The following is a brief description of the factor-finding process using Pollard's rho method.

1. Greatest common divisor

The Euclidean algorithm gives a very efficient method for finding the greatest common divisor between two integers. Suppose that a and b are two positive integers, not both zero. (The greatest common divisor does not depend on the of a or b.) The key fact for this algorithm is that, if $a = bq + r$, then $\gcd(a, b) = \gcd(b, r)$. If r is small compared to b, this fact simplifies the computation of $\gcd(a, b)$.

There are two standard choices for r: between 0 and b or between $-b/2$ and $b/2$. The first choice yields a simpler implementation because we can just use the modulus operator, denoted $a \pmod b$ or $a\%b$; the second choice yields a faster implementation because $|r|$ is guaranteed to be at most $b/2$.

Algorithm. Let a, b be positive integers, both nonzero. These two algorithms will return $\gcd(a, b)$.

(i) The algorithm to compute $\gcd(a, b)$ using a remainder $0 \leqslant r < b$ is as follows:

1. If $b = 0$, then return a.
2. Otherwise, while $b \neq 0$, replace a with b and replace b with $a\%b$.

(ii) The algorithm to compute $\gcd(a, b)$ using a remainder $|r|$ with $-b/2 < r \leqslant b/2$ is as follows:

1. If $b = 0$, then return a.
2. Otherwise, while $b \neq 0$, replace a with b and replace b with the minimum of $a\%b$ and $b - a\%b$.

Example. Let $a = 76$ and $b = 44$. The first version will give

$$\gcd(76, 44) = \gcd(44, 32) = \gcd(32, 12) = \gcd(12, 8) = \gcd(8, 4) = \gcd(4, 0) = 4$$

while the second version will give

$$\gcd(76, 44) = \gcd(44, 12) = \gcd(12, 4) = \gcd(4, 0) = 4.$$

2. Primality testing

A *prime number* is an integer $p \geqslant 2$ that has no nontrivial divisors. There are a few important theorems regarding prime numbers that are useful in determining primality.

Fermat's Theorem. *If p is prime and x is not divisible by p, then $x^{p-1} \equiv 1 \pmod p$.*

Lagrange's Theorem. *If p is prime, then a polynomial of degree k has at most k roots modulo p.*

Rabin's probabilistic primality test. The following test is often called the Miller-Rabin primality test, with probabilistic errors bounded by Michael Rabin.

Consider a positive odd integer a, which may be prime or composite. If a is prime, then it must satisfy both Fermat's and Lagrange's theorems. In particular, if there is some $x \not\equiv 0$ with $x^{n-1} \not\equiv 1$ or if we can exhibit three distinct roots modulo a of the polynomial $t^2 - 1$, then we can say with certainty that a is composite.

We call x with $x \not\equiv 0$ a *witness*: it is either a witness that a is definitely composite, or it is a witness that a is probably prime. If there are enough witnesses that declare a to be 'probably prime', we can be fairly certain that a is indeed prime.

We fix a witness x, and perform the following algorithm to decide whether x will declare a to be composite or probably prime.

Algorithm. Write $a-1 = 2^r m$ with m odd and form the sequence $X_k \equiv x^{2^k m} \pmod{a}$ by repeated squaring as follows: $X_0 \equiv x^m$ and $X_{k+1} \equiv X_k^2$. There are four possibilities.

1. If $X_0 \equiv 1$, then return 'probably prime'.
2. If $X_k \equiv -1$ for some $0 \leqslant k < r$, then return 'probably prime'.
3. If $X_k \equiv 1$ for some $0 < k \leqslant r$ but $X_0, \ldots, X_{k-1} \not\equiv -1$, then return 'composite'.
4. If $X_r \not\equiv 1$, return 'composite'.

By way of explanation:

1. If $X_0 \equiv 1$, then $x^{n-1} \equiv X_0^{2^r} \equiv 1$, so Fermat's theorem is satisfied.
2. If $X_k \equiv -1$, then $X_{k+1} \equiv 1$, so $x^{n-1} \equiv 1$, so Fermat's theorem is satisfied.
3. If $X_k \equiv 1$ with $X_{k-1} \not\equiv \pm 1$, then Lagrange's theorem is violated since we would have at least three roots of $t^2 - 1$, namely 1, -1, and X_{k-1}.
4. At this point, we have violated Fermat's theorem, since $x^{n-1} \equiv X_r \not\equiv 1$.

Rabin proved that the number of false positives, ie witnesses that return 'probably prime' for a composite number, make up at most 25% of the possible witnesses. (In practice, the percentage is generally much smaller for large enough a.) Therefore, if there are k witnesses that all declare a to be 'probably prime', then the probability of correctly determining primality is at least $1 - (.25)^k$.

It was proved by Pomerance, Selfridge, and Wagstaff that for $a < 10^{12}$, it is sufficient to check the witnesses $2, 3, 5, 7, 11$ for complete accuracy. See the Miller-Rabin primality test wikipedia page for precise bounds and references.

Example. Take the composite number $a = 2701 = 37 \cdot 73$ with $2700 = 2^2 \cdot 675$.

The witness $x = 2$ gives

$$\begin{aligned} X_0 &\equiv 2^{675} \equiv 2337 \\ X_1 &\equiv 2^{1350} \equiv 147 \\ X_2 &\equiv 2^{2700} \equiv 1, \end{aligned}$$

which shows that a is composite since 1, -1, and 147 are all square roots of 1.

The witness $x = 5$ gives

$$\begin{aligned} X_0 &\equiv 5^{675} \equiv 1511 \\ X_1 &\equiv 5^{1350} \equiv 776 \\ X_2 &\equiv 5^{2700} \equiv 2554, \end{aligned}$$

which shows that a is composite since $5^{2700} \not\equiv 1 \pmod{2701}$.

The witness $x = 6$ gives

$$\begin{aligned} X_0 &\equiv 6^{675} \equiv 2436 \\ X_1 &\equiv 6^{1350} \equiv 2700 \equiv -1, \end{aligned}$$

which incorrectly returns that a is 'probably prime'.

Including the bad witnesses ± 1, there are 486 false positives out of 2700, ie 18%:

- 81 of the false positives satisfy $X_0 \equiv x^{675} \equiv 1$;
- 405 of the false positives satisfy $X_k \equiv -1$ for either $k = 0$ or $k = 1$, ie either x^{675} or x^{1350} is congruent to -1;
- 810 of the accurate witnesses violate Lagrange's theorem by identifying that 147 and 2554 are square roots of 1 (in addition to ± 1);
- 1404 of the accurate witnesses violate Fermat's theorem: $X_2 \equiv x^{2700} \not\equiv 1$.

3. Pollard's rho method

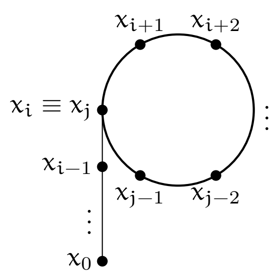
Suppose we have a positive integer a which is known to be composite with an unknown factor d. John Pollard developed a probabilistic method for finding such a factor. Let $f(x)$ be a polynomial function and define a sequence x_k with initial seed x_0 recursively by

$$x_{k+1} \equiv f(x_k) \pmod{a}.$$

Although d is unknown, we may imagine this sequence x_k modulo d. By the pigeonhole principle, there must exist indeces $i < j$ such that $x_i \equiv x_j \pmod{d}$. In particular, $x_j - x_i$ is a multiple of d, hence it must also have a factor in common with a. In other words, there exist indeces $i < j$ such that $d \leqslant \gcd(x_j - x_i, a) \leqslant a$.

Since d is unknown, the hope is that by calculating $\gcd(x_j - x_i, a)$ for various indeces, we will get a nontrivial factor of a. This method will eventually give a factor of a greater than 1, but some choices of x_0 and $f(x)$ will give the factor a itself, which is unhelpful for factoring.

It is called the 'rho' algorithm because once we find the first instance of $x_i \equiv x_j \pmod{d}$, we actually have a loop with period $j - i$, ie, $x_{i+k} \equiv x_{j+k} \pmod{d}$. The associated picture resembles the greek letter ρ.



Checking x_j against all x_i with $i < j$ can be computationally intensive, so it is instead convenient to look for an index i such that $x_i \equiv x_{2i} \pmod{d}$. This will give the result eventually since all period lengths will be accounted for, but will likely not give the first such occurrence. It is also possible to return a multiple of the period as opposed to the period itself.

Algorithm. Given a composite integer a, a polynomial $f(x)$, and an initial seed x_0, initialize with

$$r \equiv f(x_0) \pmod{a}, \quad R \equiv f(r) \pmod{a}, \quad \text{and} \quad d = \gcd(r - R, a).$$

Then, while d is equal to 1, set

$$r \equiv f(r) \pmod{a}, \quad R \equiv f\big(f(R)\big) \pmod{a}, \quad \text{and} \quad d = \gcd(r - R, a),$$

returning $1 < d \leqslant a$, which is a factor of a (possibly equal to a itself).

Example. Take $a = 13{,}118{,}851$ with $p(x) = x^2 + 1$ and $x_0 = 2$. The sequence x_k is given by

k	x_k	k	x_k
1	5	6	101502
2	26	7	4357970
3	677	8	4305221
4	458330	9	11698896
5	7346689	10	9754134

k	x_k	k	x_k
11	1335259	16	3118609
12	12270778	17	9430628
13	13044957	18	12939277
14	2881221	19	685719
15	9201956	20	4689420

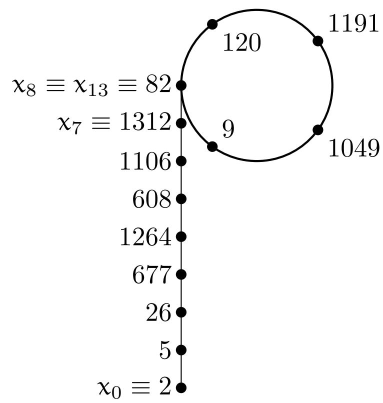
The algorithm has $\gcd(x_{2i} - x_i, a) = 1$ until $\gcd(x_{20} - x_{10}, a) = 1321$. Therefore 1321 is a nontrivial factor of a and the period modulo 1321 must divide 10. In fact, reducing modulo 1321, we find a period of 5 beginning at x_8. The table and picture modulo 1321 are given below.

k	x_k
1	5
2	26
3	677
4	1264
5	608

k	x_k
6	1106
7	1312
8	82
9	120
10	1191

k	x_k
11	1049
12	9
13	82
14	120
15	1191

k	x_k
16	1049
17	9
18	82
19	120
20	1191



Appendix

Division Algorithm. *Let a, b be integers with $b \neq 0$. There are two standard ways of choosing a quotient and remainder when dividing a by b:*

(i) There are unique integers q, r such that $a = qb + r$ and $0 \leqslant r < |b|$.

(ii) There are unique integers q, r such that $a = qb + r$ and $-\frac{|b|}{2} < r \leqslant \frac{|b|}{2}$.

Proof. Let $S = \{a - xb : x \in \mathbb{Z}, a - xb \geqslant 0\}$. Since S is a nonempty subset of the nonnegative integers, there is a minimum element $r = a - qb$. Then $0 \leqslant r < |b|$, since otherwise $r - |b|$ would be a smaller element of S.

For the second version, if $r > |b|/2$, replace r with $r - |b|$ and replace q with $q \pm 1$, depending on the sign of b. $\square$

Euclidean Algorithm. *For integers a, b with $b \neq 0$, there exist integers q_j and r_j such that*

$$\begin{aligned} a &= q_1 b + r_1 \\ b &= q_2 r_1 + r_2 \\ r_1 &= q_3 r_2 + r_3 \\ &\vdots \\ r_{n-2} &= q_n r_{n-1} + r_n \\ r_{n-1} &= q_{n+1} r_n + 0 \end{aligned}$$

Proof. Using either division algorithm repeatedly, the remainders satisfy $0 \leqslant |r_{j+1}| < |r_j|$, so this process eventually terminates with $r_{n+1} = 0$. $\square$

Definition. The *greatest common divisor* of two integers a, b (not both zero) is the largest positive integer $\gcd(a, b)$ that divides both a and b.

Proposition 1. *For a, b with $b \neq 0$, if $a = qb + r$, then $\gcd(a, b) = \gcd(b, r)$, which implies that the absolute value of the last nonzero remainder in the Euclidean algorithm is $\gcd(a, b)$.*

Proof. Assume that $a = qb + r$. If d is a common divisor of a and b, then d is also a divisor of $r = a - qb$. Conversely, if d is a common divisor of b and r, then d is also a divisor of $a = qb + r$. Therefore, the set of common divisors of a and b is equal to the set of common divisors of b and r, which implies that $\gcd(a, b) = \gcd(b, r)$.

Moreover, using the same notation from the Euclidean algorithm above, we have that

$$\gcd(a, b) = \gcd(b, r_1) = \gcd(r_1, r_2) = \cdots = \gcd(r_{n-1}, r_n) = \gcd(r_n, 0) = |r_n|,$$

which completes the proof. $\square$

Bezout's Lemma. *For integers a, b (not both zero), there exist integers x, y such that*

$$ax + by = \gcd(a, b).$$

Proof. Consider the sets $S = \{au + bv : u, v \in \mathbb{Z}\}$ and $S^+ = \{s \in S : s > 0\}$. Since S^+ is a nonempty subset of the positive integers, there is a smallest element $d = ax + by$ in S^+. By the division algorithm, there exist unique integers q, r such that $a = qd + r$ with $0 \leqslant r < d$. Then,

$$r = a - qd = a - q(ax + by) = a(1 - qx) + b(-qy),$$

which is clearly an element of S. Since $r < d$ and d is the smallest *positive* element of S, we conclude that $r = 0$, so d must be a divisor of a. Similarly, d must also be a divisor of b.

Suppose now that c is a positive common divisor of a, b. Then c divides any integer combination of a, b, in particular c must divide $d = ax + by$. This implies that $c \leqslant d$, hence $d = \gcd(a, b)$. $\square$

Proposition 2. *If p is prime, the set $\mathbb{F}_p = \{0, 1, 2, \ldots, p-1\}$ of residues modulo p form a field under addition and multiplication modulo p.*

Proof. We assume the knowledge that modular arithmetic is well-defined. The only field condition that requires verification is that all non-zero elements are invertible. A nonzero element a of $\mathbb{F}_p$ is not divisible by p, so $\gcd(a, p) = 1$. By Bezout's Lemma, there exist integers x, y such that $ax + py = 1$. Modulo p, this equation becomes $ax \equiv 1 \pmod{p}$, which proves that a is invertible. $\square$

Fermat's Theorem. *If p is prime and x is not divisible by p, the $x^{p-1} \equiv 1 \pmod{p}$.*

Proof. Suppose that x is not divisible by p. Then, modulo p, x is a nonzero element of $\mathbb{F}_p$, so it is invertible. Take the list of nonzero elements, $1, 2, \ldots, p-1$, and multiply each by x to obtain a new list modulo p,

$$x, 2x, 3x, 4x, \ldots, (p-1)x.$$

Since x is invertible modulo p, the new list is just a permutation of the original list. Therefore, the product modulo p of each list must be the same, ie,

$$\prod_{j=1}^{p-1} j \equiv \prod_{j=1}^{p-1} jx \equiv x^{p-1} \prod_{j=1}^{p-1} j.$$

Since each of $1, 2, \ldots, p-1$ is invertible modulo p, we can cancel the product from each side to obtain $1 \equiv a^{p-1} \pmod{p}$. $\square$

Lagrange's Theorem. *If p is prime, then a polynomial of degree k has at most k roots modulo p.*

Proof. Consider a polynomial equation $a_k x^k + a_{k-1} x^{k-1} + \ldots a_1 x + a_0 \equiv 0 \pmod{p}$ where a_j is in $\mathbb{F}_p$ and $a_k \not\equiv 0$. If r is a root of the polynomial, then $a_k r^k + a_{k-1} r^{k-1} + \ldots a_1 r + a_0 \equiv 0 \pmod{p}$. Therefore,

$$a_k(x^k - r^k) + a_{k-1}(x^{k-1} - r^{k-1}) + \ldots a_1(x - r) \equiv 0 \pmod{p}.$$

For each j, the polynomial $x^j - r^j$ is divisible by $x - r$ since

$$(x - r)(x^{j-1} + rx^{j-2} + \cdots + r^{j-2}x + r^{j-1}) = x^j - r^j,$$

so we can factor out $(x - r)$ from the left-hand side to get the polynomial equation

$$(x - r)(b_{n-1}x^{n-1} + \cdots + b_1 x + b_0) \equiv 0 \pmod{p}.$$

Since $\mathbb{F}_p$ is a field, there are no zero-divisors, so any other root of the original polynomial must be a root of the second factor. The remainder of the proof is a simple case of inductive reasoning. $\square$