Project Euler Problem 625: GCD Sum

Wee JunJie

July 16, 2019

1 Introduction

This question wants us to compute the following double summatory formula:

$$\text{i.e. } G(N) := \sum_{j=1}^{N} \sum_{i=1}^{j} \gcd(i, j)$$

for $N = 10^{11}$

2 Main Part of the Problem

The first thing to note for $G(N)$ is that for large N such as 10^{11}, any solution that requires a summation from 1 to N is too slow. Hence, any solution to compute $G(N)$ requires a sublinear algorithm. E.g. Computation of $O(\sqrt{N})$ complexity since $\sqrt{10^{11}} \approx 316227$ is more feasible. Knowing this problem, we require a few techniques which will be introduced.

2.1 Preliminaries

Definition 2.1.1 (Pillai's Arithmetical Function)**.** In number theory, the gcd summatory function also known as the Pillai Arithmetical Function is defined as

$$P(n) := \sum_{i=1}^{j} \gcd(i, j)$$

or equivalently,

$$P(n) = \sum_{d|j} d\varphi(j/d) \tag{1}$$

where $\varphi(n)$ is the euler totient function of n.

Definition 2.1.2 (Dirichlet Convolution)**.** Let f, g be arithmetical functions mapped from $\mathbb{N} \to \mathbb{C}$ Then the Dirichlet Convolution $f * g$ is a new arithmetical function defined as:

$$(f * g)(n) := \sum_{d|n} f(d)g(n/d) = \sum_{ab=n} f(a)g(b)$$

Remark 2.1.3. It is easy to see that the Pillai's Arithmetical Function is a Dirichlet Convolution due to (1).

Theorem 2.1.4 (Dirichlet Hyperbola Method)**.** Let f be the Dirichlet convolution of $g * h$. Then the summatory function of $f(n)$ can be written as:

$$\sum_{j=1}^{N} f(j) = \sum_{j\leq N}\sum_{ab=j} g(a)h(b) = \sum_{a\leq\sqrt{N}}\sum_{b\leq\frac{N}{a}} g(a)h(b) + \sum_{b\leq\sqrt{N}}\sum_{a\leq\frac{N}{b}} g(a)h(b) - \sum_{a\leq\sqrt{N}}\sum_{b\leq\sqrt{N}} g(a)h(b) \tag{2}$$

2.2 Manipulating the GCD Sum

Based on the preliminaries, we can work to arrive at the following:

$$\begin{aligned}
G(N) &= \sum_{j=1}^{N}\sum_{i=1}^{j} \gcd(i,j) \\
&= \sum_{j\leq N}\sum_{d\mid j} d\varphi(j/d) \quad \text{By (1)} \\
&= \sum_{j\leq N}\sum_{d(j/d)=j} d\varphi(j/d) \quad \text{(By Dirichlet Convolution)} \\
&= \sum_{d\leq\sqrt{N}}\sum_{(j/d)\leq\frac{N}{d}} d\varphi(j/d) + \sum_{(j/d)\leq\sqrt{N}}\sum_{d\leq\frac{N}{(j/d)}} d\varphi(j/d) - \sum_{d\leq\sqrt{N}}\sum_{(j/d)\leq\sqrt{N}} d\varphi(j/d) \\
&\text{(By Dirichlet Hyperbola Method)} \\
&= \sum_{d\leq\sqrt{N}} d \sum_{(j/d)\leq\frac{N}{d}} \varphi(j/d) + \sum_{(j/d)\leq\sqrt{N}} \varphi(j/d) \sum_{d\leq\frac{N}{(j/d)}} d - \sum_{d\leq\sqrt{N}} d \sum_{(j/d)\leq\sqrt{N}} \varphi(j/d) \\
&= \sum_{d\leq\sqrt{N}} d \sum_{y\leq\frac{N}{d}} \varphi(y) + \sum_{y\leq\sqrt{N}} \varphi(y) \sum_{d\leq\frac{N}{y}} d - \sum_{d\leq\sqrt{N}} d \sum_{y\leq\sqrt{N}} \varphi(y) \\
&= \sum_{d\leq\sqrt{N}} \Big(d\cdot H(N/d)\Big) + \sum_{y\leq\sqrt{N}} \Big(\varphi(y)\cdot F(N/y)\Big) - F(\sqrt{N})\cdot H(\sqrt{N})
\end{aligned}$$

where

$$F(N) = \sum_{i=1}^{N} i = \frac{N(N+1)}{2}, \quad H(N) = \sum_{i=1}^{N} \varphi(i).$$

To summarise, we essentially need to implement the following:

$$G(N) = \sum_{d\leq\lfloor\sqrt{N}\rfloor} \Big(d\cdot H(\lfloor N/d\rfloor)\Big) + \sum_{y\leq\lfloor\sqrt{N}\rfloor} \Big(\varphi(y)\cdot F(\lfloor N/y\rfloor)\Big) - F(\lfloor\sqrt{N}\rfloor)\cdot H(\lfloor\sqrt{N}\rfloor)$$

and

$$F(N) = \sum_{i=1}^{N} i = \frac{N(N+1)}{2}, \quad H(N) = \sum_{i=1}^{N} \varphi(i).$$

At least, now we can see that $G(N)$ has approximately $O(\sqrt{N})$ time complexity. However, note that in the first sum of $G(N)$, we are required to compute $H(\lfloor N/d\rfloor)$, and if $d=1$, $N=10^{11}$, then we have to compute

$$H(10^{11}) = \sum_{i=1}^{10^{11}} \varphi(i).$$

which will be still not sufficiently feasible. This $H(N)$ will require a recursive method and memoization in order to compute $G(N)$ which will be discussed in the next section.

2.3 Computing $H(N)$

In order to manipulate $H(N)$, recall that the $\varphi(n)$ is the euler totient of n which computes the number of gcd pairs of numbers $1 \leq a \leq n$ such that $\gcd(a, n) = 1$. Hence, $H(N)$ computes the number of gcd pairs of numbers $1 \leq a \leq b \leq N$ such that $\gcd(a, b) = 1$. Note that the total number of gcd pairs of $1 \leq a \leq b \leq N$ regardless of gcd value is $N(N+1)/2$. Therefore,

$$H(N) = \frac{N(N+1)}{2} - \sum_{m=2}^{N} H(\lfloor N/m \rfloor)$$

where the 2nd term sums up all the gcd pairs that are not equal to 1.

Now suppose there exist a divisor d such that $\lfloor N/m \rfloor = d$. Then

$$\begin{aligned} \lfloor N/m \rfloor = d &\iff d \leq N/m < d+1 \quad \text{(By definition of floor function.)} \\ &\iff N/(d+1) < m \leq N/d \end{aligned}$$

Note that for any $m \geq \sqrt{N}$, $d = \lfloor N/m \rfloor$ will be less than or equal to $\sqrt{N}$ and there exist more than 1 integer of m for interval $N/(d+1) < m \leq N/d$. Therefore, by doing the following, we break the sum into 2 terms and arrive at a recursive function:

$$\begin{aligned} H(N) &= \frac{N(N+1)}{2} - \sum_{m=2}^{N} H(\lfloor N/m \rfloor) \\ &= \frac{N(N+1)}{2} - \sum_{m=2}^{\sqrt{N}} H(\lfloor N/m \rfloor) - \sum_{m=\sqrt{N}+1}^{N} H(\lfloor N/m \rfloor) \\ &= \frac{N(N+1)}{2} - \sum_{m=2}^{\sqrt{N}} H(\lfloor N/m \rfloor) - \sum_{d=1}^{\sqrt{N}} (\lfloor N/(d+1) \rfloor - \lfloor N/d \rfloor) H(d) \\ &\qquad (H(d) \text{ is multiplied by the number of integer values of } m.) \end{aligned}$$

By implementing $H(N)$ using the last equality, we can compute $H(N)$ sufficiently fast to obtain $G(N)$ for $N = 10^{11}$. I computed my program in Python which took about ≈ 9 minutes.