Factoring with Fractions: On the Continued Fraction Factorization Algorithm

Anthony Dokanchi

August 2024

1 Abstract

Factoring algorithms are crucial in both theoretical and practical aspects of computer science and mathematics. At the core of many cryptographic systems, such as RSA, lies the assumption that factoring large numbers, particularly the product of two large primes, is computationally difficult. Efficient factoring algorithms could potentially break these cryptographic systems, making them a key area of research for ensuring the security of digital communications. Beyond cryptography, factoring algorithms are also important in number theory, where they help solve equations and analyze the properties of integers. Moreover, advancements in factoring can lead to improvements in algorithms for other mathematical problems, contributing to the broader field of computational mathematics.

The continuous nature of continued fractions initially seems wholly unrelated to the discrete nature of factoring. However, when attempting to factor a number n, generating the continued fraction expansion of $\sqrt{n}$ can be quite helpful, as the numerators of the convergents of $\sqrt{n}$ have a small upper bound, making them much easier to factor, and knowing the factorizations of these numerators makes it much easier to factor n.

The main body of this paper is split into two sections. In the first section, we will discuss the convergents of continued fractions and their properties. After defining the convergent, we will prove a sequence of lemmas to establish an upper bound on the numerators of the convergents. In the second section, we will detail the Continued Fraction Factoring Algorithm with examples to show how to factor any n.

2 Convergents

We define a convergent of a continued fraction as the following:

Definition 1. For an infinite simple continued fraction

$$n = a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \cdots}}} = [a_0; a_1, a_2, a_3, \ldots],$$

we define the k-th convergent C_k of $[a_0; a_1, a_2, a_3, \ldots]$ as

$$C_k = a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{\cdots + \frac{1}{a_k}}}}} = [a_0; a_1, a_2, a_3, \ldots, a_k].$$

(The same notion of convergents also exists for finite simple continued fractions, but these won't be used for the CFRAC algorithm.)

Note that each convergent is a finite simple continued fraction and is thus rational. We can simplify the convergent to obtain a ratio of two integers:

$$C_k = \frac{p_k}{q_k}.$$

Computing the first few convergents:

$$\begin{aligned}
C_0 &= [a_0] && = a_0 && = \frac{a_0}{1} = \frac{p_0}{q_0} \\
C_1 &= [a_0; a_1] && = a_0 + \frac{1}{a_1} && = \frac{a_1 p_0 + 1}{a_1} = \frac{p_1}{q_1} \\
C_2 &= [a_0; a_1, a_2] && = a_0 + \frac{1}{a_1 + \frac{1}{a_2}} = \frac{a_2(a_1 a_0 + 1) + a_0}{a_2 a_1 + 1} && = \frac{a_2 p_1 + p_0}{a_2 q_1 + q_0} = \frac{p_2}{q_2} \\
C_3 &= [a_0; a_1, a_2, a_3] && = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3}}} = \frac{a_3(a_2(a_1 a_0 + 1) + a_0) + (a_0 a_1 + 1)}{a_3(a_1 a_2 + 1) + a_1} && = \frac{a_3 p_2 + p_1}{a_3 q_2 + q_1} = \frac{p_3}{q_3} \\
C_4 &= [a_0; a_1, a_2, a_3, a_4] && = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4}}}} = \cdots && = \frac{a_4 p_3 + p_2}{a_4 q_3 + q_2} = \frac{p_4}{q_4} \\
C_5 &= [a_0; a_1, a_2, a_3, a_4, a_5] && = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \frac{1}{a_5}}}}} = \cdots && = \frac{a_5 p_4 + p_3}{a_5 q_4 + q_3} = \frac{p_5}{q_5}
\end{aligned}$$

We describe and prove this pattern in the following lemma.

Lemma 2. *For a convergent of a continued fraction $C_k = \frac{p_k}{q_k}$,*

$$p_k = a_k p_{k-1} + p_{k-2}$$

and

$$q_k = a_k q_{k-1} + q_{k-2}$$

for all $k \geq 2$. For $0 \leq k \leq 1$, we have $p_0 = a_0, p_1 = a_1 p_0 + 1, q_0 = 1, q_1 = a_1$.

Proof. This has been shown above to be true for k up to 3.

Assume that for $k = m \geq 2$, the following is true:

$$C_m = \frac{p_m}{q_m} = \frac{a_m p_{m-1} + p_{m-2}}{a_m q_{m-1} + q_{m-2}}.$$

We know $C_{m+1} = [a_0; a_1, \ldots, a_{m-1}, a_m, a_{m+1}] = [a_0; a_1, \ldots, a_{m-1}, a_m + \frac{1}{a_{m+1}}]$. By combining the m-th and $(m+1)$-th terms, we can write C_{m+1} as the m-th convergent of this new continued fraction, which we can write in the following form.

$$C_{m+1} = \frac{(a_m + \frac{1}{a_{m+1}})p_{m-1} + p_{m-2}}{(a_m + \frac{1}{a_{m+1}})q_{m-1} + q_{m-2}} = \frac{a_m p_{m-1} + p_{m-2} + \frac{p_{m-1}}{a_{m+1}}}{a_m q_{m-1} + q_{m-2} + \frac{q_{m-1}}{a_{m+1}}} = \frac{p_m + \frac{p_{m-1}}{a_{m+1}}}{q_m + \frac{q_{m-1}}{a_{m+1}}} = \frac{a_{m+1} p_m + p_{m-1}}{a_{m+1} q_m + q_{m-1}}$$

Thus, $C_k = \frac{p_k}{q_k} = \frac{a_k p_{k-1} + p_{k-2}}{a_k q_{k-1} + q_{k-2}}$ for all $k \geq 2$.

□

Our goal for the rest of this section will be to establish an upper bound on the p_k's, as CFRAC will involve factoring these p_k's. If we don't have an upper bound on the p_k's, this becomes infeasible. We will begin by finding the distance between any two consecutive convergents.

Lemma 3. $p_k q_{k-1} - p_{k-1} q_k = (-1)^{k-1}$ *for all $k \geq 1$.*

Proof. For $k = 1$: $p_1 q_0 - p_0 q_1 = (a_1 a_1 + 1)(1) - a_1 a_0 = 1 = (-1)^{1-1}$
Assume $p_j q_{j-1} - p_{j-1} q_j = (-1)^{j-1}$ holds.

$$p_{j+1} q_j + p_j q_{j+1} = (a_{j+1} p_j + p_{j-1}) q_j - p_j (a_{j+1} q_j + q_{j-1})$$

$$= a_{j+1} p_j q_j + q_j p_{j-1} - a_{j+1} p_j q_j - p_j q_{j-1}$$

$$= -(p_j q_{j-1} - p_{j-1} q_j) = -(-1)^{j-1} = (-1)^j$$

□

Corollary 3.1 (Difference of Successive Convergents)**.** *The difference between two successive convergents C_n and C_{n+1} is $\frac{1}{q_n q_{n-1}}$.*

Proof.

$$p_k q_{k-1} - p_{k-1} q_k = (-1)^{k-1}$$

Dividing both sides by $q_k q_{k-1}$:

$$\frac{p_k}{q_k} - \frac{p_{k-1}}{q_{k-1}} = \frac{(-1)^{k-1}}{q_k q_{k-1}}$$

$$|C_k - C_{k-1}| = \frac{1}{q_k q_{k-1}}$$

□

Now we can proceed to find an upper bound on the numerators of the convergents

Lemma 4. *For an irrational number $x > 1$ with convergents $C_k = \frac{p_k}{q_k}$, $|p_j^2 - x^2 q_j^2| < 2x$.*

Proof. x must always sit between C_k and C_{k+1} (a fact that will be left unproven, obtainable from Lemma 3). From this and from Corollary 3.1, we obtain the following two equations:

$$|x - \frac{p_j}{q_j}| < \frac{1}{q_{j+1} q_j}$$

$$\frac{p_j}{q_j} < x + \frac{1}{q_{j+1} q_j}$$

With these equations, we can do the following process:

$$|p_j^2 - x^2 q_j^2| = q_j^2 |x - \frac{p_j}{q_j}||x + \frac{p_j}{q_j}| < q_j^2 (\frac{1}{q_{j+1} q_j})(x + (x + \frac{1}{q_{j+1} q_j}))$$

$$= \frac{q_j}{q_{j+1}}(2x + \frac{1}{q_{j+1} q_j})$$

Thus,

$$|p_j^2 - x^2 q_j^2| < 2x(\frac{q_j}{q_{j+1}} + \frac{1}{2x q_{j+1}^2})$$

$$|p_j^2 - x^2 q_j^2| - 2x < 2x(-1 + \frac{q_j}{q_{j+1}} + \frac{1}{2x q_{j+1}^2}) < 2x(-1 + \frac{q_j}{q_{j+1}} + \frac{1}{q_{j+1}})$$

$$|p_j^2 - x^2 q_j^2| - 2x < 2x(-1 + \frac{q_j + 1}{q_{j+1}}) \leq 2x(-1 + \frac{q_{j+1}}{q_{j+1}}) = 2x(0) = 0$$

Therefore, we have

$$|p_j^2 - x^2 q_j^2| - 2x < 0 \implies |p_j^2 - x^2 q_j^2| < 2x$$

□

Theorem 5. *For any non-square integer n with $\sqrt{n}$ having convergent $C_j = \frac{p_j}{q_j}$, $-2\sqrt{n} < p_j^2 < 2\sqrt{n} \mod n$*

Proof. We apply Lemma 4 with $x = \sqrt{n}$.

$$|p_j^2 - x^2 q_j^2| < 2\sqrt{n}$$

$$|p_j^2 - n q_j^2| < 2\sqrt{n}$$

By expanding and reducing mod n,

$$-2\sqrt{n} < p_j^2 < 2\sqrt{n} \mod n$$

□

Note that this bound is entirely dependent on n, not k, which means when we're approximation an irrational $\sqrt{n}$ using an infinite continued fraction, we can generate arbitrarily many p_k's while still being sure that all p_k's fall within $2\sqrt{n}$ of a multiple of n, making them small enough to factor mod n.

3 The CFRAC Algorithm

The goal of the CFRAC algorithm is to find distinct integers x, y such that $x^2 \equiv y^2 \mod n$. In other words, we want to find an integer x such that squaring and reducing modulo n gives a perfect square. If we find such an integer, then we have that n divides $x^2 - y^2 = (x - y)(x + y)$, implying n shares factors with at least one of $x - y$ and $x + y$. From there, we can compute $\gcd(n, x - y)$ and $\gcd(n, x + y)$ to find factors of n.

For example, suppose we are trying to factor the number 9163. If we realize that $217^2 \equiv 140^2 \mod 9163$, then we know that 9163 divides $217^2 - 140^2 = (217 - 140)(217 + 140) = (77)(357)$. We can then compute the GCD of 9163 with each of these terms. $\gcd(9163, 77) = 77$ and $\gcd(9163, 357) = 119$, both of which are factors of 9163. Knowing that 77 and 119 are factors of 9163, it becomes easy to come up with the prime factorization $9163 = 7^2 * 11 * 17$.

However, this method only works if we know integers x, y such that $x^2 \equiv y^2 \mod n$. How do we find such integers? One method is to start at $x = \lceil\sqrt{n}\rceil$, which is the lowest value such that $x^2 > n$, and testing incrementally increasing choices of x until one is found that satisfies $x^2 \equiv y^2 \mod n$. Another approach is to simply choose random values of x and y. However, these brute force attacks are inefficient and unnecessary, as there is a better method of generating such values, known as the CFRAC algorithm.

We will use the example $n = 33153079$ to illustrate the algorithm. We begin by creating a convergent of the simple continued fraction expansion of $\sqrt{33153079}$:

$\sqrt{33153079} = [5757, 1, 6, 1, 3, 12, 1, 3, 1, 1, 1, 1, 1, 2, 2, 4, 2, 3, 8, 1, 1, 1, 1, 1, 1, 5, 1, 7, 44, \ldots]$

If this expansion is finite, that means $\sqrt{n}$ is rational, so n is a perfect square and it is trivial to find factors. We proceed with the assumption that n is not a perfect square, so $\sqrt{n}$ is irrational and the continued fraction is infinite.

We can use the values of p_k as choices for possible values of x in $x^2 \equiv y^2 \mod n$, creating a table of $a_k, p_k, p_k^2 \mod n$ values.

k	a_k	$p_k \mod 33153079$	$p_k^2 \mod 33153079$
0	5757	5757	$-10030 = (-1) * 2 * 5 * 17 * 59$
1	1	5758	$1485 = 3^3 * 5 * 11$
2	6	40305	$-7846 = (-1) * 2 * 3923$
3	1	46063	$2913 = 3 * 971$
4	3	178494	$-883 = (-1) * 883$
5	12	2187991	$8481 = 3 * 11 * 257$
6	1	2366485	$-2534 = (-1) * 2 * 7 * 181$
7	3	9287446	$6165 = 3^2 * 5 * 137$
8	1	11653931	$-4743 = (-1) * 3^2 * 17 * 31$
9	1	20941377	$5378 = 2 * 2689$
10	1	32595308	$-4985 = (-1) * 5 * 997$
11	1	20383606	$5709 = 3 * 11 * 173$
12	1	19825835	$-3750 = (-1) * 2 * 3 * 5^4$
13	2	26882197	$4417 = 7 * 631$
14	2	7284071	$-2374 = (-1) * 2 * 1187$
15	4	22865402	$4521 = 3 * 11 * 137$
16	2	19861796	$-3230 = (-1) * 2 * 5 * 17 * 19$
17	3	16144632	$1293 = 3 * 431$
18	8	16406536	$-6606 = (-1) * 2 * 3^2 * 367$
19	1	32551168	$4609 = 11 * 419$
20	1	15804625	$-5287 = 17 * 311$
21	1	15202714	$5250 = 2 * 3 * 5^3 * 7$
22	1	31007339	$-4683 = (-1) * 3 * 7 * 223$
23	1	13056974	$6421 = 6421$
24	1	10911234	$-1774 = (-1) * 2 * 887$

When implementing this algorithm, one should always convert large positive values of p_k^2 to small negative values. Theorem 5 guarantees that if we do this conversion, p_k^2 will fall within $2\sqrt{n}$ of zero when reduced mod n.

We observe any prime factors of $p_k^2 \mod 33153079$ that are repeated and/or are raised to an even power. Here, those are the set $\{-1, 2, 3, 5, 7, 11, 17, 137\}$, which we will call B. Moving forward, we will only consider rows from the table such that all prime factors are in B. Here, those rows are $k = 1, 7, 12, 15, 21$. For each of these rows, the prime factorization of p_k^2 can be expressed in a vector form v_k, where the power of the i-th prime in B of p_k^2 is the i-th component of v_k reduced $\mod 2$. Here, we have

$$v_1 = (0, 0, 1, 1, 0, 1, 0, 0)$$

$$v_7 = (0, 0, 0, 1, 0, 0, 0, 1)$$

$$v_{12} = (1, 1, 1, 0, 0, 0, 0, 0)$$

$$v_{15} = (0, 0, 1, 0, 0, 1, 0, 1)$$

$$v_{21} = (0, 1, 1, 1, 1, 0, 0, 0)$$

We are attempting to find a way to add any number of these vectors together such that the sum of all the terms are reduced to $0 \mod 2$. If we are able to do so, then we can take the product of the corresponding p_k values as our x and the product of the corresponding p_k^2 values as our y. This will guarantee that $x^2 \equiv y^2 \mod n$, but it does not guarantee that $x^2 \equiv y^2 \mod n$, which is what we hope for.

Here, $v_1 + v_7 + v_{15}$=(0,0,0,0,0,0,0,0), so we set $x = p_1 p_7 p_{15} = 5758 * 9287446 * 22865402 = 203445 \mod 33153079$ and $y^2 = p_1^2 p_7^2 p_{15}^2 = 1485 * 6165 * 4521 = 14825433 \mod 33153079 \implies y = 3696035 \mod 33153079$.

Thus, $x = 203445, y = 3696035$ is a solution to $x^2 \equiv y^2 \mod 33153079$ with $x \not\equiv y \mod 33153079$, which was our goal. If the particular choice of vectors yielded x and y such that $x \equiv y \mod n$, we would have to choose another set of vectors that sums to zero and repeat. If no such sets of vectors remain, we would compute a larger convergent of $\sqrt{n}$, adding more rows to the table and thus more vectors to choose from (and potentially more primes in B) and repeat.

Now that we know the values of x and y, it is easy to factor 33153079. $203445^2 \equiv 3696035^2 \mod 33153079$ implies 33153079 divides $3696035^2 - 203445^2 = (3696035+203445)(3696035-203445) = (3899480)(3492590)$. Thus, we can use the Euclidean Algorithm to compute $\gcd(33153079, 3899480) = 7499$ and $\gcd(33153079, 3492590) = 4421$. Thus, 7499 and 4421 are both factors of 8131, and from there it is easy to see that $33153079 = 7499 * 4421$.

So, in summary, the steps of the CFRAC algorithm to factor an integer n are:

1. Compute the continued fraction expansion $\sqrt{n} = [a_0; a_1, a_2, \ldots]$.
2. Compute $p_k \mod n$ and $p_k^2 \mod n$ for an arbitrary number of k's.
3. Factor each p_k^2.
4. Find a subset of the p_k^2's whose product is a perfect square mod n (possibly by converting each p_k^2 to a vector whose components are its prime factors reduced mod 2 and finding a subset of these vectors that sum to the zero vector).
5. Take x as the product of these p_k's and y^2 as the product of these p_k^2's, and find $\gcd(n, (x+y)$ and $\gcd(n, (x-y)$.

Example 6. Factor 190643

$n = 190643$

$\sqrt{n} = [436; 1, 1, 1, 2, 8, 9, 1, 2, 3, 1, 45, 5, 4, 5, 3, 2, 6, 4, 3, 1, 1, 1, 1, 5, 1, 3, 4, 3, 4, 1, 11, 2, 19, 1, 4, 1, 4, 1, 19, 2, 11, \ldots]$

k	a_k	$p_k \mod 190643$	$p_k^2 \mod 190643$
0	436	436	$-547 = (-1) * 547$
1	1	437	$326 = 2 * 163$
2	1	873	$-443 = (-1) * 443$
3	1	1310	$313 = 313$
4	2	3493	$-103 = (-1) * 103$
5	8	29254	$89 = 89$
6	9	76136	$-562 = (-1) * 2 * 281$
7	1	105390	$277 = 227$
8	2	96273	$-202 = (-1) * 2 * 101$
9	3	12923	$661 = 661$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
30	1	4783	$-71 = (-1) * 71$
31	11	45877	$409 = 409$
32	2	96537	$-43 = (-1) * 43$
33	19	164293	$694 = 2 * 347$
34	1	70187	$-151 = (-1) * 151$
35	4	63755	$622 = 2 * 311$
36	1	133942	$-151 = (-1) * 151$

$$B = \{(-1), 2, 19, 151, 227\}$$

$$v_{34} = (1, 0, 0, 1, 0)$$

$$v_{36} = (1, 0, 0, 1, 0)$$

$$v_{34} + v_{36} = (0, 0, 0, 0, 0) \mod 2$$

$$x = p_{34}p_{36} = 70187 * 133942 \equiv 190181 \mod 190643$$

$$y^2 = p_{34}^2 p_{36}^2 \equiv (-151)(-151) = 22801 = 151^2 \mod 190643 \implies y = 151$$

$$x^2 - y^2 = (x+y)(x-y) = (190332)(190030)$$

$$\gcd(n, (x+y)) = \gcd(190643, 190332) = 311$$

$$\gcd(n, (x-y)) = \gcd(190643, 190030) = 613$$

$$190643 = 311 * 613$$