# 피타고라스 쌍

kipa00

September 11, 2018

## 문제

**피타고라스 쌍**은 $n < m$인 1 이상의 자연수 $m$, $n$에 대해

$$a = m^2 - n^2$$
$$b = 2mn$$
$$c = m^2 + n^2$$

인 $(a, b, c)$를 말합니다. 이때 $(a, b, c)$는 $(m, n)$으로부터 **생성되었다**고 얘기합니다.
**원시 피타고라스 쌍** $(a, b, c)$는 피타고라스 쌍이면서 $\gcd(a, b, c) = 1$인 것입니다. 당신은 $L$이 주어졌을 때, $n < m \le L$인 $(m, n)$으로부터 생성된 원시 피타고라스 쌍의 개수를 구해야 합니다.

## 풀이

먼저 동치 조건을 생각합니다. $\gcd(n, m) =: g \ne 1$이면 $a$, $b$와 $c$ 모두 $g^2$으로 나누어떨어지기 때문에 $g = 1$이어야 합니다. 또한, $n$과 $m$이 모두 홀수라면 $a$, $b$와 $c$ 모두 2로 나누어떨어지기 때문에 이것도 피해야 합니다. 따라서, 최소한

- $\gcd(n, m) = 1$
- $n$, $m$ 중 하나 이상은 짝수

여야 합니다. 그런데 이것은 동치 조건입니다: 위 조건을 만족함을 가정합니다. 그러면 $a$는 홀수인데, $\gcd(a, b) = \gcd\left(m^2 - n^2, 2mn\right) = g' \ne 1$이면

- $g'$의 1이 아닌 square-free 약수 중 $m$을 나누는 것이 있다면 이것을 $G$라 합시다. $\left(m^2 - n^2\right)$이 $G$로 나누어떨어져야 하고, $m$은 $G$로 나누어떨어지기 때문에 $n^2$이 $G$로 나누어떨어져야 합니다. $G$가 **square-free**이므로 $n$이 $G$로 나누어떨어지며, $G$는 1이 아니므로 이는 $\gcd(n, m) = 1$에 모순입니다.
- 없다면, $g'$의 1이 아닌 square-free 약수 중 $n$을 나누는 것이 반드시 있습니다: 이것을 $G$라 하고 똑같은 논리를 적용하여 모순을 얻습니다.

따라서 $\gcd(a, b, c) = 1$일 필요충분조건은 $\gcd(n, m) = 1$이면서 $n$, $m$ 중 하나가 짝수인 것입니다. 이제 $n, m \le L$이면서 $\gcd(n, m) = 1$인 것의 개수를 센 다음, $n, m \le L$, $\gcd(n, m) = 1$이면서 $n$, $m$이 모두 홀수인 것의 개수를 빼고, 그 값을 2로 나누면 원하는 결과를 얻습니다.

## $n, m \le L$이면서 $\gcd(n, m) = 1$인 쌍의 개수

포함-배제 원리를 이용하면 다음과 같은 식을 얻을 수 있습니다. $p_i$는 소수입니다.

$$L^2 - \sum_{p_1} \left\lfloor \frac{L}{p_1} \right\rfloor^2 + \sum_{p_1, p_2} \left\lfloor \frac{L}{p_1 p_2} \right\rfloor^2 - \sum_{p_1, p_2, p_3} \left\lfloor \frac{L}{p_1 p_2 p_3} \right\rfloor^2 + \cdots$$

각각의 floor function 분모에 속하는 수들이 (소인수분해의 유일성에 의해) distinct하게 결정됩니다. 따라서, 분모를 합쳐서 이런 식으로 쓰면,

$$\sum_{i=1}^{L} \mu(i) \left\lfloor \frac{L}{i} \right\rfloor^2$$

$\mu(i)$들이 계산된 이후 linear time에 구현할 수 있을 것 같습니다. $\mu(i)$는 위 식에서 분모가 $i$일 때의 계수이고, 구체적으로는

- $i$가 square-free이고 소인수가 홀수 개이면 -1
- $i$가 1이거나, [square-free이고 소인수가 짝수 개이면] 1
- 모두 해당하지 않는 경우 0

입니다. 그런데 $i$가 $L$에 가까운 경우 $\left\lfloor \frac{L}{i} \right\rfloor$의 값은 그렇게 자주 변하지 않습니다. 그래서 만약

$$M(k) := \sum_{i=1}^{k} \mu(i)$$
$$M(x) := M(\lfloor x \rfloor) \quad (x \in \mathbb{R})$$

를 빠르게 구할 방법이 있다면, 이 식을 다음

$$\sum_{i=1}^{L} \mu(i) \left\lfloor \frac{L}{i} \right\rfloor^2 = \sum_{i=1}^{a_L - 1} \mu(i) \left\lfloor \frac{L}{i} \right\rfloor^2 + \sum_{j=1}^{\lfloor n/a_L \rfloor} \left( M\left(\frac{n}{j}\right) - M\left(\frac{n}{j+1}\right) \right) j^2$$

과 같이 변형하여 $O\left(a_L + \frac{L}{a_L}\right)$에 문제를 해결할 수 있습니다. 이때 $a_L$은 $\left\lfloor \frac{L}{a_L - 1} \right\rfloor \neq \left\lfloor \frac{L}{a_L} \right\rfloor$을 만족하는 1에서 $L$까지의 수 중에 고르면 되는데, AM-GM inequality 등을 이용해 $a_L$이 $\sqrt{L}$에 가장 가까울 때 시간복잡도 $O(\sqrt{L})$로 문제를 해결할 수 있음을 보일 수 있습니다.
이제 필요한 $M$의 값은 $a_L$보다 작은 $i$에 대해 $M(i)$, 그리고 $M(L/j)$ 꼴임을 알 수 있습니다.
$i = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k} \neq 1$라 놓고, 함수 $d$를 계산합니다:

$$\begin{aligned} d(i) &:= \sum_{j \mid i} \mu(j) \\ &= 1 - k + \binom{k}{2} - \cdots + (-1)^k \binom{k}{k} \\ &= \sum_{i=0}^{k} \binom{k}{i} (-1)^i = (1 + (-1))^k = 0. \end{aligned}$$

이제, 다음과 같은 조금 뜬금없는 식을 계산합니다.

$$\begin{aligned}\sum_{i=1}^{L} M\left(\frac{L}{i}\right) &= \sum_{i=1}^{L} \sum_{j=1}^{\lfloor L/i \rfloor} \mu(j) \\ &= \sum_{j=1}^{L} \left\lfloor \frac{L}{j} \right\rfloor \mu(j) \\ &= \sum_{j=1}^{L} \sum_{\substack{i=1 \\ j|i}}^{L} \mu(j) \\ &= \sum_{i=1}^{L} d(i) = d(1) = 1.\end{aligned}$$

**점화식이 나왔습니다!** 이 점화식 역시 비슷한 방법으로 (이전 값이 모두 계산되어 있다면) $O(\sqrt{L/i})$만에 $M\left(\frac{L}{i}\right)$를 계산할 수 있습니다.
그러나 우리는 $a_L$ 이하에서의 촘촘한 값도 필요합니다. $b_L$ 이하의 값을 촘촘하게 전부 계산한다고 하면, $b_L > a_L$이고, 에라토스테네스의 체를 이용해 $O(b_L \log\log b_L)$만에 $b_L$ 이하의 $\mu$와 $M$ 값을 전부 계산할 수 있습니다. 그 이상부터는 $1 \le i \le L/b_L$에서의 $M\left(\frac{L}{i}\right)$ 값이 필요한데요, 이것을 구하는 시간은

$$\begin{aligned}\sum_{i=1}^{\lceil L/b_L \rceil} \sqrt{\frac{L}{i}} &\approx \int_{1}^{L/b_L} \sqrt{\frac{L}{x}} dx \\ &= 2\sqrt{Lx}\Big|_{x=1}^{x=L/b_L} \\ &\approx \frac{2L}{\sqrt{b_L}}\end{aligned}$$

입니다. 따라서 원하는 모든 $M$ 값을 구하는 시간은 $O\left(b_L \log\log b_L + \frac{2l}{\sqrt{b_L}}\right)$인데요, 역시 (작은 $\log\log$ 항을 무시하고) AM-GM inequality 등을 이용해서 $b_L$가 $L^{2/3}$에 가까울 때 가장 빠르다는 것을 확인할 수 있습니다. $b_L$에 상수를 조금 붙이면 시간 복잡도가 약간 더 줄어들지만, $b_L \approx L^{2/3}$으로 놓고 풀어도 시간 복잡도는 $O\left(L^{2/3} \log\log L\right)$이라 문제를 풀기에 충분합니다.

## $n, m \le L$, $\gcd(n,m) = 1$이면서 $n$과 $m$이 모두 홀수인 쌍의 개수

역시 포함-배제 원리를 이용하면 다음과 같은 식을 얻을 수 있습니다. 이 식에서는 $p_i$는 2가 아닌 소수입니다.

$$\left\lfloor \frac{L+1}{2} \right\rfloor^2 - \sum_{p_1} \left\lfloor \frac{\left\lfloor \frac{L}{p_1} \right\rfloor + 1}{2} \right\rfloor^2 + \sum_{p_1,p_2} \left\lfloor \frac{\left\lfloor \frac{L}{p_1p_2} \right\rfloor + 1}{2} \right\rfloor^2 - \sum_{p_1,p_2,p_3} \left\lfloor \frac{\left\lfloor \frac{L}{p_1p_2p_3} \right\rfloor + 1}{2} \right\rfloor^2 + \cdots$$

마찬가지로, 다음과 같이 홀수들만 고려하는 $M$ 함수를 $M'$이라 정의하면,

$$\begin{aligned}M'(k) &:= \sum_{\substack{i=1 \\ i \textbf{ odd}}}^{k} \mu(i) \quad (k \text{ odd}) \\ M'(x) &:= M'(\lfloor x \rfloor) \quad (x \in \mathbb{R})\end{aligned}$$

이 식을 다음과 같이 정리할 수 있습니다.

$$\sum_{\substack{i=1 \\ i \text{ odd}}}^{L} \mu(i) \left\lfloor \frac{\left\lfloor \frac{L}{i} \right\rfloor + 1}{2} \right\rfloor^2 = \sum_{\substack{i=1 \\ i \text{ odd}}}^{c_L-2} \mu(i) \left\lfloor \frac{\left\lfloor \frac{L}{i} \right\rfloor + 1}{2} \right\rfloor^2 + \sum_{j=1}^{\lfloor n/c_L \rfloor} \left( M'\left(\frac{n}{j}\right) - M'\left(\frac{n}{j+1}\right) \right) \left\lfloor \frac{j+1}{2} \right\rfloor^2$$

역시 $c_L$은 $\left\lfloor \frac{L}{c_L-2} \right\rfloor \neq \left\lfloor \frac{L}{c_L} \right\rfloor$를 만족하는 1에서 $L$까지의 수 중에 고르면 됩니다. $M'$은 다음과 같은 기초적인 성질을 이용해 점화식을 세울 수 있습니다.

$$\begin{aligned} M'(k) &= \mu(1) + \mu(3) + \mu(5) + \cdots + \mu(k) \\ &= M(k) - \sum_{i=1}^{(k-1)/2} \mu(2i) \\ &= M(k) - \sum_{\substack{i=1 \\ i \text{ odd}}}^{(k-1)/2} \mu(2i) - \sum_{\substack{i=1 \\ i \text{ even}}}^{(k-1)/2} \mu(2i) \\ &= M(k) + M'\left(\frac{k}{2}\right) \quad (k \text{ odd}) \end{aligned}$$

약간 경우 분류를 하면, $M'(x) = M(x) + M'(x/2)$가 모든 실수 $x$에 대해 성립함을 알 수 있습니다.[1] 따라서 전처리를 해 둘 수 있습니다.

## 최적화

시간 제한이 매우 넉넉하기 때문에 굳이 하지 않아도 통과하지만, 약간의 최적화 기법을 소개합니다.

- $n$과 $m$이 모두 홀수인 쌍의 개수를 구할 때 사용하는 아래의 식에서, 본인의 계산 실력에 자신이 있으시다면 아래 식의 $\lfloor (j+1)/2 \rfloor$ 부분을 한 번 더 정리할 수 있고, 약간 더 빨라집니다. 이 값은 $j$가 $2k$와 $2k-1$ 꼴일 때 같은 값을 가지기 때문에 $L$의 홀짝성에 따라 약간의 경우 분류를 해 주셔야 합니다.

$$\sum_{j=1}^{\lfloor n/c_L \rfloor} \left( M'\left(\frac{n}{j}\right) - M'\left(\frac{n}{j+1}\right) \right) \left\lfloor \frac{j+1}{2} \right\rfloor^2$$

- $a_L$로 가능한 범위에 대해 생각해 봅시다. $\left\lfloor \frac{L}{a_L-1} \right\rfloor \neq \left\lfloor \frac{L}{a_L} \right\rfloor$를 만족한다면,

$$\begin{aligned} \frac{L}{a_L - 1} - \frac{L}{a_L} &\leq 1 \\ \frac{L}{a_L(a_L-1)} &\leq 1 \\ a_L \geq \frac{\sqrt{4l+1}+1}{2} &\geq \sqrt{L} \end{aligned}$$

따라서, $a_L$을 계산할 때 $\sqrt{L}$까지는 나눗셈을 건너뛸 수 있습니다. 마찬가지 방법으로 $c_L$에 대해서도 나눗셈을 건너뛸 수 있습니다.

[1] PS에서 가장 어려운 걸 하나만 꼽으라면 디버깅과 off-by-1 error라는 농담도 있죠.