一类分数问题的研究

Yang Zhe

2007 年 1 月 6 日

摘 要

本文从一个简单问题出发, 介绍一些相关知识, 并在分析与解决这个问题的同时, 尽可能引出新的问题, 并分析它与已知问题的联系, 由这些新问题又可引出其他相关问题... 运用这些知识, 本文主要讨论了这三个问题:

1. 求两实数之间分母最小的分数
2. 求 Farey 序列的第 k 项
3. 求分母不大于一给定值的与给定实数最接近的分数

目 录

1 一道经典题目——Farey 序列的定义及构造 **1**

2 Farey 序列的两个性质 **3**

3 两分数间分母最小的分数 **4**

4 简单连分数及最优渐进分数 **5**

4.1 连分数的基本概念与性质 5

4.2 求有理数的简单连分数表示 6

4.3 最优渐进分数 6

5 Farey 序列上的一些运算 **7**

5.1 求前(后)项 7

5.2 求前(后) k 项 7

5.3 Farey 序列的另一种构造方法 7

5.4 $\mathcal{F}_n$ 中不大于一给定实数的项数 7

5.5 求第 k 项 8

6 对一些最优渐近分数的讨论 **8**

6.1 分母不大于一给定值的与给定分数最接近的分数 8

6.2 分母不大于一给定值的小于给定分数的最大分数 8

6.3 分母不大于一给定值的与给定实数最接近的分数 9

7 再看两分数间分母最小的分数 **9**

7.1 大于一给定实数小于一给定分数的分母最小的分数 10

7.2 两实数间分母最小的分数 10

8 总结 **10**

1 一道经典题目——Farey 序列的定义及构造

约定 如无特殊说明, 本文中出现的字母均表示整数, 所有用作分母的字母均表示正整数. 本文所讨论的分数均满足分子为整数, 分母为正整数, 但不限制为既约分数.

题目描述 给出正整数 $n(2 \leq n \leq 5000)$, 请按从小到大的顺序输出所有分母不大于 n 的既约真分数.

这个既约分数的序列叫做第 n 个 Farey 序列, 记作 $\mathcal{F}_n$. 当 $n = 5000$ 时, $\mathcal{F}_n$ 中共有 7600457 项; 当 $n = 1000000$ 时, $\mathcal{F}_n$ 中共有 303963552391 项. 第 7 个 Farey 序列如下:

$$\frac{1}{7}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{2}{7}, \frac{1}{3}, \frac{2}{5}, \frac{3}{7}, \frac{1}{2}, \frac{4}{7}, \frac{3}{5}, \frac{2}{3}, \frac{5}{7}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{6}{7}.$$

分析 根据题目的提示, 这个序列中的项数是非常多的. 我们必须找到一个 $O(|\mathcal{F}_n|)$ 的算法. 由于这个算法不容易想到, 这里先给出这个算法, 再对它进行分析.

解法 观察这个序列, 可以发现, 其中任意连续三项 $\frac{a}{b}$, $\frac{p}{q}$, $\frac{x}{y}$ 满足

$$\frac{p}{q} = \frac{a+x}{b+y}.$$

定义两个分数 $\frac{a}{b}$, $\frac{x}{y}$ 的中项为 $\frac{a+x}{b+y}$, 那么 $\mathcal{F}_n$ 中任意连续三项的中间项为其两边两项的中项. 此性质留在后面证明.

当分数 $\frac{a}{b}$, $\frac{x}{y}$ 满足 $b, y \le n$, $0 \le \frac{a}{b} < \frac{x}{y} \le 1$, 且 $ay + 1 = bx$ 时, 定义 $\textsc{Build}(a, b, x, y)$ 为输出 $\mathcal{F}_n$ 中 $\frac{a}{b}$ 与 $\frac{x}{y}$ 之间的序列的过程, 那么 $\textsc{Build}$ 过程可递归定义如下:

$\textsc{Build}$ 过程的定义

procedure $\textsc{Build}(a, b, x, y)$
  $p \leftarrow a + x$
  $q \leftarrow b + y$
  if $q \le n$ **then**
    $\textsc{Build}(a, b, p, q)$
    $\textsc{Print}\ \frac{p}{q}$
    $\textsc{Build}(p, q, x, y)$
  end if
end procedure

也就是说, 这个过程递归地产生中项, 直到中项的分母大于 n.

正确性的证明 首先证明两个引理,

引理 1.1 *如果分数* $\frac{a}{b}$, $\frac{x}{y}$ *满足* $\frac{a}{b} < \frac{x}{y}$, *且* $ay + 1 = bx$. *设* $\frac{p}{q}$ *为满足* $\frac{a}{b} < \frac{p}{q} < \frac{x}{y}$ *的分母最小的分数, 则*

$$p = a + x, \quad q = b + y.$$

证明 首先, 因为 $\frac{a}{b} < \frac{a+x}{b+y} < \frac{x}{y}$, 即 $\frac{a+x}{b+y}$ 是一个可行解, 所以 $q \le b + y$.

另一方面, 因为 $ay + 1 = bx$, 所以 $(a, x) = 1$, $(b, y) = 1$. 并且因为 $\frac{a}{b} < \frac{p}{q}$, 所以 $1 \le bp - aq$. 同理 $1 \le qx - py$. 所以 $b + y \le b(qx - py) + y(bp - aq) = q(bx - ay) = q$.

所以 $q = b + y$, 此时有 $bp - aq = qx - py = 1$, 所以 $a + x = a(qx - py) + x(bp - aq) = p(bx - ay) = p$. 综上, $p = a + x$, $q = b + y$. ■

从这个证明中可以看出, $\frac{a}{b}$, $\frac{p}{q}$, $\frac{x}{y}$ 一定是既约分数, 并且在 $\frac{a}{b}$ 与 $\frac{x}{y}$ 之间的分母等于 $b + y$ 的分数唯一.

从引理 1.1 的证明直接可得:

引理 1.2 *如果分数* $\frac{a}{b}$, $\frac{x}{y}$ *满足* $\frac{a}{b} < \frac{x}{y}$, *且* $ay + 1 = bx$. *设* $\frac{p}{q}$ *为满足* $\frac{a}{b} < \frac{p}{q} < \frac{x}{y}$ *的分母最小的分数, 则*

$$aq + 1 = bp, \quad py + 1 = qx$$

现在我们可以证明算法的正确性了.

定理 1.3 *如果分数* $\frac{a}{b}$, $\frac{x}{y}$ *满足* $b, y \le n$, $0 \le \frac{a}{b} < \frac{x}{y} \le 1$, *且* $ay + 1 = bx$, *那么* $\textsc{Build}(a, b, x, y)$ *按顺序生成了* $\mathcal{F}_n$ *中所有在* $\frac{a}{b}$, $\frac{x}{y}$ *之间的分数.*

证明 对 $\mathcal{F}_n$ 中 $\frac{a}{b}$ 与 $\frac{x}{y}$ 之间的项数归纳.

根据引理 1.1, 这时 $\frac{a}{b}$ 与 $\frac{x}{y}$ 之间的分母最小的分数为它们的中项 $\frac{p}{q}$.

若它们的中项的分母大于 n, 根据引理 1.1, 则这两个数之间无 $\mathcal{F}_n$ 中的项, 于是归纳奠基成立.

若它们的中项的分母不大于 n, 由归纳假设以及引理 1.2, $\textsc{Build}(a,b,x,y)$ 顺次输出了 $\mathcal{F}_n$ 中 $\frac{a}{b}$ 与 $\frac{p}{q}$ 之间的项, $\frac{p}{q}$, 以及 $\mathcal{F}_n$ 中 $\frac{p}{q}$ 与 $\frac{x}{y}$ 之间的项, 故 $\textsc{Build}(a,b,x,y)$ 按顺序输出了 $\mathcal{F}_n$ 中 $\frac{a}{b}$ 与 $\frac{x}{y}$ 之间的项. ■

时间复杂度分析 设 $T(l)$ 为正确调用了一次 $\textsc{Build}$ 过程, 并且这个过程输出了 l 个分数, 所花费的时间, 则

$$T(l) = \begin{cases} 1 & l = 0; \\ \max\{1 + T(x) + T(y) : x + y = l - 1\} & l > 0 \end{cases}$$

容易证得, $T(l) = 2l + 1$, 于是 $\textsc{Build}$ 过程的时间复杂度是 $O(|\mathcal{F}_n|)$ 的.

也可以这样理解 $\textsc{Build}$ 的过程, 它每次要么产生一个区间, 要么否定一个区间. 最初有一个区间, 最终有 $|\mathcal{F}_n| + 1$ 个区间, 并且所有的最终区间都要被否定一次, 也只有这些区间会被否定, 所以 $\textsc{Bulid}$ 所花总时间就是 $2|\mathcal{F}_n| + 1$.

至此, 本题已经完美解决. 但是我们遗留了一个问题: 证明 $\mathcal{F}_n$ 中任意连续三项的中间项为其两边两项的中项.

2 Farey 序列的两个性质

对上一节中的遗留问题以及 Farey 的构造过程进行简要分析, 可以发现, 阻碍这个问题解决的关键是 $\mathcal{F}_n$ 中并不是所有的分数都是由它两边两项求中项得到的. 即有可能会约分. 为此, 我们不妨换一个思路, 看看 $\mathcal{F}_n$ 还具有哪些性质.

性质 2.1 *如果 $\frac{a}{b}$, $\frac{x}{y}$ 是 $\mathcal{F}_n$ 中的相邻两项, 那么*

$$ay + 1 = bx$$

证明 根据 $\mathcal{F}_n$ 的构造过程易得. ■

推论 2.1 *$\mathcal{F}_n$ 中任意相邻两项之差大于等于 $\frac{1}{n(n-1)}$.*

证明 假设这两项顺次分别为 $\frac{a}{b}$ 和 $\frac{p}{q}$, 则

$$\frac{p}{q} - \frac{a}{b} = \frac{bp - aq}{bq} = \frac{1}{bq} \geq \frac{1}{n(n-1)}$$

故得结论. ■

性质 2.2 *如果 $\frac{a}{b}$, $\frac{p}{q}$, $\frac{x}{y}$ 是 $\mathcal{F}_n$ 中的连续三项, 那么*

$$\frac{p}{q} = \frac{a+x}{b+y}$$

证明一

$$\frac{p}{q} = \frac{a+x}{b+y} \iff p(b+y) = q(a+x) \iff py + 1 - qx = aq + 1 - bp$$

由 2.1 可知 $py + 1 - qx = aq + 1 - bp = 0$, 证毕. ■

事实上, 根据性质 2.1 证明过程的启发, 我们也可以从 $\mathcal{F}_n$ 的构造过程考虑, 得出另一种证明方法.

证明二 我们使用归纳法证明, 如果分数 $\frac{a}{b}$, $\frac{x}{y}$ 满足 $b, y \leq n$, $0 \leq \frac{a}{b} < \frac{x}{y} \leq 1$, 且 $ay + 1 = bx$, 那么 $\textsc{Build}(a,b,x,y)$ 产生的序列满足任意连续三项以及 $\frac{a}{b}$ 与前两项, 后两项与 $\frac{x}{y}$ 中, 中间项等于两边两项的中项.

奠基是显然的. 设 $\frac{a}{b}$ 与 $\frac{x}{y}$ 的中项为 $\frac{p}{q}$. 根据 $\textsc{Build}(a,b,p,q)$ 的过程, 可以知道 $\mathcal{F}_n$ 中 $\frac{p}{q}$ 的前项一定为 $\frac{a+up}{b+uq}$ 的形式. 同理, $\mathcal{F}_n$ 中 $\frac{p}{q}$ 的后项一定为 $\frac{x+vp}{y+vq}$ 的形式. 根据归纳法原理, 我们只需要再证明,

$\frac{p}{q}$ 是 $\frac{a+up}{b+uq}$ 与 $\frac{x+vp}{y+vq}$ 的中项. 因为 $p=a+x$, $q=b+y$, 根据中项的定义, $\frac{a+up}{b+uq}$ 与 $\frac{x+vp}{y+vq}$ 的中项为 $\frac{a+x+(u+v)p}{b+y+(u+v)q}=\frac{(u+v+1)p}{(u+v+1)q}=\frac{p}{q}$.

故由数学归纳法原理, 本性质得证. ∎

由性质 2.2 的这证明一便可以看出性质 2.1 的重要性. 事实上, 这个性质的应用不仅限于此. 在后文中, 我们将继续看到它的应用.

希望了解更多 Farey 序列性质的读者可以参考[1].

3 两分数间分母最小的分数

在使用引理 1.1 证明 $\textsc{Build}(a,b,x,y)$ 过程的正确性的时候, 我们其实想到了这样一个问题:

问题 满足 $\frac{a}{b}<\frac{p}{q}<\frac{x}{y}$ 的分母最小的分数 $\frac{p}{q}$ 是多少?

引理 1.1 对 $ay+1=bx$ 的情况给出了一个答案. 那么 $ay+1<bx$ 时又如何呢? 联系到 Farey 序列的性质 2.1, 立即得到:

引理 3.1 *如果既约分数 $\frac{a}{b}$, $\frac{x}{y}$ 满足 $\frac{a}{b}<\frac{x}{y}$, 且 $ay+1<bx$. 设 $\frac{p}{q}$ 为满足 $\frac{a}{b}<\frac{p}{q}<\frac{x}{y}$ 的分母最小的分数, 则*

$$q<\max\{b,y\},\quad p\le\max\{a,x\}$$

证明 不失一般性, 可以设 $0\le\frac{a}{b}<1$.

如果 $\frac{x}{y}\ge 1$ 或 $\frac{a}{b}=0$, 则命题显然成立.

当 $\frac{a}{b}$ 和 $\frac{x}{y}$ 都为真分数时, 根据性质 2.1, 在 $\mathcal{F}_{\max\{b,y\}}$ 中 $\frac{a}{b}$, $\frac{x}{y}$ 必不相邻. 又因为 $\frac{a}{b}$, $\frac{x}{y}$ 中分母较大者在它们之间的邻项的分母小于 $\max\{b,y\}$, 所以 $q<\max\{b,y\}$.

若 $p\ge\max\{a,x\}$, 则当 $q<y$ 时, $\frac{p}{q}\ge\frac{x}{q}>\frac{x}{y}$, 矛盾; 当 $q<b$ 时, $\frac{a}{b}<\frac{a}{q}\le\frac{p-1}{q}<\frac{p-1}{q-1}<\frac{p}{q}<\frac{x}{y}$, 与 q 的最小性矛盾. 故 $p\le\max\{a,x\}$. ∎

然而, 这个引理对我们似乎没有任何帮助. 看上去这个问题很难用 Farey 序列解决, 我们不妨换一种思路.

解法 由于给两个分数同时加上某个整数不会影响最终的结果. 设 k 为一整数, 且 $0\le a'=a+bk<b$. 令 $x'=x+yk$, 我们只需要求出 $\frac{a'}{b}$ 与 $\frac{x'}{y}$ 之间分母最小的分数 $\frac{p'}{q}$, 再令 $p=p'-qk$, 即求出了原问题的答案 $\frac{p}{q}$.

所以只用考虑 $0\le\frac{a}{b}<1$ 的情况. 当 $\frac{x}{y}>1$ 时, $p=1$, $q=1$. 所以只用考虑 $0\le\frac{a}{b}<\frac{x}{y}\le 1$ 的情况.

如果 $0=\frac{a}{b}$, 那么 $p=1$, $q=\lfloor\frac{y}{x}\rfloor+1$.

当 $0<\frac{a}{b}<\frac{x}{y}\le 1$ 时, 假设所求答案为 $\frac{p}{q}$. 首先, 容易证明, 分母最小的分数是唯一的; 其次, 又可以证明, 分母最小的分数也是分子最小的分数. 于是问题转化为求 $\frac{a}{b}$ 与 $\frac{x}{y}$ 间分子最小的分数 $\frac{p}{q}$, 也就是求出 $\frac{y}{x}$ 与 $\frac{b}{a}$ 间分母最大的分数 $\frac{q}{p}$. 如果存在一个整数 k , 满足 $\frac{y}{x}<k<\frac{b}{a}$, 那么 $q=\lfloor\frac{y}{x}\rfloor+1$. 否则, 必然有 $0\le\frac{y \bmod x}{x}<\frac{q \bmod p}{p}<\frac{b \bmod a}{a}<1$. 我们递归地求出 p, 根据分数的唯一性, 就可以求出 q, 也就求出了欲求分数.

这个过程实际上就是在我们熟知的求最大公约数的 Euclid 算法的变形, 所以这个算法的时间复杂度是 $O(\log\min\{a,x\})$ 的. 求最大公约数的 Euclid 算法时间复杂度的分析可以在[2]中找到.

4 简单连分数及最优渐进分数

4.1 连分数的基本概念与性质

定义 分数

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_3 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}}}$$

为有限连分数. 通常用 $[a_0, a_1, a_2, \ldots, a_n]$ 表示. 容易计算得

$$[a_0] = \frac{a_0}{1}, \quad [a_0, a_1] = \frac{a_0 a_1 + 1}{a_1}, \quad [a_0, a_1, a_2] = \frac{a_0 a_1 a_2 + a_2 + a_0}{a_1 a_2 + 1}$$

记

$$[a_0, a_1, \ldots, a_k] = \frac{p_k}{q_k}$$

称 $\frac{p_k}{q_k}$ 为 $[a_0, a_1, \ldots, a_n]$ 的第 k 个渐进分数.

定理 4.1 *各渐进分数有如下关系:*

$$p_0 = a_0, \quad p_1 = a_0 a_1 + 1, \quad p_k = a_k p_{k-1} + p_{k-2} \quad (2 \le k \le n)$$
$$q_0 = 1, \quad q_1 = a_1, \quad q_k = a_k q_{k-1} + q_{k-2} \quad (2 \le k \le n)$$

证明 用数学归纳法证明.

$n = 0,\ 1,\ 2$ 时, 显然成立.

假设对 $k \quad (k \ge 2)$ 成立, 则

$$\begin{aligned}
\frac{p_{k+1}}{q_{k+1}} &= [a_0, a_1, \ldots, a_k, a_{k+1}] = [a_0, a_1, \ldots, \frac{a_k a_{k+1} + 1}{a_{k+1}}] \\
&= \frac{p_{k-1}(\frac{a_k a_{k+1}+1}{a_{k+1}}) + p_{k-2}}{q_{k-1}(\frac{a_k a_{k+1}+1}{a_{k+1}}) + q_{k-2}} = \frac{a_{k+1}(a_k p_{k-1} + p_{k-2}) + p_{k-1}}{a_{k+1}(a_k q_{k-1} + q_{k-2}) + q_{k-1}} \\
&= \frac{a_{k+1} p_k + p_{k-1}}{a_{k+1} q_k + q_{k-1}}
\end{aligned}$$

即对 $k + 1$ 亦成立. ■

定理 4.2 *p_k 和 q_k 满足:*

$$p_k q_{k-1} - q_k p_{k-1} = (-1)^{k-1} \quad (k \ge 1), \tag{1}$$
$$p_k q_{k-2} - q_k p_{k-2} = (-1)^k a_k \quad (k \ge 2), \tag{2}$$

证明 (1) 当 $k = 1$ 时显然成立.

假设对 k 成立, 由定理 4.1 得

$$p_{k+1} q_k - q_{k+1} p_k = (a_{k+1} p_k + p_{k-1}) q_k - (a_{k+1} q_k + q_{k-1}) p_k = -1(p_k q_{k-1} - q_k p_{k-1}) = (-1)^k.$$

(2) 由定理 4.1 及 (1) 可得

$$\begin{aligned}
p_k q_{k-2} - q_k p_{k-2} &= (a_k p_{k-1} + p_{k-2}) q_{k-2} - (a_k q_{k-1} + q_{k-2}) p_{k-2} \\
&= a_k (p_{k-1} q_{k-2} - q_{k-1} p_{k-2}) = (-1)^k a_k.
\end{aligned}$$

故此定理成立 ■

定义 若 a_0 为整数, $a_1, a_2, \ldots$ 均为正整数, 则称 $[a_0, a_1, \ldots]$ 为简单连分数. 以下讨论的均为简单连分数.

定理 4.3

$$\frac{p_{2k}}{q_{2k}} < \frac{p_{2k+2}}{q_{2k+2}}, \quad \frac{p_{2k-1}}{q_{2k-1}} > \frac{p_{2k+1}}{q_{2k+1}}. \tag{1}$$

各渐进分数均为既约分数. (2)

证明 (1) 根据定理 4.2 即证.

(2) 根据定理 4.2 及引理 1.1 即证. ■

4.2 求有理数的简单连分数表示

假设这个有理数为 x, 它的连分数表示为 $[a_0, a_1, \dots, a_n]$, 我们发现如下性质: 若 x 不为整数, 则 $n \geq 1$, 并且

$$a_0 < x < a_0 + 1.$$

故 a_0 唯一. 且容易发现

$$1 < \frac{1}{x - a_0} = [a_1, \dots, a_n].$$

若 x 为整数, 则它有两种简单连分数表示: $[x]$ 和 $[x-1, 1]$.

设 $b_k = [a_k, \dots, a_n] \quad (0 \leq k \leq n)$, 则有:

$$b_k = \begin{cases} x & (k = 0); \\ \dfrac{1}{b_{k-1} - a_k} & (1 \leq k \leq n) \end{cases}$$

及

$$b_k > 1 \quad (1 \leq k \leq n).$$

假设 $x = \dfrac{a}{b}$, 则 x 的简单连分数表示可以这样求得:

如果 x 是整数, 那么返回 $[x]$ 和 $[x-1, 1]$.

如果 x 不是整数, 那么设 $r = a - b\lfloor\dfrac{a}{b}\rfloor$, 求出 $\dfrac{b}{r}$ 的简单连分数表示 $[a_1, \dots, a_n]$, 则 x 的简单连分数表示为 $[\lfloor\dfrac{a}{b}\rfloor, a1, \dots, a_n]$.

容易发现, 这个算法就是 Euclid 算法的变形. 同样, 这个算法的时间复杂度是 $O(\log r)$.

进一步分析这个算法, 我们发现它和前面讨论的求两分数间分母最小的分数是十分类似的. 读者可以从连分数的角度分析上文所述求两分数间分母最小的分数的算法, 以加深对此过程的理解, 这里不再赘述.

4.3 最优渐进分数

看了这么多概念, 不妨休息一下, 看看这个有趣的问题.

用上节的算法可以求出, π 的前几个渐进分数是:

$$3, \frac{22}{7}, \frac{333}{106}, \frac{355}{133}$$

这些分数我们再熟悉不过了. 为了说明它们的精确程度, 我们来看这样一个问题:

求证 在所有分母不大于 106 的分数中, $\dfrac{333}{106}$ 与 π 最接近.

证明 首先, 计算可得 $\dfrac{333}{106} < \pi < \dfrac{22}{7}$, 并且 $333 \times 7 + 1 = 106 \times 22$, 根据引理 1.1, 在 $\dfrac{333}{106}$ 与 $\dfrac{22}{7}$ 之间的分数的分母不小于 113. 又因为 $\dfrac{22}{7} - \pi > \pi - \dfrac{333}{106}$, 所以任何比 $\dfrac{333}{106}$ 更接近 π 的分数的分母都大于 113. 故所有分母不大于 106 的分数中, $\dfrac{333}{106}$ 与 π 最接近. ■

这个证明出人意料地使用了一个辅助的分数 $\dfrac{22}{7}$, 这个方法实在难以想到. 这个方法提示我们要联系地看 Farey 序列和渐进分数. 事实上, 这个方法就是这种思想的产物.

这个证明也可以引发这样的思考, 在 $\dfrac{333}{106}$ 与 $\dfrac{22}{7}$ 之间的所有分数中, 有哪些数不是任何分数的分母, 容易知到, 这些数中存在最大值(答案是 $106 \times 7 - 106 - 7$). 这个问题比较简单, 读者不妨自己解决.

定义 称一个分数 $\dfrac{p}{q}$ 是一个数 x 的最优渐进分数, 当且仅当在分母不大于 q 的分数中, $\left|\dfrac{p}{q} - x\right|$ 最小.

有了解决这个问题的经验, 相信读者已经可以自己证明这个定理了:

定理 4.4 *x 的所有渐进分数均为最优渐进分数.*

值得一提的是, 连分数与 Pell 方程有着密切的联系. 希望更进一步了解连分数的读者请参考[1]中的相关章节.

5 Farey 序列上的一些运算

在 Farey 序列上, 我们能做的不只是求出 $\mathcal{F}_n$, 并且 Farey 的应用也不仅限于它的这些操作, 下一节将介绍它的一个应用, 求出分母不大于一给定值的与给定分数最接近的分数.

5.1 求前(后)项

设 $\frac{a}{b}$ 为 $\mathcal{F}_n$ 中的某一项, 那么如何求出它的前(后)项呢?

设 $\frac{a}{b}$ 的前项为 $\frac{p}{q}$. 回顾 Farey 的几个性质, 我们就能发现, $\frac{p}{q}$ 一定满足 $pb+1=aq$. 因为 $(a,b)=1$, 所以这个方程一定有解, 并且可以在 $O(\log a)$ 的时间内求出一组最小正解 (p_0,q_0), 并且容易证明, 所有的解都可以表示为 (p_0+ka, q_0+kb) 的形式. 于是所有可能作为 $\frac{a}{b}$ 在 $\mathcal{F}_n$ 中的前项的数就是 $\{\frac{p_0+ka}{q_0+kb} : q_0+kb \le n\}$. 容易证明, $\frac{p_0+ka}{q_0+kb}$ 是关于 k 单调递增的. 设 $k=\max\{k : q_0+kb \le n\}$, 则 $\frac{a}{b}$ 在 $\mathcal{F}_n$ 中的前项就是 $\frac{p_0+ka}{q_0+kb}$.

作类似的讨论即可求出 $\frac{a}{b}$ 在 $\mathcal{F}_n$ 中的后项.

显然, 本操作的时间复杂度为 $O(\log a)$.

5.2 求前(后) k 项

直接求前(后)项的运算 k 次, 就可以求出一个分数在 $\mathcal{F}_n$ 的后 k 项. 这个方法的时间复杂度是 $O(k\log n)$, 其实我们可以进一步利用 Farey 序列中连续项之间的关系来得到一个更快的算法.

联系到的性质 2.2, 容易发现, 如果 $\frac{a}{b}$ 的后项是 $\frac{p}{q}$. 假设 $\frac{p}{q}$ 的后项是 $\frac{x}{y}$, 则必有 $x+a=mp$, $y+b=mq$. 于是, $y=\max\{mq-b : mq-b \le n\}$. 显然, x 和 y都可以在 $O(1)$ 时间内求得. 如此继续, 便可以在 $O(\log a+k)$ 时间内求出 $\frac{a}{b}$ 在 $\mathcal{F}_n$ 中的后 k 项.

$\frac{a}{b}$ 在 $\mathcal{F}_n$ 中的前 k 项可以用类似的方法求出.

5.3 Farey 序列的另一种构造方法

利用 5.2 节的方法, 容易得出另一个 $O(|\mathcal{F}_n|)$ 的方法.

显然, 当 $n \ge 3$ 时, $\mathcal{F}_n$ 的前两项为 $\frac{1}{n}, \frac{1}{n-1}$. 按照上小节所用方法, 就可以在每次 $O(1)$ 的时间内求出它的下一项, 直到到达 $\mathcal{F}_n$ 的最后一项.

5.4 $\mathcal{F}_n$ 中不大于一给定实数的项数

当 n 不大的时候, 我们可以通过顺次求出 $\mathcal{F}_n$ 的项, 至到找到一项大于这个实数. 但是, 这个方法的复杂度是 $O(n^2)$ 的, 因而效率低下.

为此, 我们考虑, 对所有分母的分数, 求出不大于这个实数(设为 x)的项的个数. 设 $t_i=\lfloor ix\rfloor$, 则分母为 i 的不大于 x 的既约分数的个数为小于等于 t_i 的且与 i 互质的数的个数 c_i. 联系到求 $\phi(i)$ 的方法, 容易使用容斥原理得到:

$$
\begin{aligned}
c_i &= \sum_{d=\prod_{j=1}^{k} p_{i_j},\ d|i} (-1)^k \lfloor \frac{t_i}{d} \rfloor \\
&= \sum_{d=\prod_{j=1}^{k} p_{i_j},\ d|i} (-1)^k \lfloor \frac{i}{d} x \rfloor
\end{aligned}
$$

其中, p_k 为质数.

那么, $\mathcal{F}_n$ 中不大于 x 的项数就等于

$$
\begin{aligned}
\sum_{i=1}^{n} c_i &= \sum_{i=1}^{n} \sum_{d=\prod_{j=1}^{k} p_{i_j},\ d|i} (-1)^k \lfloor \frac{i}{d} x \rfloor \\
&= \sum_{d=\prod_{j=1}^{k} p_{i_j}} (-1)^k \sum_{i=1}^{\lfloor n/d \rfloor} \lfloor ix \rfloor.
\end{aligned}
$$

我们只需要先预处理出来所有的符合条件的 d(就是用筛法求一次素数), 然后用 $O(n)$ 时间计算出 $s_i = \sum_{k=1}^{i} \lfloor kx \rfloor \quad (1 \le i \le n)$, 再用 $O(n)$ 的时间即可算出答案.

5.5 求第 k 项

直接利用 5.2 节介绍的方法就可以在 $O(k)$ 的时间内求出 $\mathcal{F}_n$ 的第 k 项. 但是当 k 很大的时候, 这个方法就很慢了. 其原因在于, 虽然只需要求出第 k 项, 但它却把前 k 项统统求了出来.

这时我们自然而然地想到了二分的方法. 联系到推论 2.1, 假设 $f(i) \quad (1 \le i \le n^2)$ 为 $\mathcal{F}_k$ 中不大于 $\frac{i}{n^2}$ 的项数. 则 $f(i+1) - f(i) \le 1$. 我们只需要在这些值中二分, 求出一个满足 $f(i) = k$ 的 i , 然后再求出 $\mathcal{F}_n$ 中小于等于 $\frac{i}{n^2}$ 的最大分数. 先用 $O(n \log\log n)$ 的时间作计算 $f(i)$ 的预处理, 由于每次计算 $f(i)$ 的时间复杂度为 $O(n)$, 二分的次数为 $O(\log n)$, 最后一次求最大分数时, 只要枚举所有分母即可(是否存在更好方法?). 所以本操作的时间复杂度为 $O(n \log n)$

6 对一些最优渐近分数的讨论

6.1 分母不大于一给定值的与给定分数最接近的分数

假设需要求出分母不大于 k 的, 与 $\frac{x}{y}$ 最接近的分数. 这个问题在 $k \ge y$ 时是平凡的, 因为所求分数就是 $\frac{x}{y}$.

相信读者看到这个问题的第一反应就是求其简单连分数表示, 并求出对应的渐进分数序列. 然后在这个序列中找到分母不大于 k 的分母最大的分数. 然而, 连分数对应的几个渐进分数虽然都是最优渐进分数, 但却没有包括所有最优渐进分数. 例如, 求与 $\frac{17}{19}$ 最接近的, 分母不大于 6 的分数. $\frac{17}{19} = [0, 1, 8, 2]$, 它的渐进分数为

$$0, 1, \frac{8}{9}, \frac{17}{19}$$

用这种方法我们求出的答案显然不正确.

先将此任务分为两部分, 求出小于 $\frac{x}{y}$ 的分母不大于 k 的最大分数, 再求出大于 $\frac{x}{y}$ 的分母不大于 k 的最小分数. 由于这两个问题的对称性, 我们只需要考虑如何求出小于 $\frac{x}{y}$ 的分母不大于 k 的最大分数. 事实上, 当 $\frac{x}{y}$ 的分母不大于 k 的最大分数求出以后, 我们根本不用再用相同的方法求解大于 $\frac{x}{y}$ 的分母不大于 k 的最小分数, 因为这个数一定是刚才求出的分数在 $\mathcal{F}_k$ 中的后项!

6.2 分母不大于一给定值的小于给定分数的最大分数

基于对最优渐进分数的考虑, 容易想到这种方法:

定义 称一个分数 $\frac{p}{q}$ 是一个数 x 的最优左渐进分数, 当且仅当在所有分母不大于 q 的小于 x 的分数中, $x - \frac{p}{q}$ 最小.

于是求出所有最优左渐进分数, 再在这里找分母最大的分数. 显然, 分数 $\frac{x}{y}$ 的分母(小于 y)最大的最优左渐进分数就是它在 $\mathcal{F}_n$ 中的前项(设为 $\frac{p}{q}$). 分母第二大的最优左渐进分数就是 $\frac{p}{q}$ 在 $\mathcal{F}_q$ 中的前项 $\frac{a}{b}, \ldots$

我们先抛开最优左渐进分数的总数不管, 考虑如何尽快求出这些最优左渐进分数. 显然, 如果每次求 Farey 前项, 时间复杂度就是 $O(m \log x)$ 的, 其中 m 为需要求出的最优左渐进分数的个数. 在 5.2 节中, 我们已经知道了, Farey 中连续的若干项之间是有联系的, 即使是这种若干个前项并非同一个 Farey 序列中的连续项的情况. 因为 $\mathcal{F}_n$ 中的 n, 只是用来求出最合适的分母的. 于是我们可以得到一个 $O(m + \log x)$ 的解法.

可是, m 究竟有多大呢? 考虑这个分数: $\frac{17}{19}$. 我们求出的最优左渐进分数序列就是:

$$\frac{17}{19}, \frac{8}{9}, \frac{7}{8}, \frac{6}{7}, \frac{5}{6}, \frac{4}{5}, \frac{3}{4}, \frac{2}{3}, \frac{1}{2}, \frac{0}{1}$$

这个最优左渐进分数序列的长度是 $O(y)$ 的! 这样做还不如直接枚举分母.

但是我们忽略了最重要的一点, 相邻的项之间是有联系的. 在这个更为特殊的连续求 Farey 前项的过程中, 我们只需要求出 $py+1=qx$ 的最小正解 (p,q), 就求出了 $\frac{x}{y}$ 在 $\mathcal{F}_y$ 的前项 $\frac{p}{q}$. 此时如果 $2q>y(2q>y \iff 2p>x$, 想想为什么?), 那么 $\frac{p}{q}$ 在 $\mathcal{F}_q$ 中的前项一定是 $\frac{2p-x}{2q-y}$, 也就是它们的分母和分子都成等差数列. 这是因为最小正解是唯一的.

我们考虑将这些小于 $\frac{x}{y}$ 的最优左渐进分数分成若干段, 如果一个两个连续的分数 $\frac{x}{y}$ 和 $\frac{p}{q}$, 并不满足 $2q>y$, 那么 $\frac{p}{q}$ 的下一个分数就应当和它们被划分到不同的两段. 根据这个分段规则, 可以发现, 每一段的分母和分子是成等差数列的.

于是我们可以轻松的求得每一段的最后一个分数. 如果它的分母不大于 k, 那么所求的分数一定在这一段中, 我们可以用 $O(1)$ 的时间求出. 如果它的分母大于 k, 那么可以在 $O(1)$ 时间内求出它的 Farey 后项, 以及下一段的最后一项. 如此下去直到某一段的最后一项的分母不大于 k.

假设在停止之前一共计算了 m 段的最后一项, 那么此过程的时间复杂度是 $O(m+\log x)$ 的. 那么, 这回 m 有多大呢? 我们考虑最初的一段, 假设 $\frac{x}{y}$ 在 $\mathcal{F}_y$ 中的前项的分母为 $y-b$, 那么这一段的最后一项的分母应当是 $y \bmod b$. 然而复杂度分析的一个难点是: 这里 b 是不确定的. 但是, 我们可以预感到, 这里 $y \bmod b$ 一定不会大! 因为如果 b 很小, 那么 $y \bmod b<b$, 也很小. 如果 b 很大, 那么 $y \bmod b<y-b$, 也会很小. 于是, 我们发现了一个令人欣喜的结论:

定理 6.1 *如果 $b \le y$, 那么 $y \bmod b<\frac{y}{2}$.*

证明 因为 $b\le y$, $y \bmod b=(y-b) \bmod b$, 所以 $(y-b) \bmod b<y-b$. 而$(y-b) \bmod b<b$. 所以 $y \bmod b<\min\{b,y-b\}\le \frac{b+y-b}{2}=\frac{y}{2}$. ■

这个看似平凡的, 不证自明的结论, 却是本题解决的关键! 由于每次分母至多是原来的一半, 所以至多 $\lg y$ 次以后, 分母就会不大于 1, 也就不大于 k. 所以这些最优左渐进分数中至多存在 $\lg y$ 段. 根据上面的复杂度分析, 这个方法的时间复杂度是 $O(\log y)$ 的. 我们完美地解决了此问题! 但还有更一般的问题等着我们解答.

6.3 分母不大于一给定值的与给定实数最接近的分数

上面的方法对实数则无法奏效. 我们试图通过根据简单连分数表示的渐进分数的最优性, 找到第一个分母大于等于 k 的渐近分数. 由于这个渐进分数是最优的, 所以找小于等于 x 的分母不大于 k 的最大分数等价与找分母不大于 k 的小于等于这个渐进分数的最大分数. 显然, 根据求简单连分数的过程的时间复杂度分析可知, 这个步骤的时间复杂度是 $O(\log k)$ 的. 但是, 这样求出来的渐进分数的分母会多大呢? 略经思考可以发现, 这个数的分母甚至可以无限大! 我们要么对上面的求分母不大于 k 的小于给定分数的最大分数的方法加以改进, 要么就需要放弃使用连分数的方法. 然而, 我们也不得不放弃使用连分数的解法, 因为这个分母会大到难以保存.

不过我们的这个出发点是好的, 找到一个接近 x 的分数 $\frac{p}{q}$, 求出小于它的分母不大于 k 的最大分数. 其实我们完全可以跳出上面的思维定势, 我们不需要这两个问题具有相同的答案, 只需要它们接近! 而这个接近的意思是, 要么这两个问题具有相同的答案, 要么它们在 $\mathcal{F}_k$ 中是相邻两项.

具体地说, 我们可以让这个分数满足: $q=k^2$, $\frac{p}{q}\le x<\frac{p+1}{q}$. 这样, 根据推论 2.1, 在 $[\frac{p}{q},x]$ 中, 至多有一个 $\mathcal{F}_k$ 中的项. 所以我们只需要求出小于 $\frac{p}{q}$ 的分母不大于 k 的最大分数, 并求出它的后项, 如果它的后项小于等于 x, 那么它的后项就是小于等于 x 的分母不大于 k 的最大分数, 否则这个最大分数就是它本身. 再求出小于等于 x 的分母不大于 k 的最大分数在 $\mathcal{F}_k$ 中的后项, 在这两个分数中取距离 x 最小的一个, 则为所求答案. 显然, 这个方法的时间复杂度为 $O(\log k)$.

从这个更加一般性的问题中我们也可以发现, 在上面的问题中, 如果 y 过大, 那么也可以用这个方法来解决. 于是上个问题可以得到时间复杂度为 $O(\log k)$ 的解法.

7 再看两分数间分母最小的分数

看到这里, 读者明白第 6 节的真正用意了吧? 在第 3 节中, 我们换了一种思路, 得到了利用连分数解决两分数间分母最小的分数的问题. 但是, 真的不能利用 Farey 序列的相关方法么? 第 6 节就是对 Farey 序列的相关方法是否可以解决这类问题的一个尝试. 这个尝试促使我们向最优左渐进分数的方向考虑两分数间分母最小分数的问题. 我们求得的分数, 一定是某个最优左渐进分数么? 可是如果较大的分数分母小怎么办? 例如 $\frac{1}{7}<\frac{1}{6}<\frac{1}{5}$. 可是分别将这两个分数取相反数又显得麻烦. 引理 3.1 中的 max 反而成为一个思维陷阱!

7.1 大于一给定实数小于一给定分数的分母最小的分数

此问题是要求出分母最小的在 $(x, \frac{a}{b})$ 上的分数. 但直接考虑这个问题会遇到上述麻烦. 假设我们已经求出了分母最小的分数 $\frac{p}{q}$, 满足 $x < \frac{p}{q} \leq \frac{a}{b}$, 那么 $\frac{p}{q}$ 一定是 $\frac{a}{b}$ 的某个最优左渐进分数! 如果不是, 那么存在一个更接近 $\frac{a}{b}$ 的(于是一定大于 x)且分母更小的分数, 与 $\frac{p}{q}$ 的最优性矛盾. 如果求出的 $\frac{p}{q} \neq \frac{a}{b}$, 那么这个 $\frac{p}{q}$ 就是原问题的答案. 如果 $\frac{p}{q} = \frac{a}{b}$ 呢? 我们先看看如何求出这个 $\frac{p}{q}$.

对第 6.2 节中介绍的方法稍加改造, 就可以在 $O(\log b)$ 的时间内求出这个 $\frac{p}{q}$. 如果 $\frac{p}{q} = \frac{a}{b}$, 令 $\frac{p}{q}$ 为 $\frac{a}{b}$ 在 $\mathcal{F}_b$ 中的前项, 那么说明 $\frac{p}{q} \leq x$. 我们求出最小的 k, 满足 $x < \frac{p+ka}{q+kb}$, 则答案就是这个分数! 因为 $\frac{p+(k-1)a}{q+(k-1)b} \leq x < \frac{p+ka}{q+kb} < \frac{a}{b}$, 根据引理 1.1, $(x, \frac{a}{b}]$ 间分母最小的分数就是 $\frac{p+ka}{q+kb}$.

这样就在 $O(\log b)$ 的时间内解决了此问题. 由于问题的对称性, 我们也可以在 $O(\log b)$ 的时间内求出 $(\frac{a}{b}, y)$ 上的分母最小的分数.

7.2 两实数间分母最小的分数

7.1 节介绍的算法带来的好处是: 我们只需要先找到一个可行解, 就可以在很短时间内求出最优解. 假设要求出 (x, y) 上的分母最小的分数, 我们先求出一个在 (x, y) 上的分数 $\frac{a}{b}$, 就可以分别在 $O(\log b)$ 的时间内求出 $(x, \frac{a}{b}]$ 的分母最小的分数和 $[\frac{a}{b}, y)$ 上分母最小的分数, 在这里求出最好的, 就是本问题的答案.

容易求出一个满足条件的 $\frac{a}{b}$, 因为只要 $\frac{1}{b} < y - x$, 那么就一定存在一个分母为 b 的在 (x, y) 上的分数 $\frac{a}{b}$. 这样, 我们就得到了一个时间复杂度为 $O(\log \frac{1}{y-x})$ 的算法. 然而, 当 $x = 0.49999999999$, $y = 0.50000000001$ 时, 这个方法会不会很慢呢? 虽然看起来是这样, 但实际上, 这个算法的仅进行了若干次代数运算, 远小于 $\log \frac{1}{y-x}$. 这个算法是否存在一个更精确的上界(与答案的分母有关的上界)呢? 由于本人水平有限, 这个问题只能留给读者思考了.

8 总结

从上述讨论中可以看出: Euclid 算法是数论问题中的基本, 也是非常高效的方法. 很多其它用途的算法, 例如求简单连分数表示, 和两分数间分母最小的分数, 都是由此衍生而来; 最大限度利用 Farey 序列连续项的性质则是解决另一些分数问题的关键.

参考文献

[1] 华罗庚, *《数论导引》*

[2] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein, *Introduction to Algorithms, Second Edition*