Lenstra elliptic curve factorization

From Wikipedia, the free encyclopedia

The **Lenstra elliptic curve factorization** or the **elliptic curve factorization method** (**ECM**) is a fast, sub-exponential running time algorithm for integer factorization which employs elliptic curves. For general purpose factoring, ECM is the third-fastest known factoring method. The second fastest is the multiple polynomial quadratic sieve and the fastest is the general number field sieve. The Lenstra elliptic curve factorization is named after Hendrik Lenstra.

Practically speaking, ECM is considered a special purpose factoring algorithm as it is most suitable for finding small factors. Currently, it is still the best algorithm for divisors not greatly exceeding 20 to 25 digits (64 to 83 bits or so), as its running time is dominated by the size of the smallest factor p rather than by the size of the number n to be factored. Frequently, ECM is used to remove small factors from a very large integer with many factors; if the remaining integer is still composite, then it has only large factors and is factored using general purpose techniques. The largest factor found using ECM so far has 83 digits and was discovered on 7 September 2013 by R. Propper.[1] Increasing the number of curves tested improves the chances of finding a factor, but they are not linear with the increase in the number of digits.

Lenstra's elliptic curve factorization [edit]

The Lenstra elliptic curve factorization method to find a factor of the given natural number n works as follows:

1. Pick a random elliptic curve over $\mathbb{Z}/n\mathbb{Z}$, with equation of the form $y^2 = x^3 + ax + b \pmod n$ together with a non-trivial point $P(x_0, y_0)$ on it.

   This can be done by first picking random $x_0, y_0, a \in \mathbb{Z}/n\mathbb{Z}$, and then calculating $b = y_0^2 - x_0^3 - ax_0 \pmod n$.

2. 'Addition' of P and Q as points in general defines a group operation $P \oplus Q$ on the curve whose product can be computed from formulas given in the article on elliptic curves.

   Using this assumption, we can form repeated multiples of a point P: $kP = P \oplus \ldots \oplus P$ (k times). The addition formulas involve the taking the modular slope of a chord joining P and Q, and thus division between residue classes modulo n, performed using the extended Euclidean algorithm. In particular, division by some $v \pmod n$ includes calculation of the greatest common divisor $\gcd(v, n)$.
   If the slope is of the form u/v with $\gcd(u, n) = 1$, then $v = 0 \pmod n$ means that the result of the $\oplus$-addition will be ∞, the point 'at infinity' corresponding to the intersection of the 'vertical' line joining $P(x, y)$, $P'(x, -y)$ and the curve. However, if $\gcd(v, n)$ is neither 1 nor n, then the $\oplus$-addition will not produce a meaningful point on the curve, which shows that our elliptic curve is not a group (mod n), but, more importantly for now, $\gcd(v, n)$ is a non-trivial factor of n.

3. Compute eP on the elliptic curve (mod n), where e is product of many small numbers: say, a product of small primes raised to small powers, as in the $p - 1$ algorithm, or the factorial $B!$ for some not too large B. This can be done efficiently, one small factor at a time. Say, to get $B!P$, first compute $2P$, then $3(2P)$, then $4(3!P)$, and so on. Of course, B should be small enough so that B-wise $\oplus$-addition can be performed in reasonable time.

Wikipedia The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

Languages
العربية
Español
Français
Nederlands
Русский

Edit links

4.
   - If we were able to finish all the calculations above without encountering non-invertible elements (mod n), then we need to try again with some other curve and starting point.
   - If at some stage we found $kP = \infty$ (*infinity* on the elliptic curve), we should start over with a new curve and starting point, since this point ∞ is the group identity element, so is unchanged under any further addition operations.
   - If we encountered a gcd(x, n) at some stage that was neither 1 nor n, then we are done: it is a non-trivial factor of n.

The time complexity depends on the size of the factor and can be represented by $\exp((\sqrt{2} + o(1))\sqrt{(\ln p \ln \ln p)})$, where p is the smallest factor of n, or $L_p\left[\frac{1}{2}, \sqrt{2}\right]$, in L-notation.

Why does the algorithm work? [edit]

If p and q are two prime divisors of n, then $y^2 = x^3 + ax + b \pmod n$ implies the same equation also modulo p and modulo q. These two smaller elliptic curves with the $\boxplus$-addition are now genuine groups. If these groups have N_p and N_q elements, respectively, then for any point P on the original curve, by Lagrange's theorem, $k > 0$ is minimal such that $kP = \infty$ on the curve modulo p implies that k divides N_p; moreover, $N_pP = \infty$. The analogous statement holds for the curve modulo q. When the elliptic curve is chosen randomly, then N_p and N_q are random numbers close to $p + 1$ and $q + 1$, respectively (see below). Hence it is unlikely that most of the prime factors of N_p and N_q are the same, and it is quite likely that while computing eP, we will encounter some kP that is ∞ modulo p but not modulo q, or vice versa. When this is the case, kP does not exist on the original curve, and in the computations we found some v with either gcd(v,p) = p or gcd(v, q) = q, but not both. That is, gcd(v, n) gave a non-trivial factor of n.

ECM is at its core an improvement of the older $p - 1$ algorithm. The $p - 1$ algorithm finds prime factors p such that $p - 1$ is b-powersmooth for small values of b. For any e, a multiple of $p - 1$, and any a relatively prime to p, by Fermat's little theorem we have $a^e \equiv 1 \pmod p$. Then gcd($a^e - 1$, n) is likely to produce a factor of n. However, the algorithm fails when p - 1 has large prime factors, as is the case for numbers containing strong primes, for example.

ECM gets around this obstacle by considering the group of a random elliptic curve over the finite field $\mathbf{Z}_p$, rather than considering the multiplicative group of $\mathbf{Z}_p$ which always has order $p - 1$.

The order of the group of an elliptic curve over $\mathbf{Z}_p$ varies (quite randomly) between $p + 1 - 2\sqrt{p}$ and $p + 1 + 2\sqrt{p}$ by Hasse's theorem, and is likely to be smooth for some elliptic curves. Although there is no proof that a smooth group order will be found in the Hasse-interval, by using heuristic probabilistic methods, the Canfield–Erdős–Pomerance theorem with suitably optimized parameter choices, and the L-notation, we can expect to try $\mathbf{L}[\sqrt{2}/2, \sqrt{2}]$ curves before getting a smooth group order. This heuristic estimate is very reliable in practice.

An example [edit]

The following example is from Trappe & Washington (2006), with some details added.

We want to factor $n = 455839$. Let's choose the elliptic curve $y^2 = x^3 + 5x - 5$, with the point $P = (1, 1)$ on it, and let's try to compute $(10!)P$.

The slope of the tangent line at some point $A=(x, y)$ is $s = (3x^2 + 5)/(2y) \pmod n$. Using s we can compute $2A$. If the value of s is of the form a/b where $b > 1$ and gcd(a,b) = 1, we have to find the modular inverse of b. If it does not exist, gcd(n,b) is a non-trivial factor of n.

First we compute $2P$. We have $s(P) = s(1,1) = 4$, so the coordinates of $2P = (x', y')$ are $x' = s^2 - 2x = 14$ and $y' = s(x - x') - y = 4(1 - 14) - 1 = -53$, all numbers understood (mod n). Just to check that this $2P$ is indeed on the curve: $(-53)^2 = 2809 = 14^3 + 5 \cdot 14 - 5$.

Then we compute $3(2P)$. We have $s(2P) = s(14,-53) = -593/106 \pmod n$. Using the Euclidean algorithm: $455839 = 4300 \cdot 106 + 39$, then $106 = 2 \cdot 39 + 28$, then $39 = 28 + 11$, then $28 = 2 \cdot 11 + 6$, then $11 = 6 + 5$, then $6 = 5 + 1$. Hence gcd(455839, 106) = 1, and working backwards (a version of the extended Euclidean algorithm): $1 = 6 - 5 = 2 \cdot 6 - 11 = 2 \cdot 28 - 5 \cdot 11 = 7 \cdot 28 - 5 \cdot 39 = 7 \cdot 106 - 19 \cdot 39 = 81707 \cdot 106 - 19 \cdot 455839$. Hence $106^{-1} = 81707 \pmod{455839}$, and $-593/106 = -133317 \pmod{455839}$. Given this s, we can compute the coordinates of $2(2P)$, just as we did above: $4P = (259851, 116255)$. Just to check that this is indeed a point on the curve: $y^2 = 54514 = x^3 + 5x - 5 \pmod{455839}$. After this, we can compute $3(2P) = 4P \boxplus 2P$.

We can similarly compute $4!P$, and so on, but $8!P$ requires inverting 599 (mod 455839). The Euclidean algorithm gives that 455839 is divisible by 599, and we have found a factorization $455839 = 599 \cdot 761$.

The reason that this worked is that the curve (mod 599) has 640 = $2^7 \cdot 5$ points, while (mod 761) it has 777 = $3 \cdot 7 \cdot 37$ points. Moreover, 640 and 777 are the smallest positive integers k such that $kP = \infty$ on the curve (mod 599) and (mod 761), respectively. Since 8! is a multiple of 640 but not a multiple of 777, we have $8!P = \infty$ on the curve (mod 599), but not on the curve (mod 761), hence the repeated addition broke down here, yielding the factorization.

The algorithm with projective coordinates [edit]

Before considering the projective plane over $(\mathbb{Z}/n\mathbb{Z})$/~, first consider a 'normal' projective space over **I**: Instead of points, lines through the origin are studied. A line may be represented as a non-zero point (x, y, z), under an equivalence relation ~ given by: $(x, y, z) \sim (x', y', z') \Leftrightarrow \exists$ ***c*** $\neq 0$ such that *x'* = ***c****x*, *y'* = ***c****y* and *z'* = ***c****z*. Under this equivalence relation, the space is called **the projective plane** (P^2); points, denoted by $(x : y : z)$, correspond to lines in a three-dimensional space that pass through the origin. Note that the point $(0 : 0 : 0)$ does not exist in this space since to draw a line in any possible direction requires at least one of x',y' or z' ≠ 0. Now observe that almost all lines go through any given reference plane - such as the (*X*,*Y*,1)-plane, whilst the lines precisely parallel to this plane, having coordinates (*X*,*Y*,0), specify directions uniquely, as 'points at infinity' that are used in the affine (*X*,*Y*)-plane it lies above.

In the algorithm, only the group structure of an elliptic curve over the field **I** is used. Since we do not necessarily need the field **I**, a finite field will also provide a group structure on an elliptic curve. However, considering the same curve and operation over $(\mathbb{Z}/n\mathbb{Z})$/~ with n not a prime does not give a group. The Elliptic Curve Method makes use of the failure cases of the addition law.

We now state the algorithm in projective coordinates. The neutral element is then given by the point at infinity $(0 : 1 : 0)$. Let n be a (positive) integer and consider the elliptic curve (a set of points with some structure on it) $E(Z/nZ) = \{(x : y : z) \in P^2 \mid y^2z = x^3 + axz^2 + bz^3\}$.

1. Pick x_P, y_P, a in $\mathbb{Z}/n\mathbb{Z}$ ($a \neq 0$).
2. Calculate $b = y_P^2 - x_P^3 - ax_P$. The elliptic curve E is then in Weierstrass form given by $y^2 = x^3 + ax + b$ and by using projective coordinates the elliptic curve is given by the homogeneous equation $ZY^2 = X^3 + aZ^2X + bZ^3$. It has the point $P = (x_P : y_P : 1)$.
3. Choose an upperbound $B \in \mathbb{Z}$ for this elliptic curve. Remark: You will only find factors p if the group order of the elliptic curve E over $\mathbb{Z}/p\mathbb{Z}$ (denoted by $\#E(\mathbb{Z}/p\mathbb{Z})$) is B-smooth, which means that all prime factors of $\#E(\mathbb{Z}/p\mathbb{Z})$ have to be less or equal to B.
4. Calculate $k = \operatorname{lcm}(1, \ldots, B)$.
5. Calculate $kP := P + P + \cdots + P$ (k times) in the ring $E(\mathbb{Z}/n\mathbb{Z})$. Note that if $\#E(\mathbb{Z}/n\mathbb{Z})$ is B-smooth and n is prime (and therefore $\mathbb{Z}/n\mathbb{Z}$ is a field) that $kP = (0 : 1 : 0)$. However, if only $\#E(\mathbb{Z}/p\mathbb{Z})$ is B-smooth for some divisor p of n, the product might not be (0:1:0) because addition and multiplication are not well-defined if n is not prime. In this case, a non-trivial divisor can be found.
6. If not, then go back to step 2. If this does occur, then you will notice this when simplifying the product kP.

In point 5 it is said that under the right circumstances a non-trivial divisor can be found. As pointed out in Lenstra's article (Factoring Integers with Elliptic Curves) the addition needs the assumption $\gcd(x_1 - x_2, n) = 1$. If P, Q are not $(0 : 1 : 0)$ and distinct (otherwise addition works similarly, but is a little different), then addition works as follows:

- To calculate: $R = P + Q; P = (x_1 : y_1 : 1), Q = (x_2 : y_2 : 1)$,
- $\lambda = (y_1 - y_2)(x_1 - x_2)^{-1}$,
- $x_3 = \lambda^2 - x_1 - x_2$,
- $y_3 = \lambda(x_1 - x_3) - y_1$,
- $R = P + Q = (x_3 : y_3 : 1)$.

If addition fails, this will be due to a failure calculating λ. In particular, because $(x_1 - x_2)^{-1}$ can not always be calculated if n is not prime (and therefore $\mathbb{Z}/n\mathbb{Z}$ is not a field). Without making use of $\mathbb{Z}/n\mathbb{Z}$ being a field, one could calculate:

- $\lambda' = y_1 - y_2$,
- $x_3' = \lambda'^2 - x_1(x_1 - x_2)^2 - x_2(x_1 - x_2)^2$,

- $y_3' = \lambda'(x_1(x_1-x_2)^2 - x_3') - y_1(x_1-x_2)^3$,
- $R = P + Q = (x_3'(x_1-x_2) : y_3' : (x_1-x_2)^3)$, and simplify if possible.

This calculation is always legal and if the gcd of the Z-coordinate with n ≠ (1 or n), so when simplifying fails, a non-trivial divisor of n is found.

Twisted Edwards curves [edit]

The use of Edwards curves needs fewer modular multiplications and less time than the use of Montgomery curves or Weierstrass curves (other used methods). Using Edwards curves you can also find more primes.

Definition: Let k be a field in which $2 \neq 0$, and let $a, d \in k \setminus \{0\}$ with $a \neq d$. Then the twisted Edwards curve $E_{E,a,d}$ is given by $ax^2 + y^2 = 1 + dx^2y^2$. An Edwards curve is a twisted Edwards curve in which $a = 1$.

There are five known ways to build a set of point on an Edwards curve: the set of affine points, the set of projective points, the set of inverted points, the set of extended points and the set of completed points.

The set of affine points is given by: $\{(x,y) \in A^2 : ax^2 + y^2 = 1 + dx^2y^2\}$.

The addition law is given by $(e,f),(g,h) \mapsto \left(\frac{eh+fg}{1+degfh}, \frac{fh-aeg}{1-degfh}\right)$. The point (0,1) is its neutral element and the negative of (e,f) is $(-e,f)$. The other representations are defined similar to how the projective Weierstrass curve follows from the affine.

Any elliptic curve in Edwards form has a point of order 4. So the torsion group of an Edwards curve over $\mathbb{Q}$ is isomorphic to either $\mathbb{Z}/4\mathbb{Z}, \mathbb{Z}/8\mathbb{Z}, \mathbb{Z}/12\mathbb{Z}, \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$ or $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z}$.

The most interesting cases for ECM are $\mathbb{Z}/12\mathbb{Z}$ and $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z}$, since they force the group orders of the curve modulo primes to be divisible by 12 and 16 respectively. The following curves have a torsion group isomorphic to $\mathbb{Z}/12\mathbb{Z}$:

- $x^2 + y^2 = 1 + dx^2y^2$ with point (a,b) where $b \notin \{-2, -1/2, 0, \pm 1\}, a^2 = -(b^2+2b)$ and $d = -(2b+1)/(a^2b^2)$
- $x^2 + y^2 = 1 + dx^2y^2$ with point (a,b) where $a = \frac{u^2-1}{u^2+1}, b = -\frac{(u-1)^2}{u^2+1}$ and $d = \frac{(u^2+1)^3(u^2-4u+1)}{(u-1)^6(u+1)^2}, u \notin \{0, \pm 1\}$.

Every Edwards curve with a point of order 3 can be written in the ways shown above. Curves with torsion group isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z}$ and $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$ can be found on http://eprint.iacr.org/2008/016, top of page 30.

Stage 2 [edit]

The above text is about the first stage of elliptic curve factorisation. There one hopes to find a prime divisor p such that sP is the neutral element of $E(\mathbb{Z}/p\mathbb{Z})$. In the second stage one hopes to have found a prime divisor q such that sP has small prime order in $E(\mathbb{Z}/q\mathbb{Z})$.

We hope the order to be between B_1 and B_2, where B_1 is determined in stage 1 and B_2 is new stage 2 parameter. Checking for a small order of sP, can be done by computing $(ls)P$ modulo n for each prime l.

Success probability using EECM-MPFQ [edit]

For speedup techniques using Edward curves and implementation results, see: http://eprint.iacr.org/2008/016 pages 30–32.

Hyperelliptic curve method (HECM) [edit]

There are recent developments in using hyperelliptic curves to factor integers. Cosset shows in his article (of 2010) that one can build a hyperelliptic curve with genus two (so a curve $y^2 = f(x)$ with f of degree 5) which gives the same result as using two 'normal' elliptic curves at the same time. By making use of the Kummer Surface calculation is more efficient. The disadvantages of the hyperelliptic curve (versus an elliptic curve) are compensated by this alternative way of calculating. Therefore Cosset roughly claims that using hyperelliptic curves for factorization is no worse than using elliptic curves.

See also [edit]

- UBASIC for practical program (ECMX).

References [edit]

1. ^ 50 largest factors found by ECM

- Bernstein, Daniel J.; Birkner, Peter; Lange, Tanja; Peters, Christiane (2013). "ECM using Edwards curves". *Mathematics of Computation* **82** (282): 1139–1179. doi:10.1090/S0025-5718-2012-02633-0. MR 3008853.
- Bosma, W.; Hulst, M. P. M. van der (1990). *Primality proving with cyclotomy*. Ph.D. Thesis, Universiteit van Amsterdam. OCLC 256778332.
- Brent, Richard P. (1999). "Factorization of the tenth Fermat number". *Mathematics of Computation* **68** (225): 429–451. doi:10.1090/S0025-5718-99-00992-8. MR 1489968.
- Cohen, Henri (1993). *A Course in Computational Algebraic Number Theory*. Berlin: Springer-Verlag. doi:10.1007/978-3-662-02945-9. ISBN 0-387-55640-0. MR 1228206.
- Cosset, R. (2010). "Factorization with genus 2 curves". *Mathematics of Computation* **79** (270): 1191–1208. doi:10.1090/S0025-5718-09-02295-9. MR 2600562.
- Lenstra, A. K.; Lenstra Jr., H. W., eds. (1993). *The development of the number field sieve*. Lecture Notes in Mathematics **1554**. Berlin: Springer-Verlag. doi:10.1007/BFb0091534. MR 1321216.
- Lenstra Jr., H. W. (1987). "Factoring integers with elliptic curves". *Annals of Mathematics* **126** (3): 649–673. doi:10.2307/1971363. MR 0916721.
- Pomerance, Carl; Crandall, Richard (2005). *Prime Numbers: A Computational Perspective* (Second ed.). New York: Springer. ISBN 0-387-25282-7. MR 2156291.
- Pomerance, Carl (1985). "The quadratic sieve factoring algorithm". *Advances in Cryptology, Proc. Eurocrypt '84*. Lecture Notes in Computer Science **209**. Berlin: Springer-Verlag. pp. 169–182. doi:10.1007/3-540-39757-4_17. MR 0825590.
- Pomerance, Carl (1996). "A Tale of Two Sieves" (PDF). *Notices of the American Mathematical Society* **43** (12): 1473–1485. MR 1416721.
- Silverman, Robert D. (1987). "The Multiple Polynomial Quadratic Sieve". *Mathematics of Computation* **48** (177): 329–339. doi:10.1090/S0025-5718-1987-0866119-8. MR 0866119.
- Trappe, W.; Washington, L. C. (2006). *Introduction to Cryptography with Coding Theory* (Second ed.). Saddle River, NJ: Pearson Prentice Hall. ISBN 0-13-186239-1. MR 2372272.
- Watras, Marcin (2008). *Cryptography, Number Analysis, and Very Large Numbers*. Bydgoszcz: Wojciechowski-Steinhagen. PL:5324564.

External links [edit]

- Factorization using the Elliptic Curve Method, a Java applet which uses ECM and switches to the Self-Initializing Quadratic Sieve when it is faster.
- GMP-ECM, an efficient implementation of ECM.
- ECMNet, an easy client-server implementation that works with several factorization projects.
- pyecm, a python implementation of ECM. Much faster with psyco and/or gmpy.
- Distributed computing project yoyo@Home Subproject ECM is a program for Elliptic Curve Factorization which is used by a couple of projects to find factors for different kind of numbers.
- Lenstra Elliptic Curve Factorization algorithm source code Simple C and GMP Elliptic Curve Factorization Algorithm source code

v · t · e	Number-theoretic algorithms
Primality tests	AKS test · APR test · Baillie–PSW · ECPP test · Elliptic curve · Pocklington · Fermat · Lucas · *Lucas–Lehmer* · *Lucas–Lehmer–Riesel* · *Proth's theorem* · *Pépin's* · Quadratic Frobenius test · Solovay–Strassen · Miller–Rabin
Prime-generating	Sieve of Atkin · Sieve of Eratosthenes · Sieve of Sundaram · Wheel factorization
Integer factorization	Continued fraction (CFRAC) · Dixon's · **Lenstra elliptic curve (ECM)** · Euler's · Pollard's rho · $p-1$ · $p+1$ · Quadratic sieve (QS) · General number field sieve (GNFS) · *Special number field sieve (SNFS)* · Rational sieve · Fermat's · Shanks' square forms · Trial division · Shor's
Multiplication	Ancient Egyptian · Long · Karatsuba · Toom–Cook · Schönhage–Strassen · Fürer's
Discrete logarithm	Baby-step giant-step · Pollard rho · Pollard kangaroo · Pohlig–Hellman · Index calculus · Function field sieve

Greatest common divisor	Binary · Euclidean · Extended Euclidean · Lehmer's
Modular square root	Cipolla · Pocklington's · Tonelli–Shanks
Other algorithms	Chakravala · Cornacchia · Integer relation · Integer square root · Modular exponentiation · Schoof's
Italics indicate that algorithm is for numbers of special forms · Smallcaps indicate a deterministic algorithm	

Categories: Integer factorization algorithms | Finite fields

This page was last modified on 3 May 2015, at 05:00.

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

Privacy policy About Wikipedia Disclaimers Contact Wikipedia Developers Mobile view