

Sieving Techniques

Eric Landquist

October 2, 2002

Spel Chekker: Jonathan Webster

Fiesta Bowl Pick:

Virginia Tech Hokies 24

Oklahoma Sooners 17

Overview

- Motivation
- The Idea
- Sieve Sizes
- Speed-Ups
- Research Data

I want you to remember ten things.

Motivation

- Factoring: QS, NFS
- Discrete Logs: QS, NFS, FFS
- DL's over \mathbb{F}_{p^n} , $Cl(K)$, Hyperelliptic curves of high genus.
- Find smooth values of a polynomial.
- A number is B -smooth if it factors completely over the primes less than B .
- Sieving eliminates most trial division.
- Trial division is wicked slow.

One hen

The Idea

- Evaluate $f(x)$ over interval $[-M, M]$
- Find smooth values of f .
- If $p|f(x)$, then $p|f(x + kp)\forall k$
- Mark the spots in $[-M, M]$ where $p|f(x)$.
- If $x \in [-M, M]$ has enough marks, then $f(x)$ may be smooth.
- Trial divide $f(x)$ to see if it is smooth.
- Sounds easy, right? Pbtptbbtpbth!

Two ducks

Questions

- How big should the FB be?
- How big should M be?
- What is the mark?
- How many marks is enough?
- Are there other speed-ups and considerations?

Three squawking geese

The Sieving Process

- Store the roots of $f \bmod p \forall p \in FB$.
- Store the number of bits of p , or $\ln p \forall p$.
- Initialize the array $[-M, M]$ to all 0.
- For each $p \in FB$ add $\ln p$ to $SI[x]$ if $p|f(x)$.
- If $f(x) = \prod p_i^{e_i}$, then $\ln(f(x)) = \sum e_i \ln p_i$.
- If $SI[x]$ exceeds a threshold, T , test for smoothness.
- Switch polynomials and sieve again: MPQS, SIQS.

Four Limerick oysters

The Threshold

- If T is too small, many non-smooth numbers will be treated as smooth.
- If T is too large, many smooth numbers will be overlooked.
- For $Cl(K)$, $T = \ln M \sqrt{\frac{|\Delta|}{2}}$.
- For QS, $T = \ln M \sqrt{n}$
- For NFS, $T = \ln \sqrt{n}$
- These can be fudged some, as we'll see later.

Five corpulent porpoises

The Sieving Interval

- If M is too big, $f(x)$ gets really big, and smooth numbers become very sparse.
- If M is too small, the time to switch f becomes more of a factor.
- For QS, asymptotically, $M = (\#FB)^3$.
- For $Cl(K)$, $M = \sqrt{\frac{|\Delta|}{2}} / \left(\frac{p_{max}}{2}\right)^t$
- t is the number of prime (ideal)s used to form f .
- In practice, M is smaller and determined experimentally.

Six pairs of Don Alversos tweezers ouch

The Factor Base

- If $\#FB$ is very large, smoothies are found easily, but we'll need to find a lot.
- If $\#FB$ is very small, it will take a long time to find smoothies.
- For QS and variants, $\#FB = L(1/2)$.
- For NFS, $\#FB = L(1/3)$.
- In practice, $\#FB$ is smaller and determined experimentally.
- Need to consider linear algebra: larger matrices take longer to solve.

Seven thousand Macedonians in full battle array

Speed-Ups: Large Primes

- If $f(x)$ is smooth except for a prime P , $p_{max} < P < p_{max}^2$, store this prime.
- If another value of f is found with a factor of P , we can use it.
- Set an upper bound: $P < 128p_{max}$. Smaller large primes are more likely to be found.
- For QS: Size of the matrix increases.
- For $Cl(K)$: Density of the matrix increases.
- Decrease T to allow for these.

Eight brass monkeys from the ancient sacred crypts of Egypt

Speed-Ups: Double Large Prime

- If $f(x)$ is smooth except for a composite factor L , $p_{max}^2 < L < p_{max}^3$.
- Factor $L = P_1P_2$, and find other values of f with these large factors.
- Put all large primes and 1 as vertices in a graph.
- If L is a large factor, make an edge between the vertices representing the primes (or 1).
- If we have a cycle, we can use the primes.
- Kids, don't try this at home.

Speed-Ups: Small Primes

- When sieving we expect 2-15 small primes to collectively contribute some number to each spot in the sieving array.
- Don't sieve with these, and decrease T .
- Prevents several passes through the sieve array and therefore saves time.
- Ignoring too many small primes will force some non-smooth numbers to look smooth.
- Experiment to find optimal number.

Nine apathetic, sympathetic, diabetic, old men on roller skates with a marked propensity towards procrastination and sloth

Speed-Ups: Sieving Interval

- Memory is fast. Cache is faster.
- Cut the array into blocks of say 128K or 256K.
- Sieve on the smaller blocks one at a time.
- When a block of memory is used often, it is taken into the cache, but only if it fits.
- Experiment with block sizes. Other items will also be in the cache.

Ten lyrical, spherical diabolical denizens of the deep who hall stall around the corner of the quo of the quay of the quivery, all at the same time.

Research Data: The group

- $K = \mathbb{Q}(\sqrt{\Delta}), \Delta = pq^2 \approx 10^{50}$
- $p = -(10^{30} + 1443)$
- $q = 100000000993$
- $|Cl(K)| = 1039025863202650536394906$
- To form sieving polynomial f , choose four $\mathfrak{p} \in O_K$
- $f(x, y)$ is norm form of $\prod \mathfrak{p}_i^{e_i}, e_i = \pm 1$
- $\prod \mathfrak{p}_i^{e_i} = (a, b) \in Cl(K)$
- $f(x, y) = ax^2 + bxy + cy^2, c = \frac{b^2 - \Delta}{4a}$

Research Data: The polynomials

- $a = 2704396555298590819$
- $b = -1510864351163000063$
- $c = 9244208256374066219643076505651$
- $M = 42000$, $T = 79$, with about 2500 polynomials, skipping 4 small primes.
- 2129 large primes used.
- Sieving time: 40 min 35 sec.
- Linear Algebra: 58 min. 28 sec.
- Total time to solve: 1 hr. 39 min. 3 sec.