1 SIS Lattices

In this lecture we give properties and applications of SIS (short integer solution) lattices. We first recall the SIS problem.

Definition 1.1 (Shortest Integer Solution Problem). For a positive integer modulus q, dimensions n, m and a norm bound $\beta > 0$, the $\mathsf{SIS}_{n,q,\beta,m}$ problem is defined as follows: given uniformly random $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$, find a nonzero "short" solution $\mathbf{z} \in \mathbb{Z}^m$ such that $\mathbf{Az} = \mathbf{0} \in \mathbb{Z}_q^n$ and $\|\mathbf{z}\| \leq \beta$.

Equivalently, the goal is to find a non-zero vector of norm at most β in the following integer "SIS lattice" (it is easy to verify that this set is a discrete additive subgroup):

$$\mathcal{L}^\perp(\mathbf{A}) := \{\mathbf{z} \in \mathbb{Z}^m : \mathbf{Az} = \mathbf{0}\}.$$

Borrowing a term from coding theory, matrix $\mathbf{A}$ is often called a *parity-check matrix* for the lattice $\mathcal{L}^\perp(\mathbf{A})$.

We begin with some mathematical properties of SIS lattices. First, these lattices are called "q-ary" because they contain every integer vector whose entries are all multiples of q:

$$q\mathbb{Z}^m \subseteq \mathcal{L}^\perp(\mathbf{A}) \subseteq \mathbb{Z}^m.$$

So, a vector's membership in $\mathcal{L}^\perp(\mathbf{A})$ is determined solely by its entries modulo q.

Cosets. Borrowing another term from coding theory let $\mathbf{y} \in \mathbb{Z}_q^n$ be a "syndrome" in the image of $\mathbf{A}$, i.e., there exists some $\mathbf{x} \in \mathbb{Z}^m$ such that $\mathbf{Ax} = \mathbf{y}$. Then we can define the corresponding lattice coset

$$\mathcal{L}^\perp_{\mathbf{y}}(\mathbf{A}) := \{\mathbf{x}' \in \mathbb{Z}^m : \mathbf{Ax}' = \mathbf{y}\} = \mathbf{x} + \mathcal{L}^\perp(\mathbf{A}),$$

where the equality holds because every $\mathbf{x}' \in \mathbf{x} + \mathcal{L}^\perp(\mathbf{A})$ satisfies $\mathbf{Ax}' = \mathbf{Ax} = \mathbf{y}$, and for any $\mathbf{x}' \in \mathbb{Z}^m$ such that $\mathbf{Ax}' = \mathbf{y}$, we have $\mathbf{x}' = \mathbf{x} + (\mathbf{x}' - \mathbf{x}) \in \mathbf{x} + \mathcal{L}^\perp(\mathbf{A})$ because $\mathbf{A}(\mathbf{x}' - \mathbf{x}) = \mathbf{0}$.

Determinant. By the bijective correspondence between integer cosets and syndromes in the image of $\mathbf{A}$, we have that

$$\det(\mathcal{L}^\perp(\mathbf{A})) = |\mathbb{Z}^m / \mathcal{L}^\perp(\mathbf{A})| = |\text{Image}(\mathbf{A})| \leq q^n,$$

with equality if the image of $\mathbf{A}$ is all of $\mathbb{Z}_q^n$ (i.e., the columns of $\mathbf{A}$ generate $\mathbb{Z}_q^n$), which will be the setting we are almost always interested in.

If q is prime, then $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ generates all of $\mathbb{Z}_q^n$ if and only if it contains an $n \times n$ submatrix that is invertible modulo q. (It is not too hard to show that when, say, $m \geq (1+\delta)n$ for a constant $\delta > 0$, this holds with high probability over the uniformly random choice of $\mathbf{A}$.) However, if q is composite, this condition is sufficient but not necessary. For example, $\mathbf{A} = (2, 3) \in \mathbb{Z}_6^{1 \times 2}$ generates all of $\mathbb{Z}_6$, even though neither 2 nor 3 are invertible modulo 6.

Minimum distance. The minimum distance can be bounded using Minkowski's Theorem and the above bound on the determinant:

$$\begin{aligned}\lambda_1(\mathcal{L}^\perp(\mathbf{A})) &\leq \sqrt{m} \cdot \det(\mathcal{L}^\perp(\mathbf{A}))^{1/m} \\ &\leq \sqrt{m} \cdot q^{n/m}.\end{aligned}$$

However, observe that we are not required to use all m dimensions: by fixing some of the vectors' coordinates to zero (equivalently, dropping some columns of $\mathbf{A}$), we can instead work with a lattice of any dimension less than m. Assuming that the original dimension is any $m = \Omega(n \log q)$, the above bound is optimized for some dimension $\Theta(n \log q)$, and yields a minimum distance of $O(\sqrt{n \log q})$.

Also recall from last lecture that if $m > n \log_2 q$, then there exists a nonzero $\{0, \pm 1\}^m$-vector in $\mathcal{L}^\perp(\mathbf{A})$, so the ℓ_p minimum distance is $\lambda_1^{(p)}(\mathcal{L}^\perp(\mathbf{A})) \le m^{1/p}$ for any finite p, and $\lambda_1^{(\infty)}(\mathcal{L}^\perp(\mathbf{A})) \le 1$.

Equivalent representations and lattices. We observe that many different matrices $\mathbf{A}$ can define (essentially) the same lattice. The proofs of the following two lemmas are straightforward exercises.

Lemma 1.2. *Let $\mathbf{H} \in \mathbb{Z}_q^{n \times n}$ be invertible. Then*

$$\mathcal{L}^\perp(\mathbf{H} \cdot \mathbf{A}) = \mathcal{L}^\perp(\mathbf{A}).$$

Lemma 1.3. *Let $\mathbf{A}'$ be a column permutation of $\mathbf{A}$, i.e., $\mathbf{A}' = \mathbf{A}\mathbf{P}$, where $\mathbf{P} \in \{0,1\}^{m \times m}$ is a permutation matrix. Then*

$$\mathcal{L}^\perp(\mathbf{A}') = \mathbf{P}^{-1} \cdot \mathcal{L}^\perp(\mathbf{A}),$$

i.e., the lattice $\mathcal{L}^\perp(\mathbf{A}')$ is just a coordinate permutation of the lattice $\mathcal{L}^\perp(\mathbf{A})$, so it has the same determinant, successive minima, etc.

More generally, if $\mathbf{T} \in \mathbb{Z}_q^{m \times m}$ is invertible, then

$$\mathcal{L}^\perp(\mathbf{A} \cdot \mathbf{T}) = \mathbf{T}^{-1} \cdot \mathcal{L}^\perp(\mathbf{A}).$$

However, for general $\mathbf{T}$ this is not merely a coordinate permutation, so it may change the lattice's essential geometric properties.

(Canonical) basis. Suppose that $\mathbf{A}$ contains an $n \times n$ submatrix that is invertible modulo q. Without loss of generality, by permuting columns (Lemma 1.3), we can write $\mathbf{A} = [\mathbf{H} \mid \mathbf{A}']$ where $\mathbf{H} = \mathbb{Z}_q^{n \times n}$ is invertible. Then by Lemma 1.2, without changing the lattice we can write the parity-check matrix as $[\mathbf{I}_n \mid \bar{\mathbf{A}}]$ where $\bar{\mathbf{A}} = \mathbf{H}^{-1}\mathbf{A}' \in \mathbb{Z}_q^{n \times (m-n)}$. Borrowing another coding theory term, this is called the "systematic form," or more commonly "(Hermite) normal form" in the lattice context. Observe that it is a slightly more compact representation, because we can treat the identity submatrix as implicit.

We then have the following basis for the lattice $\mathcal{L} = \mathcal{L}^\perp([\mathbf{I}_n \mid \bar{\mathbf{A}}]) \subseteq \mathbb{Z}^m$:

$$\mathbf{B} = \begin{bmatrix} q\mathbf{I}_n & -\bar{\mathbf{A}} \\ \mathbf{0} & \mathbf{I}_{m-n} \end{bmatrix} \in \mathbb{Z}^{m \times m}.$$

(Formally, because $-\bar{\mathbf{A}}$ is a matrix over $\mathbb{Z}_q$ (not $\mathbb{Z}$), the basis uses any integer-matrix representative of it, e.g., with entries in $\{0, 1, \ldots, q-1\}$.) We can confirm that $\mathbf{B}$ is indeed a basis of $\mathcal{L}$:

- First, the columns are linearly independent, because $\mathbf{B}$ is upper triangular with nonzero diagonal entries.
- Second, every column vector of $\mathbf{B}$ is in the lattice, because $[\mathbf{I}_n \mid \bar{\mathbf{A}}] \cdot \mathbf{B} = \mathbf{0}$. But do these columns form a *basis* of $\mathcal{L}$?
- We have $\det(\mathbf{B}) = q^n = \det(\mathcal{L})$, so $\mathbf{B}$ is in fact a basis.

For a (full-rank) integer lattice, a basis of the above form is said to be in "Hermite normal form." The precise requirements are that the basis be upper triangular with positive entries on the diagonal, and that every other entry is reduced modulo the diagonal entry of its row. It turns out that such a basis is uniquely specified for any lattice, and is efficiently computable from any basis of the same lattice. Therefore, the Hermite normal form basis acts as a kind of "canonical" representation of an integer lattice. (See [MW01, Mic01] for further details.)

2 Applications of SIS

2.1 Collision-Resistant Hashing

Definition 2.1. A function $f\colon \mathcal{D} \to \mathcal{R}$ where $|\mathcal{D}| > |\mathcal{R}|$ is called a *hash function*. A set $\mathcal{F} = \{f_a\colon \mathcal{D} \to \mathcal{R}\}$ of such functions is called a *hash function family*.

Since the domain $\mathcal{D}$ is larger than the range $\mathcal{R}$, by the pigeonhole principle any function $f\colon \mathcal{D} \to \mathcal{R}$ must have a *collision*, i.e., some distinct $x_1, x_2 \in \mathcal{D}$ where $f_a(x_1) = f_a(x_2)$. A hash function family is said to be *collision resistant* if it is infeasible to find a collision in a randomly chosen function from the family. To make this asymptotic, we parameterize the family (and the domain and range) by a security parameter $\lambda \in \mathbb{N}$.

Definition 2.2. A hash function family $\mathcal{F}_\lambda$ is *collision resistant* if for any efficient adversary $\mathcal{A}$,

$$\Pr_{f_a \leftarrow \mathcal{F}_\lambda}[\mathcal{A}(f_a) \text{ outputs a collision in } f_a] = \mathrm{negl}(\lambda).$$

Collision resistance is a central property in cryptography. It is often used to "compress" data into a small digest, so that only the original data can be presented as consistent with that digest (because otherwise a collision has been found, which is infeasible). This can be used for detecting errors or modifications in data, authenticating arbitrary pieces of large data sets, and much more.

We can construct a collision-resistant hash function family directly from the SIS problem.

Theorem 2.3. *Let $n = \lambda$ (the security parameter), $q \geq 2$ be arbitrary, $m > n \log_2 q$, and $\beta = \sqrt{m}$. If $\mathsf{SIS}_{n,q,\beta,m}$ is hard, then the SIS function family $\mathcal{F} = \{f_{\mathbf{A}}\colon \{0,1\}^m \to \mathbb{Z}_q^n : \mathbf{A} \in \mathbb{Z}_q^{n \times m}\}$, defined as $f_{\mathbf{A}}(\mathbf{x}) := \mathbf{A}\mathbf{x} \in \mathbb{Z}_q^n$, is a collision-resistant hash function family.*

We note that the domain need not consist merely of *binary* strings; we can generalize to vectors of "small" integers with a suitable adjustment of the parameters.

Proof. First, since $|\mathcal{D}| = 2^m > q^n = |\mathcal{R}|$, we indeed have a family of hash functions. Let $\mathcal{A}$ be an efficient adversary that attempts to break the collision resistance of the hash function family. We describe an efficient reduction that attempts to solve SIS using $\mathcal{A}$. Given an SIS instance $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$, the reduction gives $\mathbf{A}$ to the adversary $\mathcal{A}$, which potentially outputs a collision in $f_{\mathbf{A}}$, i.e., distinct $\mathbf{x}, \mathbf{x}' \in \{0,1\}^m$ such that $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}'$. The reduction outputs $\mathbf{z} = \mathbf{x} - \mathbf{x}' \in \{0, \pm 1\}^m$ as its potential SIS solution.

Suppose that $\mathcal{A}$ was successful in outputting a collision. Then $\mathbf{z} \neq \mathbf{0}$, $\mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{x}' = \mathbf{0}$, and $\|\mathbf{z}\| \leq \sqrt{m} = \beta$, so the reduction successfully solves its given SIS instance. Because SIS is hard by assumption, the adversary $\mathcal{A}$ must have only negligible probability of success, as desired. □

2.2 Regular One-Way Function Family

Definition 2.4. A function family $\mathcal{F} = \{f\colon \mathcal{D} \to \mathcal{R}\}$ is *ε-regular* if

$$(f, f(x)) \approx_\varepsilon (f, y),$$

where $f \leftarrow \mathcal{F}, x \leftarrow \mathcal{D}, y \leftarrow \mathcal{R}$, and $\approx_\varepsilon$ denotes that the two distributions have statistical distance at most ε.

Definition 2.5. A function family $\mathcal{F}_\lambda = \{f_a\colon \mathcal{D}_\lambda \to \mathcal{R}_\lambda\}$ is *one way* if for all efficient adversaries $\mathcal{A}$,

$$\Pr_{f_a \leftarrow \mathcal{F}_\lambda, y \leftarrow \mathcal{R}_\lambda}[\mathcal{A}(f_a, y) \text{ outputs } x \in \mathcal{D}_\lambda \text{ s.t. } f_a(x) = y] = \mathrm{negl}(\lambda).$$

In particular, regularity (for tiny $\varepsilon \geq 0$) implies that for a randomly chosen function f_a from the family, almost every range value $y \in \mathcal{R}$ has a preimage, i.e., an $x \in \mathcal{D}$ for which $f_a(x) = y$. One-wayness can be seen as saying that finding such a pre-image is hard.

Lemma 2.6. *For $n \geq 1$, any $q \geq 2$, and any $m \geq (1+\delta)n\log_2 q$ for constant $\delta > 0$, the* SIS *family $\mathcal{F} = \{f_{\mathbf{A}}\colon \{0,1\}^m \to \mathbb{Z}_q^n\colon \mathbf{A} \in \mathbb{Z}_q^{n\times m}\}$ is $2^{-\Omega(n)}$-regular.*

Lemma 2.7. *The above* SIS *function family is one-way if $\mathsf{SIS}_{n,q,\beta,m+1}$ is hard, where $\beta = \sqrt{m+1}$.*

Proof. Let $\mathcal{A}$ be an efficient adversary that attempts to break the one-wayness of the family. We describe an efficient reduction $\mathcal{B}$ that attempts to solve $\mathsf{SIS}_{n,q,\beta,m+1}$ using $\mathcal{A}$. The reduction works as follows: given an instance $\mathbf{A}' \in \mathbb{Z}_q^{n\times(m+1)}$, $\mathcal{B}$ parses it as $\mathbf{A}' = [\mathbf{A} \mid \mathbf{y}]$ where $\mathbf{y} \in \mathbb{Z}_q^n$. Then $\mathcal{B}$ invokes $\mathcal{I}(f_{\mathbf{A}}, \mathbf{y})$, which potentially outputs some $\mathbf{x} \in \{0,1\}^m$ such that $\mathbf{A}\mathbf{x} = \mathbf{y}$. Then $\mathcal{B}$ outputs $\mathbf{z} = \binom{\mathbf{x}}{-1} \in \mathbb{Z}^{m+1}$ as a potential solution to the SIS instance.

Observe that whenever $\mathcal{A}$ successfully inverts, we have $\mathbf{A}\mathbf{x} = \mathbf{y}$ and hence $\mathbf{A}'\mathbf{z} = \mathbf{A}\mathbf{x} - \mathbf{y} = \mathbf{0}$; moreover, $\mathbf{z} \neq \mathbf{0}$ $\|\mathbf{z}\| \leq \sqrt{m+1} = \beta$, so $\mathcal{B}$ succeeds at solving its SIS instance. Because SIS is hard by assumption, $\mathcal{A}$ must have only negligible probability of success, as needed. □

Lemmas 2.6 and 2.7 together yield *the central method* (up to some minor variations) of constructing secret/public key pairs in lattice-based cryptography, which goes back to Ajtai's original work [Ajt96]. The procedure operates as follows:

1. Generate (or be given, from a trusted source) a uniformly random matrix $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$.
2. Choose a secret key $\mathbf{x} \leftarrow \{0,1\}^m = \mathcal{D}$ uniformly at random.
3. Let $\mathbf{y} = \mathbf{A}\mathbf{x} = f_{\mathbf{A}}(\mathbf{x})$ and let the public key be $\mathbf{A}' = [\mathbf{A} \mid \mathbf{y}] \in \mathbb{Z}_q^{n\times(m+1)}$ (or just $\mathbf{y}$, if others already know $\mathbf{A}$).

We can even do the same with multiple vectors $\mathbf{x}_i$, which we group into the columns of a secret-key matrix $\mathbf{X} \in \{0,1\}^{m\times k}$, yielding the public-key matrix $\mathbf{A}' = [\mathbf{A} \mid \mathbf{Y}] \in \mathbb{Z}_q^{n\times(m+k)}$ where $\mathbf{Y} = \mathbf{A}\mathbf{X} \in \mathbb{Z}_q^{n\times k}$.

The procedure above has the following properties:

- $\mathbf{A}'$ is statistically close to uniformly random, by regularity (Lemma 2.6).
- Therefore, the adversary cannot find any "short" $\mathbf{z} \in \mathcal{L}^\perp(\mathbf{A}')$, by the presumed hardness of SIS.
- However, the generating party *does* know such a short vector, namely, $\mathbf{z} = \binom{\mathbf{x}}{-1}$.

Looking ahead, this asymmetry between the user and the adversary will allow us to build more advanced cryptographic primitives, like digital signatures and public-key encryption, among many others.

Proof of Lemma 2.6. This follows immediately from Lemmas 2.9 and 2.10 below. □

Definition 2.8. A function family $\mathcal{F} = \{f_a \colon \mathcal{D} \to \mathcal{R}\}$ is *universal* if for all distinct $x, x' \in \mathcal{D}$,

$$\Pr_{f_a \leftarrow \mathcal{F}}[f_a(x) = f_a(x')] = \frac{1}{|\mathcal{R}|}.$$

Lemma 2.9. *The* SIS *function family with domain $\mathcal{D} = \{0,1\}^m$ is universal.*

Proof. Fix arbitrary distinct $\mathbf{x}, \mathbf{x}' \in \{0,1\}^m$. Without loss of generality (by rearranging coordinates and swapping $\mathbf{x}, \mathbf{x}'$ if necessary), we have that $x_1 - x'_1 = 1$. Then

$$\Pr_{\mathbf{A}}[\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}'] = \Pr_{\mathbf{A}}[\mathbf{A}(\mathbf{x} - \mathbf{x}') = \mathbf{0}] = \Pr_{\mathbf{A}}[\mathbf{a}_1 = \sum_{i=2}^{m} \mathbf{a}_i(x'_i - x_i)] = q^{-n} = \frac{1}{|\mathcal{R}|},$$

where the second-to-last equality follows by arbitrarily fixing $\mathbf{a}_i$ for $i = 2, \ldots, m$, and considering the probability over $\mathbf{a}_1$ alone. □

Lemma 2.10 (Leftover Hash Lemma [HILL99]). *Any universal hash family $\mathcal{F}$ of functions from $\mathcal{D}$ to $\mathcal{R}$ is ε-regular for $\varepsilon = \sqrt{|\mathcal{R}|/|\mathcal{D}|}$.*

Proof. Because the first components of the two distributions in Definition 2.4 are a uniformly random function $f \leftarrow \mathcal{F}$, the statistical distance in question is simply the expected (i.e., average) statistical distance between $f(x)$ and y (where $x \leftarrow \mathcal{D}$ and $y \leftarrow \mathcal{R}$), over the choice of f.

For any fixed hash function $f \in \mathcal{F}$, let D_f be the distribution of $f(x)$. The squared ℓ_2 norm of this distribution is

$$\|D_f\|_2^2 := \sum_{y \in \mathcal{R}} D_f(y)^2 = \Pr_{x,x' \leftarrow \mathcal{D}}[f(x) = f(x')] \leq \frac{1}{|\mathcal{D}|} + \Pr_{x,x'}[f(x) = f(x') \mid x \neq x'].$$

Taking the expectation over the uniformly random choice of $f \in \mathcal{F}$, and using universality, we have that

$$\mathop{\mathbb{E}}_{f \leftarrow \mathcal{F}}\left[\|D_f\|_2^2\right] \leq \frac{1}{|\mathcal{D}|} + \frac{1}{|\mathcal{R}|}.$$

So, by the relation between the ℓ_2 and ℓ_1 norms, the fact that the values of each D_f sum to unity, and Jensen's inequality, the expected statistical distance between D_f and the uniform distribution is

$$\begin{aligned}
\mathop{\mathbb{E}}_{f \leftarrow \mathcal{F}}\Big[\sum_{y \in \mathcal{R}} \big|D_f(y) - \frac{1}{|\mathcal{R}|}\big|\Big] &\leq \mathop{\mathbb{E}}_{f}\Big[|\mathcal{R}|^{1/2} \cdot \Big(\sum_{y \in \mathcal{R}} \big(D_f(y) - \frac{1}{|\mathcal{R}|}\big)^2\Big)^{1/2}\Big] \\
&= |\mathcal{R}|^{1/2} \cdot \mathop{\mathbb{E}}_{f}\Big[\Big(\sum_{y \in \mathcal{R}} D_f(y)^2 - \frac{1}{|\mathcal{R}|}\Big)^{1/2}\Big] \\
&\leq |\mathcal{R}|^{1/2} \cdot \mathop{\mathbb{E}}_{f}\Big[\|D_f\|_2^2 - \frac{1}{|\mathcal{R}|}\Big]^{1/2} \\
&\leq \sqrt{|\mathcal{R}|/|\mathcal{D}|}.
\end{aligned}$$

□

References

[Ajt96] M. Ajtai. Generating hard instances of lattice problems. *Quaderni di Matematica*, 13:1–32, 2004. Preliminary version in STOC 1996. Page 4.

[HILL99] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, 1999. Page 5.

[Mic01] D. Micciancio. Improving lattice based cryptosystems using the Hermite normal form. In *CaLC*, pages 126–145. 2001. Page 3.

[MW01] D. Micciancio and B. Warinschi. A linear space algorithm for computing the Hermite normal form. In *ISSAC*, pages 231–236. 2001. Page 3.