Electronic Colloquium on Computational Complexity, Report No. 6 (1999)

Some Recent Progress on the Complexity of Lattice Problems

Jin-Yi Cai*
Department of Computer Science and Engineering
State University of New York
Buffalo, NY 14260. USA.
cai@cse.buffalo.edu

Abstract

We survey some recent developments in the study of the complexity of lattice problems. After a discussion of some problems on lattices which can be algorithmically solved efficiently, our main focus is the recent progress on complexity results of intractability. We will discuss Ajtai's worst-case/average-case connections, NP-hardness and non-NP-hardness, transference theorems between primal and dual lattices, and the Ajtai-Dwork cryptosystem.

1 Introduction

There have been some exciting developments recently concerning the complexity of lattice problems. Research in the algorithmic aspects of lattice problems has been active in the past, especially following Lovász's basis reduction algorithm in 1982. The recent wave of activity and interest can be traced in large part to two seminal papers written by Miklós Ajtai in 1996 and in 1997 respectively.

In his 1996 paper [1], Ajtai found a remarkable worst-case to average-case reduction for some versions of the shortest lattice vector problem (SVP), thereby establishing a worst-case to average-case connection for these lattice problems. Such a connection is not known to hold for any other problem in NP believed to be outside P. In his 1997 paper [2], building on previous work by Adleman, Ajtai further proved the NP-hardness of SVP, under randomized reduction. The NP-hardness of SVP has been a long standing open problem. Stimulated by these breakthroughs, many researchers have obtained new and interesting results for these and other lattice problems [3, 11, 13, 14, 15, 16, 17, 18, 19, 23, 30, 31, 32, 33, 34, 52, 55, 57]. Our purpose in this article is to survey some of this development.

I think these lattice problems are intrinsically interesting. Moreover, the worst-case to average-case connection discovered by Ajtai also opens up possibilities regarding provably secure public-key cryptography based on only worst-case intractability assumptions. It is well known that the existence of secure public-key cryptosystems presupposes P $\neq$ NP. However the converse is far from being proven true.[1] The intractability required by cryptography is more concerned with average-case complexity rather than worst-case complexity. Even if we assume that some problem in NP is not solvable in P or BPP, this still leaves open the possibility that the problem might be rather easy on the average.

Consider the security of RSA and the intractability of factoring. First, we do not know if factoring is not solvable in P or BPP. We do not know if this is so assuming P $\neq$ NP. We do not even know whether it is NP-hard. Second, even if we assume it is NP-hard or not solvable in P or BPP, we do not know it is as hard for the special case of factoring a product of two large primes $p \cdot q$. Third, even if factoring $p \cdot q$ is hard in the worst case, we do not know if it is hard on the average, under some reasonable distribution on such numbers. Fourth, we do not know if decrypting RSA without the private key is equivalent to finding $\varphi(pq) = (p-1)(q-1)$, (although given $n = p \cdot q$, finding $\varphi(pq)$ is equivalent to factoring). Thus although RSA is believed to be an excellent public-key cryptosystem, there is a large gap between the assumption that factoring is hard in the worst-case (say it is not in BPP) and a proof that the system is secure.

Building on Ajtai's worst-case to average-case connection, Ajtai and Dwork [3] proposed a public-key cryptosystem that is provably secure, assuming only the worst case intractability of a certain version of SVP, namely to find the shortest lattice vector in a lattice with n^c-unique shortest vector, for a sufficiently large c. This is the first time that such a provable security guarantee based on the worst-case complexity alone has been established.

In Section 2 we collect some definitions. After that, I will

*Research supported in part by grants from NSF CCR-9634665 and a John Simon Guggenheim Fellowship.

[1] I do not want to say "the converse is false", since it is probably *true* for the reason that both P $\neq$NP and there exist secure public-key cryptosystems. But it is believed that it is insufficient to assume only P $\neq$NP in order to prove pseudorandom number generators exist.

ISSN 1433-8092

first discuss what is algorithmically computable efficiently for some lattice problems (Section 3), then I will discuss Ajtai's worst-case/average-case connection (Section 4), NP-hardness results (Section 5), evidence of non-NP-hardness (Section 6), transference theorems relating primal and dual lattices (Section 7), and the Ajati-Dwork cryptosystem (Section 8).

The selection of the topics is highly subjective and it reflects my limited knowledge and personal taste. They are also restrained by the space limitation. I am sure many important works have been neglected or not given its proper due. I apologize for any such omissions or mistakes.

2 Preliminaries

A lattice is a discrete additive subgroup in some $\mathbf{R}^n$. Discreteness means that every lattice point is an isolated point in the topology of $\mathbf{R}^n$. An alternative definition is that a lattice consists of all the integral linear combinations of a set of linearly independent vectors,

$$L = \{\sum_i n_i b_i \mid n_i \in \mathbf{Z}, \text{ for all } i\},$$

where the vectors b_i's are linearly independent over $\mathbf{R}$. Such a set of generating vectors are called a basis. The dimension of the linear span, or equivalently the number of b_i's in a basis is the rank (or dimension) of the lattice, and is denoted by $\dim L$. We may without loss of generality assume that $\dim L = n$, for otherwise we can replace $\mathbf{R}^n$ by its linear span. We denote L as $L(b_1, b_2, \ldots, b_n)$.

The basis of a lattice is not unique. Any two bases are related to each other by an integral matrix of determinant ± 1. Such a matrix is called a unimodular matrix. Clearly an integral matrix has an integral inverse iff it is unimodular, following Cramer's rule.

The parallelepiped

$$P(b_1, \ldots, b_n) = \{\sum x_i b_i \mid 0 \leq x_i < 1\}$$

is called the fundamental domain of the lattice.

Since basis transformation is unimodular, the determinant $|\det(b_1, \ldots, b_n)|$ which is the volume of the fundamental domain $P(b_1, \ldots, b_n)$ is independent of the basis, and is denoted by $\det(L)$.

We use lsp to denote linear span over $\mathbf{R}$. Given a basis $\{b_1, b_2, \ldots, b_n\}$ of L, let $\Pi_i = \text{lsp}\{b_1, \ldots, b_i\}$ be the linear span of $\{b_1, \ldots, b_i\}$, and $L_i = L(b_1, \ldots, b_i)$ be the sublattice generated by $\{b_1, \ldots, b_i\}$. We denote by $\Pi_i^\perp$ the orthogonal complement of Π_i. The process of Gram-Schmidt orthogonalization obtains from a basis $\{b_1, b_2, \ldots, b_n\}$ a set of orthogonal vectors $\{\widehat{b}_1, \widehat{b}_2, \ldots, \widehat{b}_n\}$, where $\widehat{b}_i$ is the orthogonal component of b_i perpendicular to Π_{i-1}:

$$\widehat{b}_i = b_i - \sum_{j<i} \frac{\langle b_i, \widehat{b}_j \rangle}{\langle \widehat{b}_j, \widehat{b}_j \rangle} \widehat{b}_j, \quad 1 \leq i \leq n,$$

where $\langle \cdot, \cdot \rangle$ denotes inner product.

The fundamental domain as well as the orthogonal "brick" $P(\widehat{b}_1, \ldots, \widehat{b}_n) = [0, \widehat{b}_1) \times \cdots \times [0, \widehat{b}_n)$ form a tessellation of $\mathbf{R}^n$ by translation. We can also tessellate $\mathbf{R}^n$ by the centralized "brick" $B = [-\frac{\widehat{b}_1}{2}, \frac{\widehat{b}_1}{2}) \times \cdots \times [-\frac{\widehat{b}_{i-1}}{2}, \frac{\widehat{b}_{i-1}}{2})$:

$$\mathbf{R}^n = \bigcup_{\ell \in L} (\ell + B).$$

We note that the volume vol $D =$ vol $B = \det L$.

The length of the shortest non-zero vector of L is denoted by $\lambda_1(L)$. In general, Minkowski's *successive minima* $\lambda_i(L)$ are defined as follows: for $1 \leq i \leq \dim L$,

$$\lambda_i(L) = \min_{v_1, \ldots, v_i \in L} \max_{1 \leq j \leq i} ||v_j||,$$

where the sequence of vectors $v_1, \ldots, v_i \in L$ ranges over all i linearly independent lattice vectors. It is not difficult to show that to get $v_i \in L$ with $||v_i|| = \lambda_i$, one can always take greedily *any* linearly independent $v_1, \ldots, v_{i-1} \in L$, with $||v_1|| = \lambda_1, \ldots, ||v_{i-1}|| = \lambda_{i-1}$.

We denote by $\text{bl}(L)$ the basis length of L

$$\text{bl}(L) = \min_{\text{all bases } b_1, \ldots, b_n \text{ for } L} \max_{i=1}^{n} ||b_i||.$$

The dual lattice L^* of a lattice L of dimension n in $\mathbf{R}^n$ is defined as those vectors $u \in \mathbf{R}^n$, such that $\langle u, v \rangle \in \mathbf{Z}$, for all $v \in L$. For a basis $\{b_1, b_2, \ldots, b_n\}$ of L, its dual basis is $\{b_1^*, b_2^*, \ldots, b_n^*\}$, where $\langle b_i^*, b_j \rangle = \delta_{ij}$. Then $L^* = L(b_1^*, b_2^*, \ldots, b_n^*)$. In particular $\det(L^*) = 1/\det(L)$, and $L^{**} = L$. For a lattice with dimension less than n, its dual is defined within its own linear span.

We let $kL = \{kv \mid v \in L\}$ be the dilatation of L for any positive $k \in \mathbf{R}$. Let $x + A = \{x + y \mid y \in A\}$ for any $x \in \mathbf{R}^n$ and $A \subseteq \mathbf{R}^n$. Let $A + B = \{a + b \mid a \in A, b \in B\}$. We denote by $\lfloor x \rfloor$ the greatest integer $\leq x$, $\lceil x \rceil$ the least integer $\geq x$, $\lceil x \rceil = -\lfloor -x \rfloor$, and $\lceil x \rfloor$ the closest integer to x, $\lceil x \rfloor = \lfloor x + \frac{1}{2} \rfloor$.

3 From Gauss to Lovász

Before we discuss intractability results on lattice problems, let us first take a look at what is algorithmically feasible. In this section we give a brief account of the motivations for the study of lattice problems, some ramifications, and the main ideas of the basis reduction algorithm of Lovász.

We should start with Gauss. The original motivation for the study of 2-dimensional lattices came from the theory of quadratic forms in number theory, which culminated in the Theory of Genus and Composition by Gauss (see e.g., [28, 22, 21]).

Gauss gave an algorithm which completely solved the classification problem of 2-dimensional lattices. The algorithm can be viewed as a 2-dimensional generalization of a version of the Euclidean algorithm, the Centralized Euclidean Algorithm (CEA). In this CEA, given two integers n and m, suppose $0 < |m| \leq |n|$, we divide n by m with a quotient q and a remainder r, such that $|r| \leq \frac{1}{2}|m|$. Thus $n = qm + r$ and r is as small as possible. If $r \neq 0$ we repeat with the substitution $n \leftarrow m, m \leftarrow r$, until the remainder is zero.

Given a 2-dimensional lattice generated by u and v. Suppose $||v|| \leq ||u||$. Gauss' algorithm "slides" u against v, i.e., it finds an integral multiple qv so that $u' = u - qv$ is as short as possible. Clearly this is the case precisely when the orthogonal projection of u' onto v is as small as possible in absolute value, and it can always be made $\leq \frac{1}{2}||v||$. This is quite obvious geometrically. Numerically, $q = \lceil \langle \frac{u}{||v||}, \frac{v}{||v||} \rangle \rfloor$ will do. In a possible tie when $\langle u - qv, \frac{v}{||v||} \rangle = \frac{1}{2}||v||$, and $\langle u - (q+1)v, \frac{v}{||v||} \rangle = -\frac{1}{2}||v||$, we can break the tie arbitrarily. Gauss' algorithm terminates if $||u'|| \geq ||v||$. Otherwise, we switch the role of u and v with the substitution $u \leftarrow v$ and $v \leftarrow u'$ and continue.

It is not difficult to show that, like CEA, Gauss' algorithm terminates in polynomial time. In fact the number of iterations is at most linear in the number of bits in the length of u and v. Moreover the precise constant in the linear rate has been determined. A worst-case bound of both $\log_{1+\sqrt{2}} M$ are given by Dupré [25] and Vallée [61], where $M = \max\{||u||, ||v||\}$.

While the Euclidean algorithm can be viewed as an algorithm for the one-dimensional lattices (generated by the two integers n and m), Gauss' algorithm finds a reduced basis for any 2-dimensional lattice, which is essentially unique. Suppose the algorithm terminates with the vectors u_0 and v_0, with $||v_0|| \leq ||u_0||$. If v_0 is scaled to unity 1 and u_0 to the upper half plane (in terms of the complex plane **C**, we apply $z \mapsto z/v_0$ or $\overline{(z/v_0)}$, as a dilatation and rotation with possibly a reflection), then u_0 is mapped to a point in the so-called *fundamental region* $\mathcal{R}$ of the upper half plane as in Figure 1. Thus up to a scaling factor the fundamental region $\mathcal{R}$ (with a suitable identification of its boundary points) is in 1-1 correspondence with the space of all 2-dimensional lattices. Of course the upper half plane with the tessellation in Figure 2, induced by the action of the unimodular group $SL_2(\mathbf{Z})$ is endowed with a hyperbolic metric. This then can be used to introduce a metric on the space of 2-dimensional lattices. The actions of $SL_2(\mathbf{Z})$ and its subgroups in the upper half plane is the starting point of a rich interplay between hyperbolic geometry, elliptic curves and modular forms [5, 42]. We will, however, leave the world of 2-dimensional lattices for higher dimensions.

The reduction theory of 2-dimensional lattices extends to 3-dimensional lattices without much difficulty. Perhaps the first indication that something non-trivial happens in higher dimensions came with a discovery by Korkin and Zolotarev [43] on shortest vectors. Originally their result is concerned with quadratic forms; we will instead present an example in the same spirit directly in terms of lattices.

Consider the lattice L generated by e_i together with $h = (\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2})$, where, $1 \leq i \leq n$, and e_i has a single 1 in the ith coordinate and 0 elsewhere. We note that $\{h, e_2, \ldots, e_n\}$ is a basis for L, for $e_1 = 2h - \sum_{i=2}^{n} e_i$. Meanwhile, $\{e_1, e_2, \ldots, e_n\}$ is not a basis for L, for h does not belong to $\mathbf{Z}^n$ which is the sublattice generated by $\{e_1, e_2, \ldots, e_n\}$. $\lambda_1(L) = \ldots = \lambda_n(L) = 1$, since they are achieved by $\{e_1, \ldots, e_n\}$. For $n > 4$, then, the shortest n linearly independent lattice vectors do not form a basis, which is rather unintuitive. The shortest basis length $\mathrm{bl}(L) = \sqrt{n}/2$.

Let L be an n-dimensional lattice in $\mathbf{R}^n$ with basis $\{b_1, b_2, \ldots, b_n\}$. Since the translations of the fundamental domain $D = P(b_1, b_2, \ldots, b_n)$ form a tiling of $\mathbf{R}^n$, the volume $\mathrm{vol}(D) = \det(L)$ provides a certain measure of the size of L. Minkowski's First Theorem makes an explicit connection of the shortest lattice vector and this quantity [54, 21, 36]:

Theorem 3.1 (Minkowski)

$$\lambda_1(L) \leq \gamma_n (\det(L))^{1/n},$$

where γ_n is some universal constant.

The smallest such constant for dimension n is denoted by γ_n and called Hermite's constant of rank n. Minkowski proved that $\gamma_n \leq \frac{2}{\sqrt{\pi}} \Gamma(\frac{n}{2} + 1)^{1/n}$, which is asymptotically $\sqrt{\frac{2n}{\pi e}}$. It is known that $\sqrt{\frac{n}{2\pi e}} \leq \gamma_n \leq \sqrt{\frac{n}{\pi e}}$. The upshot is, for a lattice with $\det(L) = 1$, (after a suitable scaling), there is always a non-zero short vector of length no more than $\sqrt{n}$.

Minkowski's First Theorem has a short and elegant proof: Consider the lattice $L' = 2L$, which is a dilatation of L by a factor of 2 in all directions. $\det(L') = 2^n \det(L)$. Consider a ball of radius r centered at every lattice point of L'. Let ω_n denote the volume of a unit ball B_n, then $\omega_n r^n$ is the volume of a ball $B_n(r)$ of radius r. Now if $\omega_n r^n > \det(L')$, there must be some overlap among two different balls, thus $\exists \ell \neq \ell'$ both $\in L$, such that $2\ell + x = 2\ell' + y$ for some $x, y \in B_n(r)$. Then $\ell - \ell' = (y - x)/2 \in B_n(r)$ by convexity. And $\ell - \ell'$ is our non-zero lattice point of L. It is known that $\omega_n = \pi^{n/2}/\Gamma(\frac{n}{2}+1)$. It follows that

$$\lambda_1(L) \leq \frac{2}{\sqrt{\pi}}\Gamma(\frac{n}{2}+1)^{1/n}(\det(L))^{1/n} = \Theta(\sqrt{n})(\det(L))^{1/n}.$$

Theorem 3.1 follows.

A more general theorem, also due to Minkowski, is concerned with successive minima:

Theorem 3.2 (Minkowski)

$$\left(\prod_{i=1}^{n} \lambda_i(L)\right)^{1/n} \leq \Theta(\sqrt{n})(\det(L))^{1/n}.$$

While Minkowski's theorem guarantees the existence of vectors as short as $\sqrt{n}\det(L)^{1/n}$, there is no polynomial-time algorithm to find such a vector. Minkowski's proof is decidedly non-constructive. The Shortest Vector Problem (SVP) is the following: Given a basis of L, find a vector $v \in L$ such that $||v|| = \lambda_1(L)$. One can also define various approximate short vector problems, seeking a non-zero $v \in L$ with $||v||$ bounded by some approximation factor, $||v|| \leq f(n)\lambda_1(L)$ or $||v|| \leq f(n)(\det(L))^{1/n}$.

The celebrated Lovász basis reduction algorithm is one such algorithm that finds some approximate short vector for any lattice in dimension n. This algorithm has proven to be widely applicable, so that it forms a benchmark against which claims of intractability has to be measured.

Theorem 3.3 (Lovász) *Given any basis $\{b_1, \ldots, b_n\}$ of a lattice, Lovász's basis reduction algorithm finds a new basis $\{b'_1, \ldots, b'_n\}$, such that*

(i) $||b'_1|| \leq 2^{\frac{n-1}{2}}\lambda_1(L)$;

(ii) $||b'_1|| \leq 2^{\frac{n-1}{4}}\sqrt[n]{\det(L)}$;

(iii) $||b'_1|| \cdots ||b'_n|| \leq 2^{\frac{1}{2}\binom{n}{2}}\det(L)$.

We will sketch this algorithm. Given a basis $\{b_1, \ldots, b_n\}$, we consider the "brick tiling" of $\mathbf{R}^n$ induced by the Gram-Schmidt orthogonalization $\{\widehat{b}_1, \ldots, \widehat{b}_n\}$. Recall that $\widehat{b}_1 = b_1$, $\widehat{b}_2 = b_2 - \frac{\langle b_2, \widehat{b}_1\rangle}{\langle \widehat{b}_1, \widehat{b}_1\rangle}\widehat{b}_1$, e.t.c. We may "slide" b_2 against $b_1 = \widehat{b}_1$, i.e., replace b_2 by $b_2 - qb_1$ so that we can assume that $|\frac{\langle b_2, \widehat{b}_1\rangle}{\langle \widehat{b}_1, \widehat{b}_1\rangle}| \leq \frac{1}{2}$. In general we want to "slide" b_i against $b_1, \ldots, b_{i-1}$, so that $\widehat{b}_i = b_i - \sum_{k<i}\mu_{ik}\widehat{b}_k$, with all $|\mu_{ik}| \leq 1/2$, for all $k < i$. Suppose we have taken care of all $b_1, b_2, \ldots, b_{i-1}$. Consider the orthogonal projection of b_i to the linear span Π_{i-1} of $b_1, \ldots, b_{i-1}$. We can "slide" b_i against $b_{i-1}, \ldots, b_1$, in that order, so that the projection of b_i lies in the orthogonal box $[-\frac{\widehat{b}_1}{2}, \frac{\widehat{b}_1}{2}] \times \cdots \times [-\frac{\widehat{b}_{i-1}}{2}, \frac{\widehat{b}_{i-1}}{2}]$. The following steps are natural. We can replace b_i by $b_i - \lceil \mu_{i,i-1} \rfloor b_{i-1}$, which can be expressed as $b_i - \lceil \mu_{i,i-1} \rfloor \widehat{b}_{i-1} + \sum_{k<i-1}\nu_k\widehat{b}_k$. We then repeat this for $\mu_{i,i-2}, \ldots, \mu_{i,1}$, in that order. Note that for $k < j$, "sliding" against b_k later will not change any previous $\mu_{i,j}$ which has already been made to have absolute value at most $1/2$. Thus we finally have made all $|\mu_{ik}| \leq 1/2$, for $k < i$. Such a basis is called weakly reduced, and can be achieved in polynomial time.

Geometrically these steps are rather obvious and unremarkable. What makes Lovász's algorithm remarkable is the following requirement which is best visualized in a faux 3-dimensional picture as in Figure 3. Suppose we have a weakly reduced basis $\{b_1, \ldots, b_n\}$. Consider the linear span Π_{i-1} of $\{b_1, \ldots, b_{i-1}\}$. Let π_{i-1} be the orthogonal projection to Π_{i-1}. Let $v(i) = v - \pi_{i-1}(v)$ be the orthogonal component of v perpendicular to Π_{i-1}. A basis satisfies the following condition is called Lovász reduced

$$||b_i(i)|| \leq \frac{2}{\sqrt{3}}||b_{i+1}(i)||, \qquad \text{for all } 1 \leq i \leq n.$$

Some explanations are in order. Note that for i, everything happens in $\Pi_{i+1} = \text{lsp}\{b_1, \ldots, b_{i+1}\}$, which is also the linear span of Π_{i-1} and $\{b_i(i), b_{i+1}(i)\}$. In $\text{lsp}\{b_i(i), b_{i+1}(i)\}$ there is a 2-dimensional lattice $L(b_i(i), b_{i+1}(i))$, which is the orthogonal projection of $L(b_1, \ldots, b_{i+1})$ along $\Pi_{i-1}^{\perp}$. Thus it is natural to perform Gauss' 2-dimensional lattice basis reduction on $L(b_i(i), b_{i+1}(i))$. Note that the Gaussian steps on $b_i(i)$ and $b_{i+1}(i)$ can be easily lifted to be performed on the pair b_i and b_{i+1}.

Since our basis is already weakly reduced, it is easy to see that the only Gaussian step that is possibly applicable is to swap $b_i(i)$ and $b_{i+1}(i)$, *if* $||b_{i+1}(i)|| < ||b_i(i)||$. This should ideally be performed, had it not been for the desire that this procedure be guaranteed to terminate quickly. Thus for efficiency considerations we swap b_i and b_{i+1} only when $||b_{i+1}(i)|| < \frac{\sqrt{3}}{2}||b_i(i)||$. Thus, we can show that when a swap takes place, a significant gain is realized. We note that after the swap, $b_{i+1}(i)$ is the new $b_i(i)$, and the previous $b_i(i)$ is the new $b_{i+1}(i)$, and hence it satisfies the Lovász condition at i. (The constant $\frac{2}{\sqrt{3}}$ is just for convenience; it can be replaced by any other constant between 1 and 1.5.)

The Lovász basis reduction algorithm then consists of the following steps being alternated. Step (I): Achieve weakly reducedness. Step (II): If there is any i violating Lovász's condition then swap b_i and b_{i+1}.

The proof of convergence relies on the potential function $D = D(b_1, \ldots, b_n) = \prod_{i=1}^{n}||\widehat{b}_i||^{n-i}$. Note that $\det(L) = \prod_{i=1}^{n}||\widehat{b}_i||$, $\det(L(b_1, \ldots, b_i)) = \prod_{k=1}^{i}||\widehat{b}_k||$, and $D \cdot \det(L) = \prod_{i=1}^{n}(\prod_{k=1}^{i} ||\widehat{b}_k||)$. Since Step (I) preserves each $\det(L(b_1, \ldots, b_i))$, D is invariant under Step (I). What happens to D under Step (II) swapping b_i and b_{i+1}? Clearly Π_{i-1} is unaffected, so $\widehat{b}_1, \ldots, \widehat{b}_{i-1}$ are the same. Since everything happens in Π_{i+1}, $\widehat{b}_{i+2}, \ldots, \widehat{b}_n$ are also the same. Let $a = b_i(i)$ and $b = b_{i+1}(i)$. Let the angle between a and b be θ. Then the current $\widehat{b}_i = a$ and the current $||\widehat{b}_{i+1}|| = ||b|| \sin\theta$. After the swap the updated $\widehat{b}_i = b$ and $||\widehat{b}_{i+1}|| = ||a|| \sin\theta$. Hence $D_{\text{new}}/D_{\text{old}} = ||b||/||a|| \leq \sqrt{3}/2$. Clearly the initial $D \leq (\max ||b_i||)^{\binom{n}{2}}$, and for any integral lattice $\det(L(b_1, \ldots, b_i)) \geq 1$, thus D is always at least one. It follows that Lovász's algorithm terminates in polynomial time. (A slight extension of the argument handles the rational case. For more details on this and the issue of bit size, see [47].) Once the algorithm terminates, we have

$$\begin{aligned} ||\widehat{b}_i||^2 = ||b_i(i)||^2 &\leq \frac{4}{3}||b_{i+1}(i)||^2 \\ &= \frac{4}{3}\left[||\widehat{b}_{i+1}||^2 + \mu_{i+1,i}^2||\widehat{b}_i||^2\right] \\ &\leq \frac{4}{3}||\widehat{b}_{i+1}||^2 + \frac{1}{3}||\widehat{b}_i||^2. \end{aligned}$$

It follows that $||\widehat{b}_{i+1}||^2 \geq \frac{1}{2}||\widehat{b}_i||^2$. By induction $||\widehat{b}_i||^2 \geq \frac{1}{2^{i-1}}||\widehat{b}_1||^2 = \frac{1}{2^{i-1}}||b_1||^2$.

Let any $v = \sum n_i b_i \in L$. Suppose $v \neq 0$. Then not all $n_i = 0$, and let j be the maximum such i. Then $v = \sum_{i \leq j} n_i b_i = n_j \widehat{b}_j + \sum_{i<j} \xi_i \widehat{b}_i$, and by orthogonality $||v|| \geq |n_j| ||\widehat{b}_j|| \geq ||\widehat{b}_j||$ since n_j is integral. In particular $\lambda_1(L) \geq \min_i ||\widehat{b}_i||$. It follows that (i) $||b_1|| \leq 2^{\frac{n-1}{2}} \lambda_1(L)$. Similarly $||b_1||^{2n} \leq 2^{n(n-1)} \prod_{i=1}^{n} ||\widehat{b}_i||^2 = 2^{\binom{n}{2}}(\det(L))^2$. Thus (ii) follows. (iii) also follows similarly.

The bound $2^{\frac{n-1}{2}}$ can be improved to $(1+\epsilon)^n$ for any fixed $\epsilon > 0$, within polynomial time. (The polynomial of course depends on ϵ.) This is due to Schnorr [58] and is accomplished by a k-dimensional variant of Lovász's reduction, for some large constant k.

The main application of Lovász's algorithm originally in [47] was a solution to a centuries old problem: How to factor a polynomial into irreducible polynomials over the rationals **Q**. The LLL algorithm has had a tremendous impact in the field. Another celebrated result is Lenstra's polynomial time algorithm [48] for integer programming for fixed dimensions.

Babai [7] used Lovász's algorithm to find an approximate closest vector: Given L and a vector $y \in \mathbf{R}^n$, one can find in polynomial time a vector $b \in L$ such that

$$||y - b|| \leq \left(\frac{3}{\sqrt{2}}\right)^n \min_{v \in L} ||y - v||.$$

Håstad [37] also proved an interesting related result.

The basis reduction algorithm has been one of the most important algorithms. It has been used successfully in a variety of context, including the attack on knapsack based cryptosystems by Lagarias and Odlyzko [46], algebraic computations [40], the disproof of Merten's conjecture by Odlyzko and te Riele [56]. Other important results can be found in [41, 50]. In practice, Lovász's algorithm and its variants have performed rather well for moderate dimensions (up to 100), and much better than the theoretical upper bound (see [59]). Thus, any claim of intractability should bear in mind this computational experience.

4 Ajtai's worst-case to average-case connection

Let n, m and q be arbitrary integers. Let $\mathbf{Z}_q^{n \times m}$ denote the set of $n \times m$ matrices over $\mathbf{Z}_q$, and let $\Omega_{n,m,q}$ denote the uniform distribution on $\mathbf{Z}_q^{n \times m}$. For any $X \in \mathbf{Z}_q^{n \times m}$, the set $\Lambda(X) = \{y \in \mathbf{Z}^m \mid Xy \equiv 0 \bmod q\}$ (where the congruence is component-wise) defines a lattice of dimension m. Let $\Lambda = \Lambda_{n,m,q}$ denote the probability space of lattices consisting of $\Lambda(X)$ by choosing X according to $\Omega_{n,m,q}$.

We note that indeed $\Lambda(X)$ is a lattice of dimension m, since it is clearly a discrete additive subgroup of $\mathbf{Z}^m$, and each $qe_i \in \Lambda(X)$, where e_i has a single 1 at the ith position and 0 elsewhere. It also follows that $\Lambda(X)$ repeats itself within each $q \times q \times \cdots \times q$ box. In other words, $\Lambda(X)$ is invariant under the translations $y \mapsto y + qe_i$, for each $1 \leq i \leq m$.

By Minkowski's First Theorem, it can be shown that

$$\forall c \; \exists c' \text{ s.t. } \forall \Lambda(X) \in \Lambda_{n,c'n,n^c} \; \exists v \; (v \in \Lambda(X) \text{ and } 0 < \|v\| \leq n).$$

In fact the bound n can be reduced to $n^{\frac{1}{2}+\epsilon}$. The bound $\|v\| \leq n$ is needed to ensure that the assumption on the hypothetical algorithm $\mathcal{A}$ below is non-vacuous.

Theorem 4.1 (Ajtai) *Suppose there is a probabilistic polynomial time algorithm $\mathcal{A}$ such that for all n, when given a random lattice $\Lambda(X) \in \Lambda_{n,m,q}$ where $m = \alpha n \log n$ and $q = n^\beta$ for appropriate constants α, β, returns with probability $\frac{1}{n^{O(1)}}$, a vector of $\Lambda(X)$ of length $\leq n$, then there exists a probabilistic polynomial time algorithm $\mathcal{B}$ such that for all n, when given a basis $\{a_1, \ldots, a_n\}$ for an arbitrary lattice $L = L(a_1, \ldots, a_n)$, performs the following with high probability:*

1) Finds a basis $\{b_1, \ldots, b_n\}$ for L such that

$$\max_{i=1}^{n} \|b_i\| \leq n^{c_1} \cdot \text{bl}(L),$$

2) Finds an estimate $\tilde{\lambda}$ of $\lambda_1(L)$ such that,

$$\frac{\lambda_1(L)}{n^{c_2}} \leq \tilde{\lambda} \leq \lambda_1(L),$$

3) Finds the unique shortest vector $\pm v$ of L, if L has an n^{c_3} unique shortest vector, i.e. $\lambda_2(L) \geq n^{c_3} \cdot \lambda_1(L)$,

where c_1, c_2, c_3 are absolute constants.

Remark: This is the first such worst-case to average-case connection proved for a problem in NP believed not in P. While random-self-reducibilities were known for other problems, such as Quadratic Residuosity (QR), there is a technical difference. In QR, one must fix a modulus, then there is a worst-case to average-case connection for this modulus. But no such reduction is known among different moduli. The permanent is another example where there is a certain worst-case to average-case connection (see [29, 27, 20, 35]), but the permanent is not believed to be in NP.

Items 2) and 3) are derived from item 1) via a transference type argument, about which we will say more later in Section 7. Here we will focus on the ideas in the proof of item 1). Without loss of generality, we can assume that the lattice consists of integral vectors. The same result also holds for lattices with rational entries or with entries from any subfield of $\mathbf{C}$, as long as there is an effective bit representation for the lattice.

As Ajtai related to me, a guiding philosophical idea is the perspective that when you look from sufficiently afar, all lattices tend to look more alike.

We will now present some ideas from the proof.

Suppose we currently have a basis $\{b_1, \ldots, b_n\}$, where $\max_{i=1}^n \|b_i\|$ is greater than $\mathrm{bl}(L)$ by a large polynomial factor n^{c_1}, i.e.

$$\mu \equiv_{\mathrm{def}} \max_{i=1}^{n} \|b_i\| > n^{c_1}\mathrm{bl}(L).$$

The main procedure of $\mathcal{B}$ is iterative. Let S be a set of n independent vectors of L (initially $S = \{b_1, \ldots, b_n\}$). If the length of the elements of S at the start of the current iteration is large enough, the algorithm finds a set of independent vectors, each of at most half the length, with high probability. This means, in a polynomial number of steps we will have a set of short enough vectors, which can then be converted to a short basis with a loss of a factor $\leq \sqrt{n}$.

The fundamental domain $D = P(b_1, \ldots, b_n)$ forms a tiling of $\mathbf{R}^n$ via translations under L,

$$\mathbf{R}^n = \bigcup_{l \in L} (l + D),$$

as a disjoint union.

Consider a large cube

$$Q = \{x \in \mathbf{R}^n \mid x = \sum_{i=1}^{n} x_i e_i, 0 \leq x_i < M\},$$

where M is a certain polynomial factor greater than μ, say, $M = n^\gamma \mu$. For each i, we can "round" the corner point Me_i to a lattice point according to which translate $l_i + D$ it belongs to. This only involves solving a linear system expressing Me_i as a rational linear combination of the basis $\{b_1, \ldots, b_n\}$ and then rounding the coordinates. Thus for each $i = 1, \ldots, n$, let

$$Me_i = \sum_{j=1}^{n} \alpha_{ij} b_j \quad \text{and} \quad l_i = \sum_{j=1}^{n} \lfloor \alpha_{ij} \rfloor b_j.$$

Now

$$Q' = \{x \in \mathbf{R}^n \mid x = \sum_{i=1}^{n} x_i l_i, 0 \leq x_i < 1\},$$

is a reasonably good approximation of Q; we will call it a pseudocube. Note that the corner vertices of Q' are all lattice points. To ensure that Q' looks reasonably close to a cube, Ajtai chose $\gamma = 3$.

In the next step we subdivide Q' into a family of disjoint sub-pseudocubes, by subdividing Q' along each direction l_i into q subintervals, where q is polynomially bounded in n.

$$Q' = \bigcup_{0 \leq k_1, \ldots, k_n < q} \left(\sum_{i=1}^{n} \frac{k_i}{q} l_i + Q'' \right),$$

where the basic sub-pseudocube

$$Q'' = \{x \in \mathbf{R}^n \mid x = \sum_{i=1}^{n} x_i l_i, 0 \leq x_i < \frac{1}{q}\}.$$

We will make sure that the length of a side of Q'', which is roughly $\frac{M}{q}$, is still larger than $\mathrm{bl}(L)$ by a significant polynomial factor.

Suppose this is the case. Then with a series of technical lemmas, Ajtai shows that the number of lattice points within each translate $Q'' + \sum_{i=1}^{n} \frac{k_i}{q} l_i$ is roughly the same. This is intuitively quite plausible. But the technical details are not straightforward, especially if one wants a reasonably good bound. (See below.)

Once this approximate equi-distribution of lattice points is established, one can sample the "addresses" $(k_1, \ldots, k_n)$ of sub-pseudocubes, by uniformly sampling a lattice point in Q'. Once a lattice point v is picked, we decide to which sub-pseudocube it belongs by expressing v as a linear combination $\sum_{i=1}^{n} \frac{x_i}{q} l_i$, where $0 \leq x_i < q$, by solving a linear system. Then, we round off x_i and set $k_i = \lfloor x_i \rfloor$.

More generally, suppose we get m such samples, $v_j \in L$, $1 \leq j \leq m$. We decompose v_j as follows, (See Figure 4)

$$v_j = \sum_{i=1}^{n} \frac{k_{ij}}{q} l_i + r_j,$$

where r_j is a vector in Q''. Note that $\|r_j\|_2$ is $O(\frac{\sqrt{n}M}{q})$.

Here is the key observation: Suppose we are able to obtain an integral solution $X = (\xi_1, \dots, \xi_m)$ to

$$\sum_{j=1}^{m} k_{ij}\xi_j \equiv 0 \bmod q,$$

then $\sum_{j=1}^{m} \xi_j v_j$ would be a lattice point which has an interesting decomposition,

$$\sum_{j=1}^{m} \xi_j v_j = \sum_{i=1}^{n} \left(\frac{\sum_{j=1}^{m} k_{ij}\xi_j}{q} \right) l_i + \sum_{j=1}^{m} \xi_j r_j. \qquad (1)$$

We note that the quantity $\frac{\sum_{j=1}^{m} k_{ij}\xi_j}{q}$ is actually an integer, which makes the first term in (1) a lattice vector. Hence $\sum_{j=1}^{m} \xi_j r_j$, being the difference of two lattice points, must be a lattice point itself, (even though each r_j is probably not a lattice point.)

Suppose the integral solution X has every $|\xi_j| \leq n$, then

$$\begin{aligned} \|\sum_{j=1}^{m} \xi_j r_j\| &\leq m \cdot n \cdot O(\frac{\sqrt{n}M}{q}) \\ &= O\left(\frac{m \cdot n^{1.5+\gamma}}{q}\mu\right). \qquad (2) \end{aligned}$$

Now q can be chosen $\Theta(n^6)$ so that $\| \sum_{j=1}^{m} \xi_j r_j\| < \frac{\mu}{2}$, which is at most half of every $\|b_i\|$.

With the choice of $\gamma = 3$, Ajtai showed that the shape of the pseudocube and thus that of the sub-pseudocubes is very close to a perfect cube. With a choice of $q = \Theta(n^6)$, and a corresponding $m = O(n \log n)$, Minkowski's Theorem applies. Hence the assumption on $\mathcal{A}$ is non-vacuous and the newly produced lattice vector $\sum_{j=1}^{m} \xi_j r_j$ has length $< \frac{\mu}{2}$. On the other hand, the length of a side of a sub-pseudocube is approximately $\frac{M}{q}$ which is bounded below by $\frac{n^{\gamma+c_1}}{q}\text{bl}(L) = \Theta(n^{c_1-3}\,\text{bl}(L))$.

With the shape of the pseudocube approximately a perfect cube, and with a sufficiently large c_1, which makes each side of the sub-pseudocube sufficiently larger than $\text{bl}(L)$, Ajtai showed that the distribution induced on the address space $\{(k_1, \dots, k_n) \mid 0 \leq k_i < q\}$ by uniformly sampling lattice points from L is close to uniform. In fact, not only must the distribution of each sample $(k_1, \dots, k_n)$ be close to uniform, but also the joint distribution on all the m samples forming the matrix (k_{ij}) must be close to the uniform distribution $\Omega_{n,m,q}$. Only then can one legitimately invoke the assumed algorithm $\mathcal{A}$ and be guaranteed to obtain a short vector X with $\sum_{j=1}^{m} k_{ij}\xi_j \equiv 0 \bmod q$, and $\|X\| \leq n$, with nontrivial probability.

So far we have only produced one lattice vector $b'_1 = \sum_{j=1}^{m} \xi_j r_j$, which is shorter than $\mu = \max \|b_i\|$ by a factor of 2. We continue this process to produce n linearly independent lattice vectors $\{b'_1, \dots, b'_n\}$ to replace $\{b_1, \dots, b_n\}$. To show that these successive b'_i are linearly independent demands another set of technical lemmas which ultimately depend on the fact that c_1 is sufficiently large. In that case, Ajtai showed that within each sub-pseudocube the lattice is quite dense. It follows that, for every $n-1$ dimensional hyperplane Π, the number of lattice points on $\Pi \cap Q''$ is much smaller compared to the total number of lattice points in Q''. Moreover, this is true for every translate of Q''. It follows that the successive b'_i's are not likely to be linearly dependent on $\{b'_1, \dots, b'_{i-1}\}$. We will not provide any more technical details of Ajtai's proof. The interested reader is referred to [1].

Improving Ajtai's connection factors

What is outlined above is essentially Ajtai's proof [1], where some universal constants c_1, c_2 and c_3 are shown. Although no explicit values for these c_i's were given, and apparently no special effort was made to minimize them, implicitly a factor less than 8, 10 and 19, respectively, can be derived from the proofs of [1].

The factors n^{c_i} are called Ajtai's connection factors; they provide a measure of the tightness of the worst-case to average-case connection. The smaller the constants are, the tighter the connection one gets. As 2) and 3) are derived through 1) (see Section 7), n^{c_1} is the crucial factor. Cai and Nerurkar [18] obtained a substantial improvement to n^{c_1}, and consequently to the other factors as well. Here we give an overview of some of the ideas involved in this improvement. As is the case with Ajtai's proof [1], there are a number of technical points we have to gloss over due to limited space.

The general structure of the procedure of Cai and Nerurkar [18] closely follows Ajtai's proof, but much of the technical justification is different. As we saw above, the general idea is to sample lattice points, in order to induce an almost uniform distribution on a set of "address" vectors, which form the columns of a matrix that is close to uniformly distributed. The assumed algorithm $\mathcal{A}$ is applied to this matrix. By hypothesis, this algorithm performs well on the average, and thus we get a short vector which can be turned into a short vector of the original lattice.

In the choice of $M = n^{\gamma}\mu$, we need γ to be a sufficiently large constant in order to ensure that the resulting pseudocube is reasonably close to a perfect cube. We call this the shape condition. Then, we need to choose an integer q to be a sufficiently large polynomial (in n) in order to ensure that the newly produced remainder vector is shorter than the previous $||b_i||$. This involves m in the numerator in $n^{\gamma+1.5}m/q$ in (2), which has to be chosen after q in order to ensure that short vectors exist by Minkowski's First Theorem. Fortunately, this is not circular; for any polynomially bounded q, m only needs to be $O(n)$. But still q must depend on γ. Finally, given q, we must ensure that the length of a side of a sub-pseudocube M/q is sufficiently large compared to $\mathrm{bl}(L)$. We know that,

$$\frac{M}{q} = \frac{n^{\gamma}\mu}{q} > \frac{n^{\gamma+c_1}}{q}\mathrm{bl}(L)$$

This is where $\mu > n^{c_1}\mathrm{bl}(L)$ is used and c_1 has to be large. Cai and Nerurkar [18] achieve $c_1 = 3 + \epsilon$ for linearly independent vectors, and $c_1 = 3.5 + \epsilon$ for basis length.

The algorithmic improvement by Cai and Nerurkar [18] starts with a tiling of $\mathbf{R}^n$ by orthogonal "bricks" of sides at most μ, via Gram-Schmidt orthogonalization. This is in contrast to the tiling by fundamental domains in [1]. The advantage is that one can round off from a perfect cube to a lattice pseudocube with less error. Thus, for $M = n^{1.5}\mu$ and $w_i = Me_i$, we can round off w_i to a lattice point l_i such that $w_i = l_i + \delta_i$ and $\|\delta_i\| \leq \frac{\sqrt{n}\mu}{2}$. This implies $\|l_i\| \leq (n^{1.5} + \frac{\sqrt{n}}{2})\mu$. $P(l_1, \ldots, l_n)$ is the pseudocube constructed.

Secondly, in [18], the pseudocube is positioned centrally and subdivided. Each sub-pseudocube will have an address vector at the center. More precisely we will take $Q' = P(2l_1, \ldots, 2l_n) - \sum_{i=1}^{n} l_i = \{\sum_{i=1}^{n} z_i l_i \mid -1 \leq z_i < 1\}$. We partition Q' into q^n sub-pseudo-cubes, (where q is odd, say), such that the basic sub-pseudocube is $Q' = \{\sum_{i=1}^{n} z_i l_i \mid -\frac{1}{q} \leq z_i < \frac{1}{q}\}$. We will sample lattice points uniformly in the pseudocube Q'. This induces an almost uniform distribution on the address space. But this time we consider each address as corresponding to the *center* of the sub-pseudocube. When we express a sample lattice point v_j as the sum of this address vector and a remainder vector r_j, these remainder vectors tend to be symmetrically distributed with respect to the address vector at the center. (See Figure 5) Here an address vector is of the form $\sum_{i=1}^{n} \frac{k_{ij}}{q} l_i$, where each k_{ij} is even, $-(q-1) \leq k_{ij} \leq q-1$. The corresponding "address" is $(k_{1j}, k_{2j}, \ldots, k_{nj})$ reduced modulo q. Thus, when we estimate $|| \sum_{j=1}^{m} \xi_j r_j||$ probabilistically, the independent r_j's tend to cancel out instead of adding up. Note that $X = (\xi_1, \ldots, \xi_m)$ is a (short) solution obtained by the algorithm $\mathcal{A}$ given only the address matrix (k_{ij}). Given such a matrix one must ensure that the r_j are almost independently and centrally symmetrically distributed. This is geometrically quite intuitive, given a sufficiently large ratio of the sides of the sub-pseudocube to $\mathrm{bl}(L)$. But the hard part is to minimize this notion of "sufficiently large". It turns out that $q = n^{3+\epsilon}$ and $\mu > n^{3+\epsilon}\mathrm{bl}(L)$ will do. The technical part of the proof is rather involved.

There is one more idea in [18] in the improvements in terms of the algorithmic steps. It turns out to be insufficient to guarantee the generation of one almost uniform address vector, which makes up one column of the matrix. We must be able to generate m columns to form an almost uniformly generated matrix. This more stringent requirement is needed to apply the algorithm $\mathcal{A}$. In [18] we used an idea to amplify the "randomness" in each column vector generated, by adding together $\lceil 2/\epsilon \rceil$ copies of independent samples

$$\begin{aligned} v &= \sum_{i=1}^{n} \frac{k_i}{q} l_i + r, \\ v' &= \sum_{i=1}^{n} \frac{k_i'}{q} l_i + r', \text{ etc.} \end{aligned}$$

This gives a lattice point

$$v + v' + \cdots = \sum_{i=1}^{n} \frac{k_i + k_i' + \cdots}{q} l_i + (r + r' + \cdots).$$

Starting from the column vector $(k_1, k_2, \ldots, k_n)$ being $n^{-\epsilon}$-close to uniform, we show that the address vector

$$(k_1 + k_1' + \cdots, k_2 + k_2' + \cdots, \ldots, k_n + k_n' + \cdots) \bmod q$$

is n^{-2}-close to uniform, which would be sufficient to ensure that the matrix is close to being uniform. The price we pay for this is that each remainder vector is enlarged by a factor at most $\lceil 2/\epsilon \rceil$.

The more difficult part of the proof is to show that the lattice samples do induce a distribution that is $n^{-\epsilon}$-close to uniform on the address space. In addition to our "shape condition", which is accomplished by $\gamma = 1.5$, we need to estimate the volume of each sub-pseudocube to ensure that the number of lattice points within each sub-pseudocube is almost identical. Moreover, in order to obtain independent lattice vectors, we need to ensure that the proportion of lattice points in a sub-pseudocube that lie on any (co-1 dimensional) hyperplane is negligible.

The bounds in [18] use eigenvalues and singular values, and a theorem of K. Ball [8]. We cannot go into much detail here, but the following lemmas give a flavor of it.

Lemma 4.1 *Let $e_1, \ldots, e_n$ be the standard unit vectors. Let $u_1, \ldots, u_n$ be linearly independent vectors such that $\|u_i - e_i\| \leq \epsilon$. Then the parallelepiped $\mathcal{P}(u_1, \ldots, u_n)$ has volume*

$$1 - n\epsilon \leq \mathrm{vol}(\mathcal{P}(u_1, \ldots, u_n)) \leq (1 + \epsilon)^n.$$

(One cannot improve the lower bound to $(1-\epsilon)^n$ for large n.)

Lemma 4.2 *Let $e_1, \ldots, e_n$ and $u_1, \ldots, u_n$ be as above. Let H be a hyperplane. Then the $(n-1)$-dimensional volume of $P(u_1, \ldots, u_n) \cap H$ is at most $\sqrt{2e}(1+\epsilon)^{n-1}$.*

5 NP-hardness

Lagarias [44] showed that SVP, under the l_∞-norm, is NP-hard. For the related Closest Vector Problem (CVP), Van Emde Boas [62] showed it to be NP-hard for all l_p-norms, $p \geq 1$. Arora et al. [6] showed that, under any l_p-norm, CVP is NP-hard to approximate within any constant factor, and that if it can be approximated within a factor of $2^{\log^{1/2-\epsilon} n}$, then NP is in quasi-polynomial time.

It had long been thought that the Shortest Vector Problem for the natural l_2-norm is NP-hard. This was conjectured e.g., by Lovász [49]. It remained a major open problem until, in 1997, Ajtai [2] proved the NP-hardness of the SVP for this norm, under randomized reductions. Moreover, Ajtai showed that to approximate the shortest vector of an n-dimensional lattice within a factor of $\left(1+\frac{1}{2^{n^k}}\right)$ (for a sufficiently large constant k) is also NP-hard under randomized reductions. This was improved to $\left(1+\frac{1}{n^\epsilon}\right)$ for any constant $\epsilon > 0$ by Cai and Nerurkar [19], and then to any constant smaller than $\sqrt{2}$ by Micciancio [52].

Theorem 5.1 *It is NP-hard, under randomized polynomial time reductions, to find a shortest lattice vector, even to approximate it within a factor of $\sqrt{2}-\epsilon$, for any $\epsilon > 0$.*

In the next subsection we outline Ajtai's result. The presentation incorporates the simplifications and improvements of [19] but the main ideas are due to Ajtai. After that we present Micciancio's improvement.

Ajtai's result

Ajtai gave a randomized reduction from the following variant of the subset sum problem to SVP.

The restricted subset sum problem Given integers $a_1, \ldots, a_l, A$, each of bit-length $\leq l^3$, find a 0-1 solution to the system $\sum_{i=1}^{l} a_i x_i = A$ and $\sum_{i=1}^{l} x_i = \lfloor \frac{l}{2} \rfloor$.

We first define a lattice which will play a crucial role in the proof. This lattice is a modified version of the one used by Adleman (unpublished) in his reduction from factoring to the SVP, under some unproven assumptions. For this lattice, we need to choose several parameters depending on the l in the restricted subset sum instance.

- n is chosen to be a sufficiently large polynomial in l.
- m is chosen to be a sufficiently large polynomial in n. $m \gg n \gg l$.
- b is chosen randomly from the set of products of n distinct elements of $\{p_1, \ldots, p_m\}$, the first m primes.
- ω is chosen a constant root of b.
- B is polynomial in ω.

Clearly, B, b and ω are exponential in n. We will not be overly precise here about the values of these parameters in order not to obscure the main points. Using these parameters, Ajtai defines the following matrix, whose $m+2$ columns generate a lattice.

$$\begin{pmatrix} \sqrt{\log p_1} & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \sqrt{\log p_m} & 0 & 0 \\ 0 & \cdots & 0 & 0 & \omega^{-2} \\ B\log p_1 & \cdots & B\log p_m & B\log b & B\log\left(1+\frac{\omega}{b}\right) \end{pmatrix}$$

Lattice L_A

This lattice is then normalized. The normalized lattice has every vector of length at least 1 and a lot of vectors of length very close to 1. We will denote by v_i, the columns of the basis matrix for this modified lattice. We will denote this normalized matrix, as well as the lattice it generates, by L. With high probability, this lattice, $L = L(v_1, \ldots, v_{m+2})$, has the interesting properties we outline next. These properties are a consequence of the way primes are distributed and the convexity of the logarithm function.

1. All non-zero vectors have length at least 1.

2. There are a lot of vectors of small norm with the property that their first m basis coefficients $\in \{0, -1\}$. More precisely, let Y be the set of all $v \in L$, $v = \sum_{i=1}^{m+2} \alpha_i v_i$ with $\sum_{i=1}^{m} |\alpha_i| = n$, $\alpha_i \in \{0,-1\}$ for $i \in \{1, \ldots, m\}$, and $\|v\|^2 < 1+\delta$. Then $|Y| \geq 2^{n \log n}$. Here, δ is an exponentially small quantity.

3. Any two distinct elements of Y differ in their first m basis coefficients.

4. If v is a non-zero vector of L of squared norm less than $1+\frac{2}{m^{3\epsilon/4}}$, then the first $m+1$ coefficients of v have a special form. More precisely, if $v = \sum_{i=1}^{m+2} \alpha_i v_i$, $\|v\|^2 < 1+\frac{2}{m^{3\epsilon/4}}$, and $\alpha_{m+1} \geq 0$, then $\alpha_1, \ldots, \alpha_m \in \{0,-1\}$ and $\alpha_{m+1} = 1$.

This lattice is now extended in the following random manner depending on the given instance of the restricted subset sum problem. With high probability, given a reasonably short vector in this extended lattice, a solution to the instance can be produced.

Let $\sum_{i=1}^{l} a_i x_i = A$ be the given instance of the restricted subset sum problem. Let $\epsilon > 0$ be any constant. Let $\tau = 2/m^\epsilon$ and $\beta = \sqrt{\tau}$. Let $C = C_1, \dots, C_l$ be a random sequence of pairwise disjoint subsets of $\{1, \dots, m\}$. Define an $(l+2) \times (m+2)$ matrix D as follows. The $(m+2)^{\text{nd}}$ column is all zeros. The $(m+1)^{\text{st}}$ column is $(Al\beta, \lfloor \frac{l}{2} \rfloor l\beta, 0, \dots, 0)^T$. The other entries of the matrix are defined in the following manner.

1. The first row has the entry $a_i l\beta$ in the j^{th} position if $j \in C_i$, and otherwise has zero.
2. The second row has the entry $l\beta$ in the j^{th} position if j is in some C_i, and otherwise has zero.
3. For i from 3 to $l+2$, row i has β in the j^{th} position if $j \in C_{i-2}$ and otherwise has zero.

If $C_1, \dots, C_l$ are consecutive intervals of $\{1, \dots, m\}$, then D is the following matrix,

$$\begin{pmatrix} a_1 l\beta & \cdots & a_1 l\beta & \cdots & a_l l\beta & \cdots & a_l l\beta & \cdots & Al\beta & 0 \\ l\beta & \cdots & l\beta & \cdots & l\beta & \cdots & l\beta & \cdots & \lfloor \frac{l}{2} \rfloor l\beta & 0 \\ \beta & \cdots & \beta & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & \beta & \cdots & \beta & \cdots & 0 & 0 \end{pmatrix}$$

The extended lattice is the lattice $L^{(D)}$ generated by the columns of the matrix $\begin{pmatrix} L \\ D \end{pmatrix}$. A vector $\bar{v} \in L^{(D)}$ can be written $\begin{bmatrix} v \\ v' \end{bmatrix}$, where for some integral column vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{m+2})^T$, $v = \sum_{i=1}^{m+2} \alpha_i v_i \in L$ and $v' = D\boldsymbol{\alpha}$. Each v uniquely determines its $\boldsymbol{\alpha}$ and thus uniquely determines v'.

Ajtai uses a constructive variant of the following combinatorial lemma, due to Sauer, to show that any solution to a subset sum instance can be produced from the coefficients of some short vector. A proof of this lemma can be found, for example, in [4].

Lemma 5.1 (Sauer) *Let S be a finite set and $\mathcal{S}$ be a set of subsets of S. If for some k, $|\mathcal{S}| > \sum_{i=1}^{k} \binom{|S|}{i}$, then there is a $X \subseteq S$ with k elements such that $2^X = \{X \cap Z \mid Z \in \mathcal{S}\}$.*

That is, every subset of X can be realized by intersecting it with some element of $\mathcal{S}$. A consequence of Ajtai's constructive lemma is that a random sequence $C = C_1, \dots C_l$ of subsets of $\{1, \dots, m\}$, has the following property:

$\forall s \in \{0,1\}^l$, $\exists v = \sum_{j=1}^{m+2} \alpha_j v_j \in Y$ such that,
$\forall i \in \{1, \dots, l\}, s_i = -\sum_{j \in C_i} \alpha_j$.

This property implies that if there is a solution to the restricted subset sum instance then there is a vector in the set Y that gives rise to it. That is, suppose $\sum_{i=1}^{l} a_i x_i = A$ has a solution $x_i = s_i$, i.e.

$$s_i \in \{0,1\}, \quad \sum_{i=1}^{l} a_i s_i = A \quad \text{and} \quad \sum_{i=1}^{l} s_i = \lfloor \frac{l}{2} \rfloor.$$

Then, $\exists v \in Y$, $v = \sum_{j=1}^{m+2} \alpha_j v_j$, such that $\forall i \in \{1, \dots, l\}$,

$$s_i = -\sum_{j \in C_i} \alpha_j.$$

Since $v \in Y$, $0 < \|v\|^2 \le 1 + \delta$. Let $\bar{v} \in L^{(D)}, \bar{v} = \begin{bmatrix} v \\ v' \end{bmatrix}$, where $v' = D\boldsymbol{\alpha}$ and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{m+2})^T$. Let $v' = (v'_1, \dots, v'_{l+2})$. Then

$$\|\bar{v}\|^2 = \|v\|^2 + \|v'\|^2 \le (1 + \delta) + \tau \lfloor \frac{l}{2} \rfloor < 1 + \tau l. \quad (3)$$

The first inequality holds because $v'_1 = v'_2 = 0$, and exactly $\lfloor \frac{l}{2} \rfloor$ of v'_i for $i \ge 3$ are $-\beta$, the rest being zero. The last inequality holds because δ is exponentially small. Also, since v is a non-zero vector, so is $\bar{v}$, which implies

$$\lambda_1(L^{(D)}) \le \|\bar{v}\|. \quad (4)$$

We now prove that, assuming a solution to the restricted subset sum instance exists, one such solution can be constructed from an approximate shortest vector. Let $\bar{w} = \begin{bmatrix} w \\ w' \end{bmatrix}$ be a $(1 + \frac{\tau}{2})$ approximate shortest non-zero vector of $L^{(D)}$, i.e.

$$\lambda_1(L^{(D)})^2 \le \|\bar{w}\|^2 \le \left(1 + \frac{\tau}{2}\right) \lambda_1(L^{(D)})^2. \quad (5)$$

We will construct a solution to the subset sum instance, given $\bar{w}$. Since $\tau = 2/m^\epsilon$, this shows that it is NP-hard to approximate the shortest vector within a factor $\left(1 + \frac{1}{\text{dim}^\epsilon}\right)$, for any constant $\epsilon > 0$, where dim stands for the dimension of the lattice.

From (3), (4) and (5) we get

$$\|\bar{w}\|^2 \le \left(1 + \frac{\tau}{2}\right) (1 + \tau l), \quad (6)$$

and by the choice of τ and m ($m^{\epsilon/4} \gg l$), one can show that

$$\|\bar{w}\|^2 \le 1 + \frac{2}{m^{3\epsilon/4}}.$$

This matches the bound in property 4 of L. Let $w = (w_1, \dots, w_{m+2})$, $w' = (w'_1, \dots, w'_{l+2})$ and $w =$

$\sum_{j=1}^{m+2} \gamma_j v_j$. By property 4, replacing w by $-w$ if necessary, $\gamma_{m+1} = 1$. We now prove that

$$y_i = -\sum_{j \in C_i} \gamma_j$$

is also a solution by showing that, if not, the length of $\bar{w}$ would be too large. It is easy to see that since $\gamma_{m+1} = 1$,

$$\begin{aligned} w'_1 &= \beta l(A - \sum_{i=1}^{l} a_i y_i), \\ w'_2 &= \beta l(\lfloor \frac{l}{2} \rfloor - \sum_{i=1}^{l} y_i), \end{aligned}$$

and for $1 \leq j \leq l$,

$$w'_{j+2} = -\beta y_j. \tag{7}$$

Assume the y_i are not a solution. Then, at least one of the following three conditions must hold.

1) $\sum_{i=1}^{l} a_i y_i \neq A$, or

2) $\sum_{i=1}^{l} y_i \neq \lfloor \frac{l}{2} \rfloor$, or

3) $\exists i \; y_i \notin \{0,1\}$.

If 1) holds, then $|w'_1| \geq \beta l$, which means

$$\|\bar{w}\|^2 = \|w\|^2 + \|w'\|^2 \geq 1 + \beta^2 l^2 = 1 + \tau l^2,$$

where $\|w\| \geq 1$ holds by property 1 of L. This contradicts (6). If 2) holds, then $|w'_2| \geq \beta l$, and we get a similar contradiction again. Finally, it can be shown that if for some i, $y_i \notin \{0,1\}$ and $\sum_{j=1}^{l} y_j = \lfloor \frac{l}{2} \rfloor$, then

$$\sum_{j=1}^{l} y_j^2 \geq \left\lfloor \frac{l}{2} \right\rfloor + 2.$$

This means, by (7) and property 1 of L,

$$\|\bar{w}\|^2 = \|w\|^2 + \|w'\|^2 \geq 1 + \tau \left(\left\lfloor \frac{l}{2} \right\rfloor + 2 \right).$$

Since $\|\bar{v}\|^2 \leq (1+\delta) + \tau \lfloor \frac{l}{2} \rfloor$ (see (3)),

$$\|\bar{w}\|^2 - \|\bar{v}\|^2 \geq 2\tau - \delta \geq \tau.$$

Due to our choice of m as a sufficiently large polynomial in l, we have

$$\tau l = \frac{2}{m^\epsilon} l < 1.$$

Thus by (3), $\|\bar{v}\|^2 < 2$, and so

$$\|\bar{w}\|^2 - \|\bar{v}\|^2 > \frac{\tau}{2} \|\bar{v}\|^2.$$

Therefore,

$$\|\bar{w}\|^2 > \left(1 + \frac{\tau}{2}\right) \|\bar{v}\|^2 \geq \left(1 + \frac{\tau}{2}\right) \lambda_1(L^{(D)})^2,$$

which contradicts (5).

This completes the proof of Ajtai's result.

Micciancio's improvement

With the same basic framework, but using the closest vector problem instead of the restricted subset sum problem, Micciancio [52] got an improved hardness result for the SVP. He showed that it is NP-hard, under randomized reductions, to approximate the SVP to within any constant smaller than $\sqrt{2}$, using the fact that it is NP-hard to approximate the CVP to within any constant. (In fact, it is even NP-hard to do so to within a factor $2^{\log^{1-\epsilon} n}$, for an $\epsilon = o(1)$ [23], but this does not seem to lead to any improvement in his proof.)

To describe this result, it is convenient to formalize the approximation problems as promise problems [26]. The following defines the problem to approximate the closest vector within a factor $c \geq 1$.

CVP Promise Problem

Given an instance (B, y, d), where $B \in \mathbf{Z}^{n \times k}$ is a basis matrix, $y \in \mathbf{Z}^n$ is a target vector, and $d \in \mathbf{R}$, with the promise that either $\|Bx - y\| \leq d$ for some $x \in \mathbf{Z}^k$, or $\|Bx - y\| > cd$ for all $x \in \mathbf{Z}^k$, decide which is the case.

Arora et.al.[6] showed that for all constants $c \geq 1$, this promise problem is NP-hard. From the proof in [6] one gets that even the following modified version of the above problem is NP-hard for all constants $c \geq 1$.

Modified CVP Promise Problem

Given an instance (B, y, d), where $B \in \mathbf{Z}^{n \times k}$, $y \in \mathbf{Z}^n$, and $d \in \mathbf{R}$, with the promise that either $\|Bx - y\| \leq d$ for some $x \in \{0,1\}^k$, or $\|Bx - \alpha y\| > cd$ for all $x \in \mathbf{Z}^k$ and for all $\alpha \in \mathbf{Z} \setminus \{0\}$, decide which is the case.

We will call instances that satisfy the first alternative, YES instances, and those that satisfy the second one, NO instances. Note that, in the modified problem, a YES instance has a 0-1 solution and a NO instance has no solution even for arbitrary integral x and arbitrary (non-zero) multiples of the target vector.

Here is the definition of the corresponding SVP promise problem. It formalizes the problem of approximating the SVP within a factor c'.

SVP Promise Problem

> Given an instance (V, t), where V is a basis matrix, and $t \in \mathbf{R}$, with the promise that either $\|Vw\| \leq t$ for some non-zero integral w, or $\|Vw\| > c't$ for all non-zero integral w, decide which is the case.

We define YES and NO instances in a similar manner.

Micciancio gave a randomized many-one reduction that reduces the modified CVP promise problem with $c = \sqrt{2/\epsilon}$ to the SVP promise problem with $c' = \sqrt{2/(1+2\epsilon)}$, for any constant $\epsilon > 0$, mapping YES instances to YES instances and NO instances to NO instances. This shows that the SVP is NP-hard to approximate within any constant smaller than $\sqrt{2}$.

The heart of his proof is a technical lemma that asserts the existence of a probabilistic algorithm that on input 1^k, where k is from the CVP promise problem instance, constructs a lattice $L \in \mathbf{R}^{(m+1)\times m}$, a matrix $C \in \mathbf{Z}^{k\times m}$, and an $s \in \mathbf{R}^{m+1}$, such that with high probability,

- For every non-zero $z \in \mathbf{Z}^m$, $\|Lz\|^2 > 2$, and
- For all $x \in \{0,1\}^k$, $\exists z \in \mathbf{Z}^m$, such that $Cz = x$ and $\|Lz - s\|^2 < 1+\epsilon$.

Here, m depends polynomially on k.

The lattice L above is essentially the same as Ajtai's lattice L_A and C can be thought of as representing the 0-1 vector x by z. The existence of such a C and the fact that such a C can be randomly constructed depends on a version of Sauer's Lemma.

Let (B, y, d) be a given instance to the CVP promise problem with $c = \sqrt{2/\epsilon}$. The reduction maps it to the instance (V, t) of the SVP promise problem with $c' = \sqrt{2/(1+2\epsilon)}$, where

$$V = \left(\begin{array}{c|c} L & -s \\ \frac{\sqrt{\epsilon}}{d}BC & -\frac{\sqrt{\epsilon}}{d}\cdot y \end{array}\right)$$

and $t = \sqrt{1+2\epsilon}$. Note that $c't = \sqrt{2}$.

Let (B, y, d) be a YES instance. That is, $\|Bx - y\| \leq d$ for some $x \in \{0,1\}^k$. Then $\exists z \in \mathbf{Z}^m$, such that $\|(BC)z - y\| \leq d$ and $\|Lz - s\|^2 < 1+\epsilon$. Let w be the vector $\begin{pmatrix} z \\ 1 \end{pmatrix}$. Then

$$\|Vw\|^2 \leq (1+\epsilon) + \frac{\epsilon}{d^2}\cdot d^2 = 1 + 2\epsilon = t^2.$$

Let (B, y, d) be a NO instance. Let $w = \begin{pmatrix} z \\ \alpha \end{pmatrix}$ be a non-zero vector in $\mathbf{Z}^{m+1}$, where $z \in \mathbf{Z}^m$ and $\alpha \in \mathbf{Z}$. If $\alpha = 0$, then $z \neq 0$ and so

$$\|Vw\| \geq \|Lz\| > \sqrt{2} = c't.$$

If $\alpha \neq 0$, then

$$\|Vw\| \geq \frac{\sqrt{\epsilon}}{d}\|B(Cz) - \alpha y\| > \frac{\sqrt{\epsilon}}{d}\sqrt{\frac{2}{\epsilon}}\,d = \sqrt{2} = c't\,.$$

This completes the description of Micciancio's result.

Other hardness results

Dinur, Kindler and Safra [23] have recently improved the hardness factor for CVP. They show that CVP is NP-hard to approximate within a factor $2^{\log^{1-\epsilon} n}$, for an $\epsilon = o(1)$. Blömer and Seifert [11] study two problems considered by Ajtai in his worst-case/ average-case connection. These are the problems of computing a shortest set of independent lattice vectors and a shortest basis. Using the result of [23], they prove that both these problems are hard to approximate within a factor $n^{c/\log\log n}$, for some constant $c < 1$. Goldreich et al [34] show a reduction from the CVP to the SVP. While this reduction does not give us an improved hardness result, it has the properties of preserving the factor of approximation for the two problems and the dimension of the lattice.

Ravikumar and Sivakumar [57] consider the problem of deciding whether a lattice vector shorter than a given bound exists, under the promise that there is at most one such vector (not counting its negation). They prove a randomized reduction from the decision version of the general shortest vector problem to this problem, in the style of Valiant and Vazirani [60]. Lattice problems for a special kind of lattice defined by certain graphs have been studied in [17].

6 Non-NP-hardness results

To what extent can we expect to improve further the approximation factor for SVP and remain NP-hard? The current proof appears not feasible beyond $\sqrt{2}$. On the other hand, the best polynomial time approximation algorithms of Lovász and Schnorr are exponential in the approximation factor.

For polynomially bounded factors, transference theorems provide evidence that beyond a factor of $\Theta(n)$, the approximate SVP is not NP-hard. This is a result of Lagarias, Lenstra and Schnorr [45]. Transference theorems in the Geometry of Numbers give bounds to quantities such λ_i of the primal and the dual lattice. In [45] the following theorem is proved

$$1 \leq \lambda_i(L)\lambda_{n-i+1}(L^*) \leq \frac{1}{6}n^2,$$

for $n \geq 7, 1 \leq i \leq n$. This already gives an "NP proof" for a lower bound for $\lambda_1(L)$ up to a factor of $\Theta(n^2)$ by guessing an appropriate set of linearly independent lattice vectors of L^* all with length at most $\lambda_n(L^*)$.

Lagarias, Lenstra and Schnorr [45] proved more. A basis $\{b_1, b_2, \ldots, b_n\}$ is said to be reduced in the sense of Korkin and Zolotarev, if the following hold:

1. $||b_1|| = \lambda_1(L)$.
2. Let $\{\widehat{b}_1, \widehat{b}_2, \ldots, \widehat{b}_n\}$ be the Gram-Schmidt orthogonalization of $\{b_1, b_2, \ldots, b_n\}$,

   $$\widehat{b}_i = b_i - \sum_{k<i} \mu_{ik}\widehat{b}_k, \quad 1 \leq i \leq n.$$

   Then $|\mu_{ik}| \leq 1/2, 1 \leq k < i \leq n$.
3. If $L^{(n-i+1)}$ is the orthogonal projection of L to $(\mathrm{lsp}\{b_1, \ldots, b_{i-1}\})^\perp$ then $||\widehat{b}_i|| = \lambda_1(L^{(n-i+1)})$.

Essentially, a Korkin-Zolotarev basis is one which is *weakly reduced*, and the orthogonal projection of b_i is a vector of minimum length in the orthogonal projection of L in the complement of $\{b_1, \ldots, b_{i-1}\}$. In terms of Lovász's algorithm, if instead of comparing $b_i(i)$ and $b_{i+1}(i)$, we searched for a vector of minimum length in $\mathrm{lsp}\{b_i(i), \ldots, b_n(i)\}$, and called it b_i, we would have obtained a Korkin-Zolotarev basis. (Of course then this algorithm would have run in exponential time.)

Let B^* be a Korkin-Zolotarev basis of L^*. Then its dual basis $B = \{b_1, b_2, \ldots, b_n\}$ is called a dual Korkin-Zolotarev basis of L. Let $\lambda(B) = \min\{||\widehat{b}_i|| \mid 1 \leq i \leq n\}$, where $\{\widehat{b}_1, \widehat{b}_2, \ldots, \widehat{b}_n\}$ is the Gram-Schmidt orthogonalization of B. Then it is shown in [45] that

$$\lambda(B) \leq \lambda_1(L) \leq n\lambda(B).$$

In particular this gives a way to provide an "NP proof" of a lower bound for $\lambda_1(L)$ up to a factor of n by guessing an appropriate basis B^* and then calculating B. This places the promise problem of approximating $\lambda_1(L)$ up to a factor n within coNP.[2] Thus if NP $\neq$ coNP, then approximating $\lambda_1(L)$ up to a factor n is not NP-hard in the sense of Karp reductions. More precisely, if NP $\neq$ coNP, then there is no deterministic polynomial time reduction σ from SAT, $\sigma(\varphi) = (L, \lambda)$, such that if $\varphi \in$ SAT, then $\lambda_1(L) \leq \lambda$, and if $\varphi \notin$ SAT, then $\lambda_1(L) \geq n\lambda$.

Theorem 6.1 (Lagarias, Lenstra, Schnorr)
If NP $\neq$ coNP, then the problem of approximating $\lambda_1(L)$ within a factor n is not NP-hard.

The interplay between the primal and dual lattices and the related transference theorems play important roles in Ajtai's worst-case to average-case connection as well. We will discuss this topic in more detail in the next section. Here we present the following rather pretty result due to Goldreich and Goldwasser which improved the approximation factor for non-NP-hardness to $\sqrt{n}$.

The proof of Goldreich and Goldwasser [30] is based on constant round interactive proof systems. More precisely, they give a bounded round interactive proof system for proving a lower bound up to a factor $\sqrt{n}$ for both SVP as well as CVP. Of course the number of rounds can be reduced to one, either by standard techniques or by directly parallelizing their IP protocol. Also by standard techniques private coins can be replaced by public coins, so that what they showed can be stated as follows:

Theorem 6.2 (Goldreich, Goldwasser)
The problem of approximating $\lambda_1(L)$ within a factor $\sqrt{n}$ is in NP $\cap$ coAM. Thus if this problem is NP-hard under Karp reductions in the sense given above, then $\Sigma_2^p = \Pi_2^p$.

The last statement follows from a well-known result of Bopanna et. al. [12] which states that if coNP $\subseteq$ AM, then $\Sigma_2^p = \Pi_2^p$.

The basic idea of the IP protocol of [30] is rather simple and elegant and we will describe it here.

Suppose L satisfies the promise of either $\lambda_1(L) \leq t$ or $\lambda_1(L) > t \cdot \sqrt{n}$, and the prover claims that $\lambda_1(L) > t \cdot \sqrt{n}$. Imagine we surround each lattice point $p \in L$ a ball $B_p(r)$ centered at p with radius $r = t \cdot \sqrt{n}/2$. If the prover P is correct, then all such balls are disjoint. Now the verifier randomly picks a lattice point p in secret, and randomly picks a point z in $B_p(r)$. The verifier presents z to the prover, who should respond with p, the center of the ball from which z was chosen. It is clear that for an honest prover P with unlimited computing power, since all the balls $B_p(r)$ are disjoint, he has no difficulty meeting his obligation. However, suppose the prover P$'$ is dishonest, so that in fact $\lambda_1(L) \leq t$. Then for any lattice point p picked by the verifier, there is at least one nearby lattice point p' with $||p - p'|| \leq t$. Then $B_p(r)$ and $B_{p'}(r)$ would have a large intersection. This follows from the fact that the radius is almost $n^{1/2}$ times the distance of their respective centers. It follows that there is a significant probability that a dishonest prover will be caught, since in case a point $z \in B_p(r) \cap B_{p'}(r)$ is chosen, the verifier could equally have chosen p or p'.

The exponent $1/2$ in this interactive proof protocol

[2] Of course, technically a promise problem is not a decision problem while coNP is a decision problem class. But the meaning of this is clear and one can always modify the definitions slightly to make it proper.

comes from the well known fact that in n-dimensional space, two unit balls with center distance d have a significant intersection if $d < 1/\sqrt{n}$, and a negligible intersection if $d > 1/n^{1/2-\epsilon}$, for any $\epsilon > 0$. With some care the proof in [30] can improve the factor $\sqrt{n}$ to $\sqrt{n/\log n}$. It also shows the same bound for the Closest Vector Problem.

What about some other problems? The problem of n^c-unique shortest vector problem is prominent in the Ajtai worst-case to average-case connection. It also plays an important role in the Ajtai-Dwork public-key cryptosystem (see Section 8). Recall that a lattice is said to have an n^c-unique shortest vector if $\lambda_2(L)/\lambda_1(L) \geq n^c$. Equivalently, there exists $v \in L$, $v \neq 0$, such that for all $v' \in L$, if $||v'|| < n^c \cdot ||v||$, then v' is an integral multiple of v.

Define the following promise problem:

The n^c-unique shortest lattice vector problem:
Given a lattice with a n^c-unique shortest vector v, find the shortest vector $\pm v$.

Building on the idea of Goldreich and Goldwasser [30], Cai [15] proved the following:

Theorem 6.3 *The n^c-unique shortest lattice vector problem for $c \leq 1/4$ is not NP-hard unless the polynomial time hierarchy collapses to $\Sigma_2^p = \Pi_2^p$.*

7 Transference theorems

We have already mentioned the transference theorem of Lagarias, Lenstra and Schnorr [45] in the last section. There is a long history in geometry of numbers to study relationships between various quantities such as the successive minima associated with the primal and dual lattices, L and L^*. Such theorems are called transference theorems. The estimate for the product

$$\lambda_i(L)\lambda_{n-i+1}(L^*)$$

has a illustrious history: Mahler [51] proved that the upper bound $(n!)^2$ holds for all lattices. This was improved by Cassels [21] to $n!$. The first polynomial upper bound was obtained by Lagarias, Lenstra and Schnorr [45] as mentioned. The best estimate for this product is due to Banaszczyk [10], who showed that

$$1 \leq \lambda_i(L)\lambda_{n-i+1}(L^*) \leq Cn,$$

for some universal constant C. The Banaszczyk bound is optimal up to a constant, for Conway and Thompson (see [53]) showed that there exists a self-dual lattice family $\{L_n\}$ with $\lambda_1(L_n) = \Omega(\sqrt{n})$.

Part 2) and part 3) of Ajtai's worst-case to average-case connection in Theorem 4.1 are proved via transference type argument. Basically, if one can get a good estimate for the basis length for any lattice, one can apply this to the dual L^*. From a good estimate for $\mathrm{bl}(L^*)$, thus $\lambda_n(L^*)$, a transference theorem gives estimate for $\lambda_1(L)$. This is part 2) in Theorem 4.1. Part 3) employs some additional argument also of a transference type. We will discuss these matters in more detail. But first we take a closer look at transference theorems.

In addition to λ_i, there are several other lattice quantities that have been studied. The covering radius of L is defined to be the minimum radius of balls centered at each lattice point whose union covers $\mathbf{R}^n$.

$$\mu(L) = \min\{r \mid L + B(0;r) = \mathbf{R}^n\}.$$

Also if $d(u, L)$ denotes the minimum distance from a point u in $\mathbf{R}^n$ to a point in L, then

$$\mu(L) = \max\{d(u, L) \mid u \in \mathbf{R}^n\}.$$

(The minimum and maximum are obvoiusly achieved.)

We have seen the quantity

$$\xi = \sup_L \max_{1 \leq i \leq n} \lambda_i(L)\lambda_{n-i+1}(L^*),$$

where the supremum is taken over all n-dimensional lattices. Regarding covering radius $\mu(L)$ the relevant quantity is

$$\eta = \sup_L \mu(L)\lambda_1(L^*).$$

By triangle inequality $\mu(L) \leq \frac{1}{2}n\lambda_n(L)$, so that

$$\eta \leq \frac{1}{2}n\xi.$$

Given any L, we say a sublattice $L' \subseteq L$ is a *saturated sublattice* if $L' = L \cap \Pi$, where Π is the linear subspace of $\mathbf{R}^n$ spanned by L'. Saturated sublattices of dimension $n-1$ are in one-to-one correspondence with primitive vectors of L^*. (A lattice vector $v \neq 0$ is primitive if it is not an integral multiple of any other vector in the lattice except $\pm v$.) The correspondence is simply $L' = L \cap \{v\}^\perp$ and $\{v\}^\perp = \mathrm{lsp}(L')$. For any L and a saturated sublattice L' of dimension $n-1$ with normal (and primitive) vector $v \in L^*$, L is a disjoint union of parallel translations of L',

$$L = \bigcup_{k \in \mathbf{Z}} (L' + ku),$$

for some $u \in L$ such that $\langle u, v \rangle = 1$. Thus, each pair of nearest hyperplanes $\{v\}^\perp + ku$ and $\{v\}^\perp + (k+1)u$ has orthogonal distance $\langle u, \frac{v}{||v||}\rangle = \frac{1}{||v||}$. We call this a parallel decomposition of L.

For any L and any $u \in \mathbf{R}^n \backslash L$, we can compare $d(u, L)$, to the distance from u to the closest parallel translation of some $\{v\}^\perp = \mathrm{lsp}(L')$ which intersects L, over all such L'. Let

$$d_{\mathbf{Z}}(\langle u, v\rangle) = |\langle u, v\rangle - \lceil \langle u, v\rangle \rfloor|$$

be the fractional part of $\langle u, v\rangle$ rounded to the nearest integer, then we consider

$$\delta = \sup_{v \in L^*,\ \langle u,v\rangle \notin \mathbf{Z}} \frac{d_{\mathbf{Z}}(\langle u, v\rangle)}{||v||},$$

which measures the distance from u to the closest parallel translation, maximized among all directions $v \in L^*$. Now the following quantity is defined

$$\zeta = \sup_L \sup_{u \in \mathbf{R}^n \backslash L} \frac{d(u, L)}{\delta}.$$

By definition $d_{\mathbf{Z}}(\langle u, v\rangle) \leq 1/2$ and $||v|| \geq \lambda_1(L^*)$, so that $\delta \leq \frac{1}{2\lambda_1(L^*)}$. Hence

$$\zeta \geq 2\eta.$$

An upper bound $\zeta \leq \beta$ says that $\forall L$ and $\forall u \notin L$, there exists a parallel decomposition where the distance from u to the nearest lattice hyperplane is $\geq \beta d(u, L)$.

Lagarias et al. [45] proved that $\xi \leq \frac{1}{6}n^2$ and $\eta \leq \frac{1}{2}n^{3/2}$. Babai [7] proved that $\zeta \leq C^n$ for some universal constant C. Håstad [37] showed that $\zeta \leq 6n^2 + 1$. Similar bounds for ξ, η and ζ were also shown by Banaszczyk [9]. The best bounds for ξ, η and ζ were shown later by Banaszczyk [10], where ξ, η and ζ are all bounded by $O(n)$. The Banaszczyk bounds are all optimal up to a constant by the Conway-Thompson family of lattices (see [53]).

In [13] an extension of Banaszczyk's theorem of [10] is proved. Define $g_i(L)$ to be the minimum r such that the sublattice generated by $L \cap B(0; r)$ contains an i-dimensional saturated sublattice L', where $1 \leq i \leq n$. When $i = n$, it is called the *generating radius* and is denoted by $g(L)$. Clearly $g(L)$ is the minimum r such that a ball $B(0; r)$ centered at 0 with radius r contains a set of lattice vectors generating L. The study of $g(L)$ is motivated by the investigation of $\mathrm{bl}(L)$ and its relation to $\lambda_n(L)$. Clearly

$$\lambda_n(L) \leq g(L) \leq \mathrm{bl}(L).$$

The following inequality is shown in [13] for every lattice L of dimension n, using and extending the techniques of [10]:

$$g_i(L) \cdot \lambda_{n-i+1}(L^*) \leq Cn, \tag{8}$$

for some universal constant C, and for all i, $1 \leq i \leq n$. We will sketch the proof for the case $i = n$ for the generating radius $g(L)$.

The main tools of the proof are Gaussian-like measures on a lattice, and their Fourier transforms. For a given lattice L we define

$$\sigma_L(\{v\}) = \frac{e^{-\pi||v||^2}}{\sum_{x \in L} e^{-\pi||x||^2}}. \tag{9}$$

The Fourier transform of σ_L is

$$\widehat{\sigma_L}(u) = \int_{x \in \mathbf{R}^n} e^{2\pi i\langle u,x\rangle} d\sigma_L = \sum_{v \in L} e^{2\pi i\langle u,v\rangle}\sigma_L(\{v\}), \tag{10}$$

where $u \in \mathbf{R}^n$. Note that σ_L is an even function, so that

$$\widehat{\sigma_L}(u) \quad = \quad \sum_{v \in L} \sigma_L(\{v\}) \cos(2\pi\langle u, v\rangle). \tag{11}$$

Define

$$\tau_L(u) = \frac{\sum_{y \in L+u} e^{-\pi||y||^2}}{\sum_{x \in L} e^{-\pi||x||^2}}. \tag{12}$$

Then the following identity holds

Lemma 7.1

$$\widehat{\sigma_L}(u) = \tau_{L^*}(u). \tag{13}$$

The proof of Lemma 7.1 uses Poisson summation formula, see [39, 10]. The following lemma is proved in [10] and is crucial:

Lemma 7.2 *For each $c \geq 1/\sqrt{2\pi}$,*

$$\sigma_L(L \backslash B(0; c\sqrt{n})) < \left(c\sqrt{2\pi e}e^{-\pi c^2}\right)^n, \tag{14}$$

and for all $u \in \mathbf{R}^n$,

$$\frac{\sum_{v \in (L+u)\backslash B(0;c\sqrt{n})} e^{-\pi||v||^2}}{\sum_{x \in L} e^{-\pi||x||^2}} < 2\left(c\sqrt{2\pi e}e^{-\pi c^2}\right)^n, \tag{15}$$

where $B(0; c_1\sqrt{n})$ is the n-dimensional ball of radius $c_1\sqrt{n}$ centered at 0.

This lemma basically says that the total weight under σ_L of all lattice (or affine lattice) points outside of radius $c\sqrt{n}$ is exponentially small.

Now we prove (8) for $i = n$ and $C = 3/(2\pi)$. Suppose $g(L)\lambda_1(L^*) > 3n/2\pi$. Let c_1 and c_2 be two constants, such that $c_1c_2 > 3/2\pi$ and $c_1 > 1/\sqrt{2\pi}$ and $c_2 > 3/\sqrt{2\pi}$. By substituting L with sL for a suitable scaling factor s, we may assume that

$$g(L) > c_1\sqrt{n} \quad \text{and} \quad \lambda_1(L^*) > c_2\sqrt{n}.$$

Let L' be the sublattice of L generated by the intersection $L \cap B(0; c_1\sqrt{n})$. Then L' is a proper sublattice of L, since $g(L) > c_1\sqrt{n}$. If $\dim L' < n$, then let P be the linear span of L', and let $b_1, \ldots, b_i$ be a lattice basis of $L \cap P$, where $i = \dim L' < n$. This can be extended to a lattice basis $b_1, \ldots, b_i, \ldots, b_n$ for L and we may replace L' by the sublattice generated by $b_1, \ldots, b_i, \ldots, 2b_n$, say. Thus without loss of generality we may assume L' is of dimension n. The important point is that we have a proper sublattice $L' \subset L$, which is of dimension n and contains $L \cap B(0; c_1\sqrt{n})$.

For any fixed $u \in \mathbf{R}^n$,

$$\begin{aligned}
\widehat{\sigma_L}(u) &= \sum_{v\in L} \sigma_L(\{v\}) \cos(2\pi\langle u, v\rangle) \\
&= \sum_{v\in L'} \sigma_{L'}(\{v\}) \cos(2\pi\langle u, v\rangle) \\
&\quad + \sum_{v\in L'} (\sigma_L(\{v\}) - \sigma_{L'}(\{v\})) \cos(2\pi\langle u, v\rangle) \\
&\quad + \sum_{v\in L\setminus L'} \sigma_L(\{v\}) \cos(2\pi\langle u, v\rangle) \\
&= \widehat{\sigma_{L'}}(u) + A + B, \quad \text{say.}
\end{aligned}$$

Since $L \cap B(0; c_1\sqrt{n}) \subset L'$, the last term

$$\begin{aligned}
|B| &\leq \sum_{v\in L\setminus B(0;c_1\sqrt{n})} \sigma_L(\{v\}) \\
&< \left(c_1\sqrt{2\pi e} e^{-\pi c_1^2}\right)^n,
\end{aligned}$$

by Lemma 7.2 inequality (14). Denote the last term by ϵ_1^n, say.

For the other error term A, we can show similarly that

$$|A| < \epsilon_1^n.$$

Hence

$$\widehat{\sigma_L}(u) > \widehat{\sigma_{L'}}(u) - 2\epsilon_1^n. \tag{16}$$

Our next task is to show that we can choose an appropriate u so that $\widehat{\sigma_L}(u)$ is small yet $\widehat{\sigma_{L'}}(u)$ is large. By Lemma 7.1, we have $\widehat{\sigma_L}(u) = \tau_{L^*}(u)$, and $\widehat{\sigma_{L'}}(u) = \tau_{(L')^*}(u)$. Thus we only need to choose a u such that $\tau_{L^*}(u)$ is small and $\tau_{(L')^*}(u)$ is large.

The following lemma is proved in [13].

Lemma 7.3 *Suppose L_1 is a proper sublattice of L_2, then there exists a $p \in L_2$, such that*

$$\min_{q\in L_1} ||p - q|| \geq \frac{\lambda_1(L_1)}{3}.$$

(Since a lattice is a discrete subset of $\mathbf{R}^n$, the above minimum over q clearly exists.)

Now we note that since L' is a full ranked proper sublattice of L, L^* is a proper sublattice of $(L')^*$. That it is proper follows from the identity of index

$$\det((L')^*)/\det(L^*) = \det(L)/\det(L') > 1.$$

By Lemma 7.3, take a $u \in (L')^*$, such that $\min_{q\in L^*} ||u - q|| \geq \frac{\lambda_1(L^*)}{3}$. Then since $u \in (L')^*$, we have $(L')^* + u = (L')^*$, and

$$\tau_{(L')^*}(u) = \frac{\sum_{x\in(L')^*+u} e^{-\pi||x||^2}}{\sum_{x\in(L')^*} e^{-\pi||x||^2}} = 1.$$

On the other hand, since

$$\min_{q\in L^*} ||u - q|| \geq \frac{\lambda_1(L^*)}{3} > \frac{c_2}{3}\sqrt{n},$$

we note that no point in $L^* + u$ is within $\frac{c_2}{3}\sqrt{n}$ in norm, and so

$$\begin{aligned}
\tau_{L^*}(u) &= \frac{\sum_{x\in L^*+u} e^{-\pi||x||^2}}{\sum_{x\in L^*} e^{-\pi||x||^2}} \\
&< 2\left(\frac{c_2}{3}\sqrt{2\pi e} e^{-\pi(\frac{c_2}{3})^2}\right)^n = 2\epsilon_2^n \quad \text{say,}
\end{aligned}$$

by Lemma 7.2 inequality (15). Since both c_1 and $c_2/3 > 1/\sqrt{2\pi}$, we have both ϵ_1 and $\epsilon_2 < 1$ by elementary estimate. Thus it follows from (16) that

$$2\epsilon_2^n > 1 - 2\epsilon_1^n,$$

which is a contradiction for large n.

For the special class of lattices possessing n^ϵ-unique shortest vector, a stronger bound is proved [14], which lead to a further improvement in the Ajtai connection factors of part 2) and 3) in Theorem 4.1.

Theorem 7.1 *For every lattice L of dimension n, if L^* has an n^c-unique shortest vector, then*

$$1 \leq \lambda_n(L)\lambda_1(L^*) \leq O(n^\delta),$$

where

$$\delta = \begin{cases} 1 - c & \textit{if } 0 < c \leq 1/2, \\ 1/2 & \textit{if } 1/2 < c \leq 1, \\ 3/2 - c & \textit{if } 1 < c \leq 3/2, \\ 0 & \textit{if } c > 3/2. \end{cases}$$

In terms of the Ajtai connection factors in Theorem 4.1—in part 2) and part 3)—these new transference theorems improve all the factors to the range of approximately 3 and 4. Details can be found in [14]. Here we outline the general idea to derive parts 2) and 3) from 1).

The idea for the estimation of $\lambda_1(L)$ is relatively straightforward. From an estimate of the maximum length of a set of linearly independent vectors from L^*, one gets an estimate of $\lambda_1(L)$, via transference theorem.

To actually compute the shortest vector, the following idea is due to Ajtai [1]. If L^* has an n^c-unique shortest vector v, then L admits a parallel decomposition

$$L = \bigcup_{k \in \mathbf{Z}} (L' + ku),$$

where the parallel hyperplanes containing $L' + ku$ have orthogonal distance much larger than the basis length of L'. Now randomly sample a large polynomial number of lattice points within a certain bound. A $1/n^{O(1)}$ fraction of samples fall on the same parallel hyperplane, and the difference vector of such a pair belongs to the hyperplane $\mathrm{lsp}(L')$. If we can distinguish such pairs from the rest, then we can identify the normal vector for the hyperplane $\mathrm{lsp}(L')$, and by taking out the gcd, we can recover the shortest vector $\pm v$.

For two sample lattice points x and y, if they belong to the same parallel hyperplane, then by including a small fractional vector $(x-y)/N$ to the generating set of L, one does not change $\mathrm{bl}(L)$, since this is controlled by the distance between the parallel hyperplanes.

But if x and y belong to different parallel hyperplanes, then by including $(x-y)/N$ to the generating set of L, the new lattice will have many additional parallel translations of L' between any two originally adjacent parallel hyperplanes $\mathrm{lsp}(L') + ku$ and $\mathrm{lsp}(L') + (k-1)u$. This will reduce the basis length significantly.

Thus to be able to compute a good estimate of the basis length for L (actually an estimate of $\lambda_n(L)$ will do) leads to the identification of the unique shortest vector for L^*. Clearly improved transference theorem bounds sharpen the provable estimates in Ajtai's worst-case to average-case connection factors.

8 Lattice based cryptosystems

The Ajtai-Dwork public-key cryptosystem is based on the intractability of SVP for lattices with n^c-unique shortest vectors. Their cryptosystem has the provable property that if for a random instance the probability that an encryption of a zero can be distinguished from an encryption of a one is at least $1/2 + 1/n^{O(1)}$, then the worst-case n^c-unique SVP can be solved in probabilistic polynomial time. This is the only public-key cryptosystem with the property that breaking a random instance is as hard as solving the worst-case instance of the problem on which the cryptosystem is based.

Their cryptosystem is best viewed in terms of the dual lattice of a lattice possessing an n^c-unique shortest vector. For notational simplicity we will assume L^* has an n^c-unique shortest vector u. Then $\{u\}^\perp$ is a hyperplane whose intersection with $L^{**} = L$ is a saturated $n-1$ dimensional sublattice L_1 of L,

$$L_1 = L \cap \{u\}^\perp.$$

L then admits a parallel decomposition

$$L = \bigcup_{k \in \mathbf{Z}} (L_1 + kv),$$

where $v \in L$ and $\langle u, v \rangle = 1$. The (affine) hyperplanes $\{\{u\}^\perp + kv\}_{k \in \mathbf{Z}}$ have orthogonal distance $\frac{1}{||u||}$. Let π be the orthogonal projection to $\{u\}^\perp$. Then it can be shown that

$$\pi(L^*) = L_1^*.$$

(The dual of L_1 is defined within its own linear span $\{u\}^\perp$.) It follows that L_1^* has no short vectors compared to u. More precisely every non-zero $w \in L_1^*$ can be lifted to a vector $w' = w + \alpha u \in L^*$, where $|\alpha| \leq 1/2$. Since w' is not parallel to u,

$$n^{2c}||u||^2 \leq ||w'||^2 \leq ||w||^2 + \frac{1}{4}||u||^2.$$

Thus $||w|| \geq ||u||\sqrt{n^{2c} - 1/4} \approx n^c||u||$. By the transference theorems of Section 7, L_1 has a generating set of vectors of length $O\left(\frac{n^{1-c}}{||u||}\right)$.

Let $c > 5$ and let d, μ be real numbers such that $d > n^c \mu$. Ajtai and Dwork consider lattices with the following properties:

1. L has an $n-1$ dimensional sublattice L' with basis length at most μ;
2. If $H = \mathrm{lsp}(L')$ and $H' \neq H$ is a coset of H intersecting L, then the orthogonal distance d_L of H and H' is at least d.

Such a lattice is called a (d, μ) lattice. Clearly for a (d, μ) lattice, every $v \in L \backslash H$ has $||v|| \geq d_L \geq d > n^c \mu$. It follows that the $n-1$ dimensional saturated sublattice $L \cap H$ is uniquely determined by L. This is denoted by $L^{(d,\mu)}$.

Let $\mathcal{L}$ be a distribution on the set of (d, μ) lattices where $d \leq d_L \leq 2d$. Then the hidden hyperplane assumption for

$\mathcal{L}$ says that given a random (d,μ) lattice $L \in_R \mathcal{L}$, it is computationally infeasible to compute the hyperplane H (equivalently $L^{(d,\mu)} = L \cap H$).

Ajtai and Dwork [3] actually present three cryptosystems. In all three systems, the value 0 is encrypted by a point in $\mathbf{R}^n$ which is obtained as a small perturbation of a random lattice point, and the value 1 is encrypted as a totally random point in $\mathbf{R}^n$ (in an exponentially large region). The idea is that an encrypted 0 is near *some* affine translation of the hyperplane H intersecting the lattice, while an encrypted 1 most likely is not near such an affine hyperplane.

To decrypt, with the private key which is the normal vector to H (the unique shortest vector in L^*), one finds the distance of the ciphertext (a point in $\mathbf{R}^n$) from the closest such affine hyperplane and decodes it as a 0 if it is near enough and as a 1 otherwise. Clearly, there is a very small chance that a 1 may be decrypted as a 0. Goldreich et al [33] modify the cryptosystem to make the decryption error-free.

The exact details of the three cryptosystems differ somewhat. In the first and second systems security is proved under a distributional intractability for some hidden hyperplane assumption.

In the third system, no explicit lattice is presented. The private key is a randomly chosen vector in the unit ball. The public key is a set of random points near some regularly spaced affine hyperplanes in $\{H_i\}$, where $H_i = \{x \mid \langle x, u\rangle = i\}$ is the family of hyperplanes induced by u. The sum of a random subset of these points is itself close to some H_i. An encryption of 0 is a small perturbation on such a random subset sum, reduced modulo a certain parallelepiped determined by the public key. The encryption of 1 is still a totally random point in $\mathbf{R}^n$ in an exponentially large region.

A rough idea of the security of Ajtai-Dwork cryptosystem is the following: Suppose one can distinguish whether a point is near one of the (affine) hyperplanes. Then one can, with non-trivial probability, identify the normal vector to H. They [3] showed that the third cryptosystem is provably secure assuming only the worst case intractability of the n^c-unique shortest vector problem for some constant c.

Nguyen and Stern [55] have shown a converse to this. They prove that if the CVP can be approximated within a factor $cn^{4/3}$ then one can distinguish between encryptions of 0 and 1 with a constant advantage d, where d depends on c. Note that, Goldreich and Goldwasser [30] show that approximating the CVP within a factor $\sqrt{n}$ is unlikely to be NP-hard. It follows that to break the Ajtai-Dwork cryptosystem is also unlikely to be NP-hard. Nguyen and Stern also show that if the SVP can be approximated within a factor $n^{1/2-\epsilon}$, for any constant $\epsilon > 0$, then a distinguishing algorithm with an inverse polynomial advantage is possible.

Goldreich, Goldwasser and Halevi [32] proposed another cryptosystem based on the hardness of lattice problems. Another system based on the hardness of lattice problems was proposed by Cai and Cusick [16]. However no average-case/worst-case proof is known for either of these two systems.

Acknowledgements

This article was written while I was visiting the Computer Sciences Department at the University of Toronto. I wish to thank my colleagues at Toronto, Allan Borodin, Steve Cook, and especially Charlie Rackoff, for their hospitality. I thank Laci Lovász and Herb Scarf who introduced me to the beauty of lattice problems. I thank Miki Ajtai, Tom Cusick, Alan Frieze, Oded Goldreich, Shafi Goldwasser, Ravi Kannan, Janos Komlos, Jeff Lagarias, Ajay Nerurkar, Endre Szemeredi and Andy Yao for interesting discussions. I especially thank Ajay Nerurkar for his valuable assistance to me in the preparation of this article.

References

[1] M. Ajtai. Generating hard instances of lattice problems. In *Proc. 28th ACM Symposium on the Theory of Computing*, 1996, 99–108. Full version available from ECCC as TR96-007.

[2] M. Ajtai. The shortest vector problem in L_2 is NP-hard for randomized reductions. In *Proc. 30th ACM Symposium on the Theory of Computing*, 1998, 10–19. Full version available from ECCC as TR97-047.

[3] M. Ajtai and C. Dwork. A public-key cryptosystem with worst-case/average-case equivalence. In *Proc. 29th ACM Symposium on the Theory of Computing*, 1997, 284–293. Full version available from ECCC as TR96-065.

[4] N. Alon and J. Spencer. The Probabilistic Method (with an appendix on open problems by Paul Erdös). Wiley, 1992.

[5] T. Apostol. *Modular Functions and Dirichlet Series in Number Theory.* Springer-Verlag GTM 41 (Second Edition), 1990.

[6] S. Arora, L. Babai, J. Stern, and Z. Sweedyk. The hardness of approximate optima in lattices, codes, and systems of linear equations. In *Proc. 34th IEEE Symposium on Foundations of Computer Science*, 1993, 724-733.

[7] L. Babai. On Lovász' lattice reduction and the nearest lattice point problem. *Combinatorica*, 6:1–13, 1986.

[8] K. Ball. Cube slicing in $\mathbf{R}^\mathbf{n}$. *Proceedings of the American Mathematical Society*, 97(3):465–473, 1986.

[9] W. Banaszczyk. Polar Lattices from the point of view of nuclear spaces. *Rev. Mat. Univ. Complutense Madr.* 2 (special issue):35–46, 1989.

[10] W. Banaszczyk. New Bounds in Some Transference Theorems in the Geometry of Numbers. *Mathematische Annalen*, 296:625–635, 1993.

[11] J. Blömer and J-P. Seifert. On the complexity of computing short linearly independent vectors and short bases in a lattice. In *Proc. 31st ACM Symposium on Theory of Computing*, 1999. To appear.

[12] R. Boppana, J. Håstad and S. Zachos. Does Co-NP have Short Interactive Proofs? *Information Processing Letters*, 25:127–132, 1987.

[13] J-Y. Cai. A New Transference Theorem in the Geometry of Numbers. Submitted to *The 5th International Computing and Combinatorics Conference*, (COCOON) 1999, Tokyo, Japan.

[14] J-Y. Cai. Applications of a New Transference Theorem to Ajtai's Connection Factor. In these proceedings.

[15] J-Y. Cai. A Relation of Primal-Dual Lattices and the Complexity of Shortest Lattice Vector Problem. *Theoretical Computer Science* 207:105–116, 1998.

[16] J-Y. Cai and T. Cusick. A Lattice-Based Public-Key Cryptosystem. To appear in *Information and Computation*.

[17] J-Y. Cai, G. Havas, B. Mans, A. Nerurkar, J-P. Seifert and I. Shparlinski. On routing in circulant graphs. Submitted to *The 5th International Computing and Combinatorics Conference*, (COCOON) 1999, Tokyo, Japan.

[18] J-Y. Cai and A. Nerurkar. An Improved Worst-Case to Average-Case Connection for Lattice Problems. In *Proc. 38th IEEE Symposium on Foundations of Computer Science*, 1997, 468–477.

[19] J-Y. Cai and A. Nerurkar. Approximating the SVP to within a factor $\left(1 + \frac{1}{\dim^{\epsilon}}\right)$ is NP-hard under randomized reductions. In *Proc. of the 13th IEEE Conference on Computational Complexity*, 1998, 46–55.

[20] J-Y. Cai, A. Pavan and D. Sivakumar. On the Hardness of Permanent. In *Proc. of the 16th International Symposium on Theoretical Aspects of Computer Science*, 1999.

[21] J. W. S. Cassels. *An Introduction to the Geometry of Numbers.* Berlin Göttingen Heidelberg: Springer 1959.

[22] H. Cohn. *Advanced Number Theory.* Dover Publications, Inc.

[23] I. Dinur, G. Kindler and S. Safra. Approximating CVP to within almost-polynomial factors is NP-hard. In *Proc. 39th IEEE Symposium on Foundations of Computer Science*, 1998, 99–109.

[24] P. G. L. Dirichlet. Über die Reduktion der positiven quadratischen Formen mit drei unbestimmten ganzen Zahlen. *Journal für die Reine und Angewandte Mathematik*, 40:209–227, 1850.

[25] A. Dupré. Sur le nombre de divisions à effectuer pour obtenir le plus grande commun diviseur entre deux nombres entiers. *Journal de Mathématiques*, 11:41–74, 1846.

[26] S. Even, A. L. Selman and Y. Yacobi. The Complexity of Promise Problems with Applications to Public-key Cryptography. *Information and Control* 61:159–173, 1984.

[27] U. Feige and C. Lund. On the hardness of computing permanent of random matrices. In *Proc. 14th ACM Symposium on Theory of Computing*, 1982, 643–654.

[28] C. F. Gauss. *Disquisitiones Arithmeticae*. Transl. by A. A. Clarke. Yale University Press, 1966.

[29] P. Gemmell and M. Sudan. Highly resilient correctors for polynomials. *Information Processing Letters*, 43:169–174, 1992.

[30] O. Goldreich and S. Goldwasser. On the Limits of Non-Approximability of Lattice Problems. In *Proc. 30th ACM Symposium on Theory of Computing*, 1998, 1–9.

[31] O. Goldreich, S. Goldwasser, and S. Halevi. Collision-free hashing from lattice problems. Available from ECCC as TR96-042.

[32] O. Goldreich, S. Goldwasser, and S. Halevi. Public-key cryptosystems from lattice reduction problems. In *Advances in Cryptology – CRYPTO '97*, Burton S. Kaliski Jr. (Ed.), Lecture Notes in Computer Science, 1294:112-131, Springer-Verlag, 1997.

[33] O. Goldreich, S. Goldwasser, and S. Halevi. Eliminating decryption errors in the Ajtai-Dwork cryptosystem. In *Advances in Cryptology – CRYPTO '97*, Burton S. Kaliski Jr. (Ed.), Lecture Notes in Computer Science, 1294:105-111, Springer-Verlag, 1997.

[34] O. Goldreich, D. Micciancio, S. Safra and J-P. Seifert Approximating shortest lattice vectors is not harder than approximating closest lattice vectors. Available from ECCC as TR99-002.

[35] O. Goldreich and D. Ron and M. Sudan. Chinese remaindering with errors. Available from ECCC as TR 98-062.

[36] P. M. Gruber and C. G. Lekkerkerker. *Geometry of Numbers*. North-Holland, 1987.

[37] J. Håstad. Dual Vectors and Lower Bounds for the Nearest Lattice Point Problem. *Combinatorica*, 8:75–81, 1988.

[38] C. Hermite. Extraits de lettres de M. Ch. Hermite à M. Jacobi sur différents objets de la théorie des nombres. *Journal für die Reine und Angewandte Mathematik*, 40:261–278, 279–290, 291–307, 308–315, 1850.

[39] E. Hewitt and K. A. Ross. *Abstract Harmonic Analysis*, Vol II. Berlin Göttingen Heidelberg: Springer 1970.

[40] E. Kaltofen. Polynomial factorization 1987–1991. *LATIN '92*, I. Simon (Ed.), Lecture Notes in Computer Science, 583:294–313, Springer, 1992.

[41] R. Kannan. Minkowski's convex body theory and integer programming. *Mathematics of Operations Research*, 12:415–440, 1987.

[42] N. Koblitz. *Introduction to Elliptic Curves and Modular Forms*. Springer-Verlag, GTM 97, 1984.

[43] A. Korkin and G. Zolotarev. Sur les formes quadratiques positives quaternaires. *Mathematische Annalen*, 5:581–583, 1872.

[44] J. C. Lagarias. The computational complexity of simultaneous diophantine approximation problems. *SIAM Journal of Computing*, 14:196–209, 1985.

[45] J. C. Lagarias, H. W. Lenstra, and C. P. Schnorr. Korkin-Zolotarev Bases and Successive Minima of a Lattice and its Reciprocal Lattice. *Combinatorica*, 10(4):333-348, 1990.

[46] J. C. Lagarias and A. M. Odlyzko. Solving low-density subset sum problems. In *Proc. 24th IEEE Symposium on Foundations of Computer Science*, 1983, 1 – 10.

[47] A. K. Lenstra, H. W. Lenstra, and L. Lovász. Factoring polynomials with rational coefficients. *Mathematische Annalen*, 261:515–534, 1982.

[48] H. W. Lenstra, Jr. Integer programming with a fixed number of variables. *Mathematics of Operations Research*, 8:538–548, 1983.

[49] L. Lovász. *An Algorithmic Theory of Numbers, Graphs and Convexity*. SIAM, Philadelphia, 1986.

[50] L. Lovász and H. Scarf. The generalized basis reduction algorithm. *Mathematics of Operations Research*, 17(3):751–764, 1992.

[51] K. Mahler. Ein Übertragungsprinzip für konvexe Körper. *Čas. Pěstoväní Mat. Fys.* 68:93–102, 1939.

[52] D. Micciancio. The Shortest Vector in a Lattice is Hard to Approximate to within Some Constant. In *Proc. 39th IEEE Symposium on Foundations of Computer Science*, 1998, 92–98.

[53] J. Milnor and D. Husemoller. *Symmetric Bilinear Forms*. Berlin Heidelberg New York: Springer 1973.

[54] H. Minkowski. Über die positiven quadratischen Formen und über kettenbruchähnliche Algorithmen. *Crelles Journal für die Reine und Angewandte Mathematik*, 107:278–297, 1891.

[55] P. Nguyen and J. Stern. A converse to the Ajtai-Dwork security proof and its cryptographic implications. Available from ECCC as TR98-010.

[56] A. Odlyzko and H.J.J. te Riele. Disproof of the Mertens conjecture. *Journal für die Reine und Angewandte Mathematik*, 357:138–160, 1985.

[57] S. Ravikumar and D. Sivakumar. A note on the shortest lattice vector problem. In *Proc. 14th IEEE Conference on Computational Complexity*, 1999. To appear.

[58] C. P. Schnorr. A hierarchy of polynomial time basis reduction algorithms. *Theory of Algorithms*, 375–386, 1985.

[59] C. P. Schnorr and M.Euchner. Lattice basis reduction: Improved practical algorithms and solving subset sum problems. *Mathematical Programming*, 66:181–199, 1994.

[60] L. Valiant and V. Vazirani. NP is as easy as detecting unique solutions. *Theoretical Computer Science*, 47:85–93, 1986.

[61] B. Vallée. Un problème central en géométrie algorithmique des nombres: la réduction des réseaux;atour de l'algorithme LLL. *Inform. Théor. Appl.*, 345–376, 1989. English transl. by E. Kranakis, *CWI Quart* 3:95–120, 1990.

[62] P. van Emde Boas. Another NP-complete partition problem and the complexity of computing short vectors in lattices. Technical Report 81-04, Mathematics Department, University of Amsterdam, 1981.

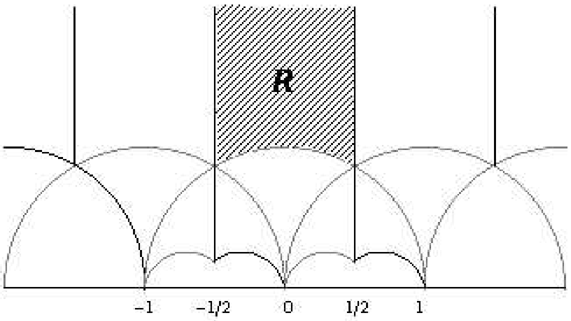

Figure 1

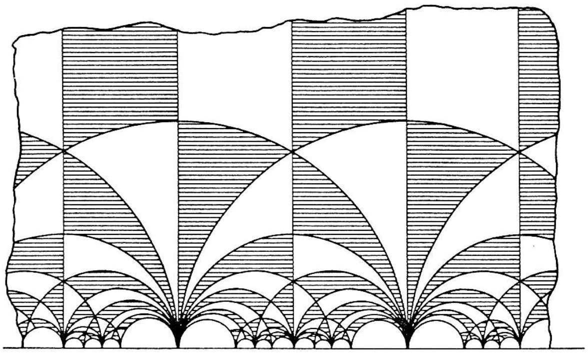

Figure 2

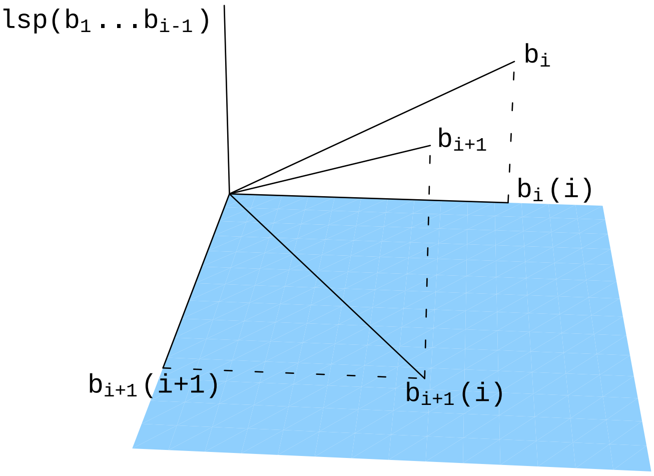

Figure 3

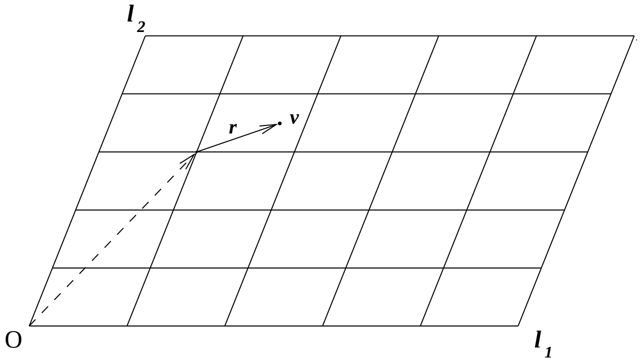

Figure 4

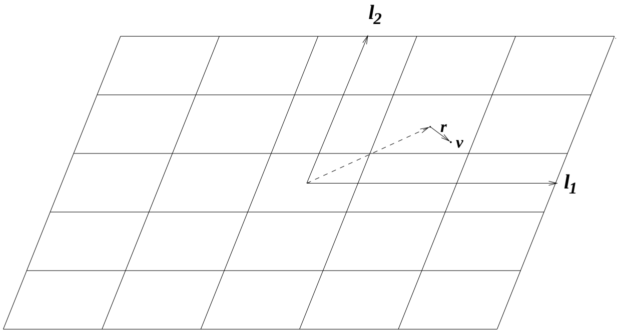

Figure 5

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'