

The role of semismooth numbers in factoring large numbers

W.H. Ekkelkamp

21st June 2007

1 Introduction

One of the popular systems for encrypting and decrypting messages is the RSA cryptosystem. Its safety is based on the assumption that factoring large numbers is hard. Asymptotically, the best algorithm known for factoring large numbers is the number field sieve, which has a sub-exponential running time.

The most time consuming step in factoring with the number field sieve or the (related) quadratic sieve is the sieving. After initializing the algorithm with some suitable polynomials, we factor many polynomial values and keep the values that are B -smooth. A B -smooth value is a number with all its prime factors up to B . As soon as we have little more B -smooth values than the number of primes below B , we can factor the number we started with.

If we know more about the density of B -smooth numbers, we can estimate how many polynomial values we need to factor and how long the sieving step will take. A well-known approximation of the number $\Psi(x, x^\alpha)$ - values at most x that are x^α -smooth with $0 < \alpha < 1$ - is given by $x\rho(1/\alpha)$, where ρ is the so-called Dickman ρ function, which is the unique continuous solution of the differential-difference equation

$$\begin{cases} \rho(t) = 1 & 0 \leq t \leq 1 \\ \rho'(t) = -\rho(t-1)/t & t \geq 1. \end{cases}$$

The results based on this approximation are reasonable, but for more accuracy it is better also to use the second order term, introduced by Ramaswami [8]:

$$\Psi(x, x^\alpha) = x\rho(1/\alpha) + (1 - \gamma)\frac{x}{\log x}\rho\left(\frac{1 - \alpha}{\alpha}\right) + o\left(\frac{x}{\log x}\right), \quad x \rightarrow \infty,$$

where γ is Euler's constant.

A more efficient way of factoring uses B -smooth numbers with additionally one or two prime factors between B and a larger bound L , the so-called large prime(s). Such numbers are called *semismooth*. Of course, one would like to know how many such numbers one may expect and whether it is better to include the second order term.

If we take a closer look at semismooth numbers with one large prime, include the second order term in our analysis and take $B = x^\alpha$, $L = x^\beta$, we find for

their number

$$\Psi_1(x, x^\beta, x^\alpha) = x \int_\alpha^\beta \rho\left(\frac{1-\lambda}{\alpha}\right) \frac{d\lambda}{\lambda} + (1-\gamma) \frac{x}{\log x} \int_\alpha^\beta \rho\left(\frac{1-\lambda-\alpha}{\alpha}\right) \frac{d\lambda}{\lambda(1-\lambda)} + o\left(\frac{x}{\log x}\right), x \rightarrow \infty.$$

In Section 3 we give a more detailed error term. We have derived a similar expression for semismooth numbers with two large primes, as we will see in Section 4. Note that we have an extra degree of freedom here, as we can take different upper bounds for the two large primes.

Zhang [9] has proved that there exist functions $F_1(\alpha, \beta)$ and $G_1(\alpha, \beta)$ such that $\Psi_1(x, x^\beta, x^\alpha) = xF_1(\alpha, \beta) + \frac{x}{\log x}G_1(\alpha, \beta) + O\left(\frac{x}{\log^2 x}\right)$, but his functions F_1 and G_1 are presented in a more complicated way. His error analysis is a generalization of the work of Knuth and Trabb Pardo [5], whereas our analysis is a generalization of the work of de Bruijn, Ramaswami, and Bach and Peralta.

2 Smooth numbers

Let $n = n_1 n_2 \dots$ with n_i prime and $n_1 \geq n_2 \geq \dots$ (where $n_j = 1$ if n has fewer than j prime factors). Then the number of values at most x that are y -smooth can be written as $\Psi(x, y) = \#\{n \leq x : n_1 \leq y\}$. Based on work of De Bruijn [2], we have

Theorem 1 *For any fixed $\epsilon > 0$ the relation*

$$\Psi(x, x^\alpha) = x\rho\left(\frac{1}{\alpha}\right) \left(1 + O\left(\frac{\log(1/\alpha + 1)}{\alpha \log x}\right)\right), \text{ as } x \rightarrow \infty,$$

holds uniformly in the range $x^\alpha \geq 2, 1 \leq \frac{1}{\alpha} \leq \exp((\alpha \log x)^{3/5-\epsilon})$.

This theorem is due to Hildebrand [3, 4]. A more precise result was obtained by Ramaswami [8]. We follow the formulation of Norton ([7], p. 12).

Theorem 2 *For $x > 1, 0 < \alpha < 1$, and $x^\alpha > 2$ we have*

$$\Psi(x, x^\alpha) = x\rho\left(\frac{1}{\alpha}\right) + (1-\gamma) \frac{x}{\log x} \rho\left(\frac{1-\alpha}{\alpha}\right) + O(\Delta(x, x^\alpha)), \text{ as } x \rightarrow \infty,$$

where

$$\Delta(x, x^\alpha) = \begin{cases} \frac{x}{(\log x)^{3/2}} & \text{for } 0 < \alpha < 1/2, \\ \frac{x^\alpha}{\log x} + \frac{x}{\log^2 x} & \text{for } 1/2 \leq \alpha < 1. \end{cases}$$

In this theorem γ is Euler's constant. Our results on semismooth numbers are based on these two theorems.

3 1-Semismooth numbers

A 1-semismooth number is a smooth number with all its prime factors below a certain bound y_2 , except for one prime factor $> y_2$, but smaller than a larger

bound y_1 . The analogue of the Ψ -function for smooth numbers is defined for 1-semismooth numbers as

$$\Psi_1(x, y_1, y_2) = \#\{n \leq x : y_2 < n_1 \leq y_1, n_2 \leq y_2\}.$$

An approximating function is given by the following theorem, which follows directly from Theorem 3.1 in an article of Bach and Peralta [1].

Theorem 3 *If $0 < \alpha < \beta < 1$ and $x^\alpha \geq 2$, then*

$$\Psi_1(x, x^\beta, x^\alpha) = x \int_\alpha^\beta \rho\left(\frac{1-\lambda}{\alpha}\right) \frac{d\lambda}{\lambda} + O\left(\frac{\log(1/\alpha)}{\alpha(1-\beta)} \frac{x}{\log x}\right).$$

Compared with Theorem 3.1 in [1], where only the condition $0 < \alpha < \beta < 1$ is stated, we have an additional condition. This condition originates from the application of a result of de Bruijn [2]. The proof of Theorem 3 can be found in [1]. Here we present the following refinement.

Theorem 4 *If $0 < \alpha < \beta < 1$, $\alpha + \beta < 1$, and $x^\alpha \geq 2$, then we have for $x \rightarrow \infty$*

$$\begin{aligned} \Psi_1(x, x^\beta, x^\alpha) = x \int_\alpha^\beta \rho\left(\frac{1-\lambda}{\alpha}\right) \frac{d\lambda}{\lambda} + \frac{(1-\gamma)x}{\log x} \int_\alpha^\beta \rho\left(\frac{1-\lambda-\alpha}{\alpha}\right) \frac{d\lambda}{\lambda(1-\lambda)} + \\ O\left(\left(\frac{\log(\beta/\alpha)}{\alpha^{3/2}} \frac{x}{\log^{3/2} x}\right) + \left(\frac{x^{\alpha+\beta}}{\alpha \log x}\right)\right). \end{aligned}$$

The main ingredients of the proof are the definition of 1-semismooth numbers, Theorem 2, partial integration, the prime number theorem $\pi(x) = \text{li}(x) + O(x/\log^c x)$ for any $c > 1$, and careful estimations of the error terms.

We compared both approximating functions with experimental results for the multiple polynomial quadratic sieve (MPQS). We computed the expected number of 1-semismooth numbers after sieving 10096 polynomials and compared this with the real sieving experiment. The second term adds about 10 % to the main term.

x	x^α	x^β	1 term	2 terms	experiment
9.26 E44	2.5 E5	2.5 E7	13 205	14 657	14 884
1.94 E50	3.0 E5	3.0 E7	935	1040	929
2.16 E55	2.5 E5	5.0 E7	25	28	29
3.81 E60	7.5 E5	3.0 E8	63	70	72

4 2-Semismooth numbers

In this section we extend the definition of a 1-semismooth number to two large primes. We recall that a 2-semismooth number is a number with all but two of its prime factors below a certain bound y_2 , whereas the other two prime factors are $> y_2$, but $\leq y_1$. The definition of the corresponding Ψ -function is

$$\Psi_2(x, y_1, y_2) = \#\{n \leq x : y_2 < n_2 \leq n_1 \leq y_1, n_3 \leq y_2\}.$$

Lambert has given an approximating function for $\Psi_2(x, y_1, y_2)$ in his thesis [6], consisting of a main term and an error term. However, it may be useful to

choose a smaller upper bound for the second largest prime. The corresponding Ψ -function becomes

$$\Psi_2(x, y_1, y_2, y_3) = \#\{n \leq x : n_2 < n_1 \leq y_1, y_3 < n_2 \leq y_2, n_3 \leq y_3\},$$

with $y_3 < y_2 \leq y_1$. If $y_2 = y_1$, we have the same upper bound for both large primes, so it suffices to give the results for the last Ψ -function. Based on Theorem 1, we have the following theorem.

Theorem 5 *Let $\epsilon > 0$ be fixed. If $0 < \alpha < \beta_2 < \beta_1$, $\alpha + \beta_2 + \beta_1 \leq 1$, $x^\alpha \geq 2$, and $\frac{1-2\alpha}{\alpha} \leq \exp((\frac{\alpha}{1-2\alpha} \log x)^{3/5-\epsilon})$, then we have for $x \rightarrow \infty$,*

$$\Psi_2(x, x^{\beta_1}, x^{\beta_2}, x^\alpha) = x \left(\int_\alpha^{\beta_2} \int_{\lambda_2}^{\beta_1} \rho \left(\frac{1 - \lambda_1 - \lambda_2}{\alpha} \right) \frac{d\lambda_1}{\lambda_1} \frac{d\lambda_2}{\lambda_2} \right) \left(1 + O \left(\frac{\log(\frac{1}{\alpha})}{\alpha \log x} \right) \right).$$

We will compute the second order term in the next theorem, based on Theorem 2.

Theorem 6 *If $0 < \alpha < \beta_2 < \beta_1$, $\alpha + \beta_2 + \beta_1 < 1$, and $x^\alpha \geq 2$, then we have for $x \rightarrow \infty$*

$$\begin{aligned} \Psi_2(x, x^{\beta_1}, x^{\beta_2}, x^\alpha) &= x \int_\alpha^{\beta_2} \int_{\lambda_2}^{\beta_1} \rho \left(\frac{1 - \lambda_1 - \lambda_2}{\alpha} \right) \frac{d\lambda_1}{\lambda_1} \frac{d\lambda_2}{\lambda_2} + \\ &(1 - \gamma) \frac{x}{\log x} \int_\alpha^{\beta_2} \int_{\lambda_2}^{\beta_1} \rho \left(\frac{1 - \lambda_1 - \lambda_2 - \alpha}{\alpha} \right) \frac{1}{1 - \lambda_1 - \lambda_2} \frac{d\lambda_1}{\lambda_1} \frac{d\lambda_2}{\lambda_2} + \\ &O \left(\left(\frac{\log(\beta_1/\alpha) \log(\beta_2/\alpha)}{\alpha^{3/2}} \frac{x}{\log^{3/2} x} \right) + \left(\frac{x^{\alpha+\beta_1+\beta_2}}{\alpha \log x} \right) \right). \end{aligned}$$

The proof consists of the same ingredients as the proof of Theorem 4. We start with the largest prime and establish an expression for it. Then we repeat the arguments for the second largest prime and get the approximating function as stated in Theorem 6.

We expect similar improvements as for 1-semismooth numbers, when using the second order term.

Zhang [9] has proved that there exist functions $F_2(\alpha, \beta)$ and $G_2(\alpha, \beta)$ such that $\Psi_2(x, x^\beta, x^\alpha) = xF_2(\alpha, \beta) + \frac{x}{\log x}G_2(\alpha, \beta) + O(\frac{x}{\log^2 x})$. The functions F_2 and G_2 are not given explicitly, but defined recursively. Zhang has not considered the case that the large primes have different upper bounds.

5 Conclusions

We have estimated numbers of smooth and semismooth numbers. We have extended these results even further to k large primes with $k \in \mathbb{N}$, but will publish this elsewhere.

Experiments with MPQS indicate that the use of the second order terms contributes about 10 % for x of the size 10^{50} . Most likely the same is true for the number field sieve, but we have not yet verified this.

References

- [1] Bach, E., Peralta, R.: Asymptotic semismoothness probabilities. *Math. Comp.* **65** (1996) 1701–1715
- [2] de Bruijn, N.G.: On the number of positive integers $\leq x$ and free of prime factors $> y$. *Proc. Kon. Ned. Akad. Wet.* **A54** = *Indag. Math.* **13** (1951) 50–60
- [3] Hildebrand, A.: On the number of positive integers $\leq x$ and free of prime factors $> y$. *J. Number Theory* **22** (1986) 289–307
- [4] Hildebrand, A., Tenenbaum, G.: Integers without large prime factors. *J. Théor. Nombres Bordeaux* **5** (1993) 411–484
- [5] Knuth, D.E., Trabb Pardo, L.: Analysis of a simple factorization algorithm. *Theoret. Comput. Sci.* **3** (1976) 321–348
- [6] Lambert, R.: Computational aspects of discrete logarithms. Ph.D. thesis, University of Waterloo (1996)
- [7] Norton, K.K.: Numbers with Small Prime Factors, and the Least k th Power Non-Residue. *Memoirs of the American Mathematical Society*, No. 106 (1971)
- [8] Ramaswami, V.: The number of positive integers $\leq x$ and free of prime divisors $> x^c$, and a problem of S.S. Pillai. *Duke Math. J.* **16** (1949) 99–109
- [9] Zhang, C.: An extension of the Dickman function and its application. Ph.D. thesis, Purdue University (2002)