1 Prime Factorization

Prime factorization is the method by which we factor an integer into its canonical prime factorization. For relatively small integers, this is trivial to do by hand. As these integers scale, however, this task becomes prohibitively expensive to compute by our modern day algorithms. In fact, the security of the RSA cryptographic algorithm relies on the difficulty of factoring large algorithms.

Because factoring large integers is so computationally expensive, it is better to first determine that the number we want to factor is, indeed, a composite. To do this, we can use the variety of primality tests to ensure that our number is factorable before we start down this path.

Choosing the correct factorization method for our number is also another way to lessen the cost of factoring. If N can be identified to have some special mathematical form that lends itself to a specialized factorization method, then it is best to use that. The problem with special factorization methods is that they may work excellently for specific kinds of numbers, but do not work for the general case. If our N is not a number with any special form, then we have to use general factoring methods. There are two kinds of general factoring

2 Classical Factorization Methods

2.1 Trial Division

As a factorization method, trial division is repeated division of our number N by small primes. We store the pre-computed primes and their number in some table. This would be very speedy, albeit at the cost of the storage. Alternatively, we could save space by generating our primes along the way, using the form $6k \pm 1$ (including 2 and 3).

2.2 Euclid's Algorithm

Interestingly enough, we can also use Euclid's algorithm to search for the factors of a number. Euclid's algorithm will give us the prime factors of our number N between g and G, where g is our lower search limit and G is our upper search limit.

To use Euclid's algorithm to factor a number, start by multiplying together all primes between the two limits. In fact, precomputed products for primes within certain ranges were used before computers were invented. Next, we apply Euclid's algorithm on the product of the primes and our number N. Prime factors of N within the search limits g and G will be found.

Application of Euclid's is fast because it simply is repeated division, multiplication, and subtraction. The only possible tricky thing here is the initial division of the (potentially) large product of primes by N. Aside from that, it is not too difficult to perform in step-wise fashion.

Although one may assume Euclid's is no longer used, it is not obsolete. In fact, it is used when N is too large to be easily divided by a small prime with the arithmetic computers provide. In that case, it is faster to divide the huge product of primes by N than to divide N over and over by small primes. It is also used when only small divisors are sought, like in the continued fraction method. To use this method to obtain small divisors, we can store products of small primes.

2.3 Fermat

Fermat's uses Legendre's congruence, which states that every number has the solutions $x \equiv \pm y \pmod{M}$ to $x^2 \equiv y^2 \pmod{M}$. These are the trivial solutions. If M is composite, however, then other solutions exist that can be used to factor M. Many other classical factorization methods also use Legendre's congruence to find the factorization of a composite number, but the way in which they differ is in how solutions to the congruence are discovered.

Here, the idea is that we can write an odd composite number as a difference between two nonconsecutive squares. Once we have that, it naturally gives us a factorization, in that: $N = ab = x^2 - y^2 = (x-y)(x+y)$. To find the two square numbers, we start by making the observation that any x that satisfies $N = x^2 - y^2$ must be $> \sqrt{N}$.

With that observation, we can start with $m = \sqrt{N} + 1$, which is the smallest x possible (aside from the case where N is x^2). Being a relatively simple factorization, we do not need to consider the case where N is x^2. We also want to discard trivial factorizations, like $N = ab = x^2 - 1^2$. To that end, we can set the restriction that the difference of our x, y must be greater than 1.

We consider $x = m^2 - N$, checking to see if it is a square. If it is, then we have found our x and y, y being m. If x is not a square, then we try the next possible x by incrementing our m by 1, and testing to see if that is a square. Repeat the process until our difference of squares is found. After the difference of squares is found, we can find our a, b by substituting our known x and y in.

For example, take the number 1625. If we wanted to write it as a difference between two nonconsecutive squares, we can start by confirming that 1625 is not a square. $\sqrt{N}$ is not an integer, so we can proceed: m $= \sqrt{N} + 1 = 41$
$x = m^2 - N = 41^2 - 1625 = 56$ 56 is not a square, and so we repeat the process until we arrive at $x = 45, y = 20$. From there, we know the two factors are $(x+y) = 65$ and $(x-y) = 25$. Notice that our factors are not primes. However, we do know that $x + y$ is the smallest factor that is $\geq \sqrt{N}$, and $a - b = \frac{N}{a+b}$ is the largest factor $\leq \sqrt{N}$. We would need to continue factoring for a prime factorization.

Note that the method is not very efficient. Fermat's method performs best in cases where N is a product of two nearby integers, because the factorization can then be found in a relatively small number of rounds of the process described above.

We can see this in the example number 1127843. Then, we have the search for our x, y below:

m	z
1062	1

Thus, we know that $1127843 = (1062^2 - 1^2) = (1062 - 1)(1062 + 1) = 1061 \cdot 1063$. $1061, 1063$ are the prime factors of our example.

However, we can consider the shortcuts that have been developed since. One of them uses the observation that the last two digits of squares cannot be just any number. To use this fact, we can test only the numbers that have a possible square ending.

Another one of the shortcuts that has been developed is to use a multiplier. The idea behind this is that we can find a factorization for a number whose prime factors are relatively close quickly. Although the number N we are trying to factor may not be a product of two nearby integers, the number n multiplied by some other integer k may be. In that case, we could take the gcd of n and kN to obtain the prime factorization. To find a multiplier k, we could use the Lehman method, which is a formalization of the discovery of these multipliers k given N.

2.4 Euler

Euler's method uses Legendre's congruence as well, but does not start with attempting to directly search for the difference of squares. In fact, it starts with the Lagrange identity, which states that:

$$(x^2+Dy^2)(u^2+Dv^2) = \{ (xu+Dyv)^2+D(yu-xv)^2(xu-Dyv)^2+D(yu+xv)^2$$

In other words, the identity states that a product of two integers of the form $a^2 + Db^2$ is itself of that form, and also has two different representations in that form. Note that D has to remain the same throughout all representations. Because this method relies on this identity, this method is only applicable to integers that are of this form.

We can use the converse of this identity, which Euler proved, to find how to write N as a product of two numbers of the form. The converse states that if we can find two different representations of the form $a^2 + Db^2$, with the gcd of the product of the b values of each representation and N being 1, then N can be written as a product of two integers of that same form.

Euler's method starts with the congruence $a^2d^2 \equiv -Db^2d^2 \equiv b^2c^2 \pmod{N}$. From here, we can see that Legendre's congruence can be used here. Then, we know the factors of N will be $\gcd(N, ad - bc)$ and $\gcd(N, ad + bc)$.

Because this method first requires us to know that our N has two different representations of the form $a^2 + Db^2$, we can start by searching for one representation first, using Fermat's method to do this. If no representations can be found, we can stop here and use another method to factorize N, for Euler's method is not applicable here.