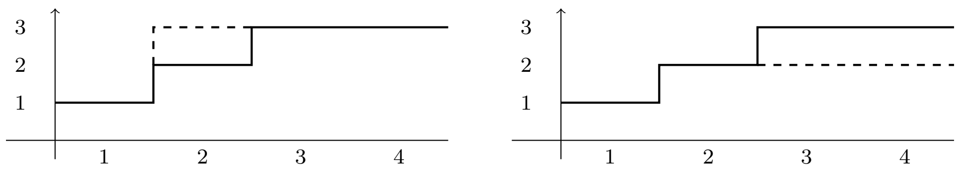
Task: DEV
Developer (author: Elias Rasmussen Lolck)

BOI 2025, Day 1: Analysis.

A first attempt would be to construct the solution greedily. It is easy to see that this is not optimal. Indeed, on the following example, the optimal solution is to raise property 2. However, for the first two properties no change is needed; going from left to right, we will rather lower properties 3 and 4, obtaining a cost of 2.



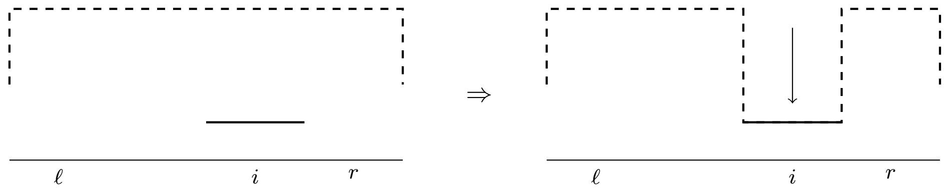
This suggests that some dynamic programming may be necessary. The difficulty is that, a priori, the number of possible levels for each of the properties is too large to be analysed. We first have to exclude most solutions by proving that they are surely not optimal. We will establish that, if we consider an optimal solution $b_1, \ldots, b_n$ in which the number of positions i with unchanged elevation (i.e., such that $a_i = b_i$) is maximal, then each b_i equals a_j, or $a_j + 1$, or $a_j - 1$ for some j in distance at most 4 (i.e., $|j - i| \leq 4$). Before we move to proving this, we will explain how this is sufficient to design an efficient dynamic programming solution.

For each property i we compute the set of candidate elevations $C_i = \{a_j - 1, a_j, a_{j+1} \mid i - 4 \leq j \leq i + 4\}$. Then, for every property i, for every elevation $y \in C_i$, and for every type of plateau $p \in \{\mathit{peak}, \mathit{valley}\}$ we compute $dp[i][y][p]$ to be the minimal cost of a solution $b_1, \ldots, b_i$ in which $b_i = y$ and in which b_i belongs to a plateau of type p. Such a solution either continues the solution given by $dp[i-1][y][p]$ (possible only if $y \in C_{i-1}$), or a solution given by $dp[i-1][y'][p']$, where $p' \neq p$, and $y' \in C_{i-1}$ satisfies $y' > y$ for $p = \mathit{valley}$ and $y' < y$ for $p = \mathit{peak}$; we have to find the minimum of those costs, and increase it by $|y - a_i|$.

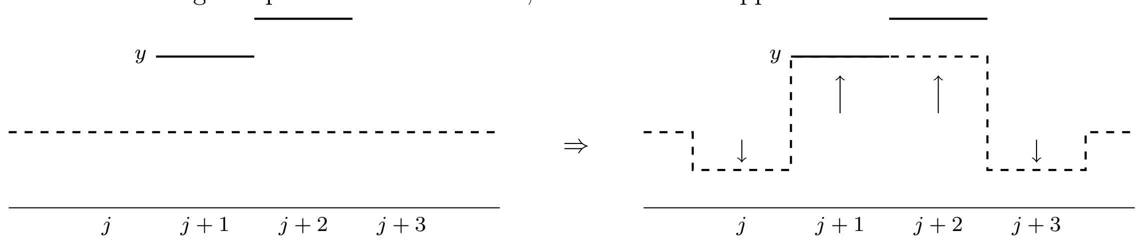
The algorithm works in time $\mathcal{O}(n)$. The limit for n in this task is relatively low, so the procedure for each i may be implemented naively (a loop over y' inside a loop over y), and it is also fine to use a larger constant instead of 4.

We now move to establishing the claimed property concerning an optimal solution $b_1, \ldots, b_n$ in which the number of positions with unchanged elevation is maximal. We first need some definitions. An *interval* $I = [\ell, r]$ is the set of properties $\{\ell, \ell + 1, \ldots, r\}$; its *length* is $r - \ell + 1$. An interval I is *malicious*, if for each $i \in I$ we have $b_i \neq a_i$. A maximal interval such that $b_\ell = b_{\ell+1} = \ldots = b_r$ is called a *plateau*. A plateau is a *peak* if $b_{\ell-1} < b_\ell > b_{r+1}$ (including the situations when $\ell = 1$ or $r = n$), and otherwise it is a *valley*. We now make several observations concerning the structure of the considered optimal solution:

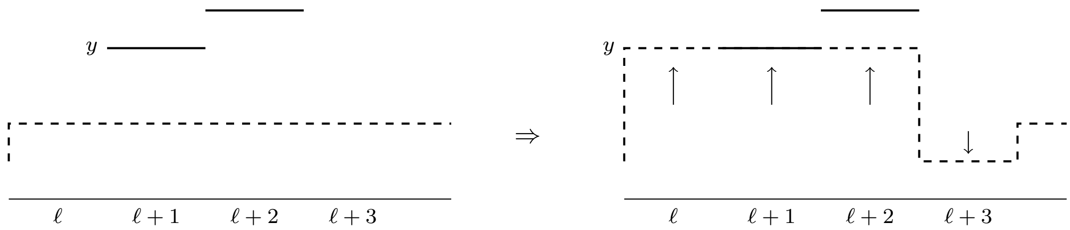
1. Suppose that a property i in the interior of a peak $[\ell, r]$ (i.e., satisfying $\ell < i < r$) is such that $a_i < b_i$. Then we can decrease b_i to a_i, obtaining a solution of better cost, which contradicts optimality. Thus, whenever $\ell < i < r$ for a peak $[\ell, r]$, we have $a_i \geq b_i$.



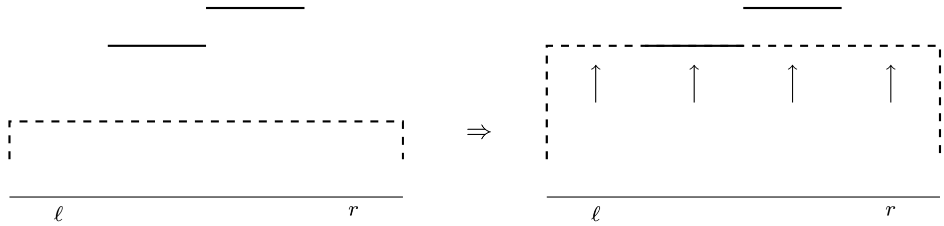
2. Suppose now that we have a malicious interval $[j, j+3]$ of length 4 strictly inside a peak $[\ell, r]$, so that $\ell < j$ and $j + 3 < r$. Let $y = \min\{a_{j+1}, a_{j+2}\}$; by Item 1 we have $y > b_j$. We can then change $b_j, b_{j+1}, b_{j+2}, b_{j+3}$ to $b_j - 1, y, y, b_j - 1$. This does not increase the cost (two properties are further from initial elevations by 1, the other two are closer by $y - b_j \geq 1$), but increases the number of properties with unchanged elevation; we have a contradiction with the assumption that the number of properties with unchanged elevation was maximal among all optimal solutions. Thus, this could not happen.



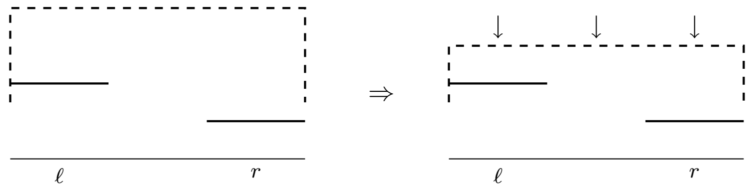
3. Next, suppose that we have a malicious interval $[\ell, \ell+3]$ of length 4 at the beginning of a peak $[\ell, r]$, so that $\ell+3 < r$. Let $y = \min\{a_{\ell+1}, a_{\ell+2}\}$; by Item 1 we have $y > b_\ell$. We can then change $b_\ell, b_{\ell+1}, b_{\ell+2}, b_{\ell+3}$ to $y, y, y, b_\ell-1$. This does not increase the cost (at least the properties $\ell+1$ and $\ell+2$ are shifted in direction of their original elevations), but increases the number of properties with unchanged elevation, which again contradicts with our assumptions. The same happens if a malicious interval $[r-3, r]$ of length 4 is at the end of a peak $[\ell, r]$, where $\ell < r-3$.



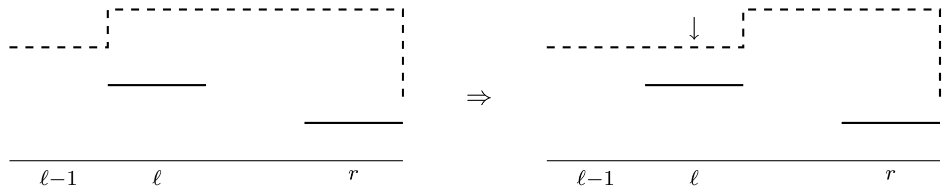
4. Next, suppose that a whole peak $P = \{\ell, \ell+1, \ldots, r\}$ is malicious. If the number of indices $i \in P$ such that $a_i > b_i$ is not smaller than the number of indices $i \in P$ such that $a_i < b_i$, then we can raise the whole peak to the smallest level of $a_i > b_i$; this either preserves or even decreases the cost, and increases the number of properties with unchanged elevation, leading to a contradiction.



Otherwise, there are more indices $i \in P$ such that $a_i < b_i$ than indices $i \in P$ such that $a_i > b_i$. If possible, we then lower the whole peak by 1, decreasing the cost; this contradicts optimality.



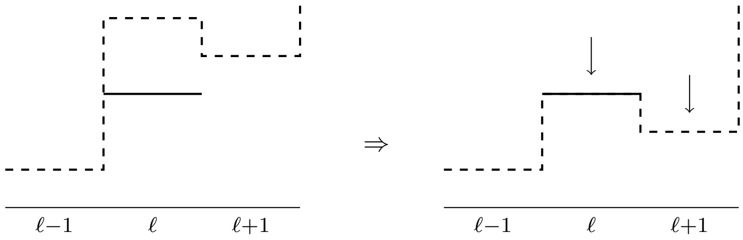
Such lowering does not lead to a valid solution if $b_{\ell-1} = b_\ell - 1$ (in particular $\ell > 1$) or if $b_{r+1} = b_r - 1$ (in particular $r < n$), so it is still possible that such a malicious peak exists. But recall that $a_i > b_i$ whenever $\ell < i < r$, and simultaneously there are strictly more indices $i \in P$ such that $a_i < b_i$ than indices $i \in P$ such that $a_i > b_i$; it follows that $a_\ell < b_\ell$ and $a_r < b_r$. If $r > \ell$ and $b_{\ell-1} = b_\ell - 1$, we can decrease the cost by lowering b_ℓ to $b_\ell - 1$; it contradicts optimality.



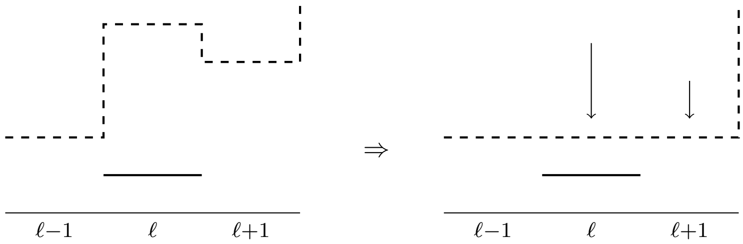
Likewise, if $r > \ell$ and $b_{r+1} = b_r - 1$, we can decrease the cost by lowering b_r to $b_r - 1$. It follows that the only possibility for a malicious peak is a peak $P = \{\ell\}$ of length 1, where moreover $a_\ell < b_\ell$ and either $b_{\ell-1} = b_\ell - 1$ or $b_{\ell+1} = b_\ell - 1$.

5. By symmetry, in the same way we can show that every malicious interval contained in a valley has length at most 3, and that a whole valley M can be malicious only if $M = \{\ell\}$ has length 1, where moreover $a_\ell > b_\ell$ and either $b_{\ell-1} = b_\ell + 1$ or $b_{\ell+1} = b_\ell + 1$.

6. Finally, suppose that we have a malicious peak $\{\ell\}$ followed by a malicious valley $\{\ell+1\}$ such that $b_{\ell+1} = b_\ell - 1$; let us see that this is not possible. Indeed, if $a_\ell > b_{\ell-1}$ (or if $\ell = 1$), then we can lower b_ℓ and $b_{\ell+1}$ to levels a_ℓ and $a_\ell - 1$ without changing the cost, but increasing the number of properties with unchanged elevation; a contradiction.



But if $a_\ell \leq b_{\ell-1}$, then we can lower both b_ℓ and $b_{\ell+1}$ to the level of $b_{\ell-1}$, strictly decreasing the cost; again, a contradiction.



By symmetry, it is also impossible to have a malicious valley $\{\ell\}$ followed by a malicious peak $\{\ell+1\}$ such that $b_{\ell+1} = b_\ell + 1$. It follows that a malicious peak cannot be surrounded by two malicious valleys: we said that a malicious peak is one level above some of its neighbours, and we now excluded this possibility if the neighbours are malicious. Likewise, a malicious valley cannot be surrounded by two malicious peaks.

7. Concluding, consider any malicious interval. How can it look like? Its first elements, at most 3, may belong to some plateau, which earlier has some non-malicious elements. Likewise, its last elements, at most 3, may belong to some plateau, which later has some non-malicious elements. In the middle, we may have at most 2 single-element plateaus. Moreover, the elevation of each such single-element plateau differs by one from the elevation of the neighbouring non-malicious plateau. It follows that each b_i equals a_j, or $a_j + 1$, or $a_j - 1$ for some j in distance at most 4 (i.e., $|j - i| \leq 4$).