SEAARC 解题报告

江苏省常州高级中学 张志俊

题目大意

给定依次排列在一条直线上的 n 个点，第 i 号点具有颜色 a_i。任意一对同色点之间都恰好连了一条圆弧，求有多少对异色圆弧相交。

时间限制：5 s

数据规模

- 对于 10% 的数据，$n \le 100$；
- 对于 40% 的数据，$|c| \le 100$；
- 对于 100% 的数据，$1 \le n \le 10^5$，$1 \le |c|, a_i \le 10^5$。

算法分析

为了方便之后的讨论，不妨记 cnt_i 为第 i 种颜色的总点数，记 $pre_{i,j}$ 为前 i 个点中具有颜色 j 的点数。由于题中所求的相交圆弧情形比较难以处理，因此我们尝试补集转化思想：记所有异色圆弧对总数为 s，形如 $AABB$ 的异色圆弧对数为 s_1，形如 $ABBA$ 的异色圆弧对数为 s_2，形如 $ABAB$ 的异色圆弧对数为 s_3，则易知 $s = s_1 + s_2 + s_3$，即 $s_3 = s - s_1 - s_2$，而 s_3 即为所求。下面依次讨论 s、s_1 和 s_2 的求解。

异色圆弧对总数 s

由于已知每种颜色的点数 cnt_i, 则第 i 种颜色的圆弧数为 $\binom{cnt_i}{2} = \frac{cnt_i(cnt_i-1)}{2}$，于是 $s = \sum_{i<j} \binom{cnt_i}{2}\binom{cnt_j}{2}$，易知这在线性时间内即可求得。

形如 $AABB$ 的异色圆弧对数 s_1

对于形如 $AABB$ 的异色圆弧对，考虑以 i 作为右侧圆弧的左端点，则其对 s_1 的贡献为 $(cnt_{a_i} - pre_{i,a_i}) \sum_{j \ne a_i} \binom{pre_{i,j}}{2}$，其中 $(cnt_{a_i} - pre_{i,a_i})$ 表示 i 右侧与其同色的点数，而 $\sum_{j \ne a_i} \binom{pre_{i,j}}{2}$ 则表示 i 左侧与其异色的圆弧数。借助上式，我们同样可以线性求得 s_1。

形如 $ABBA$ 的异色圆弧对数 s_2

通过一些简单的尝试，我们发现形如 $ABBA$ 的异色圆弧对性质较为复杂，难以只用较小的代价进行统计。因此，我们采取平衡规划的思想，对 cnt_i 以 k 为界进行划分并分别处理。

情形一：$cnt_A \geq k$

由于满足 $cnt_A \geq k$ 的颜色数最多只有 $\frac{n}{k}$ 种，因此这部分的处理较为简单：对于任意一条颜色异于 A 的圆弧 (x, y)，显然其贡献为 $pre_{x,A}(cnt_A - pre_{y,A})$。因此，若已知所有 $pre_{i,A}$，则对于任意一种不同于 A 的颜色 B，只需在顺次扫描所有具有颜色 B 的点的同时，再利用前缀和加速处理即可。整个过程的总代价为 $\Theta(\frac{n}{k} \cdot n) = \Theta(\frac{n^2}{k})$。

情形二：$cnt_A < k$ **且** $cnt_B \geq k$

这种情形与上种情形相类似，只不过此时只能枚举颜色 B 作为基础。为了方便之后的表述与计算，对于确定的颜色 A，记 pos_i 为第 i 个具有颜色 A 的点，记 $z = cnt_A$，并简记 $pre_{pos_1,B}, pre_{pos_2,B}, \ldots, pre_{pos_z,B}$ 为 $x_1, x_2, \ldots, x_z$。此时，可以得到颜色对 (A, B) 的贡献为：

$$\begin{aligned}
\sum_{i<j} \binom{x_j - x_i}{2} &= \frac{1}{2}\sum_{i<j}(x_j - x_i)^2 - \frac{1}{2}\sum_{i<j}(x_j - x_i) \\
&= \frac{1}{2}(z-1)\sum_i x_i^2 - \sum_{i<j} x_i x_j - \frac{1}{2}\sum_{i<j}(x_j - x_i) \\
&= \frac{1}{2}(z-1)\sum_i x_i^2 - \frac{1}{2}((\sum_i x_i)^2 - \sum_i x_i^2) - \frac{1}{2}\sum_{i<j}(x_j - x_i) \\
&= \frac{1}{2}z\sum_i x_i^2 - \frac{1}{2}(\sum_i x_i)^2 - \frac{1}{2}\sum_{i<j}(x_j - x_i)
\end{aligned}$$

由于上式中每部分的计算都只需对具有同种颜色的点顺序扫描即可，因此整个过程的总代价也是 $\Theta(\frac{n^2}{k})$。

情形三：$cnt_A < k$ **且** $cnt_B < k$

当两种颜色的出现次数均较小时，则上述的统计方式都不再适用，于是尝试另一种复杂度与同种颜色点数相关的方式。对于一条具有颜色 B 的圆弧 (x, y)，相应的方案数为颜色异于 B 的所有圆弧 (x', y')，且满足 $x' < x$ 及 $y < y'$。此时，如果我们采取类似于扫描线的方式从小到大枚举左端点 x，然后再枚举右端点 y，则方案数的计算可以借助后缀和求得，而加入一条圆弧的操作也能一并进行。由于此处后缀和的维护需要基于数据结构，因此整个过程的总代价为 $\Theta(nk \log n)$。

由于前两种情形的总代价均为 $\Theta(\frac{n^2}{k})$，而最后一种情形的代价为 $\Theta(nk \log n)$，因此，为了使得整个算法的总复杂度最优，易知应取 $k = \sqrt{\frac{n}{\log n}}$。此时，总复杂度为 $\Theta(n\sqrt{n \log n})$。在实际测试中，当 k 取在 10^2 级别时算法表现较好，更精确的取值则与具体实现有一定关联。

时间复杂度：$\Theta(n\sqrt{n \log n})$

空间复杂度：$\Theta(n)$