Smooth numbers, sieve theory and the quadratic sieve

Torben Hansen, 20091714

Aarhus University - Institute of Mathematics

In this measure point I will discuss smooth numbers and how they are used in the quadratic sieve. A chapter of the general sieve problem is also included.

The quadratic sieve originates from the ancient problem: "Given an integer n, how do we factor n?". Actually this question isn't difficult to answer; just use trial division. So let me restate the problem: "Given an integer n, how do we factor n **fast**?". Anyone reading this would know that this turns out to be an extremely difficult task. We will use the Quadratic sieving factorisation method for "motivating" our treatment of smooth numbers, in particular the counting function for smooth numbers. In the end we will calculate a measure of the likelihood that we pick a smooth number in an interval assuming a special bound.

1 The quadratic sieve

In this paragraph we will introduce the fundamentals of the quadratic sieve factorisation method. Take an integer n that we would like to factor, how could we proceed? Well, the first thing we could do is to assume that n is odd, otherwise we could divide out the two's. The following lemma is the basic of a century old factoring method by Fermat.

Lemma 1. *If $n \in \mathbb{N}$ is odd, there exists $a, b \in \mathbb{N}$ such that $n = a^2 - b^2$.*

Proof. We can assume $n = 2r + 1$ for $r \in \mathbb{N}$. Pick $a = r + 1$ and $b = r$ then

$$a^2 - b^2 = (r+1)^2 - r^2 = r^2 + 1 + 2r - r^2 = 2r + 1 = n$$

□

With the formula $a^2 + b^2 = (a - b)(a + b)$ we wish to obtain a nontrivial factorization of n. If we picked a and b as in the above lemma we would get the boring factorization $n = 1 \cdot n$. But if $n = 8051$ then $n = 90^2 - 7^2$ and this give $n = (90 - 7)(90 + 7) = 83 \cdot 97$ that is there are other possible picks of a and b in lemma 1.

Definition 1. If the pair (a, b) with $a, b \in \mathbb{N}$ satisfy $a \not\equiv \pm b \mod n$ we shall call the pair (a, b) interesting and uninteresting if $a \equiv \pm b \mod n$.

The next theorem is extremely promising.

Theorem 1. *If there exists $a, b \in \mathbb{N}$ with (a, b) interesting and $a^2 \equiv b^2 \mod n$ then $\gcd(a \pm b, n) > 1$ and further $\gcd(a \pm b, n)$ are nontrivial factors of n.*

Proof. $a^2 \equiv b^2 \mod n$ is equivalent to n dividing the product $(a-b)(a+b)$. Now if $\gcd(a-b, n) = 1$ then n would divide the other factor, i.e. $n|a + b$. But this cant be, since (a, b) is an interesting pair that is $\gcd(a - b, n) > 1$. With the same argument we see that $\gcd(a + b, n) > 1$. Now since $\gcd(a \pm b, n)$ divides both $a \pm b$ and n and is greater than one, $\gcd(a \pm b, n)$ has to be nontrivial factors of n. □

But how to find a and b? We can venture a guess and hope for the best. But this is as slow, well actually slower, than doing trial division. Remember we wish to find an interesting pair (a, b) such that $a^2 - b^2 = n$. Rearranging we get $a^2 - n = b^2$. That is we can recast the previous problem to the problem of figuring out when $a^2 - n$ is a perfect square which at least would give us a pair (a, b) - interesting or not. Lets try it

Example 1. Let $n = 1649$ and let us work through the sequence $x^2 - n$ with $x = \lceil n^{\frac{1}{2}} \rceil, \lceil n^{\frac{1}{2}} \rceil + 1, \ldots$:

$$41^2 - n = 32, \quad 42^2 - n = 115, \quad 43^2 - n = 200$$

This didn't gave any perfect squares in immediate sight. Of course we could proceed by taking bigger and bigger x but if the factors aren't near the square root of n, we will have to iterate through many a to find b. And if we find a pair (a, b) this could also turn out to be uninteresting! But notice now the fantastic property: $32 \cdot 200 = 6400 = 80^2$ a perfect square! And since $41^2 \equiv 32 \mod n$ and $43^2 \equiv 200 \mod n$ we have $(41 \cdot 43)^2 = 41^2 \cdot 43^2 \equiv 80^2 \mod n$ which is a solution to $a^2 \equiv b^2 \mod n$.

This is a promising method, but it will break down if we can't ensure the existence of a pair (a, b) satisfying the conditions in theorem 1. Luckily we have the following lemma. Which we state without proof

Lemma 2. *If n has at least two different odd prime factors, then more than half of the solutions to $a^2 \equiv b^2 \mod n$ with $\gcd(ab, n) = 1$ satisfy $a \equiv \pm b \mod n$.*

We can now comfortably pursue what looked like godlike luck above. Look at the sequence $\left\{x^2 - n\right\}_{x=\lceil n^{\frac{1}{2}} \rceil, \lceil n^{\frac{1}{2}} \rceil + 1, \ldots}$ From this we would like to pick out a subsequence with product a perfect square. This yield a couple of questions: "Do such a subsequence exists?", "and how do we find the subsequence and how many terms of the sequence do we need?"

To answer the above questions we need the notation of smooth numbers, specifically B-smooth numbers.

Definition 2. A number $m \in \mathbb{N}$ is B-smooth if all of its prime factors are $\leq B$

The motivation for introducing smooth numbers is the following: Assume m is a number in a our sequence $\left\{x^2 - n\right\}_{x=\lceil n^{\frac{1}{2}} \rceil, \lceil n^{\frac{1}{2}} \rceil + 1, \ldots}$, which is not B-smooth. Then it is divisible by a large prime, say $p > B$. If m has to be contained in a subsequence product a square either $p^2 | m$ or another number in the subsequence has to be divisible by p i.e. a multiple of p. Recall the way we are searching for these numbers: $x^2 - n$, for $x = \lceil n^{\frac{1}{2}} \rceil, \lceil n^{\frac{1}{2}} \rceil + 1, \ldots$. Then Since p is large either m has to be large because $p^2 | m$ or we have to find a mate for m but these are few and far between.

The next lemma, which is crucial, shows that in a big enough sequence it will always be possible to find a subsequence that has product a square.

Lemma 3. *Let $\{m_i\}_{i=1}^k$ be a sequence of B-smooth numbers for $k \in \mathbb{N}$ and let $\pi(B) = \#$primes in the interval $[1, B]$. If $k > \pi(B)$ then there exists a subsequence $\left\{m_{i_j}\right\}_{j=1}^s$ with product a perfect square(i.e. $\prod_{j=1}^s m_{i_j} =$perfect square)*

Proof. Let m be a B-smooth number and write m in its prime decomposition

$$m = \prod_{i=1}^{\pi(B)} p_i^{v_{p_i}(m)}$$

where p_i is the i'th prime and $v_{p_i}(m) = \#\ p_i$ *in the prime decomposition of* m. Notice that we only need the primes up to $\pi(B)$ because m is B-smooth. Write $\mathbf{v}(m) = (v_{p_1}(m), v_{p_2}(m), \ldots, v_{p_{\pi(B)}}(m))^T$. Consider an arbitrary subsequence $\{m_{i_r}\}_{r=1}^{r'}$ $(1 \leq r' \leq \pi(B))$ of the sequence $\{m_i\}_{i=1}^k$. $\{m_{i_r}\}_{r=1}^{r'}$ has product a square if and only if every entrance in

$$\mathbf{v}(m_{i_1}) + \mathbf{v}(m_{i_2}) + \cdots + \mathbf{v}(m_{i_{r'}}) = \begin{pmatrix} v_{p_1}(m_{i_1}) + v_{p_1}(m_{i_2}) + \cdots + v_{p_1}(m_{i_{r'}}) \\ v_{p_2}(m_{i_1}) + v_{p_2}(m_{i_2}) + \cdots + v_{p_2}(m_{i_{r'}}) \\ \vdots \\ v_{p_{\pi(B)}}(m_{i_1}) + v_{p_{\pi(B)}}(m_{i_2}) + \cdots + v_{p_{\pi(B)}}(m_{i_{r'}}) \end{pmatrix}$$

is even. But this is true if and only if

$$v_{p_j}(m_{i_1}) + v_{p_j}(m_{i_2}) + \cdots + v_{p_j}(m_{i_{r'}}) = \sum_{r=1}^{r'} v_{p_j}(m_{i_r}) \equiv 0 \mod 2 \quad \forall j = 1, 2, \ldots, \pi(B)$$

if and only if

$$\mathbf{v}(m_{i_1}) + \mathbf{v}(m_{i_2}) + \cdots + \mathbf{v}(m_{i_{r'}}) \equiv 0 \mod 2$$

where we interpret this as taking module 2 in every entrance. We get that if we can find a subsequence with the sum of the associated exponent vectors equalling 0(mod 2) we know that the subsequence has product a square.

Let $\mathbb{F}_2$ be the finite field of 2 elements. Then $\mathbb{F}_2^{\pi(B)}$ is a vectorspace over $\mathbb{F}_2$ and $\dim \mathbb{F}_2^{\pi(B)} = \pi(B)$. $\{\mathbf{v}(m_i)\}_{i=1}^k$ is contained in $\mathbb{F}_2^{\pi(B)}$ if we take mod 2 in every entrance in every vector in the sequence. But since this vectorspace only has dimension $\pi(B)$ and we have strictly more vectors in the sequence $\{\mathbf{v}(m_i)\}_{i=1}^k$ they must be linearly dependent! In other words there must exists a subsequence $\left\{\mathbf{v}(m_{i_j})\right\}_{j=1}^{k'}$ with $k' < k$ and a vector $\mathbf{v}(m_{i_{k'+1}})$ not contained in the subsequence such that

$$\mathbf{v}(m_{i_1}) + \mathbf{v}(m_{i_2}) + \cdots + \mathbf{v}(m_{i_{k'}}) + \mathbf{v}(m_{i_{k'+1}}) = 0$$

in $\mathbb{F}_2^{\pi(B)}$. This means

$$\mathbf{v}(m_{i_1}) + \mathbf{v}(m_{i_2}) + \cdots + \mathbf{v}(m_{i_{k'}}) + \mathbf{v}(m_{i_{k'+1}}) \equiv 0 \mod 2$$

and by the above condition $\left\{m_{i_j}\right\}_{j=1}^{k'+1}$ must be a subsequence with product a square. □

This subsequence could return a uninteresting pair (a, b), but by theorem 1, this wont happen that often.

Notice how the lemma uses linear algebra and our sequence $\left\{x^2 - n\right\}_{x=\lceil n^{\frac{1}{2}} \rceil, \lceil n^{\frac{1}{2}} \rceil + 1, \ldots}$ to find the subsequence we want so desperately and it is done by only knowing the B-smooth numbers in some sequence. We can now write down a prototype algorithm:

(1) Choose the bound B, and look for B-smooth numbers in the sequence $\left\{x^2 - n\right\}_{x=\lceil n^{\frac{1}{2}} \rceil, \lceil n^{\frac{1}{2}} \rceil + 1, \ldots}$.

(2) Collect $\pi(B)$ B-smooth numbers from the sequence. Now lemma 3 imply that we can use linear algebra to find a subsequence $x_1^2 - n, x_2^2 - n, \ldots, x_k^2 - n$ that has product a square, say a^2.

(3) Let $[r]_n$ be the smallest nonnegative(hence unique) remainder of r module n. Calculate $[a]_n$ and $[x_1 x_2 \cdots x_k]_n$.

(4) We have $[a]_n^2 \equiv [x_1 x_2 \cdots x_k]_n^2 \mod n$. If $([a]_n, [x_1 x_2 \cdots x_k]_n)$ is an interesting pair, then compute $\gcd([a]_n \pm [x_1 x_2 \cdots x_k]_n, n)$ otherwise find more B-smooth numbers from $\left\{x^2 - n\right\}_{x=\lceil n^{\frac{1}{2}} \rceil, \lceil n^{\frac{1}{2}} \rceil + 1, \ldots}$ and return to step 2.

The above algorithm is not complete. First of all there is a shit lode of technical stuff regarding which data structure to use so the algorithm will be efficient. This is of course interesting and most of all also an important ingredient in the algorithm, but it is something we won't discuss further other than stating that it can be done. Secondly is how we efficiently can calculate the residue in step 3. Still this is important but again not something we will care about here. Third, how do we choose the bound B and how can we examine the sequence in step 1 for B-smooth numbers and discard the rest? The bound question is a pretty hard optimization problem while the other question uses a sieving process much similar to the Sieve of Eratosthenes which is known by any competent mathematician. The sieving process won't be discussed, but we will turn our attention to sieving theory in general and analyse the B-smooth numbers, especially the number of B-smooth numbers.

2 General sieve problem and a couple of useful theorems

Here we will shortly describe the general sieve problem and state and prove some useful theorems.

In the up most generality the sieve problem can be stated as follow: Let A be a finite set of integers(In the extreme general case, this i taken as a finite set of objects) and P an index set of primes such that for each $p \in P$ there is associated $A_p \subset A$. The goal is to produce upper and lower bounds of the set $S(A,P) = A \setminus \bigcup_{p\in P} A_p$. Recall the inclusion-exclusion principle: Let $A_1, A_2, \ldots, A_m$ be finite sets then

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n} |A_i| - \sum_{1\le i<j\le n} |A_i \cap A_j| + \sum_{1\le i<j<k\le n} |A_i \cap A_j \cap A_k| - \cdots + (-1)^{n-1}|A_1 \cap \cdots \cap A_n|$$

If we assume that the index set P is finite we can easily apply this to the set $S(A,P)$ in the following manner: Let $I \in \mathcal{P}(P)$ and define $A_I := \bigcap_{p\in I} A_p$ with the convention $A_\emptyset = A$. Write $P = \{p_1, p_2, \ldots, p_n\}$ then

$$\begin{aligned}
|S(A,P)| &= \left| A \setminus \bigcup_{p\in P} A_p \right| = |A| - \left| \bigcup_{p\in P} A_p \right| \\
&= |A| - \left(\sum_{i=1}^{n} |A_{p_i}| - \sum_{1\le i<j\le n} |A_{p_i} \cap A_{p_j}| + \cdots + (-1)^{n-1}|A_{p_1} \cap \cdots \cap A_{p_n}| \right) \\
&= |A| - \left(\sum_{\substack{I\subset P \\ |I|=1}} |A_I| - \sum_{\substack{I\subset P \\ |I|=2}} |A_I| + \cdots + (-1)^{|P|-1}|A_P| \right) \\
&= (-1)^{|\emptyset|}|A_\emptyset| + \sum_{\substack{I\subset P \\ |I|=1}} (-1)^{|I|}|A_I| + \sum_{\substack{I\subset P \\ |I|=2}} (-1)^{|I|}|A_I| + \cdots + \sum_{\substack{I\subset P \\ |I|=n}} (-1)^{|P|}|A_P| \\
&= \sum_{I\subset P} (-1)^{|I|}|A_I|
\end{aligned}$$

Notice how this gives an explicit formula for the cardinality of $S(A,P)$ but sadly the information needed for calculating this is (very) rarely known.

Example 2. A concrete construction could be the following: Fix $x \in \mathbb{Z}$ with $x > 0$, let $A = [1,x] \cap \mathbb{N}$, P the set of primes and $A_p = \{y \in A \mid p|y\}$ then $|S(A,P)| = \varphi(x)$ where φ denotes the Euler totient function.

Let me also present another example

Example 3. Assume you are giving an interval $[1,x]$ for some positive integer x followed with the question: "How many primes are there in this interval?". Let us derive a rough upper bound. Let $P_z = \prod_{p<z} p$ and define the function

$$\phi(x,z) = |\{n \le x \mid \gcd(n,P_z) = 1\}|$$

where x, z are positive real numbers. Now let μ denote the Möbius function and recall the fundamental property

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{otherwise} \end{cases}$$

By this we can write

$$\sum_{d|gcd(y,P_z)} \mu(d) = \begin{cases} 1 & \text{if } y \in \{n \le x \mid \gcd(n,P_z) = 1\} \\ 0 & \text{otherwise} \end{cases}$$

In sieve theory this kind of function is often called the sifting function. We can now calculate

$$\begin{aligned}\phi(x,z) &= \sum_{n\leq x}\sum_{d|gcd(n,P_z)}\mu(d)\\ &= \sum_{d|P_z}\mu(d)\sum_{\substack{n\leq x\\ d|n}}1\\ &= \sum_{d|P_z}\mu(d)\left\lfloor\frac{x}{d}\right\rfloor\\ &= \sum_{d|P_z}\mu(d)\left(\frac{x}{d}+\left\lfloor\frac{x}{d}\right\rfloor-\frac{x}{d}\right)\\ &= \sum_{d|P_z}\mu(d)\frac{x}{d}+\sum_{d|P_z}\mu(d)\left(\left\lfloor\frac{x}{d}\right\rfloor-\frac{x}{d}\right)\end{aligned}$$

Notice $\frac{x-1}{d}\leq\left\lfloor\frac{x}{d}\right\rfloor\leq\frac{x}{d}$ imply $\left|\left\lfloor\frac{x}{d}\right\rfloor-\frac{x}{d}\right|\leq 1$ hence $\left|\mu(d)\left(\left\lfloor\frac{x}{d}\right\rfloor-\frac{x}{d}\right)\right|\leq 1$ and we get a bound on the last sum

$$\phi(x,z)=\sum_{d|P_z}\mu(d)\frac{x}{d}+O(2^{\pi(z)})$$

Another property of the Möbius function is the identity $\prod_{p<z}\left(1-\frac{1}{p}\right)=\sum_{d|P_z}\mu(d)\frac{1}{d}$. Using this on the above we get

$$\phi(x,z)=x\prod_{p<z}\left(1-\frac{1}{p}\right)+O(2^{\pi(z)})$$

The inequality $1-x\leq e^{-x}$ is valid for all positive x which applied to $\prod_{p<z}\left(1-\frac{1}{p}\right)$ implies

$$\prod_{p<z}\left(1-\frac{1}{p}\right)\leq\prod_{p<z}e^{-\frac{1}{p}}=e^{-\sum_{p<z}\frac{1}{p}}$$

Theorem 5 below now yield the bound

$$\phi(x,z)\leq xe^{-\sum_{p<z}\frac{1}{p}}+O(2^{\pi(z)})=x(\log z)^{-1}e^{O(1)}+O(2^{\pi(z)})$$

This estimate has a huge error term but nonetheless by picking $z=O(\log x)$ and observing $\pi(x)=(\pi(x)-\pi(z))+\pi(z)\leq\phi(x,z)+\pi(z)\leq\phi(x,z)+z$ we get

$$\pi(x)=O\left(\frac{x}{\log\log x}\right)$$

which yield a bound on the number of primes in the interval $[1,x]$. This bound is pretty awful and a much much better bound is known; $\pi(x)=O\left(\frac{x}{\log x}\right)$ often referred the prime number theorem.

We now present some useful results beginning with a technique called partial summation

Theorem 2. *Suppose $\{a_n\}_{n=1}^{\infty}$ is a sequence of real numbers and define $S(x)=\sum_{n\leq x}a_n$ for x a positive integer. Also suppose n_0 is a fixed positive integer with $a_i=0$ for $j<n_0$ and $x>n_0$ then if $f\in\mathcal{C}^1([n_0,\infty))$ we have*

$$\sum_{n\leq x}a_nf(n)=S(x)f(x)-\int_{n_0}^{x}S(t)\frac{d}{dt}f(t)dt$$

Proof. Notice $a_n = S(n) - S(n-1)$ so we can write

$$\begin{aligned}
\sum_{n\leq x} a_n f(n) &= \sum_{n\leq x} (S(n) - S(n-1))f(n) \\
&= \sum_{n\leq x} S(n)f(n) - \sum_{n\leq x} S(n-1)f(n) \\
&= \sum_{n\leq x} S(n)f(n) - \sum_{n\leq x-1} S(n)f(n+1) \\
&= S(x)f(x) + \sum_{n\leq x-1} S(n)f(n) - \sum_{n\leq x-1} S(n)f(n+1) \\
&= S(x)f(x) - \sum_{n\leq x-1} S(n)\left(f(n+1) - f(n)\right) \\
&= S(x)f(x) - \sum_{n\leq x-1} S(n)\int_n^{n+1} \frac{d}{dt}f'(t)dt \\
&= S(x)f(x) - \left(S(1)\int_1^2 \frac{d}{dt}f'(t)dt + S(2)\int_2^3 \frac{d}{dt}f'(t)dt + \cdots + S(x-1)\int_{x-1}^x \frac{d}{dt}f'(t)dt\right)
\end{aligned}$$

Since $S(n) = 0$ when $n < n_0$ we obtain

$$\begin{aligned}
\sum_{n\leq x} a_n f(n) &= S(x)f(x) - \Bigg(S(n_0)\int_{n_0}^{n_0+1} \frac{d}{dt}f'(t)dt \\
&\qquad + S(n_0+1)\int_{n_0+1}^{n_0+2} \frac{d}{dt}f'(t)dt + \cdots + S(x-1)\int_{x-1}^x \frac{d}{dt}f'(t)dt\Bigg) \\
&= S(x)f(x) - \sum_{n_0\leq n\leq x-1} S(n)\int_n^{n+1} \frac{d}{dt}f'(t)dt \\
&= S(x)f(x) - \int_{n_0}^x S(t)\frac{d}{dt}f'(t)dt
\end{aligned}$$

because $S(t)$ is constant on $[n, n+1)$(If t is not a integer we can interpret the sum as $S(t) = \sum_{n\leq \lfloor t \rfloor} a_n$). □

The above technique can be used to deduce

Theorem 3.

$$\sum_{n\leq x} \log n = x\log x - x + O(\log x)$$

Below is two theorems often used when working with sieve theory. We proof the first and give the idea of the second.

Theorem 4.

$$\sum_{p\leq n} \frac{\log p}{p} = \log n + O(1)$$

Proof. Consider $n!$ and notice by definition of the faculty function only primes below n can divide $n!$. Now decompose $n!$ into its prime factorization

$$n! = \prod_{p\leq n} p^{e_p}$$

For some power p^a we have $p^a > n$ that imply $\left\lfloor \frac{n}{p^a} \right\rfloor = 0$ that is e_p in the prime decomposition is giving by (if a_p is the largest positive integer such that $p^{a_p} \leq n$)

$$e_p = \sum_{i=1}^{a_p} \left\lfloor \frac{n}{p^i} \right\rfloor$$

Taking logarithm we obtain

$$\begin{aligned}
\log n! &= \log \left(\prod_{p \leq n} p^{e_p} \right) \\
&= \sum_{p \leq n} \log p^{e_p} \\
&= \sum_{p \leq n} e_p \log p \\
&= \sum_{p \leq n} \left(\log p \sum_{i=1}^{a_p} \left\lfloor \frac{n}{p^i} \right\rfloor \right)
\end{aligned}$$

We can get the bound

$$\begin{aligned}
\sum_{p \leq n} \left(\log p \sum_{i=2}^{a_p} \left\lfloor \frac{n}{p^i} \right\rfloor \right) &\leq \sum_{p \leq n} \left(\log p \sum_{i=1}^{a_p} \frac{n}{p^i} \right) \\
&= n \sum_{p \leq n} \left(\log p \sum_{i=2}^{a_p} \frac{1}{p^i} \right) \\
&= n \sum_{p \leq n} \log p \left(\frac{1 - p^{1-n}}{p(p-1)} \right) \\
&\leq n \sum_{p \leq n} \log p \left(\frac{1}{p(p-1)} \right) \\
&= O(n)
\end{aligned}$$

Using theorem 3 we also have

$$\log n! = \log 1 + \log 2 + \cdots + \log n = \sum_{i=1}^{n} \log i = n \log n - n + O(\log n)$$

Combining we get

$$\sum_{n \leq n} \left\lfloor \frac{n}{p} \right\rfloor \log p = n \log n + O(n)$$

and hence the result. □

By setting

$$a_n = \begin{cases} \frac{\log p}{p} & \text{if } n = p \\ 0 & \text{otherwise} \end{cases}$$

and using partial summation with $f(t) = \frac{1}{\log t}$ one can show

Theorem 5.

$$\sum_{p \leq n} \frac{1}{p} = \log \log n + O(1)$$

We will use the last two theorems in this chapter to derive an upper bound for the cardinality of smooth numbers in an interval.

3 Counting smooth numbers

In this paragraph we treat the smooth numbers more thoroughly. We have already defined smooth numbers in the former paragraph and therefore we go straight for the definition of the counting function for B-smooth numbers

Definition 3. Let $\psi(x,B)$ denote the number of B-smooth numbers in the interval $[1,x]$. More explicitly:

$$\psi(x,B) = |\{m \,|\, 1 \leq m \leq x,\ m \text{ is } B\text{-smooth}\}|$$

which could also be stated as

$$\psi(x,B) = |\{m \,|\, 1 \leq m \leq x, \text{ if } p|m \text{ then } p < B\}|$$

We are now interested in finding an upper for this function.

Theorem 6. *With the above definition we have*

$$\psi(x,B) = O\left(x(\log B)e^{-\frac{\log x}{\log B}}\right)$$

Proof. Pick $\delta > 0$ then $1 = 1^\delta \leq \left(\frac{x}{n}\right)^\delta$ for all $n \leq x$ so we can get

$$\psi(x,z) = \sum_{\substack{n\leq x \\ n \text{ B-smooth}}} 1 \leq \sum_{\substack{n\leq x \\ n \text{ B-smooth}}} \left(\frac{x}{n}\right)^\delta = x^\delta \sum_{\substack{n\leq x \\ n \text{ B-smooth}}} \frac{1}{n^\delta}$$

Now let $p_1, p_2, \ldots, p_k$ be the primes strictly below B then

$$\begin{aligned}
\prod_{p<B}\left(1-\frac{1}{p^\delta}\right)^{-1} &= \left(1-\frac{1}{p_1^\delta}\right)^{-1}\left(1-\frac{1}{p_2^\delta}\right)^{-1}\cdots\left(1-\frac{1}{p_k^\delta}\right)^{-1} \\
&= \left(\sum_{n=0}^{\infty}\left(\frac{1}{p_1^\delta}\right)^n\right)\left(\sum_{n=0}^{\infty}\left(\frac{1}{p_2^\delta}\right)^n\right)\cdots\left(\sum_{n=0}^{\infty}\left(\frac{1}{p_k^\delta}\right)^n\right) \\
&= \left(1+\frac{1}{p_1^\delta}+\frac{1}{p_1^{2\delta}}+\ldots\right)\left(1+\frac{1}{p_2^\delta}+\frac{1}{p_2^{2\delta}}+\ldots\right)\cdots\left(1+\frac{1}{p_k^\delta}+\frac{1}{p_k^{2\delta}}+\ldots\right) \\
&= 1+\sum_{\substack{1\leq i_1\leq k \\ j_1\in\mathbb{N}}}\frac{1}{p_{i_1}^{j_1\delta}}+\sum_{\substack{1\leq i_1<i_2\leq k \\ j_1,j_2\in\mathbb{N}}}\frac{1}{p_{i_1}^{j_1\delta}p_{i_2}^{j_2\delta}}+\cdots+\sum_{\substack{1\leq i_1<i_2<\cdots<i_k\leq k \\ j_1,j_2,\ldots,j_k\in\mathbb{N}}}\frac{1}{p_{i_1}^{j_1\delta}p_{i_2}^{j_2\delta}\cdots p_{i_k}^{j_k\delta}} \\
&\geq \sum_{\substack{n\leq x \\ n \text{ B-smooth}}}\frac{1}{n^\delta}
\end{aligned}$$

were we used the fundamental theorem of arithmetic and in the inequality discarded all $\frac{1}{y^\delta}$ with either $y > x$ or y not B-smooth. The above yield the bound

$$\psi(x,B) \leq x^\delta \prod_{p<B}\left(1-\frac{1}{p^\delta}\right)^{-1}$$

Notice

$$\left(1+\frac{1}{p^\delta}\right)\left(1-\frac{1}{p^{2\delta}}\right)^{-1} = \frac{\left(\frac{p^\delta+1}{p^\delta}\right)}{\left(\frac{p^{2\delta}-1}{p^{2\delta}}\right)} = \frac{\left(p^\delta+1\right)p^{2\delta}}{p^\delta\left(p^{2\delta}-1\right)} = \frac{\left(p^\delta+1\right)p^\delta}{\left(p^\delta-1\right)\left(p^\delta+1\right)} = \left(\frac{p^\delta-1}{p^\delta}\right)^{-1} = \left(1-\frac{1}{p^\delta}\right)^{-1}$$

Since $\prod_p\left(1-\frac{1}{p^{2\delta}}\right)^{-1} = \zeta(2\delta)$ and we know that the Riemann zeta function converges for $2\delta > 1$ (δ is real) we get for $\delta > 1/2$

$$\prod_{p<B}\left(1-\frac{1}{p^\delta}\right)^{-1} = \prod_{p<B}\left(1-\frac{1}{p^{2\delta}}\right)^{-1}\prod_{p<B}\left(1+\frac{1}{p^\delta}\right) = O\left(\prod_{p<B}\left(1+\frac{1}{p^\delta}\right)\right)$$

Combining the above we thus have

$$\psi(x,B) = O\left(x^{\delta}\prod_{p<B}\left(1+\frac{1}{p^{\delta}}\right)\right)$$

Using the inequality $1+x \leq e^x$ we obtain

$$x^{\delta}\prod_{p<B}\left(1+\frac{1}{p^{\delta}}\right) \leq x^{\delta}\prod_{p<B}e^{\frac{1}{p^{\delta}}} = e^{\log x^{\delta}}e^{\frac{1}{p_1^{\delta}}}e^{\frac{1}{p_2^{\delta}}}\cdots e^{\frac{1}{p_k^{\delta}}} = e^{\delta\log x+\sum_{p<B}\frac{1}{p^{\delta}}} \tag{1}$$

Choose $\delta = 1-\frac{1}{\log B}$ with B so big that $\delta > 1/2$ and write $p^{-\delta} = p^{-1}e^{\frac{1}{\log B}\log p}$. By applying the inequality $e^x \leq 1+xe^x$ we deduce

$$\begin{aligned}
\sum_{p<B}\frac{1}{p^{\delta}} &= \sum_{p<B}\frac{1}{p}e^{\frac{1}{\log B}\log p}\\
&\leq \sum_{p<B}\frac{1}{p}\left(1+\left(\frac{1}{\log B}\log p\right)e^{\frac{1}{\log B}\log p}\right)\\
&= \sum_{p<B}\frac{1}{p}\left(1+\left(\frac{1}{\log B}\log p\right)p^{\frac{1}{\log B}}\right)\\
&\leq \sum_{p<B}\frac{1}{p}\left(1+\left(\frac{1}{\log B}\log p\right)B^{\frac{1}{\log B}}\right)\\
&= \sum_{p<B}\frac{1}{p}+B^{\frac{1}{\log B}}\frac{1}{\log B}\sum_{p<B}\frac{\log p}{p}
\end{aligned}$$

By applying this to (1) we get

$$\begin{aligned}
x^{\delta}\prod_{p<B}\left(1+\frac{1}{p^{\delta}}\right) &\leq e^{\left(1-\frac{1}{\log B}\right)\log x+\sum_{p<B}\frac{1}{p}+B^{\frac{1}{\log B}}\frac{1}{\log B}\sum_{p<B}\frac{\log p}{p}}\\
&= e^{\log x}\cdot e^{-\frac{\log x}{\log B}}\cdot e^{\sum_{p<B}\frac{1}{p}}\cdot e^{B^{\frac{1}{\log B}}\frac{1}{\log B}\sum_{p<B}\frac{\log p}{p}}\\
&\leq x\cdot e^{-\frac{\log x}{\log B}}\cdot e^{\log\log B+O(1)}\cdot e^{B^{\frac{1}{\log B}}\frac{1}{\log B}(\log B+O(1))}\\
&= x\cdot e^{-\frac{\log x}{\log B}}(\log B)\cdot e^{O(1)}\cdot e^{B^{\frac{1}{\log B}}\frac{1}{\log B}(\log B+O(1))}
\end{aligned}$$

Hence

$$\psi(x,B) = O\left(x\cdot e^{-\frac{\log x}{\log B}}(\log B)\right)$$

□

The above proof is also an example of the use of Rankins trick; we use the constant δ to obtain a valid upper bound.

4 The likelihood of picking a B-smooth number in the quadratic sieve

In this section we calculate a measure of the likelihood that a value in our sequence is B-smooth. But as a matter of fact this is essentially an unsolved problem in the interesting ranges[1] so we restrict ourselves to a special case. Another thing that complicate things is that we are not seeking a likelihood in an ordinary interval say $[1,y]$ for some positive integer y, but in the sequence $\left\{x^2-n\right\}_{\lceil\sqrt{x}\rceil,\lceil\sqrt{x}\rceil+1,\ldots}$. To counter this problem heuristics have shown that a polynomial value is just a likely to be smooth as a random number of the same magnitude. Thus to simplify we make

[1] See p. 75 in [Pomerance]

the approximation that the numbers $x^2 - n$ are all smaller than $X = 2n^{1/2-\epsilon}$ for some $0 < \epsilon < 1/2$. Let us calculate the likelihood that a number in the interval $[1, X]$ is B-smooth when $B = X^{1/u}$ for some $1 \leq u \leq 2$(calculating the probability when u is not in this interval require more fancy stuff, see [Pomerance] or [Granville]). For this we need to calculate the number of $X^{1/u}$-smooth numbers in the interval $[1, X]$. We have already done this in 3, but for this we will use another approach. Since a number in $[1, X]$ cant be divisible by two primes strictly larger than $X^{1/u}$ (if $y \in [1, X]$ with $y = pq$ and $p, q > X^{1/u}$ then $y = pq > X^{2/u} \geq X$) we must exclude all numbers in $[1, X]$ that has a prime divisor strictly larger than $X^{1/u}$. The numbers left must be the $X^{1/u}$-smooth numbers. Writing in math this is

$$\psi(X, X^{1/u}) = \lfloor X \rfloor - \sum_{X^{1/u} < p \leq X} \left\lfloor \frac{X}{p} \right\rfloor$$

because there are exactly $\left\lfloor \frac{X}{p} \right\rfloor$ multiples of p in $[1, X]$. By removing the floor functions, we will in the first part get an error bounded by 1, while in the other part, we subtract a number between 0 and 1 to much from X for every p between $X^{1/u}$ and X hence this error term is clearly bounded by the number of primes in $[X^{1/u}, X]$. Using the prime number theorem stating that $\pi(X) = O\left(\frac{X}{\log X}\right)$ we obtain

$$\psi(X, X^{1/u}) = X - \sum_{X^{1/u} < p \leq X} \frac{X}{p} + O\left(\frac{X}{p}\right) = X\left(1 - \sum_{X^{1/u} < p \leq X} \frac{1}{p}\right) + O\left(\frac{X}{p}\right) \tag{2}$$

Mertens proved a theorem similar to theorem 5

Theorem 7.

$$\sum_{p \leq n} \frac{1}{p} = \log\log n + C + O\left(\frac{1}{\log n}\right)$$

for at constant C.

Using this we get

$$\begin{aligned}
\sum_{X^{1/u} < p \leq X} \frac{1}{p} &= \sum_{p \leq X} \frac{1}{p} - \sum_{p \leq X^{1/u}} \frac{1}{p} \\
&= \log\log X + C + O\left(\frac{1}{\log X}\right) - \left(\log\log X^{1/u} + C + O\left(\frac{1}{\log X^{1/u}}\right)\right) \\
&= \log\left(\frac{\log X}{\log X^{1/u}}\right) + O\left(\frac{1}{\log X}\right) \\
&= \log\left(u\frac{\log X}{\log X}\right) + O\left(\frac{1}{\log X}\right) \\
&= \log u + O\left(\frac{1}{\log X}\right)
\end{aligned}$$

inserting this in (2) we get

$$\psi(X, X^{1/u}) = X\left(1 - \log u\right) + O\left(\frac{X}{\log X}\right)$$

Hence the probability that a given number in $[1, X]$ is $X^{1/u}$-smooth is

$$\frac{\psi(X, X^{1/u})}{X} = (1 - \log u) + O\left(\frac{1}{\log X}\right)$$

and since

$$\frac{\frac{\psi(X, X^{1/u})}{X}}{(1 - \log u)} \to 1$$

as $X \to \infty$ we have

$$\frac{\psi(X, X^{1/u})}{X} \sim 1 - \log u \quad X \to \infty$$

Literature

- Pomerance, Carl. *Smooth numbers and the quadratic sieve.*

- Granville, Andrew. *Smooth numbers: computational number theory and beyond.*

- Cojocaru, Alina Carmen & Murty, M. Ram. *An introduction to sieve methods and their applications.*