Proof for why HPP does not work if samples are normally distributed

Eirik D. Skjerve

December 13, 2024

Notation:

- V matrix
- $\mathbf{v}$ row vector
- $\mathcal{P}(V)$ fundamental parallelepiped
- $\mathcal{N}(\mu, \sigma^2)$ continuous normal distribution
- $\mathcal{D}_{2\mathbb{Z}+c,\sigma}$ discrete Gaussian distribution
- $\mathbb{E}[X]$ expectance of a random variable X
- $\mathbb{V}[X]$ variance of a random variable X

1 HPP against normally distributed samples

In the following, we see what happens to the computations the *Learning a parallelepiped* attack is based on if we replace the uniform distribution by a normal distribution. The key component and assumption of the *Learning a parallelepiped* attack is that the provided samples are distributed uniformly over $\mathcal{P}(V)$. One runs into trouble if the sampled vectors are on the form $\mathbf{v} = \mathbf{x}V$ where $\mathbf{x}$ follows a normal distribution, i.e. $x_i \sim \mathcal{N}(\mu, \sigma^2)$. Although one might be able to approximate the covariance matrix $V^t V$ and transform the hidden parallelepiped to a hidden hypercube, one can not do a gradient descent based on the fourth moment given such samples using the method from the original attack [2]. We will show that if samples follow a normal distribution, the fourth moment of $\mathcal{P}(C)$ over $\mathbf{w}$ on the unit circle is constant, and therefore a gradient descent can not reveal any information about the secret key V.

Adapting the definition of $\mathcal{P}(V)$ Recall that $\mathcal{P}(V)$ is defined as $\{\sum_{i=1}^{n} x_i \mathbf{v}_i : x_i \in [-1,1]\}$ where $\mathbf{v}_i$ are rows of V and x_i is uniformly distributed over $[-1,1]$ (generally one can take another interval than $[-1,1]$ and do appropriate scaling). Firstly, the normal distribution $\mathcal{N}(\mu, \sigma^2)$ is defined over $(-\infty, \infty)$, so it does not make sense to talk about samples "normally distributed over $\mathcal{P}(V)$" without tweaking any definitions. Therefore, let $[-\eta, \eta]$ be a finite interval on which to consider a truncated normal distribution $\mathcal{N}_\eta(\mu, \sigma^2)$ such that $\int_{-\eta}^{\eta} f_X(x)dx = 1 - \delta$ for some negligibly small δ where $f_X(x)$ is the probability density function of $\mathcal{N}(\mu, \sigma^2)$, given by $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Now we consider $\mathcal{P}_\eta(V) = \{\sum_{i=1}^{n} x_i \mathbf{v}_i : x_i \in [-\eta, \eta]\}$ where $x_i \sim \mathcal{N}_\eta(\mu, \sigma^2)$ with $\mu = 0$ and proceed as in the original HPP with $\mathcal{P}_\eta(V)$ instead of $\mathcal{P}(V)$.

Approximating $V^t V$ Let $V \in \mathcal{GL}_n(\mathbb{R})$. Let $\mathbf{v}$ be chosen from a truncated normal distribution $\mathcal{N}_\eta(0, \sigma^2)$ over $\mathcal{P}_\eta(V)$. Then $\lim_{\eta\to\infty} \mathbb{E}[\mathbf{v}^t\mathbf{v}] = V^t V \cdot \sigma^2$.

Proof. Let samples be on the form $\mathbf{v} = \mathbf{x}V$, where $\mathbf{x}$ is a row vector where each element $x_i \sim \mathcal{N}_\eta(0, \sigma^2)$. Then $\mathbf{v}^t\mathbf{v} = (\mathbf{x}V)^t(\mathbf{x}V) = (V^t\mathbf{x}^t)(\mathbf{x}V) = V^t\mathbf{x}^t\mathbf{x}V$. Considering $\mathbb{E}[\mathbf{x}^t\mathbf{x}]$, we see that for $i \neq j$, $\mathbb{E}[x_i x_j] = \mathbb{E}[x_i]\mathbb{E}[x_j] = 0 \cdot 0 = 0$ due to independent random variables. For $i = j$, $\lim_{\eta\to\infty} \mathbb{E}[x_i^2] = \mathbb{V}[x_i] = \sigma^2$ since $\mathbb{V}[x_i] = \mathbb{E}[x_i^2] - \mathbb{E}[x_i]^2 = \mathbb{E}[x_i^2] - 0 = \sigma^2$. Therefore, $\lim_{\eta\to\infty} \mathbb{E}[\mathbf{x}^t\mathbf{x}] = I_n \cdot \sigma^2$, i.e. the matrix with σ^2 on the diagonal and 0 otherwise. Consequently, $\lim_{\eta\to\infty} \mathbf{v}^t\mathbf{v} = V^t\mathbb{E}[\mathbf{x}^t\mathbf{x}]V = V^t(I_n \cdot \sigma^2)V = (V^t V) \cdot \sigma^2$ and conversely $\lim_{\eta\to\infty} V^t V = (\mathbf{v}^t\mathbf{v})/\sigma^2$. $\square$

This means that we can in theory approximate the covariance matrix $V^t V$ by averaging over $\mathbf{v}^t\mathbf{v}$ and dividing by σ^2. However, it is not immediately clear if one needs more samples for this approximation than in the original attack due to the difference in distributions.

Hypercube transformation Assume now that we know $V^t V$. Consider instead of $\mathcal{P}(V)$, $\mathcal{P}_\eta(V)$. Then by following part 1 of **Lemma 2** and its proof from [2] we can transform our hidden parallelepiped $\mathcal{P}_\eta(V)$ into $\mathcal{P}_\eta(C)$, a hidden hypercube, since this does not depend on the distribution of the samples - it only assumes one knows $V^t V$. For completeness, by adapting the second part of **Lemma 2** to our case:

Proof. Let $\mathbf{v} = \mathbf{x}V$ where $\mathbf{x}$ is normally distributed according to $\mathcal{N}_\eta(0, \sigma^2)$. Then samples $\mathbf{v}$ are distributed according to $\mathcal{N}_\eta(0, \sigma^2)$ over $\mathcal{P}_\eta(V)$. It then follows that $\mathbf{v}L = \mathbf{x}VL = \mathbf{x}C$ has a truncated uniform distribution over $\mathcal{P}_\eta(C)$. $\square$

Thus, we should be able to map our normally distributed samples from the hidden parallelepiped to the hidden hypercube.

Learning a hypercube It is clear that samples normally distributed over $\mathcal{P}_\eta(C)$ centered at the origin form a hypersphere for which any orthogonal rotation leaves the sphere similar in shape. As a consequence, the fourth moment of $\mathcal{P}_\eta(C)$ is constant over the unit circle. Analogous to [2] we compute the 2nd and 4th moment of $\mathcal{P}_\eta(V)$ over a vector $\mathbf{w} \in \mathbb{R}^n$.
The k-th moment of $\mathcal{P}_\eta(V)$ over a vector $\mathbf{w}$ is defined as $mom_{V,k} = \mathbb{E}[\langle \mathbf{u}, \mathbf{w} \rangle^k]$ where $\mathbf{u} = \sum_{i=1}^n x_i \mathbf{v}_i$ and $x_i \sim \mathcal{N}_\eta(0, \sigma^2)$.
First we consider $\langle \mathbf{u}, \mathbf{w} \rangle = \langle \sum_{i=1}^n x_i \mathbf{v}_i, \mathbf{w} \rangle = \sum_{i=1}^n x_i \langle \mathbf{v}_i, \mathbf{w} \rangle$.
Then for $k = 2$, $\mathbb{E}[(\sum_{i=1}^n x_i \langle \mathbf{v}_i, \mathbf{w} \rangle)^2] = \mathbb{E}[\sum_{i=1}^n \sum_{j=1}^n x_i x_j \langle \mathbf{v}_i, \mathbf{w} \rangle \langle \mathbf{v}_j \mathbf{w} \rangle]$.
Due to independent random variables, $\mathbb{E}[x_i x_j] = 0$ when $i \neq j$ as previously shown, so we have $\sum_{i=1}^n \mathbb{E}[x_i^2] \langle \mathbf{v}_i, \mathbf{w} \rangle^2 = \sigma^2 \sum_{i=1}^n \langle \mathbf{v}_i, \mathbf{w} \rangle^2$ for sufficiently large η due to the well known result that $\mathbb{E}[X^2] = \sigma^2$ for $X \sim \mathcal{N}(0, \sigma^2)$.
Thus, we end up with:

$$mom_{V,2}(\mathbf{w}) = \sigma^2 \mathbf{w} V^t V \mathbf{w}^t \tag{1}$$

We observe that if $V \in \mathcal{O}(\mathbb{R})$, $mom_{V,2}(\mathbf{w}) = \sigma^2 \|\mathbf{w}\|^2$

For $k = 4$:

$$\mathbb{E}[(\sum_{i=1}^n x_i \langle \mathbf{v}_i, \mathbf{w} \rangle)^4] = \mathbb{E}[\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n x_i x_j x_k x_l \langle \mathbf{v}_i, \mathbf{w} \rangle \langle \mathbf{v}_j, \mathbf{w} \rangle \langle \mathbf{v}_k, \mathbf{w} \rangle \langle \mathbf{v}_l, \mathbf{w} \rangle]$$

There are three cases for the indices i, j, k, and l:

1. **None equal**: if i, j, k, and l are different, the expression equals 0 due to independent random variables.
2. **All equal**: if $i = j = k = l$, then we have $\sum_{1=1}^n \mathbb{E}[x_i^4] \langle \mathbf{v}_i, \mathbf{w} \rangle^4$. A well known result for the normal distribution $\mathcal{N}(0, \sigma^2)$ is that $\mathbb{E}[x^4] = 3\sigma^4$.
3. **Pairwise equal**: if either
   - $i = j \neq k = l$
   - $i = l \neq j = k$
   - $i = k \neq j = l$

we have the following:

$$\sum_{i \neq j} \mathbb{E}[x_i^2 x_j^2] \langle \mathbf{v}_i, \mathbf{w} \rangle^2 \langle \mathbf{v}_j, \mathbf{w} \rangle^2$$

Since $\mathbb{E}[x_i^2 x_j^2] = \mathbb{E}[x_i^2]\mathbb{E}[x_j^2] = \sigma^4$ due to independent random variables, we have

$$\sigma^4 \sum_{i \neq j} \langle \mathbf{v}_i, \mathbf{w} \rangle^2 \langle \mathbf{v}_j, \mathbf{w} \rangle^2$$

For large enough η in our distribution $\mathcal{N}_\eta(0, \sigma^2)$, putting together the expressions above we have

$$mom_{V,4}(\mathbf{w}) = 3\sigma^4 \sum_{i=1}^{n} \langle \mathbf{v}_i, \mathbf{w} \rangle^4 + 3(\sigma^4 \sum_{i \neq j} \langle \mathbf{v}_i, \mathbf{w} \rangle^2 \langle \mathbf{v}_j, \mathbf{w} \rangle^2)$$

since there are three cases where indices pair up two and two. The final result becomes:

$$mom_{V,4}(\mathbf{w}) = 3\sigma^4 (\sum_{i=1}^{n} \langle \mathbf{v}_i, \mathbf{w} \rangle^4 + \sum_{i \neq j} \langle \mathbf{v}_i, \mathbf{w} \rangle^2 \langle \mathbf{v}_j, \mathbf{w} \rangle^2) \tag{2}$$

Claim: If $V \in \mathcal{O}(\mathbb{R})$, and $\mathbf{w}$ is on the unit sphere, $mom_{V,4}(\mathbf{w})$ is constant.

Proof. This can be shown by rewriting (2) as

$$mom_{V,4}(\mathbf{w}) = 3\sigma^4 (\sum_{i=1}^{n} \langle \mathbf{v}_i, \mathbf{w} \rangle^4 + \sum_{i=1}^{n} \langle \mathbf{v}_i, \mathbf{w} \rangle^2 \sum_{j=1}^{n} \langle \mathbf{v}_j, \mathbf{w} \rangle^2 - \sum_{i=1}^{n} \langle \mathbf{v}_i, \mathbf{w} \rangle^4)$$

$$mom_{V,4}(\mathbf{w}) = 3\sigma^4 (\sum_{i=1}^{n} \langle \mathbf{v}_i, \mathbf{w} \rangle^2 \sum_{j=1}^{n} \langle \mathbf{v}_j, \mathbf{w} \rangle^2) = 3\sigma^4 (\sigma^2 \|\mathbf{w}\|^2)^2 = 3\sigma^8$$

because $mom_{V,2}(\mathbf{w}) = \sum_{i=1}^{n} \langle \mathbf{v}_i, \mathbf{w} \rangle^2 = \sigma^2 \|\mathbf{w}\|^2$ when $V \in \mathcal{O}(\mathbb{R})$ and $\|\mathbf{w}\|^2 = 1$ when $\mathbf{w}$ lies on the unit sphere. □

In conclusion, if samples over the secret parallelepiped $\mathcal{P}_\eta(V)$ follow a continuous normal distribution, a gradient descent based on the fourth moment described in [2] is impossible because the fourth moment is constant over the unit sphere of $\mathbb{R}^n$.

The discrete Gaussian distribution Consider now the discrete Gaussian distribution $\mathcal{D}_{2\mathbb{Z}+c,\sigma}$ as described in [1]. If $X \sim \mathcal{D}_{2\mathbb{Z}+c,\sigma}$, we have $\mathrm{Supp}(X) = 2\mathbb{Z} + c$ where $c \in \{0, 1\}$. $\Pr[X = x] = \frac{\rho_\sigma(x)}{\sum_{y \in 2\mathbb{Z}+c} \rho_\sigma(y)}$ where $\rho_\sigma(x) = e^{-\frac{x^2}{2\sigma^2}}$. For $c \in \{0, 1\}$, $\mathbb{E}[X] = 0$ and $\mathbb{V}[X] \approx \sigma^2$ for appropriate choices of σ. Naturally, we have that $\sum_{x \in 2\mathbb{Z}+0} \frac{\rho_\sigma(x)}{\sum_{y \in 2\mathbb{Z}+0} \rho_\sigma(y)} = 1$ and $\sum_{x \in 2\mathbb{Z}+1} \frac{\rho_\sigma(x)}{\sum_{y \in 2\mathbb{Z}+1} \rho_\sigma(y)} = 1$ and that $\sum_{x \in \mathbb{Z}} \frac{\rho_\sigma(x)}{\sum_{y \in \mathbb{Z}} \rho_\sigma(y)} = 1$.

Proof sketch Firstly, we need to show whether we can estimate the covariance matrix of the secret V with a method similar to the continuous case. We also need to show that we can map our samples from the hidden parallelepiped to the hidden hypercube via a transformation L similarly to the continuous case.
Next we want to prove or disprove that $\mathbb{E}[X^2] = \sigma^2$ and $\mathbb{E}[X^4] = 3\sigma^4$ or more generally that $3\mathbb{E}[X^2]^2 = \mathbb{E}[X^4]$. If either statement is true, we will be in the case of the normal distribution, i.e. that $mom_{V,4}(\mathbf{w})$ is constant when $V \in \mathcal{O}(\mathbb{R})$ and $\mathbf{w}$ is on the unit circle. If the statements are false, $mom_{V,4}(\mathbf{w})$ might not be constant, and we might be able to do a gradient descent optimization like in the HPP attack [2].

References

[1] Joppe W. Bos, Olivier Bronchain, Léo Ducas, Serge Fehr, Yu-Hsuan Huang, Thomas Pornin, Eamonn W. Postlethwaite, Thomas Prest, Ludo N. Pulles, and Wessel van Woerden. Hawk. Technical report, NXP Semiconductors, Centrum Wiskunde & Informatica, Mathematical Institute at Leiden University, NCC Group, PQShield, Institut de Mathématiques de Bordeaux, September 2024.

[2] Phong Q. Nguyen and Oded Regev. Learning a parallelepiped: Cryptanalysis of ggh and ntru signatures, 2009.