NOIP2023 模拟赛 R1 题解

By YeahPotato

2023.9

T1

题意

一排 n 个点，每次可以从开头向右边不断跳。求最少几轮（几个原子）可以把所有的区间跳到。$n \le 2000$。

原题中讲的是从 n 开始，下文记为从 1 开始。将单步跳跃（跃迁）$i \to j$ 记作区间 $[i, j)$。

idea from zrz(wishapig).

算法零

打表、搜索或随机化。期望得分 $20 \sim 40$。

算法一

先尝试把最少轮数算出来。

一个显然的事实是，如果 $[a, b) \cap [c, d) \ne \varnothing$，那么它们不能在一轮内同时解决。

因此考虑找出一堆区间，它们两两有非空交，最大化区间的数量。这样能较好地 bound 一个下界。设找出的这些区间形成集合 S。

首先贪心地考虑。可以证明：

1. 若 $I \in S$，则 $\forall I' \supseteq I, I' \in S$。
2. 若 $I, I' \in S$，则 $I \cap I' \in S$。

这两条性质决定了最优的 S 一定是形如 $\{[i, j) \mid i \le a, j \ge b\}$ 的结构，其中 $a < b$。

由均值不等式，取 $a = \lfloor n/2 \rfloor, b = a + 1$，可以达到 $|S|_{\max} = \lfloor n/2 \rfloor \cdot \lceil n/2 \rceil$。当然找规律也行。

事实上，如果将每个区间视作一个节点，连有向边 $[i, j) \to [j, k)$（$i < j < k$），那么原问题就是一个最小可重链覆盖，而上文中的 S 就是最长反链，由 Dilworth 定理两者相等。

但是本题不是必须用该定理。只需构造出一组解就可以证明答案恰好为 $\lfloor n/2 \rfloor \cdot \lceil n/2 \rceil$ 了。

算法二

建二分图跑最大流。直接连边数量为 $\Theta(n^3)$，要做到 $\Theta(n^2)$，可以对于每个 j 建立中介点 t_j，所有左部 $[i, j)$ 连它，它连右部 $[j, k)$。

或者将每个点（能级）视作节点，直接连边，跑有源汇上下界最小流即可。

不推荐写。期望得分 76。

算法三

直接构造。先把 S 中的那些区间取出来，然后将 $1 \le i < j \le \lfloor n/2 \rfloor$ 的 $[i, j)$ 和 $\lfloor n/2 \rfloor < i < j \le n$ 的 $[i, j)$ 分别“依附”到 S 中的区间的前后即可。由于 S 中每种左/右端点都对应着 $\ge \lfloor n/2 \rfloor$ 个右/左端点，因此只需对于 $[i, j) \in S$ 选择某尚覆盖的 $[x, i)$ 和 $[j, y)$，跳 $x \to i \to j \to y$ 就行了（x/y 不够用就不管），一定覆盖得全。期望得分 100。

T2

题意

给定一张类似于树的图，每条边都是多重边。求删去尽量少的边，使新图存在一条欧拉路径。$n \le 10^6$。

算法一

这里不把重边的每一条分开来看。对于有边的一对点，称 w 为偶数的为“偶重边”，否则为“奇重边”，特殊地，$w = 1$ 的奇重边称为单重边。

考虑刻画最终解的结构。先考虑 $w \ge 2$ 情况。

看到欧拉路径的第一反应是至多两个点度数为奇。但这题不能看点的度数，要考虑边。假设欧拉路径的端点为 s, t，那么对于 $s \rightsquigarrow t$（s 到 t 的简单路径）上的偶重边需要减一重，因为开始是在这条偶重边的一侧，结束时在另一侧，一定经过了奇数次。反之，不在上面的奇重边需要减一重。这样减完之后一定存在欧拉路径，并且这样一定是最优的。

若存在 $w = 1$，考虑把所有单重边断掉，原图变为若干连通块，简称块。所有 $s \rightsquigarrow t$ 经过的块包括路上的单重边都按原方法算，其余块是根本无法进入的，因为进入了就出不来了。

因此问题就在于选端点，不要想复杂。这题可以用类似求直径的方法，记录当前点的最大、次大直链（从子树中某个点出发到当前点，路径上的偶重边减一重，经过的所有块的非路径上的奇重边减一重后，除当前块外经过的所有块的边数和）即可。

时间复杂度 $\mathrm{O}(\sum n)$，期望得分 100。

T3

题意

求一个序列的所有本质不同子序列的本质不同子序列个数之和 $\bmod 1000000007$。$n\le 5000$。

算法一

计算一个序列的本质不同子序列个数的方法如下：

由于会计重，考虑对于每一种子序列，钦定只对它最靠左（贪心）的匹配计数。容易证明存在唯一的这样的子序列到下标的映射。

令 dp_i 表示末尾选 i 的本质不同子序列数，则 $dp_0=1,dp_i=\sum_{j=pre_i}^{i-1}dp_j$，$pre_i$ 表示上一个与 a_i 相等的位置，如果不存在则为 0。

或者，令 dp_i 表示末尾选 $\le i$ 的本质不同子序列数，则 $dp_0=1,dp_i=2dp_{i-1}-dp_{pre_i-1}$。

外层暴枚可做到 $\mathrm{O}(n2^n)$，期望得分 30。

算法二

直接讲正解。

设选择的外层子序列下标为集合 I，内层为集合 $J\subseteq I$。为了方便表述，设占位下标 $0\in I,J$。同样只计贪心匹配的情况，限制如下：

1. I 中相邻两个数 i,i'，$a_{i+1\sim i'-1}$ 中不存在 $=a_{i'}$ 的值。
2. J 中相邻两个数 j,j'，$a_{I\cap(j,j')}$ 中不存在 $=a_{j'}$ 的值。

考虑对 J dp。f_i 表示目前考虑到 i 且内外层末尾均选 i 的答案。如果要从 f_j 转移过来，那么就要决定 $a_{j+1\sim i-1}$ 这部分如何选外层，设选择了集合 K，限制如下；

1. K 中相邻两个数 k,k'，$a_{k+1\sim k'-1}$ 中不存在 $=a_{k'}$ 的值。
2. K 中最大值 k_r，$a_{k_r+1\sim i-1}$ 中不存在 $=a_i$ 的值。
3. K 中任意 k，$a_k\ne a_i$。

一个简洁的处理方法是，对于每一个 i，dp 出 $>$ 每个 j 的只需满足 1、3 条件的本质不同子序列个数 $g_{i,j}$，真正转移时 $f_i\xleftarrow{+}(g_{i,j}-g_{pre_i,j})\cdot f_j$ 即可。最后汇总答案可以弄一个必选的占位下标 $n+1$。

g 是 2D/0D，f 是 1D/1D，时间复杂度 $\mathrm{O}(n^2)$，期望得分 100。

如果您想到了低于平方的解法，请联系 YeahPotato！

T4

- ### 题意

对于初始为 1 的变量 m，q 次每次乘或除以一个质数 p，求 A111725(m)。$q \le 5 \times 10^5, p \le 10^7$。

- ### 算法一

求阶可以不用暴力枚举，而是使用试除法：已知 $x^{\varphi(m)} \equiv 1 \pmod m$，每次尝试将指数除以一个质因子。这样可以做到 $\mathrm{O}(qm \log^2 m)$，期望得分 20。

- ### 算法二

将 m 质因数分解为 $\prod_{i=1}^{k} p_i^{c_i}$，则 $\delta_m(x) = \mathrm{lcm}_{i=1}^{k}(\delta_{p_i^{c_i}}(x))$。于是可以确定模 m 的最大阶：A002322，记为 $\psi(m)$。其中 $\psi(2) = 1, \psi(4) = 2, \psi(2^c) = 2^{c-2}(c \ge 3), \psi(p^c) = (p-1)p^{c-1}(p \ge 3, p \in \mathbb{P})$。

> 定理：对于奇质数 p，模 p^c 阶为 d（$d \mid \varphi(p^c)$）的剩余类有 $\varphi(d)$ 个；对于 $c \ge 3$，模 2^c 阶为 d（$d \mid 2^{c-2}$）的剩余类，$d = 1$ 有 1 个，$d = 2$ 有 3 个，$d \ge 4$ 有 d 个。

证明可以考虑取原根，2 的次幂可以考虑分模 4 同余 1 和 3 的讨论。

因此可以暴搜模每个 $p_i^{c_i}$ 的阶 $\{d_i\}$，判断其 lcm 是否为 $\psi(m)$ 即可。期望得分 30。

- ### 算法三

测试点 6，设 $m = p^c$：

$$ans = \begin{cases} 1 & m \le 4 \\ 3 & m = 8 \\ \varphi(\varphi(m)) & \text{otherwise} \end{cases}$$

期望得分 $+10$。

- ### 算法四

下文中复杂度里的 p 都表示最极端情况的 p。

设定义在质数幂上的校正函数 $C_{m,p_i}(p^k) = \begin{cases} 2p^k & p_i = p = 2 \land k \ge 1 \land 8 \mid m \\ p^k & \text{otherwise} \end{cases}$，则模 m 意义下方程 $x^d \equiv 1$ 的解数为 $f(d) = \prod_{i=1}^{k} C_{m,p_i}(\gcd(\psi(p_i^{c_i}), d))$。原问题的答案为 $(f * \mu)(\psi(m))$。

这里注意，只有 $p_i = 2$ 时才需要校正，奇质数 p_i 出现 $2 \mid \psi(p_i^{c_i})$ 的情况是不需要校正的。

对于固定的 m，f 为积性函数，故 $f * \mu$ 也是积性函数，故：

$$ans = \prod_{p^c \| \psi(m), p \in \mathbb{P}} \left(\prod_{i=1}^{k} C_{m,p_i}(\gcd(\psi(p_i^{c_i}), p^c)) - \prod_{i=1}^{k} C_{m,p_i}(\gcd(\psi(p_i^{c_i}), p^{c-1})) \right)$$

gcd 里有两种情况：

1. $v_p(\psi(p_i^{c_i})) < c$；
2. $v_p(\psi(p_i^{c_i})) = c$。

其中第二种情况至少出现一次。考虑先都按第一种情况算，然后一部分校正成第二种。一个关键的观察是：

$$\prod_{p^c \| \psi(m), p \in \mathbb{P}} \prod_{i=1}^{k} C_{m,p_i}(\gcd(\psi(p_i^{c_i}), p^c)) = \varphi(m)$$

直观上是好理解的，就是 2 的次幂的部分要仔细讨论一下。因此：

$$ans = \varphi(m) \cdot \prod_{p|\psi(m), p\in\mathbb{P}} \frac{p^{g_m(p)} - 1}{p^{g_m(p)}}$$

其中 g 表示第二种情况出现的次数，具体来说：

$$g_m(p) = [p = 2 \wedge v_2(\psi(m)) = 1 \wedge v_2(m) = 3] + \sum_{i=1}^{k} [v_p(\psi(p_i^{c_i})) = v_p(\psi(m))]$$

开头艾弗森括号里这个奇怪的特判实质上就是 $p_i^{c_i} = 2^3, p^c = 2^1$ 的时候 $C_{m,2}(2^1) = 4, C_{m,1}(2^0) = 1$ 的情况。

因此只需对于每个质数 p 维护 $\max_{i=1}^{k} \{v_p(\psi(p_i^{c_i}))\}$ 、取到 $\max$ 的数量 g 以及 $\frac{p^g-1}{p^g} = 1 - p^{-g}$ 即可。

对于有删的情况，线段树分治可能可以卡过。正解：暴力移动 $\max$ 指针的时间复杂度是正确的，因为对于 $p^c \parallel (p_i - 1)$，p_i 删去时移动量至多为 c，而质因数分解的指数之和为 $\mathrm{O}(\log p)$，故总共是 $\mathrm{O}(q \log p)$ 的。

出题人的实现方法如下：离线分别对于每个质数处理。对于质数 p，记录使得某个 $v_p(\psi(p_i^{c_i}))$ 变动的修改，一共有两类：

1. 某个 $p \mid (p_i - 1)$ 的 p_i 一次方加入或删除。
2. p 本身至少二次方的加入或删除。

其中第二种不需要把原先的 p^c 删去再加入 $p^{c\pm1}$，可以直接加入或者删除，对于 $p = 2, c \geq 3$ 的情况可以直接加一个 $v_2 = c - 2$ 的修改，这样就可以模拟 $g_m(2)$ 中第一个特判的效果了。

例如以下输入：

```
10
+3
+2
+5
+2
+3
+2
+11
-3
+2
-3
```

对于每个质数开一个链表或 vector，记录 pair (修改时间，增或删的指数即 v_p)：

```
p=2: (1,1), (3,2), (4,1), (6,1), (7,1), (9,2), (10,-1)
p=3: (5,1), (8,-1)
p=5: (7,1)
```

最终涉及到的就是区间乘一段 $\frac{p^g-1}{p^g}$，这样的乘的次数一共有 $\mathrm{O}(q \cdot \omega(p-1))$，为了避免求逆元再乘一个 $\log$，可以进行差分，同个右端点的区间先乘起来再求逆元，这样可以做到 $\mathrm{O}(q \log Mod)$；使用类似离线 $\mathrm{O}(1)$ 逆元的思路可以做到这部分线性。您可能会担心 $p^g - 1 \equiv 0$，但事实上当输入允许的质数 $\leq 10^7$ 时不存在这样的情况，即 $\forall\, p \in \mathbb{P}, \sum_{i\leq 10^7, i\in\mathbb{P}} [p \parallel (i-1)] + 1 < \delta_{998244353}(p)$。不过标程里还是象征性地判了一下。

总时间复杂度 $\mathrm{O}(p + q \log p)$，空间复杂度 $\mathrm{O}(p + q \cdot \omega(p-1))$。期望得分 100。想到部分正确思路可以获得 $50 \sim 90$ 不等的分数。实际上数据很水，复杂度写得不是特别好也能过。