Integer factorization,
part 1: the $\mathbf{Q}$ sieve

Integer factorization,
part 2: detecting smoothness

Integer factorization,
part 3: the number-field sieve

Integer factorization, part 4: polynomial selection

D. J. Bernstein

NFS tries to factor n by inspecting values of a polynomial.

Consider, e.g., poly degree $d = 5$. Select integer $m \in [n^{1/6}, n^{1/5}]$; find integers $f_5, f_4, \ldots, f_0$ with $n = f_5 m^5 + f_4 m^4 + \cdots + f_0$; for various integers i, j inspect $(i - jm)(f_5 i^5 + f_4 i^4 j + \cdots + f_0 j^5)$.

Practically every choice of m will succeed in factoring n. For speed want small values $(i - jm)(f_5 i^5 + f_4 i^4 j + \cdots + f_0 j^5)$.

e.g. $n = 314159265358979323$:

Can choose $m = 1000$,

$f_5 = 314$, $f_4 = 159$, $f_3 = 265$,

$f_2 = 358$, $f_1 = 979$, $f_0 = 323$.

NFS succeeds in factoring n by inspecting values $(i - 1000j)(314i^5 + \cdots + 323j^5)$ for various integer pairs (i, j).

But NFS succeeds more quickly using $m = 1370$, inspecting $(i - 1370j)(65i^5 + 130i^4j + 38i^3j^2 + 377i^2j^3 + 127ij^4 + 33j^5)$.

Consider, e.g.,
2^{45} possible choices of m.
Quickly identify, e.g.,
2^{25} attractive candidates.
Will choose one m later.

If $|i| \le SR$ and $|j| \le S^{-1}R$ then $\left|(i - jm)(f_5 i^5 + \cdots + f_0 j^5)\right| \le \mu(m, S)R^6$ where $\mu(m, S) = (mS^{-1} + S)(\left|f_5 S^5\right| + \cdots + \left|f_0 S^{-5}\right|)$.

Attractive m, S: small $\mu(m, S)$.

Choosing one typical $m \approx n^{1/6}$ produces $\mu(m, 1) \approx n^{2/6}$.

Question: How much time do we need to save factor of B—to find m, S with $\mu(m, S) \approx B^{-1}n^{2/6}$?

This has as much impact as chopping $\approx 3 \lg B$ bits out of n.

Searching for good values of m takes noticeable fraction of total time of optimized NFS. (If not, consider more m's!) End up with rather large B.

Conjectured time $B^{7.5+o(1)}$:
Enumerate many possibilities
for m near $B^{0.25}\boldsymbol{n}^{1/6}$.

Have $f_5 \approx B^{-1.25}\boldsymbol{n}^{1/6}$.
f_4, f_3, f_2, f_1, f_0 could be
as large as $B^{0.25}\boldsymbol{n}^{1/6}$.
Hope that they are smaller,
on scale of $B^{-1.25}\boldsymbol{n}^{1/6}$,
so $\mu(m, 1) \approx B^{-1}\boldsymbol{n}^{2/6}$.

Conjecturally this happens
within roughly $B^{7.5}$ trials.

Conjectured time $B^{6+o(1)}$:
Skip through m's with small f_4.

Say $n = f_5 m^5 + f_4 m^4 + \cdots + f_0$.

Choose integer $k \approx f_4/5f_5$.
Write n in base $m + k$:

$$n = f_5(m+k)^5 + (f_4 - 5kf_5)(m+k)^4 + \cdots.$$

Now degree-4 coefficient
is on same scale as f_5.

Hope for small f_3, f_2, f_1, f_0.

Conjectured time $B^{4.5+o(1)}$:
Increase S.

Enumerate many possibilities for m near $Bn^{1/6}$.

Have $f_5 \approx B^{-5}n^{1/6}$.
f_4, f_3, f_2, f_1, f_0 could be as large as $Bn^{1/6}$.

Force small f_4. Hope for
f_3 on scale of $B^{-2}n^{1/6}$,
f_2 on scale of $B^{-0.5}n^{1/6}$.
Then $\mu(m, B^{0.75}) \approx B^{-1}n^{2/6}$.

Conjectured time $B^{3.5+o(1)}$:
Partly control f_3.

Say $n = f_5 m^5 + f_4 m^4 + \cdots + f_0$.

Choose integer $k \approx f_4/5f_5$
and integer $\ell \approx m/5f_5$.

Find all short vectors
in lattice generated by
$(m/B^3, 0, 0, 10f_5k^2 - 4f_4k + f_3)$,
$(0, m/B^4, 0, 20f_5k\ell - 4f_4\ell)$,
$(0, 0, m/B^5, 10f_5\ell^2)$,
$(0, 0, 0 \qquad\quad, m)$.

Hope for v below B^1
with $(10f_5k^2 - 4f_4k + f_3)$
$\quad + (20f_5k\ell - 4f_4\ell)v$
$\quad + (10f_5\ell^2)v^2$
below m/B^3 modulo m.

Write n in base $m + k + v\ell$.
Obtain degree-5 coefficient
on scale of $B^{-5}n^{1/6}$;
degree-4 coefficient
on scale of $B^{-4}n^{1/6}$;
degree-3 coefficient
on scale of $B^{-2}n^{1/6}$.
Hope for good degree 2.

After selecting attractive m's,
how to identify best (m, y)?

Could check smoothness of
some congruences for each m
to estimate smoothness chance.

But this is expensive:
smooth congruences are rare;
need quite a few of them
before estimate is reliable.

Want something faster,
to test more (m, y)'s.

Given $H, m, f_5, \ldots, f_0$:
How many congruences survive initial selection of small congruences?

Consider integer pairs (i, j) with $i\mathbf{Z} + j\mathbf{Z} = \mathbf{Z}$ and $j > 0$.
How many values $(i - jm)(f_5 i^5 + \cdots + f_0 j^5)$ are in $[-H, H]$?

μ bound is quite crude.
Can instead enumerate j's, count i's for each j.

Faster: Numerically approximate the area of $\{(i, j) \in \mathbf{R} \times \mathbf{R} : \cdots \in [-H, H]\}$.

Number of qualifying pairs is extremely close to $(3/\pi^2) H^{2/6} \int_{-\infty}^{\infty} dx/(F(x)^2)^{1/6}$ where $F(x) = (x - m)(f_5 x^5 + \cdots + f_0)$.

Evaluate superelliptic integral by standard techniques: partition, use series expansions.

What is chance that a uniform random integer in $[1, H]$ is, e.g., 1000000-smooth?

Define S as the set of 1000000-smooth integers $n \geq 1$.

The Dirichlet series for S is $\sum [n \in S] x^{\lg n} =$

$$(1 + x^{\lg 2} + x^{2 \lg 2} + x^{3 \lg 2} + \cdots)$$
$$(1 + x^{\lg 3} + x^{2 \lg 3} + x^{3 \lg 3} + \cdots)$$
$$(1 + x^{\lg 5} + x^{2 \lg 5} + x^{3 \lg 5} + \cdots)$$
$$\cdots$$
$$(1 + x^{\lg 999983} + x^{2 \lg 999983} + \cdots).$$

Replace primes $2, 3, 5, \ldots, 999983$ with slightly larger real numbers $\overline{2} = 1.1^8$, $\overline{3} = 1.1^{12}$, $\overline{5} = 1.1^{17}$, $\ldots$, $\overline{999983} = 1.1^{145}$.

Replace each $2^a 3^b \cdots$ in S with $\overline{2}^a \overline{3}^b \cdots$, obtaining multiset $\bar{S}$.

The Dirichlet series for $\bar{S}$ is $\sum [n \in \bar{S}] x^{\lg n} =$

$$(1 + x^{\lg \overline{2}} + x^{2 \lg \overline{2}} + x^{3 \lg \overline{2}} + \cdots)$$
$$(1 + x^{\lg \overline{3}} + x^{2 \lg \overline{3}} + x^{3 \lg \overline{3}} + \cdots)$$
$$(1 + x^{\lg \overline{5}} + x^{2 \lg \overline{5}} + x^{3 \lg \overline{5}} + \cdots)$$
$$\ldots$$
$$(1 + x^{\lg \overline{999983}} + x^{2 \lg \overline{999983}} + \cdots).$$

This is simply a power series
$c_0 y^0 + c_1 y^1 + \cdots =$
$(1 + y^8 + y^{2\cdot 8} + y^{3\cdot 8} + \cdots)$
$(1 + y^{12} + y^{2\cdot 12} + y^{3\cdot 12} + \cdots)$
$(1 + y^{17} + y^{2\cdot 17} + y^{3\cdot 17} + \cdots)$
$\cdots (1 + y^{145} + y^{2\cdot 145} + \cdots)$
in the variable $y = x^{\lg 1.1}$.

Compute series mod (e.g.) y^{2910};
i.e., compute $c_0, c_1, \ldots, c_{2909}$.
$\bar{S}$ has $c_0 + \cdots + c_{2909}$ elements
$\leq 1.1^{2909} < 2^{400}$, so S has
at least that many elements $<$
2^{400}.

Can modify Dirichlet series
to modify notion of smoothness.

Use $1 + x^{\lg \overline{999983}}$ instead of
$(1 + x^{\lg \overline{999983}} + x^{2 \lg \overline{999983}} + \cdots)$
to throw away n's having
more than one factor 999983.

Multiply $c_0 y^0 + \cdots + c_{2909} y^{2909}$
by $x^{\lg \overline{1000003}} + \cdots + x^{\lg \overline{999999937}}$
to allow n's that are
1000000-smooth integers $< 2^{400}$
times one prime in $[10^6, 10^9]$.

Number-field smoothness: replace $1 + x^{\lg p} + x^{2 \lg p} + \cdots$ with $1 + x^{\lg N(P)} + x^{2 \lg N(P)} + \cdots$ where P is ideal, N is norm.

In all of these situations, can compute an upper bound on number of smooth values to check tightness of lower bound.

If looser than desired, move 1.1 closer to 1. Achieve any desired accuracy.

Smoothness chance for $i - j\alpha$ in $\mathbf{Q}(\alpha)$ is, conjecturally, very close to smoothness chance for ideals of the same size.

Same for $(i - jm, i - j\alpha)$ in $\mathbf{Q} \times \mathbf{Q}(\alpha)$.

Integrate size distribution of $(i - jm)(i - j\alpha)$ against smoothness distribution of ideals.