quadratic sieve

Canonical name	QuadraticSieve
Date of creation	2013-03-22 12:48:14
Last modified on	2013-03-22 12:48:14
Owner	patrickwonders (217)
Last modified by	patrickwonders (217)
Numerical id	9
Author	patrickwonders (217)
Entry type	Algorithm
Classification	msc 11Y05
Classification	msc 11A51

Algorithm To factor a number n using the quadratic sieve, one seeks two numbers x and y which are not congruent modulo n with x not congruent to $-y$ modulo n but have $x^2 \equiv y^2 \pmod{n}$. If two such numbers are found, one can then say that $(x+y)(x-y) \equiv 0 \pmod{n}$. Then, $x+y$ and $x-y$ must have non-trivial factors in common with n.

The quadratic sieve method of factoring depends upon being able to create a set of numbers whose factorization can be expressed as a product of pre-chosen primes. These factorizations are recorded as vectors of the exponents. Once enough vectors are collected to form a set which contains a linear dependence, this linear dependence is exploited to find two squares which are equivalent modulo n.

To accomplish this, the quadratic sieve method uses a set of prime numbers called a factor base. Then, it searches for numbers which can be factored entirely within that factor base. If there are k prime numbers in the factor base, then each number which can be factored within the factor base is stored as a k-dimensional vector where the i-th component of the vector for y gives the exponent of the i-th prime from the factor base in the factorization of y. For example, if the factor base were $\{2, 3, 5, 7, 11, 13\}$, then the number $y = 2^3 \cdot 3^2 \cdot 11^5$ would be stored as the vector $\langle 3, 2, 0, 0, 5, 0\rangle$.

Once $k+1$ of these vectors have been collected, there must be a linear dependence among them. The $k+1$ vectors are taken modulo 2 to form vectors in $\mathbb{Z}_2^k$. The linear dependence among them is used to find a combination of the vectors which sum up to the zero vector in $\mathbb{Z}_2^k$. Summing these vectors is equivalent to multiplying the y's to which they correspond. And, the zero vector in $\mathbb{Z}_2^k$ signals a perfect square.

To factor n, chose a factor base $\mathbf{B} = \{p_1, p_2, \ldots, p_k\}$ such that $2 \in \mathbf{B}$ and for each odd prime p_j in $\mathbf{B}$, n is a quadratic residue of p_j. Now, start picking x_i near $\sqrt{n}$ and calculate $y_i = x_i^2 - n$. Clearly $y_i \equiv x_i^2 \pmod{n}$. If y_i can be completely factored by numbers in $\mathbf{B}$, then it is called $\mathbf{B}$-smooth. If it is not $\mathbf{B}$-smooth, then discard x_i and y_i and move on to a new choice of x_i. If it is $\mathbf{B}$-smooth, then store x_i, y_i, and the vector of its exponents for the primes in $\mathbf{B}$. Also, record a copy of the exponent vector with each component taken modulo 2.

Once $k+1$ vectors have been recorded, there must be a linear dependence among them. Using the copies of the exponent vectors that were taken modulo 2, determine which ones can be added together to form the zero vector. Multiply together the x_i that correspond to those chosen vectors—call this x. Also, add together the original vectors that correspond to the chosen vectors to form a new vector $\vec{v}$. Every component of this vector will be even. Divide each element of $\vec{v}$ by 2. Form $y = \prod_{i=1}^{k} p_i^{v_i}$.

Because each $y_i \equiv x_i^2 \pmod{n}$, $x^2 \equiv y^2 \pmod{n}$. If $x \equiv y \pmod{n}$, then find some more **B**-smooth numbers and try again. If x is not congruent to y modulo n, then $(x + y)$ and $(x - y)$ are factors of n.

Example Consider the number $n = 16843009$ The integer nearest its square root is 4104. Given the factor base

$$\mathbf{B} = \{2, 3, 5, 7, 13\}$$

, the first few **B**-smooth values of $y_i = f(x_i) = x_i^2 - n$ are:

x_i	$y_i = f(x_i)$	2	3	5	7	13
4122	147875	0	0	3	1	2
4159	454272	7	1	0	1	2
4187	687960	3	3	1	2	1
4241	1143072	5	6	0	2	0
4497	3380000	5	0	4	0	2
4993	8087040	9	5	1	0	1

Using $x_0 = 4241$ and $x_1 = 4497$, one obtains:

$$y_0 = 1143072 = 2^5 \cdot 3^6 \cdot 5^0 \cdot 7^2 \cdot 13^0$$

$$y_1 = 3380000 = 2^5 \cdot 3^0 \cdot 5^4 \cdot 7^0 \cdot 13^2$$

Which results in:

$$x = 4241 \cdot 4497 = 19071777$$

$$y = 2^5 \cdot 3^3 \cdot 5^2 \cdot 7^1 \cdot 13^1 = 1965600$$

From there:

$$\gcd(x - y, n) = 257$$

$$\gcd(x + y, n) = 65537$$

It may not be completely obvious why we required that n be a quadratic residue of each p_i in the factor base **B**. One might intuitively think that we actually want the p_i to be quadratic residues of n instead. But, that is not the case.

We are trying to express n as:

$$(x+y)(x-y) = x^2 - y^2 = n$$

where

$$y = \prod_{i=1}^{k} p_i^{v_i}$$

Because we end up squaring y, there is no reason that the p_i would need to be quadratic residues of n.

So, why do we require that n be a quadratic residue of each p_i? We can rewrite the $x^2 - y^2 = n$ as:

$$x^2 - \prod_{i=1}^{k} p_i^{2v_i} = n$$

If we take that expression modulo p_i for any p_i for which the corresponding v_i is non-zero, we are left with:

$$x^2 \equiv n \pmod{p_i}$$

Thus, in order for p_i to show up in a useful solution, n must be a quadratic residue of p_i. We would be wasting time and space to employ other primes in our factoring and linear combinations.