsymbol{w}_i} = \phi_{\boldsymbol{w}_i}^{-1}$. The result follows by recalling from Item 3 of Lemma 2.3.3 that $\phi_{\boldsymbol{w}_i}(\overline{R}_{W,\boldsymbol{w}_i,\varepsilon}) = \overline{R}_{W,\boldsymbol{w}_i,\varepsilon}$. □

2.3.2 Perturbations of polytopes

The purpose of this subsection is to prove Lemma 2.3.8. The next claim shows that a small change to W corresponds to a small change to $K(W)$.

Claim 2.3.5. *For any finite set $W = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k\} \subset \mathbb{R}^n$ such that $K(W)$ is a polytope, sufficiently small $\varepsilon > 0$, and any $W' := \{\boldsymbol{w}_1', \ldots, \boldsymbol{w}_k'\} \subset \mathbb{R}^n$ such that $\|\boldsymbol{w}_i - \boldsymbol{w}_i'\| \leq \varepsilon$,*

$$\mathrm{vol}(K(W) \Delta K(W')) \leq O(\varepsilon) \ ,$$

where the $O(\varepsilon)$ term hides factors that depend on W.

Proof. Suppose that $\boldsymbol{x} \in K(W) \setminus K(W')$. (The case when $\boldsymbol{x} \in K(W') \setminus K(W)$ is essentially identical.) Then, there exists some index i such that

$$\langle \boldsymbol{w}_i, \boldsymbol{x} \rangle \leq 1 < \langle \boldsymbol{w}_i', \boldsymbol{x} \rangle \ .$$

Using $\langle \boldsymbol{w}_i' - \boldsymbol{w}_i, \boldsymbol{x} \rangle \leq \varepsilon \|\boldsymbol{x}\| \leq O(\varepsilon)$, we see that

$$1 - O(\varepsilon) < \langle \boldsymbol{w}_i, \boldsymbol{x} \rangle \leq 1 \ .$$

I.e., $\boldsymbol{x}$ lies in one of k slabs with width proportional to ε. The result follows by noting that

the intersection of any such slab with the bounded body $K(W)$ has volume $O(\varepsilon)$. $\square$

Claim 2.3.6. *For any non-singular matrix $A \in \mathrm{GL}_n(\mathbb{R})$ and any $W \subset \mathbb{R}^n$, $AK(W) = K(A^{-T}W)$.*

Proof. It suffices to note that $\boldsymbol{x} \in AK(W)$ if and only if $\langle \boldsymbol{w}, A^{-1}\boldsymbol{x}\rangle = \langle A^{-T}\boldsymbol{w}, \boldsymbol{x}\rangle \leq 1$ for all $\boldsymbol{w} \in W$. $\square$

Recall that the Voronoi cell of a lattice $\mathcal{L}$ is given by $\mathcal{V}(\mathcal{L}) := K(W)$, where $W := \{2\boldsymbol{y}/\|\boldsymbol{y}\|^2 \ : \ \boldsymbol{y} \in \mathcal{L} \setminus \{\boldsymbol{0}\}\}$. We therefore define $\psi(\boldsymbol{y}) := 2\boldsymbol{y}/\|\boldsymbol{y}\|^2$. We will need the following technical claim, which shows how ψ behaves under small linear perturbations.

Claim 2.3.7. *For any $\boldsymbol{y} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$, matrix $M \in \mathbb{R}^{n\times n}$, and sufficiently small $\varepsilon > 0$, we have*

$$\|M_\varepsilon^{-1}\psi(\boldsymbol{y}) - \psi(\phi_{\boldsymbol{y}}(M_\varepsilon \boldsymbol{y}))\| \leq O(\varepsilon^2) \ ,$$

where $M_\varepsilon := I_n + \varepsilon M$ and the $O(\varepsilon^2)$ notation hides dependence on $\boldsymbol{y}$ and $\|M\|$.

Proof. We have

$$M_\varepsilon^{-1}\psi(\boldsymbol{y}) = (I_n - \varepsilon M)\psi(\boldsymbol{y}) + O(\varepsilon^2)\cdot \boldsymbol{u} = 2\cdot \frac{\boldsymbol{y} - \varepsilon M\boldsymbol{y}}{\|\boldsymbol{y}\|^2} + O(\varepsilon^2)\cdot \boldsymbol{u} \ ,$$

where $\boldsymbol{u} \in \mathbb{R}^n$ is some unit vector that depends on $\boldsymbol{y}$ and M. Similarly, we have

$$\begin{aligned}
\psi(\phi_{\boldsymbol{y}}(M_\varepsilon \boldsymbol{y})) &= 2\cdot \frac{\phi_{\boldsymbol{y}}(M_\varepsilon \boldsymbol{y})}{\|M_\varepsilon \boldsymbol{y}\|^2} \\
&= 2\cdot \frac{\boldsymbol{y} + 2\varepsilon\langle \boldsymbol{y}, M\boldsymbol{y}\rangle \boldsymbol{y}/\|\boldsymbol{y}\|^2 - \varepsilon M\boldsymbol{y}}{\|\boldsymbol{y}\|^2 + 2\varepsilon\langle \boldsymbol{y}, M\boldsymbol{y}\rangle + \varepsilon^2\|M\boldsymbol{y}\|^2} \\
&= 2\cdot \frac{(\boldsymbol{y} - \varepsilon M\boldsymbol{y})\cdot(1 + 2\varepsilon\langle \boldsymbol{y}, M\boldsymbol{y}\rangle/\|\boldsymbol{y}\|^2)}{\|\boldsymbol{y}\|^2\cdot(1 + 2\varepsilon\langle \boldsymbol{y}, M\boldsymbol{y}\rangle/\|\boldsymbol{y}\|^2)} + O(\varepsilon^2)\cdot \boldsymbol{u}' \\
&= 2\cdot \frac{\boldsymbol{y} - \varepsilon M\boldsymbol{y}}{\|\boldsymbol{y}\|^2} + O(\varepsilon^2)\cdot \boldsymbol{u}' \ ,
\end{aligned}$$

where $\boldsymbol{u}'$ is some unit vector that depends on $\boldsymbol{y}$ and M. The result follows. $\square$

With this, we can prove the analogue of Claim 2.3.6 for $K(\psi(Y))$ instead of $K(W)$. (Note the similarity between the set Y' here and the set W'' in Corollary 2.3.4.)

Lemma 2.3.8. *For any finite set $Y := \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k\} \subset \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that $K(\psi(Y))$ is a polytope, linear transformation $M \in \mathbb{R}^{n \times n}$, and sufficiently small $\varepsilon > 0$,*

$$\mathrm{vol}\big((M_\varepsilon K(\psi(Y)))\Delta K(\psi(Y'))\big) \leq O(\varepsilon^2) \,,$$

where $M_\varepsilon := I_n + \varepsilon M$, $Y' := \{\phi_{\boldsymbol{y}_1}(M_\varepsilon^T \boldsymbol{y}_1), \ldots, \phi_{\boldsymbol{y}_k}(M_\varepsilon^T \boldsymbol{y}_k)\}$, and the $O(\varepsilon^2)$ term hides dependence on Y and $\|M\|$.

Proof. By Claim 2.3.6, $M_\varepsilon K(\psi(Y)) = K(M_\varepsilon^{-T}\psi(Y))$. And, by Claim 2.3.7, the vectors in $\psi(Y')$ differ from the vectors in $M_\varepsilon^{-T}\psi(Y)$ by vectors of length $O(\varepsilon^2)$. The result then follows from Claim 2.3.5. □

2.3.3 Gradient equivalence for polytopes with normally symmetric facets

We can now prove a more general variant of Theorem 2.3.1.

Theorem 2.3.9 ([RS17b, Theorem 3.9])**.** *For any continuously differentiable function $f : \mathbb{R}^{\geq 0} \to \mathbb{R}$ and finite set $Y := \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k\} \subset \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that $K(\psi(Y))$ is a polytope with normally symmetric facets, let*

$$g(A) := \frac{1}{|\det(A)|} \cdot \int_{K(\psi(A^T Y))} f(\|\boldsymbol{x}\|^2)\mathrm{d}\boldsymbol{x}, \quad \text{and} \quad h(A) := \frac{1}{|\det(A)|} \cdot \int_{AK(\psi(Y))} f(\|\boldsymbol{x}\|^2)\mathrm{d}\boldsymbol{x} \,.$$

Then, g and h are differentiable at $A = I_n$, with

$$\nabla_A g(A)|_{A=I_n} = \nabla_A h(A)|_{A=I_n} = 2\int_{K(\psi(Y))} f'(\|\boldsymbol{x}\|^2)\boldsymbol{x}\boldsymbol{x}^T \mathrm{d}\boldsymbol{x} \,,$$

where $f'(x) := \frac{\mathrm{d}}{\mathrm{d}x} f(x)$.

Proof. Since f is continuous and we are only interested in its value over a bounded region, we may assume without loss of generality that f is bounded. I.e., $|f(x)| \le C_f$ for some finite $C_f > 0$. We have already computed the gradient of h in Claim 2.3.2, so we only need to show that $\nabla_A g(A)|_{A=I_n} = \nabla_A h(A)|_{A=I_n}$. For a convex body $K \subset \mathbb{R}^n$, let $F(K) := \int_K f(\|x\|)^2 \mathrm{d}\boldsymbol{x}$. (E.g., $g(A) = F(K(\psi(A^T Y)))/|\det(A)|$ and $h(A) = F(AK(\psi(Y)))/|\det(A)|$.)

Let $\varepsilon > 0$ be sufficiently small, and let $M \in \mathbb{R}^{n\times n}$ such that $\|M^T \boldsymbol{y}_i\| \le 1$ and $\|\psi(\boldsymbol{y}_i + \varepsilon M^T \boldsymbol{y}_i) - \psi(\boldsymbol{y}_i)\| \le \varepsilon$ for all i. Let $M_\varepsilon := I_n + \varepsilon M$. It suffices to show that

$$\left|\det(M_\varepsilon) g(M_\varepsilon) - \det(M_\varepsilon) h(M_\varepsilon)\right| = \left|F(K(\psi(M_\varepsilon^T Y))) - F(M_\varepsilon K(\psi(Y)))\right| \le O(\varepsilon^2) \ .$$

We first move from $M_\varepsilon K(\psi(Y))$ to $K(\psi(Y'))$, where $Y' := \{\phi_{\boldsymbol{y}_1}(M_\varepsilon^T \boldsymbol{y}_1), \ldots, \phi_{\boldsymbol{y}_k}(M_\varepsilon^T \boldsymbol{y}_k)\}$:

$$\left|F(M_\varepsilon K(\psi(Y))) - F(K(\psi(Y')))\right| \le C_f \cdot \mathrm{vol}((M_\varepsilon K(\psi(Y)))\Delta K(\psi(Y'))) \le O(\varepsilon^2) \ ,$$

where the last inequality follows from Lemma 2.3.8.

Let $\boldsymbol{w}_i := \psi(\boldsymbol{y}_i)$ and $W := \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k\}$. Since the cones $R_{W,\boldsymbol{w}_i}$ cover space, we have

$$\left|F(K(\psi(Y'))) - F(K(\psi(M_\varepsilon^T Y)))\right| \le \sum_i \left|F(R_{W,\boldsymbol{w}_i} \cap K(\psi(Y'))) - F(R_{W,\boldsymbol{w}_i} \cap K(\psi(M_\varepsilon^T Y)))\right| \ .$$

Let $\overline{R}_i := \overline{R}_{W,\boldsymbol{w}_i,\varepsilon}$ be the "protected cones" from Lemma 2.3.3. Then,

$$\begin{aligned}
&\left|F(R_{W,\boldsymbol{w}_i} \cap K(\psi(Y'))) - F(R_{W,\boldsymbol{w}_i} \cap K(\psi(M_\varepsilon^T Y)))\right| \\
&\le \left|F(\overline{R}_i \cap K(\psi(Y'))) - F(\overline{R}_i \cap K(\psi(M_\varepsilon^T Y)))\right| \\
&\qquad + C_f \cdot \mathrm{vol}\Big((R_{W,\boldsymbol{w}_i} \setminus \overline{R}_i) \cap (K(\psi(Y'))\Delta K(\psi(M_\varepsilon^T Y)))\Big) \\
&\le \left|F(\overline{R}_i \cap K(\psi(Y'))) - F(\overline{R}_i \cap K(\psi(M_\varepsilon^T Y)))\right| + O(\varepsilon^2) \ ,
\end{aligned}$$

where we have applied Item 2 of Lemma 2.3.3 by noting that

$$\begin{aligned}\mathrm{vol}\big((R_{W,\boldsymbol{w}_i} \setminus \overline{R}_i) \cap (K(\psi(Y'))\Delta K(\psi(M_\varepsilon^T Y)))\big) &\leq \mathrm{vol}\big((R_{W,\boldsymbol{w}_i} \setminus \overline{R}_i) \cap (K(\psi(M_\varepsilon^T Y))\Delta K(\psi(Y)))\big) \\ &\quad + \mathrm{vol}\big((R_{W,\boldsymbol{w}_i} \setminus \overline{R}_i) \cap (K(\psi(Y'))\Delta K(\psi(Y)))\big) \\ &\leq O(\varepsilon^2) \ .\end{aligned}$$

Finally, we claim that

$$F(\overline{R}_i \cap K(\psi(Y'))) = F(\overline{R}_i \cap K(\psi(M_\varepsilon^T Y))) \ .$$

Indeed, by Corollary 2.3.4, we have that $\overline{R}_i \cap K(\psi(M_\varepsilon^T Y)) = \phi_{\boldsymbol{w}_i}(\overline{R}_i \cap K(\psi(Y')))$. (Here, we have used the fact that $\phi_{\boldsymbol{y}_i} = \phi_{\boldsymbol{w}_i}$ together with the fact that $\phi_{\boldsymbol{w}_i}$ commutes with ψ.) Recalling that $\phi_{\boldsymbol{w}_i}$ is an isometry (and that it therefore preserves volume), we have

$$\begin{aligned}F(\overline{R}_i \cap K(\psi(Y'))) &= \int_{\overline{R}_i \cap K(\psi(Y'))} f(\|\phi_{\boldsymbol{w}_i}(\boldsymbol{x})\|^2)\mathrm{d}\boldsymbol{x} \\ &= F(\phi_{\boldsymbol{w}_i}(\overline{R}_i \cap K(\psi(Y')))) \\ &= F(\overline{R}_i \cap K(\psi(M_\varepsilon^T Y))) \ ,\end{aligned}$$

as claimed. Combining everything together gives the result. □

2.3.4 Proof of Theorem 2.3.1

We now prove Theorem 2.3.1 as a relatively straightforward corollary of Theorem 2.3.9. To do this, we first recall the following well known fact.

Lemma 2.3.10. *For any lattice $\mathcal{L} \subset \mathbb{R}^n$, the Voronoi cell $\mathcal{V}(\mathcal{L})$ has normally symmetric facets.*

Proof. Let $W := \psi(\mathcal{L} \setminus \{\mathbf{0}\}) = \{2\boldsymbol{y}/\|\boldsymbol{y}\|^2 \ : \ \boldsymbol{y} \in \mathcal{L} \setminus \{\mathbf{0}\}\}$ so that $\mathcal{V}(\mathcal{L}) = K(W)$, and let $\boldsymbol{y}, \boldsymbol{y}' \in \mathcal{L} \setminus \{\mathbf{0}\}$ be distinct. Let $\boldsymbol{w} := 2\boldsymbol{y}/\|\boldsymbol{y}\|^2 \in W$, and $\boldsymbol{w}' := 2\boldsymbol{y}'/\|\boldsymbol{y}'\|^2 \in W$ be the corresponding points in $\boldsymbol{w}$. Suppose $\boldsymbol{x} \in R_{W,\boldsymbol{w}}$. It suffices to show that $\langle \boldsymbol{w}', \phi_{\boldsymbol{w}}(\boldsymbol{x})\rangle \leq \langle \boldsymbol{w}, \phi_{\boldsymbol{w}}(\boldsymbol{x})\rangle$. Equivalently, it suffices to show that

$$\frac{\langle \boldsymbol{y}', \phi_{\boldsymbol{y}}(\boldsymbol{x})\rangle}{\|\boldsymbol{y}'\|^2} \leq \frac{\langle \boldsymbol{y}, \phi_{\boldsymbol{y}}(\boldsymbol{x})\rangle}{\|\boldsymbol{y}\|^2} \ .$$

We consider the inner product of $\boldsymbol{y} - \boldsymbol{y}'$ with $\boldsymbol{x}$. In particular, since $\boldsymbol{y} - \boldsymbol{y}'$ is a non-zero lattice vector and $\boldsymbol{x} \in R_{W,\boldsymbol{w}}$, we have

$$\langle \boldsymbol{y} - \boldsymbol{y}', \boldsymbol{x}\rangle \leq \|\boldsymbol{y} - \boldsymbol{y}'\|^2 \cdot \frac{\langle \boldsymbol{y}, \boldsymbol{x}\rangle}{\|\boldsymbol{y}\|^2} = \left(1 - \frac{2\langle \boldsymbol{y}, \boldsymbol{y}'\rangle}{\|\boldsymbol{y}\|^2} + \frac{\|\boldsymbol{y}'\|^2}{\|\boldsymbol{y}\|^2}\right) \cdot \langle \boldsymbol{y}, \boldsymbol{x}\rangle \ .$$

Rearranging, we have

$$2 \cdot \frac{\langle \boldsymbol{y}, \boldsymbol{y}'\rangle\langle \boldsymbol{y}, \boldsymbol{x}\rangle}{\|\boldsymbol{y}\|^2\|\boldsymbol{y}'\|^2} - \frac{\langle \boldsymbol{y}', \boldsymbol{x}\rangle}{\|\boldsymbol{y}'\|^2} \leq \frac{\langle \boldsymbol{y}, \boldsymbol{x}\rangle}{\|\boldsymbol{y}\|^2} \ .$$

The result follows by noting that the right-hand side is equal to $\frac{\langle \boldsymbol{y}, \phi_{\boldsymbol{y}}(\boldsymbol{x})\rangle}{\|\boldsymbol{y}\|^2}$ and the left-hand side is equal to $\frac{\langle \boldsymbol{y}', \phi_{\boldsymbol{y}}(\boldsymbol{x})\rangle}{\|\boldsymbol{y}'\|^2}$. □

With this, we can prove the theorem.

Proof of Theorem 2.3.1. Let $U \subset \mathrm{GL}_n(\mathbb{R})$ be some bounded open neighborhood around I_n. It suffices to show that there exists a *finite* set Y satisfying that for all $A \in U$, the Voronoi cell $\mathcal{V}(A^T\mathcal{L})$ is equal to $K(\psi(A^T Y))$. (Without the finiteness assumption, we could simply take $Y = \mathcal{L} \setminus \{\mathbf{0}\}$.) The result then follows from Theorem 2.3.9 applied to Y together with Lemma 2.3.10.

Note that we only need to take $\boldsymbol{y} \in Y$ if $\|A^T\boldsymbol{y}\| \leq 2\mu(A^T\mathcal{L})$ for some $A \in U$. Let $s := \sup_{A \in U} \mu(A^T\mathcal{L})$, and notice that $s < \infty$ since U is bounded and the covering radius function μ is continuous. Let $\alpha := \inf_{A \in U, \boldsymbol{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \|A^T\boldsymbol{x}\|/\|\boldsymbol{x}\|$. We may take U small

enough that $\alpha > 0$. Then, let $Y := (\mathcal{L} \setminus \{\mathbf{0}\}) \cap (2s/\alpha)B_2^n$ and notice that it is a finite set. If $\|A^T\boldsymbol{y}\| \leq 2\mu(A^T\mathcal{L}) \leq 2s$ for some $A \in U$, then $\|\boldsymbol{y}\| \leq \|A^T\boldsymbol{y}\|/\alpha \leq 2s/\alpha$. So, $K(\psi(A^T Y)) = \mathcal{V}(A^T\mathcal{L})$, as needed. □

2.4 Proof of the Reverse Minkowski Theorem

In this section, we prove our main theorem, Theorem 2.1.2. Recall that the Voronoi cell $\mathcal{V}(\mathcal{L})$ of a lattice $\mathcal{L} \subset \mathbb{R}^n$ is the symmetric polytope of all vectors in $\mathbb{R}^n$ that are closer to $\mathbf{0}$ than to any other lattice vector,

$$\mathcal{V}(\mathcal{L}) := \{\boldsymbol{x} \in \mathbb{R}^n \ : \ \forall \boldsymbol{y} \in \mathcal{L},\ \langle \boldsymbol{y}, \boldsymbol{x}\rangle \leq \|\boldsymbol{y}\|^2/2\}\ .$$

Also recall that for parameter $s > 0$, $\gamma_s(\cdot)$ is the Gaussian measure on $\mathbb{R}^n$ given by

$$\gamma_s(S) := \int_{S/s} e^{-\pi\|\boldsymbol{x}\|^2} \mathrm{d}\boldsymbol{x}$$

for any measurable set $S \subseteq \mathbb{R}^n$. (Some authors prefer to parametrize γ in terms of the standard deviation $\sigma := s/\sqrt{2\pi}$.) We are interested in the Gaussian mass $\gamma_s(\mathcal{V}(\mathcal{L}))$ of the Voronoi cell because, as the following lemma due to Chung, Dadush, Liu, and Peikert shows, this can be used to obtain an upper bound on the mass $\rho_s(\mathcal{L})$ of the lattice itself [CDLP13]. We include a proof for completeness.

Lemma 2.4.1 ([CDLP13, Lemma 3.4])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and $s > 0$,*

$$\rho_s(\mathcal{L}) \cdot \gamma_s(\mathcal{V}(\mathcal{L})) \leq 1\ .$$

Proof. By scaling appropriately, we may assume without loss of generality that $s = 1$. Note that the Voronoi cell tiles space with respect to $\mathcal{L}$. I.e., $\bigcup_{\boldsymbol{y}\in\mathcal{L}}(\mathcal{V}(\mathcal{L}) + \boldsymbol{y}) = \mathbb{R}^n$, where the

union is disjoint except on a measure-zero set. So,

$$\begin{aligned}
1 &= \int_{\mathbb{R}^n} e^{-\pi\|\boldsymbol{x}\|^2}\mathrm{d}\boldsymbol{x} \\
&= \sum_{\boldsymbol{y}\in\mathcal{L}} \int_{\mathcal{V}(\mathcal{L})} e^{-\pi\|\boldsymbol{y}+\boldsymbol{t}\|^2}\mathrm{d}\boldsymbol{t} \\
&= \sum_{\boldsymbol{y}\in\mathcal{L}} e^{-\pi\|\boldsymbol{y}\|^2} \int_{\mathcal{V}(\mathcal{L})} e^{-\pi\|\boldsymbol{t}\|^2} e^{2\pi\langle\boldsymbol{y},\boldsymbol{t}\rangle}\mathrm{d}\boldsymbol{t} \\
&= \sum_{\boldsymbol{y}\in\mathcal{L}} \rho(\boldsymbol{y}) \int_{\mathcal{V}(\mathcal{L})} e^{-\pi\|\boldsymbol{t}\|^2} \cosh(2\pi\langle\boldsymbol{y},\boldsymbol{t}\rangle)\mathrm{d}\boldsymbol{t} \\
&\geq \sum_{\boldsymbol{y}\in\mathcal{L}} \rho(\boldsymbol{y}) \int_{\mathcal{V}(\mathcal{L})} e^{-\pi\|\boldsymbol{t}\|^2}\mathrm{d}\boldsymbol{t} \\
&= \rho(\mathcal{L})\gamma(\mathcal{V}(\mathcal{L}))\ ,
\end{aligned}$$

where the fourth line follows from the fact that the Voronoi cell is symmetric. □

Therefore, in order to prove Theorem 2.1.2, it suffices to show that $\gamma_{1/t}(\mathcal{V}(\mathcal{L})) \geq 2/3$ for every lattice $\mathcal{L} \subset \mathbb{R}^n$ with $\det(\mathcal{L}') \geq 1$ for all sublattices $\mathcal{L}' \subseteq \mathcal{L}$, where $t := 10(\log n + 2)$. As we explained in the introduction, we will reduce this to studying local minima of the function $\mathcal{L} \mapsto \gamma_{1/t}(\mathcal{V}(\mathcal{L}))$ over the set of determinant-one lattices. (We do not know whether such local minima actually exist.)

In Section 2.4.1, we collect some facts about the Gaussian mass of convex bodies. In Section 2.4.2, we apply these facts to the Voronoi cell to prove Theorem 2.1.2.

2.4.1 Gaussian mass of convex bodies

We say that a measurable set $U \subset \mathbb{R}^n$ is in *isotropic Gaussian position for parameter s* if

$$\int_{U/s} e^{-\pi\|\boldsymbol{x}\|^2} \boldsymbol{x}\boldsymbol{x}^T \mathrm{d}\boldsymbol{x} = \alpha \cdot I_n$$

for some scalar $\alpha > 0$. If $s = 1$, we simply say that U is in *isotropic Gaussian position*. Such a position has been considered elsewhere (e.g., [Bob11]), but as far as we know, it did not have a name before [RS17b].

The main goal of this section is to prove the following theorem. We will also include a standard fact in Lemma 2.4.7 towards the end of this section.

Theorem 2.4.2 ([RS17b, Theorem 4.2])**.** *For any symmetric convex body $K \subset \mathbb{R}^n$ with $\mathrm{vol}(K) \geq 1$, if K is in isotropic Gaussian position for some parameter $0 < s \leq 1/t$, then $\gamma_s(K) \geq 2/3$ where $t := 10(\log n + 2)$.*

Our proof of Theorem 2.4.2 proceeds in two parts. The first part is a result due to Bobkov [Bob11] (Proposition 2.4.3 below), showing that an isotropic Gaussian position of a convex body has maximal Gaussian mass. We include a proof for completeness. In the second part (Theorem 2.4.6 below), we show that any volume-one convex body $K \subset \mathbb{R}^n$ has a position such that $\gamma_s(K) \geq 2/3$.

Proposition 2.4.3 ([Bob11, Proposition 3.1])**.** *For any symmetric convex body $K \subset \mathbb{R}^n$, if K is in isotropic Gaussian position for some parameter $s > 0$, then $\gamma_s(K) \geq \gamma_s(AK)$ for any determinant-one matrix $A \in \mathrm{SL}_n(\mathbb{R})$.*

We start by observing that isotropic Gaussian positions correspond to critical points of the Gaussian mass function over positions. (The simple proof appears in [RS17b].)

Fact 2.4.4. *For any measurable set $U \subset \mathbb{R}^n$, let*

$$h(A) := \frac{\gamma(AU)}{|\det(A)|} ,$$

where $A \in \mathrm{GL}_n(\mathbb{R})$ ranges over the non-singular matrices. Then,

$$\nabla_A h(A)|_{A=I_n} = -2\pi \int_U e^{-\pi\|\boldsymbol{x}\|^2} \boldsymbol{x}\boldsymbol{x}^T \mathrm{d}\boldsymbol{x} .$$

In particular, $A \mapsto \gamma(AU)$ has a critical point at I_n when restricted to determinant-one matrices if and only if U is in isotropic Gaussian position.

We will also need the following result due to Cordero-Erausquin, Fradelizi, and Maurey [CFM04], which is related to the so-called (B) conjecture due to Banaszczyk (see [Lat02]).

Theorem 2.4.5 ([CFM04])**.** *For any symmetric convex body $K \subset \mathbb{R}^n$, the function $\gamma(e^D K)$, where $D \in \mathbb{R}^{n \times n}$ ranges over all diagonal matrices, is log-concave.*

Proof of Proposition 2.4.3. By scaling K, we may assume that $s = 1$. Let $A = UDV$ be the singular-value decomposition of A. (I.e., D is a diagonal matrix and U and V are orthogonal matrices.) Note that the Gaussian measure is invariant under orthogonal transformations, so that $\gamma(AK) = \gamma(UDVK) = \gamma(DVK)$. Let $K' := VK$, and note that $\gamma(K') = \gamma(K)$ and that K' is in isotropic Gaussian position, since V is an orthogonal transformation.

Let $\widehat{h}(M) := \gamma(e^M K')/|\det(e^M)|$. By Fact 2.4.4 and the chain rule, we have

$$\nabla_M \widehat{h}(M)|_{M=0} = -2\pi \int_{K'} e^{-\pi\|\boldsymbol{x}\|^2} \boldsymbol{x}\boldsymbol{x}^T \mathrm{d}\boldsymbol{x} = -\alpha \cdot I_n$$

for some scalar $\alpha \in \mathbb{R}$, where the second equality is simply the fact that K' is in isotropic Gaussian position. Let $X \subset \mathbb{R}^{n \times n}$ be the set of trace-zero diagonal matrices. Then, the function $\widehat{h}_X$ obtained by restricting $\widehat{h}$ to X has a critical point at zero, since $\mathrm{Tr}(I_n M) = 0$ for any $M \in X$. By Theorem 2.4.5, $\widehat{h}_X$ is log-concave, so that this critical point must be a global maximum. Therefore, $\gamma(AK) = \gamma(DK') \leq \gamma(K') = \gamma(K)$, as needed. □

The second part of the proof of Theorem 2.4.2 requires the following theorem. The proof is based on an important theorem that follows from the work of Figiel and Tomczak-Jaegermann [FT79], Lewis [Lew79], and Pisier [Pis82].

Theorem 2.4.6. *For any symmetric convex body $K \subset \mathbb{R}^n$ with volume one, there is a determinant-one matrix $A \in \mathrm{SL}_n(\mathbb{R})$ such that $\gamma_{1/t}(AK) \geq 2/3$, where $t := 2\sqrt{3e}(\log_2 n + 2) < 10(\log n + 2)$.*

We now obtain Theorem 2.4.2 as an immediate corollary of Proposition 2.4.3 and Theorem 2.4.6.

Proof of Theorem 2.4.2. By Theorem 2.4.6, there is some $A \in \mathrm{SL}_n(\mathbb{R})$ such that $\gamma_s(AK) \geq 2/3$, and by Proposition 2.4.3, $\gamma_s(K) \geq \gamma_s(AK) \geq 2/3$, as needed. □

2.4.1.1 Concentration of measure

We will also need a standard lemma about the concentration of Gaussian measure. Recall that the *inradius* of a convex body K is defined as $\max\{r \geq 0 \ : \ rB_2^n \subseteq K\}$, i.e., the radius of the largest ball contained in the body.

Lemma 2.4.7. *If $K \subset \mathbb{R}^n$ is a convex body with $\gamma_{1/t}(K) \geq 2/3$ for some $t > 0$, then*

$$\gamma_{1/(t+\tau)}(K) \geq 1 - e^{-\pi r^2 \tau^2}/3 \,,$$

for any $\tau \geq 0$, where $r \geq 0$ is the inradius of K.

2.4.2 Proof of Theorem 2.1.2

We now use Theorem 2.3.1 and Theorem 2.4.2 to characterize local minima of $\gamma_s(\mathcal{V}(\mathcal{L}))$.

Theorem 2.4.8 ([RS17b, Theorem 4.12])**.** *If $\mathcal{L} \subset \mathbb{R}^n$ corresponds to a local minimum (or maximum) of $\gamma_{1/t}(\mathcal{V}(\mathcal{L}))$ over the set of determinant-one lattices, then $\mathcal{V}(\mathcal{L})$ is in isotropic Gaussian position with parameter $1/t$, and*

$$\gamma_{1/t}(\mathcal{V}(\mathcal{L})) \geq 2/3 \,,$$

where $t := 10(\log n + 2)$.

Proof. By Theorem 2.3.1 with $f(x) = t^n \cdot e^{-\pi t^2 x}$, we have

$$\begin{aligned}\nabla_A\big(\gamma_{1/t}(\mathcal{V}(A^T\mathcal{L}))/|\det(A)|\big)|_{A=I_n} &= 2\int_{\mathcal{V}(\mathcal{L})} f'(\|\boldsymbol{x}\|^2)\boldsymbol{x}\boldsymbol{x}^T\mathrm{d}\boldsymbol{x} \\ &= -2\pi t^{n+2}\cdot\int_{\mathcal{V}(\mathcal{L})} e^{-\pi t^2\|\boldsymbol{x}\|^2}\boldsymbol{x}\boldsymbol{x}^T\mathrm{d}\boldsymbol{x}\ .\end{aligned}$$

Recall that I_n corresponds to a local extremum of a differentiable function $g(A)$ restricted to the manifold of determinant-one matrices only if $\nabla_A g(A)|_{A=I_n}$ is a scalar multiple of the identity. So, the above expression must be a multiple of the identity. I.e., $\mathcal{V}(\mathcal{L})$ is in isotropic Gaussian position. The result then follows from Theorem 2.4.2. □

Before moving to the proof of our main theorem, we need the following claim.

Claim 2.4.9. *For any* $x > 1$,

$$e^{-2\log^2 x} + e^{-2\log^2(x/(x-1))} < 1\ .$$

Proof. By symmetry, we may assume that $x \geq 2$. (Otherwise, we can replace x with $x/(x-1)$.) If $2 \leq x \leq 2.5$, then

$$e^{-2\log^2 x} + e^{-2\log^2(x/(x-1))} < e^{-2\log^2 2} + e^{-2\log^2(5/3)} < 1\ .$$

A similar computation works if $2.5 \leq x \leq e$. Finally, using the fact that $\log(x/(x-1)) = -\log(1-1/x) > 1/x$ for $x > 1$, we have for any $x \geq e$ that

$$e^{-2\log^2 x} + e^{-2\log^2(x/(x-1))} < \frac{1}{x^2} + e^{-2/x^2} < \frac{1}{x^2} + 1 - \frac{1}{x^2} = 1\ .$$ □

We now prove the main theorem of this chapter in the special case when $\mathcal{L}$ is a stable

lattice. The full result will follow as a relatively straightforward corollary.

Proposition 2.4.10 ([RS17b, Proposition 4.14])**.** *For any stable lattice* $\mathcal{L} \subset \mathbb{R}^n$*,* $\rho_{1/t}(\mathcal{L}) \le \frac{3}{2}$*, where* $t := 10(\log n + 2)$*.*

Proof. By Lemma 2.4.1, it suffices to show that $\gamma_{1/t}(\mathcal{V}(\mathcal{L})) \ge 2/3$. We assume for induction that $\gamma_{1/(10(\log d+2))}(\mathcal{V}(\mathcal{L}')) \ge 2/3$ for any stable lattice $\mathcal{L}'$ of rank $d < n$. (A quick check shows that this is true for $d = 1$.) Since the set of stable lattices is compact by Item (ii) of Proposition 2.2.2 and the function $\gamma_{1/t}(\mathcal{V}(\mathcal{L}))$ is continuous, we may assume without loss of generality that $\mathcal{L}$ corresponds to a global minimum of $\gamma_{1/t}(\mathcal{V}(\mathcal{L}))$ over the set of stable lattices. If this global minimum is also a *local* minimum over the set of determinant-one lattices, then by Theorem 2.4.8, $\gamma_{1/t}(\mathcal{V}(\mathcal{L})) \ge 2/3$, and we are done.

Otherwise, $\mathcal{L}$ lies on the boundary of the set of stable lattices. I.e., there is some primitive sublattice $\mathcal{L}' \subset \mathcal{L}$ of rank $d < n$ such that $\mathcal{L}'$ and $\mathcal{L}/\mathcal{L}'$ are stable. (See Item (iv) of Proposition 2.2.2.) By Corollary 2.2.5 together with Claim 2.2.3, we have

$$\gamma_{1/t}(\mathcal{V}(\mathcal{L})) \ge \gamma_{1/t}(\mathcal{V}(\mathcal{L}/\mathcal{L}' \oplus \mathcal{L}')) = \gamma_{1/t}(\mathcal{V}(\mathcal{L}/\mathcal{L}')) \cdot \gamma_{1/t}(\mathcal{V}(\mathcal{L}')) \,. \tag{2.6}$$

Let $t_1 := 10(\log d + 2)$ and $t_2 := 10(\log(n-d) + 2)$. By the induction hypothesis, we see that $\gamma_{1/t_1}(\mathcal{V}(\mathcal{L}')) \ge 2/3$ and $\gamma_{1/t_2}(\mathcal{V}(\mathcal{L}/\mathcal{L}')) \ge 2/3$. By Lemma 2.4.7, we therefore have

$$\gamma_{1/t}(\mathcal{V}(\mathcal{L}')) \ge 1 - \frac{1}{3} \cdot e^{-2\log^2(n/d)}, \text{ and } \gamma_{1/t}(\mathcal{V}(\mathcal{L}/\mathcal{L}')) \ge 1 - \frac{1}{3} \cdot e^{-2\log^2(n/(n-d))} \,,$$

where we have used the fact that the inradius of the Voronoi cell is $\lambda_1(\mathcal{L})/2$, which is at least $1/2$ for a stable lattice (and the constant in the exponent is very loose). Therefore,

using (2.6),

$$\begin{aligned}\gamma_{1/t}(\mathcal{V}(\mathcal{L})) &\geq \left(1 - \frac{1}{3} \cdot e^{-2\log^2(n/d)}\right) \cdot \left(1 - \frac{1}{3} \cdot e^{-2\log^2(n/(n-d))}\right) \\ &> 1 - \frac{1}{3} \cdot \left(e^{-2\log^2(n/d)} + e^{-2\log^2(n/(n-d))}\right) \\ &> \frac{2}{3}\,,\end{aligned}$$

where the last inequality follows from Claim 2.4.9 with $x := n/d$.

So, for every stable lattice $\mathcal{L}$, we have $\gamma_{1/t}(\mathcal{V}(\mathcal{L})) \geq 2/3$, and the result then follows from Lemma 2.4.1. □

We now derive our main theorem as a corollary.

Proof of Theorem 2.1.2. Let $\{\mathbf{0}\} = \mathcal{L}_0 \subset \cdots \subset \mathcal{L}_k = \mathcal{L}$ be the canonical filtration of $\mathcal{L}$. Recall from Item 2 of Proposition 2.2.2 that $\alpha_i \cdot (\mathcal{L}_i/\mathcal{L}_{i-1})$ is a stable lattice, where $\alpha_i := \det(\mathcal{L}_i/\mathcal{L}_{i-1})^{-1/\mathrm{rank}(\mathcal{L}_i/\mathcal{L}_{i-1})} \leq 1$. Therefore, by Claim 1.3.1,

$$\rho_{1/t}(\mathcal{L}) \leq \rho_{1/t}\Big(\bigoplus_{i=1}^{k} \mathcal{L}_i/\mathcal{L}_{i-1}\Big) \leq \rho_{1/t}\Big(\bigoplus_{i=1}^{k} \alpha_i(\mathcal{L}_i/\mathcal{L}_{i-1})\Big)\,.$$

By Item (iii) of Proposition 2.2.2, this direct sum of stable lattices is itself a stable lattice. The result then follows from Proposition 2.4.10. □

2.5 Bounds on $\rho_s(\mathcal{L})$ for all parameters and point-counting bounds

We first give the proof of Corollary 2.1.4, which follows immediately from Theorem 2.1.3.

Proof of Corollary 2.1.4. For any $r > 0$

$$|\mathcal{L} \cap (rB_2^n + \boldsymbol{u})| \leq e^{\pi r^2/s^2} \rho_s(\mathcal{L} - \boldsymbol{u}) \leq e^{\pi r^2/s^2} \rho_s(\mathcal{L}) \ ,$$

where the last inequality follows from Eq. (1.2) (and in particular, the observation afterwards that $\rho_s(\mathcal{L} - \boldsymbol{t}) \leq \rho_s(\mathcal{L})$ for all $\boldsymbol{t} \in \mathbb{R}^n$). Item 1 then follows by plugging in $s = 1/t$ and applying Item 1 of Theorem 2.1.3. Item 2 follows by taking $s = r\sqrt{2\pi/n}$ and applying Item 2 of Theorem 2.1.3. Finally, Item 3 follows by taking $s = r\sqrt{2\pi/n}$ and applying Item 3 of Theorem 2.1.3. □

We now prove Theorem 2.1.3, which gives bounds on the Gaussian mass for all parameters. We start with Item 1, addressing parameters $s \leq 1/t$. We actually prove a slightly stronger result than the one presented in the introduction.

Theorem 2.5.1 (Slight strengthening of Item 1 of Theorem 2.1.3)**.** *For any lattice* $\mathcal{L} \subset \mathbb{R}^n$ *with* $\det(\mathcal{L}') \geq 1$ *for all sublattices* $\mathcal{L}' \subseteq \mathcal{L}$*,*

$$\rho_s(\mathcal{L}) \leq 1 + e^{-\pi\lambda_1(\mathcal{L})^2(1/s^2 - t^2)}/2 \leq 1 + e^{-\pi(1/s^2 - t^2)}/2$$

for any $s \leq 1/t$*, where* $t := 10(\log n + 2)$*.*

Proof. Note that for any $\boldsymbol{y} \in \mathcal{L} \setminus \{\boldsymbol{0}\}$,

$$\rho_s(\boldsymbol{y}) = \rho_{1/t}(\boldsymbol{y}) \cdot e^{-\pi\|\boldsymbol{y}\|^2(1/s^2 - t^2)} \leq \rho_{1/t}(\boldsymbol{y}) e^{-\pi\lambda_1(\mathcal{L})^2(1/s^2 - t^2)} \ .$$

The result follows by summing over all $\boldsymbol{y} \in \mathcal{L} \setminus \{\boldsymbol{0}\}$ and applying Theorem 2.1.2. The second inequality uses the fact that $\lambda_1(\mathcal{L}) \geq 1$. □

We now prove the "high-parameter analogue" of Theorem 2.1.2. The proof uses Theorem 2.1.2 and duality.

Theorem 2.5.2 (Item 3 of Theorem 2.1.3)**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ with $\det(\mathcal{L}') \geq 1$ for all sublattices $\mathcal{L}' \subseteq \mathcal{L}$ and any parameter $s \geq t$, $\rho_s(\mathcal{L}) \leq 2s^n$ where $t := 10(\log n + 2)$.*

Proof. Recall the Poisson Summation Formula applied to the Gaussian mass (Eq. (1.1)):

$$\rho_s(\mathcal{L}) = \frac{s^n}{\det(\mathcal{L})} \cdot \rho_{1/s}(\mathcal{L}^*) \ .$$

Assume first that $\mathcal{L}$ is stable. Then, by Theorem 2.1.2 and the fact that the dual of a stable lattice is stable (Item (i) of Proposition 2.2.2),

$$\rho_s(\mathcal{L}) = s^n \cdot \rho_{1/s}(\mathcal{L}^*) \leq s^n \cdot \rho_{1/t}(\mathcal{L}^*) \leq 2s^n \ .$$

For a general lattice $\mathcal{L} \subset \mathbb{R}^n$, let $\{\mathbf{0}\} = \mathcal{L}_0 \subset \mathcal{L}_1 \subset \cdots \subset \mathcal{L}_k = \mathcal{L}$ be the canonical filtration of $\mathcal{L}$. Recall that $\alpha_i(\mathcal{L}_i/\mathcal{L}_{i-1})$ is stable for some $\alpha_i \leq 1$. (See Item 2 of Proposition 2.2.2.) Then, by Claim 1.3.1,

$$\rho_s(\mathcal{L}) \leq \rho_s\Big(\bigoplus \mathcal{L}_i/\mathcal{L}_{i-1}\Big) \leq \rho_s\Big(\bigoplus \alpha_i \cdot \mathcal{L}_i/\mathcal{L}_{i-1}\Big) \leq 2s^n \ ,$$

where the last inequality follows from the fact that the direct sum of stable lattices is stable together with the bound proven above for stable lattices. (See Item (iii) of Proposition 2.2.2.)

□

The rest of this section is dedicated to the proof of Item 2 of Theorem 2.1.3. Note that we already have a bound on $\rho_s(\mathcal{L})$ for $s \leq 1/t$ and for $s \geq t$, but we currently have no non-trivial bound for intermediate parameters $1/t < s < t$. To remedy this, we show in Theorem 2.5.5 below that $\rho_{e^\sigma}(\mathcal{L})$ is "approximately log-convex," which allows us to interpolate between these two bounds. In the proof of Theorem 2.5.5, we are unable to work with $\rho_{e^\sigma}(\mathcal{L})$ directly, so we instead show that it can be approximated by $\gamma_{e^\sigma}(\mathcal{V}(\mathcal{L}))$ (Lemma 2.5.4). We then notice

that the latter function is log-concave by Theorem 2.4.5.

Claim 2.5.3. *For any lattice $\mathcal{L} \subset \mathbb{R}^n$, $\boldsymbol{y} \in \mathcal{L}$, and $s > 0$,*

$$\rho_s(\boldsymbol{y})\gamma_s(\mathcal{V}(\mathcal{L})) \le \gamma_s(\mathcal{V}(\mathcal{L}) + \boldsymbol{y}) \le \gamma_s(\mathcal{V}(\mathcal{L}))$$

Proof. By scaling appropriately, we may assume that $s = 1$. We have

$$\begin{aligned}
\gamma(\mathcal{V}(\mathcal{L}) + \boldsymbol{y}) &= \int_{\mathcal{V}(\mathcal{L})} e^{-\pi\|\boldsymbol{x}+\boldsymbol{y}\|^2} \mathrm{d}\boldsymbol{x} \\
&= \rho(\boldsymbol{y}) \int_{\mathcal{V}(\mathcal{L})} \rho(\boldsymbol{x}) e^{-2\pi\langle \boldsymbol{y}, \boldsymbol{x}\rangle} \mathrm{d}\boldsymbol{x} \\
&= \rho(\boldsymbol{y}) \int_{\mathcal{V}(\mathcal{L})} \rho(\boldsymbol{x}) \cosh(2\pi\langle \boldsymbol{y}, \boldsymbol{x}\rangle) \mathrm{d}\boldsymbol{x} \ ,
\end{aligned}$$

where we have used the symmetry of the Voronoi cell in the last line. The lower bound now follows from noting that $\cosh(2\pi\langle \boldsymbol{x}, \boldsymbol{y}\rangle) \ge 1$. For the upper bound, we recall that, by definition, any vector in the Voronoi cell $\boldsymbol{x} \in \mathcal{V}(\mathcal{L})$ satisfies $\langle \boldsymbol{y}, \boldsymbol{x}\rangle \le \|\boldsymbol{y}\|^2/2$ for any lattice vector $\boldsymbol{y} \in \mathcal{L}$. Therefore, $\cosh(2\pi\langle \boldsymbol{y}, \boldsymbol{x}\rangle) \le \cosh(\pi\|\boldsymbol{y}\|^2) \le 1/\rho(\boldsymbol{y})$, as needed. □

Lemma 2.5.4. *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and any $s > 0$,*

$$e^{-4n}/2 \le \gamma_s(\mathcal{V}(\mathcal{L}))\rho_s(\mathcal{L}) \le 1 \ .$$

Proof. The upper bound is Lemma 2.4.1, repeated for comparison. By scaling appropriately, we may assume that $s = 1$. Recall that $\int_{\mathbb{R}^n} \|\boldsymbol{x}\|^2 e^{-\pi\|\boldsymbol{x}\|^2} \mathrm{d}\boldsymbol{x} = n/(2\pi)$. It follows from Markov's inequality that $\int_{\sqrt{n/\pi}B_2^n} e^{-\pi\|\boldsymbol{x}\|^2} \mathrm{d}\boldsymbol{x} \ge 1/2$. Let

$$Y := \{\boldsymbol{y} \in \mathcal{L} \ : \ (\mathcal{V}(\mathcal{L}) + \boldsymbol{y}) \cap \sqrt{n/\pi}B_2^n \ne \emptyset\} \ .$$

I.e., Y is the set of vectors $\boldsymbol{y} \in \mathcal{L}$ such that there exists some $\boldsymbol{x} \in \sqrt{n/\pi}B_2^n$ with $\|\boldsymbol{y} - \boldsymbol{x}\| \le$

$\|\boldsymbol{y}' - \boldsymbol{x}\|$ for every $\boldsymbol{y}' \in \mathcal{L}$. By taking $\boldsymbol{y}' = \boldsymbol{0}$, we immediately see that $Y \subseteq \mathcal{L} \cap 2\sqrt{n/\pi}B_2^n$. Recalling that the Voronoi cell tiles space, we have

$$\begin{aligned}
1/2 &\le \int_{\sqrt{n/\pi}B_2^n} e^{-\pi\|\boldsymbol{x}\|^2} \mathrm{d}\boldsymbol{x} \\
&\le \sum_{\boldsymbol{y}\in Y} \gamma(\mathcal{V}(\mathcal{L}) + \boldsymbol{y}) \\
&\le |Y| \cdot \gamma(\mathcal{V}(\mathcal{L})) && \text{(Claim 2.5.3)} \\
&\le |\mathcal{L} \cap 2\sqrt{n/\pi}B_2^n| \cdot \gamma(\mathcal{V}(\mathcal{L})) \\
&\le e^{4n}\rho(\mathcal{L})\gamma(\mathcal{V}(\mathcal{L})) \ ,
\end{aligned}$$

as needed. □

We now prove the "approximate log-convexity" of $\rho_{e^\sigma}(\mathcal{L})$.

Theorem 2.5.5 ([RS17b, Theorem 5.5])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and any $t_1 > s > t_2 > 0$,*

$$\rho_s(\mathcal{L}) \le 2e^{4n}\rho_{t_1}(\mathcal{L})^\tau \rho_{t_2}(\mathcal{L})^{1-\tau} \ ,$$

where $\tau := \log(s/t_2)/\log(t_1/t_2)$.

Proof. We have

$$\begin{aligned}
\rho_s(\mathcal{L}) &\le \frac{1}{\gamma_s(\mathcal{V}(\mathcal{L}))} && \text{(Lemma 2.4.1)} \\
&\le \frac{1}{\gamma_{t_1}(\mathcal{V}(\mathcal{L}))^\tau \gamma_{t_2}(\mathcal{V}(\mathcal{L}))^{1-\tau}} && \text{(Theorem 2.4.5)} \\
&\le 2e^{4n}\rho_{t_1}(\mathcal{L})^\tau \rho_{t_2}(\mathcal{L})^{1-\tau} && \text{(Lemma 2.5.4)} \ ,
\end{aligned}$$

as needed. □

Corollary 2.5.6 (Item 2 of Theorem 2.1.3)**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ with $\det(\mathcal{L}') \geq 1$ for all $\mathcal{L}' \subseteq \mathcal{L}$ and any parameter $1/t < s < t$, we have*

$$\rho_s(\mathcal{L}) \leq 4(e^8 st)^{n/2} ,$$

where $t := 10(\log n + 2)$.

Proof. Let $\tau := (1 - \log s/\log t)/2$. Then,

$$\begin{aligned}
\rho_s(\mathcal{L}) &\leq 2e^{4n}\rho_{1/t}(\mathcal{L})^{\tau} \cdot \rho_t(\mathcal{L})^{1-\tau} && \text{(Theorem 2.5.5)} \\
&\leq 2^{1+\tau}e^{4n}\rho_t(\mathcal{L})^{1-\tau} && \text{(Theorem 2.1.2)} \\
&\leq 4e^{4n}t^{(1-\tau)n} && \text{(Corollary 2.5.2)} \\
&= 4(e^8 st)^{n/2} ,
\end{aligned}$$

as needed. □

2.6 Proof of the covering radius approximation

We now note that Theorem 2.1.2 (together with Corollary 1.3.7) immediately implies a bound on the covering radius of stable lattices.

Theorem 2.6.1 ([RS17b, Theorem 6.2])**.** *For any stable lattice $\mathcal{L} \subset \mathbb{R}^n$,*

$$\mu(\mathcal{L}) \leq 4\sqrt{n}(\log n + 10) .$$

Proof. Let $t := 10(\log n + 2)$. Since $\mathcal{L}^*$ is also stable (by Item (i) of Proposition 2.2.2),

Theorem 2.1.2 implies that $\rho_{1/t}(\mathcal{L}^*) \le 3/2$. Applying Corollary 1.3.7, we have

$$\mu(\mathcal{L}) \le (\sqrt{n/(2\pi)} + 1) \cdot t < 4\sqrt{n}(\log n + 10) \ ,$$

as needed. □

Next, we show (Proposition 2.6.3) how to reduce the case of general lattices to the stable case. We will need the following technical lemma, whose proof can be found in [RS17b].

Lemma 2.6.2 (Reverse AM-GM)**.** *Let $0 < a_1 < \cdots < a_k$ and $d_1, \ldots, d_k \in \mathbb{N}$, and for $j = 1, \ldots, k$, define $m_j := \sum_{i \ge j} d_i$. Then,*

$$\sum_{i=1}^{k} d_i a_i \le 2e \cdot \lceil \log(2m_1) \rceil \cdot \max_j m_j \Big(\prod_{i \ge j} a_i^{d_i} \Big)^{1/m_j} \ .$$

Recall that

$$\mu_{\det}(\mathcal{L}) := \max_{W \subset \mathbb{R}^n} \sqrt{\dim(W^\perp)} \cdot \det(\pi_{W^\perp}(\mathcal{L}))^{\frac{1}{\dim(W^\perp)}} \ ,$$

where the maximum is over lattice subspaces $W \subset \mathbb{R}^n$ of $\mathcal{L}$ (i.e., subspaces W spanned by up to $n-1$ lattice vectors).

Proposition 2.6.3 ([RS17b, Proposition 6.4])**.** *Let*

$$C_{\mathrm{KL}} := \max_{d \le n} \sup \mu(\mathcal{L})/\sqrt{d} \ ,$$

where the supremum is over stable lattices $\mathcal{L} \subset \mathbb{R}^d$. Then, for any lattice $\mathcal{L} \subset \mathbb{R}^n$,

$$\mu(\mathcal{L}) \le \sqrt{2e\lceil \log(2n) \rceil} \cdot C_{\mathrm{KL}} \cdot \mu_{\det}(\mathcal{L}) \ .$$

Proof. Let $\{\mathbf{0}\} = \mathcal{L}_0 \subset \mathcal{L}_1 \subset \cdots \subset \mathcal{L}_k = \mathcal{L}$ be the canonical filtration of some lattice $\mathcal{L} \subset \mathbb{R}^n$. Let $d_i := \mathrm{rank}(\mathcal{L}_i/\mathcal{L}_{i-1})$. Note that $\mathcal{L}_i/\mathcal{L}_{i-1}$ is a scaling of a stable lattice, i.e.,

$\det(\mathcal{L}_i/\mathcal{L}_{i-1})^{-1/d_i} \cdot (\mathcal{L}_i/\mathcal{L}_{i-1})$ is stable. (See Item 2 of Proposition 2.2.2.) We therefore have by Claim 2.2.3 and Lemma 2.2.4 that

$$\begin{aligned}\mu(\mathcal{L})^2 &\leq \mu\Big(\bigoplus_i \mathcal{L}_i/\mathcal{L}_{i-1}\Big)^2 \\ &= \sum_i \mu(\mathcal{L}_i/\mathcal{L}_{i-1})^2 \\ &\leq C_{\mathrm{KL}}^2 \cdot \sum_i d_i \det(\mathcal{L}_i/\mathcal{L}_{i-1})^{2/d_i} \ . \end{aligned} \tag{2.7}$$

Next, we recall from Item 3 of Proposition 2.2.2 that $a_i := \det(\mathcal{L}_i/\mathcal{L}_{i-1})^{2/d_i}$ is an increasing sequence, and we note that $\sum_{i\geq j} d_i = \operatorname{rank}(\mathcal{L}/\mathcal{L}_{j-1})$. We may therefore use Lemma 2.6.2 to bound Eq. (2.7) from above by

$$\begin{aligned}&2e\lceil\log(2n)\rceil \cdot C_{\mathrm{KL}}^2 \cdot \max_i \operatorname{rank}(\mathcal{L}/\mathcal{L}_i) \cdot \det(\mathcal{L}/\mathcal{L}_i)^{\frac{2}{\operatorname{rank}(\mathcal{L}/\mathcal{L}_i)}} \\ &\quad\leq 2e\lceil\log(2n)\rceil \cdot C_{\mathrm{KL}}^2 \max_{W\subset\mathbb{R}^n} \dim(W^\perp) \cdot \det(\pi_{W^\perp}(\mathcal{L}))^{\frac{2}{\dim(W^\perp)}} \ ,\end{aligned}$$

as needed. □

Theorem 2.1.5 now follows as an immediate corollary of the above results. In particular, we have $C_{\mathrm{KL}} \leq 4(\log n + 10)$ and therefore $\sqrt{2e\lceil\log(2n)\rceil} \cdot C_{\mathrm{KL}} \leq 10(\log n + 10)^{3/2}$. The result then follows from Proposition 2.6.3.

2.6.1 Connection with the Slicing Conjecture

In this section, we prove Theorem 2.6.7. The structure of the proof is based on the one suggested in [SW16], as was the case for the proof of our main theorem in Section 2.4.

As in Section 2.4, we are unable to work with the lattice parameter $\mu(\mathcal{L})$ that interests us

directly.[7] Instead, we work with the lattice parameter

$$\overline{\mu}(\mathcal{L}) := \sqrt{\frac{1}{\det(\mathcal{L})} \int_{\mathcal{V}(\mathcal{L})} \|\boldsymbol{x}\|^2 \mathrm{d}\boldsymbol{x}} \,,$$

which, as shown in Lemma 2.6.4 below, gives a good approximation to μ. We remark that the parameter $\overline{\mu}$ and various closely related parameters have been studied extensively (e.g., [ZF96, CS98, GMR05, HLR09]).

Lemma 2.6.4 ([HLR09, Claim 3.1])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$,*

$$\overline{\mu}(\mathcal{L}) \leq \mu(\mathcal{L}) \leq 2\overline{\mu}(\mathcal{L}) \,.$$

Proof. Let $\boldsymbol{t} \in \mathbb{R}^n$ be such that $\mathrm{dist}(\boldsymbol{t}, \mathcal{L}) = \mu(\mathcal{L})$. I.e., $\boldsymbol{t}$ is a "deep hole." For any $\boldsymbol{x} \in \mathbb{R}^n$, we have

$$\mu(\mathcal{L}) = \mathrm{dist}(\boldsymbol{t}, \mathcal{L}) \leq \mathrm{dist}(\boldsymbol{x}, \mathcal{L}) + \mathrm{dist}(\boldsymbol{x} + \boldsymbol{t}, \mathcal{L})$$

(since by the triangle inequality, for all $\boldsymbol{y}, \boldsymbol{z} \in \mathcal{L}$, $\|\boldsymbol{x} - \boldsymbol{y}\| + \|\boldsymbol{x} + \boldsymbol{t} - \boldsymbol{z}\| \geq \|\boldsymbol{t} - (\boldsymbol{z} - \boldsymbol{y})\|$). Integrating, we have

$$\mu(\mathcal{L}) \leq \frac{1}{\det(\mathcal{L})} \int_{\mathcal{V}(\mathcal{L})} \mathrm{dist}(\boldsymbol{x}, \mathcal{L}) \mathrm{d}\boldsymbol{x} + \frac{1}{\det(\mathcal{L})} \int_{\mathcal{V}(\mathcal{L})} \mathrm{dist}(\boldsymbol{x} + \boldsymbol{t}, \mathcal{L}) \mathrm{d}\boldsymbol{x} = \frac{2}{\det(\mathcal{L})} \int_{\mathcal{V}(\mathcal{L})} \|\boldsymbol{x}\| \mathrm{d}\boldsymbol{x} \,,$$

where we have simply observed that the integral is invariant under shifts (and that, for $\boldsymbol{x} \in \mathcal{V}(\mathcal{L})$, $\mathrm{dist}(\boldsymbol{x}, \mathcal{L}) = \|\boldsymbol{x}\|$ by definition). The result then follows by Jensen's inequality, which in particular tells us that

$$\Big(\frac{1}{\det(\mathcal{L})} \cdot \int_{\mathcal{V}(\mathcal{L})} \|\boldsymbol{x}\| \mathrm{d}\boldsymbol{x}\Big)^2 \leq \frac{1}{\det(\mathcal{L})} \cdot \int_{\mathcal{V}(\mathcal{L})} \|\boldsymbol{x}\|^2 \mathrm{d}\boldsymbol{x} = \overline{\mu}(\mathcal{L})^2 \,. \qquad \square$$

[7] While [DSV12] give a characterization of lattices corresponding to local maxima of μ, we are unable to obtain a sufficiently strong bound on the covering radius of these lattices. See [SW16] for more about this question.

We now observe that Theorem 2.3.1 is applicable to the function $\overline{\mu}(\mathcal{L})^2$. Recall that a symmetric convex body $K \subset \mathbb{R}^n$ is said to be *isotropic* if $\int_K \boldsymbol{x}\boldsymbol{x}^T \mathrm{d}\boldsymbol{x} = \alpha \cdot I_n$ for some scalar $\alpha > 0$.

Proposition 2.6.5 ([RS17b, Proposition 6.6])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$,*

$$\nabla_A \overline{\mu}(A^T\mathcal{L})^2|_{A=I_n} = \frac{2}{\det(\mathcal{L})} \int_{\mathcal{V}(\mathcal{L})} \boldsymbol{x}\boldsymbol{x}^T \mathrm{d}\boldsymbol{x} \ ,$$

where $A \in \mathrm{GL}_n(\mathbb{R})$ ranges over non-singular matrices. In particular, if $\mathcal{L}$ corresponds to a local maximum (or local minimum) of $\overline{\mu}(\mathcal{L})$ over the set of determinant-one lattices, then $\mathcal{V}(\mathcal{L})$ is isotropic.

Proof. To compute the gradient, we simply apply Theorem 2.3.1 with $f(x) := x$, and recall that

$$\overline{\mu}(A^T\mathcal{L})^2 = \frac{1}{\det(\mathcal{L})} \cdot \frac{1}{|\det(A)|} \int_{\mathcal{V}(A^T\mathcal{L})} f(\|\boldsymbol{x}\|^2) \mathrm{d}\boldsymbol{x} \ .$$

The "in particular" follows from the fact that a differentiable function $g(A)$ restricted to the set of determinant-one matrices has a critical point at $A = I_n$ if and only if $\nabla_A g(A)|_{A=I_n}$ is a scalar multiple of the identity. □

We define the (symmetric) isotropic constant

$$L_n^2 := \max_{d \le n} \frac{1}{d} \cdot \sup_K \int_K \|\boldsymbol{x}\|^2 \mathrm{d}\boldsymbol{x} \ ,$$

where the supremum is taken over all isotropic symmetric convex bodies $K \subset \mathbb{R}^d$ of volume one. It is known to satisfy $1/(2\sqrt{3}) \le L_n \le Cn^{1/4}$, and the Slicing Conjecture implies that L_n is bounded by a universal constant [Bou91, Kla06]. (The lower bound is due to the hypercube, $[-1/2, 1/2]^n$.) We note in passing that we are only concerned with the isotropic constant for Voronoi cells, which could conceivably be easier to bound than the isotropic

constant for arbitrary convex bodies.

Theorem 2.6.6. *[RS17b, Theorem 6.7] For any stable lattice $\mathcal{L} \subset \mathbb{R}^n$,*

$$\mu(\mathcal{L}) \leq 2\overline{\mu}(\mathcal{L}) \leq 2\sqrt{n}L_n \ .$$

Proof. By Lemma 2.6.4, it suffices to prove that $\overline{\mu}(\mathcal{L}) \leq \sqrt{n}L_n$. Note that this is trivially true for $n = 1$. We assume for induction that $\overline{\mu}(\mathcal{L}') \leq \sqrt{d}L_d \leq \sqrt{d}L_n$ for all stable lattices $\mathcal{L}'$ of rank $d < n$. Recall that the set of stable lattices is compact (Item (ii) of Proposition 2.2.2), so that we may assume without loss of generality that $\mathcal{L}$ corresponds to a global maximum of the function $\overline{\mu}$ over this set. If this is also a *local* maximum over the set of determinant-one lattices, then by Proposition 2.6.5, the Voronoi cell is isotropic, and we have $\overline{\mu}(\mathcal{L}) \leq \sqrt{n}L_n$ by the definition of $\overline{\mu}$ and L_n. Otherwise, $\mathcal{L}$ must lie on the boundary of the set of stable lattices. I.e., there is some primitive sublattice $\mathcal{L}' \subset \mathcal{L}$ of rank $0 < d < n$ such that $\mathcal{L}'$ and $\mathcal{L}/\mathcal{L}'$ are both stable. (See Item (iv) of Proposition 2.2.2.) Applying the induction hypothesis and Corollary 2.2.5 (together with Claim 2.2.3), we have

$$\overline{\mu}(\mathcal{L})^2 \leq \overline{\mu}(\mathcal{L}' \oplus \mathcal{L}/\mathcal{L}')^2 = \overline{\mu}(\mathcal{L}')^2 + \overline{\mu}(\mathcal{L}/\mathcal{L}')^2 \leq dL_n^2 + (n-d)L_n^2 = nL_n^2 \ ,$$

as needed. □

As far as we know, it is entirely possible that $L_n = 1/(2\sqrt{3})$, i.e., that the hypercube $[-1/2, 1/2]^n$ is the worst symmetric body for the Slicing Conjecture. If true, this would imply that for any stable $\mathcal{L} \subset \mathbb{R}^n$, $\mu(\mathcal{L}) \leq 2\overline{\mu}(\mathcal{L}) \leq \sqrt{n/3}$. Moreover, it is possible that the constant 2 in Lemma 2.6.4 can be replaced with $\sqrt{3}$, which would be tight for $\mathbb{Z}^n$ (this was already mentioned in [HLR09, Conjecture 1.3]). If this is also true, then we get that for any stable $\mathcal{L} \subset \mathbb{R}^n$, $\mu(\mathcal{L}) \leq \sqrt{3}\overline{\mu}(\mathcal{L}) \leq \sqrt{n}/2$, which is tight for $\mathbb{Z}^n$. Apart from being an interesting statement in its own right, it was shown by Shapira and Weiss [SW16] that such

a result would imply the so-called Minkowski conjecture (see there for more information).

We can now use Proposition 2.6.3 to extend Theorem 2.6.6 to all lattices $\mathcal{L} \subset \mathbb{R}^n$.

Theorem 2.6.7. *[RS17b, Theorem 6.8] For any lattice $\mathcal{L} \subset \mathbb{R}^n$,*

$$\mu(\mathcal{L}) \leq 10\sqrt{\log n + 1} \cdot L_n \cdot \mu_{\det}(\mathcal{L}) \ .$$

As we observed in Footnote 5, there are lattices with $\mu(\mathcal{L}) \geq C\sqrt{\log n} \cdot \mu_{\det}(\mathcal{L})$. So, Theorem 2.6.7 is tight up to a constant, assuming the Slicing Conjecture. (We made no attempt to optimize the constant in Theorem 2.6.7.)

2.7 An optimal bound for extreme parameters

We now prove Theorem 2.1.6, which says that $\mathbb{Z}^n$ has maximal Gaussian mass amongst all lattices $\mathcal{L}$ with $\det(\mathcal{L}') \geq 1$ for all sublattices $\mathcal{L}' \subseteq \mathcal{L}$, for very small parameters $s \leq \sqrt{2\pi/(n+2)}$ *and* for very large parameters $s \geq \sqrt{(n+2)/(2\pi)}$. The proof is similar to that of Theorem 2.1.2, except here we work directly with $\rho_s(\mathcal{L})$ (instead of the proxy $\gamma_s(\mathcal{V}(\mathcal{L}))$). Moreover, we show that $\rho_s(\mathcal{L})$ has *no stable local maxima* for those values of s, which leads to a simpler proof and the clearly tight result. In order to show that local maxima do not exist, we will show that the Laplacian of $\rho_s(\mathcal{L})$ is always positive when $\mathcal{L}$ is stable.

In more detail, for a lattice $\mathcal{L}$ and $s > 0$, let $f_{\mathcal{L},s} : X \to \mathbb{R}$ be given by

$$f_{\mathcal{L},s}(A) := \rho_s(e^{A/2}\mathcal{L}) = \sum_{\boldsymbol{y} \in \mathcal{L}} e^{-\pi \boldsymbol{y}^T e^A \boldsymbol{y}/s^2} \ ,$$

where $X \subset \mathbb{R}^{n \times n}$ is the linear space of all symmetric matrices with zero trace. Notice that as A ranges over X, $e^A := I_n + \sum_{i=1}^{\infty} A^i/i!$ ranges over all determinant-one positive-definite matrices. (In particular, $e^{A/2}\mathcal{L}$ ranges over all lattices of fixed determinant, up to orthogonal

transformations.) See [Ter16, Section 1.1.3] for a more in-depth treatment of the space of determinant-one matrices.

Recall that the *Laplacian* of a twice differentiable function $g : X \to \mathbb{R}$ is given by

$$\Delta_X g(A) := \sum_i \frac{\partial^2}{\partial E_i^2} g(A) \ ,$$

where the E_i form an orthonormal basis of X, and

$$\frac{\partial^2}{\partial M^2} g(A) := \frac{\partial^2}{\partial r^2} g(A + rM)|_{r=0}$$

is the directional second derivative of g in the M direction. One can show that the Laplacian does not depend on the choice of basis. Clearly, if the Laplacian is positive at A, then A cannot correspond to a local maximum of g, since there must be at least one direction in which the second derivative is positive.

The Laplacian of $f_{\mathcal{L},s}$ is straightforward (but tedious) to calculate. It can be found, e.g., in the work by Sarnak and Strömbergsson [SS06] who used it to study local minima of $\rho_s(\mathcal{L})$.

Claim 2.7.1 ([SS06, Eq. (46)])**.** *Let $X \subset \mathbb{R}^{n \times n}$ be the space of trace-zero symmetric matrices. Then, for any lattice $\mathcal{L} \subset \mathbb{R}^n$ and any parameter $s > 0$,*

$$\Delta_X f_{\mathcal{L},s}(0) = \frac{\pi}{s^2} \cdot \frac{n-1}{n} \cdot \sum_{\boldsymbol{y} \in \mathcal{L}} \rho_s(\boldsymbol{y}) \|\boldsymbol{y}\|^2 \Big(\frac{\pi}{s^2} \cdot \|\boldsymbol{y}\|^2 - \frac{n+2}{2} \Big) \ .$$

Proposition 2.7.2 ([RS17b, Proposition 7.2])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and*

$$0 < s \leq \sqrt{\frac{2\pi}{n+2}} \cdot \lambda_1(\mathcal{L}) \ ,$$

$\mathcal{L}$ cannot correspond to a local maximum of $\rho_s(\mathcal{L})$ over the set of determinant-one lattices. In particular, since stable lattices have $\lambda_1(\mathcal{L}) \geq 1$, a stable lattice cannot correspond to a local

maximum for $s \leq \sqrt{2\pi/(n+2)}$.

Proof. It suffices to show that the Laplacian given in Claim 2.7.1 is positive for such $\mathcal{L}$. Indeed, the summand is zero for $\boldsymbol{y} = \mathbf{0}$, and since

$$\frac{\pi}{s^2} \cdot \lambda_1(\mathcal{L})^2 \geq \frac{n+2}{2} \ ,$$

the summand is non-negative for all non-zero $\boldsymbol{y} \in \mathcal{L}$. Finally, since any lattice contains vectors of arbitrarily large length, there must be some strictly positive terms in the sum. Therefore, the full sum is strictly positive, as needed. □

From this, we derive our main result for the special case of stable lattices.

Proposition 2.7.3 ([RS17b, Proposition 7.3])**.** *For any* $0 < s \leq \sqrt{2\pi/(n+2)}$ *and stable lattice* $\mathcal{L} \subset \mathbb{R}^n$, $\rho_s(\mathcal{L}) \leq \rho_s(\mathbb{Z}^n)$.

Proof. Note that the result is trivial for $n = 1$. We assume for induction that the result holds for all dimensions less than n. Since the set of stable lattices is compact and $\rho_s(\mathcal{L})$ is a continuous function, we may assume that $\mathcal{L}$ corresponds to a global maximum of $\rho_s(\mathcal{L})$ over the set of stable lattices. By Proposition 2.7.2, this cannot be a local maximum over the set of determinant-one lattices. So, $\mathcal{L}$ must be on the boundary of the set of stable lattices. I.e., there is a non-trivial primitive sublattice $\mathcal{L}' \subset \mathcal{L}$ with $d := \mathrm{rank}(\mathcal{L}')$ such that $\mathcal{L}'$ and $\mathcal{L}/\mathcal{L}'$ are themselves stable lattices of rank strictly less than n. (See Item (iv) of Proposition 2.2.2.) Applying the induction hypothesis, we have by Claim 1.3.1 that

$$\rho_s(\mathcal{L}) \leq \rho_s(\mathcal{L}') \cdot \rho_s(\mathcal{L}/\mathcal{L}') \leq \rho_s(\mathbb{Z}^d) \cdot \rho_s(\mathbb{Z}^{n-d}) = \rho_s(\mathbb{Z}^n) \ ,$$

where we have used the fact that $s \leq \sqrt{2\pi/(n+2)} \leq \min\{\sqrt{2\pi/(d+2)},\ \sqrt{2\pi/(n-d+2)}\}$ in order to apply the induction hypothesis. □

We now "invert the parameter" using duality.

Corollary 2.7.4 ([RS17b, Corollary 7.4])**.** *For any $s \geq \sqrt{(n+2)/(2\pi)}$ and stable lattice $\mathcal{L} \subset \mathbb{R}^n$, $\rho_s(\mathcal{L}) \leq \rho_s(\mathbb{Z}^n)$.*

Proof. Recall that the dual $\mathcal{L}^*$ of a stable lattice is itself stable. (See Item (i) of Proposition 2.2.2.) Furthermore, by the Poisson Summation Formula for the discrete Gaussian (Eq. (1.1)),

$$\rho_s(\mathcal{L}) = \frac{s^n}{\det(\mathcal{L})} \cdot \rho_{1/s}(\mathcal{L}^*) \leq \frac{s^n}{\det(\mathcal{L})} \cdot \rho_{1/s}(\mathbb{Z}^n) = \rho_s(\mathbb{Z}^n) \ ,$$

as needed, where the inequality follows from Proposition 2.7.3, and the last equality follows from the Poisson Summation Formula applied to $\mathbb{Z}^n$. □

We can now prove Theorem 2.1.6.

Proof of Theorem 2.1.6. Let $\{\mathbf{0}\} = \mathcal{L}_0 \subset \mathcal{L}_1 \subset \cdots \subset \mathcal{L}_k = \mathcal{L}$ be the canonical filtration of $\mathcal{L}$, and let $d_i := \operatorname{rank}(\mathcal{L}_i/\mathcal{L}_{i-1}) \leq n$. Then, by Claim 1.3.1, we have

$$\rho_s(\mathcal{L}) \leq \prod_i \rho_s(\mathcal{L}_i/\mathcal{L}_{i-1}) \ .$$

Note that, if $s \leq \sqrt{2\pi/(n+2)}$, then we also have $s \leq \sqrt{2\pi/(d_i+2)}$ for all i. And, $\alpha_i \cdot (\mathcal{L}_i/\mathcal{L}_{i-1})$ is a stable lattice for some $\alpha_i \leq 1$. (See Item 2 of Proposition 2.2.2.) So, in this case we may apply Proposition 2.7.3 to obtain

$$\rho_s(\mathcal{L}) \leq \prod_i \rho_s(\alpha_i \cdot (\mathcal{L}_i/\mathcal{L}_{i-1})) \leq \prod_i \rho_s(\mathbb{Z}^{d_i}) = \rho_s(\mathbb{Z}^n) \ .$$

If, on the other hand, $s \geq \sqrt{(n+2)/(2\pi)}$, then $s \geq \sqrt{(d_i+2)/(2\pi)}$ for all i, so we may similarly apply Corollary 2.7.4 to obtain the same result. □

Remark. *It is possible to show that, in the setting of Theorem 2.1.6, $\rho_s(\mathcal{L}) = \rho_s(\mathbb{Z}^n)$ if and only if $\mathcal{L}$ is an orthogonal transformation of $\mathbb{Z}^n$. To see this, first notice that in order to get equality, all the α_i in the proof above must be one, i.e., $\mathcal{L}$ must be stable. Next, we follow the induction argument in the proof of Proposition 2.7.3, and recall the case of equality in Lemma 1.3.1.*

2.8 Tightness of our bounds

In this section, we discuss the tightness of our bounds by considering some classes of lattices $\mathcal{L} \subset \mathbb{R}^n$.

2.8.1 Tightness of Item 3 of Theorem 2.1.3 for stable lattices

It is an immediate consequence of the Poisson Summation Formula (Eq. (1.1)) that $\rho_s(\mathcal{L}) \geq s^n / \det(\mathcal{L})$ for any $s > 0$ and $\mathcal{L} \subset \mathbb{R}^n$. Combining this with Item 3 of Theorem 2.1.3, we see that

$$s^n \leq \rho_s(\mathcal{L}) \leq 2s^n$$

for any *stable* lattice $\mathcal{L} \subset \mathbb{R}^n$ and any $s \geq 10(\log n + 2)$. I.e., Item 3 of Theorem 2.1.3 is tight for all stable lattices up to a factor of two in the mass.

2.8.2 The integer lattice $\mathbb{Z}^n$

We first prove bounds on the Gaussian mass of $\mathbb{Z}^n$. In particular, the lower bound in Eq. (2.8) below shows that $\rho_{\sqrt{\pi/\log n}}(\mathbb{Z}^n) \geq 3/2$, so that Theorem 2.1.2 is tight for $\mathbb{Z}^n$ up to a factor of $C\sqrt{\log n}$ in t. Similar bounds hold for Items 1 and 2 of Theorem 2.1.3.

Claim 2.8.1. *For any $n \geq 1$ and parameter $s > 0$,*

$$\left(1 + 2e^{-\pi/s^2}\right)^n \leq \rho_s(\mathbb{Z}^n) \leq \left(1 + (2+s)e^{-\pi/s^2}\right)^n , \tag{2.8}$$

and

$$s^n \cdot \left(1 + 2e^{-\pi s^2}\right)^n \leq \rho_s(\mathbb{Z}^n) \leq s^n \cdot \left(1 + (2 + 1/s)e^{-\pi s^2}\right)^n . \tag{2.9}$$

Proof. Note that $\rho_s(\mathbb{Z}^n) = \rho_s(\mathbb{Z})^n$. So, it suffices to bound $\rho_s(\mathbb{Z})$. Furthermore, Eq. (2.9) follows from Eq. (2.8) and the Poisson Summation Formula (Eq. (1.1)). So, it suffices to prove Eq. (2.8) for the case $n = 1$. For the lower bound, we have

$$\rho_s(\mathbb{Z}) = 1 + 2\sum_{z=1}^{\infty} e^{-\pi z^2/s^2} \geq 1 + 2e^{-\pi/s^2} .$$

For the upper bound, we write

$$\rho_s(\mathbb{Z}) = 1 + 2e^{-\pi/s^2} + 2\sum_{z=2}^{\infty} e^{-\pi z^2/s^2} \leq 1 + 2e^{-\pi/s^2} + 2\int_1^{\infty} e^{-\pi x^2/s^2} \mathrm{d}x \leq 1 + (2+s)e^{-\pi/s^2} ,$$

where we have used [AS64, Eq. 7.1.13] to bound the error function. □

We now bound $|\mathbb{Z}^n \cap rB_2^n|$. Note that the lower bound in the next claim, which shows that $|\mathbb{Z}^n \cap rB_2^n| \geq e^{Cr^2 \log(n/r^2)}$ for $r \leq \sqrt{n}$, is relatively close to the upper bound $|\mathbb{Z}^n \cap rB_2^n| \leq e^{C'r^2 \log^2 n}$ given by Item 1 of Corollary 2.1.4. (We include a better upper bound on $|\mathbb{Z}^n \cap rB_2^n|$ below for completeness. In particular, the two bounds match up to a factor of $2^{o(r^2)}$ for $\omega(1) < r < o(\sqrt{n})$. See [MO90] for tighter bounds for $r = \Theta(\sqrt{n})$.)

Proposition 2.8.2. *For any $n \geq 1$ and any radius $1 \leq r \leq \sqrt{n/2}$ with $r^2 \in \mathbb{Z}$,*

$$|\mathbb{Z}^n \cap rB_2^n| = (2ne^{1+\chi}/r^2)^{r^2} ,$$

where

$$-\frac{r^2}{n} - \frac{\log(Cr)}{r^2} \le \chi \le \sqrt{\frac{C}{\log(n/r^2)}} \,.$$

Proof. For the lower bound, we note that the number of vectors of length r whose coordinates lie in the set $\{-1, 0, +1\}$ is

$$2^{r^2}\binom{n}{r^2} \ge \frac{1}{\sqrt{2\pi e^{1/6} r}} \cdot (2e^{1-r^2/n} n/r^2)^{r^2} \,,$$

where we have used Corollary 1.4.2.

For the upper bound, using Eq. (2.8) with $s := \sqrt{\pi/\log(2n/r^2)}$,

$$\begin{aligned} |\mathbb{Z}^n \cap rB_2^n| &\le e^{\pi r^2/s^2} \rho_s(\mathbb{Z}^n) \\ &\le (2n/r^2)^{r^2} \cdot \left(1 + \frac{r^2(2+s)}{2n}\right)^n \\ &\le (2ne^{1+s/2}/r^2)^{r^2} \,, \end{aligned}$$

as needed. □

2.8.3 Random lattices

There exists a unique probability measure $\mathscr{L}_n$ over the set of full-rank determinant-one lattices in $\mathbb{R}^n$ that is invariant under $\mathrm{SL}_n(\mathbb{R})$ [Sie45]. (See, e.g., [Ter16] or [GL87, Chapter 3].) We call a random variable sampled from $\mathscr{L}_n$ a *random lattice*. The purpose of this section is to prove the following result.

Proposition 2.8.3. *For any sufficiently large n and any $r \ge \sqrt{n}\log n$,*

$$\Pr_{\mathcal{L}\sim\mathscr{L}_n}\left[\mathcal{L} \text{ is stable and } |\mathcal{L} \cap rB_2^n| \ge \mathrm{vol}(rB_2^n)/2\right] \ge 1 - (Cn/r^2)^{n/2} - (C/n)^{n/2} \,,$$

where $C > 0$ is some universal constant. In particular, there exists a stable lattice $\mathcal{L}$ satisfying

$$|\mathcal{L} \cap rB_2^n| \geq \mathrm{vol}(rB_2^n)/2 = (4\pi n)^{-1/2}(2\pi e r^2/n)^{n/2}(1 + o(1)) \,, \tag{2.10}$$

where the $o(1)$ term approaches zero as n approaches ∞.

Note that the lower bound in Eq. (2.10) is within a factor of $C\sqrt{n}$ of the upper bound in Item 3 of Corollary 2.1.4, which applies to stable lattices.

The proof of Proposition 2.8.3 uses the following three results.

Theorem 2.8.4 ([Sie45])**.** *For any $n \geq 2$ and any measurable set $S \subset \mathbb{R}^n$,*

$$\mathop{\mathbb{E}}_{\mathcal{L} \sim \mathscr{L}_n} [|(\mathcal{L} \setminus \{\mathbf{0}\}) \cap S|] = \mathrm{vol}(S) \,.$$

Theorem 2.8.5 ([Rog55, Sch60]; see [Gru07, Theorem 24.3])**.** *For $n \geq 3$ and any Borel set $S \subset \mathbb{R}^n$,*

$$\mathop{\mathbb{E}}_{\mathcal{L} \sim \mathscr{L}_n} \left[\left(|(\mathcal{L} \setminus \{\mathbf{0}\}) \cap S| - \mathrm{vol}(S) \right)^2 \right] \leq C\mathrm{vol}(S) \,,$$

where $C > 0$ is some universal constant.

Theorem 2.8.6 ([SW14])**.** *For any sufficiently large n, an n-dimensional random lattice is stable with probability at least $1 - (C/n)^{n/2}$, where $C > 0$ is some universal constant.*

Proof of Proposition 2.8.3. By Chebyshev's inequality, Theorem 2.8.4, and Theorem 2.8.5, there is some universal constant $C > 0$ such that

$$\Pr_{\mathcal{L} \sim \mathscr{L}_n} \left[|\mathcal{L} \cap rB_2^n| < \mathrm{vol}(rB_2^n)/2 \right] \leq \frac{C}{\mathrm{vol}(rB_2^n)} \leq (C'n/r^2)^{n/2} \,.$$

The result then follows by Theorem 2.8.6 and union bound. □

Chapter 3

A "Rotation" Identity and Related Inequalities[1]

3.1 Introduction

In spite of their importance, there is still a lot that we do not know about $\rho_s(\mathcal{L} - \boldsymbol{x})$, $f_{\mathcal{L},s}(\boldsymbol{t})$, and $D_{\mathcal{L}-\boldsymbol{t},s}$. Here, we prove several basic inequalities concerning these objects, as described below. All of these inequalities (with the minor exception of Theorem 3.2.2) follow without too much effort from one main inequality (Theorem 3.2.1), which is closely related to Riemann's theta relations (see [Mum07]). Namely, in terms of the periodic Gaussian $f_{\mathcal{L}}(\boldsymbol{t})$, our main inequality says that

$$f_{\mathcal{L}}(\boldsymbol{t})^2 f_{\mathcal{L}}(\boldsymbol{u})^2 \leq f_{\mathcal{L}}(\boldsymbol{t}+\boldsymbol{u}) f_{\mathcal{L}}(\boldsymbol{t}-\boldsymbol{u}) \ .$$

Note that the Gaussian function $\rho(\boldsymbol{t})$ over $\mathbb{R}^n$ satisfies the "rotation" identity

$$\rho(\boldsymbol{t})^2 \rho(\boldsymbol{u})^2 = \rho(\boldsymbol{t}+\boldsymbol{u}) \rho(\boldsymbol{t}-\boldsymbol{u}) \ ,$$

[1]This chapter is primarily based on joint work with Oded Regev that appeared in the SIAM Journal of Discrete Mathematics (SIDMA), 31(2) 2017 [RS17a], and passages have been taken verbatim from this source. This work was supported by the National Science Foundation (NSF) under Grant No. CCF-1320188.

so that our main inequality can be viewed as a relaxation of this identity to the periodic case. From this (perhaps rather opaque) inequality, we derive many natural statements concerning $\rho_s(\mathcal{L} - \boldsymbol{t})$, $f_{\mathcal{L},s}(\boldsymbol{t})$, and $D_{\mathcal{L}-\boldsymbol{t},s}$.

First, we show in Corollary 3.3.2 that the covariance of $D_{\mathcal{L}-\boldsymbol{t}}$ is minimized when $\boldsymbol{t} = \boldsymbol{0}$, answering a natural question communicated to us by Dadush [Dad12a]. (We note in passing that closely related questions are still open, e.g., whether $\mathbb{E}_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t}}}[\|\boldsymbol{X}\|]$ is minimized when $\boldsymbol{t} = \boldsymbol{0}$.) Along the way, we derive an interesting inequality concerning the "shape" of $f_{\mathcal{L}}(\boldsymbol{t})$ (Proposition 3.3.1). We also analyze the fourth moment, showing in particular that the discrete Gaussian is "leptokurtic" (Proposition 3.3.3)—i.e., its kurtosis is at least that of the *continuous* Gaussian distribution.

Second, in Section 3.4, we show various monotonicity results concerning $f_{\mathcal{L},s}$, answering a natural open question due to Price [Pri14b] in the affirmative. In particular, in Proposition 3.4.1 we show that $f_{\mathcal{L},s}$ is monotonic in s and in Proposition 3.4.2, we extend this to the non-spherical Gaussian case. (Recently, Price showed how to derive from this an analogous monotonicity result for Abelian Cayley graphs [Pri16]. A further extension to arbitrary Cayley graphs, previously suggested by Peres [Per13], turns out to be false [RS16].) Additionally, in Proposition 3.4.3, we show that $f_{\mathcal{L},s}$ is monotonic under taking sublattices of $\mathcal{L}$.

Finally, in Section 3.5, we show that sublattices of a lattice $\mathcal{L}$ are positively correlated under the normalized Gaussian measure on $\mathcal{L}$. This result answers another open question due to Price [Pri14a], and was recently used by him in his work on cohomology [Pri15]. It has a (possibly superficial) resemblance to the recently proven Gaussian correlation conjecture on symmetric convex bodies [Roy14]. In fact, we note in passing that our main inequality can also be viewed as a correlation result. In particular, it shows that $\cos(2\pi\langle \boldsymbol{X}, \boldsymbol{t}\rangle)$ and $\cos(2\pi\langle \boldsymbol{X}, \boldsymbol{y}\rangle)$ are positively correlated when $\boldsymbol{t}$ is sampled from $D_{\mathcal{L}}$. (See Eq. 3.4e.)

3.2 The main inequality (and a variant)

The following is our main theorem. The proof is essentially a combination of a certain identity related to Riemann's theta relations (see [Mum07, Chapter 1, Section 5]) and the Cauchy-Schwarz inequality.

Theorem 3.2.1 ([RS17a])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and any two vectors $\boldsymbol{t}, \boldsymbol{u} \in \mathbb{R}^n$, we have*

$$\rho(\mathcal{L} - \boldsymbol{t})^2 \rho(\mathcal{L} - \boldsymbol{u})^2 \leq \rho(\mathcal{L})^2 \rho(\mathcal{L} - \boldsymbol{t} - \boldsymbol{u}) \rho(\mathcal{L} - \boldsymbol{t} + \boldsymbol{u}) \, .$$

Proof. Let $\mathcal{L}^{\oplus 2} := \mathcal{L} \oplus \mathcal{L}$ be the lattice in $\mathbb{R}^{2n}$ formed by taking all pairs of lattice elements. We can then write $\rho(\mathcal{L} - \boldsymbol{t})\rho(\mathcal{L} - \boldsymbol{u}) = \rho(\mathcal{L}^{\oplus 2} - (\boldsymbol{t}, \boldsymbol{u}))$. Consider the $2n \times 2n$ matrix

$$T := \begin{pmatrix} I_n & I_n \\ I_n & -I_n \end{pmatrix} \, ,$$

where I_n is the $n \times n$ identity matrix. Note that $T/\sqrt{2}$ is an orthogonal matrix so that $\|T\boldsymbol{x}\| = \sqrt{2}\|\boldsymbol{x}\|$ for any $\boldsymbol{x} \in \mathbb{R}^{2n}$. We therefore have

$$\rho(\mathcal{L} - \boldsymbol{t})\rho(\mathcal{L} - \boldsymbol{u}) = \rho_{\sqrt{2}}\big(T(\mathcal{L}^{\oplus 2} - (\boldsymbol{t}, \boldsymbol{u}))\big) = \rho_{\sqrt{2}}\big(T\mathcal{L}^{\oplus 2} - (\boldsymbol{t} + \boldsymbol{u}, \boldsymbol{t} - \boldsymbol{u})\big) \, . \tag{3.1}$$

For any $\boldsymbol{y} := (\boldsymbol{y}_1, \boldsymbol{y}_2) \in \mathcal{L}^{\oplus 2}$, we have $T\boldsymbol{y} = (\boldsymbol{w}_1, \boldsymbol{w}_2)$ where $\boldsymbol{w}_1 := \boldsymbol{y}_1 + \boldsymbol{y}_2$ and $\boldsymbol{w}_2 := \boldsymbol{y}_1 - \boldsymbol{y}_2 = \boldsymbol{w}_1 - 2\boldsymbol{y}_2$. It follows that

$$\begin{aligned} T\mathcal{L}^{\oplus 2} &= \{(\boldsymbol{w}_1, \boldsymbol{w}_2) \in \mathcal{L}^2 : \boldsymbol{w}_1 \equiv \boldsymbol{w}_2 \bmod 2\mathcal{L}\} \\ &= \bigcup_{\boldsymbol{c} \in \mathcal{L}/(2\mathcal{L})} (2\mathcal{L} + \boldsymbol{c})^2 \, , \end{aligned}$$

where the union is disjoint. Plugging in to Eq. (3.1), we have

$$\rho(\mathcal{L}-\boldsymbol{t})\rho(\mathcal{L}-\boldsymbol{u}) = \sum_{\boldsymbol{c}\in\mathcal{L}/(2\mathcal{L})} \rho_{\sqrt{2}}(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t}-\boldsymbol{u})\cdot\rho_{\sqrt{2}}(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t}+\boldsymbol{u})\ . \tag{3.2}$$

Note that, by the right-hand side of (3.2), we can view $\rho(\mathcal{L}-\boldsymbol{t})\rho(\mathcal{L}-\boldsymbol{u})$ as the inner product of two 2^n-dimensional vectors as

$$\rho(\mathcal{L}-\boldsymbol{t})\rho(\mathcal{L}-\boldsymbol{u}) = \langle \boldsymbol{h}(\boldsymbol{t}+\boldsymbol{u}), \boldsymbol{h}(\boldsymbol{t}-\boldsymbol{u})\rangle\ , \tag{3.3}$$

where

$$h(\boldsymbol{x}) := \left(\rho_{\sqrt{2}}(2\mathcal{L}+\boldsymbol{c}_1-\boldsymbol{x}), \rho_{\sqrt{2}}(2\mathcal{L}+\boldsymbol{c}_2-\boldsymbol{x}), \ldots, \rho_{\sqrt{2}}(2\mathcal{L}+\boldsymbol{c}_{2^n}-\boldsymbol{x})\right) \in \mathbb{R}^{2^n}\ ,$$

for some ordering of the cosets $\boldsymbol{c}_i \in \mathcal{L}/(2\mathcal{L})$. Then, by Cauchy-Schwarz, we have

$$\rho(\mathcal{L}-\boldsymbol{t})^2\rho(\mathcal{L}-\boldsymbol{u})^2 \le \|\boldsymbol{h}(\boldsymbol{t}+\boldsymbol{u})\|^2\|\boldsymbol{h}(\boldsymbol{t}-\boldsymbol{u})\|^2 = \rho(\mathcal{L})^2\rho(\mathcal{L}-\boldsymbol{t}-\boldsymbol{u})\rho(\mathcal{L}-\boldsymbol{t}+\boldsymbol{u})\ ,$$

where the last equality follows from plugging in $\boldsymbol{u}=\boldsymbol{0}$ to Eq. (3.3) which tells us that $\|\boldsymbol{h}(\boldsymbol{t})\|^2 = \rho(\mathcal{L})\rho(\mathcal{L}-\boldsymbol{t})$. □

We remark that using the same proof with other transformations T might lead to other such inequalities. We leave this for future work, but we do note here an additional inequality with a very similar proof that originally appeared in [ADS15] (in a less general form).

Theorem 3.2.2. *For any lattice $\mathcal{L}\subset\mathbb{R}^n$ and any two vectors $\boldsymbol{t},\boldsymbol{u}\in\mathbb{R}^n$, we have*

$$\rho(\mathcal{L}-\boldsymbol{t})\rho(\mathcal{L}-\boldsymbol{u}) \le \rho_{\sqrt{2}}(\mathcal{L}-\boldsymbol{t}+\boldsymbol{u})\cdot\max_{\boldsymbol{c}\in\mathcal{L}/(2\mathcal{L})}\rho_{\sqrt{2}}(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t}-\boldsymbol{u})\ .$$

Proof. Recall from Eq. (3.2) that

$$\rho(\mathcal{L}-\boldsymbol{t})\rho(\mathcal{L}-\boldsymbol{u}) = \sum_{\boldsymbol{c}\in\mathcal{L}/(2\mathcal{L})} \rho_{\sqrt{2}}(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t}-\boldsymbol{u})\cdot\rho_{\sqrt{2}}(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t}+\boldsymbol{u})\ .$$

In the proof of Theorem 3.2.1, we applied Cauchy-Schwarz to the right-hand side of this equation. Here, we instead simply note the trivial inequality

$$\rho_{\sqrt{2}}(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t}-\boldsymbol{u}) \le \max_{\boldsymbol{c}'\in\mathcal{L}/(2\mathcal{L})} \rho_{\sqrt{2}}(2\mathcal{L}+\boldsymbol{c}'-\boldsymbol{t}-\boldsymbol{u})\ .$$

Plugging this in yields

$$\begin{aligned}\rho(\mathcal{L}-\boldsymbol{t})\rho(\mathcal{L}-\boldsymbol{u}) &\le \max_{\boldsymbol{c}\in\mathcal{L}/(2\mathcal{L})} \rho_{\sqrt{2}}(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t}-\boldsymbol{u})\cdot\sum_{\boldsymbol{c}'\in\mathcal{L}/(2\mathcal{L})}\rho_{\sqrt{2}}(2\mathcal{L}+\boldsymbol{c}'-\boldsymbol{t}+\boldsymbol{u})\\ &= \rho_{\sqrt{2}}(\mathcal{L}-\boldsymbol{t}+\boldsymbol{u})\cdot\max_{\boldsymbol{c}\in\mathcal{L}/(2\mathcal{L})}\rho_{\sqrt{2}}(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t}-\boldsymbol{u})\ ,\end{aligned}$$

as needed. □

3.2.1 Corollaries and reformulations

We now proceed to list a few immediate corollaries of Theorem 3.2.1.

Corollary 3.2.3 ([RS17a]). *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and any two vectors $\boldsymbol{t}, \boldsymbol{u} \in \mathbb{R}^n$, we have*

$$f_{\mathcal{L}}(\boldsymbol{t})^2 f_{\mathcal{L}}(\boldsymbol{u})^2 \leq f_{\mathcal{L}}(\boldsymbol{t}+\boldsymbol{u}) f_{\mathcal{L}}(\boldsymbol{t}-\boldsymbol{u}) \tag{3.4a}$$

$$f_{\mathcal{L}}(\boldsymbol{t})^4 \leq f_{\mathcal{L}}(2\boldsymbol{t}) \tag{3.4b}$$

$$f_{\mathcal{L}}(\boldsymbol{t}) f_{\mathcal{L}}(\boldsymbol{u}) \leq (f_{\mathcal{L}}(\boldsymbol{t}+\boldsymbol{u}) + f_{\mathcal{L}}(\boldsymbol{t}-\boldsymbol{u}))/2 \tag{3.4c}$$

$$\begin{aligned} \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\cos(2\pi\langle \boldsymbol{X}, \boldsymbol{t}\rangle)]^2 \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\cos(2\pi\langle \boldsymbol{X}, \boldsymbol{u}\rangle)]^2 \leq{}& \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\cos(2\pi\langle \boldsymbol{X}, \boldsymbol{t}\rangle)\cos(2\pi\langle \boldsymbol{X}, \boldsymbol{u}\rangle)]^2 \\ &- \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\sin(2\pi\langle \boldsymbol{X}, \boldsymbol{t}\rangle)\sin(2\pi\langle \boldsymbol{X}, \boldsymbol{u}\rangle)]^2 \end{aligned} \tag{3.4d}$$

$$\mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\cos(2\pi\langle \boldsymbol{X}, \boldsymbol{t}\rangle)] \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\cos(2\pi\langle \boldsymbol{X}, \boldsymbol{u}\rangle)] \leq \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\cos(2\pi\langle \boldsymbol{X}, \boldsymbol{t}\rangle)\cos(2\pi\langle \boldsymbol{X}, \boldsymbol{u}\rangle)] \,. \tag{3.4e}$$

Proof. Eq. (3.4a) follows from the definition of $f_{\mathcal{L}}$. Eq. (3.4b) follows from plugging in $\boldsymbol{u} = \boldsymbol{t}$ to Eq. (3.4a). Eq. (3.4c) follows from the fact that $\sqrt{ab} \leq (a+b)/2$ for all $a, b \geq 0$. For Eq. (3.4d), use the Poisson Summation Formula (Eq. (1.1)) to write $f_{\mathcal{L}^*}(\boldsymbol{t})$ in its dual form as

$$f_{\mathcal{L}^*}(\boldsymbol{t}) = \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\cos(2\pi\langle \boldsymbol{X}, \boldsymbol{t}\rangle)] \,.$$

We can then apply the identity $\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$ to derive Eq. (3.4d) from Eq. (3.4a). Finally, Eq. (3.4e) follows from applying the same analysis to (3.4c). □

3.3 Moments of the discrete Gaussian distribution

We will need the Hessian product identity

$$H\big(f(\boldsymbol{x})g(\boldsymbol{x})\big) = f(\boldsymbol{x})Hg(\boldsymbol{x}) + g(\boldsymbol{x})Hf(\boldsymbol{x}) + \nabla f(\boldsymbol{x})(\nabla g(\boldsymbol{x}))^T + \nabla g(\boldsymbol{x})(\nabla f(\boldsymbol{x}))^T \,. \tag{3.5}$$

We next show an inequality concerning the Hessian of $f_{\mathcal{L}}$. In particular, this inequality constrains the shape of the local maxima of $f_{\mathcal{L}}$. (In [DRS14], we showed that $f_{\mathcal{L}}$ can in fact

have local maxima at non-lattice points.)

Proposition 3.3.1 ([RS17a, Proposition 3.1])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and any vector $\boldsymbol{t} \in \mathbb{R}^n$, we have the positive semidefinite inequality*

$$\frac{Hf_{\mathcal{L}}(\boldsymbol{t})}{f_{\mathcal{L}}(\boldsymbol{t})} \succeq Hf_{\mathcal{L}}(\boldsymbol{0}) + \frac{\nabla f_{\mathcal{L}}(\boldsymbol{t})(\nabla f_{\mathcal{L}}(\boldsymbol{t}))^T}{f_{\mathcal{L}}(\boldsymbol{t})^2} \,.$$

Proof. By Eq. (3.4a), we have

$$f_{\mathcal{L}}(\boldsymbol{t}+\boldsymbol{u})f_{\mathcal{L}}(\boldsymbol{t}-\boldsymbol{u}) - f_{\mathcal{L}}(\boldsymbol{t})^2 f_{\mathcal{L}}(\boldsymbol{u})^2 \geq 0 \,.$$

Note that we have equality when $\boldsymbol{u} = \boldsymbol{0}$. It follows that, for any $\boldsymbol{t}$, the left-hand side has a local minimum at $\boldsymbol{u} = \boldsymbol{0}$, and therefore the Hessian with respect to $\boldsymbol{u}$ at $\boldsymbol{0}$ must be positive semidefinite. The result follows by using Eq. (3.5) to take the Hessian and rearranging. □

As a corollary, we obtain that the covariance matrix of $D_{\mathcal{L}-\boldsymbol{t}}$ is minimized at $\boldsymbol{t} = \boldsymbol{0}$. (Notice that the expectation of the centered Gaussian $D_{\mathcal{L}}$ is zero because the lattice is symmetric.) The corollary follows immediately from Proposition 3.3.1 and the following two identities:

$$\frac{\nabla f_{\mathcal{L}}(\boldsymbol{t})}{f_{\mathcal{L}}(\boldsymbol{t})} = \frac{\nabla \rho(\mathcal{L}-\boldsymbol{t})}{\rho(\mathcal{L}-\boldsymbol{t})} = -2\pi \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t}}}[\boldsymbol{X}] \,, \text{ and} \tag{3.6}$$

$$\frac{Hf_{\mathcal{L}}(\boldsymbol{t})}{f_{\mathcal{L}}(\boldsymbol{t})} = \frac{H\rho(\mathcal{L}-\boldsymbol{t})}{\rho(\mathcal{L}-\boldsymbol{t})} = 4\pi^2 \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t}}}[\boldsymbol{X}\boldsymbol{X}^T] - 2\pi I_n \,. \tag{3.7}$$

Corollary 3.3.2 ([RS17a, Corollary 3.2])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and vector $\boldsymbol{t} \in \mathbb{R}^n$, we have the positive semidefinite inequality*

$$\mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t}}}[\boldsymbol{X}\boldsymbol{X}^T] - \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t}}}[\boldsymbol{X}] \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t}}}[\boldsymbol{X}^T] \succeq \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\boldsymbol{X}\boldsymbol{X}^T] \,.$$

In particular,

$$\mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t}}}[\|\boldsymbol{X}\|^2] - \Big\| \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t}}}[\boldsymbol{X}]\Big\|^2 \geq \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\|\boldsymbol{X}\|^2] \,.$$

The following proposition (with $\boldsymbol{u} = \boldsymbol{v}$) implies that the one-dimensional projections of the discrete Gaussian distribution are "leptokurtic," i.e., have kurtosis at least 3, the kurtosis of a normal variable. We remark that the case $n = 1$ follows from a known inequality related to the Riemann zeta function [Chu76, New76] (see also [BPY01, Section 2.2]).

Proposition 3.3.3 ([RS17a, Proposition 3.3])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$,*

$$\mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\langle \boldsymbol{X}, \boldsymbol{u}\rangle^2 \langle \boldsymbol{X}, \boldsymbol{v}\rangle^2] \geq \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\langle \boldsymbol{X}, \boldsymbol{u}\rangle^2] \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\langle \boldsymbol{X}, \boldsymbol{v}\rangle^2] + 2 \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\langle \boldsymbol{X}, \boldsymbol{u}\rangle \langle \boldsymbol{X}, \boldsymbol{v}\rangle]^2 \,.$$

Proof. From Corollary 3.3.2, we have

$$\mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t}}}[\langle \boldsymbol{X}, \boldsymbol{u}\rangle^2] - \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t}}}[\langle \boldsymbol{X}, \boldsymbol{u}\rangle]^2 - \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\langle \boldsymbol{X}, \boldsymbol{u}\rangle^2] \geq 0 \,.$$

The same is true if we multiply through by $\rho(\mathcal{L} - \boldsymbol{t})^2$, which leads to the inequality

$$\sum_{\boldsymbol{y},\boldsymbol{y}'\in\mathcal{L}} \rho(\boldsymbol{y} - \boldsymbol{t})\rho(\boldsymbol{y}' - \boldsymbol{t}) \cdot \Big(\langle \boldsymbol{y} - \boldsymbol{y}', \boldsymbol{u}\rangle^2/2 - \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\langle \boldsymbol{X}, \boldsymbol{u}\rangle^2]\Big) \geq 0 \,.$$

(To see that, use $\langle \boldsymbol{y} - \boldsymbol{y}', \boldsymbol{u}\rangle = \langle \boldsymbol{y} - \boldsymbol{t}, \boldsymbol{u}\rangle - \langle \boldsymbol{y}' - \boldsymbol{t}, \boldsymbol{u}\rangle$, and expand the square.) Note that the left-hand side equals zero when $\boldsymbol{t} = \boldsymbol{0}$ since $\mathcal{L} = -\mathcal{L}$. Therefore, as in the proof of Proposition 3.3.1, the Hessian of the left-hand side with respect to $\boldsymbol{t}$ at $\boldsymbol{t} = \boldsymbol{0}$ must be positive semidefinite. Using Eqs. (3.5), (3.6), and (3.7), we see that

$$H(\rho(\boldsymbol{y} - \boldsymbol{t})\rho(\boldsymbol{y}' - \boldsymbol{t}))|_{\boldsymbol{t}=\boldsymbol{0}} = 4\pi^2 \rho(\boldsymbol{y})\rho(\boldsymbol{y}')\big((\boldsymbol{y} + \boldsymbol{y}')(\boldsymbol{y} + \boldsymbol{y}')^T - I_n/\pi\big) \,.$$

Therefore,

$$\begin{aligned}
0 &\preceq \mathop{\mathbb{E}}_{\boldsymbol{X},\boldsymbol{X}'\sim D_{\mathcal{L}}}\left[\left((\boldsymbol{X}+\boldsymbol{X}')(\boldsymbol{X}+\boldsymbol{X}')^T - I_n/\pi\right)\cdot\left(\langle \boldsymbol{X}-\boldsymbol{X}',\boldsymbol{u}\rangle^2/2 - \mathop{\mathbb{E}}_{\boldsymbol{X}''\sim D_{\mathcal{L}}}[\langle \boldsymbol{X}'',\boldsymbol{u}\rangle^2]\right)\right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{X},\boldsymbol{X}'\sim D_{\mathcal{L}}}\left[(\boldsymbol{X}+\boldsymbol{X}')(\boldsymbol{X}+\boldsymbol{X}')^T\cdot\left(\langle \boldsymbol{X}-\boldsymbol{X}',\boldsymbol{u}\rangle^2/2 - \mathop{\mathbb{E}}_{\boldsymbol{X}''\sim D_{\mathcal{L}}}[\langle \boldsymbol{X}'',\boldsymbol{u}\rangle^2]\right)\right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\boldsymbol{X}\boldsymbol{X}^T\langle \boldsymbol{X},\boldsymbol{u}\rangle^2] - \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\boldsymbol{X}\boldsymbol{X}^T]\mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\langle \boldsymbol{X},\boldsymbol{u}\rangle^2] - 2\mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\boldsymbol{X}\langle \boldsymbol{X},\boldsymbol{u}\rangle]\mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}}}[\boldsymbol{X}^T\langle \boldsymbol{X},\boldsymbol{u}\rangle],
\end{aligned}$$

as needed. □

3.4 Monotonicity of the periodic Gaussian function

The next proposition shows that $f_{\mathcal{L},s}(\boldsymbol{t})$ is non-decreasing as a function of s. This (and the more general statement in Proposition 3.4.2) answers a question of Price [Pri14b], who proved it for the one-dimensional case $n = 1$ (illustrated in Figure 3.1).

One might wonder if such a monotonicity property is specific to flat tori or whether it is a special case of a more general phenomenon. Namely, Peres [Per13] asked whether for any vertex transitive graph G it holds that for any two vertices u, v, the ratio $\Pr[X_s = v]/\Pr[X_s = u]$ is non-decreasing as a function of s, where X_s is a continuous-time random walk on G starting at u after time s. Recently, using our result, Price showed how to prove this for Abelian Cayley graphs [Pri16]. Interestingly, a further extension to arbitrary Cayley graphs turns out to be false [RS16].

Proposition 3.4.1 ([RS17a, Proposition 4.1])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and vector $\boldsymbol{t} \in \mathbb{R}^n$,*

$$\frac{\frac{d}{ds}f_{\mathcal{L},s}(\boldsymbol{t})}{f_{\mathcal{L},s}(\boldsymbol{t})} \geq \frac{s}{2\pi}\cdot\frac{\|\nabla f_{\mathcal{L},s}(\boldsymbol{t})\|^2}{f_{\mathcal{L},s}(\boldsymbol{t})^2}.$$

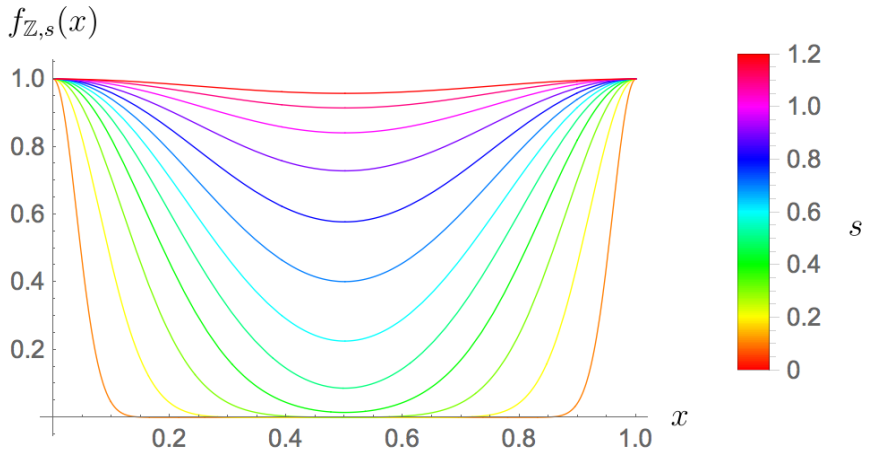


Figure 3.1: $f_{\mathbb{Z},s}(t)$ for various values of s and $t \in [0,1]$.

Proof. A straightforward computation shows that

$$\begin{aligned}\frac{d}{ds} f_{\mathcal{L},s}(\boldsymbol{t}) &= \frac{2\pi f_{\mathcal{L},s}(\boldsymbol{t})}{s^3} \cdot \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t},s}} [\|\boldsymbol{X}\|^2] - \frac{2\pi f_{\mathcal{L},s}(\boldsymbol{t})}{s^3} \cdot \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L},s}} [\|\boldsymbol{t}\|^2] \\ &\geq \frac{2\pi f_{\mathcal{L},s}(\boldsymbol{t})}{s^3} \left\| \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t},s}} [\boldsymbol{X}] \right\|^2 ,\end{aligned}$$

where we have applied Corollary 3.3.2. The result then follows from the fact that (see Eq.(3.6))

$$\frac{\nabla f_{\mathcal{L},s}(\boldsymbol{t})}{f_{\mathcal{L},s}(\boldsymbol{t})} = -\frac{2\pi}{s^2} \cdot \mathop{\mathbb{E}}_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t},s}} [\boldsymbol{X}] . \qquad \square$$

We now extend this monotonicity result by replacing the scalar variance parameter s^2 by a positive-definite matrix Σ. In particular, we define

$$f_{\mathcal{L},\Sigma}(\boldsymbol{t}) = \frac{\sum_{\boldsymbol{y}\in\mathcal{L}} \exp(-\pi(\boldsymbol{y}-\boldsymbol{t})^T\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{t}))}{\sum_{\boldsymbol{y}\in\mathcal{L}} \exp(-\pi\boldsymbol{y}^T\Sigma^{-1}\boldsymbol{y})} .$$

Equivalently,

$$f_{\mathcal{L},\Sigma}(\boldsymbol{t}) = f_{\Sigma^{-1/2}\mathcal{L}}(\Sigma^{-1/2}\boldsymbol{t}) ,$$

where $\Sigma^{1/2}$ is the unique positive-definite square root of Σ.

Proposition 3.4.2 ([RS17a, Proposition 4.2])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$, $\boldsymbol{t} \in \mathbb{R}^n$, and positive-definite matrices $\Sigma, \Sigma' \in \mathbb{R}^{n \times n}$ satisfying the positive semidefinite inequality $\Sigma' \preceq \Sigma$,*

$$f_{\mathcal{L},\Sigma'}(\boldsymbol{t}) \leq f_{\mathcal{L},\Sigma}(\boldsymbol{t}) \, .$$

Proof. We may replace $\mathcal{L}$ by $\Sigma'^{-1/2}\mathcal{L}$, $\boldsymbol{t}$ by $\Sigma'^{-1/2}\boldsymbol{t}$, and Σ by $\Sigma'^{-1/2}\Sigma\Sigma'^{-1/2}$ so that we can assume without loss of generality that $\Sigma' = I_n$. Moreover, by a change of basis, we may take Σ to be diagonal. (Here, we have used the fact that the Gaussian is invariant under orthogonal transformations.)

So, it suffices to show that $f_{\mathcal{L}}(\boldsymbol{t}) \leq f_{\mathcal{L},\Sigma}(\boldsymbol{t})$ when $\Sigma \in \mathbb{R}^{n \times n}$ is a diagonal matrix with $\Sigma \succeq I_n$. Let $s_1^2, \dots, s_n^2 \geq 1$ be the entries along the diagonal of Σ. The proof now proceeds nearly identically to the proof of Proposition 3.4.1. Differentiating with respect to s_i, we have

$$\frac{\mathrm{d}}{\mathrm{d}s_i} f_{\mathcal{L},\Sigma}(\boldsymbol{t}) = \frac{2\pi f_{\mathcal{L},\Sigma}(\boldsymbol{t})}{s_i^3} \Big(\mathop{\mathbb{E}}_{\boldsymbol{X} \sim D_{\Sigma^{-1/2}(\mathcal{L}-\boldsymbol{t})}} [X_i^2] - \mathop{\mathbb{E}}_{\boldsymbol{X} \sim D_{\Sigma^{-1/2}\mathcal{L}}} [X_i^2] \Big) \, ,$$

where X_i is the ith coordinate of $\boldsymbol{X}$. The result follows by noting that Corollary 3.3.2 implies that this derivative is positive for all $s_i > 0$, so that $f_{\mathcal{L},\Sigma}(\boldsymbol{t})$ is an increasing function of s_i. □

Finally, we prove our last monotonicity result, now with respect to taking sublattices.

Proposition 3.4.3 ([RS17a, Proposition 4.3])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$, sublattice $\mathcal{L}' \subseteq \mathcal{L}$, and vector $\boldsymbol{t} \in \mathbb{R}^n$,*

$$f_{\mathcal{L}'}(\boldsymbol{t}) \leq f_{\mathcal{L}}(\boldsymbol{t}) \, .$$

Proof.

$$\begin{aligned}
\rho(\mathcal{L}' - \boldsymbol{t})\rho(\mathcal{L}) &= \sum_{\boldsymbol{c} \in \mathcal{L}/\mathcal{L}'} \rho(\mathcal{L}' - \boldsymbol{t})\rho(\mathcal{L}' + \boldsymbol{c}) \\
&\leq \sum_{\boldsymbol{c} \in \mathcal{L}/\mathcal{L}'} \rho(\mathcal{L}')(\rho(\mathcal{L}' + \boldsymbol{c} - \boldsymbol{t}) + \rho(\mathcal{L}' - \boldsymbol{c} - \boldsymbol{t}))/2 \qquad \text{(Eq. (3.4c))} \\
&= \sum_{\boldsymbol{c} \in \mathcal{L}/\mathcal{L}'} \rho(\mathcal{L}')\rho(\mathcal{L}' + \boldsymbol{c} - \boldsymbol{t}) \\
&= \rho(\mathcal{L}')\rho(\mathcal{L} + \boldsymbol{x}) \ .
\end{aligned}$$

The result follows. □

3.5 Positive correlation of the Gaussian measure on lattices

The following shows that sublattices are positively correlated under the normalized Gaussian measure on a lattice. (Price asked whether this holds in the special case when $\mathcal{N} := \mathcal{L} \cap V$ for some subspace $V \subseteq \mathbb{R}^n$ [Pri14a].)

Theorem 3.5.1 ([RS17a, Theorem 5.1])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and sublattices $\mathcal{M}, \mathcal{N} \subseteq \mathcal{L}$,*

$$\frac{\rho(\mathcal{M})}{\rho(\mathcal{L})} \cdot \frac{\rho(\mathcal{N})}{\rho(\mathcal{L})} \leq \frac{\rho(\mathcal{M} \cap \mathcal{N})}{\rho(\mathcal{L})} \ .$$

Proof. Note that the natural mapping from $\mathcal{M}/(\mathcal{M} \cap \mathcal{N})$ to $\mathcal{L}/\mathcal{N}$ given by $\boldsymbol{c} \mapsto \mathcal{N} + \boldsymbol{c}$ is

injective. So,

$$\begin{aligned}
\frac{\rho(\mathcal{L})}{\rho(\mathcal{N})} &= \sum_{\boldsymbol{c}\in\mathcal{L}/\mathcal{N}} \frac{\rho(\mathcal{N}+\boldsymbol{c})}{\rho(\mathcal{N})} \\
&\geq \sum_{\boldsymbol{c}\in\mathcal{M}/(\mathcal{M}\cap\mathcal{N})} \frac{\rho(\mathcal{N}+\boldsymbol{c})}{\rho(\mathcal{N})} \\
&\geq \sum_{\boldsymbol{c}\in\mathcal{M}/(\mathcal{M}\cap\mathcal{N})} \frac{\rho((\mathcal{M}\cap\mathcal{N})+\boldsymbol{c})}{\rho(\mathcal{M}\cap\mathcal{N})} && \text{(Prop. 3.4.3)} \\
&= \frac{\rho(\mathcal{M})}{\rho(\mathcal{M}\cap\mathcal{N})} \,.
\end{aligned}$$

The result follows by rearranging. □

Chapter 4

An Algorithm for DGS (and SVP and CVP)[1]

4.1 Introduction

The two most important computational problems on lattices are the Shortest Vector Problem (SVP) and the Closest Vector Problem (CVP). Given a basis for a lattice $\mathcal{L} \subseteq \mathbb{R}^n$, SVP asks us to compute a non-zero vector in $\mathcal{L}$ of minimal length, and CVP asks us to compute a lattice vector nearest in Euclidean distance to a target vector $\boldsymbol{t}$. I.e., SVP asks for a lattice vector with length $\lambda_1(\mathcal{L})$, and CVP asks for a lattice vector $\boldsymbol{y} \in \mathcal{L}$ with $\|\boldsymbol{y} - \boldsymbol{t}\| = \text{dist}(\boldsymbol{t}, \mathcal{L})$.

Starting with the seminal work of [LLL82], algorithms for solving these problems either exactly or approximately have been studied intensely. Such algorithms have found applications in factoring polynomials over rationals [LLL82], integer programming [Len83, Kan87, DPV11], cryptanalysis [Odl90, JS98, NS01], checking the solvability by radicals [LM83], and solving

[1]This chapter is primarily based on joint work with Divesh Aggarwal, Daniel Dadush, and Oded Regev, which appeared in the Symposium on the Theory of Computing (STOC), 2015 [ADRS15] and joint work with Divesh Aggarwal and Daniel Dadush, which appeared in the Symposium on the Foundations of Computer Science (FOCS), 2015 [ADS15]. Some passages have been taken verbatim from these sources. This material is based upon work supported by the National Science Foundation under Grant No. CCF-1320188.

low-density subset-sum problems [CJL+92]. More recently, many powerful cryptographic primitives have been constructed whose security is based on the *worst-case* hardness of these or related lattice problems [Ajt04, MR07, Gen09, Reg09, BV11, BLP+13, BV14], and some of these cryptosystems are nearing widespread deployment [ADPS16, BCD+16, NIS16].

In their exact forms, both problems are known to be NP-hard (although SVP is only known to be NP-hard under randomized reductions), and they are even hard to approximate to within a factor of $n^{O(1/\log\log n)}$ under reasonable complexity-theoretic assumptions [ABSS93, Ajt98, CN98, BS99, DKRS03, Mic01, Kho05, HR12]. CVP is thought to be the "harder" of the two problems, as there is a simple reduction from SVP to CVP that preserves the dimension n of the lattice [GMSS99], even in the approximate case, while there is no known reduction in the other direction that preserves the dimension.[2] Indeed, CVP is in some sense nearly "complete for lattice problems," as there are known dimension-preserving reductions from nearly all important lattice problems to CVP [Mic08]. (The Lattice Isomorphism Problem [HR14] and the related Lattice Distortion Problem [BDS16] are important exceptions.) None of these problems has a known dimension-preserving reduction to SVP.

Exact algorithms for CVP and SVP have a rich history. Kannan initiated their study with an enumeration-based $n^{O(n)}$-time algorithm for CVP [Kan87], and many others improved upon his technique to lower the constant in the exponent [Hel85, HS07, MW15]. Since these algorithms solve CVP, they also imply solutions for SVP and all of the problems listed above. (Notably, these algorithms use only polynomial space.)

For over a decade, these $n^{O(n)}$-time algorithms remained the state of the art until, in a major breakthrough, Ajtai, Kumar, and Sivakumar (AKS) published the first $2^{O(n)}$-time algorithm for SVP [AKS01]. The AKS algorithm is based on "randomized sieving," in which

[2]Since both problems are NP-complete, there is necessarily an efficient reduction from CVP to SVP. However, all known reductions either blow up the approximation factor or the dimension of the lattice by a polynomial factor [Kan87, DH11]. Since we are interested in an algorithm for solving exact CVP whose running time is exponential in the dimension, such reductions are not useful for us.

many randomly generated lattice vectors are iteratively combined to create successively shorter lattice vectors. The work of AKS led to two major questions: First, can CVP be solved in a similar amount of time? And second, what is the best achievable constant in the exponent? Much work went into solving both of these problems using AKS's sieving technique [AKS01, AKS02, NV08, AJ08, BN09, PS09, MV10, HPS11], culminating in a $\widetilde{O}(2^{2.456n})$-time algorithm for SVP and a $2^{O(n)}(1+1/\varepsilon)^{O(n)}$-time algorithm for $(1+\varepsilon)$-approximate CVP. But, algorithms for exact CVP remained out of reach.

The celebrated algorithm of Micciancio and Voulgaris [MV13] (MV), which built upon the approach of Sommer, Feder, and Shalvi [SFS09], addressed this issue while simultaneously achieving a lower constant in the exponent than prior techniques. Indeed, MV showed a *deterministic* $\widetilde{O}(4^n)$-time and $\widetilde{O}(2^n)$-space algorithm for exact CVP (and thus SVP as well), using an entirely new technique based on the Voronoi cell of the lattice. Until very recently, this algorithm had the best known asymptotic running time for *both* SVP and CVP. And, the MV algorithm was the only $2^{O(n)}$-time algorithm for *exact* CVP. (Indeed, there are inherent barriers to extending sieving results to exact CVP. See [ADS15] for a brief discussion.)

4.1.1 Our contribution

We show a $2^{n+o(n)}$-time (and space) algorithm for SVP (originally due to [ADRS15]) and a $2^{n+o(n)}$-time (and space) algorithm for CVP (originally due to [ADS15]). Both of these results follow from a new $2^{n+o(n)}$-time (and space) algorithm for discrete Gaussian sampling (DGS) with very low parameters $s > 0$. A crucial property of this algorithm is that it outputs *many* independent discrete Gaussian samples in $2^{n+o(n)}$ (i.e., the algorithm is in some sense amortized), and we use this property to prove our main result.

Theorem 4.1.1 ([ADRS15, ADS15])**.** *There is a $2^{n+o(n)}$-time algorithm that takes as input a (basis for a) lattice $\mathcal{L} \subset \mathbb{R}^n$, a shift vectors $\boldsymbol{t} \in \mathbb{R}^n$, and any parameter $s > \operatorname{dist}(\boldsymbol{t}, \mathcal{L})/2^{o(n/\log n)}$*

and outputs many independent samples from $D_{\mathcal{L}-\boldsymbol{t},s}$.

In fact, for the special case when $\boldsymbol{t} = \boldsymbol{0}$*, the algorithm outputs* $2^{n/2}$ *independent samples from* $D_{\mathcal{L},s}$ *(in the same running time), and for any* $\boldsymbol{t} \in \mathbb{R}^n$*, the algorithm outputs at least*

$$\frac{\rho_s(\mathcal{L} - \boldsymbol{t})}{\max_{\boldsymbol{c} \in \mathcal{L}/(2\mathcal{L})} \rho_s(2\mathcal{L} - \boldsymbol{c} - \boldsymbol{t})} \geq 1$$

independent samples from $D_{\mathcal{L}-\boldsymbol{t},s}$.

It is relatively straightforward to obtain an algorithm for SVP from Theorem 4.1.1. One simply needs to show that there exists some $s > 0$ such that $D_{\mathcal{L},s}$ is a shortest non-zero vector with probability greater than $2^{-n/2}$. (See Section 4.4.)

Theorem 4.1.2 ([ADRS15])**.** *There is a (randomized)* $2^{n+o(n)}$*-time algorithm for SVP.*

A similar argument about the shifted Gaussian $D_{\mathcal{L}-\boldsymbol{t},s} + \boldsymbol{t}$ shows that Theorem 4.1.1 immediately implies a $2^{n+o(n)}$-time algorithm that approximates CVP to within any approximation factor γ with $\gamma > 1 + 2^{-o(n/\log n)}$. With a lot more work, in [ADS15] we show how to use recursive calls to the sampler from Theorem 4.1.1 to solve *exact* CVP with the same asymptotic running time.

Theorem 4.1.3 ([ADS15])**.** *There is a (randomized)* $2^{n+o(n)}$*-time algorithm for CVP.*

We also show a $2^{n/2+o(n)}$-time algorithm that samples from the discrete Gaussian above the smoothing parameter.

Theorem 4.1.4 ([ADRS15])**.** *There is a* $2^{n/2+o(n)}$*-time algorithm that takes as input a (basis for a) lattice* $\mathcal{L} \subset \mathbb{R}^n$*, a shift vector* $\boldsymbol{t} \in \mathbb{R}^n$*, and parameter* $s \geq \sqrt{2}\eta_{1/2}(\mathcal{L})$ *and outputs* $2^{n/2}$ *independent samples from* $D_{\mathcal{L}-\boldsymbol{t},s}$.

Theorem 4.1.4 is already enough to approximate the decision version of SVP to within a small constant factor in $2^{n/2+o(n)}$ time. (See [ADRS15].)

4.1.2 Our techniques

A $2^{n+o(n)}$-time combiner for DGS. Recall that efficient algorithms are known for sampling from the discrete Gaussian at very high parameters [GPV08]. Indeed, by using prior work, we can sample from $D_{\mathcal{L}-\boldsymbol{t},s}$ in, say, $2^{n/10}$ time for any $s \geq n^{10} \max\{\lambda_1(\mathcal{L}),\ \mathrm{dist}(\boldsymbol{t},\mathcal{L})\}$. (This is a bit of an oversimplification. See Corollary 4.2.2.) It therefore suffices to find a way to *convert* samples from the discrete Gaussian with a high parameter to samples with a parameter lowered by a constant factor. By repeating this "conversion" many times, we can obtain samples with much lower parameters.

Note that this is trivial to do for the *continuous* Gaussian: if we divide a vector sampled from the continuous Gaussian distribution by 2, the result is distributed as a continuous Gaussian with half the width. Of course, half of a vector in $\mathcal{L} - \boldsymbol{t}$ is typically not contained in $\mathcal{L} - \boldsymbol{t}$, so this method fails spectacularly when applied to the *discrete* Gaussian. In the centered case, when $\boldsymbol{t} = \boldsymbol{0}$, we can try to fix this by conditioning on the result staying in $\mathcal{L}$. I.e., we can sample many vectors from $D_{\mathcal{L},s}$, keep those that are in $2\mathcal{L}$, and divide them by two. This method does work, but it is terribly inefficient—there are 2^n cosets of $2\mathcal{L}$, and for some typical parameters, a sample from $D_{\mathcal{L},s}$ will land in $2\mathcal{L}$ with probability as small as 2^{-n}. I.e., our "loss factor," the ratio of the number of output vectors to the number of input vectors, can be as bad as 2^{-n} for a single step. If we wish to iterate this k times, we could need 2^{kn} input vectors for each output vector, resulting in a very slow algorithm!

We can be much more efficient, however, if we instead look for *pairs* of vectors sampled from $D_{\mathcal{L},s}$ whose *sum* is in $2\mathcal{L}$, or equivalently pairs of vectors that lie in the same coset $\boldsymbol{c}$ mod $2\mathcal{L}$. Taking our intuition from the continuous Gaussian, we might hope that the *average* of two such vectors will be distributed as $D_{\mathcal{L},s/\sqrt{2}}$. And, this process has the additional benefit that it has the potential to work in the general case when $\boldsymbol{t} \neq \boldsymbol{0}$ as well. In particular, if we take the average of vectors from $D_{\mathcal{L}-\boldsymbol{t},s}$ conditioned on them lying in the same coset mod $2\mathcal{L}$,

the result will at least land in $\mathcal{L} - \boldsymbol{t}$. So, there is at least hope that the resulting distribution will be $D_{\mathcal{L}-\boldsymbol{t},s/\sqrt{2}}$.

This suggests an *amortized* algorithm, in which we sample many vectors from $D_{\mathcal{L}-\boldsymbol{t},s}$, place them in "buckets" according to their coset mod $2\mathcal{L}$, and then take the average of disjoint pairs of elements in the same bucket. We call such an algorithm a "combiner." The most natural combiner to consider is the "greedy combiner," which simply pairs as many vectors in each bucket as it can, leaving at most one unpaired vector per bucket. Since there are 2^n cosets, if we take, say, $\Omega(2^n)$ samples from $D_{\mathcal{L}-\boldsymbol{t},s}$, almost all of the resulting vectors will be paired. A lemma due to Peikert [Pei10] shows that the resulting distribution will be statistically close to the desired distribution, $D_{\mathcal{L}-\boldsymbol{t},s/\sqrt{2}}$, *provided that the parameter s is above the smoothing parameter.* (I.e., $s \gtrsim \eta_{2^{-n}}(\mathcal{L})$.)

At this point, we can already build a roughly 2^n-time algorithm for DGS that works for such parameters. (Namely, use prior work to sample at some very high parameter and iteratively apply the combiner described above.) But, in order to move *below smoothing* (which is necessary, e.g., for solving SVP and CVP), we need to do something else.

In particular, below the smoothing parameter, combining discrete Gaussian vectors "greedily" as above will not typically give a result that is statistically close to a Gaussian distribution. However, all is not lost. Recall that our algorithm works by picking pairs of vectors sampled independently from $D_{\mathcal{L}-\boldsymbol{t},s}$ that are in the same coset $\boldsymbol{c}$ mod $2\mathcal{L}$, and then taking the average of each pair. So, the algorithm effectively samples a vector $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ from *some* distribution over a coset $\overline{\mathcal{L}} - (\boldsymbol{t}, \boldsymbol{t})$ of the $2n$-dimensional lattice $\overline{\mathcal{L}}$ of pairs of vectors that are in the same coset mod $2\mathcal{L}$,

$$\overline{\mathcal{L}} := \{(\boldsymbol{y}_1, \boldsymbol{y}_2) \in \mathcal{L}^2 \ : \ \boldsymbol{y}_1 = \boldsymbol{y}_2 \bmod 2\mathcal{L}\} = \bigcup_{\boldsymbol{c} \in \mathcal{L}/(2\mathcal{L})} \boldsymbol{c} \times \boldsymbol{c}\,,$$

and then outputs $(\boldsymbol{X}_1 + \boldsymbol{X}_2)/2$. We claim that *assuming that that distribution is* $D_{\overline{\mathcal{L}}-(\boldsymbol{t},\boldsymbol{t}),s}$,

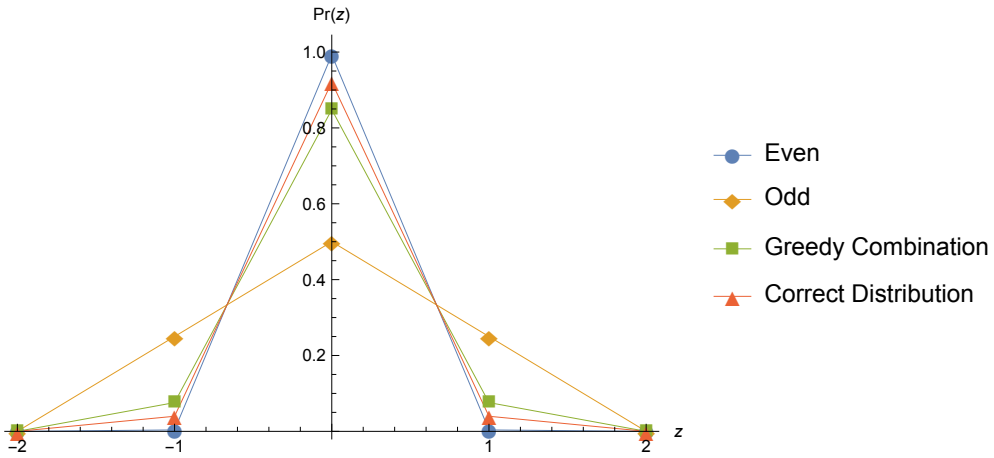


Figure 4.1: The distribution of averages of pairs of integers sampled from $D_{\mathbb{Z},\sqrt{2}}$ resulting from taking (1) only even pairs; (2) only odd pairs; (3) even and odd pairs with "greedy" weights proportional to $\rho_{\sqrt{2}}(2\mathbb{Z})$ and $\rho_{\sqrt{2}}(2\mathbb{Z}+1)$ respectively; and (4) even and odd pairs with "squared" weights proportional to $\rho_{\sqrt{2}}(2\mathbb{Z})^2$ and $\rho_{\sqrt{2}}(2\mathbb{Z}+1)^2$ respectively. The fourth distribution is exactly $D_{\mathbb{Z}}$.

the output $(\boldsymbol{X}_1 + \boldsymbol{X}_2)/2$ is distributed *exactly* as $D_{\mathcal{L}-\boldsymbol{t},s/\sqrt{2}}$. In fact, this is just a special case of the rotation identity from Chapter 3, Eq. (3.1).

However, note that if the combiner just greedily paired as many vectors from each coset as possible, it would *not* yield samples from $D_{\overline{\mathcal{L}-(\boldsymbol{t},\boldsymbol{t}),s}}$. In particular, the probability that a sample from $D_{\overline{\mathcal{L}-(\boldsymbol{t},\boldsymbol{t}),s}}$ will land in $(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t}) \times (2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t})$ for some coset $\boldsymbol{c} \in \mathcal{L}/(2\mathcal{L})$ is proportional to the "squared weight" of the coset $\rho_s(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t})^2$. But, the greedy approach pairs vectors from $2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t}$ with probability essentially proportional to $\rho_s(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t})$. (Figure 4.1 shows how the resulting distributions differ in the one-dimensional case.) For parameters above smoothing, these distributions are roughly the same, but to go below smoothing (and to avoid the statistical error resulting from the greedy approach), we need a way to sample pairs from this "squared distribution" directly.

In [ADRS15], we showed a generic solution for "converting any probability distribution to its square" relatively efficiently, which we call the "square sampler." Informally, the square sampler is given access to samples from some probability distribution that assigns respective

(unknown) probabilities $(p_1, \ldots, p_N)$ to the elements in some (large) finite set $\{1, \ldots, N\}$. It uses this to efficiently sample a large collection of *independent* coin flips $b_{i,j}$ such that $b_{i,j} = 1$ with probability proportional to p_i. Then, using these coins, it applies rejection sampling to the input samples (accepting the jth instance of input value i if $b_{i,j} = 1$) in order to obtain the desired "squared distribution." If $\Pr[b_{i,j} = 1] = Tp_i$ for some proportionality factor T, it is not hard to see that the expected "loss factor" of this process is $T \sum p_i^2$. We therefore take T to be as large as possible by setting $T \approx 1/\max p_i$ (if we took T to be any larger, we would need a coin that lands on heads with probability greater than one!), making the loss factor of the square sampler approximately $\sum p_i^2 / \max p_i$. (See Section 4.3.1 and Corollary 4.3.4 in particular.)

In particular, when combining discrete Gaussian vectors, the loss factor is approximately the *collision probability* over the cosets, $\sum \rho_s(2\mathcal{L} + \boldsymbol{c} - \boldsymbol{t})^2 / \rho_s(\mathcal{L} - \boldsymbol{t})^2$, divided by the maximal probability of a single coset. As a result, if one coset has a $2^{-n/2}$ fraction of the total weight and the other cosets split the remaining weight roughly evenly, then the loss factor is roughly $2^{-n/2}$ *for a single step of the combiner*. This looks terrible for us, as it could be the case that k applications of the combiner could yield a loss factor of $2^{-kn/2}$! Surprisingly, we show that (1) in the centered case when $\boldsymbol{t} = \boldsymbol{0}$, the product of all loss factors for an arbitrarily long sequence of applications of the combiner is at worst $2^{-n/2}$ (ignoring loss due to other factors); and (2) in the general case, the product of all loss factors is at worst 2^{-n} times the reciprocal of the probability of sampling the coset with maximal mass.[3] As a result, our sampler always returns many independent Gaussian samples—at least $2^{n/2}$ when $\boldsymbol{t} = \boldsymbol{0}$; and at least the reciprocal of the probability of the maximal coset in general.

[3]While the purely algebraic proofs of these facts are quite simple (see the proof of Theorem 4.3.7), we do not yet have good intuitive understanding of it. Indeed, we have found ourselves referring to the remarkable cancellation in these proofs as the "magic cancellation."

A $2^{n/2+o(n)}$-time combiner for DGS above smoothing. We now present a faster algorithm that works as long as the parameter s is just slightly above the smoothing parameter. Here, we focus only on the centered case in which $\boldsymbol{t} = \mathbf{0}$, since as we show in Section 4.5.4, the shifted case is equivalent to the centered case above smoothing.

Recall that the general combiner described above starts with many vectors and then repeatedly takes the average of pairs of vectors that lie in the same coset of $2\mathcal{L}$. We observed that this combiner necessarily needs over 2^n vectors "just to get started" because it works over the 2^n cosets of $2\mathcal{L}$. To get a faster combiner, we therefore try pairing vectors according to the cosets of some sublattice $2\mathcal{L}'$ that "lies between" $\mathcal{L}$ and $2\mathcal{L}$ such that $2\mathcal{L} \subseteq 2\mathcal{L}' \subset \mathcal{L}$. If we simply take many samples from $D_{\mathcal{L},s}$, group them according to their cosets mod $2\mathcal{L}'$, and take their average, analogy with the continuous Gaussian suggests that the resulting vectors will be distributed as roughly $D_{\mathcal{L}',s/\sqrt{2}}$. Note that the parameter has decreased, which is what we wanted, but we are now sampling from a denser lattice. In particular, suppose that we apply this combiner twice, so that in the second step we obtain vectors from some $\mathcal{L}''$. We then expect to obtain samples from roughly $D_{\mathcal{L}'',s/2}$. So, intuitively, if we take $\mathcal{L}''$ to be a sublattice of $\mathcal{L}/2$, we have "made progress." Our running time will be proportional to the index of $2\mathcal{L}'$ over $\mathcal{L}$ (assuming that the index of $2\mathcal{L}''$ over $\mathcal{L}'$ is the same, etc.), so we should take the index of $2\mathcal{L}'$ over $\mathcal{L}$ to be as small as possible. More specifically, we can build a "tower" of progressively denser lattices $(\mathcal{L}_0, \ldots, \mathcal{L}_\ell)$ with the index of $2\mathcal{L}_i$ over $\mathcal{L}_{i-1}$ taken to be slightly larger than $2^{n/2}$.[4] If we take $\mathcal{L}_\ell$ to be the lattice from which we wish to obtain samples with parameter s and $\mathcal{L}_0$ to be a sparse lattice from which we can sample efficiently with parameter $2^{\ell/2}s$, we can hope that iteratively applying such a combiner "up the tower" will yield a sampling algorithm.

As in the description of our 2^n-time combiner, the lemma from [Pei10] shows that the above approach, when instantiated with the "greedy combiner," will yield an algorithm

[4]We note that Becker et al. [BGJ14] also use a tower of lattices in their heuristic algorithm.

that can output vectors whose distribution is statistically close to the discrete Gaussian for parameters s that are above the smoothing parameter. Though this statistical distance can be made small, it is large enough to break applications such as our approximation algorithm for decision SVP.

To avoid this error, the natural hope is that the same combiner used in the 2^n-time algorithm above (the one with the "square sampler") will suffice. Unfortunately, this gives the wrong distribution. In particular, we obtain a distribution in which the cosets of $\mathcal{L}'$ over $\mathcal{L}$ have weight that is proportional to the *square* of their weights over the discrete Gaussian. (See Lemma 4.5.2. Note that when $\mathcal{L}' = \mathcal{L}$ there is only one such coset, which is why our 2^n-time combiner does not run into this problem.) In [ADRS15], we get around this problem by using a "square root sampler," in analogy to our square sampler. In this dissertation, we observe that, since our combiner requires sampled that are "squared" over the cosets of $2\mathcal{L}''$, it suffices to find an algorithm that converts the "squared" distribution over the cosets of $\mathcal{L}'$ to the squared distribution over the cosets of $2\mathcal{L}''$. We show that it suffices to run the square sampled "inside these cosets." (See Section 4.5.) This alternative strategy seems more likely to extend below the smoothing parameter, though we are still unable to achieve this.

Solving *exact* CVP. It is relatively easy to show that $2^{n/2}$ samples from $D_{\mathcal{L},s}$ for an appropriate parameter $s > 0$ will contain a shortest vector. (See Section 4.4.) Similarly, if we could obtain just one sample from $D_{\mathcal{L}-\boldsymbol{t},s}$ for sufficiently small $s > 0$, we would easily be able to solve *exact* CVP. However, we are only able to handle parameters $s > 2^{-o(n/\log n)} \operatorname{dist}(\boldsymbol{t}, \mathcal{L})$. This immediately allows us to approximate CVP up to an extremely good approximation factor $1 + 2^{-o(n/\log n)}$, but it does not solve the exact problem. In [ADS15], we resolve this issue by developing a recursive algorithm that uses certain special properties of our discrete Gaussian sampler. This is outside of the scope of this dissertation, so we do not include the details here.

4.1.3 Open problems and directions for future work

Of course, the most natural and important open problem is whether a faster algorithm for these problems is possible. In recent work with Bennett and Golovnev, we showed some (debatable, but rather convincing to the author) evidence that a $2^{n+o(n)}$-time algorithm for CVP might be optimal [BGS17].[5]

In contrast, it seems very unlikely that the algorithm presented in this work is optimal for SVP. Indeed, there exist certain reasonable heuristics that imply significantly faster sieving algorithms. Assuming these heuristics, the current fastest running time is $(3/2)^{n/2+o(n)} \approx 2^{0.29n}$ [BDGL16]. Furthermore, the techniques used to prove Theorem 4.1.4 seem tantalizingly close to yielding a provable $2^{n/2+o(n)}$-time algorithm for SVP (or perhaps SVP with a relatively small approximation factor), though we have been unable to obtain this result thus far.

In another direction, a long-standing open problem is to find an algorithm that solves SVP or CVP in $2^{O(n)}$ time but *polynomial* space. Currently, the only known algorithms that run in polynomial space are the enumeration-based method of Kannan and its variants, which run in $n^{O(n)}$ time. This is part of the reason why $n^{O(n)}$-time enumeration-based methods are often used in practice to solve large instances of CVP and SVP, in spite of their much worse asymptotic running time.

My co-authors and I are particularly interested in finding a better explanation for why "everything seems to work out" so remarkably well in the analysis of our algorithms. It seems almost magical that we end up with exactly as many samples as we need for our CVP to DGS reduction to go through. We do not have a good intuitive understanding of why our sampler returns the number of samples that it does, but it seems largely unrelated to the reason that

[5] In particular, in [BGS17], we proved that if the Strong Exponential-Time Hypothesis holds, then no faster algorithm exists for CVP in the ℓ_p norm for "almost all" values of p. Unfortunately, our proof technique cannot work for $p = 2$, so that it does not directly apply to the problem considered in this work. We also provide other evidence that significantly faster algorithms—say, $2^{2n/3}$-time algorithms—for CVP in the ℓ_2 norm are unlikely [BGS17].

our CVP algorithm needs as many samples as it does. The fact that these two numbers are the same is remarkable, and we would love a clear explanation.[6] A better understanding of this would be interesting in its own right, and it could lead to an improved algorithm.

4.2 Preliminaries

4.2.1 Sampling with Large Parameter

If we naively applied Corollary 1.3.15 or 1.3.16 and then used the tricks in the sequel to repeatedly lower the parameter by some constant factor, we would still necessarily arrive at an algorithm that only worked for parameters larger than some value proportional to $\lambda_n(\mathcal{L})$ or $\eta_{1/2}(\mathcal{L})$. The following proposition and corollary are what allow us to obtain our main result (Theorem 4.1.1), with a lower bound on the parameter that depends only on $\mathrm{dist}(\boldsymbol{t}, \mathcal{L})$. The trick is simply to work over a sublattice $\mathcal{L}' \subseteq \mathcal{L}$ with the property that (1) $\eta_{1/2}(\mathcal{L}')$ is proportional to $\mathrm{dist}(\boldsymbol{t}, \mathcal{L})$; and (2) all of the short vectors in $\mathcal{L} - \boldsymbol{t}$ are also in $\mathcal{L}' - \boldsymbol{t}$, so that $D_{\mathcal{L}'-\boldsymbol{t},s} \approx D_{\mathcal{L}-\boldsymbol{t},s}$ for sufficiently small parameters $s > 0$.

Proposition 4.2.1 ([ADS15, Proposition 4.5])**.** *There is an algorithm that takes as input a lattice $\mathcal{L} \subset \mathbb{R}^n$, shift $\boldsymbol{t} \in \mathbb{R}^n$, $r > 0$, and parameter $u \geq 2$, such that if*

$$r \geq u^{n/u}(1 + \sqrt{n}u^{n/u}) \cdot \mathrm{dist}(\boldsymbol{t}, \mathcal{L}) \ ,$$

then the output of the algorithm is $\boldsymbol{y} \in \mathcal{L}$ and a basis $\mathbf{B}'$ of a (possibly trivial) sublattice $\mathcal{L}' \subseteq \mathcal{L}$ such that all vectors from $\mathcal{L} - \boldsymbol{t}$ of length at most $r/u^{n/u} - \mathrm{dist}(\boldsymbol{t}, \mathcal{L})$ are also contained in $\mathcal{L}' - \boldsymbol{y} - \boldsymbol{t}$, and $\|\widetilde{\mathbf{B}'}\| \leq r$. The algorithm runs in time $\mathrm{poly}(n) \cdot 2^{O(u)}$.

[6] It is worth noting that our sampling algorithm and our reduction from CVP to DGS both work over the cosets $\mathcal{L}/(2\mathcal{L})$. But, the way in which they use these cosets seems rather different. E.g., the CVP algorithm "views the cosets algebraically," while the reduction "views them geometrically." (In fact, the CVP algorithm due to Micciancio and Voulgaris also works over the cosets $\mathcal{L}/(2\mathcal{L})$.)

Proof. On input a lattice $\mathcal{L} \subset \mathbb{R}^n$, $\boldsymbol{t} \in \mathbb{R}^n$, and $r > 0$, the algorithm behaves as follows. First, it calls the procedure from Theorem 1.2.3 to compute a $u^{n/u}$-HKZ basis $\mathbf{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ of $\mathcal{L}$. Let $(\widetilde{\boldsymbol{b}}_1, \ldots, \widetilde{\boldsymbol{b}}_n)$ be the corresponding Gram-Schmidt vectors. Let $k \geq 0$ be maximal such that $\|\widetilde{\boldsymbol{b}}_i\| \leq r$ for $1 \leq i \leq k$, and let $\mathbf{B}' = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k)$. Let $\pi_k = \pi_{\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k\}^\perp}$ and $\mathcal{M} = \pi_k(\mathcal{L})$. The algorithm then calls the procedure from Theorem 1.2.3 again with the same s and input $\pi_k(\boldsymbol{t})$ and $\mathcal{M}$, receiving as output $\boldsymbol{x} = \sum_{i=k+1}^{n} a_i \pi_k(\boldsymbol{b}_i)$ where $a_i \in \mathbb{Z}$, a $\sqrt{n}u^{n/u}$-approximate closest vector to $\pi_k(\boldsymbol{t})$ in $\mathcal{M}$. Finally, the algorithm returns $\boldsymbol{y} = -\sum_{i=k+1}^{n} a_i \boldsymbol{b}_i$ and $\mathbf{B}' = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k)$.

The running time is clear, as is the fact that $\|\widetilde{\mathbf{B}'}\| \leq r$. It remains to prove that $\mathcal{L}' - \boldsymbol{y} - \boldsymbol{t}$ contains all sufficiently short vectors in $\mathcal{L} - \boldsymbol{t}$. If $k = n$, then $\mathcal{L}' = \mathcal{L}$ and $\boldsymbol{y}$ is irrelevant, so we may assume that $k < n$. Note that, since $\mathbf{B}$ is a $u^{n/u}$-HKZ basis, $\lambda_1(\mathcal{M}) \geq \|\widetilde{\boldsymbol{b}}_{k+1}\|/u^{n/u} > r/u^{n/u}$. In particular, $\lambda_1(\mathcal{M}) > (1 + \sqrt{n} \cdot u^{n/u}) \cdot \operatorname{dist}(\boldsymbol{t}, \mathcal{L}) \geq (1 + \sqrt{n} \cdot u^{n/u}) \cdot \operatorname{dist}(\pi_k(\boldsymbol{t}), \mathcal{M})$. So, there is a unique closest vector to $\pi_k(\boldsymbol{t})$ in $\mathcal{M}$, and by triangle inequality, the next closest vector is at distance greater than $\sqrt{n} \cdot u^{n/u} \operatorname{dist}(\pi_k(\boldsymbol{t}), \mathcal{M})$. Therefore, the call to the subprocedure from Theorem 1.2.3 will output the exact closest vector $\boldsymbol{x} \in \mathcal{M}$ to $\pi_k(\boldsymbol{t})$.

Let $\boldsymbol{w} \in \mathcal{L} \setminus (\mathcal{L}' - \boldsymbol{y})$ so that $\pi_k(\boldsymbol{w}) \neq \pi_k(-\boldsymbol{y}) = \boldsymbol{x}$. We need to show that $\boldsymbol{w} - \boldsymbol{t}$ is relatively long. Since $\mathbf{B}$ is a $s^{n/s}$-HKZ basis, it follows that

$$\|\pi_k(\boldsymbol{w}) - \boldsymbol{x}\| \geq \lambda_1(\mathcal{M}) > r/u^{n/u} \, .$$

Applying triangle inequality, we have

$$\|\boldsymbol{w} - \boldsymbol{t}\| \geq \|\pi_k(\boldsymbol{w}) - \pi_k(\boldsymbol{t})\| \geq \|\pi_k(\boldsymbol{w}) - \boldsymbol{x}\| - \|\boldsymbol{x} - \pi_k(\boldsymbol{t})\| > r/u^{n/u} - \operatorname{dist}(\boldsymbol{t}, \mathcal{L}) \, ,$$

as needed. □

Corollary 4.2.2 ([ADS15, Corollary 4.6])**.** *There is an algorithm that takes as input a lattice $\mathcal{L} \subset \mathbb{R}^n$ with $n \geq 2$, shift $\boldsymbol{t} \in \mathbb{R}^n$, $M \in \mathbb{N}$ (the desired number of output vectors), and*

parameters $u \geq 2$ and $\hat{s} > 0$ and outputs $\boldsymbol{y} \in \mathcal{L}$, a (possibly trivial) sublattice $\mathcal{L}' \subseteq \mathcal{L}$, and M vectors from $\mathcal{L}' - \boldsymbol{y} - \boldsymbol{t}$ such that if

$$\hat{s} \geq C\sqrt{n \log n} \cdot u^{2n/u} \cdot \operatorname{dist}(\boldsymbol{t}, \mathcal{L}) \ ,$$

then the output vectors are distributed as M independent samples from $D_{\mathcal{L}'-\boldsymbol{y}-\boldsymbol{t},\hat{s}}$, and $\mathcal{L}'-\boldsymbol{y}-\boldsymbol{t}$ contains all vectors in $\mathcal{L}-\boldsymbol{t}$ of length at most $C\hat{s}/(u^{n/u}\sqrt{\log n})$. The algorithm runs in time $\operatorname{poly}(n) \cdot 2^{O(u)} + \operatorname{poly}(n) \cdot M$.

Proof. The algorithm first calls the procedure from Proposition 4.2.1 with input $\mathcal{L}$, $\boldsymbol{t}$, and

$$r := \frac{C\hat{s}}{\sqrt{\log n}} \geq u^{n/u}(1 + \sqrt{n}u^{n/u}) \cdot \operatorname{dist}(\boldsymbol{t}, \mathcal{L}) \ ,$$

receiving as output $\boldsymbol{y} \in \mathcal{L}$ and a basis $\mathbf{B}'$ of a sublattice $\mathcal{L}' \subset \mathcal{L}$. It then runs the algorithm from Theorem 1.3.14 M times with input $\mathcal{L}'$, $\boldsymbol{y} + \boldsymbol{t}$, and $\hat{s}$ and outputs the resulting vectors, $\boldsymbol{y}$, and $\mathcal{L}'$.

The running time is clear. By Proposition 4.2.1, $\mathcal{L}' - \boldsymbol{y} - \boldsymbol{t}$ contains all vectors of length at most $r/u^{n/u} - \operatorname{dist}(\boldsymbol{t}, \mathcal{L}) \geq C\hat{s}/(u^{n/u}\sqrt{\log n})$ in $\mathcal{L} - \boldsymbol{t}$, and $\|\widetilde{\mathbf{B}}'\| \leq r \leq C\hat{s}/\sqrt{\log n}$. So, it follows from Theorem 1.3.14 that the output has the correct distribution. □

4.2.2 Some properties of Poisson distributions

Definition 4.2.3 (Poisson distribution)**.** *The Poisson distribution with parameter $\lambda > 0$ is the distribution defined by*

$$\Pr_{X \sim \operatorname{Pois}(\lambda)}[X = r] = \frac{\lambda^r}{r!} \cdot e^{-\lambda}$$

for all $m \in \mathbb{N}$.

Intuitively, the Poisson distribution is the distribution obtained by, e.g., counting the

number of decay events over some fixed time period in some large, homogenous radioactive source. The parameter λ is just the expected count.

Lemma 4.2.4 (Poisson tail bounds [Gly87])**.** *For $\lambda > 0$ let X be a* Pois(λ) *random variable. Then,*

- *for any $0 \leq m < \lambda$,*

$$\Pr(X \leq m) \leq \frac{\exp(-\lambda)\lambda^m}{m!(1-(m/\lambda))} ,$$

- *and for any $m > \lambda - 1$,*

$$\Pr(X \geq m) \leq \frac{\exp(-\lambda)\lambda^m}{m!(1-(\lambda/(m+1)))} .$$

Corollary 4.2.5. *For any $\alpha > 0$, there exist $C_1, C_2 > 0$ such that the following holds for all $m \geq 1$. If X is a* Pois(λ) *random variable for some $\lambda < (1-\alpha)m$ then*

$$\Pr(X \geq m) \leq C_1 \exp(-C_2 m) ,$$

and similarly, if $\lambda > (1+\alpha)m$ then

$$\Pr(X \leq m) \leq C_1 \exp(-C_2 m) .$$

Proof. Stirling's approximation implies the inequality $m! \geq (m/e)^m$ valid for all $m \geq 1$, which together with Lemma 4.2.4 implies in both cases the upper bound

$$C \exp(-m(\lambda/m - \log(\lambda/m) + 1)) .$$

The function $x - \log x + 1$ is non-negative and strictly convex on $x > 0$ and obtains its minimum of 0 at $x = 1$. As a result, it is uniformly bounded away from 0 for all x satisfying

$|x-1| \geq \alpha$. □

Lemma 4.2.6 (Multinomial to independent Poisson)**.** *Let $\lambda > 0$ and $\boldsymbol{p} \in [0,1]^N$ with $\sum p_i = 1$. Consider the process that first samples $r \sim \mathrm{Pois}(\lambda)$ and then samples $X_1, \ldots, X_r$ independently with $\Pr[X_j = i] = p_i$. For each i, let Y_i be the number of occurrences of i in the sequence $X_1, \ldots, X_r$. Then, Y_i is distributed as $\mathrm{Pois}(\lambda p_i)$ independently of the other Y_j.*

Proof. Considering the joint distribution, we have

$$\begin{aligned}
\Pr[\boldsymbol{Y} = \boldsymbol{a}] &= \Pr[r = \|\boldsymbol{a}\|_1] \cdot \Pr[\boldsymbol{Y} = \boldsymbol{a} | r = \|\boldsymbol{a}\|_1] \\
&= \lambda^{\|\boldsymbol{a}\|_1} e^{-\lambda} \prod_i \frac{p_i^{a_i}}{a_i!} \\
&= \prod_i \left(\frac{(\lambda p_i)^{a_i}}{a_i!} \cdot e^{-\lambda p_i} \right) ,
\end{aligned}$$

as needed. □

Claim 4.2.7 (Poisson to Bernoulli)**.** *For $\lambda \leq 1$ and $\kappa \geq 2$, consider the procedure obtained by sampling r from $\mathrm{Pois}(\lambda)$ and then outputting 1 with probability $\min\{1,\ r/\kappa\}$ and 0 otherwise. The output of this procedure is within statistical distance $1/(\lfloor \kappa \rfloor !)$ of the Bernoulli distribution $\mathrm{Bern}(\lambda/\kappa)$.*

Proof. If X is distributed like $\mathrm{Pois}(\lambda)$, the statistical distance is given by

$$\begin{aligned}
\mathbb{E}[X/\kappa - \min\{1,\ X/\kappa\}] &= \mathbb{E}[\max\{0,\ X/\kappa - 1\}] \\
&\leq \kappa^{-1} \, \mathbb{E}[1_{X > \kappa} \cdot X] \\
&= \kappa^{-1} \sum_{r = \lfloor \kappa \rfloor + 1}^{\infty} r \lambda^r \exp(-\lambda)/r! \\
&= \kappa^{-1} \lambda \sum_{r = \lfloor \kappa \rfloor}^{\infty} \lambda^r \exp(-\lambda)/r! \, ,
\end{aligned}$$

which is at most $1/(\lfloor \kappa \rfloor !)$ by Lemma 4.2.4 and our choice of parameters. □

4.3 Sampling from the discrete Gaussian

4.3.1 Sampling from the square

Recall that a naive bucketing procedure does not weight cosets in the way that we would like. In particular, the resulting number of vectors in the cosets is distributed with probabilities essentially proportional to $\rho_s(2\mathcal{L} + \boldsymbol{c} - \boldsymbol{t})$, while we would like the probabilities to be proportional to $\rho_s(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t})^2$. Corollary 4.3.4 shows how to use samples from any multinomial distribution to sample from the "squared distribution" (with small error).

Our presentation is slightly different than that of [ADRS15] and [ADS15]. In particular, these prior works present Corollary 4.3.4 directly, without going through the intermediate result in Theorem 4.3.2. However, we will need Theorem 4.3.2 later.

Proposition 4.3.1 (Estimating $p_{\max}$, [ADRS15, Proposition 3.1])**.** *There is an algorithm that takes as input $\kappa \geq 1$ (the confidence parameter) and a sequence of M elements from $\{1, \ldots, N\}$ and outputs a value $\widetilde{p}_{\max}$ such that, if the input consists of $M \geq \kappa/p_{\max}$ independent samples from the distribution that assigns probability p_i to element i, then*

$$p_{\max} \leq \widetilde{p}_{\max} \leq 4p_{\max}$$

except with probability at most $C_1 N \log N \exp(-C_2\kappa)$, where $p_{\max} := \max p_i$. The algorithm runs in time $M \cdot \mathrm{poly}(\log \kappa, \log N)$.

We now wish to show that, given access to independent samples from a multinomial distribution $(p_1, p_2, \ldots, p_N)$ and a scaling factor $1 \leq T \leq 1/\max p_i$, we can generate samples from $b_i = \mathrm{Bern}(Tp_i)$ *very* efficiently. In particular, we can generate N coins $b_1, \ldots, b_N \in \{0, 1\}$ (i.e., a single coin flip for each i) using just $O(T)$ samples and essentially $O(T)$ time. (Notice that, in order for this to make sense when $N \gg T$, we must represent the coins $b_1, \ldots, b_N$ in some sparse representation—i.e., by simply listing the indices of the non-zero entries.)

Theorem 4.3.2. *There is an algorithm that takes as input $T \geq 1$ (the scaling parameter), $\kappa \geq 2$ (the confidence parameter), and $\lceil \kappa T \rceil$ elements from $\{1, \ldots, N\}$ and outputs a (sparse) vector $\boldsymbol{b} = (b_1, \ldots, b_N) \in \{0,1\}^N$, such that if the input consists of $M \geq \kappa T$ independent samples from the distribution that assigns probability $p_i \leq 1/T$ to element i, then up to statistical distance $C_1 N \exp(-C_2 \kappa)$, the b_i are independently distributed with $b_i = \mathrm{Bern}(Tp_i/\kappa)$. Furthermore, the algorithm runs in time $T \cdot \mathrm{poly}(\log \kappa, \log N, \log T)$ and a coin b_i is non-zero only if the element i appears in the input.*

Proof. The algorithm first samples r according to the distribution $\mathrm{Pois}(T)$. It then restricts its attention to the first r input samples. (If $r > 2T$, then the algorithm simply fails.) For each $i \in \{1, \ldots, N\}$ that appears in these samples, let a_i be the number of appearances. The algorithm then sets b_i to be one with probability $\min\{1, a_i/\kappa\}$ and zero otherwise. (If i did not appear in the first r samples, then $b_i = 0$.)

The running time is clear, as is the fact that a coin b_i is non-zero only if the element i appears in the input. By Corollary 4.2.5, the probability of failure is at most $C_1 \exp(-C_2 \kappa)$. By Lemma 4.2.6, the a_i are distributed exactly as $\mathrm{Pois}(Tp_i)$. Then, by Claim 4.2.7, each b_i is within statistical distance $1/\lfloor \kappa \rfloor ! \leq C \exp(-\kappa)$ of $\mathrm{Bern}(Tp_i/\kappa)$. The result follows by applying union bound. □

From this, we derive the following corollary.

Definition 4.3.3. *For a vector $\boldsymbol{p} \in [0,1]^N$ with $\sum p_i = 1$, let $\boldsymbol{p}^2 = (p_1^2/p_{\mathrm{col}}, \ldots, p_N^2/p_{\mathrm{col}})$ where $p_{\mathrm{col}} := \sum p_i^2$.*

Corollary 4.3.4 (Square sampler, [ADRS15, Theorem 3.3])**.** *There is an algorithm that takes as input $\kappa \geq 2$ (the confidence parameter) and M elements from $\{1, \ldots, N\}$ and outputs a sequence of elements from the same set such that*

1. *the running time is $M \cdot \mathrm{poly}(\log \kappa, \log N)$;*

2. *each* $i \in \{1, \dots, N\}$ *appears at least twice as often in the input as in the output; and*

3. *if the input consists of* $M \geq 10\kappa^3 / \max p_i$ *independent samples from the distribution that assigns probability* p_i *to element* i*, then the output is within statistical distance* $C_1 MN \log N \exp(-C_2\kappa)$ *of* m *independent samples with respective probabilities* $\boldsymbol{p}^2$ *where* $m \geq M \cdot \sum p_i^2/(32\kappa^2 \max p_i)$ *is a random variable.*

Proof. The algorithm uses its first $M/10$ samples to run the procedure from Proposition 4.3.1, to compute $\widetilde{p}_{\max}$ such that $p_{\max} \leq \widetilde{p}_{\max} \leq 4p_{\max}$. It then runs the procedure from Theorem 4.3.2 a total of $\lceil \widetilde{p}_{\max} M/(2\kappa) \rceil$ times, each time with $T := 1/\widetilde{p}_{\max}$, to obtain (a sparse representation of) coins $b_{i,j} \in \{0, 1\}$ for $1 \leq i \leq N$ and $1 \leq j \leq \lceil \widetilde{p}_{\max} M/(2\kappa) \rceil$.

Finally, the algorithm looks through the next $\lceil M/(3\kappa) \rceil$ elements, one element at a time. When it sees element i, it adds it to its output if $b_{i,j} = 1$ where $j \geq 1$ is the smallest index such that $b_{i,j}$ is unused (or it fails if there is no unused $b_{i,j}$).

The running time of the algorithm is clear. And, since the procedure from Theorem 4.3.2 only sets $b_{i,j} = 1$ if there is at least one element i in its input, it follows that there are at least twice as many instances of i in the input to this algorithm as there are in its output.

By Proposition 4.3.1, we have $p_{\max} \leq \widetilde{p}_{\max} \leq 4p_{\max}$ except with probability at most $C_1 N \log N \exp(-C_2\kappa)$. By Theorem 4.3.2 and union bound, we may assume that the $b_{i,j}$ are independently distributed exactly as $\mathrm{Bern}(p_i T/\kappa)$, introducing statistical distance that is at most $C_1 NM \exp(-C_2\kappa)$. Then, in the final stage of the algorithm, the probability of outputting i at each step is $p_i^2/(\kappa \max p_i)$ Hence, the output samples have the correct distribution.

And, by the Chernoff-Hoeffding bound (Lemma 1.4.3), the number of coins $b_{i,j}$ used for some fixed i will not be larger than $Mp_i/(2\kappa) \leq M\widetilde{p}_{\max}/(2\kappa)$ except with probability at most $\exp(-C\kappa)$. It follows from union bound that the total probability of failure is at most $N \exp(-C\kappa)$.

Finally, by the Chernoff-Hoeffding bound again, the size of the output will be at least $M \sum p_i^2/(8\kappa^2 \widetilde{p}_{\max}) \geq M \sum p_i^2/(\kappa^2 \max p_i)$, except with probability at most $\exp(-C\kappa)$. □

4.3.2 A discrete Gaussian combiner

Ideally, we would like the average of two vectors sampled from $D_{\mathcal{L}-\boldsymbol{t},s}$ to be distributed as $D_{\mathcal{L}-\boldsymbol{t},s'}$ for some $s' < s$. Unfortunately, this is false for the simple reason that the average of two vectors in $\mathcal{L} - \boldsymbol{t}$ may not be in $\mathcal{L} - \boldsymbol{t}$! The following lemma shows that we do obtain the desired distribution if we condition on the result being in $\mathcal{L} - \boldsymbol{t}$. Indeed, this lemma is simply a special case of Eq. (3.2), but we give a direct proof anyway. (Note that, for two vectors $\boldsymbol{X}_1, \boldsymbol{X}_2 \in \mathcal{L} - \boldsymbol{t}$, we have $(\boldsymbol{X}_1 + \boldsymbol{X}_2)/2 \in \mathcal{L} - \boldsymbol{t}$ if and only if $\boldsymbol{X}_1 \equiv \boldsymbol{X}_2 \bmod 2\mathcal{L}$.)

Lemma 4.3.5 ([ADS15, Lemma 4.1])**.** *Let $\mathcal{L} \subset \mathbb{R}^n$, $s > 0$ and $\boldsymbol{t} \in \mathbb{R}^n$. Then for all $\boldsymbol{y} \in \mathcal{L} - \boldsymbol{t}$,*

$$\Pr_{(\boldsymbol{X}_1,\boldsymbol{X}_2)\sim D^2_{\mathcal{L}-\boldsymbol{t},s}}[\boldsymbol{X}_1 + \boldsymbol{X}_2 = 2\boldsymbol{y} \mid \boldsymbol{X}_1 \equiv \boldsymbol{X}_2 \bmod 2\mathcal{L}] = \Pr_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t},s/\sqrt{2}}}[\boldsymbol{X} = \boldsymbol{y}]\,. \tag{4.1}$$

Proof. Multiplying the left-hand side of (4.1) by $\Pr_{(\boldsymbol{X}_1,\boldsymbol{X}_2)\sim D^2_{\mathcal{L}-\boldsymbol{t},s}}[\boldsymbol{X}_1 + \boldsymbol{X}_2 \in 2\mathcal{L} - 2\boldsymbol{t}]$, we get for any $\boldsymbol{y} \in \mathcal{L} - \boldsymbol{t}$,

$$\begin{aligned}
\Pr_{(\boldsymbol{X}_1,\boldsymbol{X}_2)\sim D^2_{\mathcal{L}-\boldsymbol{t},s}}[(\boldsymbol{X}_1 + \boldsymbol{X}_2)/2 = \boldsymbol{y}] &= \frac{1}{\rho_s(\mathcal{L}-\boldsymbol{t})^2} \cdot \sum_{\boldsymbol{x}\in\mathcal{L}-\boldsymbol{t}} \rho_s(\boldsymbol{x})\rho_s(2\boldsymbol{y}-\boldsymbol{x}) \\
&= \frac{\rho_{s/\sqrt{2}}(\boldsymbol{y})}{\rho_s(\mathcal{L}-\boldsymbol{t})^2} \cdot \sum_{\boldsymbol{x}\in\mathcal{L}-\boldsymbol{t}} \rho_{s/\sqrt{2}}(\boldsymbol{x}-\boldsymbol{y}) \\
&= \frac{\rho_{s/\sqrt{2}}(\boldsymbol{y})}{\rho_s(\mathcal{L}-\boldsymbol{t})^2} \cdot \rho_{s/\sqrt{2}}(\mathcal{L})\,.
\end{aligned}$$

Hence both sides of (4.1) are proportional to each other. Since they are probabilities, they are actually equal. □

Proposition 4.3.6. *There is an algorithm that takes as input a lattice $\mathcal{L} \subset \mathbb{R}^n$, $\boldsymbol{t} \in \mathbb{R}^n$,*

$\kappa \geq 2$ (the confidence parameter), and a sequence of vectors from $\mathcal{L} - \boldsymbol{t}$, and outputs a sequence of vectors from $\mathcal{L} - \boldsymbol{t}$ such that, if the input consists of

$$M \geq 10\kappa^3 \cdot \frac{\rho_s(\mathcal{L} - \boldsymbol{t})}{\max_{\boldsymbol{c} \in \mathcal{L}/(2\mathcal{L})} \rho_s(2\mathcal{L} + \boldsymbol{c} - \boldsymbol{t})}$$

independent samples from $D_{\mathcal{L}-\boldsymbol{t},s}$ for some $s > 0$, then the output is within statistical distance $M \exp(C_1 n - C_2 \kappa)$ of m independent samples from $D_{\mathcal{L}-\boldsymbol{t},s/\sqrt{2}}$ where m is a random variable with

$$m \geq M \cdot \frac{1}{32\kappa^2} \cdot \frac{\rho_{s/\sqrt{2}}(\mathcal{L}) \cdot \rho_{s/\sqrt{2}}(\mathcal{L} - \boldsymbol{t})}{\rho_s(\mathcal{L} - \boldsymbol{t}) \max_{\boldsymbol{c} \in \mathcal{L}/(2\mathcal{L})} \rho_s(2\mathcal{L} + \boldsymbol{c} - \boldsymbol{t})} .$$

The running time of the algorithm is at most $M \cdot \mathrm{poly}(n, \log \kappa)$.

Proof. Let $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M)$ be the input vectors. For each i, let $\boldsymbol{c}_i \in \mathcal{L}/(2\mathcal{L})$ be such that $\boldsymbol{X}_i \in 2\mathcal{L} + \boldsymbol{c}_i - \boldsymbol{t}$. The algorithm runs the procedure from Corollary 4.3.4 with input κ and $(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_M)$, receiving output $(\boldsymbol{c}'_1, \ldots, \boldsymbol{c}'_m)$. (Formally, we must encode the cosets as integers in $\{1, \ldots, 2^n\}$.) Finally, for each $\boldsymbol{c}'_i$, it chooses a pair of unpaired vectors $\boldsymbol{X}_j, \boldsymbol{X}_k$ with $\boldsymbol{c}_j = \boldsymbol{c}_k = \boldsymbol{c}'_i$ and outputs $\boldsymbol{Y}_i = (\boldsymbol{X}_j + \boldsymbol{X}_k)/2$.

The running time of the algorithm follows from Item 1 of Corollary 4.3.4. Furthermore, we note that by Item 2 of the same corollary, there will always be a pair of indices j, k for each i as above.

To prove correctness, we observe that for $\boldsymbol{c} \in \mathcal{L}/(2\mathcal{L})$ and $\boldsymbol{y} \in 2\mathcal{L} + \boldsymbol{c} - \boldsymbol{t}$,

$$\Pr[\boldsymbol{X}_i = \boldsymbol{y}] = \frac{\rho_s(2\mathcal{L} + \boldsymbol{c} - \boldsymbol{t})}{\rho_s(\mathcal{L} - \boldsymbol{t})} \cdot \Pr_{\boldsymbol{X} \sim D_{2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t},s}}[\boldsymbol{X} = \boldsymbol{y}] .$$

In particular, we have that $\Pr[\boldsymbol{c}_i = \boldsymbol{c}] = \rho_s(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t})/\rho_s(\mathcal{L}-\boldsymbol{t})$. Then, the cosets $(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_M)$ satisfy the conditions necessary for Item 3 of Corollary 4.3.4.

Applying Corollary 4.3.4, up to statistical distance $M \exp(C_1 n - C_2 \kappa)$, we have that the

output vectors are independent, and

$$\begin{aligned} m &\geq M \cdot \frac{1}{32\kappa^2} \cdot \frac{\sum_{\boldsymbol{c} \in \mathcal{L}/(2\mathcal{L})} \rho_s(2\mathcal{L} + \boldsymbol{c} - \boldsymbol{t})^2}{\rho_s(\mathcal{L} - \boldsymbol{t}) \max_{\boldsymbol{c} \in \mathcal{L}/(2\mathcal{L})} \rho_s(2\mathcal{L} + \boldsymbol{c} - \boldsymbol{t})} \\ &= M \cdot \frac{1}{32\kappa^2} \cdot \frac{\rho_{s/\sqrt{2}}(\mathcal{L}) \cdot \rho_{s/\sqrt{2}}(\mathcal{L} - \boldsymbol{t})}{\rho_s(\mathcal{L} - \boldsymbol{t}) \max_{\boldsymbol{c} \in \mathcal{L}/(2\mathcal{L})} \rho_s(2\mathcal{L} + \boldsymbol{c} - \boldsymbol{t})} , \end{aligned}$$

where the equality follows from Eq. (3.2) by setting $\boldsymbol{u} = \boldsymbol{0}$. Furthermore, we have $\Pr[\boldsymbol{c}'_i = \boldsymbol{c}] = \rho_s(2\mathcal{L} + \boldsymbol{c} - \boldsymbol{t})^2 / \sum_{\boldsymbol{c}'} \rho_s(2\mathcal{L} + \boldsymbol{c}' - \boldsymbol{t})^2$ for any coset $\boldsymbol{c} \in \mathcal{L}/(2\mathcal{L})$. Therefore, for any $\boldsymbol{y} \in \mathcal{L}$,

$$\begin{aligned} \Pr[\boldsymbol{Y}_i = \boldsymbol{y}] &= \frac{1}{\sum \rho_s(2\mathcal{L} + \boldsymbol{c} - \boldsymbol{t})^2} \cdot \sum_{\boldsymbol{c} \in \mathcal{L}/(2\mathcal{L})} \rho_s(2\mathcal{L} + \boldsymbol{c} - \boldsymbol{t})^2 \cdot \Pr_{(\boldsymbol{X}_j, \boldsymbol{X}_k) \sim D^2_{2\mathcal{L} + \boldsymbol{c} - \boldsymbol{t}, s}} [(\boldsymbol{X}_j + \boldsymbol{X}_k)/2 = \boldsymbol{y}] \\ &= \Pr_{(\boldsymbol{X}_1, \boldsymbol{X}_2) \sim D^2_{\mathcal{L} - \boldsymbol{t}, s}} [(\boldsymbol{X}_1 + \boldsymbol{X}_2)/2 = \boldsymbol{y} \mid \boldsymbol{X}_1 + \boldsymbol{X}_2 \in 2\mathcal{L} - 2\boldsymbol{t}] . \end{aligned}$$

The result then follows from Lemma 4.3.5. □

We will show in Theorem 4.3.7 that by calling the algorithm from Proposition 4.3.6 repeatedly, we obtain a general discrete Gaussian combiner.

Theorem 4.3.7. *There is an algorithm that takes as input a lattice $\mathcal{L} \subset \mathbb{R}^n$, $\ell \in \mathbb{N}$ (the step parameter), $\kappa \geq 2$ (the confidence parameter), $\boldsymbol{t} \in \mathbb{R}^n$, and $M = (32\kappa)^{2\ell+2} \cdot 2^n$ vectors in $\mathcal{L}$ such that, if the input vectors are distributed as $D_{\mathcal{L}-\boldsymbol{t},s}$ for some $s > 0$, then the output is a list of vectors whose distribution is within statistical distance $\ell M \exp(C_1 n - C_2 \kappa)$*

$$m \geq \begin{cases} 2^{n/2} & \boldsymbol{t} = \boldsymbol{0} \\ \frac{\rho_{2^{-\ell/2}s}(\mathcal{L} - \boldsymbol{t})}{\max_{\boldsymbol{c} \in \mathcal{L}/(2\mathcal{L})} \rho_{2^{-\ell/2}s}(2\mathcal{L} + \boldsymbol{c} - \boldsymbol{t})} & \text{otherwise} \end{cases}$$

independent samples from $D_{\mathcal{L}-\boldsymbol{t}, 2^{-\ell/2}s}$. The algorithm runs in time $\ell M \cdot \mathrm{poly}(n, \log \kappa)$.

Proof. Let $\mathcal{X}_0 = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M)$ be the sequence of input vectors. For $i = 0, \ldots, \ell - 1$, the algorithm calls the procedure from Proposition 4.3.6 with input $\mathcal{L}$, κ, and $\mathcal{X}_i$, receiving an

output sequence $\mathcal{X}_{i+1}$ of length M_{i+1}. Finally, the algorithm outputs the sequence $\mathcal{X}_\ell$.

The running time is clear. Fix $\mathcal{L}$, s, $\boldsymbol{t}$ and ℓ. Define $\theta(i) := \rho_{2^{-i/2}s}(\mathcal{L})$, $\phi(i) := \max_{\boldsymbol{c}\in\mathcal{L}/(2\mathcal{L})} \rho_{2^{-i/2}s}(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t})$, and $\psi(i) := \rho_{2^{-i/2}s}(\mathcal{L}-\boldsymbol{t})$.

We will first prove the result for the case $\boldsymbol{t} \neq \boldsymbol{0}$. We wish to prove by induction that $\mathcal{X}_i$ is within statistical distance $iM\exp(C_1 n - C_2\kappa)$ of $D^{M_i}_{\mathcal{L}-\boldsymbol{t},2^{-i/2}s}$ with

$$M_i \geq (32\kappa)^{2\ell-2i+2} \cdot \frac{\psi(i)}{\phi(i)} \,, \tag{4.2}$$

for all $i \geq 1$. In particular, with $i = \ell$, this implies the result.

Let

$$L(i) := \frac{\theta(i+1)\psi(i+1)}{\psi(i)\phi(i)} \,,$$

be the "loss factor" resulting from the $(i+1)$st run of the combiner, ignoring the factor of $32\kappa^2$. By Theorem 3.2.2, we have

$$L(i) \geq \frac{\psi(i+1)}{\phi(i+1)} \cdot \frac{\phi(i)}{\psi(i)} \,. \tag{4.3}$$

By Proposition 4.3.6, up to statistical distance $M\exp(C_1 n - C_2\kappa)$, we have that $\mathcal{X}_1$ has the right distribution with

$$\begin{aligned} M_1 &\geq \frac{1}{32\kappa^2} \cdot M_0 \cdot L(0) \\ &\geq (32\kappa)^{2\ell} \cdot 2^n \cdot \frac{\psi(1)}{\phi(1)} \cdot \frac{\phi(0)}{\psi(0)} \,, \end{aligned}$$

where we used Eq. (4.3) with $i = 0$. By noting that $\psi(0) \leq 2^n\phi(0)$ (by Lemma 1.3.3), we see that (4.2) holds when $i = 1$.

Suppose that $\mathcal{X}_i$ has the correct distribution and (4.2) holds for some i with $0 \leq i < \ell$. In particular, we have that M_i is at least $10\kappa^2\psi(i)/\phi(i)$. This is precisely the condition necessary

to apply Proposition 4.3.6. So, we can apply the proposition and the induction hypothesis and obtain that (up to statistical distance at most $(i+1)M\exp(C_1 n - C_2\kappa)$), $\mathcal{X}_{i+1}$ has the correct distribution with

$$M_{i+1} \geq \frac{1}{32\kappa^2}\cdot M_i \cdot L(i) \geq (32\kappa)^{2\ell-2i}\cdot\frac{\psi(i)}{\phi(i)}\cdot\frac{\phi(i)}{\psi(i)}\cdot\frac{\psi(i+1)}{\phi(i+1)} = (32\kappa)^{2\ell-2i}\cdot\frac{\psi(i+1)}{\phi(i+1)}\,,$$

where in the second inequality we used the induction hypothesis and Eq. (4.3).

To prove that $m \geq 2^{n/2}$ when $\boldsymbol{t}=\boldsymbol{0}$, we observe that when $\boldsymbol{t}=\boldsymbol{0}$, $\theta(i)=\psi(i)$ and $\phi(i) = \max_{\boldsymbol{c}\in\mathcal{L}/(2\mathcal{L})}\rho_{2^{-i/2}s}(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t}) = \rho_{2^{-i/2}s}(2\mathcal{L}) = \psi(i+2)$. Therefore, in this case, the loss factor $L(i)$ is

$$L(i) = \frac{\theta(i+1)\psi(i+1)}{\psi(i)\phi(i)} = \frac{\psi(i+1)^2}{\psi(i)\psi(i+2)}\,.$$

Therefore, the product of the $L(i)$ is a telescoping product, with

$$\prod_i L(i) = \frac{\psi(2)}{\psi(1)}\cdot\frac{\psi(\ell+1)}{\psi(\ell+2)} \geq 2^{-n/2}\,,$$

and the result follows. □

4.3.3 Finishing the proof of Theorem 4.1.1

Theorem 4.1.1 follows more-or-less immediately from Theorem 4.3.7 together with Corollary 4.2.2. Here, we present a more formal version of Theorem 4.1.1.

Theorem 4.3.8 ([ADRS15, ADS15]). *For any efficiently computable function $f(n) \geq n^C$, let σ be the function defined by $\sigma(\mathcal{L}-\boldsymbol{t}) := \mathrm{dist}(\boldsymbol{t},\mathcal{L})/f(n)$ for any lattice $\mathcal{L}\subset\mathbb{R}^n$ and $\boldsymbol{t}\in\mathbb{R}^n$, and let*

$$m(\mathcal{L}-\boldsymbol{t},s) := \begin{cases} 2^{n/2} & \boldsymbol{t}\in\mathcal{L} \\ \frac{\rho_s(\mathcal{L}-\boldsymbol{t})}{\max_{\boldsymbol{c}\in\mathcal{L}/(2\mathcal{L})}\rho_s(2\mathcal{L}+\boldsymbol{c}-\boldsymbol{t})} & \textit{otherwise.}\end{cases}$$

Then, there is an algorithm that solves ε-DGS$^m_\sigma$ with $\varepsilon(n) := 2^{-f(n)}$ in time $2^{n+O(\log n \log f(n))}$.

Proof. We assume without loss of generality that $f(n) > 10$. The algorithm behaves as follows on input a lattice $\mathcal{L} \subset \mathbb{R}^n$, a shift $\boldsymbol{t} \in \mathcal{L}$, and a parameter $s > \sigma(\mathcal{L} - \boldsymbol{t})$. First, it runs the procedure from Corollary 4.2.2 with input $\mathcal{L}$, $\boldsymbol{t}$, $M := (Cf(n))^{2\ell+2} \cdot 2^n$ with $\ell := C\lceil \log f(n) \rceil$, $u := Cn \log n / \log f(n) + 2$, and

$$\hat{s} := 2^\ell s > C\sqrt{n \log n} \cdot u^{2n/u} \cdot \operatorname{dist}(\boldsymbol{t}, \mathcal{L}) \ .$$

(Note that $u^{n/u} \le f(n)^C$.) It receives as output $\mathcal{L}' \subset \mathbb{R}^n$, $\boldsymbol{y} \in \mathcal{L}$, and $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M) \in \mathcal{L}' - \boldsymbol{y} - \boldsymbol{t}$. It then runs the procedure from Theorem 4.3.7 *twice*, first with input $\mathcal{L}'$, ℓ, $\kappa := 10f(n)$, $\boldsymbol{t}$, and the first half of the vectors, $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{M/2})$; and next with input $\mathcal{L}'$, ℓ, κ, $\boldsymbol{t}$, and the second half of the vectors, $(\boldsymbol{X}_{M/2+1}, \ldots, \boldsymbol{X}_M)$. Finally, it outputs the resulting vectors. (We run the procedure twice simply to double the output size.)

The running time follows from the respective running times of the two subprocedures. In particular, the procedure from Corollary 4.2.2 runs in time $\operatorname{poly}(n) \cdot (2^{O(u)} + M) = n^{O(n/\log f(n))} + 2^{n+O(\log n \log f(n))} = 2^{n+O(\log n \log f(n))}$, and the procedure from Theorem 4.3.7 runs in time $\ell M \cdot \operatorname{poly}(n, \log \kappa) = 2^{n+O(\log n \log f(n))}$.

By Corollary 4.2.2, the $\boldsymbol{X}_i$ are M independent samples from $D_{\mathcal{L}'-\boldsymbol{y}-\boldsymbol{t},\hat{s}}$ and $\mathcal{L}' - \boldsymbol{y} - \boldsymbol{t}$ contains all vectors in $\mathcal{L} - \boldsymbol{t}$ of length at most $C\hat{s}/(u^{n/u}\sqrt{\log n})$. By Theorem 4.3.7, the output contains at least $2m(\mathcal{L}' - \boldsymbol{t}, s)$ vectors whose joint distribution is within statistical distance $2^{-2f(n)}$ of independent samples from $D_{\mathcal{L}'-\boldsymbol{y}-\boldsymbol{t},s}$.

We now show that $D_{\mathcal{L}'-\boldsymbol{y}-\boldsymbol{t},s}$ is statistically close to $D_{\mathcal{L}-\boldsymbol{t},s}$. Let $d := \operatorname{dist}(\boldsymbol{t}, \mathcal{L})$ and

$$r := \frac{C2^\ell}{u^{n/u}\sqrt{n \log n}} \ge f(n)^C \ge \frac{1}{\sqrt{2\pi}} \ .$$

The statistical distance is exactly

$$\begin{aligned}\Pr_{\boldsymbol{w}\sim D_{\mathcal{L}-\boldsymbol{t},s}}\left[\boldsymbol{w}\notin\mathcal{L}'-\boldsymbol{y}-\boldsymbol{t}\right] &< \Pr_{\boldsymbol{w}\sim D_{\mathcal{L}-\boldsymbol{t},s}}\left[\|\boldsymbol{w}\|>C\hat{s}/(u^{n/u}\sqrt{\log n})\right]\\ &= \Pr_{\boldsymbol{w}\sim D_{\mathcal{L}-\boldsymbol{t},s}}\left[\|\boldsymbol{w}\|>rs\sqrt{n}\right]\\ &< e^{\pi d^2/s^2}e^{-f(n)^C}\\ &< \varepsilon/M\ ,\end{aligned}$$

where we have used Corollary 1.3.11. It follows that the output has the correct size and distribution. In particular, it follows from applying union bound over the output samples that the joint distribution of the output samples is within statistical distance ε of independent samples from $D_{\mathcal{L}-\boldsymbol{t},s}$, and an easy calculation shows that $2m(\mathcal{L}'-\boldsymbol{t},s)>m(\mathcal{L}-\boldsymbol{t},s)$. □

4.4 Solving SVP and (approximate) CVP in $2^{n+o(n)}$ time

4.4.1 A bound on the Gaussian mass

In this section, we prove a bound on the Gaussian mass of a lattice that follows from an upper bound on the kissing number due to Kabatjanskiĭ and Levenšteĭn [KL78]. In particular, we use the following lemma from [PS09] based on [KL78]. For convenience, we define $\beta := 2^{0.401}$, and we use this notation throughout this section.

Lemma 4.4.1 ([PS09, Lemma 3])**.** *Let $\mathcal{L}\subseteq\mathbb{R}^n$ be a lattice with $\lambda_1(\mathcal{L})=1$. Then for any $r\geq 1$, the number of lattice vectors of length at most r is at most $\beta^{n+o(n)}r^n$.*

We now use Lemma 4.4.1 to bound $\rho_s(\mathcal{L})$.

Lemma 4.4.2 ([ADRS15, Lemma 4.2])**.** *Let $\mathcal{L}\subset\mathbb{R}^n$ be a lattice of rank at least one. Then*

for any $s > \sqrt{2\pi/n} \cdot \lambda_1(\mathcal{L})$,

$$\rho_s(\mathcal{L}) \le 1 + \left(\frac{\beta^2 s^2 n}{2\pi e \cdot \lambda_1(\mathcal{L})^2}\right)^{n/2+1} 2^{o(n)} , \tag{4.4}$$

and for $s \le \sqrt{2\pi/n} \cdot \lambda_1(\mathcal{L})$, *we have*

$$\rho_s(\mathcal{L}) \le 1 + e^{-\pi\lambda_1(\mathcal{L})^2/s^2} \cdot \beta^{n+o(n)} . \tag{4.5}$$

We note that an easy calculation shows that the right-hand side of Eq. (4.4) is never smaller than the right-hand side of Eq. (4.5). In particular, this means that Eq. (4.4) actually applies for all s.

Proof of Lemma 4.4.2. We assume without loss of generality that $\mathcal{L}$ is normalized so that $\lambda_1(\mathcal{L}) = 1$. Let $t := 1 + 1/n$. For $r \ge 1$, define $T_r := \{\boldsymbol{x} \in \mathbb{R}^n : r \le \|\boldsymbol{x}\| < tr\}$. By Lemma 4.4.1, $|\mathcal{L} \cap T_r| \le \beta^{n+o(n)} r^n$, and, for any vector $\boldsymbol{y} \in \mathcal{L} \cap T_r$, $\rho_s(\boldsymbol{y}) \le e^{-\pi r^2/s^2}$. Therefore,

$$\rho_s(\mathcal{L} \cap T_r) \le e^{-\pi r^2/s^2} \cdot \beta^{n+o(n)} \cdot r^n .$$

So, we have

$$\begin{aligned}
\rho_s(\mathcal{L}) &= 1 + \sum_{i=0}^{\infty} \rho_s(\mathcal{L} \cap T_{t^i}) \\
&\le 1 + \beta^{n+o(n)} \cdot \sum_{i=0}^{\infty} e^{-\pi t^{2i}/s^2} t^{in} \\
&\le 1 + (1+s)\beta^{n+o(n)} \cdot \max_{r \ge 1} e^{-\pi r^2/s^2} r^n ,
\end{aligned}$$

where we have used the fact that $e^{-\pi t^{2i}/s^2} t^{in}$ decays geometrically when i is at least, say, $(1+s) \cdot \mathrm{poly}(n)$, and so the sum up to that point is the same as the infinite sum up to a constant factor. Note that for any $a, b > 0$, the maximum of $e^{-ar^2} r^b$ over the interval $r \ge 1$

is obtained at $r = \sqrt{b/(2a)}$ if this value is at least 1 or at $r = 1$ otherwise. The result follows. □

Proposition 4.4.3 ([ADRS15, Proposition 4.3])**.** *Let $\mathcal{L} \subset \mathbb{R}^n$ be a lattice of rank at least one. Let*

$$s = \sqrt{\frac{2\pi e}{\beta^2 n}} \cdot \lambda_1(\mathcal{L}) \ .$$

Then,

$$\Pr_{\boldsymbol{X} \sim D_{\mathcal{L},s}}[\|\boldsymbol{X}\| = \lambda_1(\mathcal{L})] \geq e^{-\beta^2 n/(2e) - o(n)} \approx 1.38^{-n-o(n)} \ .$$

Proof. By Lemma 4.4.2, we have that $\rho_s(\mathcal{L}) = 2^{o(n)}$. Therefore,

$$\Pr_{\boldsymbol{X} \sim D_{\mathcal{L},s}}[\|\boldsymbol{X}\| = \lambda_1(\mathcal{L})] \geq e^{-\pi/s^2}/\rho_s(\mathcal{L}) \geq e^{-\pi/s^2 - o(n)} = e^{-\beta^2 n/(2e) - o(n)} \ ,$$

as needed. □

An easy calculation shows that the probability in Proposition 4.4.3 is maximized to within a factor of two when $s = 1/\eta_1(\mathcal{L}^*)$. I.e., for any shortest non-zero vector $\boldsymbol{y} \in \mathcal{L}$,

$$\max_s \Pr_{\boldsymbol{X} \sim D_{\mathcal{L},s}}[\boldsymbol{X} = \boldsymbol{y}] \leq 2 \Pr_{\boldsymbol{X} \sim D_{\mathcal{L},1/\eta_1(\mathcal{L}^*)}}[\boldsymbol{X} = \boldsymbol{y}] = \exp(-\pi \eta_1(\mathcal{L}^*)^2 \cdot \lambda_1(\mathcal{L})^2) \ .$$

4.4.2 A reduction from SVP to DGS

Theorem 4.4.4 ([ADRS15, Theorem 4.4])**.** *There is a reduction from SVP to $\frac{1}{2}$-DGS$^{2^{n/2}}$. The reduction makes $O(n)$ calls to the DGS oracle, preserves the dimension of the lattice, and runs in time $2^{n/2} \cdot \mathrm{poly}(n)$.*

Proof. Let $\mathcal{D}$ be an oracle solving $\frac{1}{2}$-DGS$^{2^{n/2}}$. The reduction first runs the procedure from Theorem 1.2.3 on $\mathcal{L}$ with $r = 2$. Let d be the length of the first basis vector in the output. For $i = 0, \ldots, 100n$, the reduction calls $\mathcal{D}$ on $\mathcal{L}$ with parameter $s_i = 1.01^{-i} \cdot d$. Let $\boldsymbol{x}_i$ be

a shortest non-zero vector in the output. Finally, the reduction outputs a shortest vector among the $\boldsymbol{x}_i$.

The running time of the reduction is clear. By Theorem 1.2.3, we have $d \leq 2^{n/2}\lambda_1(\mathcal{L})$. It follows that there exists some i such that $\hat{s} \leq s_i \leq 1.01\hat{s}$ where $\hat{s} = \sqrt{\pi e/n} \cdot 2^{0.099} \cdot \lambda_1(\mathcal{L})$ (i.e., $\hat{s}$ is the parameter from Proposition 4.4.3). We assume that the output of $\mathcal{D}$ is exactly $D_{\mathcal{L},s_i}^{2^{n/2}}$ when called on s_i, incurring statistical distance at most $1/2$. For such i, by Lemma 1.3.3 we have

$$\begin{aligned}\Pr_{\boldsymbol{X}\sim D_{\mathcal{L},s_i}}[\|\boldsymbol{X}\| = \lambda_1(\mathcal{L})] &\geq 1.01^{-n} \cdot \Pr_{\boldsymbol{X}\sim D_{\mathcal{L},\hat{s}}}[\|\boldsymbol{X}\| = \lambda_1(\mathcal{L})] \\ &\geq 1.4^{-n-o(n)} \qquad \text{(Proposition 4.4.3)} .\end{aligned}$$

The result follows by noting that $1.4 < \sqrt{2}$, so $2^{n/2}$ samples from $D_{\mathcal{L},s_i}$ will contain a shortest vector with probability at least $1 - \exp(-\Omega(n))$. □

Corollary 4.4.5. *There is an algorithm that solves SVP in time $2^{n+o(n)}$.*

Proof. Combine the reduction from Theorem 4.4.4 with the algorithm from Theorem 4.3.8. □

4.4.3 Approximate CVP

There is no analogue of Proposition 4.4.3 for CVP. In particular, for any $n \geq 2$, $\varepsilon > 0$, and $\alpha > 0$, there is a lattice $\mathcal{L} \subset \mathbb{R}^n$ and target $\boldsymbol{t} \in \mathbb{R}^n$ such that

$$\Pr_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t},\alpha\,\mathrm{dist}(\boldsymbol{t},\mathcal{L})}}[\|\boldsymbol{X}\| = \mathrm{dist}(\boldsymbol{t},\mathcal{L})] < \varepsilon .$$

This is because there is in general no bound on the number of γ-approximate closest vectors for any $\gamma > 1$. (Contrast this with Lemma 4.4.1.) So, it is not immediately obvious how to turn Theorem 4.3.8 into an algorithm for *exact* CVP.

Nevertheless, in [ADS15], we show how to obtain such an algorithm that runs in time $2^{n+o(n)}$. This requires a lot more work and relies crucially on the fact that Theorem 4.3.8 guarantees many samples from $D_{\mathcal{L}-\boldsymbol{t},s}$ whenever many cosets contain close vectors. This algorithm is outside of the scope of this thesis, so here we simply observe that Theorem 4.3.8 yields an approximate CVP algorithm with an extremely good approximation factor.

Theorem 4.4.6. *For any efficiently computable function $f(n) \geq n^C$, there is an algorithm that solves γ-CVP in time $2^{n+O(\log(n)\log f(n))}$ for*

$$\gamma := 1 + \frac{2+\sqrt{2n/\pi}}{f(n)} \ .$$

Proof. The algorithm takes as input a lattice $\mathcal{L} \subset \mathbb{R}^n$ and target $\boldsymbol{t} \in \mathbb{R}^n$ and uses Babai's algorithm (Theorem 1.2.4) to find $\widetilde{d}$ with $\operatorname{dist}(\boldsymbol{t},\mathcal{L}) \leq \widetilde{d} \leq 2^{n/2}\operatorname{dist}(\boldsymbol{t},\mathcal{L})$. If $\widetilde{d} = 0$ (i.e., if $\boldsymbol{t} \in \mathcal{L}$), then the algorithm simply outputs $\boldsymbol{t}$. Otherwise, for $i = 0,\ldots,n$, let $d_i := 2^{-i/2}\widetilde{d}$ and let $s_i := d_i/f(n)$. For each i, the algorithm calls the procedure from Theorem 4.3.8 with input $\mathcal{L}$, $\boldsymbol{t}$, and s_i. Let $\boldsymbol{x} \in \mathcal{L}-\boldsymbol{t}$ be the shortest vector in the output. The algorithm outputs $\boldsymbol{x}+\boldsymbol{t} \in \mathcal{L}$.

The running time is clear. Let i such that $\operatorname{dist}(\boldsymbol{t},\mathcal{L}) < d_i < 2\operatorname{dist}(\boldsymbol{t},\mathcal{L})$. By Theorem 4.3.8, the output of the subprocedure is at least one vector that is statistically close to $D_{\mathcal{L}-\boldsymbol{t},s_i}$. By Corollary 1.3.11,

$$\Pr_{\boldsymbol{X}\sim D_{\mathcal{L}-\boldsymbol{t},s_i}}[\|\boldsymbol{X}\| \leq \gamma\|\operatorname{dist}(\boldsymbol{t},\mathcal{L})\|] < \exp(-\pi x^2) \ ,$$

where $x := (\gamma-1)\operatorname{dist}(\boldsymbol{t},\mathcal{L})/s_i - \sqrt{n/(2\pi)} \geq 1$. Therefore, with at least constant probability, the output will be a solution to γ-CVP. □

4.5 Sampling $2^{n/2}$ vectors above smoothing in $2^{n/2}$ time

In this section we present a $2^{n/2}$-time algorithm for DGS_σ with σ approximately the smoothing parameter. Here, we focus on the *centered* case when the shift $\boldsymbol{t}$ is zero because, as we show in Section 4.5.4, the shifted case reduces to the centered case above smoothing.

Recall that, in Section 4.3, we showed that the average of two independent samples from $D_{\mathcal{L},s}$ conditioned on the result landing in $\mathcal{L}$ is distributed exactly as $D_{\mathcal{L},s/\sqrt{2}}$ (Lemma 4.3.5). We then used the "square sampler" (Corollary 4.3.4) to sample efficiently from this conditional distribution. Indeed, we observed that sampling from this conditional distribution boiled down to sampling from the "squared" distribution, $D^{(2)}_{\mathcal{L},2\mathcal{L},s}$ defined by

$$\Pr_{\boldsymbol{X}\sim D^{(2)}_{\mathcal{L}/(2\mathcal{L}),s}}[\boldsymbol{X}=\boldsymbol{y}] := \frac{\rho_s(2\mathcal{L}+\boldsymbol{y})\rho_s(\boldsymbol{y})}{\sum_{\boldsymbol{d}\in\mathcal{L}/(2\mathcal{L})}\rho_s(2\mathcal{L}+\boldsymbol{d})^2} = \frac{\rho_s(2\mathcal{L}+\boldsymbol{y})^2}{\sum_{\boldsymbol{d}\in\mathcal{L}/(2\mathcal{L})}\rho_s(2\mathcal{L}+\boldsymbol{d})^2}\cdot\Pr_{\boldsymbol{X}\sim D_{2\mathcal{L}+\boldsymbol{y},s}}[\boldsymbol{X}=\boldsymbol{y}]\,.$$

Our algorithm necessarily ran in at least 2^n time because it had to work over the 2^n cosets of $2\mathcal{L}$. It is therefore natural to ask what happens when we try to work over cosets of some sublattice with lower index. Notice that, if we take the average of two lattice vectors that are not in the same coset of $2\mathcal{L}$, the result will lie in some *superlattice* $\mathcal{L}'$ of $\mathcal{L}$. In particular, for some superlattice $\mathcal{L}'\supseteq\mathcal{L}$, the average of two vectors from $\mathcal{L}$ will lie in $\mathcal{L}'$ if and only if they lie in the same coset mod $2\mathcal{L}'$. Furthermore, $2\mathcal{L}'$ will be a sublattice of $\mathcal{L}$ if and only if $\mathcal{L}'\subseteq\mathcal{L}/2$. So, we consider lattices $\mathcal{L}'$ with $\mathcal{L}\subseteq\mathcal{L}'\subseteq\mathcal{L}/2$ and study the distribution of the average of two discrete Gaussian vectors *conditioned on the result landing in* $\mathcal{L}'$. We will show in Section 4.5.2 that the resulting distribution is the discrete Gaussian "squared" over the cosets of $\mathcal{L}'/\mathcal{L}$. I.e., for $\boldsymbol{y}\in\mathcal{L}'$,

$$\Pr_{\boldsymbol{X}_1,\boldsymbol{X}_2\sim D_{\mathcal{L},s}}[\boldsymbol{X}_1+\boldsymbol{X}_2=2\boldsymbol{y}\mid\boldsymbol{X}_1\equiv\boldsymbol{X}_2 \bmod 2\mathcal{L}'] = \Pr_{\boldsymbol{X}\sim D^{(2)}_{\mathcal{L}',\mathcal{L},s/\sqrt{2}}}[\boldsymbol{X}=\boldsymbol{y}]\,. \tag{4.6}$$

Notice in particular that our $2^{n+o(n)}$-time algorithm relies on Eq. (4.6) in the special case when $\mathcal{L}' = \mathcal{L}$, in which case the squared distribution $D^{(2)}_{\mathcal{L},\mathcal{L},s}$ is the same as the discrete Gaussian. But, intuitively, as long as $\mathcal{L}'$ is not too much denser than $\mathcal{L}$, Eq. (4.6) should allow us to "make progress," in the sense that the drop in parameter by a factor of $\sqrt{2}$ should more than compensate for the fact that we are now sampling over a denser lattice.

In order to sample from the conditional distribution in Eq. (4.6), it suffices to sample from the squared distribution $D^{(2)}_{\mathcal{L},2\mathcal{L}',s}$. So, Eq. (4.6) gives us a way to convert samples from $D^{(2)}_{\mathcal{L},2\mathcal{L}',s}$ into samples from $D^{(2)}_{\mathcal{L}',\mathcal{L},s/\sqrt{2}}$. If we wanted to repeat the process with some new superlattice $\mathcal{L}'$ with $\mathcal{L}' \subseteq \mathcal{L}'' \subseteq \mathcal{L}'/2$, we would have to obtain samples from $D^{(2)}_{\mathcal{L}',2\mathcal{L}'',s/\sqrt{2}}$. In [ADRS15], we showed a "square root sampler," which allowed us to convert "squared" samples from $D^{(2)}_{\mathcal{L}',\mathcal{L},s/\sqrt{2}}$ to "unsquared" samples from $D_{\mathcal{L}',s/\sqrt{2}}$. We then used the square sampler to convert these samples into samples from $D^{(2)}_{\mathcal{L}',2\mathcal{L}'',s/\sqrt{2}}$. We can then repeat this process to keep lowering the parameter.

Here, we observe that we can move directly from $D^{(2)}_{\mathcal{L}',\mathcal{L},s/\sqrt{2}}$ to $D^{(2)}_{\mathcal{L}',2\mathcal{L}'',s/\sqrt{2}}$, without bothering to "take the square root" first. In particular, if we take $2\mathcal{L}'' \subseteq \mathcal{L}$, then $\mathcal{L}'/(2\mathcal{L}'')$ is a *refinement* of $\mathcal{L}'/\mathcal{L}$. I.e., each coset $\boldsymbol{c} \in \mathcal{L}'/(2\mathcal{L}'')$ is contained inside a coset in $\mathcal{L}'/\mathcal{L}$, $2\mathcal{L}'' + \boldsymbol{c} \subseteq \mathcal{L} + \boldsymbol{c}$. The distribution $D^{(2)}_{\mathcal{L}',\mathcal{L},s/\sqrt{2}}$ assigns to an element $\boldsymbol{y} \in 2\mathcal{L}'' + \boldsymbol{c}$ weight proportional to $\rho_{s/\sqrt{2}}(\boldsymbol{y}) \cdot \rho_{s/\sqrt{2}}(\mathcal{L}' + \boldsymbol{c})$, whereas we want weight proportional to $\rho_{s/\sqrt{2}}(\boldsymbol{y})\rho_{s/\sqrt{2}}(2\mathcal{L}'' + \boldsymbol{c})$. We therefore focus on flipping coins with probability proportional to $\rho_{s/\sqrt{2}}(2\mathcal{L}'' + \boldsymbol{c})/\rho_{s/\sqrt{2}}(\mathcal{L} + \boldsymbol{c})$. In the next section, we make this abstract and show how to do this efficiently by using Theorem 4.3.2.

4.5.1 Running the square sampler "inside buckets"

Consider a multinomial distribution over $\{1, \ldots, N_1\} \times \{1, \ldots, N_2\}$ that assigns to element (i, j) probability $p_{i,j}$. For such a distribution, we can think of the first index i as representing

a "big bucket" containing all elements (i, j). We write $p_i := \sum_j p_{i,j}$ for the probability of sampling an element from the big bucket i.

In this language, the goal that we described in the previous section corresponds to converting the distribution that assigns to (i, j) probability $p_i p_{i,j} / \sum_{i'} p_{i'}^2$ into the distribution that assigns it probability $p_{i,j}^2 / \sum_{i',j'} p_{i',j'}^2$, which is what the following theorem accomplishes. Equivalently, we need to show how to convert a distribution given by $q_{i,j}$ into one given by $q_{i,j}^2/q_i \cdot \sum_{i',j'} q_{i',j'}^2/q_{i'}$. Notice that "inside each bucket," this corresponds to simply running the square sampler from Corollary 4.3.4. The following theorem shows how to do this, by using Theorem 4.3.2.

Theorem 4.5.1. *There is an algorithm that takes as input $\kappa \geq 100$ (the confidence parameter) and M elements from $\{1, \ldots, N_1\} \times \{1, \ldots, N_2\}$ and outputs a sequence of elements from the same set such that*

1. *the running time is $M \cdot \mathrm{poly}(\log \kappa, \log N_1, \log N_2)$;*
2. *each $(i, j) \in \{1, \ldots, N_1\} \times \{1, \ldots, N_2\}$ appears at least twice as often in the input as in the output; and*
3. *if the input consists of $M \geq 10\kappa^3 / \min_i \max_j p_{i,j}$ independent samples from the distribution that assigns probability $p_{i,j}$ to element (i, j) with $p_1 \geq p_i$ and $p_{1,1}/p_1 \geq p_{i,j}/p_i$ for all i, j, then the output is within statistical distance $C_1 N_1 N_2 \log(N_2) \exp(-C_2\kappa)$ of m independent samples with respective probabilities proportional to $p_{i,j}^2/p_i$, where*

$$m \geq \frac{M}{C\kappa^2 p_{\max}} \cdot \sum_{i,j} \frac{p_{i,j}^2}{p_i}$$

is a random variable with $p_{\max} := \max_{i,j} p_{i,j}/p_i$.

Proof. Notice that the distribution given by conditioning on bucket i assigns probability $p_{i,j}/p_i$ to each element (i, j).

The algorithm uses its first $\lceil M/5 \rceil$ samples to estimate both $p_{\max}$ and p_i. In particular, it takes these samples, $(i_1, j_1), \ldots, (i_{\lceil M/5 \rceil}, j_{\lceil M/5 \rceil})$, and groups them according to their first coordinate i_k. For each $i = 1, \ldots, N_1$, it then runs Proposition 4.3.1 on the samples (i_k, j_k) with $i_k = i$ in order to obtain an estimate $\widetilde{p}_{\max,i}$ with $\max_j p_{i,j}/p_i \leq \widetilde{p}_{\max,i} \leq \max_j 4p_{i,j}/p_i$. Finally, it sets $\widetilde{p}_{\max} := \max_i \widetilde{p}_{\max,i}$ and sets $\widetilde{p}_i$ to the fraction of these samples with $i_k = i$. (I.e., $\widetilde{p}_i := |\{1 \leq k \leq \lceil M/5 \rceil \ : \ i_k = i\}|/\lceil M/5 \rceil$.)

The algorithm then uses its next $\lceil 2M/3 \rceil$ samples as follows. It again groups the elements according to their first coordinate i_k. For $i = 1, \ldots, N_1$, the algorithm uses the samples with $i_k = i$ to run the procedure from Theorem 4.3.2 with $T := 1/\widetilde{p}_{\max}$ a total of $\lceil \widetilde{p}_{\max}\widetilde{p}_i M/(20\kappa) \rceil$ times to obtain coins $b_{i,j,k} \approx \mathrm{Bern}(p_{i,j}/(\kappa \widetilde{p}_{\max} p_i))$ for $i = 1, \ldots, N_1$, $j = 1, \ldots, N_2$, and $k = 1, \ldots, \lceil \widetilde{p}_{\max}\widetilde{p}_i M/(20\kappa) \rceil$. (If it ever runs out of samples, it simply fails.)

Finally, the algorithm goes through its next $\lceil M/(50\kappa) \rceil$ samples. When it encounters a sample (i, j), it outputs it if and only if $b_{i,j,k} = 1$, where k is chosen to be minimal so that $b_{i,j,k}$ has not been used previously. (If no such coin exists, the algorithm fails.)

Notice that each call to the procedure from Proposition 4.3.1 and each call to the procedure from Theorem 4.3.2 runs in time $M' \cdot \mathrm{poly}(\log \kappa, \log N_1, \log N_2)$, where M' is the number of input samples for this call. Since the total number of input samples used by these subprocedures is $O(M)$, it follows that the total running time is as claimed. Furthermore, it is clear that the input always contains at least twice as many elements of the form (i, j) as the output.

We now turn to proving Item 3. We first show that the estimates $\widetilde{p}_{\max}$ and $\widetilde{p}_i$ obtained from the first $\lceil M/5 \rceil$ samples are accurate. For each i, Proposition 4.3.1 requires $\kappa/p_{\max,i}$ samples in the ith bucket to successfully approximate $p_{\max,i}$, where $p_{\max,i} := \max_j p_{i,j}/p_i$. We have at least $M/5 \geq \max_i 2\kappa^2/(p_i p_{\max,i})$ samples total, so by the Chernoff-Hoeffding bound (Lemma 1.4.3), we expect to see at least $\kappa/p_{\max,i}$ samples in the ith bucket except with probability at most $\exp(-C\kappa)$. By the union bound, this holds for all buckets simultaneously

except with probability at most $N_1 \exp(-C\kappa)$. Similarly, by Proposition 4.3.1 and union bound, we have that $p_{\max,i} \le \widetilde{p}_{\max,i} \le 4p_{\max,i}$ for all i, except with probability at most $C_1 N_1 N_2 \log(N_2) \exp(-C_2\kappa)$. So, we may assume that $\max_{i,j} p_{i,j}/p_i \le \widetilde{p}_{\max} \le 4 \max_{i,j} p_{i,j}/p_i$. Finally, by the Chernoff-Hoeffding bound and union bound again, we may similarly assume that $p_i/2 \le \widetilde{p}_i \le 2p_i$, since this holds except with probability at most $N_1 \exp(-C\kappa)$.

We now show that the coins procedure from Theorem 4.3.2 generates the coins $b_{i,j,k}$ successfully Indeed, for this procedure to generate all coins for a fixed i, it requires $\kappa T \lceil \widetilde{p}_{\max} \widetilde{p}_i M/(20\kappa) \rceil \le p_i M/2$ samples for each i. By the Chernoff-Hoeffding bound again and union bound, we will have enough samples for each i except with probability at most $N_1 \exp(-C\kappa)$. Therefore, by applying Theorem 4.3.2 and union bound, we may therefore assume that the coins $b_{i,j,k}$ are distributed exactly as $\mathrm{Bern}(p_{i,j}/(\kappa \widetilde{p}_{\max} p_i))$, incurring statistical distance at most $C_1 N_1 N_2 M \exp(-C_2\kappa)$.

It follows that the output samples are distributed correctly. To see that the algorithm rarely runs out of coins, we simply observe that by the Chernoff-Hoeffding bound, the number of coins $b_{i,j,k}$ needed for any fixed pair (i,j) is at most $p_{i,j} M/(40\kappa) \le \widetilde{p}_{\max} \widetilde{p}_i M/(20\kappa)$ except with probability at most $\exp(-C\kappa)$. So, by union bound, the algorithm will have enough coins for all (i,j) except with probability at most $N_1 N_2 \exp(-C\kappa)$.

Finally, notice that the algorithm outputs each element in the final step with probability equal to

$$\sum_{i,j} p_{i,j} \cdot \frac{p_{i,j}}{\kappa \widetilde{p}_{\max} p_i} \ge \sum_{i,j} \frac{p_{i,j}^2}{4\kappa p_{\max} p_i} \,.$$

It follows from one final application of the Chernoff-Hoeffding bound that the number of output samples is at least

$$\frac{M}{60\kappa} \cdot \sum_{i,j} \frac{p_{i,j}^2}{4\kappa p_{\max} p_i} = \frac{M}{240\kappa^2 p_{\max}} \cdot \sum_{i,j} \frac{p_{i,j}^2}{p_i} \,,$$

except with probability $\exp(-C\kappa)$, as needed. $\square$

4.5.2 A more efficient combiner that works above smoothing

The following lemma generalizes the first part of Lemma 4.3.5. In particular, we recover Lemma 4.3.5 when $\mathcal{L}' = \mathcal{L}$. (Again, this lemma follows rather directly from Eq. (3.2), but we instead give a more direct proof.)

Lemma 4.5.2 ([ADRS15, Lemma 5.6])**.** *Let $\mathcal{L} \subset \mathbb{R}^n$ be a lattice, and let $\mathcal{L}' \supseteq \mathcal{L}$ be a superlattice with $2\mathcal{L}' \subseteq \mathcal{L}$. Then for any $\boldsymbol{y} \in \mathcal{L}'$ and $s > 0$, we have*

$$\Pr_{\boldsymbol{X}_1, \boldsymbol{X}_2 \sim D_{\mathcal{L},s}}[\boldsymbol{X}_1 + \boldsymbol{X}_2 = 2\boldsymbol{y} \mid \boldsymbol{X}_1 \equiv \boldsymbol{X}_2 \bmod 2\mathcal{L}'] = \Pr_{\boldsymbol{X} \sim D^{(2)}_{\mathcal{L}', \mathcal{L}, s/\sqrt{2}}}[\boldsymbol{X} = \boldsymbol{y}] \, .$$

Furthermore,

$$\sum_{\boldsymbol{c} \in \mathcal{L}/(2\mathcal{L}')} \rho_s(2\mathcal{L}' + \boldsymbol{c})^2 = \sum_{\boldsymbol{d} \in \mathcal{L}'/\mathcal{L}} \rho_{s/\sqrt{2}}(\mathcal{L} + \boldsymbol{d})^2 \, .$$

Proof. To prove the first equation, it suffices to show that the probability on the left-hand side is proportional to $\rho_{s/\sqrt{2}}(\boldsymbol{y})\rho_{s/\sqrt{2}}(\mathcal{L} + \boldsymbol{y})$. Indeed,

$$\begin{aligned}
\Pr_{\boldsymbol{X}_1, \boldsymbol{X}_2 \sim D_{\mathcal{L},s}}[\boldsymbol{X}_1 + \boldsymbol{X}_2 = 2\boldsymbol{y}] &= \frac{1}{\rho_s(\mathcal{L})^2} \cdot \sum_{\boldsymbol{x} \in \mathcal{L}} \rho_s(\boldsymbol{x})\rho_s(2\boldsymbol{y} - \boldsymbol{x}) \\
&= \frac{\rho_{s/\sqrt{2}}(\boldsymbol{y})}{\rho_s(\mathcal{L})^2} \cdot \sum_{\boldsymbol{x} \in \mathcal{L}} \rho_{s/\sqrt{2}}(\boldsymbol{x} - \boldsymbol{y}) \\
&= \frac{\rho_{s/\sqrt{2}}(\boldsymbol{y})}{\rho_s(\mathcal{L})^2} \cdot \rho_{s/\sqrt{2}}(\mathcal{L} + \boldsymbol{y}) \, .
\end{aligned}$$

The second equation follows by summing the left-hand side and the right-hand side of the above over all $\boldsymbol{y} \in \mathcal{L}'$. I.e., we have

$$\Pr[\boldsymbol{X}_1 \equiv \boldsymbol{X}_2 \bmod 2\mathcal{L}'] = \sum_{\boldsymbol{c} \in \mathcal{L}/(2\mathcal{L}')} \frac{\rho_s(\mathcal{L} + \boldsymbol{c})^2}{\rho_s(\mathcal{L})^2} \, ,$$

but from the above, this must also equal

$$\sum_{\boldsymbol{y}\in\mathcal{L}'} \frac{\rho_{s/\sqrt{2}}(\boldsymbol{y})}{\rho_s(\mathcal{L})^2} \cdot \rho_{s/\sqrt{2}}(\mathcal{L}+\boldsymbol{y}) = \sum_{\boldsymbol{d}\in\mathcal{L}'/\mathcal{L}} \frac{\rho_{s/\sqrt{2}}(\mathcal{L}+\boldsymbol{d})^2}{\rho_s(\mathcal{L})^2} \ .$$

□

Proposition 4.5.3. *There is an algorithm that takes as input three lattices $\mathcal{L} \subseteq \mathcal{L}' \subseteq \mathcal{L}'' \subset \mathbb{R}^n$ with $2\mathcal{L}' \subseteq \mathcal{L}$ and $2\mathcal{L}'' \subseteq \mathcal{L}$, $\kappa \geq Cn$ (the confidence parameter), and a sequence of vectors from $\mathcal{L}'$ such that, if the input consists of*

$$M \geq 10\kappa^3 \cdot \frac{\sum_{\boldsymbol{c}\in\mathcal{L}/\mathcal{L}'} \rho_s(\mathcal{L}'+\boldsymbol{c})^2}{\min_{\boldsymbol{c}\in\mathcal{L}'/\mathcal{L}} \max_{\boldsymbol{d}\in\mathcal{L}/(2\mathcal{L}'')} \rho_s(\mathcal{L}+\boldsymbol{c})\rho_s(2\mathcal{L}''+\boldsymbol{c}+\boldsymbol{d})}$$

independent samples from $D^{(2)}_{\mathcal{L}',\mathcal{L},s}$ for some $s > 0$, then the output distribution is $M\exp(-C\kappa)$-close to m independent samples from $D^{(2)}_{\mathcal{L}'',\mathcal{L}',s/\sqrt{2}}$, where

$$m \geq \frac{M}{C\kappa^2} \cdot \frac{\rho_s(\mathcal{L})}{\rho_{s/2}(\mathcal{L}'')} \cdot \frac{\sum_{\boldsymbol{d}\in\mathcal{L}''/\mathcal{L}'} \rho_{s/\sqrt{2}}(\mathcal{L}'+\boldsymbol{d})^2}{\sum_{\boldsymbol{c}\in\mathcal{L}'/\mathcal{L}} \rho_s(\mathcal{L}+\boldsymbol{c})^2}$$

is a random variable. Furthermore, the algorithm runs in time $M \cdot \mathrm{poly}(n, \log\kappa)$.

Proof. Let $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M)$ be the input vectors, and for each i, let $\boldsymbol{c}_i \in \mathcal{L}'/\mathcal{L}$ be the coset of $\boldsymbol{X}_i$ over $\mathcal{L}$, and let $\boldsymbol{d}_i \in \mathcal{L}/(2\mathcal{L}'')$ be the (unique) coset such that $\boldsymbol{X}_i \in 2\mathcal{L}'' + \boldsymbol{c}_i + \boldsymbol{d}_i$. The algorithm first applies the the procedure from Theorem 4.5.1 in a manner similar to that of the algorithm from Proposition 4.3.6. Namely, the algorithm runs the procedure from Theorem 4.5.1 with input κ and pairs $(\boldsymbol{c}_1, \boldsymbol{d}_1), \ldots, (\boldsymbol{c}_M, \boldsymbol{d}_M)$, receiving output pairs $(\boldsymbol{c}'_1, \boldsymbol{d}'_1), \ldots, (\boldsymbol{c}'_m, \boldsymbol{d}'_m)$. For each $i = 1, \ldots, m$, it chooses a pair of unpaired vectors $\boldsymbol{X}_j, \boldsymbol{X}_k$ with $(\boldsymbol{c}_j, \boldsymbol{d}_j) = (\boldsymbol{c}_k, \boldsymbol{d}_k) = (\boldsymbol{c}'_i, \boldsymbol{d}'_i)$ and adds $\boldsymbol{Y}_i = (\boldsymbol{X}_j + \boldsymbol{X}_k)/2 \in \mathcal{L}'$ to its output.

The running time of the algorithm follows from Item 1 of Theorem 4.5.1. Furthermore, we note that by Item 2 of Theorem 4.5.1, the algorithm will always be able to find unused

j, k satisfying $(\boldsymbol{c}_j, \boldsymbol{d}_j) = (\boldsymbol{c}_k, \boldsymbol{d}_k) = (\boldsymbol{c}'_i, \boldsymbol{d}'_i)$.

Next, we observe that the output distribution is correct. In particular, let

$$p_{\boldsymbol{c},\boldsymbol{d}} := \Pr[\boldsymbol{c}_k = \boldsymbol{c} \text{ and } \boldsymbol{d}_k = \boldsymbol{d}] = \frac{\rho_s(\mathcal{L} + \boldsymbol{c})\rho_s(2\mathcal{L}'' + \boldsymbol{c} + \boldsymbol{d})}{\sum_{\boldsymbol{c}' \in \mathcal{L}/\mathcal{L}'} \rho_s(\mathcal{L}' + \boldsymbol{c}')^2} \ .$$

Then, we see that the input satisfies the criteria necessary to apply Item 3 of Theorem 4.5.1. And, by the theorem, the distribution of pairs $(\boldsymbol{X}_j, \boldsymbol{X}_k)$ chosen by the algorithm will be within statistical distance $M \exp(-C\kappa)$ of the distribution given by sampling $\boldsymbol{X}_j$ and $\boldsymbol{X}_k$ independently from the discrete Gaussian, conditioned on $\boldsymbol{X}_j \equiv \boldsymbol{X}_k \bmod 2\mathcal{L}''$. It follows from Lemma 4.5.2 that the output distribution is correct.

Finally, again by Item 3 of Theorem 4.5.1, we see that

$$m \geq \frac{M}{C\kappa^2 \max_{\boldsymbol{c},\boldsymbol{d}} p_{\boldsymbol{c},\boldsymbol{d}}/p_{\boldsymbol{c}}} \cdot \sum_{\boldsymbol{c},\boldsymbol{d}} \frac{p_{\boldsymbol{c},\boldsymbol{d}}^2}{p_{\boldsymbol{c}}} \ .$$

By Proposition 3.4.3, $\max_{\boldsymbol{c},\boldsymbol{d}} p_{\boldsymbol{c},\boldsymbol{d}}/p_{\boldsymbol{c}} = \rho_s(2\mathcal{L}'')/\rho_s(\mathcal{L})$. So, we have

$$\begin{aligned} m &\geq \frac{M\rho_s(\mathcal{L})}{C\kappa^2\rho_s(2\mathcal{L}'')} \cdot \sum_{\boldsymbol{c},\boldsymbol{d}} \frac{p_{\boldsymbol{c},\boldsymbol{d}}^2}{p_{\boldsymbol{c}}} \\ &= \frac{M\rho_s(\mathcal{L})}{C\kappa^2\rho_s(2\mathcal{L}'')} \cdot \frac{\sum_{\boldsymbol{d} \in \mathcal{L}'/(2\mathcal{L}'')} \rho_s(2\mathcal{L}'' + \boldsymbol{d})^2}{\sum_{\boldsymbol{c} \in \mathcal{L}'/\mathcal{L}} \rho_s(\mathcal{L}' + \boldsymbol{c})^2} \\ &= \frac{M\rho_s(\mathcal{L})}{C\kappa^2\rho_s(2\mathcal{L}'')} \cdot \frac{\sum_{\boldsymbol{d} \in \mathcal{L}''/\mathcal{L}'} \rho_{s/\sqrt{2}}(\mathcal{L}' + \boldsymbol{d})^2}{\sum_{\boldsymbol{c} \in \mathcal{L}'/\mathcal{L}} \rho_s(\mathcal{L} + \boldsymbol{c})^2} \ , \end{aligned}$$

as needed, where the last equality follows from Lemma 4.5.2. □

We presented a very general version of Proposition 4.5.3 above in the hopes that it might help improve our algorithm to work below the smoothing parameter. However, the lower bounds for M and m are rather unwieldy. So, below, we present a simplified version that assumes that the parameter $s > 0$ is above the smoothing parameter.

Corollary 4.5.4. *There is an algorithm that takes as input three lattices $\mathcal{L} \subseteq \mathcal{L}' \subseteq \mathcal{L}'' \subset \mathbb{R}^n$ with $2\mathcal{L}' \subseteq \mathcal{L}$ and $2\mathcal{L}'' \subseteq \mathcal{L}$, $\kappa \geq Cn$ (the confidence parameter), and a sequence of vectors from $\mathcal{L}'$ such that, if the input consists of*

$$M \geq C\kappa^3 \cdot |\mathcal{L}'/(2\mathcal{L}'')|$$

independent samples from $D^{(2)}_{\mathcal{L}',\mathcal{L},s}$ for some $s \geq \max\{\sqrt{2}\eta_{1/2}(\mathcal{L}'), 2\eta_{1/2}(\mathcal{L}'')\}$, then the output distribution is $M\exp(-C\kappa)$-close to m independent samples from $D^{(2)}_{\mathcal{L}'',\mathcal{L}',s/\sqrt{2}}$, where

$$m \geq \frac{M}{C\kappa^2}$$

is a random variable. Furthermore, the algorithm runs in time $M \cdot \mathrm{poly}(n, \log \kappa)$.

Proof. By Eq. (1.4), we see that

$$\sum_{\boldsymbol{c} \in \mathcal{L}'/\mathcal{L}} \rho_s(\mathcal{L} + \boldsymbol{c})^2 \geq \frac{1}{3} \cdot \rho_s(\mathcal{L}) \sum_{\boldsymbol{c} \in \mathcal{L}'/\mathcal{L}} \rho_s(\mathcal{L} + \boldsymbol{c}) = \frac{1}{3} \cdot \rho_s(\mathcal{L})\rho_s(\mathcal{L}') \ .$$

(Here, we have used the fact that $2\mathcal{L}'' \subseteq \mathcal{L}$, so that $\eta_{1/2}(\mathcal{L}) \leq 2\eta_{1/2}(2\mathcal{L}'')$.) We similarly have that

$$\min_{\boldsymbol{c} \in \mathcal{L}'/\mathcal{L}} \max_{\boldsymbol{d} \in \mathcal{L}/(2\mathcal{L}'')} \rho_s(\mathcal{L} + \boldsymbol{c})\rho_s(2\mathcal{L}'' + \boldsymbol{c} + \boldsymbol{d}) \geq \frac{1}{9} \cdot \rho_s(\mathcal{L})\rho_{s/2}(\mathcal{L}'') \ .$$

Therefore, it suffices to take

$$M \geq C\kappa^3 \cdot \frac{\rho_s(\mathcal{L}')}{\rho_s(2\mathcal{L}'')} \ .$$

Finally, we observe that

$$\rho_s(\mathcal{L}') = \sum_{\boldsymbol{c} \in \mathcal{L}'/(2\mathcal{L}'')} \rho_s(2\mathcal{L}'' + \boldsymbol{c}) \leq |\mathcal{L}'/(2\mathcal{L}'')|\rho_s(2\mathcal{L}'')$$

to obtain the final bound on M. Similar analysis shows that the lower bound on m holds as

well. □

We are going to apply Corollary 4.5.4 repeatedly, to a "tower" of lattices $(\mathcal{L}_0, \ldots, \mathcal{L}_\ell)$, as defined next.

Definition 4.5.5. *For an integer a satisfying $0 \le a \le n/2$, we say that the sequence of lattices $(\mathcal{L}_0, \ldots, \mathcal{L}_\ell)$ is a* tower of lattices in $\mathbb{R}^n$ of index 2^a *if for all i we have $2\mathcal{L}_{i+1} \subseteq \mathcal{L}_i \subseteq \mathcal{L}_{i+1}$, $2\mathcal{L}_{i+2} \subseteq \mathcal{L}_i$, and the index of $\mathcal{L}_i$ over $\mathcal{L}_{i+1}$ is 2^a.*

We next observe that it is easy to construct a tower with any desired final lattice $\mathcal{L}_\ell$. In fact, one can even choose $\mathcal{L}_{\ell-1}$, the second-to-last lattice in the tower.

Claim 4.5.6 ([ADRS15, Claim 5.9])**.** *There is a polynomial-time algorithm that takes as input integers $\ell \ge 1$ and $0 \le a \le n/2$, as well as two lattices $\mathcal{L}$ and $\mathcal{L}'$ satisfying $2\mathcal{L} \subseteq \mathcal{L}' \subseteq \mathcal{L} \subset \mathbb{R}^n$ with the index of $\mathcal{L}'$ in $\mathcal{L}$ being 2^a, outputs a tower of lattices $(\mathcal{L}_0, \ldots, \mathcal{L}_\ell)$ of index 2^a with $\mathcal{L}_\ell = \mathcal{L}$, $\mathcal{L}_{\ell-1} = \mathcal{L}'$, and $\mathcal{L}_0 \supseteq 2^{\lceil \ell a/n \rceil}\mathcal{L}$.*

Proof. Let $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ be a basis of $\mathcal{L}$ chosen so that $2\boldsymbol{b}_1, \ldots, 2\boldsymbol{b}_a, \boldsymbol{b}_{a+1}, \ldots, \boldsymbol{b}_n$ is a basis of $\mathcal{L}'$. It is not difficult to see that such a basis exists. Then define the tower by "cyclically doubling a coordinates," namely,

$$\begin{aligned}
\mathcal{L}_\ell &= \mathcal{L}(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n), \\
\mathcal{L}_{\ell-1} &= \mathcal{L}(2\boldsymbol{b}_1, \ldots, 2\boldsymbol{b}_a, \boldsymbol{b}_{a+1}, \ldots, \boldsymbol{b}_n), \\
\mathcal{L}_{\ell-2} &= \mathcal{L}(2\boldsymbol{b}_1, \ldots, 2\boldsymbol{b}_{2a}, \boldsymbol{b}_{2a+1}, \ldots, \boldsymbol{b}_n), \\
\mathcal{L}_{\ell-3} &= \mathcal{L}(4\boldsymbol{b}_1, \ldots, 4\boldsymbol{b}_{3a-n}, 2\boldsymbol{b}_{3a-n+1}, \ldots, 2\boldsymbol{b}_n)\ ,
\end{aligned}$$

etc. It is easy to check that this satisfies all the required properties. □

Corollary 4.5.7. *There is an algorithm that takes as input a tower of lattices $(\mathcal{L}_0, \ldots, \mathcal{L}_\ell)$ in $\mathbb{R}^n$ of index $1 \le 2^a \le 2^{n/2}$, $\kappa \ge Cn$ (the confidence parameter), and $M = (C\kappa^4)^\ell \cdot 2^{n-a}$ vectors*

in $\mathcal{L}_1$ such that the algorithm runs in time $M \cdot \mathrm{poly}(n, \log \kappa, \ell)$ and, if the input vectors are distributed as independent samples from $D^{(2)}_{\mathcal{L}_1, \mathcal{L}_0, s}$ for some $s \geq \max\{2^{(\ell-1)/2}\eta_{1/2}(\mathcal{L}_{\ell-1}), 2^{\ell/2}\eta_{1/2}(\mathcal{L}_\ell)\}$, then the output will be $\ell M \exp(-C\kappa)$-close to 2^{n-a} independent samples from $D^{(2)}_{\mathcal{L}_\ell, \mathcal{L}_{\ell-1}, 2^{-(\ell-1)/2}s}$.

Proof. Let $\mathcal{X}_1 = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M)$ be the sequence of input vectors. For $i = 1, \ldots, \ell - 1$, the algorithm calls the procedure from Corollary 4.5.4 with input $\mathcal{L}_{i-1}$, $\mathcal{L}_i$, $\mathcal{L}_{i+1}$, κ, and $\mathcal{X}_i$, receiving output $\mathcal{X}_{i+1}$. Finally, the algorithm outputs the first 2^{n-a} vectors in $\mathcal{X}_\ell$.

The running time and correctness follow immediately from Corollary 4.5.4. □

4.5.3 Sampling above smoothing in time $2^{n/2}$

From Corollary 4.5.7, it follows more-or-less immediately that we can obtain $2^{n/2}$ independent samples from $D^{(2)}_{\mathcal{L}, \mathcal{L}', s}$ for any lattices $\mathcal{L}' \subseteq \mathcal{L} \subset \mathbb{R}^n$ with $2\mathcal{L} \subseteq \mathcal{L}'$ and, say, $|\mathcal{L}/\mathcal{L}'| = 2^a < 2^{n/2 - n/\log^2 n}$, and a parameter $s \geq \max\{\sqrt{2}\eta_{1/2}(\mathcal{L}'), 2\eta_{1/2}(\mathcal{L})\}$, and samples from $D^{(2)}_{\mathcal{L}, \mathcal{L}', s}$ in time $2^{n-a+o(n)}$. In particular, we can set up a tower of lattices as in Claim 4.5.6, start with samples from $D^{(2)}_{\mathcal{L}_1, \mathcal{L}_0, 2^{(\ell-1)/2}s}$, and then run Corollary 4.5.7.

The only question is how to obtain the samples from $D^{(2)}_{\mathcal{L}_1, \mathcal{L}_0, 2^{(\ell-1)/2}s}$. Note that, if $\ell \gg \log^2 n$ is sufficiently large, then since $\mathcal{L}_0 \supseteq 2^{\lceil \ell a/n \rceil}\mathcal{L} \supseteq 2^{\lceil \ell/2 - \ell/\log^2 n \rceil}\mathcal{L}$, it follows that $2^{(\ell-1)/2}s \gg \eta_{1/2}(\mathcal{L}_1)$. We will therefore be able to use Corollary 1.3.16 to obtain samples from $D_{\mathcal{L}_1, 2^{(\ell-1)/2}s}$. To convert these into samples from $D^{(2)}_{\mathcal{L}_1, \mathcal{L}_0, 2^{(\ell-1)/2}s}$, we can use the square sampler (Corollary 4.3.4).

Theorem 4.5.8. *There is an algorithm that takes as input two lattices $\mathcal{L}' \subseteq \mathcal{L} \subset \mathbb{R}^n$ with $2\mathcal{L} \subset \mathcal{L}'$ and $|\mathcal{L}/\mathcal{L}'| = 2^a < 2^{n/2}$, and a parameter $s \geq \max\{\sqrt{2}\eta_{1/2}(\mathcal{L}'), 2\eta_{1/2}(\mathcal{L})\}$ and outputs 2^{n-a} samples that are 2^{-n^2}-close to independent samples from $D^{(2)}_{\mathcal{L}, \mathcal{L}', s}$ in time $2^{n-a+O(\ell \log n)+o(n)}$, where $\ell := Cn \log n/(n - 2a)$.*

Proof. The algorithm first runs the procedure from Claim 4.5.6 to obtain a tower of lattices $(\mathcal{L}_0, \ldots, \mathcal{L}_\ell)$ of index 2^a such that $\mathcal{L}_\ell = \mathcal{L}$ and $\mathcal{L}_0 \supseteq 2^{\lceil \ell a/n \rceil}\mathcal{L}$.

Next, the algorithm runs the procedure from Corollary 1.3.16 with input $\mathcal{L}_1$, $\hat{s} := 2^{(\ell-1)/2}s$, $u = cn$ (where the constant $c > 0$ is chosen so that the running time of the procedure is at most $\mathrm{poly}(n)M + 2^{n-a+o(n)}$), and $M := (C_1 n)^{C_2 \ell} 2^{n-a}$, receiving as output M samples from $D_{\mathcal{L}_1,\hat{s}}$ The algorithm then runs the square sampler (Corollary 4.3.4) as in Proposition 4.3.6 with $\kappa := Cn^2$ to convert these into $M/\mathrm{poly}(n)$ samples that are within statistical distance at most $\exp(-Cn^2)$ of independent samples from $D^{(2)}_{\mathcal{L}_1,\mathcal{L}_0,\hat{s}}$. Finally, the algorithm runs the procedure from Corollary 4.5.7 on these samples to obtain 2^{n-a} samples that are 2^{-n^2}-close to independent samples from $D^{(2)}_{\mathcal{L},\mathcal{L}',s}$, as needed.

The running time is clear. By Claim 4.5.6, we have that $\mathcal{L}_1 \supseteq \mathcal{L}_0 \supseteq 2^{\lceil \ell a/n \rceil}\mathcal{L}$. It follows that

$$\hat{s} = 2^{(\ell-1)/2}s \geq 2^{(\ell+1)/2}\eta_{1/2}(\mathcal{L}) \geq 2^{\ell(1/2-a/n)-1/2}\eta_{1/2}(\mathcal{L}_1) = \mathrm{poly}(n)\eta_{1/2}(\mathcal{L}_1) \,.$$

Therefore, the procedure from Corollary 1.3.16 will succeed. The result then follows immediately from Corollary 4.3.4 and Corollary 4.5.7. □

In [ADRS15], we show a "square-root sampler," which can be used to convert the sampled from $D^{(2)}_{\mathcal{L},\mathcal{L}',s}$ from Theorem 4.5.8 into pure Gaussian samples from $D_{\mathcal{L},s}$, as claimed in Theorem 7. In this thesis, we prefer to leave this rather tedious proof out, and we simply observe that the distributions $D^{(2)}_{\mathcal{L},\mathcal{L}',s}$ and $D_{\mathcal{L},s}$ are very similar when $s \geq \eta_\varepsilon(\mathcal{L}')$.

4.5.4 Sampling from shifted Gaussians above smoothing

Here, we observe that a sampler for sampling m independent samples from centered discrete Gaussian distributions with $s \geq \sqrt{2}\eta_{1/2}(\mathcal{L})$ can actually be used to obtain m samples from the *shifted* discrete Gaussian $D_{\mathcal{L}-\boldsymbol{t},s}$ in $2^{n/2+o(n)}$ time for any parameter $s > \sqrt{2} \cdot \eta_\varepsilon(\mathcal{L})$ with $\varepsilon \approx 1/2$. We present a brief proof sketch here in case this finds applications in future work. The idea is to call our centered discrete Gaussian sampler repeatedly on the lattice $\overline{\mathcal{L}} := \mathcal{L}(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n, \bar{\boldsymbol{t}}) \subset \mathbb{R}^{n+1}$, where $\bar{\boldsymbol{t}} := (-\boldsymbol{t}, s) \in \mathbb{R}^{n+1}$. Note that the lattice hyperplane

$\mathcal{L} + \bar{\boldsymbol{t}} \subset \overline{\mathcal{L}}$ is simply a copy of $\mathcal{L} - \boldsymbol{t}$ shifted by $s\boldsymbol{e}_{n+1}$, so that $\pi_{\mathbb{R}^n}(D_{\mathcal{L}+\bar{\boldsymbol{t}}}) = D_{\mathcal{L}-\boldsymbol{t},s}$. We therefore simply return the first n coordinates of the first m vectors in $\mathcal{L} + \bar{\boldsymbol{t}}$.

To prove that this algorithm works, we simply need to show that (1) $\eta_\varepsilon(\mathcal{L}) > \eta_{1/2}(\overline{\mathcal{L}})$, so that the call to the centered sampler will be valid as long as $s > \sqrt{2} \cdot \eta_\varepsilon(\mathcal{L})$; and (2) when s is above smoothing, a vector sampled from $D_{\overline{\mathcal{L}},s}$ will land in $\mathcal{L} + \bar{\boldsymbol{t}}$ with relatively high probability, so that we will not have to make too many calls to the centered sampler in order to find m vectors in $\mathcal{L} + \bar{\boldsymbol{t}}$. Both claims follow from standard calculations. (As described above, the algorithm achieves $\varepsilon \approx 0.38$ and makes a constant number of calls to the centered DGS oracle. If we instead set $\bar{\boldsymbol{t}} := (-\boldsymbol{t}, s/\kappa)$ for $\kappa \geq 1$ and make $O(\kappa)$ oracle calls, we can obtain $\varepsilon \approx 1/2 - \exp(-C\kappa^2)$.)

Chapter 5

A Reduction from DGS to CVP (and SVP)[1]

5.1 Introduction

Given the importance of discrete Gaussian sampling (DGS) in the study of classical computational lattice problems, it is natural to ask about the complexity of DGS itself. We have seen the importance of discrete Gaussian sampling (DGS) in the study of computational lattice problems. In particular, in Chapter 4, we showed that the fastest known algorithms for SVP and CVP (originally from [ADRS15] and [ADS15]) both relied heavily on DGS as a subroutine. But, intuitively, if DGS were a much harder problem than SVP and CVP, then these techniques would seem wasteful. It is therefore natural to ask about the complexity of DGS itself.

Prior to this work, DGS was one of the only prominent lattice problems not known to

[1]This chapter is primarily based on work that appeared in the Symposium on Discrete Algorithms (SODA), 2016 [Ste16a], and passages have been taken verbatim from this source. This work was done while at the Simons Institute 2015 cryptography summer program and was partially supported by the National Science Foundation under Grant No. CCF-1320188.

be reducible to CVP via a dimension-preserving reduction. (We are particularly interested in dimension-preserving reductions because they imply *quantitative* relationships between computational problems.) In fact, previously, there was simply no known algorithm that sampled from $D_{\mathcal{L}-\boldsymbol{t},s}$ for an arbitrary shift $\boldsymbol{t}$ and parameter $s > 0$, and it was not even known whether sampling from the *centered* distribution $D_{\mathcal{L},s}$ could be efficiently reduced to a problem in NP. (Since DGS is a sampling problem, it technically cannot be placed directly in classes of decision problems or search problems like NP or FNP. But, we can still hope to reduce it to such problems. See, e.g., [Aar14] for a discussion of the complexity of sampling problems and their relationship to search problems.)

5.1.1 Our results

Our first main result is a dimension-preserving reduction from DGS to CVP. (See Theorem 5.3.6.) This immediately implies two important corollaries. Together with the relatively straightforward reduction from CVP to DGS (see Corollary 1.3.11), this shows that CVP and DGS are equivalent via efficient dimension-preserving reductions. In particular, this suggests that the approach of [ADS15] presented in Chapter 4 is in some (weak) sense the "correct" way to attack CVP, since we now know that any faster algorithm for CVP necessarily implies a similarly efficient discrete Gaussian sampler (and vice versa). Furthermore, together with the result of [ADS15], this gives a $2^{n+o(n)}$-time algorithm for discrete Gaussian sampling that works for any parameter s and shift $\boldsymbol{t}$, the first known algorithm for this problem.

Our second main result is a dimension-preserving reduction from *centered* DGS to SVP. (See Theorem 5.4.6.) As we describe below, this result requires quite a bit more work, and we consider it to be more surprising, since, in a fixed dimension, an SVP oracle seems to be significantly weaker than a CVP oracle. In contrast to the CVP case, we know of no efficient reduction from SVP to centered DGS, and we do not even know whether centered DGS is

NP-hard. (While [ADRS15] use centered DGS to solve SVP, they require exponentially many samples to do so.) We present only a much weaker reduction from γ-approximate SVP to centered DGS for any $\gamma = \Omega(\sqrt{n/\log n})$. We also show that, for any $\gamma = o(\sqrt{n/\log n})$, no "simple" reduction from γ-SVP to centered DGS will work. (See Section 5.5.)

In [Ste16a], we also note that our proofs do not make use of any unique properties of the discrete Gaussian or of the ℓ_2 norm. We therefore show a much more general result: any distribution that is close to a weighted combination of uniform distributions over balls in some norm reduces to CVP in this norm. n particular, sampling from the natural ℓ_q analogue of the discrete Gaussian is equivalent to CVP in the ℓ_q norm, under efficient dimension-preserving reductions. These results are outside of the scope of this thesis, so we refer the interested reader to [Ste16a].

5.1.2 Proof overview

We now provide a high-level description of our techniques.

Reduction from DGS to CVP. Our basic idea is to sample from the discrete Gaussian $D_{\mathcal{L}-\boldsymbol{t},s}$ in two natural steps. We first sample some radius r from a carefully chosen distribution. We then sample a uniformly random point in $(\mathcal{L}-\boldsymbol{t}) \cap rB_2^n$. In particular, the distribution on the radius should assign probability to each radius r that is roughly proportional to $\exp(-\pi r^2/s^2) \cdot |(\mathcal{L}-\boldsymbol{t}) \cap rB_2^n|$. (See the proof of Theorem 5.3.6 for the exact distribution.) So, in order to reduce DGS to CVP, it suffices to show how to use our CVP oracle (1) compute $|(\mathcal{L}-\boldsymbol{t}) \cap rB_2^n|$ for arbitrary r, and (2) sample a uniformly random point from $(\mathcal{L}-\boldsymbol{t}) \cap rB_2^n$.

We actually use the same technical tool to solve both problems: lattice sparsification, as introduced by Khot [Kho05] (though our analysis is more similar to that of Dadush and Kun [DK13] and [DRS14]). Intuitively, sparsification allows us to sample a random sublattice $\mathcal{L}' \subset \mathcal{L}$ of index p such that for any vector $\boldsymbol{x} \in \mathcal{L}$, we have $\Pr[\boldsymbol{x} \in \mathcal{L}'] \approx 1/p$. (Of course, if

we did exactly this, then $\mathcal{L}'$ would not be a lattice.) Suppose we could find a sublattice $\mathcal{L}'$ such that for the closest $N \approx p$ points to $\boldsymbol{t}$ in $\mathcal{L}$, we have $\Pr[\boldsymbol{x} \in \mathcal{L}'] = 1/p$, independently of the other close points. Then, this would suffice for our two use cases. In particular, if the lattice has N points in the ball of a given radius around $\boldsymbol{t}$, then $\mathcal{L}' - \boldsymbol{t}$ would have a point in this ball with probability very close to N/p. We can use a CVP oracle to approximate this probability empirically, and we therefore obtain a good approximation for the number of lattice points in any ball. (We achieve an approximation factor of $1 + 1/f(n)$ for any $f(n) = \text{poly}(n)$. See Theorem 5.3.5.) Similarly, if we know that the number of lattice points in a ball of radius r around $\boldsymbol{t}$ is roughly N, then we can take $p = \text{poly}(n) \cdot N$ and repeatedly sample $\mathcal{L}'$ until $\mathcal{L}'$ has a point inside the ball of radius r around $\boldsymbol{t}$. The resulting point will be a nearly uniformly random sample from the lattice points in the ball of radius r around $\boldsymbol{t}$. Combining these two operations allows us to sample from the discrete Gaussian using a CVP oracle, as described above. (See Theorem 5.3.6.)

Unfortunately, sparsification does not give us exactly this distribution. More specifically, sparsification works as follows. Given a prime p and lattice basis $\mathbf{B}$, we sample $\boldsymbol{z} \in \mathbb{Z}_p^n$ uniformly at random and define the corresponding sparsified sublattice as

$$\mathcal{L}' := \{\boldsymbol{x} \in \mathcal{L} \ : \ \langle \boldsymbol{z}, \mathbf{B}^{-1}\boldsymbol{x} \rangle \equiv 0 \bmod p\} \ . \tag{5.1}$$

Then, for any vector $\boldsymbol{x} \in \mathcal{L}$, we have $\Pr[\boldsymbol{x} \in \mathcal{L}'] = 1/p$ unless $\boldsymbol{x} \in p\mathcal{L}$ (in which case $\boldsymbol{x}$ is always in $\mathcal{L}'$). Even if we ignore the issue that points in $p\mathcal{L}$ do not behave properly, it is easy to see that these probabilities are not at all independent. For example, if $\boldsymbol{x} = \alpha\boldsymbol{y}$, then $\boldsymbol{x} \in \mathcal{L}'$ if and only if $\boldsymbol{y} \in \mathcal{L}'$. And of course, more complex dependencies can exist as well. Fortunately, we can get around this by using an idea from [DRS14] (and implicit in [DK13]). In particular, we can show that the probabilities are close to independent if we also shift the sublattice $\mathcal{L}'$ by a "random lattice vector" $\boldsymbol{w} \in \mathcal{L}$. I.e., while the distribution of the points in

$\mathcal{L}' \cap (rB_2^n + \boldsymbol{t})$ might be very complicated, each point in $\mathcal{L} \cap (rB_2^n + \boldsymbol{t})$ will land in $\mathcal{L}' - \boldsymbol{w}$ with probability $\approx 1/p$, and their distributions are nearly independent. (See Theorem 5.3.1 for the precise statement.) Our CVP oracle makes no distinction between lattices and shifted lattices (we can just shift $\boldsymbol{t}$ by $\boldsymbol{w}$), so this solution suffices for our purposes.

Reduction from centered DGS to SVP. Our reduction from centered DGS to SVP uses the same high-level ideas described above, but the details are a bit more complicated. As in the CVP case, our primary tool is lattice sparsification, in which we choose a sparsified sublattice as in Eq. (5.1). As before, we wish to control the distribution of the shortest vector in $\mathcal{L}'$, and we note that, ignoring degenerate cases, $\boldsymbol{x}$ is a shortest vector of $\mathcal{L}'$ if and only if $\boldsymbol{x} \in \mathcal{L}'$ and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N \notin \mathcal{L}'$ where the $\boldsymbol{y}_i \in \mathcal{L}$ are the non-zero lattice vectors shorter than $\boldsymbol{x}$ (up to sign). However, as in the CVP case, this probability can be affected by linear dependencies. In the CVP case, we solved this problem by considering a random shift of $\mathcal{L}'$. But, this solution clearly does not work here because an SVP oracle simply "cannot handle" shifted lattices. We therefore have to deal explicitly with these dependencies.

The most obvious type of dependency occurs when $\boldsymbol{x}$ is not *primitive*, so that $\boldsymbol{x} = \alpha \boldsymbol{y}_i$ for $|\alpha| > 1$. In this case, there is nothing that we can do—$\boldsymbol{y}_i$ is shorter than $\boldsymbol{x}$ and $\boldsymbol{y}_i \in \mathcal{L}'$ if and only if $\boldsymbol{x} \in \mathcal{L}'$, so $\boldsymbol{x}$ will *never* be a shortest non-zero vector in $\mathcal{L}'$. We therefore are forced to work with only primitive vectors (i.e., lattice vectors that are not a scalar multiple of a shorter lattice vector). Even if we only consider primitive vectors, it can still be the case that two such vectors are scalar multiples of each other mod p, $\boldsymbol{x} \equiv \alpha \boldsymbol{y}_i \bmod p\mathcal{L}$. Luckily, we show that this can only happen if there are $\widetilde{\Omega}(p)$ primitive vectors shorter than $\boldsymbol{x}$ in the lattice, so that this issue does not affect the $\widetilde{\Omega}(p)$ shortest primitive vectors. (See Lemma 5.2.10.) We also show that higher-order dependencies (e.g., equations of the form $\boldsymbol{x} \equiv \alpha \boldsymbol{y}_i + \beta \boldsymbol{y}_j \bmod p\mathcal{L}$) have little effect. (See Lemma 5.2.8.) So, the shortest non-zero vector in the sparsified lattice will be distributed nearly uniformly over the $\widetilde{\Omega}(p)$ shortest primitive vectors in the original

lattice. (See Theorem 5.4.1 and Proposition 5.4.2 for the precise statement, which might be useful in future work.)

As in the CVP case, this suffices for our purposes. In particular, if there are N *primitive* lattice vectors in the ball of radius r centered at the origin for $N \leq \tilde{O}(p)$, then there will be a non-zero vector in $\mathcal{L}' \cap rB_2^n$ with probability very close to N/p. With an SVP oracle, we can estimate this probability, and this allows us to approximate the number of primitive lattice vectors in a ball with very good accuracy. (See Theorem 5.4.5.) And, the sparsification algorithm and SVP oracle also allow us to sample a primitive lattice vector in the ball of radius r around the origin with nearly uniform probability, as in the CVP case. (See Lemma 5.4.3.)

Then, the same approach as before would allow us to use an SVP oracle to sample from the discrete Gaussian over the *primitive* lattice vectors. In order to obtain the true discrete Gaussian, we first "add $\mathbf{0}$ in" by estimating the total Gaussian mass $\rho_s(\mathcal{L})$ and returning $\mathbf{0}$ with probability $1/\rho_s(\mathcal{L})$. Second, after sampling a primitive vector $\boldsymbol{x}$ using roughly the above idea, we sample an integer coefficient $z \in \mathbb{Z} \setminus \{0\}$ according to a one-dimensional discrete Gaussian (using an algorithm introduced by [BLP+13]) and output $z\boldsymbol{x}$. If we choose the primitive vector appropriately, we show that the resulting distribution is $D_{\mathcal{L},s}$.[2]

5.2 Preliminaries

The following (loose version of a) lemma due to [BHW93] can be thought of as a simpler (and asymptotically much weaker) form of Lemma 4.4.1 that is slightly easier to work with

[2]Interestingly, the problem of sampling from the centered discrete Gaussian over the *primitive* lattice vectors, or even just the discrete Gaussian over $\mathcal{L} \setminus \{\mathbf{0}\}$ might be strictly harder than centered DGS. In particular, in Section 5.5, we show a family of lattices for which $D_{\mathcal{L},s}$ almost never returns a $o(\sqrt{n/\log n})$-approximate shortest vector. However, it is easy to see that the discrete Gaussian over the *primitive* lattice vectors or even just over the lattice without $\mathbf{0}$ will output the shortest vector with overwhelming probability if the parameter s is sufficiently small. Therefore, both of these sampling problems are actually polynomial-time equivalent to SVP, while we have some evidence to suggest that sampling from $D_{\mathcal{L},s}$ is not. Indeed, we know of no application of centered DGS in which non-primitive vectors are actually desirable.

because it does not have a factor of $2^{o(n)}$.

Lemma 5.2.1. *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and $r > 0$,*

$$|\mathcal{L} \cap rB_2^n| \leq 1 + \left(\frac{8r}{\lambda_1(\mathcal{L})}\right)^n .$$

For simplicity, we endeavor to ignore the specific representation of the basis vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n \in \mathbb{R}^n$. Here, we note a simple bound that applies if the basis vectors are rational and represented in the natural way. In the sequel, we will apply this even to vectors in $\mathbb{R}^n$, with the understanding that the same corollary applies for any "reasonable" representation of real numbers.

Corollary 5.2.2. *For any lattice $\mathcal{L} \subset \mathbb{Q}^n$ with basis $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$, $\boldsymbol{t} \in \mathbb{Q}^n$, and $r > 0$, let m be a bound on the bit length of the $\boldsymbol{b}_i$ for all i in the natural representation of rational numbers. Then,*

$$|(\mathcal{L} - \boldsymbol{t}) \cap rB_2^n| \leq 1 + (2 + r)^{\mathrm{poly}(n,m)} .$$

The following lemma is actually true for "almost all lattices," in a certain precise sense that is outside the scope of this thesis. (See, e.g., [Sie45].)

Lemma 5.2.3. *For any $n \geq 1$, there is a lattice $\mathcal{L} \subset \mathbb{R}^n$ such that for any $s > 0$, $\rho_s(\mathcal{L}) \geq 1 + s^n$ and $\lambda_1(\mathcal{L}) > \sqrt{n}/10$.*

5.2.1 Variants of DGS

We will need a slight generalization of our definition of DGS from the previous chapters. In particular, we will add a "multiplicative error term" γ.

Definition 5.2.4. *For $\gamma \geq 1$ and $\varepsilon \geq 0$, we say that a distribution X is (γ, ε)-close to a distribution Y if there is another distribution X' with the same support as Y such that*

1. *the statistical distance between X and X' is at most ε; and*

2. *for all x in the support of Y, $\Pr[Y = x]/\gamma \leq \Pr[X' = x] \leq \gamma \Pr[Y = x]$.*

Definition 5.2.5. *For any parameters $\varepsilon \geq 0$ and $\gamma \geq 1$, (γ, ε)-DGS (the Discrete Gaussian Sampling problem) is defined as follows: The input is a basis $\mathbf{B}$ for a lattice $\mathcal{L} \subset \mathbb{R}^n$, a shift $\boldsymbol{t} \in \mathbb{Q}^n$, and a parameter $s > 0$. The goal is to output a vector whose distribution is (γ, ε)-close to $D_{\mathcal{L}-\boldsymbol{t},s}$.*

We also explicitly define the problem of sampling from the *centered* discrete Gaussian $D_{\mathcal{L},s}$.

Definition 5.2.6. *For any parameters $\varepsilon \geq 0$ and $\gamma \geq 1$, (γ, ε)-cDGS (the centered Discrete Gaussian Sampling problem) is defined as follows: The input is a basis $\mathbf{B}$ for a lattice $\mathcal{L} \subset \mathbb{R}^n$ and a parameter $s > 0$. The goal is to output a vector whose distribution is (γ, ε)-close to $D_{\mathcal{L},s}$.*

5.2.2 Lattice vectors mod p and $\mathbb{Z}_p^n$

Our primary technical tool will be lattice sparsification, in which we consider the sublattice

$$\mathcal{L}' := \{\boldsymbol{y} \in \mathcal{L} \ : \ \langle \boldsymbol{z}, \mathbf{B}^{-1}\boldsymbol{y} \rangle \equiv 0 \bmod p\} \ ,$$

where p is some prime, $\boldsymbol{z} \in \mathbb{Z}_p^n$ is uniformly random, and $\mathbf{B}$ is a basis of the lattice $\mathcal{L} \subset \mathbb{R}^n$. As such, we will need some lemmas concerning the behavior of lattice vectors mod $p\mathcal{L}$. We first simply note that we can compute $\mathcal{L}'$ efficiently.

Claim 5.2.7. *There is a polynomial-time algorithm that takes as input a basis $\mathbf{B}$ for a lattice $\mathcal{L} \subset \mathbb{R}^n$, a number $p \in \mathbb{Z}^+$, and a vector $\boldsymbol{z} \in \mathbb{Z}_p^n$ and outputs a basis $\mathbf{B}'$ for*

$$\mathcal{L}' := \{\boldsymbol{y} \in \mathcal{L} \ : \ \langle \boldsymbol{z}, \mathbf{B}^{-1}\boldsymbol{y} \rangle \equiv 0 \bmod p\} \ .$$

Proof. On input $\mathbf{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$, $p \in \mathbb{Z}^+$, and $\boldsymbol{z} = (z_1, \ldots, z_n) \in \mathbb{Z}_p^n$, if $\boldsymbol{z} = \mathbf{0}$, the algorithm simply outputs $\mathbf{B}$. Otherwise, by possibly reordering the basis, we may assume without loss of generality that $z_n \neq 0$. The algorithm then computes $\mathbf{B}^* := \mathbf{B}^{-T} = (\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_n^*)$. It sets

$$\widehat{\mathbf{B}} := \left(\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_{n-1}^*, \frac{1}{p}\sum z_i \boldsymbol{b}_i^*\right).$$

Notice that, since $z_n \neq 0$ and $\mathbf{B}^*$ is non-singular, $\widehat{\mathbf{B}}$ must be non-singular. (In particular, $\det(\widehat{\mathbf{B}}) = z_n \det(\mathbf{B}^*)/p$.) Finally, it outputs $\mathbf{B}' := \widehat{\mathbf{B}}^{-T}$. A quick computation shows that $\mathbf{B}'$ is indeed a basis for $\mathcal{L}'$. □

Since we will only be concerned with the coordinates of the vectors mod p, it will suffice to work over $\mathbb{Z}_p^n$.

Lemma 5.2.8 ([Ste16a, Lemma 2.16])**.** *For any prime p and collection of vectors $\boldsymbol{x}, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_N \in \mathbb{Z}_p^n \setminus \{\mathbf{0}\}$ such that $\boldsymbol{x}$ is not a scalar multiple of any of the $\boldsymbol{v}_i$, we have*

$$\frac{1}{p} - \frac{N}{p^2} \leq \Pr\left[\langle \boldsymbol{z}, \boldsymbol{x}\rangle \equiv 0 \bmod p \text{ and } \langle \boldsymbol{z}, \boldsymbol{v}_i\rangle \not\equiv 0 \bmod p \ \forall i\right] \leq \frac{1}{p},$$

where $\boldsymbol{z}$ is sampled uniformly at random from $\mathbb{Z}_p^n$.

Proof. For the upper bound, it suffices to note that, since $\boldsymbol{x}$ is non-zero and p is prime, $\langle \boldsymbol{z}, \boldsymbol{x}\rangle$ is uniformly distributed over $\mathbb{Z}_p$. Therefore, $\Pr[\langle \boldsymbol{z}, \boldsymbol{x}\rangle \equiv 0 \bmod p] = 1/p$. For the lower bound, note that $A := \{\boldsymbol{y} \in \mathbb{Z}_p^n \ : \ \langle \boldsymbol{y}, \boldsymbol{x}\rangle \equiv 0 \bmod p\}$ and $B_i := \{\boldsymbol{y} \in \mathbb{Z}_p^n \ : \ \langle \boldsymbol{y}, \boldsymbol{v}_i\rangle \equiv 0 \bmod p\}$ are distinct subspaces of dimension $n-1$. Therefore, $A \cap B_i$ is a subspace of dimension $n-2$

with p^{n-2} elements. Let $B := \bigcup B_i$. It follows that

$$\begin{aligned}\Pr\left[\langle \boldsymbol{z}, \boldsymbol{x}\rangle \equiv 0 \bmod p \text{ and } \langle \boldsymbol{z}, \boldsymbol{v}_i\rangle \not\equiv 0 \bmod p\right] &= \frac{|A \setminus B|}{|\mathbb{Z}_p^n|} \\ &\geq \frac{|A| - \sum_i |A \cap B_i|}{|\mathbb{Z}_p^n|} \\ &= \frac{p^{n-1} - Np^{n-2}}{p^n} \\ &= \frac{1}{p} - \frac{N}{p^2}\,.\end{aligned}$$

□

Corollary 5.2.9 ([Ste16a, Corollary 2.17])**.** *For any prime p, collection of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N \in \mathbb{Z}_p^n$, and $\boldsymbol{x} \in \mathbb{Z}_p^n$ with $\boldsymbol{x} \neq \boldsymbol{v}_i$ for any i, we have*

$$\frac{1}{p} - \frac{N}{p^2} - \frac{N}{p^{n-1}} \leq \Pr\left[\langle \boldsymbol{z}, \boldsymbol{x} + \boldsymbol{c}\rangle \equiv 0 \bmod p \text{ and } \langle \boldsymbol{z}, \boldsymbol{v}_i + \boldsymbol{c}\rangle \not\equiv 0 \bmod p\ \forall i\right] \leq \frac{1}{p} + \frac{1}{p^n}\,,$$

where $\boldsymbol{z}$ and $\boldsymbol{c}$ are sampled uniformly and independently at random from $\mathbb{Z}_p^n$.

Proof. For the upper bound, it suffices to note that

$$\Pr[\langle \boldsymbol{z}, \boldsymbol{x} + \boldsymbol{c}\rangle \equiv 0 \bmod p] = \Pr[\boldsymbol{x} + \boldsymbol{c} \equiv \boldsymbol{0} \bmod p] + \frac{1}{p}\Pr[\boldsymbol{x} + \boldsymbol{c} \not\equiv \boldsymbol{0} \bmod p] = \frac{1}{p} + \frac{1}{p^n} - \frac{1}{p^{n+1}}\,.$$

Turning to the lower bound, note that for any i, we have $\Pr[\boldsymbol{v}_i + \boldsymbol{c} \equiv \boldsymbol{0} \bmod p] = 1/p^n$. By union bound, the probability that $\boldsymbol{v}_i + \boldsymbol{c} \equiv \boldsymbol{0} \bmod p$ for any i is at most N/p^n. Now, fix i, and note that if there exists some $\alpha \in \mathbb{Z}_p \setminus \{1\}$ such that $\alpha(\boldsymbol{v}_i + \boldsymbol{c}) \equiv \boldsymbol{x} + \boldsymbol{c} \bmod p$, then we must have

$$\boldsymbol{c} \equiv \frac{\alpha \boldsymbol{v}_i - \boldsymbol{x}}{1 - \alpha} \bmod p\,.$$

There are therefore at most $p - 1$ values for $\boldsymbol{c}$ that satisfy the above—one for each value of α. So, the probability that $\boldsymbol{c}$ will satisfy the above equation for any α is at most $(p-1)/p^n$.

Taking a union bound over all i, we see that the probability that $\boldsymbol{x} + \boldsymbol{c}$ is a multiple of any of the $\boldsymbol{v}_i + \boldsymbol{c}$ is at most $N(p-1)/p^n$. The result then follows from Lemma 5.2.8 and union bound. □

5.2.3 Primitive lattice vectors

For a lattice $\mathcal{L} \subset \mathbb{R}^n$, we say that $\boldsymbol{y} \in \mathcal{L}$ is non-primitive in $\mathcal{L}$ if $\boldsymbol{y} = k\boldsymbol{x}$ for some $\boldsymbol{x} \in \mathcal{L}$ and $k \geq 2$. Otherwise, $\boldsymbol{y}$ is primitive in $\mathcal{L}$. Let $\mathcal{L}_{\text{prim}}$ be the set of primitive vectors in $\mathcal{L}$. For a radius $r > 0$, let $\xi(\mathcal{L}, r) := |\mathcal{L}_{\text{prim}} \cap rB_2^n|/2$ be the number of primitive lattice vectors in a (closed) ball of radius r around the origin (counting $\boldsymbol{x}$ and $-\boldsymbol{x}$ as a single vector).

We will need the following technical lemma, which shows that relatively short primitive vectors cannot be scalar multiples of each other mod p.

Lemma 5.2.10 ([Ste16a, Lemma 2.18])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ with basis $\mathbf{B}$, suppose $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathcal{L}$ are primitive with $\boldsymbol{y}_1 \neq \pm\boldsymbol{y}_2$ and $\|\boldsymbol{y}_1\| \geq \|\boldsymbol{y}_2\|$ such that*

$$\mathbf{B}^{-1}\boldsymbol{y}_1 \equiv \alpha\mathbf{B}^{-1}\boldsymbol{y}_2 \bmod p$$

for any number $p \geq 100$ and $\alpha \in \mathbb{Z}_p$. Then, $\xi(\mathcal{L}, \|\boldsymbol{y}_1\|) > p/(20 \log p)$.

Proof. Notice that $\alpha \neq 0$, since otherwise $\boldsymbol{y}_1$ is not even primitive. So, we have that $\boldsymbol{y}_1 - q\boldsymbol{x}_2 \in p\mathcal{L} \setminus \{\mathbf{0}\}$ for some integer $q \equiv \alpha \bmod p$ with $0 < |q| \leq p/2$. Let $\boldsymbol{x} := (\boldsymbol{y}_1 - q\boldsymbol{y}_2)/p \in \mathcal{L}$ and note that $\boldsymbol{x}$ is not a scalar multiple of $\boldsymbol{y}_2$. It suffices to find at least $\lceil p/(20 \log p) \rceil$ primitive vectors in the lattice spanned by $\boldsymbol{x}$ and $\boldsymbol{y}_2$ that are at least as short as $\boldsymbol{y}_1$.

We consider two cases. If $q = \pm 1$, then for $i = 0, \ldots, p-1$, the vectors $i\boldsymbol{x} + q\boldsymbol{y}_2$ are clearly primitive in the lattice spanned by $\boldsymbol{x}$ and $\boldsymbol{y}_2$, and we have

$$\|i\boldsymbol{x} + q\boldsymbol{y}_2\| = \|i\boldsymbol{y}_1 + q(p-i)\boldsymbol{y}_2\|/p \leq \|\boldsymbol{y}_1\| \,,$$

as needed.

Now, suppose $|q| > 1$. Then, for $i = \lceil p/4 \rceil, \ldots, \lfloor p/2 \rfloor$, let k_i be an integer such that $|k_i - iq/p| \leq 1/2$ and $0 < |k_i| < i$. (Note that such an integer exists, since $1/2 \leq |iq/p| \leq i/2$). Then,

$$\|i\boldsymbol{x} + k_i\boldsymbol{y}_2\| = \|i\boldsymbol{y}_1/p + (k_i - iq/p)\boldsymbol{y}_2\| \leq \|\boldsymbol{y}_1\| .$$

When i is prime, then since $0 < |k_i| < i$, we must have $\gcd(i, k_i) = 1$. Therefore, the vector $i\boldsymbol{x} + k_i\boldsymbol{y}_2$ must be primitive in the lattice spanned by $\boldsymbol{x}$ and $\boldsymbol{y}_2$ when i is prime. It follows from a suitable effective version of the Prime Number Theorem that there are at least $\lceil p/(20 \log p) \rceil$ primes between $\lceil p/4 \rceil$ and $\lfloor p/2 \rfloor$ (see, e.g., [Ros41]), as needed.

□

We next show that we can find many primitive lattice vectors in a suitably large ball around $\mathbf{0}$.

Lemma 5.2.11 ([Ste16a, Lemma 2.19])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ and radius $r \geq \lambda_2(\mathcal{L})$,*

$$\xi(\mathcal{L}, r) > \frac{\sqrt{r^2 - \lambda_2(\mathcal{L})^2}}{\lambda_1(\mathcal{L})} + \left\lfloor \frac{r - \lambda_2(\mathcal{L})}{\lambda_1(\mathcal{L})} \right\rfloor .$$

Proof. Let $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathcal{L}$ with $\|\boldsymbol{v}_i\| = \lambda_i(\mathcal{L})$ and $\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle \geq 0$. Then, for $k = 0, \ldots, \lfloor \sqrt{r^2 - \lambda_2(\mathcal{L})^2}/\lambda_1(\mathcal{L}) \rfloor$,

$$\|\boldsymbol{v}_2 - k\boldsymbol{v}_1\|^2 = \lambda_2(\mathcal{L})^2 + k^2\lambda_1(\mathcal{L})^2 - 2k\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle \leq r^2 .$$

Similarly, for $k = 1, \ldots, \lfloor (r - \lambda_2(\mathcal{L}))/\lambda_1(\mathcal{L}) \rfloor$,

$$\|\boldsymbol{v}_2 + k\boldsymbol{v}_1\| \leq \lambda_2(\mathcal{L}) + k\lambda_1(\mathcal{L}) \leq r$$

The result follows by noting that all of these vectors are distinct and primitive in the lattice

generated by $\boldsymbol{v}_1, \boldsymbol{v}_2$ (as is $\boldsymbol{v}_1$). □

5.3 DGS to CVP reduction

5.3.1 Sparsify and shift

We now present the main sparsification result that we require. In particular, Theorem 5.3.1 (which is immediate from Section 5.2.2, and is presented in this form here for the reader's convenience) shows the generic behavior of the sparsification procedure. Proposition 5.3.2 then applies the theorem to show how sparsification interacts with a CVP oracle.

Theorem 5.3.1 ([Ste16a, Theorem 3.1])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ with basis $\mathbf{B}$, prime p, and lattice vectors $\boldsymbol{x}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_N \in \mathcal{L}$ such that $\mathbf{B}^{-1}\boldsymbol{x} \not\equiv \mathbf{B}^{-1}\boldsymbol{y}_i \bmod p$ for all i, we have*

$$\frac{1}{p} - \frac{N}{p^2} - \frac{N}{p^{n-1}} \leq \Pr[\langle \boldsymbol{z}, \mathbf{B}^{-1}\boldsymbol{x} + \boldsymbol{c}\rangle \equiv \mathbf{0} \text{ and } \langle \boldsymbol{z}, \mathbf{B}^{-1}\boldsymbol{y}_i + \boldsymbol{c}\rangle \not\equiv 0 \bmod p \ \forall i] \leq \frac{1}{p} + \frac{1}{p^n} ,$$

where $\boldsymbol{z}, \boldsymbol{c} \in \mathbb{Z}_p^n$ are chosen uniformly and independently at random.

Proof. Simply apply Corollary 5.2.9 to $\mathbf{B}^{-1}\boldsymbol{x}$ and $\mathbf{B}^{-1}\boldsymbol{y}_i$. □

Proposition 5.3.2 ([Ste16a, Proposition 3.2])**.** *There is a polynomial-time algorithm that takes as input a basis $\mathbf{B}$ for a lattice $\mathcal{L} \subset \mathbb{R}^n$ and a prime p and outputs a full-rank sublattice $\mathcal{L}' \subseteq \mathcal{L}$ and shift $\boldsymbol{w} \in \mathcal{L}$ such that, for any $\boldsymbol{t} \in \mathbb{R}^n$, $\boldsymbol{x} \in \mathcal{L}$ with $N := |\mathcal{L} \cap (\|\boldsymbol{x} - \boldsymbol{t}\| \cdot B_2^n + \boldsymbol{t})| - 1 < p$, and any CVP oracle,*

$$\frac{1}{p} - \frac{N}{p^2} - \frac{N}{p^{n-1}} \leq \Pr[CVP(\boldsymbol{t} + \boldsymbol{w}, \mathcal{L}') = \boldsymbol{x} + \boldsymbol{w}] \leq \frac{1}{p} + \frac{1}{p^n} .$$

In particular,

$$\frac{N}{p} - \frac{N^2}{p^2} - \frac{N^2}{p^{n-1}} \leq \Pr[\mathrm{dist}(\boldsymbol{t}+\boldsymbol{w}, \mathcal{L}') \leq \|\boldsymbol{x}-\boldsymbol{t}\|] \leq \frac{N}{p} + \frac{N}{p^n} \,.$$

Proof. On input $\mathcal{L} \subset \mathbb{R}^n$ with basis $\mathbf{B}$ and p, the algorithm samples $\boldsymbol{z}, \boldsymbol{c} \in \mathbb{Z}_p^n$ uniformly and independently at random. It then returns the sublattice

$$\mathcal{L}' := \{\boldsymbol{x} \in \mathcal{L} \ : \ \langle \boldsymbol{z}, \mathbf{B}^{-1}\boldsymbol{x} \rangle \equiv 0 \bmod p\} \,,$$

and the shift $\boldsymbol{w} := \mathbf{B}\boldsymbol{c}$.

By Claim 5.2.7, the algorithm can be run in polynomial time. Let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N \in \mathcal{L}$ be the unique vectors such that $\|\boldsymbol{y}_i - \boldsymbol{t}\| \leq \|\boldsymbol{x} - \boldsymbol{t}\|$ with $\boldsymbol{y}_i \neq \boldsymbol{x}$. Note that $\mathrm{CVP}(\mathcal{L}', \boldsymbol{t}+\boldsymbol{w})$ must be $\boldsymbol{x}+\boldsymbol{w}$ if $\langle \boldsymbol{z}, \mathbf{B}^{-1}\boldsymbol{y}_i + \boldsymbol{c} \rangle \not\equiv 0 \bmod p$ for all i *and* $\langle \boldsymbol{z}, \mathbf{B}^{-1}\boldsymbol{x} + \boldsymbol{c} \rangle \equiv 0 \bmod p$. We therefore wish to apply Theorem 5.3.1, which requires showing that $\mathbf{B}^{-1}\boldsymbol{y}_i \not\equiv \mathbf{B}^{-1}\boldsymbol{x} \bmod p$ for all i.

Suppose on the contrary that $\mathbf{B}^{-1}\boldsymbol{y}_i \equiv \mathbf{B}^{-1}\boldsymbol{x} \bmod p$ for some i. Then, $\boldsymbol{y} := \boldsymbol{y}_i - \boldsymbol{x} \in p\mathcal{L} \setminus \{\boldsymbol{0}\}$, and there are therefore $p+1$ lattice vectors on the line segment between $\boldsymbol{y}_i$ and $\boldsymbol{x}$ (including the two endpoints). Note that all of these vectors are at least as close to $\boldsymbol{t}$ as $\boldsymbol{x}$. But, there can be at most $N+1 < p+1$ such vectors, a contradiction. Therefore, we can apply Theorem 5.3.1, yielding the result. □

As a consequence of Proposition 5.3.2, we show that we can use a CVP oracle to sample nearly uniformly from the lattice points in a ball around $\boldsymbol{t}$. This relatively straightforward algorithm is the core idea behind our reduction. For simplicity, we provide the algorithm with an estimate of the number of points inside the ball as input. (In the next section, we show how to obtain this estimate using roughly the same techniques.)

Lemma 5.3.3 ([Ste16a, Lemma 3.3])**.** *For any efficiently computable $f(n)$ with $2 \leq f(n) \leq \mathrm{poly}(n)$, there is an algorithm with access to a CVP oracle that takes as input a lattice*

$\mathcal{L} \subset \mathbb{R}^n$, shift $\boldsymbol{t} \in \mathbb{R}^n$, radius $r > 0$, and integer $N \geq 1$ and outputs a vector $\boldsymbol{y}$ such that, if

$$N \leq |\mathcal{L} \cap (rB_2^n + \boldsymbol{t})| \leq f(n)N \ ,$$

then the algorithm runs in expected polynomial time, and for any $\boldsymbol{x} \in \mathcal{L} \cap (rB_2^n + \boldsymbol{t})$,

$$\frac{\gamma^{-1}}{|\mathcal{L} \cap (rB_2^n + \boldsymbol{t})|} \leq \Pr[\boldsymbol{y} = \boldsymbol{x}] \leq \frac{\gamma}{|\mathcal{L} \cap (rB_2^n + \boldsymbol{t})|} \ ,$$

where $\gamma := 1 + 1/f(n)$. Furthermore, all of the algorithm's oracle calls are on full-rank sublattices of the input lattice.

Proof. We assume without loss of generality that $n \geq 2$. On input $\mathcal{L} \subset \mathbb{R}^n$, $\boldsymbol{t} \in \mathbb{R}^n$, $r > 0$, and $N \geq 1$, the algorithm chooses a prime p with $10f(n)N \leq p \leq 20f(n)N$ and calls the procedure from Proposition 5.3.2 on input $\mathcal{L}$ and p, receiving as output a sublattice $\mathcal{L}' \subseteq \mathcal{L}$ and a shift $\boldsymbol{w} \in \mathcal{L}$. It then calls its CVP oracle on input $\mathcal{L}'$ and $\boldsymbol{t} + \boldsymbol{w}$, receiving as output $\boldsymbol{y}'$. If $\|\boldsymbol{y}' - \boldsymbol{w} - \boldsymbol{t}\| \leq r$, it outputs $\boldsymbol{y} := \boldsymbol{y}' - \boldsymbol{w}$. Otherwise, it repeats.

From Proposition 5.3.2, we have that, after a single run of the algorithm,

$$\frac{1}{\sqrt{\gamma} \cdot p} \leq \frac{1}{p} - \frac{N}{p^2} - \frac{N}{p^{n-1}} \leq \Pr[\boldsymbol{y}' = \boldsymbol{x} + \boldsymbol{w}] \leq \frac{1}{p} + \frac{1}{p^n} \leq \frac{\sqrt{\gamma}}{p} \ .$$

Correctness follows immediately. Furthermore, note that the reduction outputs something on each run with probability at least $\frac{N}{\sqrt{\gamma}p} \geq \frac{1}{100f(n)}$. So, in particular, the expected number of runs is polynomial in n. It is clear that a single run takes polynomial time, and the result follows. □

5.3.2 Counting the lattice vectors in a ball

We now show how to use the sparsification algorithm to approximate the number of lattice points in a ball, given access to a CVP oracle. We will use this both to instantiate the procedure from Lemma 5.3.3 and directly in our DGS sampling procedure.

Definition 5.3.4. *For any parameter $\gamma := \gamma(n) \geq 1$, γ-GapVCP (the Vector Counting Problem) is the promise problem defined as follows: the input is a (basis for a) lattice $\mathcal{L} \subset \mathbb{R}^n$, shift $\boldsymbol{t} \in \mathbb{R}^n$, radius $r > 0$, and an integer $N \geq 1$. It is a NO instance if $|\mathcal{L} \cap (rB_2^n + \boldsymbol{t})| \leq N$ and a YES instance if $|\mathcal{L} \cap (rB_2^n + \boldsymbol{t})| > \gamma N$.*

Theorem 5.3.5 ([Ste16a, Theorem 3.5])**.** *For any efficiently computable function $f(n)$ with $1 \leq f(n) \leq \mathrm{poly}(n)$, there is a polynomial-time reduction from γ-GapVCP to CVP where $\gamma := 1 + 1/f(n)$. The reduction only calls the CVP oracle on full-rank sublattices of the input lattice.*

Proof. We assume without loss of generality that $n \geq 20$ and $f(n) \geq 20$. On input a lattice $\mathcal{L} \subset \mathbb{R}^n$ with basis $\mathbf{B}$, target $\boldsymbol{t} \in \mathbb{R}^n$, $r > 0$, and $N \geq 1$, the reduction behaves as follows. First, it finds a prime p with $200f(n)N \leq p \leq 400f(n)N$. Then, for $i = 1, \ldots, \ell := \lceil 100f(n)^2p^2/N^2 \rceil$, the reduction calls the procedure from Proposition 5.3.2 on $\mathcal{L}$, $\boldsymbol{t}$, and p. It receives as output $\mathcal{L}_i$ and $\boldsymbol{w}_i$. It then calls the CVP oracle on $\mathcal{L}_i$ and $\boldsymbol{t} + \boldsymbol{w}_i$, receiving as output a vector whose distance from $\boldsymbol{t} + \boldsymbol{w}_i$ is r_i. Finally, it returns yes if $r \leq r_i$ for all but at most $\ell N/p + 2\sqrt{\ell}$ values of r_i and no otherwise.

It is clear that the reduction runs in polynomial time. Now, suppose $|\mathcal{L} \cap (rB_2^n + \boldsymbol{t})| \leq N$. By Proposition 5.3.2, we have that for each i,

$$\Pr[r_i \leq r] \leq \frac{N}{p} + \frac{N}{p^n} < \frac{N}{p} + \frac{1}{2\sqrt{\ell}} .$$

Then, applying the Chernoff-Hoeffding bound (Lemma 1.4.3), we have

$$\Pr[|\{i \ : \ r_i \le r\}| > \ell N/p + 2\sqrt{\ell}] < 1/e \ .$$

So, the reduction returns the correct answer in this case with probability at least $1 - 1/e$.

On the other hand, suppose that $|\mathcal{L} \cap (rB_2^n + \boldsymbol{t})| > \gamma N$. Using the lower bound in Proposition 5.3.2,

$$\Pr[r_i \le r] \ge \frac{\gamma N}{p} - \frac{\gamma^2 N^2}{p^2} - \frac{\gamma^2 N^2}{p^{n-1}} > \frac{N}{p} + \frac{5}{\sqrt{\ell}} \ .$$

Applying the Chernoff-Hoeffding bound again, we have

$$\Pr[|\{i \ : \ r_i \le r\}| \le \ell N/p + 2\sqrt{\ell}] < 1/e \ ,$$

as needed. □

5.3.3 The DGS algorithm

Theorem 5.3.6 ([Ste16a, Theorem 3.6])**.** *For any efficiently computable function $f(n)$ with $1 \le f(n) \le \mathrm{poly}(n)$, there exists an (expected) polynomial-time reduction from (γ, ε)-DGS to CVP, where $\varepsilon := 2^{-f(n)}$ and $\gamma := 1 + 1/f(n)$. The reduction only calls the CVP oracle on full-rank sublattices of the input lattice.*

Proof. We assume without loss of generality that $n \ge 5$ and $s = 1$. (If $s \ne 1$, we can simply rescale the lattice.) On input $\mathcal{L} \subset \mathbb{R}^n$ and $\boldsymbol{t} \in \mathbb{R}^n$, the reduction behaves as follows. It first calls its CVP oracle to compute $d := \mathrm{dist}(\boldsymbol{t}, \mathcal{L})$. For $i = 0, \ldots, \ell := \lceil 100n^2 f(n)^2 \log(10 + 10d^2/n) \rceil$, let $r_i := \sqrt{d^2 + i/(10f(n))}$. For each i, the reduction uses its CVP oracle together with the procedure given in Theorem 5.3.5 to compute N_i such that $\gamma^{-1/10} \cdot |\mathcal{L} \cap (r_i B_2^n + \boldsymbol{t})| \le N_i \le |\mathcal{L} \cap (r_i B_2^n + \boldsymbol{t})|$.

Let $w_\ell := e^{-\pi r_\ell^2}$, and for $i = 0, \ldots, \ell - 1$, let $w_i := e^{-\pi r_i^2} - e^{-\pi r_{i+1}^2}$. Let $W := \sum_{i=0}^{\ell} N_i w_i$. The reduction then chooses an index $0 \le k \le \ell$ from the distribution that assigns to index i probability $N_i w_i / W$. It then runs the procedure from Lemma 5.3.3 with input $\mathcal{L}$, $\boldsymbol{t}$, r_k, and N_k, receiving as output a vector $\boldsymbol{y} \in \mathcal{L} \cap (r_k B_2^n + \boldsymbol{t})$ whose distribution is $(\gamma^{1/10}, 0)$-close to the uniform distribution over $\mathcal{L} \cap (r_k B_2^n + \boldsymbol{t})$. It then simply returns $\boldsymbol{y}$.

It is clear that the reduction runs in polynomial time (subject to a suitable version of Corollary 5.2.2 for the specific form of the input).

We now prove correctness. Let $A := \mathcal{L} \cap (r_\ell B_2^n + \boldsymbol{t})$ be the support of $\boldsymbol{y}$. By Corollary 1.3.11, D_A is within statistical distance ε of $D_{\mathcal{L}-\boldsymbol{t}}$, so it suffices to show that the output of the reduction is $(\gamma, 0)$-close to D_A. In order to show this, it suffices to show that, for any $\boldsymbol{x} \in A$, $\Pr[\boldsymbol{y} = \boldsymbol{x}]$ is proportional to $\rho(\boldsymbol{x})$, up to a factor of $\gamma^{\pm 1/2}$. Note that

$$\Pr[\boldsymbol{y} = \boldsymbol{x}] = \frac{1}{W} \sum_{i \,:\, r_i \ge \|\boldsymbol{x} - \boldsymbol{t}\|} w_i N_i \cdot \Pr[\boldsymbol{y} = \boldsymbol{x} \mid k = i] \,. \tag{5.2}$$

For any i such that $\boldsymbol{x} \in (\mathcal{L} - \boldsymbol{t}) \cap r_i B_2^n$, by Lemma 5.3.3 we have that

$$\frac{\gamma^{-1/5}}{N_i} \le \frac{\gamma^{-1/10}}{|\mathcal{L} \cap (r_i B_2^n + \boldsymbol{t})|} \le \Pr[\boldsymbol{y} = \boldsymbol{x} \mid k = i] \le \frac{\gamma^{1/10}}{|\mathcal{L} \cap (r_i B_2^n + \boldsymbol{t})|} \le \frac{\gamma^{1/10}}{N_i} \,.$$

Let j be minimal such that $\boldsymbol{x} \in (\mathcal{L} - \boldsymbol{t}) \cap r_j B_2^n$. Plugging in the upper bound to Eq. (5.2), we have

$$\Pr[\boldsymbol{y} = \boldsymbol{x}] \le \frac{\gamma^{1/10}}{W} \cdot \sum_{i \ge j} w_i = \frac{\gamma^{1/10}}{W} \cdot e^{-\pi r_j^2} \le \frac{\sqrt{\gamma}}{W} \cdot \rho(\boldsymbol{x}) \,.$$

A nearly identical computation shows that $\Pr[\boldsymbol{y} = \boldsymbol{x}] \ge \rho(\boldsymbol{x}) / (\sqrt{\gamma} W)$, as needed. □

5.4 Centered DGS to SVP reduction

5.4.1 Sparsification

Since we are now interested in the SVP case, we can no longer handle the shifts used in Theorem 5.3.1 and Proposition 5.3.2 (neither the input shift $\boldsymbol{t}$ nor the output shifts $\boldsymbol{w}$ and $\boldsymbol{c}$). As a result, we are forced to consider the effect of sparsification on primitive vectors only, which requires new analysis. Recall that $\xi(\mathcal{L}, r) := |\mathcal{L}_{\text{prim}} \cap r B_2^n|/2$ is the number of primitive lattice vectors in a ball of radius r (counting $\pm\boldsymbol{x}$ as a single vector).

Theorem 5.4.1 ([Ste16a, Theorem 4.1])**.** *For any lattice $\mathcal{L} \subset \mathbb{R}^n$ with basis $\mathbf{B}$, primitive lattice vectors $\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_N \in \mathcal{L}_{\text{prim}}$ with $\boldsymbol{y}_i \neq \pm\boldsymbol{y}_0$ for all $i > 0$, and prime $p \geq 101$, if $\xi(\mathcal{L}, \|\boldsymbol{y}_i\|) \leq p/(20 \log p)$ for all i, then*

$$\frac{1}{p} - \frac{N}{p^2} \leq \Pr\left[\langle \boldsymbol{z}, \mathbf{B}^{-1}\boldsymbol{y}_0\rangle \equiv 0 \bmod p \text{ and } \langle \boldsymbol{z}, \mathbf{B}^{-1}\boldsymbol{y}_i\rangle \not\equiv 0 \bmod p \ \forall i > 0\right] \leq \frac{1}{p}\,,$$

where $\boldsymbol{z} \in \mathbb{Z}_p^n$ is chosen uniformly at random.

Proof. Let $\boldsymbol{v}_i := \mathbf{B}^{-1}\boldsymbol{y}_i$. By Lemma 5.2.10, we have that $\boldsymbol{v}_0$ is not a scalar multiple of $\boldsymbol{v}_i$ mod p for any $i > 0$. The result then follows from Lemma 5.2.8. □

Proposition 5.4.2 ([Ste16a, Proposition 4.2])**.** *There is a polynomial-time algorithm that takes as input a basis $\mathbf{B}$ for a lattice $\mathcal{L} \subset \mathbb{R}^n$ and a prime $p \geq 101$ and outputs a full-rank sublattice $\mathcal{L}' \subseteq \mathcal{L}$ such that for every $\boldsymbol{x} \in \mathcal{L}$ with $N := \xi(\mathcal{L}, \|\boldsymbol{x}\|) - 1 \leq p/(20 \log p)$ and $\lambda_1(\mathcal{L}) > \|\boldsymbol{x}\|/p$, we have that for any SVP oracle,*

$$\frac{1}{p} - \frac{N}{p^2} \leq \Pr[SVP(\mathcal{L}') = \pm\boldsymbol{x}] \leq \frac{1}{p}\,.$$

In particular,

$$\frac{N}{p} - \frac{N^2}{p^2} \leq \Pr\left[\lambda_1(\mathcal{L}') \leq \|\boldsymbol{x}\|\right] \leq \frac{N}{p}\,.$$

Proof. On input $\mathcal{L} \subset \mathbb{R}^n$ with basis $\mathbf{B}$ and p, the algorithm samples $\boldsymbol{z} \in \mathbb{Z}_p^n$ uniformly at random. It then returns the sublattice

$$\mathcal{L}' := \{\boldsymbol{x} \in \mathcal{L} \ : \ \langle \boldsymbol{z}, \mathbf{B}^{-1}\boldsymbol{x} \rangle \equiv 0 \bmod p\} \ .$$

It is clear that the algorithm runs in polynomial time. Since $\Pr[\boldsymbol{x} \in \mathcal{L}'] = 1/p$, the upper bound on the probability is immediate as well.

For the lower bound, let $\boldsymbol{y}_0, \ldots, \boldsymbol{y}_N \in \mathcal{L}_{\text{prim}}$ such that $\|\boldsymbol{y}_i\| \leq \|\boldsymbol{x}\|$, $\boldsymbol{y}_i \neq \pm\boldsymbol{y}_j$, and $\boldsymbol{y}_0 := \boldsymbol{x}$. Let $\boldsymbol{v}_i := \mathbf{B}^{-1}\boldsymbol{y}_i$. Note that, if $\boldsymbol{v}_0 \in \mathcal{L}'$ and $\boldsymbol{v}_i \notin \mathcal{L}'$ for $i > 0$, then $\mathrm{SVP}(\mathcal{L}') = \pm\boldsymbol{x}$. (Here, we have used the fact that $\lambda_1(\mathcal{L}) > \|\boldsymbol{x}\|/p$.) The result then follows from Theorem 5.4.1. □

Lemma 5.4.3 ([Ste16a, Lemma 4.3])**.** *For any efficiently computable $f(n)$ with $2 \leq f(n) \leq \mathrm{poly}(n)$, there is an (expected) polynomial-time algorithm with access to a SVP oracle that takes as input a lattice $\mathcal{L} \subset \mathbb{R}^n$, radius $r > 0$, and integer $N \geq 1$ and outputs a vector $\boldsymbol{y} \in \mathcal{L}$ such that, if $N \leq \xi(\mathcal{L}, r) \leq f(n)N$ and $\lambda_1(\mathcal{L}) > r/(f(n)\xi(\mathcal{L}, r))$, then for any $\boldsymbol{x} \in \mathcal{L}_{\mathsf{prim}} \cap rB_2^n$,*

$$\frac{\gamma^{-1}}{\xi(\mathcal{L}, r)} \leq \Pr[\boldsymbol{y} = \pm\boldsymbol{x}] \leq \frac{\gamma}{\xi(\mathcal{L}, r)} \ ,$$

where $\gamma := 1 + f(n)$. Furthermore, the algorithm only calls its oracle on full-rank sublattices of $\mathcal{L}$.

Proof. We assume without loss of generality that $n \geq 10$. On input $\mathcal{L} \subset \mathbb{R}^n$, $r > 0$, and $N \geq 1$, the algorithm chooses a prime p with $100f(n)N\log(10f(n)N) \leq p \leq 200f(n)N\log(10f(n)N)$ and calls the algorithm from Proposition 5.4.2 on input $\mathcal{L}$ and p, receiving as output a sublattice $\mathcal{L}' \subset \mathcal{L}$. It then calls its SVP oracle on input $\mathcal{L}'$, receiving as output $\boldsymbol{y}$. If $\|\boldsymbol{y}\| \leq r$, it outputs $\boldsymbol{y}$. Otherwise, it repeats.

From Proposition 5.4.2, we have that, after a single run of the algorithm

$$\frac{\gamma^{-1/2}}{p} \leq \frac{1}{p} - \frac{N}{p^2} - \frac{N}{p^{n-1}} \leq \Pr[\boldsymbol{y} = \pm\boldsymbol{x}] \leq \frac{1}{p} .$$

Correctness follows immediately. Furthermore, note that the algorithm terminates after a given run with probability at least $\gamma^{-1/2}N/(f(n)p) \geq 1/(1000f(n)^2 \log(Nf(n)))$. By Corollary 5.2.2, $\log(N)$ is polynomial in the length of the input (assuming a reasonable representation of the input). So, in particular, the expected number of runs is polynomial in the length of the input. It is clear that a single run takes polynomial time, and the result follows. □

5.4.2 Counting the primitive lattice vectors in a ball around the origin

Definition 5.4.4. *For any parameters $\beta := \beta(n) \geq 0$, $\gamma := \gamma(n) \geq 1$, (β, γ)-GapPVCP (the Primitive Vector Counting Problem) is the promise problem defined as follows: the input is a (basis for a) lattice $\mathcal{L} \subset \mathbb{R}^n$, radius $r > 0$, and an integer $N \geq 1$. It is a NO instance if $\xi(\mathcal{L}, r) \leq N$ or if $\lambda_1(\mathcal{L}) < \beta r/N$ and a YES instance if $\xi(\mathcal{L}, r) > \gamma N$.*

Intuitively, the condition that $\lambda_1(\mathcal{L}) < \beta r/N$ handles the degenerate case in which there are many non-primitive vectors that may "hide" the primitive vectors in the lattice. It is not clear that this should be treated as a degenerate case in general, but it is clear that our methods must treat this case differently.

Theorem 5.4.5 ([Ste16a, Theorem 4.5])**.** *For any efficiently computable $f(n)$ with $1 \leq f(n) \leq \mathrm{poly}(n)$, there is a polynomial-time reduction from (β, γ)-GapPVCP to SVP where $\beta := 1/f(n)$ and $\gamma := 1 + 1/f(n)$. The reduction only calls the SVP oracle on full-rank sublattices of the input lattice.*

Proof. On input $\mathcal{L} \subset \mathbb{R}^n$ with basis $\mathbf{B}$, $r > 0$, and $N \geq 1$, the reduction behaves as follows. It first calls its SVP oracle on $\mathcal{L}$ to compute $\lambda_1(\mathcal{L})$. If $\lambda_1(\mathcal{L}) > r$ or $\lambda_1(\mathcal{L}) < \beta r/N$, it returns no. The reduction then finds a prime p with $200f(n)N\log(10f(n)N) \leq p \leq 400f(n)N\log(10f(n)N)$, and for $i = 1, \ldots, \ell := \lceil 100f(n)^2p^2/N^2 \rceil$, it calls the procedure from Proposition 5.4.2 on $\mathcal{L}$ and p, receiving as output $\mathcal{L}_i$. It then calls the SVP oracle on each $\mathcal{L}_i$, receiving as output a vector of length r_i. Finally, it returns yes if $r \leq r_i$ for all but at most $\ell N/p + 2\sqrt{\ell}$ values of r_i and no otherwise.

It is clear that the reduction runs in polynomial time. We assume $\lambda_1(\mathcal{L}) \geq \beta r/N > r/p$ (since otherwise the reduction clearly outputs the correct answer).

Suppose $m := \xi(\mathcal{L}, r) \leq N$. By Proposition 5.4.2, we have $\Pr[r_i \leq r] \leq \frac{m}{p} \leq \frac{N}{p}$, for each i. Applying the Chernoff-Hoeffding bound (Lemma 1.4.3), we have

$$\Pr\left[|\{i \ : \ r_i \leq r\}| > \frac{N\ell}{p} + 2\sqrt{\ell}\right] < 1/e \ .$$

So, the reduction returns the correct answer in this case with probability at least $1 - 1/e$.

Now, suppose $\xi(\mathcal{L}, r) > \gamma N$. We again apply Proposition 5.4.2 to obtain

$$\Pr[r_i \leq r] \geq \frac{\gamma N}{p} - \frac{\gamma^2 N^2}{p^2} > \frac{N}{p} + \frac{5}{\sqrt{\ell}}$$

Applying the Chernoff-Hoeffding bound again, we have

$$\Pr\left[|\{i \ : \ r_i \leq r\}| \leq \frac{N\ell}{p} + 2\sqrt{\ell}\right] < 1/e \ .$$

The result follows. □

5.4.3 The centered DGS algorithm

Theorem 5.4.6 ([Ste16a, Theorem 4.6]). *For any efficiently computable function $f(n)$ with $1 \le f(n) \le \mathrm{poly}(n)$, there is an (expected) polynomial-time reduction from (γ, ε)-cDGS to SVP, where $\varepsilon := 2^{-f(n)}$ and $\gamma := 1 + 1/f(n)$. The reduction preserves dimension and only calls the SVP oracle on sublattices of the input lattice.*

Proof. We assume without loss of generality that $s = 1$. (If $s \neq 1$, we can simply scale the lattice.) On input $\mathcal{L} \subset \mathbb{R}^n$, the reduction behaves as follows. First, it computes $\lambda_1(\mathcal{L})$ using its SVP oracle. For $i = 0, \ldots, \ell := \lceil 200n^2 f(n)^2 \rceil$, let $r_i := \sqrt{\lambda_1(\mathcal{L})^2 + i/(100nf(n))}$. For each i, the reduction uses its SVP oracle together with the procedure given in Theorem 5.4.5 to compute N_i such that

$$\gamma^{-1/10} \cdot \xi(\mathcal{L}, r_i) \le N_i \le \xi(\mathcal{L}, r_i) \,, \tag{5.3}$$

or $N_i := 1$ if $\lambda_1(\mathcal{L}) < r_i/(100n^2 f(n)\xi(\mathcal{L}, r_i))$. Let $w_\ell := \rho_{1/r_\ell}(\mathbb{Z} \setminus \{\mathbf{0}\})$, and for $i = 0, \ldots, \ell - 1$, let $w_i := \rho_{1/r_i}(\mathbb{Z} \setminus \{0\}) - \rho_{1/r_{i+1}}(\mathbb{Z} \setminus \{0\})$. (Claim 1.3.13 shows one way to compute w_i efficiently.)

Let $W := \sum_{i=0}^{\ell} N_i w_i$. Then, the reduction outputs $\mathbf{0}$ with probability $1/(1 + W)$. Otherwise, it chooses an index $0 \le k \le \ell$, assigning to each index i probability $N_i w_i / W$. If $N_k > 1$, the reduction then calls the procedure from Lemma 5.4.3 on input $\mathcal{L}$, r_k, and N_k, receiving as output a vector $\boldsymbol{x} \in \mathcal{L}_{\mathsf{prim}}$ that is distributed uniformly over $\mathcal{L}_{\mathsf{prim}} \cap r_k B_2^n$, up to a factor of $\gamma^{\pm 1/10}$. If $N_k = 1$, the reduction simply sets $\boldsymbol{x} = \mathrm{SVP}(\mathcal{L})$. Finally, it uses the procedure from Lemma 1.3.12 to sample an integer z from $D_{\mathbb{Z} \setminus \{0\}, 1/\|\boldsymbol{x}\|}$ and returns $\bar{\boldsymbol{x}} := z \cdot \boldsymbol{x}$. (Lemma 1.3.12 shows how to sample such an integer efficiently.)

First, we note that the reduction runs in expected polynomial time. In particular, the N_i have polynomial bit length by Corollary 5.2.2, and the various subprocedures have expected running times that are polynomial in the length of their input.

We now prove correctness. Let A be the set of all points that are integer multiples of a

lattice vector whose length is at most $r_\ell > \sqrt{nf(n)}$. By Theorem 1.3.4, it suffices to consider the distribution D_A, as this is within statistical distance ε of $D_{\mathcal{L}}$. Then,

$$\rho(A \setminus \{\mathbf{0}\}) = \sum_{\boldsymbol{y} \in A \setminus \{\mathbf{0}\}} \rho(\boldsymbol{y}) = \sum_{\boldsymbol{y} \in \mathcal{L}_{\mathsf{prim}} \cap \sqrt{n} B_2^n} \rho_{1/\|\boldsymbol{y}\|}(\mathbb{Z} \setminus \{0\}) \ .$$

A quick computation shows that for any $\boldsymbol{y}$ with $r_{i-1} \le \|\boldsymbol{y}\| \le r_i$, we have

$$\rho_{1/r_i}(\mathbb{Z} \setminus \{\mathbf{0}\}) \le \rho_{1/\|\boldsymbol{y}\|}(\mathbb{Z} \setminus \{0\}) \le \gamma^{1/10} \cdot \rho_{1/r_i}(\mathbb{Z} \setminus \{0\}) \ .$$

Recalling the definition of the w_i, it follows that

$$\sum_{i=0}^{\ell} \xi(\mathcal{L}, r_i) w_i \le \rho(A \setminus \{\mathbf{0}\}) \le \gamma^{1/10} \cdot \sum_{i=0}^{\ell} \xi(\mathcal{L}, r_i) w_i \ .$$

Now, we would like to say that $N_i \approx \xi(\mathcal{L}, r_i)$, as in Eq. (5.3). This is of course true by definition *except* when $N_i = 1$ and $\xi(\mathcal{L}, r_i) > 1$, i.e., when $\lambda_1(\mathcal{L}) < r_i/(100n^2 f(n)\xi(\mathcal{L}, r_i))$ and $\lambda_2(\mathcal{L}) \le r_i$. But, in this case, a quick computation together with Lemma 5.2.11 shows that $\xi(\mathcal{L}, r_{i+1}) > 1/(100nf(n)\lambda_1(\mathcal{L}))$, and therefore N_j satisfies Eq. (5.3) for all $j > i$. (In other words, the N_i can only be "wrong" for at most one value of i.) It follows that, for any $i < \ell$, we have

$$\gamma^{-1/5} \cdot \sum_{j \ge i} \xi(r_j, \mathcal{L}_j) w_j \le \sum_{j \ge i} N_j w_j \le \sum_{j \ge i} \xi(r_j, \mathcal{L}_j) w_j \ .$$

(The case $N_\ell = 1$ can be handled separately. Correctness in this case follows essentially immediately from Theorem 1.3.4.) Putting everything together, we have that

$$\gamma^{-1/5} \cdot \rho(A \setminus \{\mathbf{0}\}) \le W \le \gamma^{1/5} \cdot \rho(A \setminus \{\mathbf{0}\}) \ .$$

So, in particular, the probability that the reduction outputs $\mathbf{0}$ is $1/(1+W)$, which is a good

approximation to the correct probability of $1/\rho(A)$.

Now, for any $\boldsymbol{y} \in \mathcal{L}_{\mathsf{prim}}$, it follows from Lemma 5.4.3 and the argument above that

$$\gamma^{-1/2} \cdot \frac{\rho_{1/\|\boldsymbol{y}\|}(\mathbb{Z} \setminus \{0\})}{\rho(A)} \leq \Pr[\boldsymbol{x} = \pm\boldsymbol{y}] \leq \gamma^{1/2} \cdot \frac{\rho_{1/\|\boldsymbol{y}\|}(\mathbb{Z} \setminus \{0\})}{\rho(A)} \,. \tag{5.4}$$

Finally, for any $\boldsymbol{w} \in A \setminus \{0\}$, let $\boldsymbol{y}$ be one of the two primitive lattice vectors that are scalar multiples of $\boldsymbol{w}$, and let $\bar{z}$ such that $\boldsymbol{w} = \bar{z}\boldsymbol{y}$. Then,

$$\begin{aligned}\Pr[\bar{\boldsymbol{x}} = \boldsymbol{w}] &= \Pr[\boldsymbol{x} = \pm\boldsymbol{y}] \cdot \Pr[z = \bar{z} \mid \boldsymbol{x} = \pm\boldsymbol{y}] \\ &= \Pr[\boldsymbol{x} = \pm\boldsymbol{y}] \cdot \frac{\rho(\boldsymbol{w})}{\rho_{1/\|\boldsymbol{y}\|}(\mathbb{Z} \setminus \{0\})}\end{aligned}$$

The result follows from plugging the above equation into Eq. (5.4). □

5.5 $\sqrt{n/\log n}$-SVP to centered DGS reduction and a lower bound

It is an immediate consequence of Theorem 1.3.4 that $O(\sqrt{n})$-SVP reduces to DGS. In fact, we can do a bit better.

Proposition 5.5.1 ([Ste16a, Proposition 6.1])**.** *For any efficiently computable function $10 \leq f(n) \leq \mathrm{poly}(n)$, there is a polynomial-time reduction from γ-SVP to (f, ε)-DGS, where $\gamma := 10\sqrt{\frac{n}{\log f(n)}}$, and $\varepsilon := 1/f(n)$. The reduction only calls the oracle on the input lattice.*

Proof. We assume without loss of generality that n is large enough so that $f(n) < 2^{n-1}$. On input $\mathcal{L} \subset \mathbb{R}^n$, the reduction behaves as follows. It first uses the procedure from Theorem 1.2.3

to compute $\widetilde{d}$ such that $2^{-n/2} \cdot \lambda_1(\mathcal{L}) \leq \widetilde{d} \leq \lambda_1(\mathcal{L})$. For $i = 0, \ldots, 100n^3$, let

$$s_i := (1 + 1/n^2)^i \cdot \frac{\widetilde{d}}{\sqrt{\log f(n)}} .$$

The reduction calls the DGS oracle on input $\mathcal{L}$ and s_i for each i, $\lceil 100nf(n)^2 \rceil$ times. It then returns the shortest resulting non-zero vector.

It is clear that the reduction runs in polynomial time. Let i such that $s_{i-1} \leq 10\frac{\lambda_1(\mathcal{L})}{\sqrt{\log f(n)}} < s_i$. Note that

$$\Pr_{\boldsymbol{X} \sim D_{\mathcal{L},s_i}}[\boldsymbol{X} = \mathbf{0}] < \frac{1}{1 + 4/f(n)} < 1 - 2/f(n) .$$

By Theorem 1.3.4,

$$\Pr_{\boldsymbol{X} \sim D_{\mathcal{L},s_i}}\left[\|\boldsymbol{X}\| > \gamma \cdot \lambda_1(\mathcal{L})\right] \leq \Pr_{\boldsymbol{X} \sim D_{\mathcal{L},s_i}}[\|\boldsymbol{X}\| > s_i\sqrt{n}] < 2^{-n} .$$

Therefore, if the samples were truly from $D_{\mathcal{L},s_i}$, each would be a valid approximation with probability at least $2/f(n) - 2^{-n}$. It follows that each sample from the DGS oracle is a valid approximation with probability at least $1/f(n)^2 - 2^{-n}/f(n) > 1/(2f(n)^2)$, and the result follows. □

We now show a lower bound on the length of non-zero discrete Gaussian vectors. In particular, for any approximation factor $\gamma = o(\sqrt{n/\log n})$, we show a lattice (technically, a family of lattices indexed by the dimension n) such that the probability that $D_{\mathcal{L},s}$ yields a γ-approximate shortest vector is negligible for any s. This shows that any efficient reduction from γ-SVP to DGS with $\gamma = o(\sqrt{n/\log n})$ must output a vector not returned by the DGS oracle and/or make DGS calls on a lattice other than the input lattice.

Theorem 5.5.2 ([Ste16a, Theorem 6.2]). *For any sufficiently large n and $2 < t < \sqrt{n}/10$,*

there exists a lattice $\mathcal{L} \subset \mathbb{R}^n$ *with* $\lambda_1(\mathcal{L}) = t$ *such that for any* $s > 0$,

$$\Pr_{\boldsymbol{X} \sim D_{\mathcal{L},s}}[0 < \|\boldsymbol{X}\| \leq \sqrt{n}/10] < e^{-t^2} \ .$$

In particular, for any $t = \omega(\sqrt{\log n})$, $D_{\mathcal{L},s}$ *will yield a* $\sqrt{n}/(10t)$*-approximate shortest vector with at most negligible probability.*

Proof. Fix n. Let $\mathcal{L}' \subset \mathbb{R}^{n-1}$ be an $(n-1)$-dimensional lattice with $\rho_s(\mathcal{L}') \geq 1 + s^{n-1}$ and $\lambda_1(\mathcal{L}) > \sqrt{n-1}/10$, as promised by Lemma 5.2.3. Then, let $\mathcal{L} := \mathcal{L}' \oplus t\mathbb{Z}$ be the lattice obtained by "appending" a vector of length t to $\mathcal{L}'$. Note that the only vectors of length at most $\sqrt{n-1}/10$ in $\mathcal{L}$ are those that are multiples of the "appended" vector. So,

$$\Pr_{\boldsymbol{X} \sim D_{\mathcal{L},s}}[0 < \|\boldsymbol{X}\| \leq \sqrt{n-1}/10] \leq \frac{\rho_s(t\mathbb{Z} \setminus \{\mathbf{0}\})}{\rho_s(\mathcal{L}')} \leq \frac{\rho_{s/t}(\mathbb{Z} \setminus \{\mathbf{0}\})}{1 + s^{n-1}} \ .$$

Now, if $s \leq t$, then the numerator is less than e^{-t^2}. If $s > t$, then we have

$$\frac{\rho_{s/t}(\mathbb{Z} \setminus \{\mathbf{0}\})}{1 + s^{n-1}} < \frac{s}{1 + s^{n-1}} < \frac{1}{s^{n/2}} < \frac{1}{t^{n/2}} < e^{-t^2} \ ,$$

where we have used the fact that $\rho_{s'}(\mathbb{Z} \setminus \{0\}) < s'$, and the fact that $2 < t < \sqrt{n}/10$. □

Bibliography

[Aar14] Scott Aaronson. The equivalence of sampling and searching. *Theory of Computing Systems*, 55(2):281–298, 2014.

[ABSS93] Sanjeev Arora, László Babai, Jacques Stern, and Z Sweedyk. The hardness of approximate optima in lattices, codes, and systems of linear equations. In *FOCS*, 1993.

[ADPS16] Erdem Alkim, Léo Ducas, Thomas Pöppelmann, and Peter Schwabe. Post-quantum key exchange — A new hope. In *USENIX Security Symposium*, 2016.

[ADRS15] Divesh Aggarwal, Daniel Dadush, Oded Regev, and Noah Stephens-Davidowitz. Solving the Shortest Vector Problem in 2^n time via discrete Gaussian sampling. In *STOC*, 2015.

[ADS15] Divesh Aggarwal, Daniel Dadush, and Noah Stephens-Davidowitz. Solving the Closest Vector Problem in 2^n time— The discrete Gaussian strikes again! In *FOCS*, 2015.

[AJ08] Vikraman Arvind and Pushkar S Joglekar. Some sieving algorithms for lattice problems. In *FSTTCS*, pages 25–36, 2008.

[Ajt98] Miklós Ajtai. The shortest vector problem in ℓ_2 is NP-hard for randomized reductions. In *STOC*, 1998.

[Ajt04] Miklós Ajtai. Generating hard instances of lattice problems. In *Complexity of computations and proofs*, volume 13 of *Quad. Mat.*, pages 1–32. Dept. Math., Seconda Univ. Napoli, Caserta, 2004. Preliminary version in STOC'96.

[AKS01] Miklós Ajtai, Ravi Kumar, and D. Sivakumar. A sieve algorithm for the shortest lattice vector problem. In *STOC*, pages 601–610, 2001.

[AKS02] Miklós Ajtai, Ravi Kumar, and D. Sivakumar. Sampling short lattice vectors and the closest lattice vector problem. In *CCC*, pages 41–45, 2002.

[AR05] Dorit Aharonov and Oded Regev. Lattice problems in NP intersect coNP. *Journal of the ACM*, 52(5):749–765, 2005. Preliminary version in FOCS'04.

[AS64] Milton Abramowitz and Irene A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.

[Bab86] L. Babai. On Lovász' lattice reduction and the nearest lattice point problem. *Combinatorica*, 6(1):1–13, 1986.

[Ban93] Wojciech Banaszczyk. New bounds in some transference theorems in the geometry of numbers. *Mathematische Annalen*, 296(4):625–635, 1993.

[BCD+16] Joppe W. Bos, Craig Costello, Léo Ducas, Ilya Mironov, Michael Naehrig, Valeria Nikolaenko, Ananth Raghunathan, and Douglas Stebila. Frodo: Take off the ring! Practical, quantum-secure key exchange from LWE. In *CCS*, 2016.

[BDGL16] Anja Becker, Léo Ducas, Nicolas Gama, and Thijs Laarhoven. New directions in nearest neighbor searching with applications to lattice sieving. In *SODA*, 2016.

[BDS16] Huck Bennett, Daniel Dadush, and Noah Stephens-Davidowitz. On the Lattice Distortion Problem. In *ESA*, 2016.

[BGJ14] Anja Becker, Nicolas Gama, and Antoine Joux. A sieve algorithm based on overlattices. *LMS Journal of Computation and Mathematics*, 17(A):49–70, 2014.

[BGS17] Huck Bennett, Alexander Golovnev, and Noah Stephens-Davidowitz. On the quantitative hardness of CVP. In *FOCS*, 2017.

[BHW93] U. Betke, M. Henk, and J.M. Wills. Successive-minima-type inequalities. *Discrete & Computational Geometry*, 9(1):165–175, 1993.

[BLP+13] Zvika Brakerski, Adeline Langlois, Chris Peikert, Oded Regev, and Damien Stehlé. Classical hardness of learning with errors. In *STOC*, pages 575–584, 2013.

[BN09] Johannes Blömer and Stefanie Naewe. Sampling methods for shortest vectors, closest vectors and successive minima. *Theoret. Comput. Sci.*, 410(18):1648–1665, 2009.

[Bob11] Sergey G. Bobkov. On Milman's ellipsoids and M-position of convex bodies. In *Concentration, functional inequalities and isoperimetry*, volume 545 of *Contemp. Math.*, pages 23–33. Amer. Math. Soc., Providence, RI, 2011.

[Bou91] Jean Bourgain. On the distribution of polynomials on high-dimensional convex sets. In *Geometric aspects of functional analysis (1989–90)*, volume 1469 of *Lecture Notes in Math.*, pages 127–137. Springer, Berlin, 1991.

[BPY01] Philippe Biane, Jim Pitman, and Marc Yor. Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions. *Bull. Amer. Math. Soc. (N.S.)*, 38(4):435–465, 2001.

[Bri85] Ernest F. Brickell. Breaking iterated knapsacks. In *Advances in cryptology (Santa Barbara, Calif., 1984)*, volume 196 of *Lecture Notes in Comput. Sci.*, pages 342–358. Springer, Berlin, 1985.

[BS99] Johannes Blömer and Jean-Pierre Seifert. On the complexity of computing short linearly independent vectors and short bases in a lattice. In *STOC*. ACM, 1999.

[BSW16] Shi Bai, Damien Stehlé, and Weiqiang Wen. Improved reduction from the Bounded Distance Decoding Problem to the Unique Shortest Vector Problem in lattices. In *ICALP*, 2016.

[BV11] Zvika Brakerski and Vinod Vaikuntanathan. Efficient fully homomorphic encryption from (standard) LWE. In *FOCS*, 2011.

[BV14] Zvika Brakerski and Vinod Vaikuntanathan. Lattice-based FHE as secure as PKE. In *ITCS*, pages 1–12, 2014.

[Cas04] Bill Casselman. Stability of lattices and the partition of arithmetic quotients. *Asian J. Math.*, 8(4):607–637, 2004.

[CDLP13] Kai-Min Chung, Daniel Dadush, Feng-Hao Liu, and Chris Peikert. On the lattice smoothing parameter problem. In *CCC*, 2013.

[CFM04] Dario Cordero-Erausquin, Matthieu Fradelizi, and Bernard Maurey. The (B) conjecture for the Gaussian measure of dilates of symmetric convex sets and related problems. *J. Funct. Anal.*, 214(2):410–427, 2004.

[Chu76] Kai Lai Chung. Excursions in Brownian motion. *Ark. Mat.*, 14(2):155–177, 1976.

[CJL+92] Matthijs J Coster, Antoine Joux, Brian A LaMacchia, Andrew M Odlyzko, Claus-Peter Schnorr, and Jacques Stern. Improved low-density subset sum algorithms. *computational complexity*, 2(2):111–128, 1992.

[CN98] J-Y Cai and Ajay Nerurkar. Approximating the SVP to within a factor $(1+1/\dim^{\varepsilon})$ is NP-hard under randomized conditions. In *CCC*. IEEE, 1998.

[CS98] John Conway and Neil J.A. Sloane. *Sphere Packings, Lattices and Groups*. Springer New York, 1998.

[Dad12a] Daniel Dadush. Private communication, 2012.

[Dad12b] Daniel Dadush. *Integer Programming, Lattice Algorithms, and Deterministic Volume Estimation*. PhD thesis, Georgia Institute of Technology, 2012.

[DH11] Chandan Dubey and Thomas Holenstein. Approximating the closest vector problem using an approximate shortest vector oracle. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 184–193. Springer, 2011.

[DK13] Daniel Dadush and Gabor Kun. Lattice sparsification and the approximate Closest Vector Problem. In *SODA*, 2013.

[DKRS03] Irit Dinur, Guy Kindler, Ran Raz, and Shmuel Safra. Approximating CVP to within almost-polynomial factors is NP-hard. *Combinatorica*, 23(2):205–243, 2003.

[DPV11] Daniel Dadush, Chris Peikert, and Santosh Vempala. Enumerative lattice algorithms in any norm via M-ellipsoid coverings. In *FOCS*, 2011.

[DR16] Daniel Dadush and Oded Regev. Towards strong reverse Minkowski-type inequalities for lattices. In *FOCS*, 2016. http://arxiv.org/abs/1606.06913.

[DRS14] Daniel Dadush, Oded Regev, and Noah Stephens-Davidowitz. On the Closest Vector Problem with a distance guarantee. In *CCC*, 2014.

[DSV12] Mathieu Dutour Sikirić, Achill Schürmann, and Frank Vallentin. Inhomogeneous extreme forms. *Ann. Inst. Fourier (Grenoble)*, 62(6):2227–2255 (2013), 2012.

[FT79] T. Figiel and Nicole Tomczak-Jaegermann. Projections onto Hilbertian subspaces of Banach spaces. *Israel J. Math.*, 33(2):155–171, 1979.

[Gen09] Craig Gentry. Fully homomorphic encryption using ideal lattices. In *STOC*, 2009.

[GL87] P. M. Gruber and C. G. Lekkerkerker. *Geometry of numbers*, volume 37 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, second edition, 1987.

[Gly87] Peter W. Glynn. Upper bounds on Poisson tail probabilities. *Oper. Res. Lett.*, 6(1):9–14, 1987.

[GMR05] Venkatesan Guruswami, Daniele Micciancio, and Oded Regev. The complexity of the Covering Radius Problem. *Comput. Complex.*, 14(2):90–121, 2005.

[GMSS99] O. Goldreich, D. Micciancio, S. Safra, and J.-P. Seifert. Approximating shortest lattice vectors is not harder than approximating closest lattice vectors. *Information Processing Letters*, 71(2):55 – 61, 1999.

[GPV08] Craig Gentry, Chris Peikert, and Vinod Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions. In *STOC*, pages 197–206, 2008.

[Gra84] Daniel R. Grayson. Reduction theory using semistability. *Comment. Math. Helv.*, 59(4):600–634, 1984.

[Gru07] Peter M. Gruber. *Convex and discrete geometry*, volume 336 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Berlin, 2007.

[Hel85] Bettina Helfrich. Algorithms to construct Minkowski reduced and Hermite reduced lattice bases. *Theoret. Comput. Sci.*, 41(2-3):125–139 (1986), 1985.

[HLR09] Ishay Haviv, Vadim Lyubashevsky, and Oded Regev. A note on the distribution of the distance from a lattice. *Discrete & Computational Geometry*, 41(1):162–176, 2009.

[HN75] Günter Harder and Mudumbai S. Narasimhan. On the cohomology groups of moduli spaces of vector bundles on curves. *Mathematische Annalen*, 212(3):215–248, 1975.

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[HPS11] Guillaume Hanrot, Xavier Pujol, and Damien Stehlé. Algorithms for the Shortest and Closest Lattice Vector Problems. In *Coding and Cryptology*, pages 159–190. Springer, 2011.

[HR12] Ishay Haviv and Oded Regev. Tensor-based hardness of the Shortest Vector Problem to within almost polynomial factors. *Theory of Computing*, 8(23):513–531, 2012. Preliminary version in STOC'07.

[HR14] Ishay Haviv and Oded Regev. On the Lattice Isomorphism Problem. In *SODA*, 2014.

[HS07] Guillaume Hanrot and Damien Stehlé. Improved analysis of Kannan's shortest lattice vector algorithm (extended abstract). In *CRYPTO*, 2007.

[Jac28] C.G.J. Jacobi. Suite des notices sur les fonctions elliptiques. *Journal fr die reine und angewandte Mathematik*, 3:403–404, 1828.

[JS98] Antoine Joux and Jacques Stern. Lattice reduction: A toolbox for the cryptanalyst. *Journal of Cryptology*, 11(3):161–185, 1998.

[Kam16] Ohad Kammar. A note on Fréchet diffrentiation under Lebesgue integrals, 2016. Note available at https://www.cs.ox.ac.uk/people/ohad.kammar/notes/kammar-a-note-on-frechet-differentiation-under-lebesgue-integrals.pdf.

[Kan87] Ravi Kannan. Minkowski's convex body theorem and integer programming. *Math. Oper. Res.*, 12(3):415–440, 1987.

[Kho05] Subhash Khot. Hardness of approximating the Shortest Vector Problem in lattices. *Journal of the ACM*, 52(5):789–808, September 2005. Preliminary version in FOCS'04.

[KL78] G. A. Kabatjanskiĭ and V. I. Levenšteĭn. Bounds for packings on the sphere and in space. *Problemy Peredači Informacii*, 14(1):3–25, 1978.

[KL88] Ravi Kannan and László Lovász. Covering minima and lattice-point-free convex bodies. *Ann. of Math. (2)*, 128(3):577–602, 1988.

[Kla06] Bo'az Klartag. On convex perturbations with a bounded isotropic constant. *Geom. Funct. Anal.*, 16(6):1274–1290, 2006.

[Kle00] Philip Klein. Finding the closest lattice vector when it's unusually close. In *SODA*, 2000.

[Lat02] Rafał Latała. On some inequalities for Gaussian measures. In *Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002)*, pages 813–822. Higher Ed. Press, Beijing, 2002.

[Len83] H. W. Lenstra, Jr. Integer programming with a fixed number of variables. *Math. Oper. Res.*, 8(4):538–548, 1983.

[Lew79] D. R. Lewis. Ellipsoids defined by Banach ideal norms. *Mathematika*, 26(1):18–29, 1979.

[LLL82] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász. Factoring polynomials with rational coefficients. *Math. Ann.*, 261(4):515–534, 1982.

[LLM06] Yi-Kai Liu, Vadim Lyubashevsky, and Daniele Micciancio. On Bounded Distance Decoding for general lattices. In *RANDOM*, 2006.

[LLS90] J. C. Lagarias, H. W. Lenstra, Jr., and C.-P. Schnorr. Korkin-Zolotarev bases and successive minima of a lattice and its reciprocal lattice. *Combinatorica*, 10(4):333–348, 1990.

[LM83] Susan Landau and Gary Lee Miller. Solvability by radicals is in polynomial time. In *STOC*, 1983.

[LO85] J. C. Lagarias and A. M. Odlyzko. Solving low-density Subset Sum problems. *J. Assoc. Comput. Mach.*, 32(1):229–246, 1985.

[Mic01] Daniele Micciancio. The Shortest Vector Problem is NP-hard to approximate to within some constant. *SIAM Journal on Computing*, 30(6):2008–2035, March 2001. Preliminary version in FOCS 1998.

[Mic08] Daniele Micciancio. Efficient reductions among lattice problems. In *SODA*, 2008.

[Min10] Hermann Minkowski. *Geometrie der Zahlen*. B.G. Teubner, 1910.

[MO90] J. E. Mazo and A. M. Odlyzko. Lattice points in high-dimensional spheres. *Monatsh. Math.*, 110(1):47–61, 1990.

[MR07] Daniele Micciancio and Oded Regev. Worst-case to average-case reductions based on Gaussian measures. *SIAM Journal on Computing*, 37(1):267–302 (electronic), 2007.

[Mum07] David Mumford. *Tata lectures on theta. I.* Modern Birkhäuser Classics. Birkhäuser Boston, Inc., Boston, MA, 2007. With the collaboration of C. Musili, M. Nori, E. Previato and M. Stillman, Reprint of the 1983 edition.

[MV10] Daniele Micciancio and Panagiotis Voulgaris. Faster exponential time algorithms for the Shortest Vector Problem. In *SODA*, 2010.

[MV13] Daniele Micciancio and Panagiotis Voulgaris. A deterministic single exponential time algorithm for most lattice problems based on Voronoi cell computations. *SIAM Journal on Computing*, 42(3):1364–1391, 2013.

[MW15] Daniele Micciancio and Michael Walter. Fast lattice point enumeration with minimal overhead. In *SODA*, 2015.

[New76] Charles M. Newman. Fourier transforms with only real zeros. *Proc. Amer. Math. Soc.*, 61(2):245–251 (1977), 1976.

[NIS16] NIST post-quantum standardization call for proposals. http://csrc.nist.gov/groups/ST/post-quantum-crypto/cfp-announce-dec2016.html, 2016. Accessed: 2017-04-02.

[NS01] Phong Q Nguyen and Jacques Stern. The two faces of lattices in cryptology. In *Cryptography and lattices*, pages 146–180. Springer, 2001.

[NV08] Phong Q. Nguyen and Thomas Vidick. Sieve algorithms for the Shortest Vector Problem are practical. *J. Math. Cryptol.*, 2(2):181–207, 2008.

[Odl90] Andrew M Odlyzko. The rise and fall of knapsack cryptosystems. *Cryptology and computational number theory*, 42:75–88, 1990.

[Pei09] Chris Peikert. Public-key cryptosystems from the worst-case Shortest Vector Problem. In *STOC*, 2009.

[Pei10] Chris Peikert. An efficient and parallel Gaussian sampler for lattices. In *CRYPTO*. 2010.

[Pei16] Chris Peikert. A decade of lattice cryptography. *Foundations and Trends in Theoretical Computer Science*, 10(4):283–424, 2016.

[Per13] Yuval Peres, 2013. Personal communication.

[Pis82] Gilles Pisier. Holomorphic semigroups and the geometry of Banach spaces. *Ann. of Math. (2)*, 115(2):375–392, 1982.

[Pri14a] Thomas McMurray Price. Inequality regarding sum of Gaussian on lattices. MathOverflow, 2014. http://mathoverflow.net/q/160507 (version: 2014-12-01).

[Pri14b] Thomas McMurray Price. Is the heat kernel more spread out with a smaller metric? MathOverflow, 2014. http://mathoverflow.net/q/186428 (version: 2014-12-11).

[Pri15] Thomas McMurray Price. Numerical cohomology, 2015. http://arxiv.org/abs/1509.05797.

[Pri16] Thomas McMurray Price. An inequality for the heat kernel on an Abelian Cayley graph, 2016. https://arxiv.org/abs/1612.07306.

[PRS17] Chris Peikert, Oded Regev, and Noah Stephens-Davidowitz. Pseudorandomness of Ring-LWE for any ring and modulus. In *STOC*, 2017.

[PS09] Xavier Pujol and Damien Stehlé. Solving the Shortest Lattice Vector Problem in time $2^{2.465n}$. *IACR Cryptology ePrint Archive*, 2009:605, 2009.

[Reg09] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *Journal of the ACM*, 56(6):Art. 34, 40, 2009.

[Rie57] Bernhard Riemann. Theorie der Abel'schen Functionen. *Journal für die reine und angewandte Mathematik*, 54:101–155, 1857.

[Rob55] Herbert Robbins. A remark on stirling's formula. *The American Mathematical Monthly*, 62(1):26–29, 1955.

[Rog55] C. A. Rogers. Mean values over the space of lattices. *Acta Math.*, 94:249–287, 1955.

[Ros41] Barkley Rosser. Explicit bounds for some functions of prime numbers. *American Journal of Mathematics*, 63(1):pp. 211–232, 1941.

[Roy14] Thomas Royen. A simple proof of the Gaussian correlation conjecture extended to some multivariate gamma distributions. *Far East J. Theor. Stat.*, 48(2):139–145, 2014.

[RS16] Oded Regev and Igor Shinkar. A counterexample to monotonicity of relative mass in random walks. *Electronic Communications in Probability*, 2016.

[RS17a] Oded Regev and Noah Stephens-Davidowitz. An inequality for Gaussians on lattices. *SIDMA*, 2017.

[RS17b] Oded Regev and Noah Stephens-Davidowitz. A reverse Minkowski theorem. In *STOC*, 2017.

[Sch60] Wolfgang M. Schmidt. A metrical theorem in geometry of numbers. *Trans. Amer. Math. Soc.*, 95:516–529, 1960.

[Sch87] C.P. Schnorr. A hierarchy of polynomial time lattice basis reduction algorithms. *Theoretical Computer Science*, 53(23):201 – 224, 1987.

[SFS09] Naftali Sommer, Meir Feder, and Ofir Shalvi. Finding the closest lattice point by iterative slicing. *SIAM J. Discrete Math.*, 23(2):715–731, 2009.

[Sha84] Adi Shamir. A polynomial-time algorithm for breaking the basic Merkle-Hellman cryptosystem. *IEEE Trans. Inform. Theory*, 30(5):699–704, 1984.

[Sie45] Carl Ludwig Siegel. A mean value theorem in geometry of numbers. *Ann. of Math. (2)*, 46:340–347, 1945.

[Sol16] Omri N. Solan. Intersections of diagonal orbits, 2016. https://arxiv.org/abs/1612.08765.

[SS06] Peter Sarnak and Andreas Strömbergsson. Minima of Epstein's zeta function and heights of flat tori. *Invent. Math.*, 165(1):115–151, 2006.

[Ste16a] Noah Stephens-Davidowitz. Discrete Gaussian sampling reduces to CVP and SVP. In *SODA*, 2016.

[Ste16b] Noah Stephens-Davidowitz. Search-to-decision reductions for lattice problems with approximation factors (slightly) greater than one. In *APPROX/RANDOM*, 2016.

[Stu76] Ulrich Stuhler. Eine Bemerkung zur Reduktionstheorie quadratischer Formen. *Arch. Math. (Basel)*, 27(6):604–610, 1976.

[SW14] Uri Shapira and Barak Weiss. A volume estimate for the set of stable lattices. *C. R. Math. Acad. Sci. Paris*, 352(11):875–879, 2014.

[SW16] Uri Shapira and Barak Weiss. Stable lattices and the diagonal group. *J. Eur. Math. Soc. (JEMS)*, 18(8):1753–1767, 2016.

[Ter16] Audrey Terras. *Harmonic analysis on symmetric spaces—higher rank spaces, positive definite matrix space and generalizations*. Springer, New York, second edition, 2016.

[vdC36] Johannes van der Corput. Verallgemeinerung einer Mordellschen Beweismethode in der Geometrie der Zahlen, Zweite Mitteilung. *Acta Arithmetica*, 2(1):145–146, 1936.

[ZF96] Ram Zamir and Meir Feder. On lattice quantization noise. *IEEE Transactions on Information Theory*, 42(4):1152–1159, 1996.