The Quadratic Sieve Integer Factorization Algorithm

Preliminary Examples (Trial Division)

What's the super naive way to find the prime factors of $n = 8051$?

$$8051 \div 2 = 4025.5 \text{ ✗}$$
$$8051 \div 3 = 2683.67 \text{ ✗}$$
$$8051 \div 4 = 2012.75 \text{ ✗}$$
$$8051 \div 5 = 1610.2 \text{ ✗}$$
$$\vdots$$
$$8051 \div 83 = 97 \text{ ✓}$$

$$8051 = 83 \cdot 97$$

But this is slow! In the worst case, we would have to trial divide up to $\lfloor \sqrt{n} \rfloor$ (89 trial divisions for $n = 8051$).

Preliminary Examples (Fermat's Factorization)

What is the slightly less naive way?

Use $n = a^2 - b^2 = (a-b)(a+b)$.

$$\begin{aligned} n &= 8051 \\ &= 8100 - 49 \\ &= 90^2 - 7^2 \\ &= (90-7)(90+7) \\ &= 83 \cdot 97. \end{aligned}$$

Much faster! But this only works well for $n = xy$ if x and y are close to $\sqrt{n}$.

What is the Quadratic Sieve?

- It is an integer factorization algorithm.
- Currently the second fastest factorization method, next to the number field sieve.
- But it's still the fastest for integers under 100 digits [Lan01].
- Running time to factor an integer n: $O(e^{\sqrt{ln(n)\ ln(ln(n))}})$ [Pom96] .

The General Idea of Quadratic Sieve

Given n as the integer that needs to be factored, Quadratic Sieve attempts to find x, y such that

$$\begin{aligned} x^2 &\equiv y^2 \ (mod \ \ n) \\ \Longrightarrow x^2 - y^2 &\equiv 0 \ (mod \ \ n) \\ \Longrightarrow (x-y)(x+y) &\equiv 0 \ (mod \ \ n) \end{aligned}$$

Then we can just compute $\gcd(x \pm y, \ n)$ to find the two factors!

Remark

We might get a trivial solution that we don't care about, i.e. when $\gcd(x \pm y, \ n) = 1$ or n.

The **Kraitchik function** is defined as

$$Q(x) = (x + \lfloor \sqrt{n} \rfloor)^2 - n.$$

We want to compute **Kraitchik's sequence**,

$$K = (Q(x_1), Q(x_2), \ ..., Q(x_i)),$$

with the values of $x \in \mathbb{Z}$ from a given interval $[-M, M]$, called the **sieving interval**.

Then choose a subsequence of K such that the products of the elements of that subsequence, $Q(x_{K_1}) \cdot Q(x_{K_2}) \cdot \ldots \cdot Q(x_{K_j})$, is a perfect square.

Furthermore, note that

$$\begin{aligned} & Q(x) = (x + \lfloor \sqrt{n} \rfloor)^2 - n \\ \Longrightarrow \ & Q(x) \equiv (x + \lfloor \sqrt{n} \rfloor)^2 \ (mod \ \ n). \end{aligned}$$

This means that

$$\underbrace{Q(x_{K_1})Q(x_{K_2}) \ \ldots \ Q(x_{K_j})}_{x^2} \equiv \underbrace{(x_{K_1} \ x_{K_2} \ \ldots \ x_{K_j})^2}_{y^2} \ (mod \ \ n),$$

which is precisely what we want.

Quadratic Sieve Algorithm Outline

1. Data Collection
   - Generate a factor base
   - Sieving to get smooth numbers
2. Data Processing
   - Build the matrix
   - Process the matrix
   - Factor n

Q: How do we find the product $Q(x_{K_1}) \cdot Q(x_{K_2}) \cdot \ldots \cdot Q(x_{K_j})$ and make sure it's a perfect square?

A: We must first find the prime factors of each element of K. The product $Q(x_{K_1}) \cdot Q(x_{K_2}) \cdot \ldots \cdot Q(x_{K_j})$ is a perfect square if the sum of the exponents of a given base from their prime factorization are all even.

Example

Q: Is $29 \cdot 782 \cdot 22678$ a perfect square?
First, calculate the prime factors of 29, 782, and 22678:

$$29 = 29^1$$
$$782 = 2^1 \cdot 17^1 \cdot 23^1$$
$$22678 = 2^1 \cdot 17^1 \cdot 23^1 \cdot 29^1$$

Next, add the exponents of the matching bases:

$$($$

All of the exponents are even, therefore $29 \cdot 782 \cdot 22678$ is a perfect square.

To speed up calculations, we factor over a fixed set of primes, called the **factor base**.

The criteria for choosing a factor base:

1. The factor base should always include -1 to handle cases when $Q(x)$ is negative.
2. Each prime p should be less than or equal to a bound B, called the smoothness bound. This value is dependent on the size of n.
3. The prime p must satisfy the Legendre symbol $\left(\frac{n}{p}\right) = 1$.

Definition of the Legendre Symbol

$$\left(\frac{n}{p}\right) \equiv n^{\frac{p-1}{2}} \ (mod \ p)$$

$$\left(\frac{n}{p}\right) = \begin{cases} 1, & \text{if n is a quadratic residue } \pmod{p} \\ -1, & \text{if n is a non-quadratic residue } \pmod{p} \\ 0, & \text{if } n \equiv 0 \pmod{p} \end{cases}$$

Generally, the size of the factor base should increase with the size of n, meaning that the size of B also increases. For small n, it often suffices to use trial and error when choosing B because the run-time of quadratic sieve will be insignificant.

However, in order to optimize efficiency in the case for a large n, the most optimal size of the factor base (**not** the smoothness bound) is approximately

$$\left(e^{\sqrt{\ln(n)\ln(\ln(n))}}\right)^{\sqrt{2}/4}\text{[Lan01]}.$$

Similar to factor bases, the sieving interval $[-M, M]$ should also increase with the size of n. Using trial and error when choosing M suffices for small n.

For large n, the most optimal value for the size of the sieving interval is approximately the cube of the factor base size:

$$\left(e^{\sqrt{\ln(n)\ln(\ln(n))}}\right)^{3\sqrt{2}/4} \text{[Lan01]}.$$

Since $[-M, M]$ is symmetric about zero, we can say that the optimal value for M is

$$M = \frac{1}{2}\left(e^{\sqrt{\ln(n)\ln(\ln(n))}}\right)^{3\sqrt{2}/4}.$$

Data Collection: Sieving

Sieving begins by calculating Kraitchik's function $Q(x_i)$ for all integers x_i in the sieving interval $[-M, M]$.

- If a given $Q(x_i)$ *does* factor completely over the factor base, it is said to be B-smooth. We store the values of $Q(x_i)$ and $x_i + \lfloor \sqrt{n} \rfloor$ for further use in the data processing portion of the algorithm.

- If $Q(x_i)$ *does not* factor completely over the factor base, we throw this number away and move on to $Q(x_{i+1})$.

Data Processing: Building the Matrix

After all elements of K have been processed, we now have a list of $Q(x_i)$ that are B-smooth along with a list of their respective values for $x_i + \lfloor \sqrt{n} \rfloor$.

In the data processing part of the algorithm, the goal is to find a subsequence of K such that the product of the elements of that subsequence $Q(x_{K_1}) \cdot Q(x_{K_2}) \cdot \ldots \cdot Q(x_{K_j})$ is a perfect square.

Reminder

Recall that $Q(x_{K_1}) \cdot Q(x_{K_2}) \cdot \ldots \cdot Q(x_{K_j})$ is a perfect square if the sum of the exponents of matching bases in their prime factorization are all even.

Data Processing: Building the Matrix

An easy way to do this is to first calculate the prime factorization of each element $Q(x_i) \in K$.Then, create an exponent matrix from each prime factorization.

Example

Let $K = \{19343, 114376, 225998\}$. The prime factorization of each element is:

$$\begin{aligned} 19343 &= 2^0 \cdot 17^0 \cdot 23^1 \cdot 29^2 \\ 114376 &= 2^3 \cdot 17^1 \cdot 23^0 \cdot 29^2 \\ 225998 &= 2^1 \cdot 17^3 \cdot 23^1 \cdot 29^0 \end{aligned}$$

The resulting exponent matrix is $\begin{bmatrix} 0 & 0 & 1 & 2 \\ 3 & 1 & 0 & 2 \\ 1 & 3 & 1 & 0 \end{bmatrix}$.

We can simplify calculations by working in (mod 2) since all we care about is finding even sums. For example, the previously calculated matrix becomes:

$$\begin{bmatrix} 0 & 0 & 1 & 2 \\ 3 & 1 & 0 & 2 \\ 1 & 3 & 1 & 0 \end{bmatrix} \equiv \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix} \pmod 2.$$

Data Processing: Processing the Matrix

Now that we have created the matrix, we can finally process the matrix to attempt to find the subsequence whose product is a square.

We do this by observing the exponent matrix, and choosing rows whose sum is the zero vector in (mod 2).

Example

Given the previous matrix $\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$, we can see that $R_0 + R_1 + R_2 = \vec{0}$.

R_0 corresponds to 19343, R_1 corresponds to 114376, R_2 corresponds to 225998.

Thus, we know that $19343 \cdot 114376 \cdot 225998$ is square.

If we check just to make sure, we can see that

$$19343 \cdot 114376 \cdot 225998 = 22360508^2.$$

Data Processing: Factoring n

We finally have what we've been wanting this whole time:

$$\underbrace{Q(x_{K_1})Q(x_{K_2})\ \dots\ Q(x_{K_j})}_{x^2} \equiv \underbrace{(x_{K_1}\ x_{K_2}\ \dots\ x_{K_j})^2}_{y^2} \pmod{n}.$$

Now, we calculate $\gcd(x \pm y, n)$ to find the two factors of n.

Remark

There is a 50% chance that you find a trivial factor, i.e. n or 1. If this happens, just choose another subsequence of K whose product is a square. If another square-product subsequence does not exist, try adjusting the smoothness bound B or the sieving interval $[-M, M]$.

What is the factorization of $n = 87463$?

Factor Base:
Since n is small, we can just choose the smoothness bound $B = 37$. Below is a table of the Legendre symbol $\left(\frac{n}{p}\right)$ calculations for every prime less than or equal to 37:

2	3	5	7	11	13	17	19	23	29	31	37
1	1	-1	-1	-1	1	1	1	-1	1	-1	-1

Recall that -1 is included in the factor base, and that we want $\left(\frac{n}{p}\right) = 1$.

Thus, our factor base is $\{-1, 2, 3, 13, 17, 19, 29\}$.

Sieving:
We choose the sieving interval to be [-30, 30] because of how small n is. For each integer value x_i in [-30, 30], we calculate the Kraitchik function

$$Q(x_i) = (x_i + \lfloor \sqrt{n} \rfloor)^2 - n$$

and see if $Q(x_i)$ factors completely over the factor base. If it does, we keep track of the values $Q(x_i)$ and $x_i + \lfloor \sqrt{n} \rfloor$.

Sieving (cont.):

For the sake of brevity, $Q(x_i)$ calculations for every x_i in [-30, 30] are not shown. Only the ones that factor completely over the factor base are shown below:

$Q(x_i)$	$x_i + \lfloor\sqrt{n}\rfloor$
-17238	265
-10179	278
153	296
1938	299
6786	307
12393	316

Building the Matrix:
We calculate the prime factorization of each $Q(x_i)$ in the table from the previous slide, and build the exponent matrix (mod 2) from that.

$$\begin{aligned}
-17238 &= -1^1 \cdot 2^1 \cdot 3^1 \cdot 13^2 \cdot 17^1 \cdot 19^0 \cdot 29^0 \\
-10179 &= -1^1 \cdot 2^0 \cdot 3^3 \cdot 13^1 \cdot 17^0 \cdot 19^0 \cdot 29^1 \\
153 &= -1^0 \cdot 2^0 \cdot 3^2 \cdot 13^0 \cdot 17^1 \cdot 19^0 \cdot 29^0 \\
1938 &= -1^0 \cdot 2^1 \cdot 3^1 \cdot 13^0 \cdot 17^1 \cdot 19^1 \cdot 29^0 \\
6786 &= -1^0 \cdot 2^1 \cdot 3^2 \cdot 13^1 \cdot 17^0 \cdot 19^0 \cdot 29^1 \\
12393 &= -1^0 \cdot 2^0 \cdot 3^6 \cdot 13^0 \cdot 17^1 \cdot 19^0 \cdot 29^0
\end{aligned}$$

Building the Matrix (cont.):
From those prime factorizations we get the resulting exponent matrix (mod 2):

$Q(x_i)$	-1	2	3	13	17	19	29
-17238	1	1	1	0	1	0	0
-10179	1	0	1	1	0	0	1
153	0	0	0	0	1	0	0
1938	0	1	1	0	1	1	0
6786	0	1	0	1	0	0	1
12393	0	0	0	0	1	0	0

Processing the Matrix:
Now, we find a combination of rows in the matrix that sum to the zero vector in (mod 2). Given the exponent matrix from the previous slide

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

we can observe that $R_0 + R_1 + R_2 + R_4 = \vec{0}$.

Processing the Matrix (cont.):
Since

$$\begin{gathered} R_0 \text{ corresponds to -17238,} \\ R_1 \text{ corresponds to -10179,} \\ R_2 \text{ corresponds to 153,} \\ \text{and } R_4 \text{ corresponds to 6786,} \end{gathered}$$

we know that $-17238 \cdot -10179 \cdot 153 \cdot 6786$ is a perfect square.

Finding the Factors:
In the end, we get the congruence

$$-17238 \cdot -10179 \cdot 153 \cdot 6786 \equiv (265 \cdot 278 \cdot 296 \cdot 307)^2 \pmod{87463}.$$

Therefore,

$$x = \sqrt{-17238 \cdot -10179 \cdot 153 \cdot 6786}$$
$$y = 265 \cdot 278 \cdot 296 \cdot 307.$$

We then calculate the factors $\gcd(x - y, n) = 149$ and $\gcd(x + y, n) = 587$. Thus, we get our result:

$$n = 87463 = 149 \cdot 587.$$

References

[Lan01] Eric Landquist. *The Quadratic Sieve Factoring Algorithm*. Paper. Charlottesville VA: University of Virginia, 2001.

[Pom96] Carl Pomerance. "A Tale of Two Sieves". In: *Notices of the American Mathematical Society* 43.12 (1996), pp. 1473–85.