VALLÉE'S TWO-THIRDS ALGORITHM

PAUL CARR

1. Introduction

(Brigitte) Vallée's Two-Thirds Algorithm is an optimization of the Dixon's Random Squares factoring algorithm.

In Dixon's Algorithm, the goal is to factor an integer n. We choose a factor base B consisting of all primes less than some bound. Then we select random integers from $\mathbb{Z}_n$ looking for those whose square mod n is B-smooth. Once we have enough of these, we solve a linear system to construct a congruence of squares that can then be used to generate a factorization of n.

The essential observation of Vallée is that smaller numbers are more likely to be B-smooth. A polynomial-time algorithm is thus developed that can convert a random element of $\mathbb{Z}_n$ into a number less than $4n^{2/3}$ that is still "sufficiently random" so as to not perturb any of the results that Dixon is based on. For large n this is a massive decrease in size, with corresponding improvement in odds of being B-smooth.

Using this, the number of random selections that must be made to generate a set of B-smooth squares is reduced, improving the complexity results. "Vanilla" Dixon has a time complexity of $L[1/2, 2\sqrt{2}]$. Vallée's Algorithm improves this to $L[1/2, \sqrt{4/3}]$. At the time of its publication in 1989 this was the best known rigorous complexity bound for integer factorization.

2. Tools

Let n be the integer to be factored. Let $h = 4n^{2/3}$. Let $k = n/h = \frac{1}{4}n^{1/3}$. Finally, let $D = \{x \in \mathbb{Z}_n : x \leq h\}$.

The first goal is to develop a covering of $\mathbb{Z}_n$ such that the distribution of D is quasi-uniform with respect to it. The **Farey covering** turns out to fit the bill. This covering consists of intervals $I(p, q)$ centered at $\frac{pn}{2q}$ of radius $\frac{h}{2q}$, for positive integers p, q with $p \leq q \leq k$ and $gcd(p, q) = 1$. We'll skip over the proofs of this (they're in [3]).

Next, we want a decent way to find elements of D in a particular element of a Farey covering. We will do this in a geometric way. Let $Q(x)$ be the squaring operation modulo n. Given a point x_0 in $\mathbb{Z}_n$, we're going to want a fairly small u such that $|Q(x_0 + u)| \leq h$. That is, $-h \leq x_0^2 + 2x_0u + u^2 \leq h$. Let $w = Q(x_0 + u) - u^2 - x_0^2$. Then what we're looking for is a point (u, w) such that $-h \leq w + u^2 + x_0^2 \leq h$ subject to $w = 2x_0u$, and $u \in \mathbb{Z}_n$. That

Date: April 29, 2009.

is, (u, w) should be on the lattice generated by $(1, 2x_0)$ and $(0, n)$ (call this lattice $L(x_0)$).

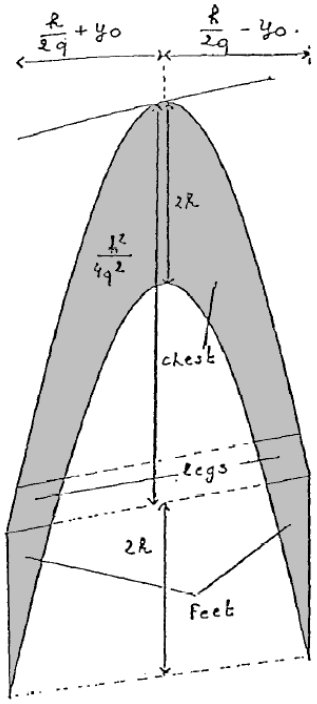
Thus, given an interval $I(p, q)$ in a Farey covering, we can describe all the elements of D inside the interval as the intersection of the lattice $L(x_0)$ and the region bounded above and below by two parabolas, and on the sides by vertical lines determined by $I(p, q)$. We will call this intersection $P(p, q)$

3. Chest, Legs, Feet

We need a better handle on this set. Assume x_0 is the closest integer to the rational point $\frac{pn}{2q}$ at the center of the Faring interval $I(p, q)$. Let $y_0 = x_0 - pn/2q$. Then the region we wish to intersect with D is

$$P(p, q) = \{(u, w) \in L(x_0) : |u + y_0| \leq \frac{h}{2q}, |w + u^2 + x_0^2| \leq h\}$$

This is illustrated in the included diagram (from Vallée: [3]):



Now, consider the lattice $L(x_0)$. It contains a more convenient vector to work with: $r = q(1, 2x_0) - p(0, n) = (q, 2qy_0)$. This is an approximately horizontal line (slope is $2y_0 \leq 1$), and has a small horizontal component (q). By using lines with this slope to divide up the total region into managable pieces (called, descriptively, the chest, legs, and feet), it becomes reasonable to quantify with accuracy the number of lattice points contained within the region, and enumerate them. The "chest" is sized to contain four quasi-horizontal lines of the lattice, the "feet" contain two, and the "legs" have the rest, but a restricted number of lattice points per line.

Using this construction, it is possible to prove solid bounds on the number of lattice points within the region, and thus in $P(p,q)$. If the distribution of D were exactly uniform, we would expect to find $2h^2/qn$ points (call this N_e). In fact, the actual number can be bounded to be between $\frac{1}{5}N_e$ and $4N_e$.

4. Example

To get a vague handle on this, a small example. Take $n = 10403 = 101 \cdot 103$. Then $h = 4n^{2/3} \approx 1906.2$, $k = n/h \approx 5.45$. The Faring cover will have values of q ranging from 1 to 5. Choose the $I(3,5)$ cover element centered at 3120.9 with radius 190.62. Then $x_0 = 3121$ and $y_0 = .1$. Also $x_0^2 = 3422 \pmod n$.

Then the bounding parabolas in the (u,w) plane are given by:

$$w = -u^2 + 1906.2 - 3433 = -u^2 - 1526.8 w = -u^2 - 1906.2 - 3433 = -u^2 - 5339.2$$

The bounds on the left and right are $u = -190.62 - .1$ and $u = 190.62 - .1$ The lowest point in the region (where the lower parabola intersects the left bounding line) is $w = -41675.07$. The region of the "legs" is bounded between roughly $w = -9151.6$ and $w = -37862.7$

Locating the lattice points involves finding the w coordinate of the intersection of each quasi-horizontal lattice line with the w axis, and then counting off the lattice points on that line between the two parabolas.

5. Algorithm

Thus fortified, we shall lay out the algorithm explicitly.

Input: A random point $x \in \mathbb{Z}_n$

Output: A random point $x_0 + u \in D$ that lies in the same Farey interval as x.

Step 1: Determine the Farey interval $I(p,q)$ which contains x. If it is contained in the overlap of two, determine which side of the median $((p_1 + p_2)/(q1 + q2))$ between the two centers x lies on and choose that interval.

Step 2: Select x_0 as the nearest integer to the center of $I(p,q)$, and construct the two parabolas and two bounding lines delimiting the region $P(p,q)$. Also determine the boundaries between the "chest", "legs", and "feet". Finally, construct the lattice $L(x_0)$.

Step 3: Determine the number of lattice points contained in the intersection of $L(x_0)$ and $P(p,q)$, enumerate them, select one number randomly, and locate that lattice point. If it is in the "chest" or "feet", simply locate all points and pick off the chosen one. If it is in the "legs", use an approximation to determine the quasi-horizontal lattice line desired, locate the lattice points on it that are in the region, and choose one at random.

Step 4: The u-coordinate of the chosen lattice point gives the output $x_0 + u$.

6. Implementation

The goal was to get this algorithm going in Sage (or even just C++ or something, since there's nothing particularly high-order in it as far as mathematical machinery goes). After playing around with it for a while, though, I haven't been able to get the lattice-point locating portion to work, from [3]. The claim is that the quasi-horizontal lattice lines will cut the vertical axis every n/q units, but I don't see quite where that's coming from, and the "lattice points" I pick up under that assumption aren't lattice points (nonintegral u). So clearly I'm doing something wrong, and I haven't been able to figure out what, heh. It's a shame, because it would be a cute thing to play with and time out.

7. Bottom Line

This algorithm produces a significant (asymptotic-wise) improvement on Dixon, or any algorithm based on random squares. It can be improved further to smaller exponents than 2/3, at the expense of turning it heuristic (quasi-uniformity is lost on the "legs"). Ultimately, though, it's been obsoleted along with its whole class of algorithms by the Quadratic and Number Field Sieves, which both have fundamentally better L function parameters. And there's enough overhead in the map from random x to $x_0 + u$ that I doubt seriously it could contend against any of the current methods at smallish integer inputs.

It's a cute trick, though.

References

[1] R. Crandall, C. Pomerance. Prime Numbers: A computational Perspective. Springer, 2001.

[2] A. K. Lenstra, H. w. Lenstra, Jr. Algorithms in Number Theory, in *Handbook of Theoretical Computer Science*, pp 673-712, MIT Press, 1990.

[3] B. Vallée. Provably fast integer factoring with quasi-uniform small quadratic residues, in *Proceedings of the Twenty-First Annual ACM Symposium on theory of Computing* (Seattle, Washington, United States, May 14-17, 1989). pp 98-106.