《数组》解题报告

杭州学军中学 金策

1 试题来源

PA 2014 Round 5 Ciągi

提交地址：http://main.edu.pl/pl/archive/pa/2014/cia

BZOJ：http://www.lydsy.com/JudgeOnline/problem.php?id=3710

2 试题大意

对于两个长度为n的整数数组$A = (a[1], a[2], \cdots, a[n]), B = (b[1], b[2], \cdots, b[n])$，定义它们的距离为$d(A, B) = \sum_{i=1}^{n} |a[i] - b[i]|$。

给定k个整数数组$A_1, A_2, \cdots, A_k$，求出它们的中心。中心定义为这样一个整数数组B，使得$\max\{d(A_i, B) | i = 1, 2, \cdots, k\}$尽可能小。如有多个解可以任意输出一个。

数据规模：$2 \le k \le 5, 2 \le n \le 10^5, |\text{数组元素}| \le 10^9$。

3 算法介绍

3.1 一个优化

我们先只考虑求解最小距离。

对于每一个$1 \le i \le n$，考虑$A_1[i], A_2[i], \cdots, A_k[i]$。设这些数排序后组成序列$A_{ind_i[1]}[i] \le \cdots \le A_{ind_i[k]}[i]$。那么在最优解中显然有$A_{ind_i[1]}[i] \le b[i] \le A_{ind_i[k]}[i]$。

如果对于i, j，数组$ind_i[]$和数组$ind_j[]$是完全相同的，则可以将i, j合并，以$A_1[i] + A_1[j], \cdots, A_k[i] + A_k[j]$来代替原先的$A_1[i], \cdots, A_k[i]$和$A_1[j], \cdots, A_k[j]$，

并用$b[i]+b[j]$代替$b[i]$和$b[j]$，最小距离不会改变。于是我们可以按$ind_i[]$分类进行合并，总共不超过$k!$种。

另外容易发现，将$A_1[i],A_2[i],\cdots,A_k[i]$同时乘上$-1$，最小距离也不会改变（只需对$b[i]$同样操作），而$ind_i[]$数组会被翻转。于是$k!$种排列中有一半是可以通过对称得到的，所以可以将$n$的规模缩减到$k!/2\le 60$。

3.2 一些定义和结论

对于$k=2,3,4$，可以通过将数组复制几遍使得$k=5$。下面只讨论$k=5$的情形。

假设我们的解为$B=(b[1],b[2],\cdots,b[n])$，对于每个$1\le i\le n$，定义：

定义1. i是**中立**的，若$A_{ind_i[2]}[i]\le b[i]\le A_{ind_i[4]}[i]$；

定义2. i**偏向**$t(1\le t\le 5)$，当$t=ind_i[1]$时，$A_{ind_i[1]}[i]\le b[i]\le A_{ind_i[2]}[i]$；当$t=ind_i[5]$时，$A_{ind_i[4]}[i]\le b[i]\le A_{ind_i[5]}[i]$；当$t\neq ind_i[1],ind_i[5]$时，$A_t[i]=b[i]$。

命题1. *最优解只存在两种情况：*

- *情况1：存在某个$t(1\le t\le 5)$，使得对于所有$1\le i\le n$，i偏向t；*
- *情况2：对于所有$1\le i\le n$，i是中立的。*

证明. 假如目前的解对于上述两种情况都不满足，那么一定存在非中立的i，根据对称性，不妨假设所有这样的i都满足$A_{ind_i[1]}[i]\le b[i]<A_{ind_i[2]}[i]$。如果存在$i,j$，满足$i,j$都是非中立的，而$ind_i[1]\neq ind_j[1]$，我们可以设$delta=\min(A_{ind_i[2]}[i]-b[i],A_{ind_j[2]}[j]-b[j])$，并令$b'[i]=b[i]+delta,b'[j]=b[j]+delta$，这样调整后不会产生距离增大，且使得$i,j$中至少一个变为中立，所以这样的调整只能进行有限次。调整到无法进行后，如果没有非中立的i，则符合情况2；否则，对于所有非中立的i，其$ind_i[1]$都是相等的，我们将其记作t。

如果不存在j，使得j是中立的且$b[j]\neq A_t[j]$，那么情况1已经符合；否则取这样的j，并取任一非中立的i，设$delta=\min(A_{ind_i[2]}[i]-b[i],|A_t[j]-b[j]|)$，并令$b'[i]=b[i]+delta$，$b'[j]=b[j]+\mathrm{sgn}(A_t[j]-b[j])\cdot delta$（其中$\mathrm{sgn}(x)$表示$x$的符号正负），这样调整不会产生距离增大，且使得$i$变为中立，或使得$j$偏向$t$，所以这样的调整也只能进行有限次。调整到无法进行后，如果没有非中立的i，则符合情况2；否则，所有中立的j都偏向t，所有非中立的i也偏向t，于是符合情况1。 □

3.3 情况1的处理

先考虑较为简单的情况1。我们枚举t，并将所有$ind_i[5]=t$的i通过对称转化为$ind_i[1]=t$。

对于$t\neq ind_i[1]$的i，$b[i]=A_t[i]$是确定的。

对于剩下的i，有$A_t[i]\le b[i]\le A_{ind_i[2]}[i]$，我们令$X=\sum_{i满足ind_i[1]=t}(b[i]-A_t[i])$，则$X$的大小范围为$0\le X\le\sum_{i满足ind_i[1]=t}(A_{ind_i[2]}[i]-A_t[i])$。

于是，$d(A_t,B)=X$，$d(A_s,B)=const_s-X(s\neq t)$，其中$const_s$是常数。简单讨论一下即可求出$\max_i d(A_i,B)$的最小值。

3.4 情况2的处理

情况2较为复杂。我们再做一些定义。

定义3. i是**特别中立**的，若$b[i]=A_{ind_i[3]}$；

定义4. 对于某个中立的i，i是**靠近** $\{s,t\}$的，当$A_{ind_i[2]}[i]\le b[i]\le A_{ind_i[3]}[i]$时，$\{ind_i[1],ind_i[2]\}=\{s,t\}$；当$A_{ind_i[3]}[i]\le b[i]\le A_{ind_i[4]}[i]$时，$\{ind_i[4],ind_i[5]\}=\{s,t\}$。（特别中立的$i$既靠近$\{ind_i[1],ind_i[2]\}$，又靠近$\{ind_i[4],ind_i[5]\}$）

命题2. *对于任意i,j，i和j所靠近的集合至少有一个公共元素。*

证明. 对于当前的解，若存在i,j不满足上述条件，不妨设$A_{ind_i[2]}[i]\le b[i]<A_{ind_i[3]}[i]$，$A_{ind_j[2]}[j]\le b[j]<A_{ind_j[3]}[j]$。设$delta=\min(A_{ind_i[3]}[i]-b[i],A_{ind_j[3]}[j]-b[j])$，并令$b'[i]=b[i]+delta,b'[j]=b[j]+delta$。由于$ind_i[1],ind_i[2],ind_j[1],ind_j[2]$无重复元素，所以这样调整不会发生距离增大。调整使得i,j中至少一个变为特别中立，所以这样的调整只能进行有限次。调整到无法进行后，此条件满足。 □

根据这一结论，最优解只存在两种情况：

- 情况2.1：存在t，使得对于所有非特别中立的i，i靠近$\{t,s_i\}$，$s_i\in\{1,2,3,4,5\}-\{t\}$；
- 情况2.2：存在r,s,t，使得对于所有非特别中立的i，i靠近$\{r,s\}$或$\{s,t\}$或$\{t,r\}$。

不妨设所有非特别中立的i都靠近$\{ind_i[1],ind_i[2]\}$，$A_{ind_i[2]}[i]\le b[i]\le A_{ind_i[3]}[i]$。

3.4.1 情况2.1的处理

枚举t，在这里为了叙述方便假设$t=5$。

对于$1 \le s \le 4$，令$X_s = \sum_{i靠近\{s,5\}}(b[i] - A_{ind_i[2]}[i])$，则$X_s$的范围是$0 \le X_s \le \sum_{i靠近\{s,5\}}(A_{ind_i[3]}[i] - A_{ind_i[2]}[i])$。那么我们有，

$$\begin{cases} d(A_1,B) = X_1 - X_2 - X_3 - X_4 + const_1 \\ d(A_2,B) = -X_1 + X_2 - X_3 - X_4 + const_2 \\ d(A_3,B) = -X_1 - X_2 + X_3 - X_4 + const_3 \\ d(A_4,B) = -X_1 - X_2 - X_3 + X_4 + const_4 \\ d(A_5,B) = X_1 + X_2 + X_3 + X_4 + const_5 \end{cases}$$

为求出$\max d(A_i,B)$的最小值，分两类情况讨论，

(1) $\max d(A_i,B) = d(A_5,B)$。那么固定$X_1+X_2+X_3+X_4 = S$时，对于$i \in \{1,2,3,4\}$可以得知$X_i \le S + \lfloor (const_5 - const_i)/2 \rfloor$。二分$S$后判断是否可行即可。

(2) $\max d(A_i,B) = d(A_s,B) (1 \le s \le 4)$，为了叙述方便假设$s=4$。那么固定$X_4$时，有$X_i \le X_4 - \lceil (const_i - const_4)/2 \rceil\ (1 \le i \le 3)$，且有$X_1+X_2+X_3 \le \lfloor (const_4 - const_5)/2 \rfloor$。目标是最小化$X_4 - (X_1+X_2+X_3) + const_4$，只要三分$X_4$即可。

于是情况2.1可以解决。

3.4.2 情况2.2的处理

与上面类似。为叙述方便假设$\{r,s,t\} = \{1,2,3\}$。

对$p \in \{1,2,3\}$，设$X_p = \sum_{i靠近\{1,2,3\}-\{p\}}(A_{ind_i[3]}[i] - b[i])$，则$X_p$的范围是$0 \le X_p \le \sum_{i靠近\{1,2,3\}-\{p\}}(A_{ind_i[3]}[i] - A_{ind_i[2]}[i])$。然后可以写出每个$d(A_i,B)$的表达式，结构和情况2.1中相同，可用同样的方法处理。

3.5 具体构造

我们已经求出了最小距离。具体的构造方案可以由上述求解过程自然得到，在此不赘述。

4 总结

这道题的难点主要在于发现结论，以及细致严谨的讨论与证明。