Kangwon Science Highschool

A novel $O(|P|k \max_{p_i \in P} p_i)$-time heuristic

Construction of the perfect-square number detection heuristic

Kim Jun Hyeok

October 16, 2020

Abstract

기존에 이분 탐색과 산술(Arithmetic) 연산을 이용하여 제시되었던 k-bit 정수에 대하여 $O(k^3)$ 혹은 $O(k^2 \log k)$ 시간 결정론적 알고리즘은 매우 큰 수에 대해서는 시간이 오래 걸리는 결점이 있다. 이 연구에서는, 빅데이터 분석을 통해서 임의의 홀수인 소수에 관한 이차 잉여(Quadratic-residue)에 대하여 Euler's criterion 을 이용한 완전 제곱수 휴리스틱(Heuristic) 제시와 유의미한 높은 확률로 판정 가능한 소수들의 최소 개수를 제시하는 것을 목표로 한다. 또한, 후속 연구 방안으로 임의의 제곱수를 비트로 표현하였을 때 임의의 개수의 제곱수에 관하여 LCS(Longest-common- sequence)를 구하여 여기에서 찾은 LCS 를 통해서 매칭하는 휴리스틱을 제시하고, 위의 방법과 결합하여 더욱 판정률을 높인다.

Contents

1 서론

기존에 제시된 결정론적 다항 시간(Deterministic polynomial time)에 임의의 k-bit 정수가 제곱수(Perfect-square number)인지 결정하는 알고리즘은, 이분 탐색(Binary search)와 산술 연산(Arithmetic operation)을 통해서 $O(k^2)$ Naïve 정수 곱셈을 이용하면, $O(k^3)$ 시간에 결정할 수 있다. 여기에서, Fürer 의 $O(k \log k)$ 시간 정수 곱셈 알고리즘을 이용하면, $O(k^2 \log k)$ 시간에 문제를 해결할 수 있다. [1] 하지만 k의 크기가 매우 커지면 수행이 느려지는 결점이 있다.
이 연구에서는 이를 개선시킬, $k = 10^8$, $N = 10^{10^8}$ 에서도 빠른 수행시간이 보장되는 k에 대하여 선형 시간에 가까운 휴리스틱을 제시한다.
또한, 임의의 홀수인 소수 p에 대하여 다음과 같은 Euler's criterion을 통해서 $a \in GF(p)$에 대하여 Legendre's symbol 을 계산할 수 있다. [2]

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{\text{p}}$$

여러 소수에 대하여 반복적으로 시도해봄으로서 높은 확률로 제곱수임을 판정할 수 있다. 본 연구에서는 임의의 k-bit 정수가 제곱수(Perfect-square number)인지 결정하는 $O(k \log k)$ 시간 휴리스틱을 제시한다.

2 본론

2.1 이론적 배경

이차 잉여(Quadratic)를 통한 확률적인 알고리즘을 설계하기 위해서 다음과 같은 정의가 필요하다.

Definition 1. 임의의 홀수인 소수 p와 $a \in GF(p)$에 대하여 a가 p에 대한 이차 잉여 *(Quadratic residue)*라는 것은 $x^2 \equiv a \ (mod\ p)$인 $x \in GF(p)$가 존재하는 것이다.

Definition 2. 임의의 홀수인 소수 p와 $a \in GF(p)$에 대하여 *Legendre's symbol* 은 다음과 같이 정의된다.

$$\left(\frac{a}{p}\right) \equiv \begin{cases} 1 & a \text{ is a quadratic residue of } p \\ -1 & Otherwise \end{cases} \tag{1}$$

Theorem 1. 임의의 홀수인 소수 p와 $a \in GF(p)$에 대하여 다음과 같은 식이 성립한다.

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \ (mod\ p) \tag{2}$$

Proof. Hardy, G. H., et al., *An introduction to the Theory of numbers (Sixth ed.)* [2]를 참고하라. □

Theorem 2. $GF(p)$에서 홀수인 소수 p에 대한 이차 잉여의 개수는 $\frac{p-1}{2}$이다.

Proof. 만일 $n, m \in GF(p)$에 대하여 $n^2 \equiv m^2$을 만족하려면, $n \equiv -m$ 또는 $n \equiv m$을 만족하여야 한다. 따라서 n^2 mod p의 서로 다른 결과는 최대 $\frac{p-1}{2}$개가 될 수 있다. 따라서 성립한다. □

2.2 알고리즘 설계

임의의 홀수인 소수 p와 $a \in GF(p)$에 대하여, $\left(\frac{a}{p}\right)$를 계산하는 것은 theorem 1과 분할-정복 기법(Divide-and-conquer method)을 통한 거듭 제곱과 Fürer의 곱셈 알고리즘 등을 이용하여 $O(\log p \log\log p)$시간에 계산할 수 있다. 여기에

Algorithm 1 Legendre's symbol heuristic

```
p ← RandomPrime()
m ← Pow(a,(p−1)/2,p)
if m == 1 then
    return true
else
    return false
end if
```

서, $Pow(a, x, p)$는 a mod p를 분할-정복 기법을 통해서 충분히 작은 p에 대하여 $O(\log p)$시간에 계산하는 함수이다. 다음과 같은 정리를 이용하여 통계학적인 분류에 중요한 영향을 끼치게 될 것이다.

Lemma 1. 임의의 제곱수 n^2과 홀수인 소수 p에 대하여 $\left(\frac{n^2}{p}\right) \equiv 1$이다.

Proof. theorem 1과 Fermat's little theorem [2]에 의하여 $\left(\frac{n^2}{p}\right) \equiv n^{p-1} \equiv 1$ (mod p)가 성립한다. □

따라서 우리는 임의의 제곱수가 아닌 수에 대하여 어떤 홀수인 소수 p의 이차 잉여가 되는 경우를 확인하면 된다. 그리고, 이 과정에서 K개의 소수를 선택하도록

하자. 그리고 이 소수들의 집합을 P라고 하자. 여기에서 정상적인 판단이 안되는 두 가지의 경우가 있다.

1. $p_i \in P$가 $p_i \mid a$인 경우이다. 여기에서 모든 수에 대하여 이러한 상황을 만족할 확률은 약 $\frac{1}{\prod_i p_i}$이다.
2. a가 완전 제곱수가 아니고, $\left(\frac{a}{p_i}\right) \equiv 1$를 만족하는 경우이다.

두번째 경우에 대해서 $a = m^2 \equiv d_i \pmod{\mathrm{p}}$, $n_1 = p_i + b \rightarrow n_1^2 = p_i(p_i + 2b) + b^2$ 이고, $n_1^2 = p_i k + c$, $c < p_i$ 꼴로 나타내어지기 원하니 이런 k의 기댓값은 $p_i > 9$를 가정하여, c가 $\sqrt{p}$를 넘을 확률은 약 $\frac{p_i - \sqrt{p_i}}{p_i}$이고 이 경우에는 $k = p_i + 2b + 1$이니 k의 기댓값을 근사하면, $k_{expect} = \frac{1}{p_i}(p_i + 2b_{expect}) + \frac{p_i - \sqrt{p_i}}{p_i}(p_i + 2b_{expect} + 1) = p_i + 2b_{expect} + 1 - \frac{1}{\sqrt{p_i}} \cong p_i + 2b_{expect} + 1$이다.
또한, b의 기댓값을 근사하면, $b_{expect} = \frac{1}{\sqrt{p_i}}\sqrt{d_i} + \frac{p_i - \sqrt{p_i}}{p_i}(\sqrt{d_i} - p_i) = \frac{1}{\sqrt{p_i}}\sqrt{d_i} + \sqrt{d_i} - p_i - \frac{1}{\sqrt{p_i}}(\sqrt{d_i} - p_i) = \sqrt{d_i} - p_i + \sqrt{p_i}$이다. 결국 고정된 a에 대하여 단위 구간 $[1, n_1^2]$에서는 약 $p_i + 2b_{expect} - 1 = p_i + 2\sqrt{d_i} - 2p_i + 2\sqrt{p_i} - 1 = 2(\sqrt{d_i} + \sqrt{p_i}) - (p_i + 1)$개의 위 조건을 만족하지 않고 d_i와 합동인 수가 존재한다. 따라서 $a \geq p_i + d_i$이고 $[1, a]$에 대하여 위와 같은 조건을 만족하는 수의 개수는 약 $\frac{a(2(\sqrt{d_i} + \sqrt{p_i}) - p_i)}{p_i(p_i + 2\sqrt{d_i} - p_i + \sqrt{p_i})}$개가 존재한다. 따라서 약 $\frac{2(\sqrt{d_i} + \sqrt{p_i}) - p_i}{p_i(2(\sqrt{d_i} + \sqrt{p_i}) - (p_i + 1))}$의 확률로 고정된 d_i에 대하여 정확히 판정할 수 없게 된다.

$$Pr(p_i) = \frac{1}{p_i}\left(1 + \sum_{d_i=1}^{p_i - 1} \frac{2(\sqrt{d_i} + \sqrt{p_i}) - p_i}{2(\sqrt{d_i} + \sqrt{p_i}) - (p_i + 1)}\right) \tag{3}$$

여기에서 $\forall p_i \in P$에 대하여 실패할 확률은 $\prod_{p_i \in P} Pr(p_i)$이니 $|P|$가 커질 수록 더 줄어든다. 또한, p_i가 커질 수록 더 줄어든다. 이 결과의 가장 큰 의미는 구간에 상관없이 확률이 균등하다는 것이다. 따라서 적절하게 각 p_i와 $|P|$를 선택하면 충분히 낮은 오류로 1에 따른 제곱수를 판정할 수 있다.

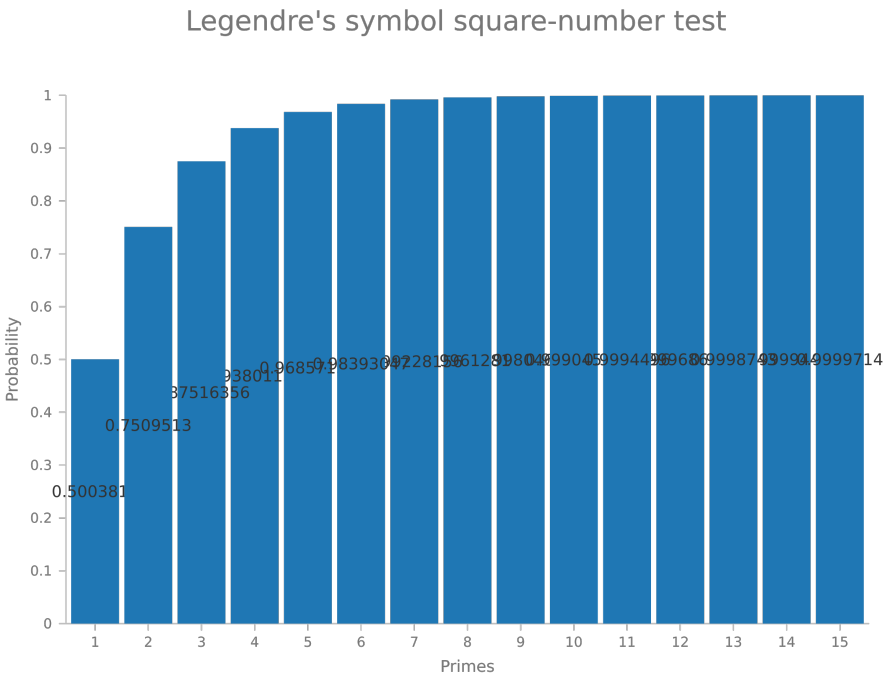

Figure 1: $|P|$의 변화 따른 성공 확률을 나타낸 히스토그램. $[1, 10^6]$ 구간에서 랜덤한 소수에 대하여 확인하였다.

fig. 1에 따르면 machine epsilon $\epsilon = 10^{-3}$이라고 가정할때 $|P| = 10$정도가 수행 시간등의 측면에서 적절하다고 볼 수 있다. 또한, 충분히 큰 랜덤한 소수로 잡는 것으로 충분하다. 여기에서 사용된 코드는 부록에 있는 링크에 존재한다. 또한, 데이터 분석을 통해서 $|P|$가 일정할 때, 임의의 p_i에 대하여 성공할 확률의 변화는 거의 없음을 다음과 같이 확인하였다.

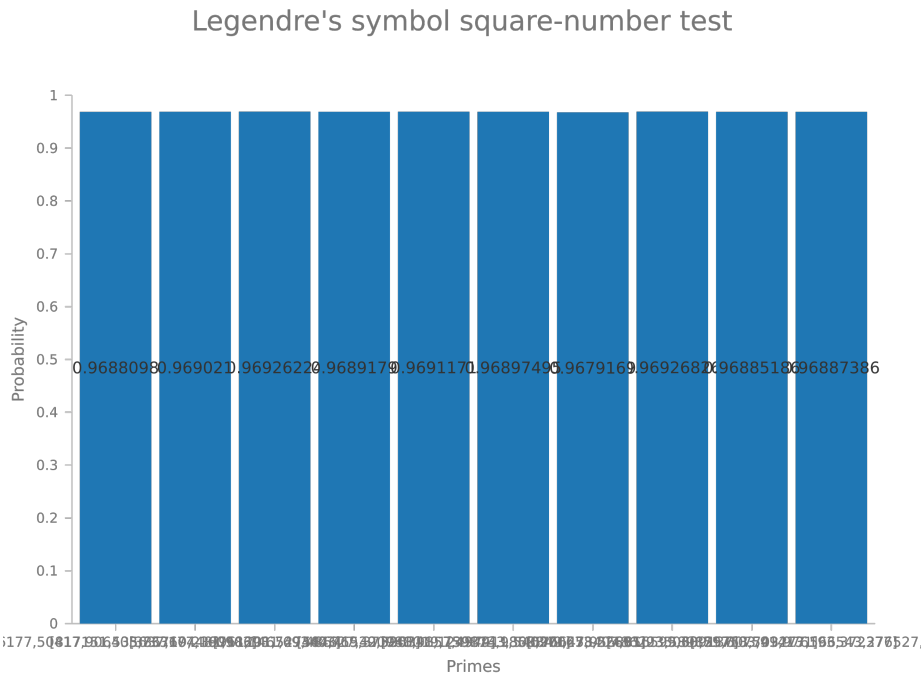

Figure 2: $|P| = 5$로 일정할 때, 여러 P에 대하여 휴리스틱의 성공 확률을 나타낸 히스토그램.

fig. 2와 같은 수행 결과를 얻을 수 있었다.

3 결론

이와 같이 제곱수를 판정하는 정수론적인 휴리스틱을 제시하였다. 하지만, 이를 통해서 완벽히 제곱수를 판정하는 것은 불가능하다. 따라서, 추후 연구 방안으로 임의의 제곱수를 비트로 표현하였을 때 임의의 개수의 제곱수에 관하여 LCS(Longest-common-sequence)를 구하여 여기에서 찾은 LCS 를 통해서 매칭하는 휴리스틱을 제시하고, 위의 방법과 결합하여 더욱 판정률을 높인다.

A 휴리스틱 구현에 대한 소스코드

https://github.com/ANEP-Research/legendre-test 링크에 `Rust` 프로그래밍 언어를 통해서 휴리스틱과 테스트를 수행하는 코드를 볼 수 있다.

References

[1] Fürer, M. Faster integer multiplication. *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing - STOC 07* (2007).

[2] Hardy, G. H., Heath-Brown, D. R., and Wright, E. M. *An introduction to the theory of numbers*. Oxford Univ. Press, 2011.