RANDOM LATTICES: THEORY AND PRACTICE

YOSHINORI AONO, THOMAS ESPITAU, AND PHONG Q. NGUYEN

ABSTRACT. In recent years, there has been renewed interest in random lattices, both from a mathematical and an algorithmic point of view. In this article, we review the main properties of random lattices, compare them with experimental results, and formulate heuristics. For instance, we apply our heuristics to Darmstadt's Shortest Vector Problem challenges, and compare recent theoretical results of Kim and Venkatesh on the number of reduced bases with low-dimensional experiments.

1. INTRODUCTION

The set of full-rank lattices in $\mathbf{R}^n$ of unit co-volume is identified with the group $\mathcal{L}_n = \mathrm{SL}(n, \mathbf{R})/\mathrm{SL}(n, \mathbf{Z})$. In 1945, Siegel [23] introduced a "natural" measure μ_n over $\mathcal{L}_n$ such that $\mu_n(\mathcal{L}_n)$ is finite to prove the existence of dense lattice packings: μ_n is the projection of the Haar measure of $\mathrm{SL}(n, \mathbf{R})$ over $\mathcal{L}_n$, normalized so that $\mu_n(\mathcal{L}_n) = 1$. Accordingly, a random lattice is a unit-volume lattice in $\mathbf{R}^n$ chosen with distribution μ_n.

In the 50s, several results on random lattices appeared, such as [16, 17, 18, 10, 19], but the topic had essentially disappeared from the mathematical literature until Södergren published [25, 24] a few years ago, where he studied the configuration (lengths and angles) of shortest vectors in a random lattice, revisiting Rogers [18].

On the other hand, there has been growing interest in random lattices from the algorithmic community in the past fifteen years. After establishing the first worst-case to average-case reductions over lattices [2], Ajtai used Siegel's measure to formulate conjectures [3] on the hardness of lattice problems, and showed in an unpublished manuscript [1] how to efficiently sample Siegel's distribution. Goldstein and Mayer [7] independently presented a simple efficiently samplable distribution, and showed that it converges in a weak sense to Siegel's distribution: this discrete distribution has been very popular among lattice practitioners [13, 6, 4] to benchmark lattice algorithms, which is useful to assess the concrete security of lattice-based cryptosystems. In fact, it is even used in an online contest [20] called *The SVP Challenge*: this contest asks to find extremely short vectors in explicit lattices of increasing dimension.

Random lattices have also been used to improve our understanding of the behaviour of lattice algorithms: if $h(n)$ denotes the radius of the unit-volume n-dimensional ball, *i.e.* $h(n) = 1/v_n^{1/n}$ where v_n is the volume of the n-dimensional unit ball, Chen and Nguyen [4] used the heuristic estimate $h(n)$ for a random lattice to predict approximately the behaviour of the BKZ reduction algorithm [21] for high blocksizes. This is somehow reminiscent of the situation of the best integer factoring algorithms like the number field sieve or the quadratic sieve, where the complexity analysis assumes heuristically that certain random numbers produced by the algorithm have the same probability of being smooth as a uniformly-distributed random number of the same bit-length.

this work has been supported in part by the european union as H2020 programme under grant agreement number ERC-669891 and JSPS KAKENHI Grant Numbers 16H02780 and 16H02830.

Though the interest in random lattices is growing, it appears that mathematical results such as [18, 25, 24] are not widely known in the algorithmic community. And several papers such as [3] give properties of random lattices without proofs, such as the existence of a basis with norms asymptotically close to $h(n)$.

Our results. We aim at clarifying what is known and not known for random lattices, and perform experiments to check theoretical results and heuristics. In Sect. 2, we show that the expectation of the first minimum of a random lattice is asymptotically equivalent to the Gaussian heuristic estimate $h(n)$. We introduce a heuristic (Lemma 2.6) to guess the distribution of the first minimum. In Sect. 3, we show how some of the previous results on random lattices can be adapted to random integer lattices, and check the experimental validity of our first minimum heuristic. We apply our heuristics to the SVP challenge, such as guessing when it is likely that a shortest vector has already been found. In Sect. 4, we focus on the low-dimensional case. We show that the expectation of various quantities related to successive minima can be explicitly computed in dimension two. In low dimension, we enumerate reduced bases to investigate the gap between experiments and theoretical results [22, 8] on the number of reduced bases in random lattices: it turns out that the experimental expectation is consistent with theory, but not so with the standard deviation. Sect. 5 deals with random bases of a given lattice: we compare the main generation algorithms used in the literature, and highlight their issues. This allows us to study the *dark bases* phenomenon introduced by Kim [22]: these are LLL-reduced bases which are output by LLL with negligible probability in practice. We uncover an extreme case: lattices of dimension ≤ 10 which only have two LLL bases, but only one can be output when taking random bases as input.

2. Random Real Lattices

2.1. **Siegel's measure.** The moduli space of full-rank lattices in $\mathbf{R}^n$ of unit covolume is homeomorphic to the topological quotient $\mathcal{L}_n = \mathrm{SL}(n, \mathbf{R})/\mathrm{SL}(n, \mathbf{Z})$. In 1945, Siegel [23] proved that this quotient is of finite mass under the the projection of the Haar measure of $\mathrm{SL}(n, \mathbf{R})$ over $\mathcal{L}_n$, yielding a natural probability distribution μ_n over $\mathcal{L}_n$. By construction this distribution is translation-invariant, that is for any measurable set $A \subseteq \mathcal{L}_n$ and all $f \in \mathrm{SL}(n, \mathbf{Z})$, $\mu_n(A) = \mu_n(Af)$. A *random (real) lattice* is a unit-covolume lattice in $\mathbf{R}^n$ drawn under the probability distribution μ_n. Macbeath and Rogers [10] proved the following averaging result on μ_n, generalizing the result appearing in [23] for Riemann-integrable function:

Theorem 1. *Let f be a compactly supported Lebesgue-integrable function over $\mathbf{R}^n$. Then: $\int_{\mathbf{R}^n} f(\mathbf{x}) d\mathbf{x} = \int_{L \in \mathcal{L}_n} f(L \setminus \{0\}) d\mu_n$, where $f(L \setminus \{0\}) = \sum_{\mathbf{x} \in L, \mathbf{x} \neq 0} f(\mathbf{x})$.*

In particular, if we take f to be the characteristic function of a bounded measurable set C, then $\int_{\mathbf{R}^n} f(\mathbf{x}) d\mathbf{x}$ is the volume of C, and $f(L \setminus \{0\})$ is the number of non-zero points in $L \cap C$. Hence, Theorem 1 states that the average number of non-zero lattice points in C is equal to $\mathrm{vol}(C)$.

2.2. **Higher Moments.** Theorem 1 is actually an estimate of the first moment of $f(L \setminus \{0\})$, when f is the characteristic function of a bounded measurable set C. If C is further assumed to be symmetrical in 0, Rogers [18] bounded arbitrary-order moments as follows:

Theorem 2. *Let f be the characteristic function of a measurable set C whose volume is V, and symmetric with respect to 0. Then, provided that $n \geq [\frac{1}{4}k^2] + 3$:*

$$0 \leq \int_{L \in \mathcal{L}_n} f(L \setminus \{0\})^k d\mu_n - 2^k e^{-V/2} \sum_{r=0}^{\infty} \frac{r^k}{r!}(V/2)^r \\ \leq \left(2(3)^{[k^2/4]}(\sqrt{3/4})^n + 21(5)^{[k^2/4]}(\frac{1}{2})^n\right)(V+1)^k.$$

Rogers deduces the following corollary:

Corollary 3. *Let C be a measurable set of fixed volume V, symmetric with respect to 0. Then the number N_n of pairs of points $\pm\mathbf{x}$ of a lattice $L \in \mathcal{L}_n$ in C has a limit distribution, as n grows to infinity, which is the Poisson distribution with mean $V/2$.*

To see this, recall that the Poisson distribution of mean $V/2$ is a discrete distribution whose probability mass function is $\lambda^r e^{-\lambda}/r!$ where $\lambda = V/2$, and whose k-th moment is therefore:

$$\sum_{r=0}^{\infty} r^k \lambda^r e^{-\lambda}\frac{1}{r!} = e^{-\lambda}\sum_{r=0}^{\infty} r^k \lambda^r \frac{1}{r!}.$$

Let N_n be the number of pairs of points $\pm\mathbf{x}$ of a random lattice $L \in \mathcal{L}_n$ in C. Then $f(L \setminus \{0\}) = 2N_n$ so $f(L \setminus \{0\})^k = 2^k N_n^k$. Hence:

$$0 \leq E(N_n^k) - e^{-V/2}\sum_{r=0}^{\infty} \frac{r^k}{r!}(V/2)^r \leq \{2(3)^{[k^2/4]}(\sqrt{3/4})^n + 21(5)^{[k^2/4]}(\frac{1}{2})^n\}((V+1)/2)^k.$$

Now, let k and V be fixed. Then, as n grows to infinity, the right-hand term converges to zero. Therefore, $E(N_n^k)$ converges, and its limit is:

$$e^{-V/2}\sum_{r=0}^{\infty} \frac{r^k}{r!}(V/2)^r = e^{-\lambda}\sum_{r=0}^{\infty} \frac{r^k}{r!}\lambda^r,$$

which is exactly the k-th moment of the Poisson distribution of mean λ. Since the Poisson distribution is uniquely determined by its moments, the method of moments implies that (N_n) converges in distribution to a Poisson-distributed random variable X of mean $\lambda = V/2$, which proves Cor. 3.

Let N_n be the number of pairs of points $\pm\mathbf{x}$ of a random lattice $L \in \mathcal{L}_n$, lying in C, then N_n converges in distribution towards a Poisson distribution of CDF $\frac{\Gamma(\lfloor k+1 \rfloor, \lambda)}{\lfloor k \rfloor !}$, or $e^{-\lambda}\sum_{i=0}^{\lfloor k \rfloor} \frac{\lambda^i}{i!}$ or $Q(\lfloor k+1 \rfloor, \lambda)$. As a direct application, we have:

Corollary 4. *Let $\alpha > 0$. Then for a random unit-volume lattice L, $\Pr(\lambda_1(L) \leq \alpha^{1/n}h(n))$ converges to $1 - e^{-\alpha/2}$ as n grows to infinity.*

Proof. Let C be the centered ball of volume α: its radius is $\alpha^{1/n}h(n)$. Then, $\lambda_1(L)\alpha^{1/n}h(n)$ if and only if there is no non-zero lattice point in C. And the CDF of the Poisson distribution at 0 is $e^{-\alpha/2}$. □

Using $\alpha = 1$, we obtain that $\Pr(\lambda_1(L) \leq h(n))$ converges to $1 - e^{-1/2} \approx 39\%$. And using $\alpha = 2\ln 2$, $\Pr(\lambda_1(L) \leq (2\ln 2)^{1/n}h(n))$ converges to $1/2$.

2.3. **Distribution of the first minimum.** Sodergren noticed in [25] that Cor. 3 shows the following:

Theorem 5. *if L is a random unit-volume full-rank lattice in $\mathbf{R}^n$, then the volume of the ball of radius $\lambda_1(L)$ has a limit distribution, which is the exponential distribution with mean 2, that is with parameter $\lambda' = 1/2$.*

Proof. Let V_n be the volume of the ball of radius $\lambda_1(L)$. Let $x > 0$ and let C be the ball of volume x. Then $V_n \leq x$ if and only if the ball C contains a non-zero lattice point, that is $N_n \geq 1$ where N_n is defined as in Cor. 3. By Cor. 3, $\Pr(N_n = 0)$ converges to $e^{-x/2}$ as n grows to infinity, so $\Pr(V_n \leq x)$ converges to $1 - e^{-x/2}$ as n grows to infinity. Hence, the CDF of V_n converges towards to the CDF of the exponential distribution with parameter $\lambda' = 1/2$, at every point of continuity. □

This suggests that $\lambda_1(L)$ might have a distribution close to the following:

Lemma 6. *Let X_n be a random variable such that the volume of the n-dimensional ball of radius X_n has exponential distribution with mean 2. Then X_n follows a Weibull distribution of shape parameter n and scale parameter $2^{1/n}h(n)$. Then:*

$$\mathrm{E}(X_n) = 2^{1/n}\Gamma(1+1/n)h(n) = (1 + (\ln 2 - \gamma)/n + O(1/n^2))h(n),$$

where γ is Euler's constant, and therefore $\ln 2 - \gamma \approx 0,1159\ldots$. And the k-th moment of X_n is:

$$\mathrm{E}(X_n^k) = (2^{1/n}h(n))^k\Gamma(1+k/n).$$

In particular:

$$\mathrm{Var}(X_n) = (2^{1/n}h(n))^2[\Gamma(1+2/n) - (\Gamma(1+1/n))^2].$$

Proof. This follows from the definition of the Weibull distribution, whose moments are known. For the asymptotic development, note that $2^{1/n} = 1 + (\ln 2)/n + O(1/n^2)$ and $\Gamma(1+1/n) = 1 - \gamma/n + O(1/n^2)$ where γ is Euler's constant. Therefore:

$$\mathrm{E}(X_n) = (1 + (\ln 2 - \gamma)/n + O(1/n^2))h(n).$$

□

Th. 2 is also useful for varying V. First, let us clarify the expression in Th. 2:

Lemma 7. *For any $V \geq 0$:*

$$e^{-V/2}\sum_{r=0}^{\infty}\frac{r^k}{r!}(V/2)^r = \begin{cases} V/2 & \text{if } k=1 \\ V/2 + (V/2)^2 & \text{if } k=2 \end{cases}.$$

Corollary 8. *There exist constants $c_1 > 0$ and $c_2 > 0$ such that for all sufficiently large n, the number N_n of pairs of points $\pm\mathbf{x}$ of a random lattice $L \in \mathcal{L}_n$ in a measurable set C_n of volume V_n, symmetric with respect to 0, satisfies:*

$$\mathrm{E}(N_n) = V/2(1 + O(2^{-c_1 n})) \quad (1)$$

$$\mathrm{Var}(N_n) = V/2 + (V+1)^2 O(2^{-c_2 n}) \quad (2)$$

Proof. This follows by combining Lemma 7 and Th. 2. □

By combining Markov's inequality with Cor. 8, we obtain:

Theorem 9. *Let L be a random unit-volume full-rank lattice in $\mathbf{R}^n$. Then, with probability at least $1 - o(1)$ as n grows to infinity:*

$$1 - (\log\log n)/n \leq \frac{\lambda_1(L)}{h(n)} \leq 1 + (\log\log n)/n.$$

Thus, $\mathrm{E}(\lambda_1(L)/h(n))$ *converges to 1, and therefore* $\mathrm{E}(\lambda_1(L))$ *is asymptotically equivalent to $h(n)$.*

Proof. Markov's inequality ensures that for any $t > 0$:

$$\Pr(|N_n - \mathrm{E}(N_n)| > t) \leq \mathrm{Var}(N_n)/t^2.$$

Let C be the centered ball of volume $V = (\log n)/2$, and let $t = (\log n)/8$. Then $E(N_n) - t = (\log n)/4(1 + O(2^{-c_1 n})) - (\log n)/8 = (\log n)/8(1 + O(2^{-c_1 n}) > 1$, for sufficiently large n. And:

$$\mathrm{Var}(N_n)/t^2 = 16((\log n)/4 + ((\log n)/2+1)^2 O(2^{-c_2 n}))/log^2 n = O(1/\log n) + O(2^{-c_2 n}).$$

Therefore, with probability at least $1 - o(1)$, $\lambda_1(L)$ is less or equal than the radius of the ball C, which is $(\log n)/2^{1/n} h(n) \leq 1 + (\log\log n)/n$ for all sufficiently large n.

To obtain the lower bound, let C be the centered ball of volume $V = 2/\log n$ and $t = 1/\log\log n$. Its radius is $(2/\log n)^{1/n} h(n) \geq 1 - (\log\log n)/n$ for all sufficiently large n. Furthermore, $E(N_n) + t = 2/(\log n)(1 + O(2^{-c_1 n})) + 1/\log\log n = o(1)$ and $\mathrm{Var}(N_n)/t^2 = o(1)$, which proves the lower bound. The bounds on $\lambda_1(L)/h(n)$ hold with probability at least $1 - o(1)$. And by Minkowski's bound, we always have $0 < \lambda_1(L)/h(n) \leq 2$. It follows that $\mathrm{E}(\lambda_1(L)/h(n))$ converges to 1. □

Th. 9 shows that the expectation of $\lambda_1(L)$ is asymptotically equivalent to that of X_n from Lemma 6.

3. Random Integer Lattices

3.1. **Definitions.** In the literature, several classes of random integer lattices have been considered: from a practical point of view, it is preferable to consider classes which are efficiently samplable. In order to facilitate comparisons with real random lattices, it is customary to scale any integer lattice L to make its co-volume equal to one, by normalizing it by a factor $\mathrm{vol}(L)^{-1/\dim(L)}$.

Goldstein and Mayer [7] considered the (finite) set $\mathcal{I}_{N,n}$ of full-rank n-dimensional integer lattices of co-volume exactly $N \geq 1$, normalized by $N^{1/n}$, so that each element of $\mathcal{I}_{N,n}$ is a lattice of unit co-volume. In [7], it is noted that if N is a sufficiently large prime number, one can efficiently sample the uniform distribution over $\mathcal{I}_{N,n}$, using the Hermite normal form: there is an efficient probabilistic algorithm which, given n and a prime N, outputs a random lattice whose distribution is statistically close to the uniform distribution over $\mathcal{I}_{N,n}$. Motivated by worst-case to average-case reductions, Gama *et al.* [5] considered partition cells of $\mathcal{I}_{N,n}$ based on the factor group $\mathbf{Z}^n/L$. More precisely, for any finite Abelian group G, let $\mathcal{I}_{G,n}$ be the (finite) set of full-rank n-dimensional integer lattices L such that $\mathbf{Z}^n/L \simeq G$, where each lattice L is divided by $\#G^{1/n}$, $\#G$ being the cardinality of G. Then the sets $\mathcal{I}_{G,n}$ form a partition of $\mathcal{I}_{N,n}$ when G runs over all finite Abelian groups of order N, up to isomorphism.

The structure theorem states that any finite Abelian group G is isomorphic to a direct product $\mathbf{Z}/q_1\mathbf{Z} \times \cdots \times \mathbf{Z}/q_k\mathbf{Z}$; the rank of G is then defined as the minimal number of cyclic groups in such a decomposition. In order for $\mathcal{I}_{G,n}$ to be non-empty, the rank of G must be $\leq n$. If the q_i's are known for one such decomposition, [5] shows how to efficiently sample the uniform distribution over $\mathcal{I}_{G,n}$. This generalizes the sampling result of [7], because if G has prime order, then $\mathcal{I}_{G,n} = \mathcal{I}_{\#G,n}$.

We note that Nguyen and Shparlinski [12] recently determined the asymptotic proportion of co-cyclic lattices (those for which $\mathbf{Z}^n/L$ is cyclic) among all full-rank integer lattices: the natural density is $1/[\zeta(6)\prod_{k=4}^{n}\zeta(k)] \approx 85\%$ (for large n). Finally, we mention that in an unpublished manuscript [1], Ajtai showed how to efficiently sample a random lattice whose distribution is statistically close to the Haar distribution over unit co-volume lattices.

3.2. **Properties.**

Goldstein and Mayer [7] proved the following equidistribution result:

Theorem 10. *Let A be a measurable subset of X_n such that the boundary of A has μ-measure zero, and let χ denote the characteristic function of A. Then the average $(1/\#\mathcal{I}_{N,n})\sum_{L\in\mathcal{I}_{N,n}}\chi(L)$ converges when N grows to ∞, and its limit is $\mu(A)$.*

As an example, let A be the subset of X_n formed by all lattices L such that $\lambda_1(L) \leq h(n)$. Cor. 4 implies that $\mu(A)$ converges to $1-e^{-1/2}$ when n grows to infinity. For any n, the average $(1/\#\mathcal{I}_{N,n})\sum_{L\in\mathcal{I}_{N,n}}\chi(L)$ converges towards $\mu(A)$ when N grows to infinity. This average is the probability that a random lattice $L \in \mathcal{I}_{N,n}$ satisfies $\lambda_1(L) \leq h(n)$.

Theorem 11. *Let $\varepsilon > 0$. There exists $n_0 > 0$ such that for all $n \geq n_0$, there exists N_0 such that for all $N \geq N_0$, a random lattice L chosen uniformly at random from $\mathcal{I}_{n,N}$ satisfies with probability at least $1-\varepsilon$:*

$$1 - (\log\log n)/n \leq \frac{\lambda_1(L)}{h(n)} \leq 1 + (\log\log n)/n.$$

Proof. Let A be the subset of X_n formed by all lattices L such that:

$$1 - (\log\log n)/n \leq \frac{\lambda_1(L)}{h(n)} \leq 1 + (\log\log n)/n.$$

By Th. 9, there exists $n_0 > 0$ such that for all $n \geq n_0$, $\mu(A) \geq 1 - \varepsilon/2$. Now, let $n \geq n_0$. By Th. 10, there exists N_0 such that for all $N \geq N_0$, $|(1/\#\mathcal{I}_{N,n})\sum_{L\in\mathcal{I}_{N,n}}\chi(L) - \mu(A)| \leq \varepsilon/2$, therefore $(1/\#\mathcal{I}_{N,n})\sum_{L\in\mathcal{I}_{N,n}}\chi(L) \geq 1 - \varepsilon$. □

3.3. **Experiments.**

3.3.1. *Equidistribution speed.* Th. 10 is an equidistribution result on $\mathcal{I}_{N,n}$, but it is only asymptotic: in particular, it does not say how large the co-volume N must be, compared to n. A closer look at [7, Th. 5.2] reveals that the proof requires that $n^2/\log N = o(1)$, which leads to very large lattices.

On the other hand, we note that if N is too small, then $\mathcal{I}_{N,n}$ cannot be equidistributed, because the distribution of $\lambda_1(L)$ is too far from that of a random lattice. Indeed, note that for any n-dimensional integer lattice L of volume N and any $n' \in \{1,\ldots,n\}$, there is a n'-dimensional sublattice $L' \subseteq L$ of volume $\leq N$: such a sublattice can easily be derived from the Hermite normal form of L. It follows that for any $L \in \mathcal{I}_{N,n}$ and any $n' \in \{1,\ldots,n\}$, we have that $\lambda_1(L) \leq N^{1/n}2h(n')N^{1/n'} = 2h(n')N^{1/n'+1/n}$ (because Minkowski's upper bound on λ_1 is exactly twice the Gaussian heuristic). If N is sufficiently small and if n' is suitably chosen, then the upper bound $2h(n')N^{1/n'+1/n}$ will be much smaller than $h(n)$. For instance, if $N \leq 2^{n/\log n}$, then $n' = \lfloor n/\log n \rceil$ implies that $2h(n')N^{1/n'+1/n} \leq O(h(n')) = O(\sqrt{n/\log n})$ while $h(n) = \Theta(\sqrt{n})$, which would contradict Cor. 4.

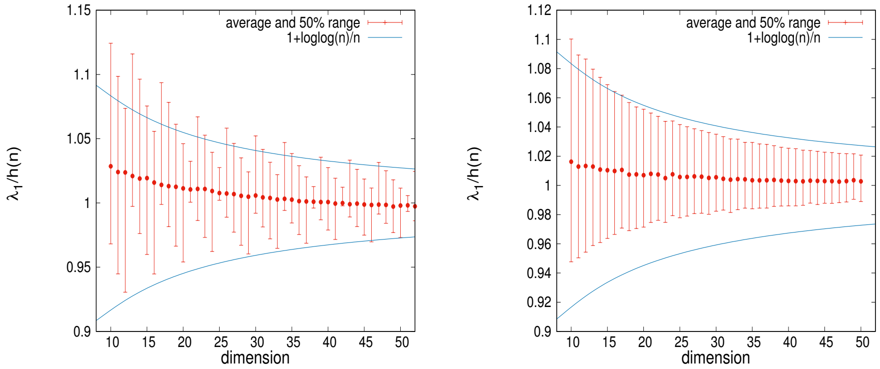


Figure 1. Average and ranges of 50% median of $\lambda_1(L)/h(n)$. Statistics over $M = 10,000$ bases from $\mathcal{EI}_{b(n),n}$. Left: $b(n) = n$. Right: $b(n) = n \log n$

We performed experiments to guess how large N must be in practice to ensure equidistribution. To do this, we compare the experimental distributions of $\lambda_1(L)$ between very large N with smaller N. Let $b(n)$ be a function of the dimension n and $\mathcal{EI}_{b(n),n}$ be the union of $\mathcal{I}_{N,n}$ for all $N \in [2^{\lfloor b(n) \rceil - 1}, 2^{\lfloor b(n) \rceil} - 1]$. For $b(n) = n, n \log n$ and some other functions, we compute several statistics over the shortest vectors of $M = 10,000$ random bases for each dimension. Figure 1 shows the averages, ranges of 50% median of $\lambda_1(L)/h(n)$, and $1 \pm (\log \log n)/n$. In the left and right figures, L is uniformly sampled from $\mathcal{EI}_{n,n}$ and $\mathcal{EI}_{n \log n,n}$ respectively. For instance, for $\varepsilon = 1/2$, $n_0 = 20$ and $N_0 = 2^{n \log n}$ seem likely to satisfy the assumption of Th. 11.

Notice that $b(n) = n$ implies $h(n) \in [2^{1-1/n}, 2] \cdot V_n(1)^{-1/n} \approx \sqrt{2n/\pi e} = 0.4839 \cdot \sqrt{n}$ whereas the length of the shortest vector is discretized as the square root of integers. Figure 2 shows the histogram of $\lambda_1/h(n)$ of $\mathcal{EI}_{b(n),n}$ for $b(n) = n, 2n, 3n$ and $n \log n$ in $n = 40$ dimension. To see the difference, we also look at the Kolmogorov-Smirnov statistics between the Weibull distribution and the sampled values of $\lambda_1(L)/h(n)$ in the right in Figure 2

To summarize, experiments suggest that $\mathcal{I}_{N,n}$ with $\log_2 N \geq 5n$ is sufficient in practice to make the distribution close to the Siegel distribution. Furthermore, in this range, we get that the distribution of $\lambda_1(L)$ is very close to the Weibull distribution of Lem 6.

3.3.2. *Equidistribution of other random integer lattices.* Th. 10 is an equidistribution result on $\mathcal{I}_{N,n}$. One might wonder if this result can be generalized to $\mathcal{I}_{G_n,n}$ for suitable sequences (G_n) of finite Abelian groups of order growing to infinity. For instance, one may consider a sequence (G_n) of cyclic groups. To this end, we performed experiments for sequences (G_n) of cyclic groups: the lattices appear to have a distribution close to the Haar distribution.

Furthermore, we experiment using lattices from higher rank cyclic groups. Fix rank r and dimension n Then select random integer q of $\lfloor n \log n/r \rceil$-bit, and select random group elements $g_1, \ldots, g_n$ in $(\mathbb{Z}/q\mathbb{Z})$. The lattice is defined by all integer tuple $(x_1, \ldots, x_n)$ s.t. satisfies $\sum x_i g_i \equiv 0 \pmod q$. Figure 3 shows the averages, ranges of 50% median of $\lambda_1(L)/h(n)$, and $1 \pm (\log \log n)/n$ for $r = 5$ and 20.

3.4. **Application to the SVP Challenge.** The SVP Challenge [20] managed by TU Darmstadt is an online contest to find a nearly shortest vector in random integer lattices. More

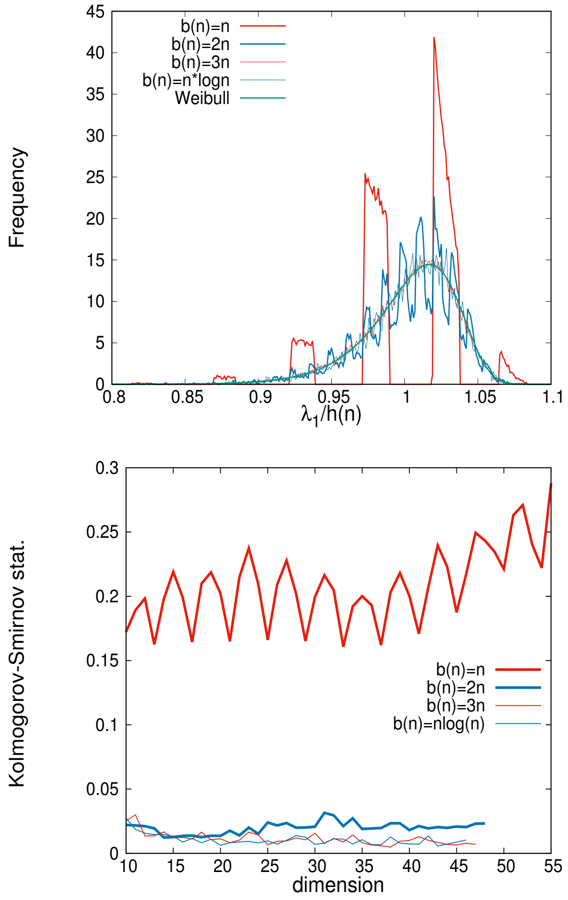


Figure 2. Left: Histogram of $\lambda_1(L)/h(n)$ for $b(n) = n, 2n, 3n$ and $n \log n$ in 40 dimension. Right: Kolmogorov-Smirnov statistics between the Weibull distribution and experimental $\lambda_1(L)/h(n)$. Statistics over $M = 10,000$ bases.

precisely, the challenge problem in dimension n is randomly chosen from $\mathcal{I}_{N,n}$ where N is a random $10n$-bit prime number determined by a seed parameter s. For each dimension, the goal of the challenge is to find a vector shorter than $1.05h(n)$: if such a vector has already been found, the goal is to find a shorter vector than the current record in the same dimension, and one is allowed to change of seed, where the default seed is zero.

3.4.1. *Distribution of the first minimum.* We discuss heuristic estimates for the first minimum of challenge lattices, based on random lattices theory. This allows to "guess" how likely one can find a better solution to a solved challenge.

The experiments of Sect. 3.3 suggest that the distribution of the SVP challenge lattices is close to the distribution of random lattices, and that $\lambda_1/h(n)$ has a distribution close to the Weibull distribution with parameter $(\eta = 2^{1/n}, m = n)$. If true, the probability that a shorter vector than $|\mathbf{v}|$ is given by the CDF (cumulative distribution function)

$$1 - e^{-(|\mathbf{v}|/h(n))^n/2}. \tag{3}$$

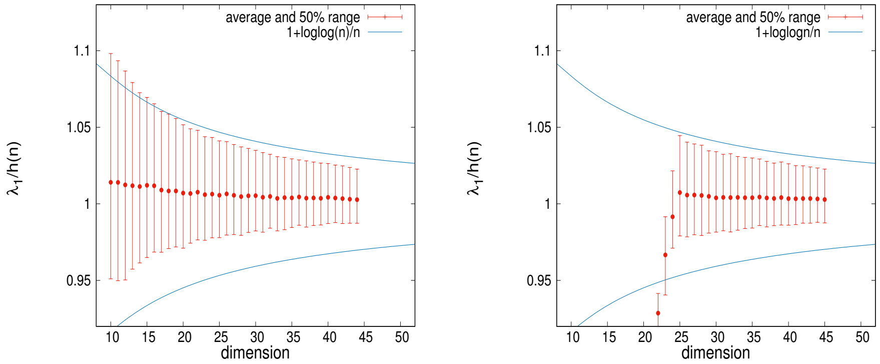


Figure 3. Average and ranges of 50% median of $\lambda_1(L)/h(n)$. Statistics over $M = 10,000$ bases from $\mathcal{I}_{G_n,n}$ whose co-volume is $\approx 2^{n\log n}$. Left: $r = 5$. Right: $r = 20$.

Recall that the main theorem of Södergren [25, Theorem 1'] states that the sequence of volumes $\{V_n(\lambda_i/h(n))\}_{i=1}^{\infty}$ converges weakly to the Poisson point process with intensity $1/2$. Our previous experiments suggest that λ_1 of the SVP challenge lattices have a distribution close to Weibull.

Assume the challenge instances follow this distribution Then, the i-th volume $V_n(\lambda_i/h(n)$ follows the $Gamma(i,2)$ whose p.d.f. (probability density function) is $\frac{2^{-i}x^{i-1}e^{-x/2}}{(i-1)!}$. In particular, the first volume follows the exponential distribution whose p.d.f. is $(1/2)e^{-x/2}$. Under this assumption, assume that a short lattice vector $\mathbf{v}$ has been found in a dimension n. Then, the probability that a shorter vector exists in the same lattice can be computed as follows.

$$\begin{aligned}
&1 - \Pr_L\left[v \text{ is the shortest in } L | L \text{ has a vector of length } |\mathbf{v}|\right] \\
&1 - \frac{v \text{ is the shortest in } L}{\Pr_L[L \text{ has a vector of length } |\mathbf{v}|]} \\
&1 - \frac{v \text{ is the shortest in } L}{\sum_{k=1}^{\infty} \Pr_L[\text{Length of } k\text{-th vector is } |\mathbf{v}|]} \\
&1 - \frac{(1/2)e^{-W/2}}{\sum_{k=1}^{\infty} 2^{-k}W^{k-1}e^{-W/2}/(k+1)!} \text{ where } W = V_n(|\mathbf{v}|) \\
&1 - e^{-W/2} = 1 - e^{-V_n(|\mathbf{v}|/2)}.
\end{aligned}$$

Table 1 shows the probability from the current records of SVP Challenge.

When this probability is small, say being lower than 10%, meaning that the found vector is probably the shortest, trying another seed is a possible strategy. We now discuss the expected necessary number of lattice bases in this case. Because the lattice co-volume is a random $10n$-bit prime, i.e. in $[w(n), 2w(n)-1]$ where $w(n) = 2^{10n-1}$, the distribution of λ_1 is not exactly a shifted Weibull distribution even if we neglect the difference between the challenge instances and truly random lattices. Such variation can be ignored by choosing special seeds that give small volumes, for instance smaller than $1.001w(n)$. Again assuming the distribution of $\lambda_1(L)/h(n)$ is Weibull, with $h(n) = w(n)^{1/n}V_n(1)^{-1/n}$, we can estimate the necessary number K of new lattice bases to find a vector shorter than $|\mathbf{v}|$. Let $x_1, \ldots, x_K$ be the points sampled from the Weibull distribution induced from the random integer lattices whose co-volume is $w(n)$. Then, the CDF of $\min\{x_1, \ldots, x_K\}$ is

Table 1. Probability that there exists a vector shorter than current records and expected number of bases to find a shorter vector.

dim	$\lvert\mathbf{v}\rvert/h(n)$ of best known vector	probability (??)	number of bases (??)
79	0.91591	0.00048444	1431
90	0.95952	0.0012056	58
108	0.95162	0.0023579	294
110	0.98355	0.077479	9
120	0.99535	0.24859	3
126	1.00556	0.63413	1
130	0.99871	0.34476	2
140	1.01139	0.91293	1
146	1.04534	≈ 1	1
150	1.04192	≈ 1	1

given by

$$C_{K-min} = 1-(1-C(x))^K = 1-e^{-K\cdot(x/h(n))^n/2}.$$

Thus, the probability that there exists a shortest vector shorter than $|\mathbf{v}|$ is $1-e^{-K\cdot(|\mathbf{v}|/h(n))^n/2}$. Hence, letting $K=-2\log(1-p)/(|\mathbf{v}|/h(n))^n$, the probability is p.

Setting the probability 1/2, we get the number:

$$K = \lceil 2\log 2/(|\mathbf{v}|/h(n))^n\rceil = \lceil \log 2/V_n(|\mathbf{v}|/2^{10})\rceil. \tag{4}$$

Table 1 shows the expected numbers for the current records in the SVP Challenge.

Remark 12. *The strategy without selecting the seeds can be considered. That is, sequentially take the integer seeds. $\lambda_1(L)$ follows the product distribution of the Weibull and the determinant uniformly distributes over the prime numbers in $[2^{10n-1}, 2^{10n}-1]$. Due to the prime number theorem, the p.d.f. of volume has a distribution close to $c/\log(x)$ for some constant c. Thus, the product distribution is given by the integral*

$$P(z) \quad = \quad n\cdot 2^{-m-1}\cdot\int_{z\cdot 2^{-m}}^{z\cdot 2^{1-m}} w^{n-2}e^{-w^n/2}\cdot\frac{c}{\log(z/w)}dw.$$

We obtain the CDF by integrating this function from $-\infty$ to z, and the distribution of the minimum of λ_1 of K bases.

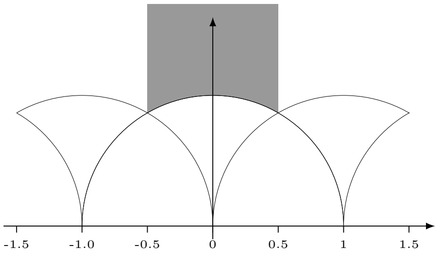
3.4.2. *Number of solutions.* Remember that the goal of the challenge is to find a vector shorter than $1.05h(n)$ for a lattice chosen uniformly at random from $\mathcal{I}_{N,n}$, where N is a random $10n$-bit prime number. We compute a heuristic estimate of the number of solutions from random lattices theory.

Let N_n be the number of pairs of points $\pm\mathbf{x}$ of a random lattice $L\in\mathcal{L}_n$ in C. Let N_n be the number of of pairs of vectors of norm $\leq 1.05h(n)$ for a random lattice $L\in\mathcal{L}_n$. By Th. 1, we know that: $\mathrm{E}(N_n) = 1.05^n/2$. Letting C be the ball of radius $1.05h(n)$, and because $\sqrt{4/3}^{n/2}\approx 1.0746^n$, Theorem. 2 implies that:

$$\lim_{n\to\infty}(\mathrm{Var}(N_n) - 1.05^n/2) = 0.$$

4. Explicit computations and experiments in low dimensions

4.1. **On the average value of the first minima in dimension 2.** In dimension two, it is possible to carry out explicitly the computations. Taken modulo rotations, the moduli space Λ_2 of two-dimensional unit-volume lattices corresponds to the $\mathrm{SL}(2,\mathbf{Z})$-equivalent classes of the upper dimension 2 upper half-plane $\mathfrak{h} = \mathrm{SL}(2,\mathbf{R})/\mathrm{SO}(2,\mathbf{R}) \cong \{x+iy \in \mathbf{C} : y > 0\}$ under the modular action of $\mathrm{SL}(2,\mathbf{Z})$ given by $\tau \mapsto (a\tau+b)/(c\tau+d)$. By the Iwasawa decomposition, an element of $\mathfrak{h}$ has a unique representative of the form $y^{-1/2}\begin{pmatrix}1 & x\\ 0 & y\end{pmatrix}$. Lagrange's reduction shows that the connected set $\mathcal{D} = \{(x,y) : x^2+y^2 \geq 1, |x| \leq 1/2, y > 0\}$ is a fundamental domain for this action. By adding the condition $x \geq 0$, one obtains a fundamental domain for $\mathrm{GL}(2,\mathbf{Z})$. For instance, if $x+iy$ is a point of this fundamental domain, the corresponding unimodular lattice has first minimum $y^{-1/2}$.



Fundamental domain of $\mathrm{SL}(2,\mathbf{Z}) \setminus \mathfrak{h}$

The Siegel measure over Λ_2 derives from the hyperbolic measure $dxdy/y^2$ invariant by the modular action of $\mathrm{SL}(2,\mathbf{Z})$. Let us explicit its mass: the corresponding fundamental domain $\mathcal{D}$, although not compact, has finite measure since $\int_{y_0}^{\infty} dy/y^2$ converges. To normalize the measure so that $\mu(\Lambda_2) = 1$, it suffices to take :

$$d\mu(x+iy) = \frac{3}{\pi}\frac{dxdy}{y^2}.$$

On can thus compute expectations over two-dimensional random lattices, by integrating $d\mu(x+iy)$ over $\{(x,y) : x^2+y^2 \geq 1, |x| \leq 1/2, y > 0\}$. As an example, we have:

Theorem 13. *Let L be a random two-dimensional unit-volume lattice. Then:*

$$\begin{array}{ll} \mathrm{E}(\lambda_1(L)) = \frac{2}{\pi}\int_{-1/2}^{1/2}\frac{dx}{(1-x^2)^{3/4}} \approx 0.6826 & \mathrm{E}(\lambda_2(L)) \approx 1.97314 \\ \mathrm{E}(\lambda_1^2(L)) = 3\ln 3(2\pi)^{-1} \approx 0.5245 & \mathrm{E}(\lambda_2(L)^2) = \infty \end{array}$$

Proof. We have:

$$\mathrm{E}(\lambda_1(L)) = \frac{3}{\pi}\int_{-1/2}^{1/2}\left(\int_{\sqrt{1-x^2}}^{\infty}\frac{dy}{y^{2+1/2}}\right)dx = \frac{2}{\pi}\int_{-1/2}^{1/2}\frac{dx}{(1-x^2)^{3/4}} \approx 0.6826.$$

□

In particular, one notes that the expectation $\mathrm{E}(\lambda_1(L))$ is much less than the maximal value of $\lambda_1(L)$, which is $\sqrt{\gamma_2} = (4/3)^{1/4} \approx 1.0746$.

4.2. On the number of Siegel and LLL bases in low dimension.

4.2.1. *Reduced basis of a lattice.* Let L be a lattice; as soon as $\mathrm{rk}L \geq 2$, it owns infinitely many bases. Among those, some have interesting and congenial properties, such as having reasonably small vectors and low orthogonality defect. They are called *reduced bases* and finding them is the goal of *lattice reduction* theory. A first notion, introduced in 1945 by Siegel in [23] to approximate a fundamental domain of the classes of generalized upper half-plane $\mathfrak{h}_n = \mathrm{SL}(n, \mathbf{R})/\mathrm{SO}(n, \mathbf{R})$ under the action of the discrete group $\mathrm{SL}(n, \mathbf{Z})$. Transposed in the lattice theory language gives the following definition:

Definition 14 (Siegel-reduced basis). *A basis $\mathcal{B} = (b_1, \ldots, b_n)$ of a lattice L is said to be (δ, η)-**Siegel-reduced** for certain parameters $0 < \delta < 1$ and $1/2 < \eta < 1$ if the following conditions are satisfied:*

$$\forall i < j, \quad |\langle b_j, \pi_i(b_i)\rangle| \leq \eta \|\pi_i(b_i)\|^2 \quad \textit{(Size-Reduction condition)} \tag{5}$$

$$\forall i, \quad \delta\|\pi_i(b_i)\|^2 \leq \|\pi_{i+1}(b_{i+1})\|^2 \quad \textit{(Siegel condition)}, \tag{6}$$

where π_i is the orthogonal projection on the subspace $(b_1, \ldots, b_{i-1})^\perp$, with the convention of π_1 being the identity map.

Even thought of very valuable theoretical interest, the Siegel reduction is quite impractical since it lacks a computational method to reduce an arbitrary basis in a Siegel-reduced one. In 1982, Lenstra, Lenstra, and Lovász designed the LLL *algorithm* [9], being the first algorithm whose running time is polynomial in the rank of the lattice.

Definition 15 (LLL-reduced basis). *A basis $\mathcal{B} = (b_1, \ldots, b_n)$ of a lattice is said to be (δ, η)-**LLL-reduced** for certain parameters $1/4 < \delta < 1$ and $1/2 < \eta < \sqrt{\delta}$ if it is size-reduced (in the sense of definition 14 and if the following condition is satisfied:*

$$\forall i, \quad \delta\|\pi_i(b_i)\|^2 \leq \left(\|\pi_{i+1}(b_{i+1})\|^2 + \frac{\langle b_{i+1}, \pi_i(b_i)\rangle}{\|\pi_i(b_i)\|^2}\right) \quad \textit{(Lovász condition)}. \tag{7}$$

4.2.2. *Number of reduced basis.* The number of Siegel-reduced bases (and therefore LLL) is finite for a given lattice. Since the Siegel measure gives a natural measure on the moduli space of lattices, one can look at the expected number of reduced bases of a random lattice. This question has been treated in the PhD. thesis ([22]) of Kim and the work of Kim and Venkatesh ([8]):

Theorem 16. *Let n be a non-negative integer.*

1. *The average number[1] of the (δ, η)-LLL bases in dimension n is:*
$$\frac{n 2^n (2\eta)^{(n-1)(n-2)/2}}{\mathfrak{m}_n n! (n-1)!} \prod_{\ell=2}^{n} \int_{-\eta}^{\eta} (\delta - \eta^2)^{-\frac{\ell(n-\ell)}{2}}.$$
2. *The average number of the (δ, η)- Siegel bases in dimension n is:*
$$\frac{n 2^n (2\eta)^{(n-1)(n-2)/2}}{\mathfrak{m}_n n! (n-1)!} \delta^{-\frac{n^3-n}{12}},$$
where $\mathfrak{m}_n = n\sqrt{\pi}^{-(\ell+2)(\ell-1)} \prod_{\ell=2}^{n} \zeta(\ell)\Gamma(\ell/2)$ is the mass of a fundamental domain of $SL(n, \mathbf{Z}) \setminus SL(n, \mathbf{Z})/SO(n, \mathbf{Z})$ for the Haar measure.

As introduced in Section 3, one can efficiently sample random integer lattices from using the Goldstein and Mayer distribution. Table 2 enumerates the expected (theoretical and empirical) number of $(0.998001, 0.5)$-LLL/Siegel bases, under the action of the group $(\mathbf{Z}/2)^d$, corresponding to the identification of vectors modulo their signs, as well as the empirical standard variation of these distributions. The experiments consists in sampling

[1] In Kim's manuscript, the definition of the δ parameter differs from the usual one: Kim's δ corresponds to an $\sqrt{\delta}$ in the standard definition. We adapt the theorem in consequence.

Table 2. Expected number of LLL reduced bases for low dimensional lattices.

dim	(empirical) mean		(empirical) std variation		(theoretical) mean	
	LLL	Siegel	LLL	Siegel	LLL	Siegel
2	1.001	1.091	0.0316	0.2878	1.001	1.104
3	1.007	1.488	0.0834	0.8606	1.008	1.479
4	1.036	2.623	0.1864	2.7273	1.036	2.666
5	1.113	7.337	0.4587	11.2318	1.116	7.133
6	1.303	33.384	0.7495	74.273	1.318	31.374
7	1.889	303.449	1.828	746.6434	1.949	251.5
8	3.376		5.8490		3.281	
9	7.467		16.7904		8.642	
10	40.03		106.4059		38.856	

1000 lattices, using at least[2] 2^{10d} as size of coefficients, and then enumerating the reduced basis with a recursive enumeration routine. The number of Siegel basis past dimension 8 was too large to get meaningful statistical results due to a very large standard deviation. The expectation is consistent between theoretical estimates and practical results, showing that the random integer lattice model is a very good approximation of random lattices even in a non-asymptotic regime. However, the standard deviation is noticeably high, though [8] showed that the standard deviation is asymptotically small: this phenomenon may only be observed in higher dimensions than can be handled in a reasonable time.

5. Random basis of lattices

Though random lattices are well understood, there is *a priori* no good definition of *random bases* of a given lattice L. Any definition of random bases of L should only depend on L itself.

5.1. Generative process for random basis and generating family.

Let us fix a lattice L, with a distinguished basis $\mathcal{B} = (b_1, \ldots, b_n) \in L^n$, represented as a matrix $B \in \mathrm{GL}(n, \mathbf{Z})$. As a preliminary remark and since it useful in different generating procedure, we start with a brief preliminary discussion on vectors generation in L.

5.1.1. *Sampling a vector in L.* Let us suppose that we want to sample a vector in L according to a distribution that is *isotropic* in the sense that its conditioning by the event of getting a prescribed norm ℓ yields the uniform distribution over $L \cap B(0, \ell)$. In order to be

[2] Remark that Kim conducted experiments in low dimensions as well in his dissertation. Nonetheless he used an insufficient number of bits to follow the Goldstein-Mayer distribution. Hence one can not consider the lattices he sampled as ”random” in this sense.

of practical interest, we also require the distribution to be *effective*, that is easily algorithmically samplable. In practice, it appears that a particular distribution fulfill this requirement and are sufficiently simple to sample in addition of being very congenial[3]: the *(centered) discrete Gaussian distribution.* This is the distribution $\Theta_{L,\sigma}(x) = \Theta(L,\sigma)^{-1}e^{-\frac{\|x\|^2}{2\sigma^2}}$, where $x \in L$ and $\Theta(L,\sigma)^{-1}$ is a normalizing factor. Intuitively[4], $\Theta(L,\sigma)$ is the conditional probability that derived from the Gaussian kernel $\rho_s(x) = e^{-\frac{\|x\|^2}{2\sigma^2}}$ conditioned by the event of belonging to L. Alternatively one can think of this distribution as the normalized finite measure induced by the theta function on $\frac{1}{\sigma}L$, viewed as an Hermitian vector bundle.

5.1.2. *Generating by multiplication with random unimodular matrices.* Since any pair of basis of the lattice L are related by a unimodular transformation, a naive yet convenient way to generate a random basis of L by first sampling a unimodular matrix $U \in \mathrm{SL}(n, \mathbf{Z})$ and output the matrix $P = UB$. This is the solution chosen in most public-key lattice-based cryptosystems, where the public key P matrix is constructed from the secret key B. The generation then falls back to the sampling of a unimodular matrix.

- A first way to sample a unimodular matrix comes from the decomposition of any element of $\mathrm{SL}(n, \mathbf{Z})$ in a product of transposition and (integral) transvection matrix. It suffices indeed to randomly generate a random sequence of transposition/transvection matrices and multiply them all together. A variant of this strategy is implanted for instance in the NIST post-quantum candidate DRS ([14]).
- A second way consists in sampling a matrix $S \in \mathbf{Z}^{n\times n}$, where each vector is drawn independently from a probability distribution over $\mathbf{Z}^n$. With overwhelming probability this matrix is invertible, meaning that its row vectors $\mathbf{Z}$-span a sublattice of $\mathbf{Z}^n$. It then suffices to apply the construction of Micciancio and Goldwasser (Chap. 6 of [11]) to transform this family of vectors in a basis of $\mathbf{Z}^n$, at the cost of expanding their norm by a factor dominated by $\sqrt{n}$.

It appears in practice that as far as our experiments were involved the two distributions yield the same practical results. We thus refer this kind of generation by *unimodular generation.*

5.1.3. *On the covariance of the first vector sampled.* Let $P = U \cdot B$ where U is a random unimodular matrix. Then, by unrolling the definition, the covariance matrix of the first vector $P[1]$ of the generated basis P is:

$$\mathrm{cov}\ P[1] = B^T \cdot \mathrm{cov}\ U[1] \cdot B,$$

where $U[1]$ is the first row vector of U. By a symmetry argument, the covariance matrix of the discrete Gaussian distribution is diagonal. When generating U —with an algorithm independent of B—, it then appears clearly that the covariance of the corresponding $P[1]$ is not basis-independent in L. In subsequent paragraphs, the bias induced by this generation is studied in an more extensive fashion.

[3] Another distribution with such properties would be the uniform distribution in $L \cap B(0, \ell)$ for large enough ℓ. For well chosen parameters i the two introduced distributions can be made statistically close by using rejection sampling techniques.

[4] This intuition can actually be made rigorous by standard argument of discrete probability theory.

5.2. **Sampling a generating family.** In the previous generating procedure, the role of the initial basis B is preponderant as it twists the distribution of the vectors appearing in the basis. An orthogonal method consists in trying to suppress the dependence on this privileged basis by sampling according to an isotropic distribution, as introduced in Paragraph 5.1.1. When sampling sufficiently many vectors independently—namely $n+k$ for a small constant k— they generates the whole original lattice with high probability. This procedure doesn't give a random basis but makes arise a natural notion of random generating family. This procedure is coordinate-free in the sense that the Gaussian distribution is indeed coordinate-free: it only depends on the vectors of L. We refer to this kind of process by the generic term *Gaussian generation*.

5.3. **Experimental results on random basis with the LLL algorithm and LLL dark bases.** The LLL algorithm is widely used in many branches of mathematics and computer science, but despite its interest, its average-case analysis is still not fully understand. We use here the LLL reduction to highlight the differences between the two kind of procedures described in Section 5.1, as well as to point out a surprising result on the discrepancy that occurs between the different LLL basis. Indeed, one can study the distribution of output bases of the LLL algorithm[5].

5.3.1. *Evolution of the bias in small dimension for random lattices.* An intuitive experiment consists in studying the total variation distance between the outputs of the two distributions, once the LLL reduction is applied. Formally we generated 5000 random lattices (in the sense of Goldstein-Mayer), and for each of them generated 50000 bases, apply the LLL reduction and compute an empirical estimate of the TVD between these distributions. The results of this extensive experiment are compiled in Figure 4 as the serie Unimodular/Gaussian. This graph presents also the empirical TVD between two independant series of samples drawn with same generation (Unimodular/Unimodular and Gaussian/Gaussian series), in order to be able to differentiate a possible artifact of measure from a real difference. Eventually, it appears clearly that the TVD between the Gaussian and Unimodular generation is significantly larger and grows faster than the TVD obtained between the two same series of sample, ruling out the measurement artifact possibility. Unsurprisingly, the bias between these two processes is an increasing function of the dimension.

5.3.2. *Dark bases for LLL reduction.* Let us define the family of lattices $(L_n)_n$ as the lattices generated by the rows of the $d \times n$ matrices ℓ_n:

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ (1-\eta) & a & 0 & \cdots & \\ 0 & (1-\eta)a & a^2 & & \vdots \\ \vdots & 0 & \ddots & & \\ 0 & \cdots & 0 & (1-\eta)a^{n-2} & a^{n-1} \end{bmatrix},$$

with $a = 0.8912$. These lattices have an anomalously low number of bases: we experimentally checked that up to dimension 10, L_n has exactly 2 LLL-reduced bases, counted modulo the sign of vectors, namely: on the one hand ℓ_n and the other hand the "reversed" basis

$$\kappa_n = (\ell[n], \ell[n-1], \ell[n-1] + \ell[n-2], \ldots, \ell[n-1] + \ell[n-2] + \cdots + \ell[1])\,.$$

[5] Technically speaking, reducing a generating family is possible with a variant of the LLL algorithm, introduced by Schnorr in [15].

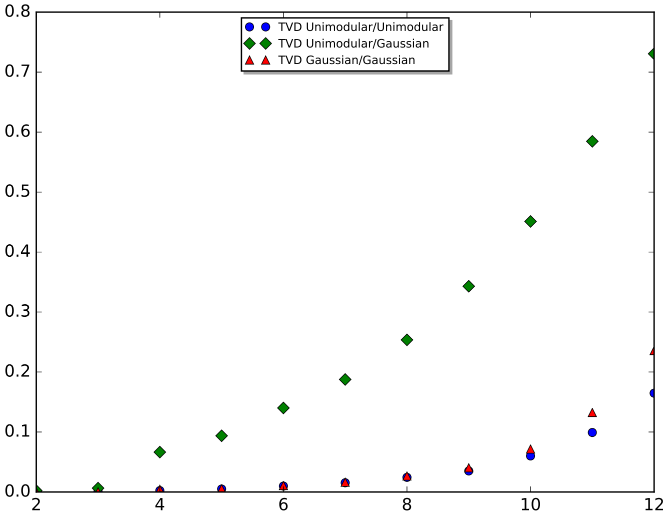


Figure 4. Bias in the repartition of LLL reduced bases.

One could suppose that there is an equiprobability of getting one or another basis from the reduction of a random generating family. But surprisingly the basis ℓ_n is *never* touched by this process whereas κ_n is *systematically* obtained. Extensive experiments points out that the probability of getting ℓ_n seems to be *at least* lower than 2^{-21} as soon as $n \geq 6$. We call such bases **dark**: informally, an LLL basis is dark if the experimental probability that it is output by LLL given as input a random basis is negligible, especially with respect to the number of LLL bases. The presence of dark bases appears not to be limited to these specificlattices and is observable even in small dimensions. Such a behavior was encoutnered and reported by Kim in [22]. Obviously, the LLL reduction *does prefer* certain bases over others in a given lattice. Understanding the distribution of dark bases may lead to a better global understanding of the average behavior of LLL.

References

[1] M. Ajtai. Generating random lattices according to the invariant distribution. March 2006.

[2] Miklós Ajtai. Generating hard instances of lattice problems. In *Proc. STOC '96*. ACM, 1996.

[3] Miklós Ajtai. Random lattices and a conjectured 0 - 1 law about their polynomial time computable properties. In *43rd Symposium on Foundations of Computer Science (FOCS 2002), 16-19 November 2002, Vancouver, BC, Canada, Proceedings*, pages 733–742, 2002.

[4] Yuanmi Chen and Phong Q. Nguyen. BKZ 2.0: better lattice security estimates. In *Advances in Cryptology – Proc. ASIACRYPT '11*, volume 7073 of *Lecture Notes in Computer Science*, pages 1–20. Springer, 2011.

[5] Nicolas Gama, Malika Izabachene, Phong Q. Nguyen, and Xiang Xie. Structural lattice reduction: Generalized worst-case to average-case reductions and homomorphic cryptosystems. In *Proc. EUROCRYPT '16*, Lecture Notes in Computer Science. Springer, 2016.

[6] Nicolas Gama and Phong Q. Nguyen. Predicting lattice reduction. In *Proc. EUROCRYPT '08*, volume 4965 of *Lecture Notes in Computer Science*, pages 31–51. Springer, 2008.

[7] Daniel Goldstein and Andrew Mayer. On the equidistribution of Hecke points. *Forum Math.*, 15(2):165–189, 2003.

[8] Seungki Kim and Akshay Venkatesh. The behavior of random reduced bases. *International Mathematics Research Notices*, 2017.

[9] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász. Factoring polynomials with rational coefficients. *Mathematische Ann.*, 261:513–534, 1982.

[10] A. M. Macbeath and C. A. Rogers. A modified form of Siegel's mean value theorem. II. *Proc. Cambridge Philos. Soc.*, 54:322–326, 1958.

[11] D. Micciancio and S. Goldwasser. *Complexity of Lattice Problems: A Cryptographic Perspective.* Springer Int. series in Engineering and Computer Science. Springer US, 2002.

[12] Phong Q. Nguyen and Igor E. Shparlinski. Counting co-cyclic lattices. *CoRR*, abs/1505.06429, 2015.

[13] Phong Q. Nguyen and Damien Stehlé. LLL on the average. In *ANTS*, pages 238–256, 2006.

[14] National Institute of Standards and Technology. Round 1 submissions of post-quantum cryptography call, 2018. Available from https://goo.gl/3DYgrK.

[15] M. Pohst. A modification of the lll reduction algorithm. *J. Symb. Comput.*, 4(1), 1987.

[16] C. A. Rogers. Mean values over the space of lattices. *Acta Math.*, 94:249–287, 1955.

[17] C. A. Rogers. The moments of the number of points of a lattice in a bounded set. *Philos. Trans. Roy. Soc. London. Ser. A.*, 248, 1955.

[18] C. A. Rogers. The number of lattice points in a set. *Proc. London Math. Soc. (3)*, pages 305–320, 1956.

[19] C. A. Rogers. The chance that a point should be near the wrong lattice point. *J. London Math. Soc.*, 37:161–163, 1962.

[20] M. Schneider and N. Gama. SVP challenge. Available at http://goo.gl/7B9zpG.

[21] Claus-Peter Schnorr and M. Euchner. Lattice basis reduction: improved practical algorithms and solving subset sum problems. *Math. Programming*, 66:181–199, 1994.

[22] Kim Seungki. *On the shape of a high-dimensional random lattice.* PhD thesis, Stanford University, 2015.

[23] Carl Ludwig Siegel. A mean value theorem in geometry of numbers. *Ann. of Math. (2)*, 46, 1945.

[24] Anders Södergren. On the distribution of angles between the N shortest vectors in a random lattice. *J. Lond. Math. Soc. (2)*, 84(3):749–764, 2011.

[25] Anders Södergren. On the Poisson distribution of lengths of lattice vectors in a random lattice. *Math. Z.*, 269(3-4):945–954, 2011.

6. More Statistical Experiments on Integer Lattices

This section shows the result of additional experiments from Section 3.3 to check the higher order moments $\mathrm{E}(X^k)$.

Figure 5 shows the 1st to 4th moments of the lattices in $\mathcal{EI}_{n,n}$ and $\mathcal{EI}_{n\log n,n}$.

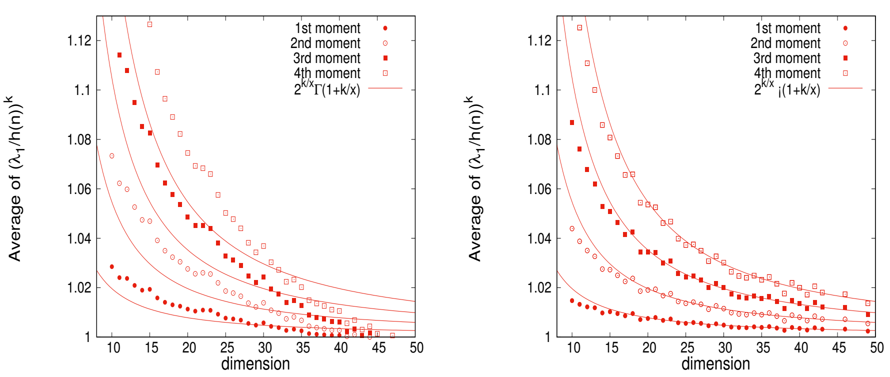


Figure 5. Experimental average of 1st to 4th moments. Statistics over $M = 10,000$ bases. Left: $b(n) = 3n$. Right: $b(n) = n\log n$.

Figure 6 shows the result of additional experiments on $\mathcal{I}_{G_n,n}$ for rank 2 and 15.

Security Fundamentals Laboratory, Cybersecurity Research Institute, National Institute of Information and Communications Technology, Japan

Sorbonne Universitées, LIP 6 UMR 7606, Paris, France

Inria Paris, France, CNRS/JFLI and the University of Tokyo, Japan

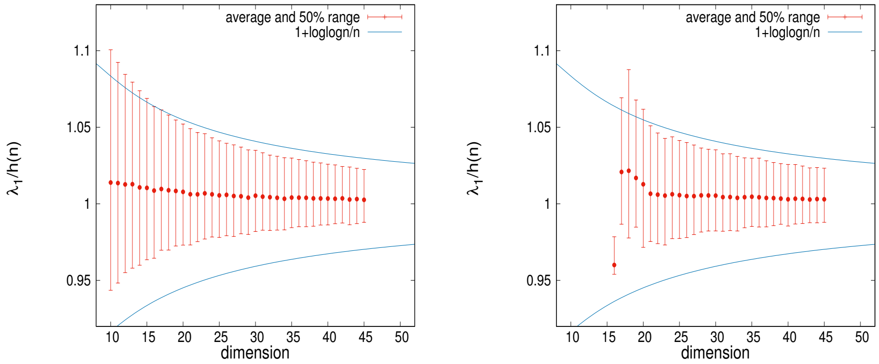


Figure 6. Average and ranges of 50% median of $\lambda_1(L)/h(n)$. Statistics over $M = 10,000$ bases from $\mathcal{I}_{G_n,n}$ whose co-volume is $\approx 2^{n\log n}$. Left: $r = 2$. Right: $r = 15$.