数理解析研究所講究録
1060巻 1998年 176-184

Recent Developments of Computational Number Theory

— A Survey on the Number Field Sieve —

東京都立大学 中村 憲 (Ken Nakamula)

Recent developments of computational number theory is very much influenced by cryptography. Especially, several methods have been introduced to factor a large integer. In this note, a brief survey on the number field sieve, shortly NFS, will be given as one of the most important methods among them. The contents are:

Notations and References.

Let n always denote a large natural number to be factored. By the run time of a method of factoring, we mean the worst case time complexity for $n \to \infty$, in many cases heuristic, conjectured or expected one. To estimate the complexity, we define

$$L_x[u,v] = \exp(v(\ln x)^u(\ln\ln x)^{1-u})$$

for real numbers x, u, v with $x > e$. When x is large, the function $L_x[u,v]$ is a monotone increasing function for $0 \leq u \leq 1$ and interpolates between powers of x and powers of

$\ln x$. More precisely, we have

$$L_x[0,v] = (\ln x)^v, \quad L_x[1,v] = x^v,$$

$$\ln\ln L_x[u,v] = u\ln\ln L_x[1,v] + (1-u)\ln\ln L_x[0,v].$$

If $0 < u < 1$, the order of $L_x[u,v]$ is said to be subexponential in the size $\ln x$ of x.

A good introduction to the NFS is given in

A. K. Lenstra, H. W. Lenstra, Jr., M. S. Manasse, J. M. Pollard, *The factorization of the ninth Fermat number*, Math. Comp. **61** (1993), 319–349.

Basic facts, improvements, related topics, references and a history of the NFS can be found in the lecture note

A. K. Lenstra, H. W. Lenstra, Jr. (eds.), *The development of the number field sieve*, Lecture Notes in Mathematics **1554**, Springer-Verlag, 1993.

About implementation of the general NFS, shortly GNFS, we reffer to

J. Buchmann, J. Loho, J. Zayer, *An implementation of the general number field sieve*, in CRYPTO '93, Lecture Notes in Computer Science **773**, Springer-Verlag, 1994, 159–165.

The multiple polynomial GNFS is introduced in

M. Elkenbracht-Huizing, *A multiple polynomial general number fields sieve*, in ANTS-II, Lecture Notes in Computer Science **1122**, Springer-Verlag, 1996, 99–114.

Factoring the 130-digit number RSA130 by the GNFS through a World Wide Web sieving project is reported in

J. Cowie, B. Dodson, R. M. Elkenbracht-Huizing, A. K. Lenstra, P. L. Montgomery, J. Zayer, *A world wide number field sieve factoring record: on to 512 bits*, in Asiacrypt '96, Lecture Notes in Computer Science **1163**, Springer-Verlag, 1996, 382–394.

An application of the NFS to the Discrete Logarithm Problem, shortly DLP, is discussed in

> D. Weber, *Computing discrete logarithms with the general number field sieve*, in ANTS-II, Lecture Notes in Computer Science **1122**, Springer-Verlag, 1996, 391–403,

by which the McCurley's challenge DLP-129 for a 129-digit prime finite field was solved on 25 January, 1998.

For further bibliographies, see these papers.

1. A Summary of Factoring Methods.

A factoring method is classified to a deterministic one or a probabilistic one. The latter expects either some "smooth" prime factor of n or some "smooth" non-trivial congruence modulo n. We summarize here the run time of the major methods.

There are not so many deterministic methods. The run time of the classical trial division is $O(L_n[1, 1/2])$. As an application of the fast Fourier transform, there is a method of run time $L_n[1,\ 1/4 + o(1)]$. The run time of the class group method is $L_n[1,\ 1/5 + o(1)]$ under Extended Riemann Hypothesis.

Expecting a "smooth" prime factor p of n, several probabilistic methods such as ρ-method, $(p-1)$-method, $(p+1)$-method, etc. are introduced. The elliptic curve method is the most significant one among them, and its run time is $L_p[1/2,\ \sqrt{2} + o(1)]$.

All probabilistic methods expecting a "smooth" congruence modulo n have the same scheme of algorithm as will be described in the next section. The run time of the rational sieve, the continued fraction method or the (multiple polynomial) quadratic sieve, shortly QS, is respectively $L_n[1/2,\ \sqrt{2}+o(1)]$, $L_n[1/2,\ 1+o(1)]$ or $L_n[1/2,\ 1+o(1)]$. The special NFS, shortly SNFS, is first introduced for $n = r^k - s$ with small natural numbers r and $|s|$; its run time is $L_n\left[1/3,\ \sqrt[3]{32/9}+o(1)\right]$. The GNFS for an arbitrary n is then introduced; its run time is $L_n\left[1/3,\ \sqrt[3]{64/9}+o(1)\right]$, which is now slightly refined to $L_n\left[1/3,\ \sqrt[3]{(92+26\sqrt{13})/27}+o(1)\right]$. So the GNFS is expected to be the fastest factoring method for a general n at present.

2. A General Scheme of Factoring.

We have the following general procedure which is common for all probabilistic methods expecting a "smooth" congruence modulo n.

Step 0. May assume that n is odd and is not a power of a prime by some other easier precomputations.

Step 1. Select the factor base $a_j \in \mathbb{Z}/n\mathbb{Z}$ $(j \in J)$ over a certain finite index set J. May assume that a_j are invertible in $\mathbb{Z}/n\mathbb{Z}$, i.e. $a_j \in (\mathbb{Z}/n\mathbb{Z})^\times$ $(j \in J)$, because otherwise we are done.

Step 2. Collect relations

$$V \subseteq \{\ (v_j)_{j\in J} \in \mathbb{Z}^{\#J} \mid \prod_{j\in J} a_j^{v_j} = 1\ \}$$

so that the number $\#V$ of relations is slightly larger than $\#J$.

Step 3. Find dependencies $W \subseteq V$ such that

$$\sum_{v\in W} v = 2(w_j)_{j\in J} \in 2\mathbb{Z}^{\#J}.$$

For each W, calculate $x \in \mathbb{Z}$ with

$$\prod_{j \in J} a_j^{w_j} = x \bmod n.$$

Then

$$(*) \qquad x^2 \equiv 1 \pmod{n}.$$

The probability of

$$1 < \gcd(x-1, n) < n$$

is at least $1 - 2^{1-r}$ with $r > 1$ by Step 0 if x as in (*) is randomly chosen for the n with r distinct prime divisors, because every $x \bmod n \in (\mathbb{Z}/n\mathbb{Z})^\times$ of order 2 except $x = \pm 1$ gives a non-trivial factor of n.

3. The Idea of the General NFS.

The GNFS is based on the fact that it is possible to construct a monic irreducible polynomial $f \in \mathbb{Z}[X]$ and a ring homomorphism

$$\varphi : \mathbb{Z}[X]/f\mathbb{Z}[X] \longrightarrow \mathbb{Z}/n\mathbb{Z}$$

so that the sum of the absolute values of the coefficients of f and the absolute value of the representative for $\varphi(X \bmod f)$ in the open interval $(-n/2, n/2)$ are both small compared to n. Let

$$\alpha = X \bmod f, \qquad m \bmod n = \varphi(\alpha) \quad \text{with} \quad m \in \mathbb{Z}.$$

Assume $\mathbb{Z}[\alpha] = \mathbb{Z}[X]/f\mathbb{Z}[X]$ is a principal ideal domain for simplicity. Then the steps in the preceding section are as follows.

Step 1. Let $J = U \cup P \cup G$ with some smoothness bound B chosen appropriately. Here U is a set of independent generators of the group $\mathbb{Z}[\alpha]^\times$ of units, P is the set of prime numbers not exceeding B and

$$G = \{\, g \in \mathbb{Z}[\alpha] \mid \text{the ideal } g\mathbb{Z}[\alpha] \text{ is a prime of } \mathbb{Z}[\alpha] \text{ with norm in } P \,\}.$$

Set the factor base $a_j = \varphi(j)$ $(j \in J)$.

Step 2. Search for pairs (a, b) of coprime integers such that $a + bm$ is P-smooth and $a + b\alpha$ is G-smooth:

$$a + bm = \prod_{p \in P} p^{e(p)} \quad \text{with} \quad e(p) \in \mathbb{Z}_{\geq 0},$$
$$a + b\alpha = \prod_{u \in U} u^{e(u)} \prod_{g \in G} g^{e(g)} \quad \text{with} \quad e(u) \in \mathbb{Z},\ e(g) \in \mathbb{Z}_{\geq 0}.$$

Since $\varphi(a + bm) = \varphi(a + b\alpha)$, we get the relations

$$\prod_{p \in P} \varphi(p)^e(p) = \prod_{u \in U} \varphi(u)^{e(u)} \prod_{g \in G} \varphi(g)^{e(g)}.$$

Collecting them we obtain J.

Step 3. This step can be accomplished by ordinary large sparse matrix elimination over $\mathbb{Z}/2\mathbb{Z}$.

4. Related Problems.

There are many problems to be solved for practical and/or theoretical use of sieving methods:

(i) On the general scheme above:

(a) Is the map $\Phi : \mathbb{Z}^{\#J} \ni (v_j)_{j\in J} \mapsto \prod_{j\in J} a_j^{v_j} \in (\mathbb{Z}/n\mathbb{Z})^\times$ onto? Namely, does the factor base a_j $(j \in J)$ generate $(\mathbb{Z}/n\mathbb{Z})^\times$?

(b) Does the relations in V contain fundamental ones among a_j $(j \in J)$? Namely, does V generates $\mathrm{Ker}(\Phi)$?

If (i–a) and (i–b) are true, all the prime power factors of n are computed in Step 3.

(ii) On the GNFS:

(a) How are f and ϕ (or m) to be constructed? The condition $f(m) \equiv 0 \pmod{n}$ is enough. The "base m" method (or SNFS) constructs f and m with the condition $f(m) = n$.

(b) How is a set S of coprime integer pairs satisfying

$$\prod_{(a,b)\in S} (a + bm) \in (\mathbb{Q}^\times)^2$$

and

$$\gamma := \prod_{(a,b)\in S} (a + b\alpha) \in (\mathbb{Q}(\alpha)^\times)^2$$

to be found so that Steps 2 and 3 are combined?

(c) How is an element $\beta \in \mathbb{Z}[\alpha]$ to be found such that $\beta^2 = \gamma$?

(d) Of course, a more precise analysis of computational complexity is required.

(iii) Are there other formulations of a general scheme similar to that explained above?

(iv) On the GNFS again:

(a) Related to the above (ii–b) and (ii–c), refine the sieving stage! This problem will be one of the most popular problems in the near future.

(b) Apply the GNFS to the DLP!

(c) To solve small obstructions may be important! For example, we should know how to select smoothness parameters.

(d) On a practical point of view, the implementation will be the most important!!!

5. A World Wide NFS Record.

The following is extracted from the e-mail to Number Theory List, shortly NTL, by A. K. Lenstra.

On April 10, 1996, we found that

RSA-130 =

18070820886874048059516561644059055662781025167694013491701270214

50056662540244048387341127590812303371781887966563182013214880557

has the following factorization

RSA-130 =

39685999459597454290161126162883786067576449112810064832555157243 *

45534498646735972188403686897274408864356301263205069600999044599.

This factorization was found using the NFS factoring algorithm, and beats the 129-digit record that was set on April, 2, 1994, by the QS factoring algorithm. ...

> We used the polynomial
>
> $$\begin{aligned} 5748,30224,87384,05200X^5 &+ 9882,26191,74822,86102X^4 \\ -13392,49938,91281,76685X^3 &+ 16875,25245,88776,84989X^2 \\ +3759,90017,48552,08738X &- 46769,93055,39319,05995 \end{aligned}$$
>
> and its root 125,74411,16841,80059,80468 modulo RSA-130.
>
> Sieving was done on a great variety of workstations at many different locations: ...
>
> ... We can say, however, that we would have spent about 500 mips years (i.e., 10% of the computing time spent on the 129-digit QS-record) if we had done all the sieving on average workstations with at least 24 megabytes of memory. ...

Such news can be available on the NTL if you issue the command

```
echo 'subscribe nmbrthry' | mail listserv@vm1.nodak.edu
```

on a UNIX workstation. In Japanese, you can get it by

```
echo add | mail tnt-request@math.metro-u.ac.jp
```

from the mailing list of Tools on Number Theory, shortly TNT. Related programs, data and papers are stored in the URL

```
ftp://ftp.math.metro-u.ac.jp/tnt/
```

including the proceedings of "Algebra and Computation", 1995, 1997 in the directories ac95, ac97.