Comm. Korean Math. Soc. 6(1991), No. 1, pp. 159–165

AN EXPERIMENT AIDED PROOF OF LENSTRA'S CONJECTURE

조 인 호, 임 종 인, 서 광 석, 고 승 철

1. 서론

정수의 소인수분해문제는 가장 오래되었으면서도 1978년 RSA - 암호법 발견 이래로 암호학에서 가장 중요한 문제가 되어왔다 [8]. 본 논문의 목적은 정수의 인수분해 알고리즘중 현재 가장 효율적인 것의 하나로 평가되고 있는 Lenstra의 타원 곡선법 (Elliptic Curve Method)과 관련하여 제기된 문제를 풀고 이것에 의해서 타원곡선법을 재분석해 보는데 있다. 이 문제는 경험적인 사고(heuristic reasoning)에 의해서 Lenstra가 1985년 타원곡선법을 발표한 논문 [3] 에서 추론한 것으로 1990년 현재까지 사실인 것으로 생각되어왔다. 우리는 이 논문에서 위의 추론이 옳다는 것을 보이고자 한다.

2 타원 곡선법

2.1. 타원곡선

타원곡선 E를

$$E : By^2 = x^3 + Ax^2 + x \tag{1}$$

로 표시되는 곡선을 나타낸다 [6]. 타원곡선 E는 특히 유한체(finite field) GF(p) (p는 솟수) 위에서 생각해보자. 식 (1) 을 만족하는 GF(p)위의 점

Received June 8, 1990.
본 연구는 1989년도 문교부 기초과학육성연구비의 지원에 의한 것임.

들의 집합 E(GF(p))는 특히 유한 가환 덧셈군(finite abelian additive group)을 이룬다. 이때 단위원 0 은 무한 원점 $0 = (\infty, \infty)$이 되고 덧셈의 규칙은 $P_1 + P_2 = P_3 = (x_3, y_3)$라 하면

$$k = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1} & \text{if } p_1 \neq p_2 \\ \dfrac{3x_1 + 2Ax_1 + 1}{2By_1} & \text{if } P_1 = P_2 \end{cases} \tag{2}$$

라 할때

$$\begin{cases} x_3 = Bk^2 - x_1 - x_2 - A \\ y_3 = k(x_1 - x_3) - y_1 \end{cases} \tag{3}$$

으로 표시된다, 식 (3)의 덧셈규칙은 GF(p)에서의 inversion을 포함하고 있기에 컴퓨터에 많은 시행시간 (running time)을 요하고 있다. 이러한 단점은 K.Koyama와 P.L. S = —E(GF(p))—라 할때 S가 $p + 1 - \sqrt{2} < S < p + 1 + \sqrt{2}$을 만족한다는 것은 잘 알려져 있다 [6]. 또한 M이 S–M인 큰 정수라 하면 E(GF(p))의 임의의 한 점 P를 택했을 때 $MP = 0$가 된다.

2.2. 타원곡선법 (E.C.M.)

타원곡선법의 알고리즘은 Pollard의 p-1 방법 [7] 과 유사한 것으로서 이 방법이 가지는 단점을 극복한 확률론적 알고리즘(probabilistic algorithm)이다. 합성수 N의 미지의 p소인수를 구하려 할 때 점 (x , y)가 mod N 하에서 식 (1)을 만족한다면 mod p 하에서도 식 (1)을 만족하개 때문에 $E(GF(p))$의 성질을 이용할 수 있다. mod N으로서 타원곡선상의 덧셈을 계산해서 $0 = (\infty, \infty)$에 도달하면 미지 소인수 p가 구해지고 소인수 분해에 성공하게 된다 [3]. 타원곡선법의 알고리즘은 다음과 같다.

1) 소인수 분해 목표가 되는 합성수 N을 택하고, 적당한 수 M을 정한다.
2) mod N으로써 정수 A를 random하게 뽑고 타원곡선 E를 정한다. 또한 E 위의 점 P를 택한다.

3) E위의 덧셈공식 (2) - (3) 에 의해서 MP mod N을 계산한다. MP mod N을 구할때 식 (2)의 inversion operation이 실패한다는 것은 식 (2)의 분모가 N과 서로소(coprime)가 아니라는 것이므로, 이 때 우리는 분모와 N의 최대공약수를 구함으로써 N의 인수 (divisor)를 얻게된다. 만약 인수를 얻는 데 실패하면 (2)로 되돌아가 다시 시행한다.

2.3. 타원곡선법의 특징

타원곡선법 알고리즘의 성공은 찾고자하는 미지 소인수를 p라 할 때 $|E(GF(p))| = S$가 작은 수 b에 대하여 b-smooth 즉 b이하의 소인수만을 갖는 타원곡선 E를 택하는데 달려 있다. 우리가 E를 잘 선택해서 S가 b-smooth라면 예를 들면 M이 SIM이면 E의 임의의 점 p에 대하여 $MP = 0$가 되어 알고리즘이 성공하여 미지의 소인수 p를 찾게된다. Lenstra 는 총 시행시간을 최소화하는 최적화된 b의 선택방법을 추론하였고 [3] b 에 의존하는 M의 선택방법은 K.Koyama의 방법을 따르는 것이 좋다 [1] . 타원곡선법의 분석및 Lenstra의 추론에 대한 계산적 증명은 3절에서 제시하겠다.

3. 새로운 공식의 유도 및 타원곡선법의 분석

Canfield등은 1983년 두 정수 $b << x$에 대해서 x이하의 임의의 난수 (random number)가 b-smooth가 될 확률 $\phi(x,b)$는 점근적으로 $\phi(x,b) \simeq u^{-u}, u = (logx)/(logb)$, 임을 보였다 [3]. 타원곡선법의 성공을 위해서는 x근방의 임의의 난수가 b-smooth가 될 확률 $\psi(x,b)$를 구하여야 한다. Lenstra는 [3] 에서 $\phi(x,b) = \psi(x,b)$라고 추론하고 이것으로부터 $x = p$ 일 때 최적화된 b는 $b = L(p)^{1/\sqrt{2}}$ $(L(p) = exp\sqrt{\log p \log\log p})$이고 이때 $\psi(p,b) = 1/b$이며 총 시행시간은 $L(o)^{1/\sqrt{2}}$라는 것을 유도하였다. 이제 $\psi(x,b)$를 구해보자.

정리 3.1.1. 두 정수 $b << x$에 대해서 x근방의 임의의 난수가 *b-smooth*가 될 확률 $\psi(x,b)$는

$$\log\psi(x,b) = -1/(loglogx + 1.0346)(1 - u^{-u})(logx - log(b+1)) + c$$

를 만족한다. 여기서 $u = (logx)/(logb)$이고 c는 x 와 b에 의존한다고 생각되는 잉여항이다.

증명: x근방의 임의의 난수 n에 대하여 n이 b-smooth가 될 확률을 구해보자. 이제 $b < y \leq x$인 y에 대하여 y가 n의 인수가 되는 사건을 b-smooth가 사건을 C라 하면 n이 b-smooth 가 되기 위해서는 모든 y에 대해서 $y \in B^c \cup C$가 성립해야 한다. 사건들을 독립적이라 가정하면

$$\begin{aligned} Prob(B^c \cup C) &= Prob(B^c) + Prob(C) - Prob(B^c)Prob(C) \\ &= (1 - Prob(B)) + Prob(C) - (1 - Prob(B))Prob(C) \\ &= 1 - Prob(B) + Prob(B)Prob(C) \end{aligned}$$

$Prob(B) \simeq 1/[y(loglogx + 1.0346)]$ 이고 ——(9)
$Prob(C) \simeq u^{-u}(u = (logx)/(logb))$ 이므로 $Prob(B^c \cup C) \simeq 1 - [(1 - u^{-u})/(\log\log x + 1.0346)](1/y)$이다. 따라서 $\psi(x,b) = \pi_{b<y\leq x} Prob(B^c \cup C)$ 이므로

$$\begin{aligned} \log\phi(x,b) &\simeq \sum_{b<y\leq x} log(1 - \frac{a}{y})(a = (1 - u^{-u})/(\log\log x + 1.0346) \\ &\simeq -a \sum_{b<y\leq x} (\frac{1}{y}) \\ &= -a \int_{b+1}^{x} \frac{1}{y}\mathrm{dy} \ = -a(logx - log(b+1)) \end{aligned}$$

따라서

$$\log\psi(x,b) = \frac{-(1 - u^{-u})(\log x - \log(b+1))}{\log\log x + 1.0346} + c$$

라 할 수 있다.

3.2 수치실험

이 실험은 IBM 386-DX에 의해서 수행되었으며 사용 언어는 Fortran이었다. x는 10^4 부터 10^{12}까지 택했으며 근방은 타원곡선군의 범위인 $x \pm 2/\sqrt{x}$로 하였다. 표 1에는 $b = L(x)^{1/\sqrt{2}}$로 하였을 때의 결과를 나타내고 있다. 실험결과 이 때는 $c = -1 + u^{-u}$로 하였을 때 $\psi(x, b)$가 실험치에 가장 접근하였다. 표 2에는 $b = exp[(logx)^{0.67}(2.81)^{-0.33}]$으로 했을 때의 결과를 나타내고 있다. 실험결과 이때는 $C = 1.17$일때 $\psi(x, b)$가 시험치에 가장 접근하였다.

3.3 실험결과 분석 및 결론

1) 실험 결과 우리의 새로운 공식은 c를 적당히 택하면 거의 정확하다는 것을 확인 할 수 있다. 여기서 c는 x와 b에 의존할 것이라고 추정하고 있으며 현재 컴퓨터 실험을 통해서 이들과의 상관관계를 구하는 작업을 진행중이다.
2) Canfield의 공식 u^{-u}은 표 2에서와 같이 b를 택했을 때 오차가 더 커졌으나 전반적으로 작은 상수배의 오차 밖에 나지 않았다.
3) 목표 합성수 N가 $N \simeq 10^{200}$이고 미지의 소인수 p가 $p \simeq 10^{25}$일 경우에도 $b = L(s)^{1//\sqrt{2}} = 49048$이므로 $u^{-u} = 1.339 \times 10^{-4}$이고 $\psi(p, b) = 3.751 \times 10^{-5}$이다. 시행시간 계산시 상수배의 차이는 무시할 수 있으므로 우리는 Lenstra의 추론이 거의 정확하다는 것을 계산적으로 검증할 수 있다.
4) $N \simeq 10^{200}$ 이고 $p \simeq 10^{25}$이면 타원곡선을 1회 시행할 때 성공할 확률이 3.751×10^{-5}이므로 결국 평균 소요횟수는 20000회 이상이 된다. 따라서 1회 시행시간 (CRAY - II, 1시간 이상)을 감안할 때 결국 E.C.M의 한계는 p가 10^{30}정도라고 할 수 있다. 이는 [2]에서 언급한 바와 같다.
5) 최근 Lenstra등은 수체선별법이라는 새로운 소인수 분해방법을 개발하였다 [4].
6) 현재 우리들은 수체선별법과 타원곡선법을 이용해서 큰 Mersenne수의 인수분해 신기록에 도전하고 있다.

표 1 수치실험

x ± 2√x	b=L(x)$^{1/\sqrt{2}}$	수치실험 결과	$\psi(x,b)$	u^{-u}
$10^4 \pm 2x10^2$	23	29개 7.250 %	6.503 %	4.221 %
$10^5 \pm 2x316$	41	54개 4.272 %	4.334 %	2.996 %
$10^6 \pm 2x10^3$	71	117개 2.925 %	2.942 %	2.212 %
$10^7 \pm 2x3162$	113	259개 2.048 %	1.978 %	1.527 %
$10^8 \pm 2x10^4$	177	520개 1.3 %	1.349 %	1.092 %
$10^9 \pm 2x31622$	271	1213개 0.959 %	0.927 %	0.792 %
$10^{10} \pm 2x10^5$	407	2610개 0.653 %	0.641 %	0.581 %
$10^{11} \pm 2x316228$	601	6056개 0.479 %	0.4476 %	0.431 %
$10^{12} \pm 2x10^6$	873	14047개 0.351 %	0.311 %	0.322 %

표 2 수치시험

x ± 2√x	b	수치실험 결과	$\psi(x,b)$	u^{-u}
$10^4 \pm 2x10^2$	23	29개 7.250 %	5.26 %	4.221 %
$10^5 \pm 2x316$	37	47개 3.718 %	3.412 %	2.48 %
$10^6 \pm 2x10^3$	61	95개 2.375 %	2.301 %	1.701 %
$10^7 \pm 2x3162$	97	183개 1.447 %	1.564 %	1.183 %
$10^8 \pm 2x10^4$	150	414개 1.035 %	1.071 %	0.834 %
$10^9 \pm 2x31622$	225	839개 0.663 %	0.736 %	0.589 %
$10^{10} \pm 2x10^5$	335	1910개 0.478 %	0.498 %	0.368 %
$10^{11} \pm 2x316228$	492	4525개 0.358 %	0.3566 %	0.318 %
$10^{12} \pm 2x10^6$	714	10205개 0.255 %	0.2489 %	0.238 %

References

1. K.Koyama, *Factoring Using a Fast Elliptic Curve Method*, 신학론D **J70 - D** (12) (1987), 2730 – 2738.
2. K.Koyama, *Speeding the Elliptic Method and examination of factoring*, 신학기보, ISEC (1988), 88 – 19.
3. H.N.Lenstra Jr, *Factoring integers with elliptic curves*, Annals of Math. **126** (1987), 649 – 673.
4. L.K.Lenstra, H.W.Lenstra, M.S.Manasse and J.M.Pollard, *The number field sieve*, preprint to appear at Proc. of STOC (1990).
5. P.L.Montgomery, *Speeding the Pollard and elliptic curve method of factorization*, Math. Comp. **48** (1987), 243 – 264.
6. L.J.Mordell, *Diophantine Equations A.P*, 1969.
7. J.M.Pollard, *Theorems on Factorization and primality testing*, Proc. Camb. Philos. Soc. **76** (1974), 521 – 528.
8. R.L.Rivest, A.shamir and L.Adleman, *A Method of obtaining digital signatures and public key cryptosystem*, Comm. of ACM (1978), 120 – 126.
9. M.R.Shroder, *Number theory in science and communication*, Springer – Verlag, 1986.

고려대학교 이과대학 수학과
고려대학교 자연과학대학 수학과
서남대학 수학과
포항공대 수학과