# 2020 年 第 5 問

**ジャンル: 算法**
**難易度: Lunatic**

> **問題の概要**
> 非常に難しい。筆者は丸 3 日考えた末に (1-2) までは思いついた。競技プログラミングなどをやっていれば試験中に思いつくことも可能なのだろうか...? (1-2) も (2) も難しく、悪問どころの話ではない。これを解いても勉強にはならないが、興味がある人は見てみればいいだろう。 解けた人は GitHub でぜひとも教えてほしい。

以降、以下の記号を用いる。

$$d_n(A, B) = \sum_{i=1}^{n} (A[i] - B[i])^2$$

## (1-1)

どうやら問題文の不備らしく、十分ではあるが必要ではないことに注意。

問題文で与えられている単調非減少配列を $C$ とする。すなわち、$C[i] = B[i] \quad (1 \leq i \leq n), C[n+1] = A'[n+1]$ とする。このとき、$d_{n+1}(A', C) = d_n(A, B)$ である。一方で、$A'$ に対する最適な単調非減少配列が $D$ で、$C$ は最適ではない $d_{n+1}(A', D) < d_{n+1}(A', C)$ とすると、

$$d_n(A, D \text{ から末尾を削ったもの}) \leq d_{n+1}(A', D) < d_{n+1}(A', C) = d_n(A, B)$$

これは $B$ の $A$ に対する最適性に矛盾する。よって $C$ は最適である。

## (1-2)

記号を定義する。

$$b_r := \frac{1}{r} \sum_{i=1}^{r} A'[i] \quad (\text{先頭 } r \text{ 個の平均})$$

いくつか補題を示す。

**Lemma 1**: $A$ に対する近似配列 $B$ が、ある $r$ $(1 \leq r \leq n)$ で、$B[1] = \cdots = B[r]$ であり、$B[r+1]$ 以降の要素は既に決まっているものとする。
このとき、$B[r+1] \geq b_r$ ならば、$B[1] = \cdots = B[r] = b_r$ とするのが最適である。 $B[r+1] < b_r$ ならば、$B[1] = \cdots = B[r] = B[r+1]$ とするのが最適である。

*Proof*: $d_r(A, c) = \sum_{i=1}^{r} (A[i] - c)^2$ を $c$ で微分することで、最小化する定数 $c$ は $b_r$ であることがわかる。$d_r(A, c)$ が $c$ に対する二次関数なので、$B[r+1] < b_r$ の場合もできるだけ $b_r$ に近い場所が最適である。これは $c = B[r+1]$ を意味する。 □

**Lemma 2**: 長さ $n$ の配列 $A$ に対して、最適な近似配列 $B$ が $B[1] = \cdots = B[n]$ であるとする。
このとき、$b_r \geq b_n$ $(1 \leq r \leq n)$

*Proof*: Lemma 1 より、$B[1] = \cdots = B[n] = b_n$ である。もし、$B[1] = \cdots = B[r] = c \neq b_r$ と変更しても、$B$ の最適性より悪化しかしない。Lemma 1 より、$b_r < b_n$ ならば $c = b_r$ とすると改善してしまうので、$b_r \geq b_n$ である。 □

準備は終わったので、示していく。

**Theorem 1**: $B[1] = \cdots = B[n]$ かつ $A'[n+1] < B[n]$ とする。このとき、$B'[1] = \cdots = B'[n+1]$ かつ $B'[n+1] < B[n]$

*Proof*: Lemma 1 より、$B[1] = \cdots = B[n] = b_n$ である。さらに問題文より、

$$b_{n+1} = \frac{1}{n+1}A'[n+1] + \frac{n}{n+1}b_n < \frac{1}{n+1}B[n] + \frac{n}{n+1}b_n = b_n$$

であり、Lemma 2 と合わせて、$b_r \geq b_n > b_{n+1}$ $(1 \leq r \leq n)$。

さて、$B'$ の最後の変化点について考える。$B'[1] = \cdots = B'[r] < B'[r+1]$ なる最後(最初?)の変化点 $r$ $(1 \leq r \leq n-1)$ が存在するとする。Lemma 1 より、変化点が存在するならば、そこでは $B'[1] = \cdots = B'[r] = b_r < B'[r+1]$ でなくてはいけない。$B'$ は単調非減少なので、$B'[r+1] \leq B'[n] \leq B'[n+1]$。 よって、

$$B'[n+1] \geq B'[n] \geq B'[r+1] > b_r \geq b_n$$

だが、これは $B$ の最適性に反する。長さ $n$ の近似配列は $B[i] = b_n$ が最適で、$B'[n+1] > b_n$ なので実行可能解に入っている。よって矛盾。$r = n$ か $r = n+1$ であるが、$r = n$ のときは明らかに $B'[n+1]$ をより小さくして $B'[n+1] = B'[n]$ にしたほうが改善するので、$r = n$ ではない。 よって、$r = n+1$ であり、$B'[1] = \cdots = B'[n+1]$。Lemma 1 より、$B'[1] = \cdots = B'[n+1] = b_{n+1} < b_n = B[n]$ より、示された。 □

## (2)

(1-1) のケースはいいが、 (1-1) でないときに (1-2) に帰着する方法を思いつくかどうか。

Monotonic Regression, Isotonic regression という問題が関連していそう。pool adjacent violators algorithm (PAVA) というアルゴリズムがあるらしい。また、この問題は狭義凸な二次計画問題なので、最適解は一意だそうです。だから(1-1)は間違っていないんですね。