

Number Theoretic Algorithms

작은 수: 소수 전부 구하기

April 11, 2019

1 체 이론

체 이론(sieve theory)은 “체”라고 불리는 방법으로 생성된 일정한 정수의 집합(sifted set)에 대해 연구하는 분야입니다. 1주차 보강 자료에서 강조했던 추상화¹의 일환이라고 할 수 있습니다. 이 분야는 많이 연구된 분야는 아니나, 쌍둥이 소수나 소수의 합 등을 체계적으로 다룰 수 있게 해 주는, “초등학생에게도 문제를 이해시킬 수 있지만 증명은 온 인류가 하지 못하는” 문제들을 한꺼번에 풀어 버릴 수 있는 강력한 가능성을 가진 분야입니다.

Sifted set을 일반적으로 다루는 것은 상당히 어려우므로 우리는 sifted set의 대표적인 예인 **특정 자연수 이하의 모든 소수의 집합**을 다루고자 합니다. 자연수 n 에 대해 S_n 을 n 이하인 모든 소수의 집합으로 정의하겠습니다.

2 S_n 구하기

여러분은 S_n 을 구하는 빠른 알고리즘인 에라토스테네스의 체를 알고 있습니다.

이 알고리즘의 시간 복잡도는 $\mathcal{O}(n \log \log n)$ 으로 알려져 있습니다. 그런데 이 알고리즘은

$$\sum_{p \leq n} \frac{n}{p} = n \cdot \sum_{p \leq n} \frac{1}{p}$$

시간을 사용하므로, $\sum_{p \leq n} \frac{1}{p} = \mathcal{O}(\log \log n)$ 이라는 말이 됩니다. 이 식의 증명을 시도해 보고자 합니다. 먼저,

$$\begin{aligned} \frac{\pi^2}{6} \cdot \exp\left(\sum_{p \leq n} \frac{1}{p}\right) &\geq \sum_{i=1}^n \frac{1}{i^2} \cdot \prod_{p \leq n} \left(1 + \frac{1}{p}\right) \geq \sum_{i=1}^n \frac{1}{n} \\ &\geq \log(n+1) = \log\left(1 + \frac{1}{n}\right) \log n \end{aligned}$$

에서 양 변에 로그를 취해 $\log \log n + C \leq \sum_{p \leq n} \frac{1}{p}$ 을 얻습니다. 반대 방향 부등식을 증명하

¹같은 것을 같게 보기.

려고 하면

$$\begin{aligned}
 \exp\left(\sum_{p \leq n} \frac{1}{p}\right) &\leq \prod_{p \leq n} \left(1 + \frac{1}{p} + \frac{1}{p^2}\right) \\
 &\leq \frac{\pi^2}{6} \prod_{p \leq n} \left(1 + \frac{1}{p}\right) \\
 &= \frac{\pi^2}{6} \frac{\prod_{p \leq n} \left(1 - \frac{1}{p^2}\right)}{\prod_{p \leq n} \left(1 - \frac{1}{p}\right)} \\
 &\leq \frac{\pi^4}{36} \left(\prod_{p \leq n} \left(1 - \frac{1}{p}\right)\right)^{-1}
 \end{aligned}$$

와 같이 많이 tough합니다. 증명하고자 한 부등식은 Mertens theorem으로 잘 알려져 있는 결과입니다. 각 부등호의 값 차이가 얼마 나지 않는 것 **같은**으로, 남은 내용에서 $\sum_{p \leq n} \frac{1}{p} \approx \log \log n$ 는 증명 없이 사용합니다.

에라토스테네스의 체는 n 이하의 모든 소수를 판별하는 데 각 수 당 $\mathcal{O}(\log \log n)$ 에 확
인합니다. Atkin이 제안한 더 빠른 알고리즘이 있는데, 다음 세 개의 결과를 사용합니다.
squarefree m 에 대해,

- m 이 $4k+1$ 꼴 수이면 $4x^2 + y^2 = m$ 의 양의 정수해의 개수가 홀수이면 소수, 짝수이면
합성수이다.
- m 이 $12k+7$ 꼴 수이면 $3x^2 + y^2 = m$ 의 양의 정수해의 개수가...
- m 이 $12k+11$ 꼴 수이면 $3x^2 - y^2 = m$ 의 $x > y$ 인 양의 정수해의 개수가...

따라서, 위의 두 개에 대해서는 모든 $x \leq \sqrt{m}$, $y \leq \sqrt{m}$ 에 대해 소수/소수 아님 여부를 flip
하는 방식으로 $\mathcal{O}(m)$ 시간 복잡도를 달성할 수 있습니다. 마지막에 대해서는, $r \geq 1$ 에 대해
 $x = r\sqrt{m}$ 이라면,

$$y = \sqrt{3x^2 - m} = \sqrt{3r^2 - 1}\sqrt{m} > \sqrt{r^2}\sqrt{m} = x$$

로부터 모순을 이끌어낼 수 있으므로, $y < x < \sqrt{m}$ 을 알게 되어, 역시 $\mathcal{O}(m)$ 시간 복잡도입
니다. 이 방법을 Sieve of Atkin이라고 부릅니다.

실제 구현 시 이 알고리즘은 cache hit rate가 에라토스테네스의 체에 비해 현저하게 떨어
지기 때문에 10^8 크기에서도 에라토스테네스의 체가 더 빠릅니다. 10^9 크기부터는 Atkin이
에라토스테네스를 앞지르기 시작합니다.

3 $|S_n|$ 구하기

$|S_n|$ 은 보통 $\pi(n)$ 으로 씁니다.

앞 절의 결과 $\sum_{p \leq n} \frac{1}{p} \approx \log \log n$ 을 이용하면 놀라운 사실을 증명할 수 있는데,

$$\begin{aligned} \sum_{p \leq n} \frac{1}{p} &= \sum_{i=1}^n \frac{\pi(i) - \pi(i-1)}{i} \\ &= \frac{\pi(n)}{n} + \sum_{i=1}^{n-1} \frac{\pi(i)}{i(i+1)} \\ &\approx \frac{\pi(n)}{n} + \int_1^n \frac{\pi(x)}{x(x+1)} dx = \log \log n. \end{aligned}$$

물론 $\pi(n)$ 은 적분 가능하지도 않지만, $\pi(n)$ 의 경향성을 잘 따르는 미분가능한 함수가 존재한다고 하면,

$$\frac{x\pi'(x) - \pi(x)}{x^2} + \frac{\pi(x)}{x(x+1)} = \frac{1}{x \log x}$$

와 같은 미분방정식을 풀어서 $\pi(n)$ 의 경향을 추측할 수 있습니다. 이를 조금 더 엄밀하게 증명한 것을 소수 정리라고 하는데, 컴퓨터가 나오고 나서 $\pi(n)$ 을 정확히 구해 이 소수 정리의 물리적인 확증을 얻을 수 있었습니다.

$\pi(n)$ 을 정확하게 구하는 기본적인 idea는 소수를 구하는 것이 합성수를 구하는 것보다 훨씬 어렵다는 것입니다. 예를 들어, $\varphi(x, a)$ 를 x 이하이면서 소인수가 모두 p_a 보다 큰 수의 개수로 정의하면,

$$\varphi(x, a) = \varphi(x, a-1) - \varphi\left(\frac{x}{p_a}, a-1\right)$$

임을 쉽게 확인할 수 있고, 이 식에 square root decomposition을 이용하면 $\varphi(x, a)$ 를 $\mathcal{O}(a\sqrt{x})$ 에 구할 수 있습니다.

$a = \pi(x^{1/2})$ 이면 $\varphi(x, a) = \pi(x) - \pi(x^{1/2}) + 1$ 이고, $\pi(x^{1/2})$ 은 체 등을 이용하여 구하면 S_n 을 정확하게 구하지 않고 그 개수를 구하는 데에는 $\mathcal{O}(x^{1/2}\pi(x^{1/2}))$ 복잡도를 달성할 수 있습니다. 이 방법을 조금 최적화한 것이 Meissel-Lehmer 알고리즘입니다. 이 알고리즘은 $a = \pi(x^{1/3})$ 을 사용합니다.

$P_k(x, a)$ 를 x 이하이면서, 소인수가 정확히 k 개이고 모두 p_a 보다 큰 수의 개수로 정의합니다. 그러면 a 를 잘 골랐기에,

$$\varphi(x, a) = P_0(x, a) + P_1(x, a) + P_2(x, a) = 1 + (\pi(x) - a) + P_2(x, a)$$

가 되고, $\varphi(x, a)$ 와 $P_2(x, a)$ 를 빠르게 구하기만 하면 $\pi(x)$ 를 빠르게 구할 수 있습니다.

$P_2(x, a)$ 는 대단히 쉽게 구할 수 있습니다.

$$\begin{aligned}
P_2(x, a) &= \sum_{a < i \leq j} \mathbf{1}(p_i p_j \leq x) \\
&= \sum_{i=a+1}^{\pi(\sqrt{x})} \sum_{p_i \leq p_j \leq x/p_i} 1 \\
&= \sum_{i=a+1}^{\pi(\sqrt{x})} \left(\pi\left(\frac{x}{p_i}\right) - i + 1 \right)
\end{aligned}$$

체 등을 이용하여 $x^{2/3}$ 이하의 모든 π 들을 구하면 위 값은 $\mathcal{O}(\sqrt{x})$ 시간에 계산할 수 있습니다.

문제는 $\varphi(x, a)$ 인데, 우리는 이미 $x^{2/3}$ 이하의 모든 소수들에 대한 정보를 알고 있으므로, 점화식을 반복하여 적용하다 $x^{2/3}$ 이하로 떨어지게 되면 그때 멈추는 방법을 생각해 볼 수 있습니다. 즉, 위 점화식을 **top-down**으로² 계산하면서 $\varphi(x/n, b)$ 에 도달했을 때 다음의 두 조건 중 하나

- (평범) $n \leq x^{1/3}$ 이면서 $b = 0$
- (특별) $n \geq x^{1/3}$

를 만족하면 멈춥니다. 그러면 φ 의 트리 구조에서 이 값들은 leaf가 됩니다.

먼저 두 leaf $\varphi(x/n, b)$ 와 $\varphi(x/n', b')$ 에 대해 $n \neq n'$ 이라는 관찰이 필요합니다. 이는 n 의 소인수분해의 유일성과 b 의 크기 비교로부터 얻습니다. 그러면 평범한 leaf는 많아 봐야 $x^{1/3}$ 개가 된다는 것을 바로 알 수 있습니다. 특별한 노드는 n 을 높이면서 도달해야 하기 때문에, 부모 노드는 $n^* p_{b+1} = n$ 과 $n^* < x^{1/3}$ 을 만족하는 n^* 에 대해 $\varphi(x/n^*, b+1)$ 입니다. n 의 선택은 많아 봐야 $x^{1/3}$ 개인 n^* 와 많아 봐야 $x^{1/3}$ 개인 b 의 선택³에 의존하기 때문에, 특별한 leaf도 많아 봐야 $x^{2/3}$ 개임을 알 수 있습니다. φ 의 트리는 full binary tree이므로, 이 값을 그냥 계산해도 중간 노드들의 개수까지 $\mathcal{O}(x^{2/3})$ 개 안으로 들어오게 됩니다!

각 노드들의 계수를 구한 다음에는 계산을 해야 하는데, 이는 위 증명에서 n^* 와 b 를 선택하면, 각각의 n 에 대해 b 가 유일하게 결정됨을 알 수 있고, LCA의 “구간 길이를 두 배씩 늘리는 technique”처럼 체의 원소를 차례로 구하면서 필요한 값을 구하면 필요한 모든 값을 $\mathcal{O}(n^{2/3} \log n)$ 에 구할 수 있습니다.⁴

²확신이 없을 땐 상당히 위험한 접근이라는 것을 저보다도 더 잘 알고 있을 것입니다.

³ $a = \pi(x^{1/3})$ 이므로.

⁴ $n^{2/3}$ 까지 모두 한 번씩만 update되기 때문에, sieving 이후 updating을 하는 것과 시간 복잡도 상 차이가 없습니다.

4 문제

1. (Basel Problem)⁵

(a) $\sin z$ 는 복소수 전체에서 **아름답게** 정의되므로, 그 모든 해 $z = n\pi$ 에 대해,

$$\sin z = \prod_{n=-\infty}^{\infty} C_n (z - n\pi)$$

와 같이 쓸 수 있습니다. C_n 을 구하고, 이를 바탕으로 우변을 다시 쓰세요. (Hint: $\lim_{z \rightarrow 0} \frac{\sin z}{z} = 1$ 이후 “합차 공식.”)

(b) $\sin z$ 의 Taylor 전개와 오른쪽 무한 곱의 삼차항을 비교하여,

$$\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$$

을 얻으세요. 이 증명은 gap이 매우 많지만, 수렴값의 납득에는 충분히 유용합니다.

(c) 리만 제타 함수 $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ 로 정의됩니다. 따라서 우리는 $\zeta(2) = \pi^2/6$ 임을 납득했습니다. 이제 $\zeta(4)$ 와 $\zeta(6)$ 에 대해 납득할 만한 값을 찾아보세요. 원한다면 $\zeta(2m)$ 의 일반식을 찾아보세요.

2. (Mertens Theorems)⁶ Mertens는 1874년에 *Ein Beitrag zur analytischen Zahlentheorie*라는 제목의 논문을 발표하는데,⁷ 이 논문의 주요 정리 세 개를 소개하면 다음과 같습니다.

- (First) 다음과 같이 쓰면, 모든 자연수 n 에 대해 $|M(n)| < 2$ 입니다.

$$\sum_{p \leq n} \frac{\log p}{p} = \log n + M(n)$$

- (Second) 아래 극한이 수렴합니다. M 은 Meissel-Mertens 상수라고 합니다.

$$\lim_{n \rightarrow \infty} \left(\sum_{p \leq n} \frac{1}{p} - \log \log n \right) =: M \approx 0.261497$$

- (Third) 아래 극한이 수렴합니다. γ 는 Euler-Mascheroni 상수입니다.

$$\lim_{n \rightarrow \infty} \log n \cdot \prod_{p \leq n} \left(1 - \frac{1}{p} \right) = e^{-\gamma} \approx 0.561459$$

⁵혹시 전혀 모르고 계셨을까봐 덧붙인 문제입니다.

⁶Analytic Number Theory에서 하는 계산에 대한 실마리를 드리기 위해 만든 문제입니다. 건너뛰어도 좋습니다.

⁷해석적 정수론에의 기여라는 제목입니다. 한 분야의 판도를 바꿀 정도로 강력한 논문이었는데, 제목을 정말 겸손하게 지었다고 생각합니다.

- (a) First를 알고 있다고 하고, Second를 증명하세요. (Hint: Abel Summation.)
- (b) Third를 알고 있다고 하고, $\sum_{p \leq n} \frac{1}{p} = \mathcal{O}(\log \log n)$ 을 증명하세요.
- (c) Second를 알고 있다고 하고, Third의 좌변이 수렴함을 알고 있다고 합시다. 이때, 그 수렴값이 $e^{-M-\pi^2/6} \approx 0.148610$ 과 $e^{-M} \approx 0.769898$ 사이에 있음을 보이세요.

3. (Quadratic Form)

- (a) 양수 a 와 squarefree m 에 대해, $x^2 + ay^2 = m$ 의 양수 해 (x, y) 가 존재하는 경우 $xy^{-1} \pmod{m}$ 이 유일함을 보이세요. (Hint: (x, y) 와 (x', y') 으로 둘 있다고 하고, $\frac{x+\sqrt{ai}y}{x'+\sqrt{ai}y'}$ 을 계산하세요.)
- (b) (a)에서, $xy^{-1} \pmod{m}$ 과 $-xy^{-1} \pmod{m}$ 에 대해 동시에 해가 존재할 수 없음을 보이세요.
- (c) 실제로는 모든 가능한 $z \leq |m|/2$ 인 $z^2 = -a \pmod{m}$ 에 대해 $x^2 + ay^2 = \lambda m$ 의 해가 존재하도록 하는 $\lambda \leq \sqrt{\frac{4a}{3}}$ 이 존재합니다. 이를 통해, $n \in 4\mathbb{Z} + 1$ 이거나 $n \in 12\mathbb{Z} + 7$ 인 경우 Atkin의 소수판별법이 옳음을 보이세요. 소수 p 에 대해 $z^2 = a \pmod{p}$ 의 해가 없거나 둘 존재한다는 사실은 자유롭게 사용해도 좋습니다.

4. (Prime Number Theorem) Mertens Second Theorem을 이용하여 $\pi(n)$ 의 대략적인 추세를 알 수 있습니다.

- (a) 위에서 구해놓은 미분방정식을 마저 풀어

$$\pi(n) \approx \int_1^n \frac{1}{\log x} dx \approx \frac{n}{\log n}$$

혹은 $\pi'(n) \approx 1/\log n$ 을 보이세요. 즉 자릿수가 두 배가 되면 소수일 확률은 반이 된다는 의미입니다. Hint.⁸ 원하면 부등호 계산을 이용해서 엄밀하게 증명할 수도 있겠지만 저도 안 해 봤습니다.

- (b) (Erdős) 실제 Prime Number Theorem은

$$\lim_{n \rightarrow \infty} \frac{\pi(n)}{\frac{n}{\log n}} = 1$$

을 얘기합니다. 이를 통해 Erdős 정리:

임의의 자연수 k 에 대해 $n > N$ 일 때마다 n 과 $2n$ 사이에 소수가 k 개 이상 있도록 하는 적당한 자연수 N 이 존재한다

를 증명하세요.

5. Binary search나 Newton's method 등을 이용하여 31415926536번째 소수를 구하세요.⁹

⁸ $u(x)\pi(x)$ 를 미분하여, $u(x)$ 가 어떤 함수여야 하는지 먼저 알아내세요. 물론 여기에도 미분방정식이 들어갑니다. 적분식과 $\frac{n}{\log n}$ 을 근사할 수 있음을 보이려면, 비의 극한이 1임을 보이면 됩니다. 이를 위해 로피탈의 법칙을 이용하세요.

⁹정답은 829470385207이며, 제 컴퓨터에서는 1초 미만의 시간이 걸렸습니다.