Mean Squared Length of Vectors in the Approximate Greatest Common Divisor Lattice

S. Murphy

Technical Report
RHUL–MA–2012–06–v2
5 September 2012

Minor Correction to
RHUL-MA-2012-06
21 May 2012

Department of Mathematics
Royal Holloway, University of London
Egham, Surrey TW20 0EX, England

http://www.rhul.ac.uk/mathematics/techreports

Abstract

This paper derives the mean squared length of vectors in a lattice relevant to the Approximate GCD problem and the related fully homomorphic encryption scheme.

1 Introduction

The *Approximate GCD* problem is loosely speaking to find an integer p given a collection of noisy multiples of p. A fully homomorphic encryption scheme based on the *Approximate GCD* problem is given in [2] based on the difficulty of the *Approximate GCD* problem for certain parameter sizes. One suggested method of solving the *Approximate GCD* problem is to set up a particular lattice and find short vectors in this lattice. We give a rigourous derivation for the mean square length of vectors in this lattice and briefly consider some implications relevant to the security analysis given for the fully homomorphic encryption scheme of [2].

2 The Approximate GCD Problem

We give the following formulation of *Approximate GCD* or `AGCD` problem. We suppose that we have $(t+1)$ noisy multiples of the odd integer p, where the multipliers of p have integer mean μ and variance κ^2. We can model this situation statistically in the following way. We let $Q_0, Q_1, \ldots, Q_t$ be independent and identically distributed integer-valued random variables with mean μ, variance κ^2 and distribution function F_Q. Similarly, we let $Z_0, Z_1, \ldots, Z_t$ be independent and identically distributed integer-valued random variables with zero mean, variance σ^2 and distribution function F_Z, where Q_i and Z_j are independent $(i, j = 0, 1, \ldots, t)$. We then define the noisy multiples $X_0, X_1, \ldots, X_t$ of p by

$$X_i = pQ_i + Z_i \qquad [i = 0, 1, \ldots, t],$$

so we have $\mathbf{E}(X_i) = p\mu$ and $\mathrm{Var}(X_i) = p^2\kappa^2 + \sigma^2$.

For fairly obvious reasons, p can be described as an *Approximate GCD* (AGCD) of the integers $X_0, X_1, \ldots, X_t$. The notation $X' = (X_1, \ldots, X_t)^T$ denotes the t-dimensional vector obtained by deleting the first component of $(t+1)$-dimensional vector X and so on. In some discussions of the `AGCD` problem, X_0 is taken to be the maximum of $X_0, X_1, \ldots, X_t$, though this does not materially affect the conclusions of our analysis. The `AGCD` problem is loosely speaking to find p given the noisy multiples $X_0, X_1, \ldots, X_t$ of p.

2.1 The AGCD Lattice

The `AGCD` problem is often analysed by constructing a particular lattice based on the noisy observations X. The LLL algorithm is then applied to this lattice in order to find short vectors in the lattice. Under certain circumstances such a short vector can directly give the vector of multiples Q, so immediately giving p as the Approximate GCD or AGCD of $X_0, X_1, \ldots, X_t$.

The usual lattice used to analyse the `AGCD` problem is constructed in the following. Firstly, a bound $a\sigma$ on the error size of the X_i is given, that is to say a small multiple a of the standard deviation σ of the errors and we may typically take $a = 1$. The *AGCD lattice* is then defined by the *AGCD lattice basis matrix* A given by

$$A = \begin{pmatrix} a\sigma & X_1 & X_2 & \ldots & X_t \\ 0 & -X_0 & 0 & \ldots & 0 \\ 0 & 0 & -X_0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & -X_0 \end{pmatrix}.$$

The high-level rationale usually for the use of the AGCD lattice is that in many circumstances the so-called *target vector* $Q^T A$ is a short vector in the AGCD lattice and may even be the shortest. Short vectors in a lattice can often be identified by applying the LLL algorithm to the AGCD lattice basis matrix A. The LLL algorithm finds a unimodular matrix L and an alternative basis matrix $B = LA$ for the lattice in which the basis vectors are loosely speaking short vectors and roughly orthogonal. In terms of the LLL-reduced basis matrix B, the target vector $Q^T A = \left(Q^T L^{-1}\right) B$. If the short target vector $Q^T A$ is the short first vector in the LLL-reduced basis for example, then $Q^T L^{-1} = (1, 0, \ldots, 0)$, so $Q^T = (1, 0, \ldots, 0)L$. In this case, the vector of multipliers Q^T is given as the first row of the unimodular matrix L calculated by the LLL algorithm.

If U is a generic integer random vector, then a generic vector in the AGCD lattice is the random vector $U^T A$ with squared length given by the random variable $|U^T A|^2 = U^T(AA^T)U$. A generic random vector in this AGCD lattice is therefore given by the random vector

$$U^T A = \left(a\sigma U_0 \;\middle|\; U_0 X'^T - X_0 U'^T\right),$$

which has squared length given by the random variable

$$\left|U^T A\right|^2 = a^2\sigma^2 U_0^2 + \left|U_0 X' - X_0 U'\right|^2.$$

3 Random AGCD Lattice Vectors

The random vectors in the AGCD lattice we consider are those that arise from random vectors of the form $U = lQ + v$, where l is an integer and v is an integer vector. Thus we consider those random vectors of the form $U^T A = (lQ + v)^T A$ in the AGCD lattice. In order to calculate the squared length of such a lattice vector $(lQ + v)^T A$, we need to consider the vector $U_0 X' - X_0 U'$, whose i^{th} component is given by

$$\begin{aligned} U_0 X_i - X_0 U_i &= (lQ_0 + v_0)(pQ_i + Z_i) - (pQ_0 + Z_0)(lQ_i + v_i) \\ &= (pv_0 - lZ_0)Q_i + (lQ_0 + v_0)Z_i - v_i(pQ_0 + Z_0) \\ &= \alpha(Z_0)Q_i + \beta(Q_0)Z_i - v_i\gamma(Q_0, Z_0), \end{aligned}$$

where $\alpha(Z_0) = pv_0 - lZ_0$, $\beta(Q_0) = lQ_0 + v_0$ and $\gamma(Q_0, Z_0) = pQ_0 + Z_0$. Thus we have

$$\left|(lQ + v)^T A\right|^2 = a^2\sigma^2\beta(Q_0)^2 + \sum_{i=1}^{t} \left(\alpha(Z_0)Q_i + \beta(Q_0)Z_i - v_i\gamma(Q_0, Z_0)\right)^2$$

3.1 Conditional Mean Squared Length

In order to calculate the mean squared length $\mathbf{E}\left(\left|(lQ + v)^T A\right|^2\right)$ of a generic vector in the standard lattice, we first calculate this quantity conditional on the joint value of (Q_0, Z_0). Thus if we write

$$\Psi(q_0, z_0) = \mathbf{E}\left(\left|(lQ + v)^T A\right|^2 \;\middle|\; Q_0 = q_0, Z_0 = z_0 \right)$$

for this conditional expectation, then we can obtain

$$\Psi(q_0, z_0) = a^2\sigma^2\beta(q_0)^2 + \sum_{i=1}^{t} \mathbf{E}\left((\alpha(z_0)Q_i + \beta(q_0)Z_i - v_i\gamma(q_0, z_0))^2 \right).$$

If we write

$$W_i(q_0, z_0) = \alpha(z_0)Q_i + \beta(q_0)Z_i - v_i\gamma(q_0, z_0)$$

for a generic random variable in the above summand expectation, then the conditional expectation is given as

$$\Psi(q_0, z_0) = a^2\sigma^2\beta(q_0)^2 + \sum_{i=1}^{t} \mathbf{E}\left(W_i(q_0, z_0)^2 \right).$$

We note that

$$\mathbf{E}\left(W_i(q_0, z_0)^2\right) = \mathrm{Var}\left(W_i(q_0, z_0)\right) + \mathbf{E}\left(W_i(q_0, z_0)\right)^2,$$

where the mean and variance of $W_i(q_0, z_0)$ are given by

$$\begin{aligned}\mathbf{E}\left(W_i(q_0,z_0)\right) &= \alpha(z_0)\mathbf{E}(Q_i)+\beta(q_0)\mathbf{E}(Z_i)-v_i\gamma(q_0,z_0)\\ &= \mu\alpha(z_0)-v_i\gamma(q_0,z_0),\\ \text{and } \mathrm{Var}\left(W_i(q_0,z_0)\right) &= \alpha(z_0)^2\mathrm{Var}(Q_i)+\beta(q_0)^2\mathrm{Var}(Z_i)\\ &= \kappa^2\alpha(z_0)^2+\sigma^2\beta(q_0)^2.\end{aligned}$$

Thus the conditional expectation is given by

$$\Psi(q_0,z_0)=(a^2+t)\sigma^2\beta(q_0)^2+t\kappa^2\alpha(z_0)^2+\sum_{i=1}^{t}\left(\mu\alpha(z_0)-v_i\gamma(q_0,z_0)\right)^2.$$

3.2 Mean Squared Length of a AGCD Lattice Vector

The (unconditional) mean squared length of the random AGCD lattice vector $(lQ+v)^TA$ is given by

$$\int_{\mathbb{R}}\int_{\mathbb{R}}\mathbf{E}\left(\left|(lQ+v)^TA\right|^2\Big|\, Q_0=q_0,\ Z_0=z_0\right)\,dF_Z(z_0)dF_Q(q_0),$$

so we have

$$\mathbf{E}\left(\left|(lQ+v)^TA\right|^2\right)=\int_{\mathbb{R}}\int_{\mathbb{R}}\Psi(q_0,z_0)\,dF_Z(z_0)dF_Q(q_0).$$

However, this integral is itself an expectation, so we obtain

$$\mathbf{E}\left(\left|(lQ+v)^TA\right|^2\right)=\mathbf{E}\left(\Psi(Q_0,Z_0)\right),$$

where this expectation is given by

$$(a^2+t)\sigma^2\mathbf{E}\left(\beta(Q_0)^2\right)+t\kappa^2\mathbf{E}\left(\alpha(Z_0)^2\right)+\sum_{i=1}^{t}\mathbf{E}\left(\left(\mu\alpha(Z_0)-v_i\gamma(Q_0,Z_0)\right)^2\right).$$

We now evaluate each term in this expression. The random variable $\beta(Q_0)=lQ_0+v_0$ has mean $l\mu+v_0$ and variance $l\kappa^2$, and the random variable $\alpha(Z_0)=pv_0-lZ_0$ has mean pv_0 and variance $l^2\sigma^2$, so

$$\begin{aligned}\mathbf{E}\left(\beta(Q_0)^2\right) &= \mathrm{Var}\left(\beta(Q_0)\right)+\mathbf{E}\left(\beta(Q_0)\right)^2 &= l^2\kappa^2+(l\mu+v_0)^2,\\ \mathbf{E}\left(\alpha(Z_0)^2\right) &= \mathrm{Var}\left(\alpha(Z_0)\right)+\mathbf{E}\left(\alpha(Z_0)\right)^2 &= l^2\sigma^2+p^2v_0^2.\end{aligned}$$

The other random variable in the above expression is

$$\mu\alpha(Z_0)-v_i\gamma(Q_0,Z_0)=\mu(pv_0-lZ_0)-v_i(pQ_0+Z_0)=p\mu v_0-pv_iQ_0-(l\mu+v_i)Z_0,$$

so has mean $p\mu(v_0 - v_i)$ and variance $p^2 v_i^2 \kappa^2 + (l\mu + v_i)^2\sigma^2$, which gives

$$\mathbf{E}\left((\mu\alpha(Z_0) - v_i\gamma(Q_0, Z_0))^2\right) = p^2 v_i^2 \kappa^2 + (l\mu + v_i)^2\sigma^2 + p^2\mu^2(v_0 - v_i)^2.$$

Thus the mean squared length of the lattice vector $(Q + v)^T A$ is given by

$$\begin{aligned}\mathbf{E}\left(\left|(lQ + v)^T A\right|^2\right) \;&=\; (a^2 + t)\sigma^2\left(l^2\kappa^2 + (l\mu + v_0)^2\right) + t\kappa^2\left(l^2\sigma^2 + p^2 v_0^2\right) \\ &\quad + \textstyle\sum_{i=1}^{t} p^2 v_i^2 \kappa^2 + (l\mu + v_i)^2\sigma^2 + p^2\mu^2(v_0 - v_i)^2.\end{aligned}$$

If we now define the function

$$\eta(l, v_0) = l^2(a^2 + 2t)\sigma^2\kappa^2 + tp^2\kappa^2 v_0^2 + (a^2 + t)\sigma^2(l\mu + v_0)^2,$$

then the mean squared length of the lattice vector $(lQ + v)^T A$ is given by

$$\mathbf{E}\left(\left|(lQ + v)^T A\right|^2\right) = \eta(l, v_0) + p^2\kappa^2|v'|^2 + \sigma^2|v' + l\mu\mathbf{1}|^2 + p^2\mu^2|v' - v_0\mathbf{1}|^2.$$

3.3 Mean Squared Length of the Target Vector

The mean squared length $Q^T A$ of the target vector is given by taking $l = 1$ and $v = 0$ in the above expression for the mean of $\left|(lQ + v)^T A\right|^2$. We note that

$$\eta(1, 0) + \sigma^2|\mu\mathbf{1}|^2 = (a^2 + 2t)\sigma^2\kappa^2 + (a^2 + t)\sigma^2\mu^2 + \sigma^2 t\mu^2,$$

so the mean squared length of the target vector $Q^T A$ is given by

$$\mathbf{E}\left(\left|Q^T A\right|^2\right) = (a^2 + 2t)\sigma^2\left(\mu^2 + \kappa^2\right).$$

4 Fully Homomorphic Encryption and AGCD

The AGCD problem is used as the basis of the fully homomorphic encryption process given in [2]. We base our discussion of this encryption process and the corresponding AGCD problem on the version of this process given in Section 3.1 of [2]. In this AGCD problem, the AGCD p is an η-bit odd integer. If we let $m = \lfloor \frac{2^\gamma}{p} \rfloor$ for some integer $\gamma > \eta$, then m is a positive integer of length $(\gamma - \eta)$ bits. Furthermore, the multipliers $Q_0, Q_1, \ldots, Q_t$ are independent and identically distributed integer random variables, where

$$Q_0, Q_1, \ldots, Q_t \sim \mathrm{Uni}\left(\{0, \ldots, m\}\right).$$

The noise random variables $Z_0, Z_1, \ldots, Z_t$ are independent and identically distributed integer random variables, where

$$Z_0, Z_1, \ldots, Z_t \sim \mathrm{Uni}\left(\{-2^\rho, \ldots, 2^\rho\}\right) \quad [\rho << \eta].$$

The noisy multiples $X_i = pQ_i + Z_i$ $(i = 0, 1, \ldots, t)$ are therefore positive integers of length at most γ bits.

We now give the parameters of our discussion for the `AGCD` problem used by [2]. The mean and variance of the multipliers are given by

$$\mu = \mathbf{E}(Q_i) = \frac{m}{2} \quad \text{and} \quad \kappa^2 = \mathrm{Var}(Q_i) = \frac{m(m+2)}{12} \approx \frac{m^2}{12}.$$

Thus we have $\mu^2 \approx 3\kappa^2$, so μ^2 and κ^2 are comparable quantities. Similarly, the noise random variables have mean and variance given by

$$\mathbf{E}(Z_i) = 0 \quad \text{and} \quad \sigma^2 = \mathrm{Var}(Z_i) = \frac{2^{2\rho}}{3}.$$

4.1 Target Vector in Full Homomorphic Encryption

The security analysis of the fully homomorphic encryption scheme of [2] considers the AGCD lattice basis matrix denoted by M in Section 5.2 of [2]. We first note that the top-left entry in the matrix M of Section 5.2 of [2] is $2^\rho = \sqrt{3}\sigma$, so we need to take $a \approx \sqrt{3}$ in the AGCD lattice basis matrix A of Section 2.1 to obtain this matrix M.

We can now calculate the mean squared length of the target vector used in the AGCD lattice basis matrix M of [2], or equivalently A with $a \approx \sqrt{3}$ so $a\sigma = 2^\rho$. The mean squared length of this target vector $Q^T A$ is given by

$$\mathbf{E}\left(\left|Q^T A\right|^2\right) = (3+2t)\sigma^2\left(\frac{m^2}{4} + \frac{m^2}{12}\right) = (2t+3)\sigma^2\frac{m^2}{3}.$$

Extensive simulations have demonstrated the accuracy of the above result.

It is asserted in Section 5.2 of [2] that this target vector $Q^T A$ has length roughly

$$(t+1)^{\frac{1}{2}}\ 2^{\rho+\gamma-\eta}$$

We now express this true length of the target vector $Q^T A$ in terms of the parameters of [2], where we have $\sigma^2 = \frac{1}{3}2^{2\rho}$ and $m < 2^{\gamma-\eta}$. The mean squared length of the target vector $Q^T A$ is essentially bounded by $\frac{2}{9}(t+1)2^{2(\rho+\gamma-\eta)}$ for large t. Thus the root mean squared length of the target vector $Q^T A$ is essentially bounded for large t by

$$\frac{1}{2}\ (t+1)^{\frac{1}{2}}\ 2^{\rho+\gamma-\eta}.$$

In summary, the length of the target vector asserted by [2] is essentially more than double the true root mean squared length of the target vector $Q^T A$.

5 Conclusions

The treatment of this paper for the mean square length of an AGCD lattice vector illustrates the need for rigourous rather than *ad hoc* analysis of statistical quantities deriving from random lattices. For example, the analysis of the length of the target vector given by [2] at no point refers to the variability of the multipliers $(Q_0, Q_1, \ldots, Q_t)$, even though the expression in Section 3.3 clearly shows that this variablilty (given by κ) is potentially a major component of the mean square length.

The rigourous derivation of the mean square length of AGCD lattice vectors also raises questions about the heuristic security analysis of the corresponding fully homomorphic encryption scheme given by [2] in respect of using the LLL algorithm [1] to find small vectors in this lattice. It is asserted in Section 5.2 of [2] that the LLL algorithm cannot efficiently find the target vector $Q^T A$ for large t. However, the mean square length of the target vector is of the order of $t\sigma^2\mu^2$ (Section 3.3), whereas the mean square length of the non-target vector $(lQ + v)^T A$ is generally of the order of $tp^2\mu^2$ (Section 3.2). Loosely speaking, a non-target vector is therefore generally of the order of $\frac{p}{\sigma}$ larger than the target vector. However, the ratio $\frac{p}{\sigma}$ is essentially the "security parameter" λ of the fully homomorphic encryption scheme of [2], which can be made arbitrarily large. It would therefore seem that the target vector is generally an isolated vector in the lattice, of the order of λ^{-1} of the length of the shortest non-target vectors. In such a situation with a large security parameter λ, the LLL algorithm would be guarenteed to find this single very short target vector [1].

The heuristic security analysis of the fully homomorphic AGCD encryption scheme of Section 5.2 of [2] in respect of AGCD lattice seems incomplete as it lacks a rigourous discussion of lattice vector lengths. In particular, the claim that (for large t) "known lattice reduction algorithms will not be able to find [the target vector] efficiently" cannot be regarded as having been substantiated.

References

[1] A.K. Lenstra, H.W. Lenstra, and L. Lovasz. Factoring Polynomials with Rational Coefficients. *Mathematische Annalen*, 261:515–534, 1982.

[2] M. van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan. Fully Homomorphic Encryption over the Integers. In H. Gilbert, editor, *Advances in Cryptology – EUROCRYPT 2010*, volume 6110 of *LNCS*, pages 24–43, 2010.