STOCHASTIC ANALYSIS OF THE SMOOTH NUMBER PROPERTIES AND THEIR SEARCH

G. Vostrov, O. Ponomarenko
Odessa national politechnic university
Odessa, UKRAINE
e-mail: vostrov@gmail.com, ponomarenkoelena1997@gmail.com

Abstract

The paper introduces the concept of smooth primes and their classification, depending on the properties of their simple factors, into perfectly smooth and partially smooth. It is shown that in order to search for smooth primes and their analysis, it is necessary to know how primes are distributed depending on the number of simple factors $(p-1)$. The problem of constructing a measure of smoothness is considered. Results are shown showing how the first 10 million numbers $(p-1)$ are distributed by the number of simple factors.
Keywords: smooth number, data science, stochastic analysis

1 Introduction

At the moment, in number theory, there are many open problems that have not been solved for decades or centuries. Unsolved problems in number theory significantly limit the development of mathematics in theoretical and applied aspects. One of the fundamental problems is unproven hypotheses about prime numbers. According to which law the numbers $(p-1)$ are distributed according to the number of simple factors, is still unknown.

Prime numbers are widely used in public key cryptography, as well as directly related to the problem of discrete logarithm, which is currently not solved either. There are no efficient algorithms that allow solving a problem with polynomial complexity. Silver, Pohlig, and Hellman in [5] proposed an algorithm that promises polynomial complexity when using numbers of a special type, called "smooth". A more accurate description of the algorithm is given in [1]. However, the authors begin the description of the algorithm with a remark that it is not known that, for simplicity, we assume that the prime number p is smooth, and b is its primitive root. Such an assumption greatly simplifies the solution of the problem. At the same time, the authors completely ignore the rationale for such an assumption from the point of view, but how to find such numbers and how great is the probability of their being in a certain interval and what is the most important and what is the probability of such an option in the entire problem of the discrete logarithm. In modern number theory there is no clear definition of "smooth" numbers and their classification, and the laws of their distribution are completely unknown. This is especially important in modern cryptography. Along with this, it is necessary to solve the problem of finding smooth numbers of sufficiently large dimension.

2 The definition of the concept of smoothness of numbers and their classification

The general form of the discrete logarithm equation is represented by the expression: $c \equiv b^x(modm)$ and $c, b, x, m \in N$. There are four possible solutions to this equation: 1) set b, x, m, find c; 2) set c, b, m, find x; 3) set c, b, x, find m; 4) set c, x, m, find b.

The first equation is solved relatively simply, but it should be noted that if $c = 1$, m is a prime number, $m \in \mathrm{P}$, b is its primitive root and it is necessary to find the distribution law of primes having the same primitive root b, then we come to the solution of Artin's hypothesis(1927), $\pi(x, a) = c(b) \cdot \pi(x)$, where $\pi(x, a)$ is the number of primes $p \leq x$ with a primitive root b, $c(b)$ is Artin's constant, and $\pi(x)$ is the general number of primes $p \leq x$. However, this is true if $x \to \infty$.

The second problem is the discrete logarithm problem. As stated in [1] in general, the problem may turn out to be algorithmically unsolvable. If m is a prime number, that is $m = p \in \mathrm{P}$, then the algorithm exists, but its complexity depends on the complexity of factorization $p - 1 = \prod_{i=1}^{k} p_i^{\alpha_i}$, where p_i are prime factors, k their number. In case if the number to be factorized is a Sophie Germain number $p = 2\tilde{p}+1$ or $p = 2^{\alpha}\tilde{p}+1$, where $\tilde{p}$ there is also a prime number, the solution of the discrete logarithm problem becomes very complex in terms of computational costs. Note that if m the composite number and does not belong to the final fields $F(2^p)$ or $F(p^k)$, then non-simple algorithmic problems arisen in [2]. Cases 3 and 4 are practically not considered in the literature. It is easy to show that they are more complex than the discrete logarithm problem.

Pohlig and Hellman [5] proposed an algorithm for solving the problem of the discrete logarithm, provided that it p is a simple smooth number. In [3], the problem of smoothness is discussed. The authors note that there is still no clear interpretation of the concept of smoothness. One option is that the decomposition $p - 1 = \prod_{i=1}^{k} p_i^{\alpha_i}$ has the property that $p_k < \tilde{p}$, where the $\tilde{p}$ given prime number.

As noted above in the work [1], the author literally write: "for simplicity, we suppose that $(p - 1)$ is smooth and b its primitive root". This is where a really complicated problem arises. Since p can be arbitrarily large, it is not known how smooth primes are distributed. A search algorithm is needed for a smooth prime number of arbitrarily large size, and then to find its primitive root satisfying certain conditions. This is especially important if it is related to cryptographic tasks. But this is no less important from the point of view of the whole variety of problems formulated in the "MathOverflow" project for the discrete logarithm. For this reason, the problem of the distribution of smooth primes on the set of all primes was considered. In the work of E. Kowalski [6] the Erdцs-Kac theorem is given, which states that for any natural number $n \in [1; N]$, for any real $a < b$, the following holds: $\lim\limits_{N\to\infty} \frac{1}{N}|\{1 \leq n \leq N | a \leq \frac{\omega(n)-\log\log N}{\sqrt{\log\log N}} \leq b\}| = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$, where $\omega(n)$ the number of prime dividers, without taking into account their degrees.

From the theorem it follows that by the number of simple factors in their factorization on the set of natural numbers by the frequency of their appearance the natural numbers n obey the normal distribution law. But the question arises to what extent

is this true for numbers like $(p-1)$, where $p \in \mathrm{P}$. On the basis of data related to the solution by computer and analytical methods of the generalized Artin hypothesis, it was established [7] that for smooth prime numbers $p-1=\prod_{i=1}^{k} p_i^{\alpha_i}$ with increasing k, the value $\omega(p-1)$ obeys most likely the lognormal distribution law. It is possible that this may turn out to be another distribution law, similar to a lognormal one. This fact requires clarification. From the experimental data on the set of the first 10 million primes it follows that smooth numbers are very rare. Therefore, Koblitz's statement [1] "for simplicity, we suppose that $(p-1)$ is smooth ..." is completely incorrect, since the search for such a simple number may be more difficult than solving the problems of the discrete logarithm. In addition, to find a primitive root for p from $\varphi(p-1)$ (Euler function) possible variants from the set of natural numbers $\{2, ..., p-1\}$ is not at all easy, it follows from the theorem on the cyclic group [4]. For this reason, the study of this problem is no less relevant than the solution of the problem of the discrete logarithm. We present the results of a statistical analysis of the first 10 million primes and decomposition $(p-1)$ into prime factors. On the horizontal axis, the number

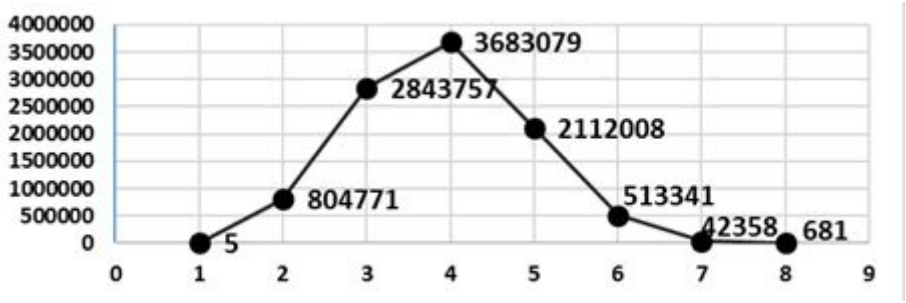

Figure 1: Results of decomposition $(p-1)$ into simple factors

of simple factors is plotted, on the vertical, the number of the corresponding prime numbers. Based on the obtained results, it can be seen that with increasing values of prime numbers, the number of factors increases, from which it can be concluded that the set of numbers $(p-1)$ does not obeys a normal distribution law. Figure 1 confirms that the lognormal form of distribution is clear.

Depending on the sequence of increasing simple factors, smooth numbers can be classified as follows:

1) Perfectly smooth primes - such numbers as for $p-1=\prod_{i=1}^{k} p_i^{\alpha_i}$, $p_i = 2, 3, 5, ..., p_k$. This means that all prime factors are consecutive prime numbers.

Pohlig and Hellman argue that such a type of smooth numbers will allow solving the discrete logarithm problem using an algorithm with polynomial complexity [5]. However, along with the definition of a perfectly smooth number, the question arises as to the number of such numbers. It may happen that the search for a smooth number of large dimensionality will have a high algorithmic complexity, which in turn makes the use of the Silver-Pohlig-Hellman algorithm impractical.

2) Partially smooth prime numbers. It is assumed that for such numbers the sequence of simple factors is not necessarily consecutive prime numbers, but the difference between the consecutive factors should not be too large. Partially smooth primes can

also be used for the algorithm of Silver, Pohlig, and Hellman, but its effectiveness will depend on a measure of the smoothness of such numbers.

Among the first 10 million $(p-1)$, perfectly smooth numbers with more than 6 factors, there are no more than a few dozen. From this we can conclude that the assumption of smoothness does not simplify the solution of the discrete logarithm problem, but even complicates if we take into account the iterations for each factor $p_i^{\alpha_i}$ with α_i substantially more than 1 and systematic use of the Chinese theorem on residuals with a large number of simple factors.

In conclusion, we note that this paper shows that when solving the problem of a discrete logarithm, it is necessary to take into account the properties of some types of prime numbers. To solve the problem of discrete logarithm using the Silver-Pohlig-Hellman algorithm, an important point is the use of smooth numbers that must satisfy $(p-1) > 10^{300}$. However, it is still not known how to find such numbers. This makes the task difficult to compute.

The next important factor in solving the discrete logarithm problem is a measure of the smoothness of a prime number, since the speed of the algorithm execution directly depends on this. The construction of a measure of smoothness and an analysis of the algorithm depending on the smoothness of the number is a topic for further study. The problem of constructing a measure of smoothness must be considered from two sides: from the point of view of the difference between adjacent factors of simple smooth numbers and depending on the exponent of each of the factors.

References

[1] Koblitz N. (2001). *Course of number theory and cryptography*. Scientific publishing house PTA, Moscow.

[2] Pomerance C. (2008). Elementary thoughts on discrete logarithms. *Algorithmic Number Theory, MSRI Publications*. Vol. **44**

[3] Crandall R., Pomerance C. (2001). *Prime numbers: A Computational Perspective.*. Springer-Verlag, New York.

[4] Lidl R., Niederreiter G. (1997). *Finite Fields*. Cambridge University Press.

[5] Pohlig S.C., Hellman M.E. (1978). An Improved Algorithm for Computing Logarithms Over GF(p) and its Cryptographic Significance. *IEEE Transactions on Information Theory*. Vol. **1**, pp. 106-110.

[6] Kowalski E. (2018). Arithmetic Randomnnee. An introduction to probabilistic number theory. Available at: https://people.math.ethz.ch/ kowalski/probabilistic-number-theory.pdf

[7] Vostrov G., Opiata R. (2018). Generalized Artina hypothesis and computer information model its solutions. *ELTECS*. Vol. **29**, pp. 120-126.