CR10: Cryptanalysis Autumn 2018

Lecture V — 23 October, 2018

Lecturer G. Hanrot, E. Kirshanova *Scribe: Nidhi Purohit*

EUCLIDEAN LATTICES: An algorithm for Low- Density Sum Subset Algorithm

The subset sum problem is NP-Complete in the worst case [4], where a proof for NP-Completeness is shown. Therefore in this lecture we will be discussing an algorithm which in polynomial time solve almost all subset sum problems of sufficiently low density. Before, we need to define various terms briefly.

1 Definitions

Definition 1. ***Euclidean Lattices***: Let $\left(\vec{b_{\mathrm{i}}}\right)_{i\leq n}$ be linearly independent vectors in $\mathbb{R}^m$. The lattice L of rank n spanned by $\left(\vec{b_{\mathrm{i}}}\right)$ is

$$\begin{aligned} L\left(\left(\vec{b_{\mathrm{i}}}\right)\right)_{i\leq n} &= \sum_{n=1}^{\infty} \mathbb{Z}\cdot\vec{b_{\mathrm{i}}} \\ &= \left\{\sum x_{\mathrm{i}}\cdot\vec{b_{\mathrm{i}}}, x_{\mathrm{i}}\in\mathbb{Z}\right\} \end{aligned} \tag{1}$$

The $\vec{b_{\mathrm{i}}}$ are called a basis of L. Denote the lattice L by $L(B)$, where

$$\begin{bmatrix} | & . & . & . & . & | \\ \vec{b_1} & \vec{b_2} & .. & . & . & \vec{b_{\mathrm{n}}} \\ | & . & . & . & . & | \end{bmatrix} \in \mathbb{R}^{m\times n}$$

(*In this lecture, we consider basis-vector into a matrix column-wise (i.e, L is $Im_{\mathbb{Z}}(B)$.)*

Bases of L are related by a unimodular transformation, $(\mathrm{B} = \mathrm{B}'\cdot U, U \in GL_{\mathrm{n}}(\mathbb{Z}))$.

Main lattice invariants:

- The determinant of L: $\det(L) = \sqrt{\det(B^{\mathrm{T}})\cdot B}$
  If B is square i.e, $B\in\mathbb{R}^{n\times n}$, $\det(L) = |\det(B)|$.
  Geometric interpretation : determinant is the inverse of the density of the lattice L points in $\mathbb{R}^{\mathrm{m}}$.
  Algebraic interpretation: $|\mathbb{R}^{\mathrm{m}}/L(B| = \det(L)$.
- The first minimum of a lattice $L(B)$:

$$\lambda_1(L) = min\left\{r : \exists\vec{b} \in L \setminus \vec{0} : \left\|\vec{b}\right\| \leq r\right\}, \text{ where } \|\cdot\| \text{ is an Euclidean norm.}$$

In other words, the length of shortest vector in $L(B)$.

Fact 1.*(Minkowski's Ist Theorem)*
Let $m = n$. for any L of rank n,

$$\lambda_1 \leq \sqrt{n}\cdot(\det(L))^{\frac{1}{n}}$$

The bound is tight up to a constant.

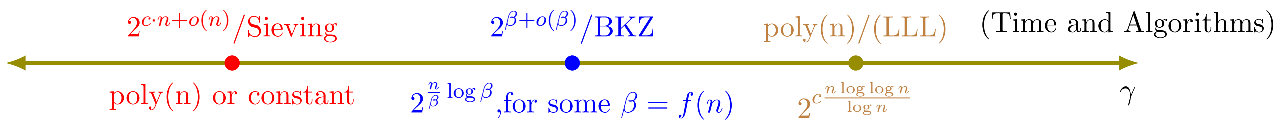
Definition 2. ***The Shortest Vector Problem*** (SVP) ***on*** L : It asks to find a non-zero vector $\vec{v} \in L$ such that $\|\vec{v}\| = \lambda_1(L)$.

- The solution to SVP is not unique.

Definition 3. ***The Approximation SVP with approximation factor*** $\gamma\geq 1$ (Approax SVP$_\gamma$) asks to find $\vec{v} \in L$ such that,

$$\|\vec{v}\| \leq\gamma\cdot\lambda_1(L)$$

In particular, the hardness of Approx SVP$_\gamma$ depends on γ.


$2^{c\cdot n+o(n)}$/Sieving
$2^{\beta+o(\beta)}$/BKZ
poly(n)/(LLL)
(Time and Algorithms)
poly(n) or constant
$2^{\frac{n}{\beta}\log\beta}$,for some $\beta = f(n)$
$2^{c\frac{n\log\log n}{\log n}}$
γ

Time/ Algorithm for Approx. SVP_γ

At some point if γ=$2^{\log^{1-\epsilon}}$, $\epsilon > 0$, *SVP is NP-Hard.*

Here,

- LLL - An algorithm due to Lenstra, Lenstra, Lovasz [5].
- BKZ - block Korkine- Zolotarev [6].

2 Estimate, Upper bounds.

2.1 Number of integers vectors in a ball:

Denote $S_{\text{n}}(R) = \left\{\vec{X} \in \mathbb{Z}^n : \left\|\vec{X}\right\|^2 \leq R\right\}$
Number of vectors inside or on the n-dimensional sphere of radius $\sqrt{R}$ centered at $\vec{O}$.

Theorem 1. *For all $n \geq 1$, $S_n(\frac{n}{2}) \leq 2^{C_o \cdot n}$, where C_o = 1.54725.*

Proof. Consider $\theta(z) = 1 + 2\sum_{n=1}^{\infty} z^{\mathrm{i}^2} = 1 + 2z + 2z^4 + ...$, $(\theta_3(z)$: Jacobi θ_3 function), $\theta(z)$ *is a θ-series of 1-dim integer lattice $\mathbb{Z}$.*

In general for a lattice Λ, θ is defined as:

$$\theta_{\mathrm{n}}(z) = \sum_{X \in \Lambda} q^{\|X\|^2 \cdot z}, where, q = e^{+\pi i} = \sum_{n=1}^{\infty} N_{\mathrm{m}} q^{\mathrm{m}} \tag{2}$$

where N_{m} counts the number of lattice points at distance m from the origin.

Let,

$$\gamma_{\mathrm{n}}(k) = \left\{X \in \mathbb{Z}^n : \|X\|^2 \leq k\right\}$$

then

$$(\theta(z))^n = \sum_{k=1}^{\infty} \gamma_{\mathrm{n}}(k) z^k.$$

Now to relate $S_{\mathrm{n}}(R)$ to $(\theta(z))^n$.

for $x \geq 0$, it holds $e^{ndx} \cdot e^{-k.x} \geq 1$, $k \leq nd$ then,

$$S_{\mathrm{n}}(\alpha_{\mathrm{n}}) = \sum_{k \leq \alpha n} \gamma_{\mathrm{n}}(k) \leq e^{ndx} \sum_{k=0}^{\infty} \gamma_{\mathrm{n}}(k).e^{-kx} = e^{ndx}[\theta(z)]^n$$

Taking ln, obtain
$\ln S_{\mathrm{n}}(\alpha_{\mathrm{n}}) \leq n.dx + n \cdot \ln(\theta(z)) = n(\alpha x + \ln \theta(z))$, Take, $\alpha x + \ln \theta(z) = \delta(x)$

Minimizing δ for $\alpha = \frac{1}{2}$, gives

$$S_{\mathrm{n}}(\alpha_{\mathrm{n}}) \leq 2^{1.54725n}$$

□

3 Low- Density Subset Sum(S)

<u>*Goal: Reduce an instance of low-density Subset Sum to SVP.*</u>

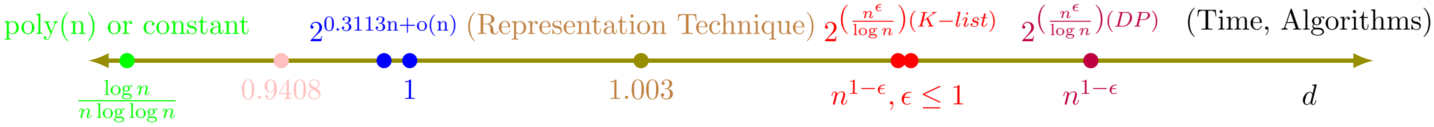
Remainder: The Subset Sum Problem: Given $a_1, a_2 \ldots\ , a_{\mathrm{n}} \xleftarrow{\$} [1,\ \mathrm{A}]$, and $s = \sum_{i=1}^{n} e_{\mathrm{i}} a_{\mathrm{i}}$ with $e_{\mathrm{i}} \in \{1, 0\}$, find $\vec{e}= (e_1, e_2 \ldots\ e_{\mathrm{n}})$.

The density of Subset Sum is defined as:

$$d = \frac{n}{\log\left(\max\{a_{\mathrm{i}}\}\right)} = \frac{n}{\log A}$$

This problem is known to be NP-complete [3](in its feasibility recognition form), and so is thought to be very hard in general. However, there are two algorithms, One by Brickell [1] and the other by Lagarias and Odlyzko [2], which in polynomial time solve almost all subset sum problems of sufficiently low density. Both methods rely on basis reduction algorithms to find short non-zero vectors in special lattices.

The interesting case is $d < 1$. For $d < 0.9408$, it can be reduced to SVP in poly(n), for $1 < d < 1.003$, the problem is NP-hard.


poly(n) or constant
$2^{0.3113\mathrm{n}+\mathrm{o}(\mathrm{n})}$ (Representation Technique)
$2^{\left(\frac{n^\epsilon}{\log n}\right)(K-list)}$
$2^{\left(\frac{n^\epsilon}{\log n}\right)(DP)}$
(Time, Algorithms)
$\frac{\log n}{n \log\log n}$
0.9408
1
1.003
$n^{1-\epsilon}, \epsilon \leq 1$
$n^{1-\epsilon}$
d

Lemma 2. *The solution vector $\hat{e} = (0, e_1, e_2 \dots, e_n) \in \mathbb{Z}^{n+1}$ is in $L(B)$.*

Proof. $B \cdot (-1, e_1, e_2 \dots, e_\mathrm{n})^t = (0, e_1, e_2 \dots, e_\mathrm{n})$
Consider a lattice $L(B)$ spanned by the columns of $B \in \mathbb{Z}^{(\mathrm{n}+1)(\mathrm{n}+1)}$

$$\begin{bmatrix} Ns & Na_1, & . & . & . & Na_\mathrm{n} & \\ 0 & 1 & 0 & 0 & . & 0 & \\ 0 & 0 & 1 & 0 & . & 0 & \\ 0 & 0 & 0 & 1 & . & 0 & \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & . & 1 & \end{bmatrix} \in \mathbb{R}^{m \times n}$$

$L(B)$ is of rank $(n+1)$.
It is not enough to show that $\hat{e} \in L(B)$.
We would like $\hat{e}$ to be the shortest vector in L. In the other words, we are interested in vectors $\hat{X} \in L$:

$$\begin{cases} \left\|\hat{X}\right\| \leq \|\hat{e}\|, \\ \hat{X} \in L, \\ \hat{X} \neq \{\vec{0}, \pm\hat{e}\} \end{cases} \tag{3}$$

Lets fix some notation,

$$\sum_{i=1}^{n} e_\mathrm{i} \leq \frac{n}{2}$$

(otherwise replace $s = \sum_{i=1}^{n} a_\mathrm{i} - s$ and the first basis vector by $(N(\sum_{i=1}^{n} a_\mathrm{i} - s), 0, 0, ., ., ., 0)^t)$.

Further, denote $T = \sum_{i=1}^{n} a_\mathrm{i}$. Then wlog, $s \geq \frac{T}{n}$ (otherwise if $s < \frac{T}{n}$, any $a_\mathrm{i} > \frac{T}{n}$ can not be in the subset), $s \leq (1 - \frac{1}{n}) \cdot T$ (otherwise, any $a_\mathrm{i} \geq \frac{T}{n}$ must be in the subset).
That is, $\frac{T}{n} \leq s \leq \frac{n-1}{n} T$ □

Lemma 3. *For any* $a_i \xleftarrow{\$} [1, A]$ *and* $N > \sqrt{n}$ *and* C_o *as in Theorem 1, it holds*

$$P := \Pr_{a \xleftarrow{\$} A} [\exists \hat{X} \mathit{that\ satisfies}(3)\ ] \le n(2n\sqrt{\tfrac{n}{2}} + 1)\tfrac{2^{C_o n}}{A}$$

In particular, if $A = 2^{Cn}$ *with* $C \ge C_o, \lim_{n \to \infty} P = 0$

It means that for

$$d \le \frac{n}{C_o.n} = 0.6463$$

the Subset Sum Problem can be reduced to SVP.

Proof. Since $N > \sqrt{n}$, for $\hat{X}$ to satisfy (3), it must hold that $\hat{X}_1 = 0$, (otherwise, $\left\|\hat{X}\right\| \ge |\ \hat{X}_1\ | \ge N > \sqrt{n}\ \ > \|\hat{e}\|\ ), (\ \|\hat{e}\| \le \sqrt{\frac{n}{2}})$.

Denote, $\vec{X} = X_1, X_1\ ...\ X_{\mathrm{n}}, \vec{e} = e_1, e_1\ ...\ e_{\mathrm{n}}$,

$$P \le \Pr_{a_{\mathrm{i}}}(\exists(\vec{X}, y) \in \mathbb{Z}^{\mathrm{n}} \times \mathbb{Z} : \left\|\vec{X}\right\| \le \|\vec{e}\|, |y| < n \cdot \sqrt{\frac{n}{2}}, \vec{X} \notin \{0, \vec{e}, -\vec{e}\}, \sum X_{\mathrm{i}} a_{\mathrm{i}} = sy)$$

$$\le \Pr\,[\sum X_{\mathrm{i}} a_{\mathrm{i}} = sy, \text{for fixed}(\exists(\vec{X}, y), \left\|\vec{X}\right\| \le \|\vec{e}\|, |y| < n \cdot \sqrt{\frac{n}{2}}, \vec{X} \notin \{0, \vec{e}, -\vec{e}\} \cdot K \cdot K^{'}] \tag{4}$$

where,

$$K = |\{\vec{X} : \left\|\vec{X}\right\| \le \|\vec{e}\||\} \tag{5}$$

and

$$K^{'} = |\{y : \|y\| < n \cdot \sqrt{\frac{n}{2}}\}| \tag{6}$$

In equation(4)

$$\sum_{i=1}^{n} a_{\mathrm{i}} \cdot x_{\mathrm{i}} = y \sum_{i=1}^{n} e_{\mathrm{i}} \cdot x_{\mathrm{i}} \iff \sum_{i=1}^{n} a_{\mathrm{i}} \cdot z_{\mathrm{i}} = 0, \text{for}, z_{\mathrm{i}} = x_{\mathrm{i}} - y e_{\mathrm{i}}$$

Since $\vec{X} \ne \vec{0}$, wlog, we assume $z_{\mathrm{i}} = 0$,

$$\implies Pr[\sum_{i=1}^{n} a_{\mathrm{i}} \cdot z_{\mathrm{i}} = 0] = Pr[a_{\mathrm{i}} - \sum_{i=1}^{n} \frac{a_{\mathrm{i}}}{z_1}]$$

Take, $z' = \sum_{i=1}^{n} \frac{a_{\mathrm{i}}}{z_1}$, we get,

$$\sum_{j=1}^{A} Pr[a_{\mathrm{i}} = j \mid z' = j] \cdot Pr[z' = j]$$

Since $a_{\mathrm{i}}'s$ are choosen uniformly at random thus,

$$\sum_{j=1}^{A} Pr[a_{\mathrm{i}} = j] \cdot Pr[z' = j] = \sum_{j=1}^{A} \frac{1}{A} Pr[z' = j]$$

As $Pr[z' = j] < 1$,
This implies

$$\sum_{j=1}^{A} Pr[a_{\mathrm{i}} = j] \cdot Pr[z' = j] \leq \frac{1}{A}$$

Now consider (5) i.e. $K = |\{\vec{X} : \left\|\vec{X}\right\| \leq \|\vec{e}\|\}| \leq |\{\vec{X} : \left\|\vec{X}\right\| \leq \|\vec{e}\|\}| \leq s \cdot (\frac{n}{2}) \leq 2^{\mathrm{C_o n}}$

Consider (5) i.e $K = |\{\vec{X} : \left\|\vec{X}\right\| \leq \|\vec{e}\||\} \leq 1 + 2 \cdot n\sqrt{\frac{n}{2}}$

Hence,

$$\implies P \leq n(1 + 2n\sqrt{\frac{n}{2}}) \cdot \frac{2^{\mathrm{C_o n}}}{A}$$

Define $y \in \mathbb{Z}$ such that,

$$y \cdot s = \sum_{i=1}^{n} \hat{X}_{\mathrm{i+1}} a_{\mathrm{i}}, i.e$$

y is the coefficient of $\hat{X}$ in its first basis vector b_1. Since $s > 0$,

$$\|y\| \cdot s = |\sum_{i=1}^{n} \hat{X}_{\mathrm{i+1}}| \leq \left\|\hat{X}\right\| \cdot |\sum_{i=1}^{n}, a_{\mathrm{i}}|$$

$$\leq T \cdot \sqrt{\frac{n}{2}}$$

$$\leq s \cdot n\sqrt{\frac{n}{2}}(\left\|\hat{X}\right\| \leq \|\hat{e}\|, T \leq s \cdot n)$$

$$\iff |y| \leq n \cdot \sqrt{\frac{n}{2}}.$$

□

Remarks:

- Setting $b_{\mathrm{i}} = (Ns, \frac{1}{2} \dots , \frac{1}{2})^t$ improves d to $d < 0.9408$ (it allows to consider only $\vec{X}'s$) such that $\left\|\vec{X}\right\| < \frac{\sqrt{n}}{2}$, in particular the optimization in Theorem(1) when $d = \frac{1}{4}$, gives $C_{\mathrm{o}} = 1.0628$; hence $d < \frac{1}{C_{\mathrm{o}}} = 0 \cdot 9408$.
- Applying Minkowski's bound to $L(B)$, we expect the length of the shortest vector in $L(B)$ to be

$$\lambda_1(L(B)) \leq \sqrt{n+1} \cdot \det(L(B))^{\frac{1}{n+1}}$$

$$\approx \sqrt{n} \cdot (Ns)^{\frac{1}{n}} \approx \sqrt{n} \cdot (\sqrt{n}A)^{\frac{1}{n}}, (\text{since}, s \approx A, N \approx \sqrt{n})$$

In poly(n), LLL applied to $L(B)$, returns a vector $\vec{e'} \in L(B)$ such that $\|\vec{e}\| \leq 2^{\frac{n \lg \lg n}{\lg n}} \lambda_1(L(B))$ As soon as $\|\vec{e}\|$ for $\hat{e}$ the Subset Sum solution, is by a factor ot $2^{\frac{-n \lg \lg n}{\lg n}}$ shorter than $\lambda_1(L(B))$, LLL will return $\|\hat{e}\|$:

$$\|\hat{e}\| \leq 2^{\frac{n \lg \lg n}{\lg n}} \cdot \lambda_1(L(B))$$

$$\sqrt{\frac{n}{2}} \leq 2^{\frac{n \lg \lg n}{\lg n}} \cdot \sqrt{n}(\sqrt{n}A)^{\frac{1}{n}}$$

on solving for A, we get:

$$A \geq O(2^{\frac{n^2 \lg \lg n}{\lg n}})$$

$$\implies \alpha = \frac{n}{\log A} < \frac{\lg n}{n \lg \lg n}$$

One can extend these arguments to BKZ algorithm

- This algorithm can be extended to non-zero weights e_i. We only require $\|e\|$ to be short.

References

[1] E. F. Brickell, *Solving low density knapsacks*, Advances in Cryptology, Proceedings of Crypto 83, Plenum Press, New York, 1984, 25-37.

[2] A. M. Frieze, *On the Lagarias-Odlyzko algorithm for the subset sum problem,* SIAM J. Comput. 15(2) (May 1986), 536-539.

[3] B. A. LaMacchia, *Basis Reduction Algorithms and Subset Sum Problems* SM Thesis, Dept. of Elect. Eng. and Comp. Sci., Massachusetts Institute of Technology, Cambridge, MA (1991).

[4] M.R. Garey, David S.Johnson, *Computer and Tractibility: A Guide to theory of NP-Completeness*, W.H.Freeman and Company (1979)

[5] A. K. Lenstra, H. W. Lenstra, and L. Lovsz, *A hierarchy of polynomial time lattice basis reduction algorithms,* Theoretical Computer Science 53 (1987), 201224.

[6] C. P. Schnorr, *Factoring polynomials with rational coefficients,* Math. Ann. 261 (1982), 515534.