Solving Random Subset Sum Problem by l_p-norm SVP Oracle

Gengran Hu
joint work with Yanbin Pan, Feng Zhang

Key Laboratory of Mathematics Mechanization, NCMIS,
Academy of Mathematics and Systems Science, Chinese Academy of Sciences

PKC2014
March 28, 2014

Outline

1. Lattices and SVP
2. Random Subset Sum Problem
3. Solving RSSP by l_p-norm SVP Oracle

Outline

1. Lattices and SVP
2. Random Subset Sum Problem
3. Solving RSSP by l_p-norm SVP Oracle

Lattices

Definition (Lattice)

Given a matrix $B = (b_{ij}) \in \mathbb{R}^{m \times n}$ with rank n, the lattice $\mathcal{L}(B)$ spanned by the columns of B is

$$\mathcal{L}(B) = \{Bx = \sum_{i=1}^{n} x_i b_i | x_i \in \mathbb{Z}\},$$

where b_i is the i-th column of B.

- Lattices can also be regarded as discrete subgroups of $\mathbb{R}^m$.

Shortest Vector Problem

Definition (l_p-norm SVP)

Given a lattice basis B, the l_p-norm SVP asks to find a nonzero vector in $\mathcal{L}(B)$ with the smallest l_p-norm.

- SVP is one of the most famous computational problems of lattice.
- SVP's hardness is important in proving the security of almost all the lattice-based cryptography.

Shortest Vector Problem

Definition (l_p-norm SVP)

Given a lattice basis B, the l_p-norm SVP asks to find a nonzero vector in $\mathcal{L}(B)$ with the smallest l_p-norm.

- SVP is one of the most famous computational problems of lattice.
- SVP's hardness is important in proving the security of almost all the lattice-based cryptography.

Hardness of SVP

- The l_∞-norm SVP is NP-hard under deterministic reduction.
- However, SVP for other norms can only be proved to be NP-hard under randomized reduction.
  (Ajtai 1998, Micciancio 2001, 2012)

Outline

Subset Sum Problem

Definition (SSP)

Given $\mathbf{a} = (a_1, a_2 \dots a_n)$ in $[1, A]^n$ and $s = \sum_{i=1}^{n} e_i a_i$ where $\mathbf{e} = (e_1 e_2 \dots e_n) \in \{0, 1\}^n$ is independent of $\mathbf{a}$, SSP refers to finding some $\mathbf{c} = (c_1 c_2 \dots c_n) \in \{0, 1\}^n$ s.t. $s = \sum_{i=1}^{n} c_i a_i$ without knowing $\mathbf{e}$.

- SSP is a well-known NP-hard problem.

Subset Sum Problem

Definition (SSP)

Given $\mathbf{a} = (a_1, a_2 \ldots a_n)$ in $[1, A]^n$ and $s = \sum_{i=1}^{n} e_i a_i$ where $\mathbf{e} = (e_1 e_2 \ldots e_n) \in \{0, 1\}^n$ is independent of $\mathbf{a}$, SSP refers to finding some $\mathbf{c} = (c_1 c_2 \ldots c_n) \in \{0, 1\}^n$ s.t. $s = \sum_{i=1}^{n} c_i a_i$ without knowing $\mathbf{e}$.

- SSP is a well-known NP-hard problem.

Random Subset Sum Problem

- When all of the elements in SSP, say $a_1, a_2 \ldots a_n$ are uniformly random over $[1, A]$, SSP becomes RSSP, which is also a significant computational problem.
- The density of such random subset sum instance is defined as

$$\delta = \frac{n}{\log_2 A}.$$

Random Subset Sum Problem

- When all of the elements in SSP, say $a_1, a_2 \dots a_n$ are uniformly random over $[1, A]$, SSP becomes RSSP, which is also a significant computational problem.
- The density of such random subset sum instance is defined as

$$\delta = \frac{n}{\log_2 A}.$$

Hardness of RSSP

The hardness of RSSP is depending on its density:

- If $\delta < 1/n$, RSSP can be efficiently solved by LLL algorithm. (Lagarias & Odlyzko, 1985)
- If $\delta > \Omega(\frac{n}{\log_2 n})$, RSSP can be efficiently solved by dynamic programming.
- The hardest instances of RSSP lie in those with $\delta = 1$. (Impagliazzo & Naor, 1996)

Hardness of RSSP

The hardness of RSSP is depending on its density:

- If $\delta < 1/n$, RSSP can be efficiently solved by LLL algorithm. (Lagarias & Odlyzko, 1985)
- If $\delta > \Omega(\frac{n}{\log_2 n})$, RSSP can be efficiently solved by dynamic programming.
- The hardest instances of RSSP lie in those with $\delta = 1$. (Impagliazzo & Naor, 1996)

Solving RSSP by SVP oracle

Given an l_p-norm SVP oracle, RSSP can be almost solved with:

- $\delta < 0.9408 (p = 2)$.(Coster et al, 1992)
- $\delta < +\infty (p = +\infty)$.
- Q1:How to improve the density bound from 0.9408 to 1 or larger?
- Q2:How to explain the gap between 0.9408 and $+\infty$?

Solving RSSP by SVP oracle

Given an l_p-norm SVP oracle, RSSP can be almost solved with:

- $\delta < 0.9408 (p = 2)$.(Coster et al, 1992)
- $\delta < +\infty (p = +\infty)$.
- Q1:How to improve the density bound from 0.9408 to 1 or larger?
- Q2:How to explain the gap between 0.9408 and $+\infty$?

Solving RSSP by SVP oracle

We answer the second question:

- For $p \in \mathbb{Z}^+, p \geq 2$, given the l_p-norm SVP oracle, almost all RSSP instances can be solved with density δ s.t.

$$\delta < \delta_p = \frac{1}{2^p} \log_2(2^{p+1} - 2) + \log_2(1 + \frac{1}{(2^p - 1)(1 - (\frac{1}{2^{p+1}-2})^{(2^p-1)})}).$$

(Asymptotically, $\delta_p \approx 2^p/(p+2)$)

Solving RSSP by SVP oracle

- The table below gives the values of δ_p for p from two to five:

p	2	3	4	5
δ_p	0.9408	1.4957	2.5013	4.3122

- More specifically, we have $\delta_p > 1 (p \geq 3)$ and $\delta_p \rightarrow +\infty (p \rightarrow +\infty)$.

Solving RSSP by SVP oracle

- The table below gives the values of δ_p for p from two to five:

p	2	3	4	5
δ_p	0.9408	1.4957	2.5013	4.3122

- More specifically, we have $\delta_p > 1 (p \geq 3)$ and $\delta_p \to +\infty (p \to +\infty)$.

Outline

Revisiting RSSP

- An RSSP instance consists of $\mathbf{a} = (a_1, a_2 \dots a_n)$ distributed uniformly in $[1, A]^n$ and $s = \sum_{i=1}^n e_i a_i$ with private $\mathbf{e} = (e_1 e_2 \dots e_n) \in \{0, 1\}^n$.
- The density of this instance is

$$\delta = \frac{n}{\log_2 A}.$$

- Our goal is to find some $\mathbf{c} = (c_1 c_2 \dots c_n) \in \{0, 1\}^n$ s.t. $s = \sum_{i=1}^n c_i a_i$.

Constructing respective lattice

- From RSSP instance, we construct the lattice basis matrix to be

$$B = \begin{pmatrix} 1 & 0 & \dots & 0 & \frac{1}{2} \\ 0 & 1 & \dots & 0 & \frac{1}{2} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & \frac{1}{2} \\ 0 & 0 & \dots & 0 & \frac{1}{2} \\ Na_1 & Na_2 & \dots & Na_n & Ns \end{pmatrix},$$

where $N > \frac{1}{2}(n+1)^{\frac{1}{p}}$ is an positive integer.

Calling SVP oracle

- we see $\mathcal{L}(B)$ contains a corresponding short lattice vector $\mathbf{e}^{'} = (e_1^{'} \ldots e_n^{'}, -\frac{1}{2}, 0)$ with $e_i^{'} = e_i - \frac{1}{2} \in \{-\frac{1}{2}, \frac{1}{2}\}$.
- If SVP oracle returns $\pm\mathbf{e}^{'}$, we can recover our $\mathbf{e}$ from $\pm\mathbf{e}^{'}$.
- In fact, Considering the set $S_n = \{(y_1, y_2 \ldots y_{n+1}, 0)^T | \quad |y_i| = \frac{1}{2}\}$, if our SVP oracle returns an $\mathbf{x} \in S_n$, we can also recover an solution $\mathbf{c}$ of RSSP.
- What if $\mathbf{x} \notin S_n$?

Calling SVP oracle

- we see $\mathcal{L}(B)$ contains a corresponding short lattice vector $\mathbf{e}^{'} = (e_1^{'} \dots e_n^{'}, -\frac{1}{2}, 0)$ with $e_i^{'} = e_i - \frac{1}{2} \in \{-\frac{1}{2}, \frac{1}{2}\}$.
- If SVP oracle returns $\pm\mathbf{e}^{'}$, we can recover our $\mathbf{e}$ from $\pm\mathbf{e}^{'}$.
- In fact, Considering the set $S_n = \{(y_1, y_2 \dots y_{n+1}, 0)^T | \quad |y_i| = \frac{1}{2}\}$, if our SVP oracle returns an $\mathbf{x} \in S_n$, we can also recover an solution $\mathbf{c}$ of RSSP.
- What if $\mathbf{x} \notin S_n$?

Failure Probability

- We fail to solve RSSP if $\mathbf{x} \notin S_n$.
- Denote P the probability of $\mathbf{x} \notin S_n$, we can still almost solve RSSP if $P \leq 1/2^{\Omega(n)}$.

Failure Probability

- We fail to solve RSSP if $\mathbf{x} \notin S_n$.
- Denote P the probability of $\mathbf{x} \notin S_n$, we can still almost solve RSSP if $P \leq 1/2^{\Omega(n)}$.

Failure Probability

- Formally,

$$P = \Pr(\exists \mathbf{x} \quad \text{s.t.} \quad 0 < \|\mathbf{x}\|_p \leq \|\mathbf{e}'\|_p, \mathbf{x} \in \mathcal{L}(B) \backslash S_n).$$

- We can bound P as

$$\begin{aligned} P &\leq \sum_{0<\|x\|_p \leq \|\mathbf{e}'\|_p} \Pr(\mathbf{x} \in \mathcal{L}(B) \backslash S_n) \\ &\leq \max_{0<\|x\|_p \leq \|\mathbf{e}'\|_p} \Pr(\mathbf{x} \in \mathcal{L}(B) \backslash S_n) \cdot \#\{\mathbf{x} \in \mathbb{Z}^{n+1} \| \|\mathbf{x}\|_p \leq \frac{1}{2}(n+1)^{\frac{1}{p}}\} \end{aligned}$$

Failure Probability

Considering any $\mathbf{x} \in \mathcal{L}(B) \backslash S_n$, taking $z_i = x_i + 2x_{n+1}e_i - x_{n+1}$, then $\sum_{i=1}^{n} z_i a_i = 0$ and $\exists j$ s.t. $z_j \neq 0$. Let $z' = -\sum_{i \neq j} z_i a_i / z_j$, then

$$\begin{aligned}
\max_{0<\|x\|_p \leq \|\mathbf{e}'\|_p} \Pr(\mathbf{x} \in \mathcal{L}(B) \backslash S_n) &\leq \Pr(\sum_{i=1}^{n} z_i a_i = 0, z_j \neq 0) \\
&= \Pr(a_j = z') \\
&= \sum_{k=1}^{A} \Pr(a_j = k) \cdot \Pr(z' = k) \\
&= \frac{1}{A} \sum_{k=1}^{A} \Pr(z' = k) \\
&\leq \frac{1}{A}.
\end{aligned}$$

Failure Probability

- Thus we've obtained

$$P \leq \frac{1}{A} \cdot \#\{\mathbf{x} \in \mathbb{Z}^{n+1} | \|\mathbf{x}\|_p \leq \frac{1}{2}(n+1)^{\frac{1}{p}}\}$$

Failure Probability

- If we find suitable u_p s.t. $\#\{\mathbf{x} \in \mathbb{Z}^n | \|\mathbf{x}\|_p \leq \frac{1}{2}n^{\frac{1}{p}}\} \leq 2^{u_p n}$ for every n, then

$$P \leq \frac{2^{u_p(n+1)}}{A} = \frac{2^{u_p(n+1)}}{2^{(\frac{1}{\delta}n)}}.$$

- When $\delta < 1/u_p \triangleq \delta_p$, $P \leq 1/2^{\Omega(n)}$, thus we can solve RSSP with high probability.

Estimating integer points in l_p ball

- We can find an upper bound

$$u_p = \frac{1}{2^p}\log_2(2^{p+1}-2) + \log_2\left(1 + \frac{1}{(2^p-1)(1-(\frac{1}{2^{p+1}-2})^{(2^p-1)})}\right)$$

(Asymptotically, $u_p \approx (p+2)/2^p$) to make sure

$$\#\{\mathbf{x} \in \mathbb{Z}^n \mid \|\mathbf{x}\|_p \leq \frac{1}{2}n^{\frac{1}{p}}\} \leq 2^{u_p n}.$$

- On the other hand, for large enough n, there is a lower bound:

$$\#\{\mathbf{x} \in \mathbb{Z}^n \mid \|\mathbf{x}\|_p \leq \frac{1}{2}n^{\frac{1}{p}}\} \geq \frac{1}{\Omega(n^{3/2})}2^{l_p n}.$$

Estimating integer points in l_p ball

- The u_p and l_p are so close:

p	2	3	4	5
u_p	1.0613	0.6686	0.3998	0.2319
l_p	1.0630	0.6686	0.3998	0.2319

- In fact, we can prove the error bound:

$$\frac{u_p - l_p}{u_p} < (2^p - 1)^{-(2^p-1)}.$$

Conclusion

- Since RSSP with density = 1 is the hardest and $\delta_p > 1$ when $p \geq 3$, we have a probabilistic reduction from RSSP to l_p-norm SVP($p \geq 3$).

Open Problems

- Proving RSSP is NP-hard will lead to another probabilistic reduction to show l_p-norm SVP($p \geq 3$) is NP-hard.
- Finding SVP algorithm for l_p-norm is also interesting.

Thanks!