积性函数前缀和——洲阁筛

问题描述

洲阁筛解决的问题主要是n范围较大的积性函数前缀和。

已知一积性函数$f(i)$，求$\sum_{i=1}^{n} f(i)$。

$n \le 10^{12}$.

求解方法

如果$f(i)$在质数处的取值比较简单，那么可以运用洲阁筛来求解。

我们需要两个辅助数组。

$g_{i,j}$

定义如下：

$$\begin{aligned} g_{i,j} &= \sum_{k=2}^{i}[k\text{与}p_1,p_2,\dots,p_j\text{互质或就是其中某个质数}]\ s(k) \\ &= \sum_{k=2}^{i}[k\text{是 }\le p_j\text{的质数或}k\text{的最小质因子大于}p_j]\ s(k) \end{aligned}$$

其中$s(x)$是一个积性函数，它可以是$s(x)=x$，或$s(x)=1$，等等。

这个数组怎么求呢？

首先边界条件比较简单：

$$g_{i,0} = \sum_{k=2}^{i} s(k)$$

考虑由$g_{i,j-1}$推出$g_{i,j}$。$g_{i,j-1}$代表着一些在$j-1$时合法的数的函数值之和，而从$j-1$变成j时，若某些原本合法的数变得不合法，显然一定是因为触犯了第二个条件：最小质因子恰好为p_j。如何算出这一部分的数的函数值之和呢？

$$g_{i,j} = g_{i,j-1} - s(p_j)(g_{\lfloor \frac{i}{p_j} \rfloor, j-1} - g_{p_j-1,j-1})$$

后面减去的部分就是这一部分数的函数值之和。首先它们都有一个最小质因子p_j，随后，它们除去p_j剩下的部分，不可以含有小于p_j的质因子，因此剩下的部分至少大于等于p_j，但要小于等于$\frac{i}{p_j}$。这恰好对应了g的定义！后面一部分算的就是

$$\sum_{k=p_j}^{\frac{i}{p_j}}[\text{最小质因子} > p_{j-1}]\ s(k)$$

$h_{i,j}$

定义如下：

$$h_{i,j}=\sum_{k=2}^{i}[k\text{的最小质因子} \ge p_j]\ f(k)$$

其递推式为：

$$\begin{aligned}h_{i,j}&=\sum_{p_k\ge p_j}\sum_{e\ge 1\text{且}\,p_k^e\le i}f(p_k^e)h_{\lfloor\frac{i}{p_k^e}\rfloor,j+1}+f(p_k^e)\\&=\sum_{p_k\ge p_j}\sum_{e\ge 1\text{且}\,p_k^e\le i}f(p_k^e)(h_{\lfloor\frac{i}{p_k^e}\rfloor,j+1}+1)\end{aligned}$$

它相当于枚举所有最小质因子大于等于p_j的数，并对函数值求和。首先枚举它们的最小质因数p_k，再枚举最小质因数p_k的幂；最后统计有多少个数，满足除去p_k^e后剩余部分最小质因子大于p_j，也就是大于等于p_{j+1}，这和h本身的定义恰好符合。如此枚举，不会算重，但是会漏掉一种情况：$f(p_k^e)$没有被算入，因为h的$\sum$是从2开始枚举的。额外加上即可。

我们使用搜索计算h。这里的搜索不需要记忆化，因为有结论是：如果要求不同的$h_{i,j}$，它们在搜索时不会搜索到重复的地方。

求解问题

以$f(x)=\varphi(x)$举例子，要求$\sum_{i=1}^{n}\varphi(i)$。求解分为两个部分，一个是$\sum_{\sqrt{n}<p\le n}f(p)$，一个是剩余部分。

对于第一部分，首先我们求出函数在所有质数处的值之和$\sum_{p\le n}f(p)$。

当x为质数p时，$\varphi(p)=p-1$。因此：

$$Part_1=\sum_{p\le n}f(p)=\sum_{p\le n}(p-1)=\sum_{p\le n}p-\sum_{p\le n}1$$

首先看第一个和式。我们把p按大小分成两部分：$p\in[1,\sqrt{n}]$和$p\in(\sqrt{n},n)$，对于后半部分质数，它们的最小质因子即本身，必定比小于等于$\sqrt{n}$的最大质因数p_m要大。两部分加起来恰好就是$s(x)=x$意义下的$g_{n,m}$；同理，后面一个和式，恰好就是$s(x)=1$意义下的$g_{n,m}$。

至于$p>\sqrt{n}$的限制，由于$\sqrt{n}$是线性筛可行的范围10^6，直接计算出$\sum_{p\le\sqrt{n}}f(p)$，从刚才的式子中减去即可。

对于第二部分中的数，要么是小于等于$\sqrt{n}$的数，要么是大于$\sqrt{n}$的合数，它们的最小质因子都必定小于等于$\sqrt{n}$。我们计算每个质数的贡献，枚举每个小于等于$\sqrt{n}$的质数p，再枚举它作为最小质因子的指数e，最后枚举有多少种x满足x除去p^e后，最小质因子大于p，这和h的运算思路非常的相像！同理，我们会漏掉p^e这个数本身，因此情况数要加一。具体的说：

$$Part_2=\sum_{p_i\le\sqrt{n}}\sum_{p_i^e\le n}\varphi(p_i^e)(h_{\lfloor\frac{n}{p_i^e}\rfloor,i+1}+1)$$

至此$Part_1+Part_2$就是所求答案。

$$\sum_{i=1}^{n}\varphi(i)=g_{n,m,s(x)=x}+g_{n,m,s(x)=1}+\sum_{p_i\le\sqrt{n}}\sum_{p_i^e\le n}\varphi(p_i^e)(h_{\lfloor\frac{n}{p_i^e}\rfloor,i+1}+1)$$
$$m\text{为满足}p_m\le\sqrt{n}\text{的最大正整数}$$

总结

整体思路稍微有点复杂，但只要明白了洲阁筛对积性函数求和的关键步骤，就可以比较好地理解。

首先，我们所讨论的积性函数，最好在质数处有简明的表达式。我们可以将表达式写出后，对于每个和式，用g在不同$s(x)$的意义下逐个求解。

然后要理解洲阁筛对分配律的利用。它通过枚举最小质因数、枚举其作为最小质因数的指数、最后统计除去最小质因数后，剩余部分最小质因数大于自己的情况数这一种枚举方法，可以结合积性函数性质、运用h数组快速枚举遗漏的数，并得到其函数值之和。

总的时间复杂度为$O(\frac{n^{\frac{3}{4}}}{log_2 n})$。