NOI 2018 Solution Presentation

NOI Scientific Committee

Asia-Pacific Informatics Olympiad Invitation

APIO Invitation

All silver and gold medallists are cordially invited to participate in the Asia-Pacific Informatics Olympiad The selection for the Singapore International Olympiad in Informatics team will be 50% NOI, 50% APIO Top 4 Singapore Citizens/Permanent Residents will be invited to form the team Foreigners are not eligible for the team, but are invited to participate in the APIO

The to-be-confirmed details are as follows:

Date: 12th May 2018 (Saturday)

TIme: 9am to 2pm

Venue: NUS COM1 Programming Lab 2

You will receive a formal invitation via email when the details are confirmed You can find out more information at apio2018.ru

Scientific Committee

Dr. Steven Halim

Head Coach of NUS ICPC Teams (2008-2014, 2017)

Team Leader of Singapore IOI Team (2009-2013, 2015-present)

Deputy Chairman for IOI 2020, Singapore



Dr Wing-Kin Sung, Ken

Has multiple awards for research contribution in Algorithms and Bioinformatics

Senior group leader in Genome Institute of Singapore



Gan Wei Liang

IOI International Scientific Committee Singapore Representative (2018 - present)

Singapore IOI Team Coach (2016)

Singapore IOI Team Member (2012 - 2013)



Dr Frank Stephan

Research Interests:

- Recursion Theory and Kolmogorov Complexity
- Learning Theory
- Computational Complexity
- Automata and Formal Languages



Bernard Teo Zhi Yi

Singapore IOI Team Member (2012 - 2013)



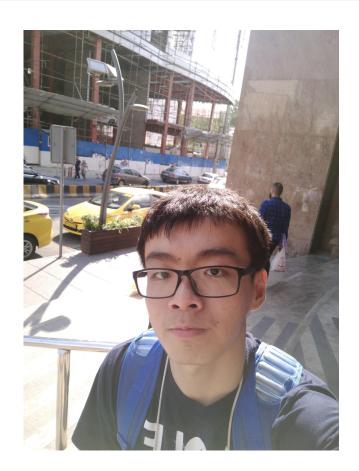
Clarence Chew Xuan Da

Singapore IOI Team Member (2016)



Zhang Guangxuan

Singapore IOI Team Member (2016 - 2017)



Pang Wen Yuen

Singapore IOI Team Member (2015 - 2017)



Lim Li

Singapore IOI Team Member (2017)



Problem Statistics

Interesting Statistics

167 participants, 14 gold, 28 silver, 35 bronze medallists

Collecting Mushrooms: 32 people scored 100

Journey: 22 people scored 100

Lightning Rod: 24 people score 100

City Mapping: highest score is 95

Safety: highest score is 40

Task 1: collectingmushrooms

Problem Author: Lim Li

Abridged Problem

Given an RxC grid, count the number of mushrooms within a D square of at least K sprinklers.

Subtask 1 (9%)

 $1 \le R,C \le 500, D = max(R,C), K=1$

Since every sprinkler can water the whole grid, each mushroom only needs 1 sprinkler to be harvestable, and the grid is guaranteed to have at least 1 sprinkler, you just need to count the number of mushrooms in the grid.

Subtask 2 (10%)

 $1 \le R,C \le 500, D = max(R,C)$

Same as subtask 1, but now you also need to count the number of sprinklers. If the number of sprinklers at least K, print the number of mushrooms. Else, print "0".

Subtask 3 (18%)

$$1 \le R,C \le 500, D = 1, K=1$$

D=1, so check the 3x3 square around each mushroom. If there is a sprinkler, then it is harvestable.

Complexity: O(9RC)

Subtask 4 (23%)

 $1 \le R,C \le 500$, no. of mushrooms ≤ 500 , no. of sprinklers ≤ 500

Since D can be big, checking the (D+1)x(D+1) square around each mushroom will TLE.

Instead, loop through the list of sprinklers to see which sprinkler can water that mushroom

Subtask 4 (23%)

```
For each mushroom:

water:=0

For each sprinkler:

if the mushroom is in the sprinklers range:

water++

if water ≥ K

This mushroom is harvestable
```

R = 1

Since R=1, the grid is now a line. The range of each sprinkler is now a line of length (2D+1).

To count the number of sprinklers that can water each mushroom, we can keep a prefix sum.

Value	1				1	1				1			1	1
Sum	1	1	1	1	2	3	3	3	3	4	4	4	5	6

Value	1				1	1				1			1	1
Sum	1	1	1	1	2	3	3	3	3	4	4	4	5	6

V	/alue	1				1	1				1			1	1
	Sum	1	1	1	1	2	3	3	3	3	4	4	4	5	6

Valu	Э	1				1	1				1			1	1
Sum	<u> </u>	1	1	1	1	2	3	3	3	3	4	4	4	5	6

This allows us to find the number of sprinklers in a range in O(1) time.

No additional constraints.

The solution for subtask 6 is a generalisation of the prefix sum idea from subtask 5. We store a 2D prefix sum.

1					
			1		
	1				
				1	
		1		1	
	1				1

1	1	1	1	1	1
1	1	1	2	2	2
1	2	2	3	3	3
1	2	2	3	4	4
1	2	3	4	6	6
1	3	4	5	7	8

1					
			1		
	1				
				1	
		1		1	
	1				1

1	1	1	1	1	1
1	1	1	2	2	2
1	2	2	3	3	3
1	2	2	3	4	4
1	2	3	4	6	6
1	3	4	5	7	8

1					
			1		
	1				
				1	
		1		1	
	1				1

1	1	1	1	1	1
1	1	1	2	2	2
1	2	2	3	3	3
1	2	2	3	4	4
1	2	3	4	6	6
1	3	4	5	7	8

1					
			1		
	1				
				1	
		1		1	
	1				1

1	1	1	1	1	1
1	1	1	2	2	2
1	2	2	3	3	3
1	2	2	3	4	4
1	2	3	4	6	6
1	3	4	5	7	8

This allows us to find the number of sprinklers in a 2D range in O(1) time.

Task 2: Journey

Problem Author: Frank Stephan

Abridged Problem

Input n (Number of Cities), m (Maximum Days used up), h (Flights out per City for Cities 0,1,...,n-2);

n x h data pairs (Next Destination, Minimum Stay).

Output: For each number of hotel nights (0,1,...,m-1), the number of itineraries which go from 0 to n-1 and go forward in each flight; Numbers above 500000001 are replaced by 500000001.

Example (Amman -> Cairo -> Dakar -> Brasilia)

In this example, n = 4 (for cities) and h = 3 (three outgoing flights per city); m is number of total nights (upper bound), say 4. Cannot use flight Dakar -> Cairo, as this is wrong direction. Cannot visit all cities in Amman - Cairo - Dakar - Brasiila, as too many nights.

From Amman	To Cairo, 2+ Nights	To Dakar, 3+ Nights	To Brasilia, 0+ Nights
From Cairo	To Dakar, 3+ Nights	To Dakar, 4+ Nights	To Brasilia, 0+ Nights
From Dakar	To Cairo, 0+ Nights	To Brasilia, 2+ Nights	To Brasilia, 0+ Nights

Algorithm and Data Types

```
Use <u>array a[i,p]</u> being the number of itineraries from City 0 to City i with p nights; Let Minadd(x,y) denote the update x = min(x+y,500000001).

Read n,m,h and create n x m array a with all entries initialised 0; let a[0,0]=1;

For i = 0,1,...,n-2 For each Flight leaving i Do Begin

Read next stop j and minimum stay k;

If i < j Then Do Begin

For p=0 to m-1-k For q=p to m-1-k

Do Begin Minadd(a[j,q+k], a[i,p]) End End;

Output a[n-1,0],a[n-1,1],...,a[n-1,m-1].
```

Improved Algorithm

```
Recall that Minadd(x,y) denotes x = min(x+y,500000001).

Read n,m,h and Initialise all entries a[i,p] with 0 except for a[0,0] which is 1;

For i = 0,1,...,n-2 For each Flight leaving i Do Begin

Read next stop j, minimum stay k;

If i < j Then Do Begin s=0;

For p=0 to m-1-k Do Begin Minadd(s,a[i,p]); Minadd(a[j,p+k],s) End End;

Output a[n-1,0],a[n-1,1],...,a[n-1,m-1].
```

Runtime Analysis

Both the Algorithm and the Improved Algorithm have an <u>outer loop</u> over h x n items and processes in each loop one outgoing flight from city i to city j and updates the entries at city j. <u>The inner loops</u> update m entries.

Original algorithm: each entry is updated O(m) times; Improved Algorithm: each entry is only updated O(1) times.

Original Algorithm has runtime $O(n \times m^2 \times h)$; Improved Algorithm has runtime $O(n \times m \times h)$.

Subtasks

<u>Subtask 1</u>: Only 1 Stop-Over possible, algorithm can be implemented without needing the array a to store intermediate information; <u>Subtask 2</u>: All flights are forward, solutions without if-statement can score; <u>Subtask 3</u>: Larger inputs; original algorithm and improved algorithm score; <u>Subtask 4</u>: Even larger inputs; as far as we tested it, only implementations of the improved algorithm scored for this subtask.

Task 3: lightningrod

Problem Author: Zhang Guangxuan

Abridged Problem

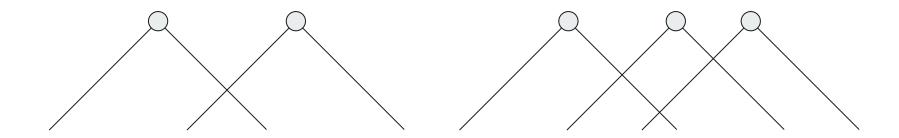
Find the minimum number of points to be chosen as lightning rods among N points, such that every point is covered by a lightning rod.

Point i is covered by Point j if and only if $|X_i - X_i| \le Y_i - Y_i$.

Subtask 1 (4%)

 $Y_i = 1$ for all i.

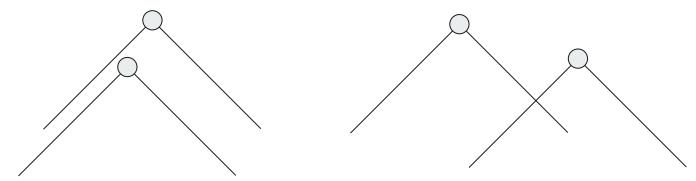
All right angle triangles cover exactly 1 point, so the answer is N



Subtask 2 (7%)

N = 2

If $|X_0 - X_1| \le |Y_0 - Y_1|$, one point can cover the other point, so the answer is 1, otherwise the answer is 2.



Subtask 3 (12%)

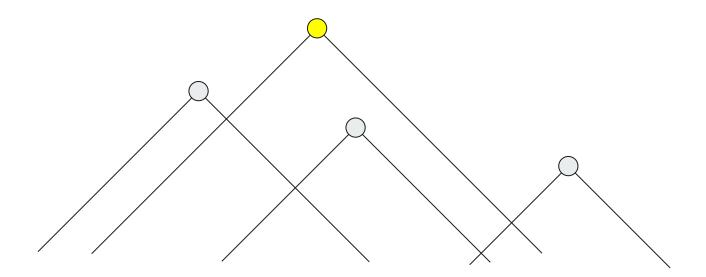
N ≤ 20

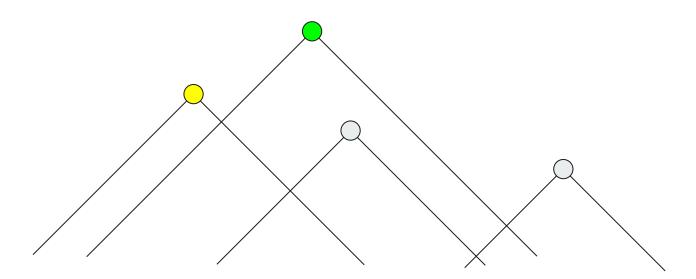
Do a $O(2^N)$ brute force for selecting a subset of points, $O(N^2)$ to check if the selection is valid.

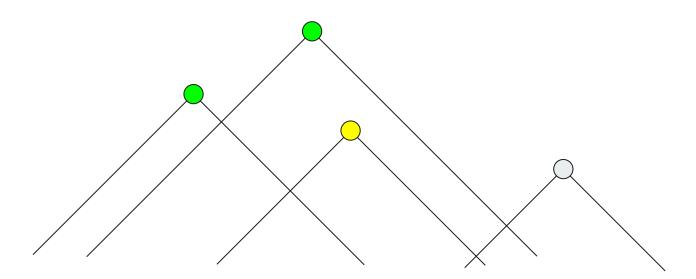
Overall complexity: $O(2^N N^2)$

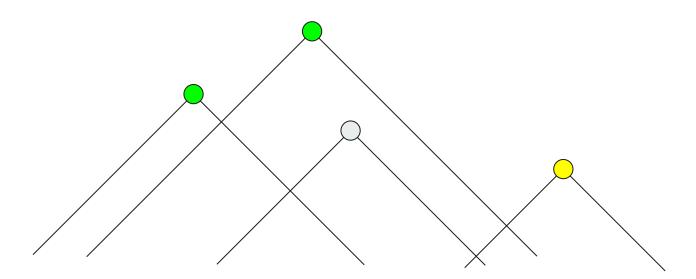
 $N \le 2000$

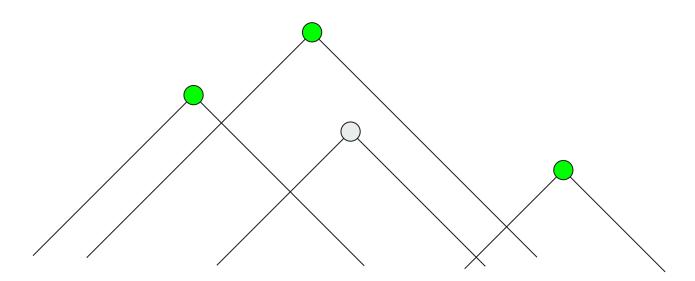
If we process points by decreasing Y_i , then we have to pick the current point if and only if previous points have not covered this point. We can check this condition for every point in O(N) by looping through previous points, giving a total complexity of O(N²).











Subtask 5 (26%)

 $N \le 200000$

We can improve on the subtask 4 solution. Instead of looping through all previous points to check if the current point is already covered, we just have to look for the highest point on the left and right of the current point.

Subtask 5 (26%)

Let the current point be (X_i, Y_i) . For points to the left, we look for the largest $X_j + Y_j$, and compare it with $X_i + Y_i$. For points to the right, we look for the largest $Y_j - X_j$, and compare it with $Y_i - X_i$. The current point is chosen if and only if no points cover the current point.

Subtask 5 (26%)

This can be in O(logN) using 2 range max point update segment trees or fenwick trees, where the first tree stores $X_i + Y_i$, second tree stores $Y_i - X_i$. For each point, we do a prefix query on the first tree, suffix query on the second tree. If the current node is chosen, update the current position with $X_i + Y_i$ in the first tree, $Y_i - X_i$ in the second tree

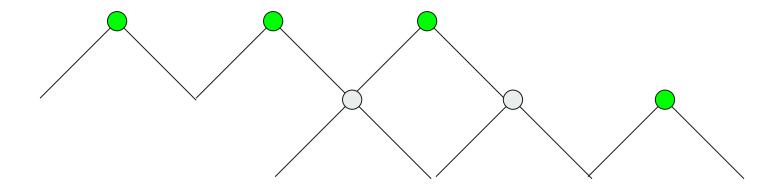
Subtask 6 (10%)

$$N \le 10000000, X_i = i, 0 \le Y_i \le 1$$

Each point can only be covered by the point directly on the left or right. It can be easily checked in O(1) for each point.

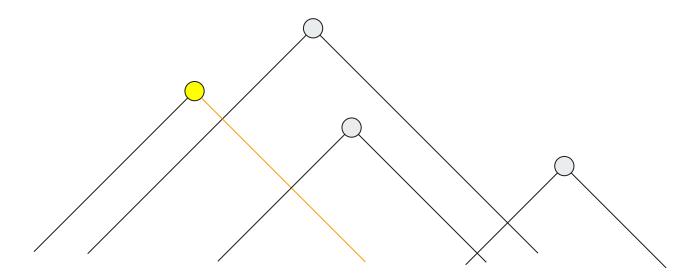
Note that for subtasks where $N \le 10\,000\,000$, fast input is required to ensure the programme runs within 1 second.

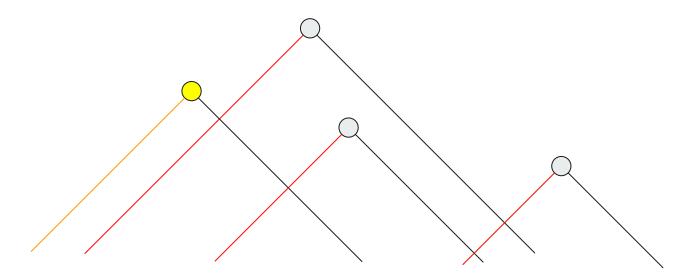
Subtask 6 (10%)

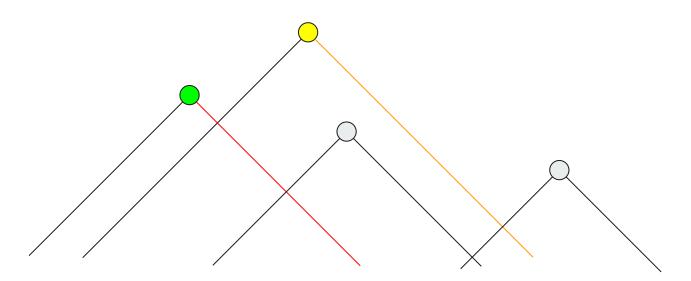


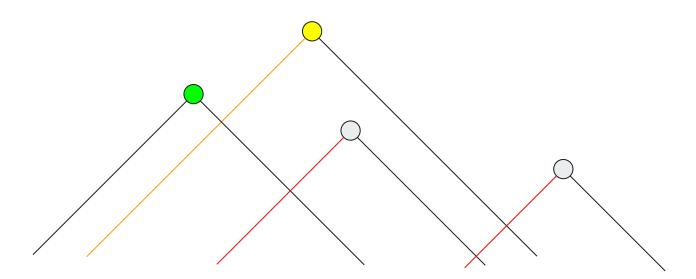
We observe that we can first assume all points are chosen, then pick out points which are not covered by other points.

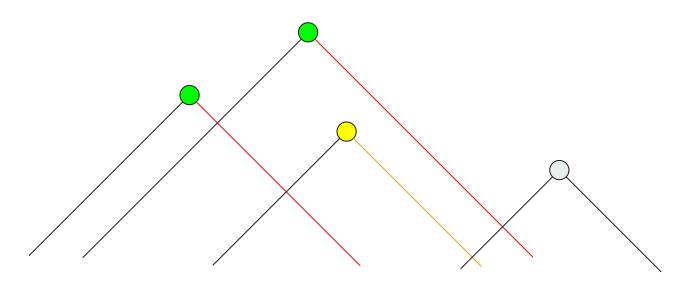
We pick a point (X_i, Y_i) if $X_i + Y_i$ is higher than all of $X_j + Y_j$ for i > j, and $Y_i - X_i$ is higher than all of $Y_j - X_j$ for i < j. This can be done with a static prefix max of $X_i + Y_i$ and suffix max of $Y_i - X_i$, for an overall complexity of O(N).

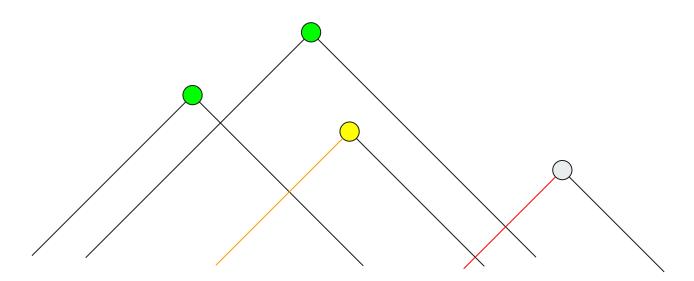


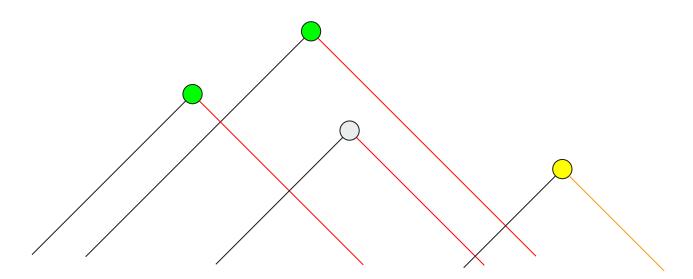


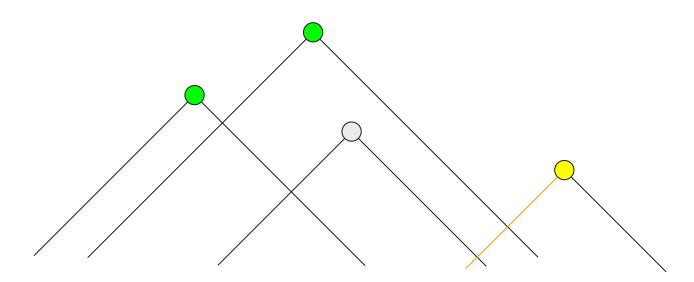












Task 4: citymapping

Problem Author: Pang Wen Yuen

Abridged Problem

You are **not** given a weighted tree of up to N = 1000 nodes, and each node degree ≤ 3 .

Use up to Q = 6500 queries of **distance between two nodes** to obtain the entire tree.

Subtask 1 (9%)

Q = 500000, all weights are 1.

Observation: Two nodes are connected by an edge if and only if their distance is 1.

Solution is to **query all pairs of nodes** and see if they are connected, then report the edge list.

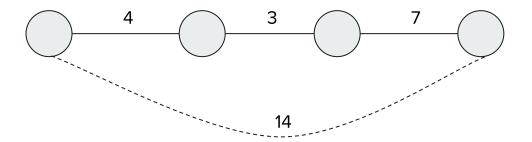
Subtask 2 (16%)

Q = 500000.

Unlimited queries!

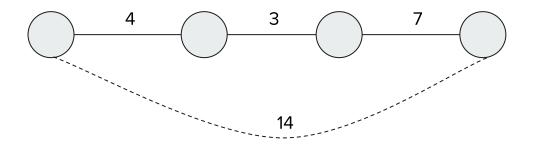
Subtask 2 (16%)

Construct a complete graph where the weight of the edge between two nodes is the distance between them.



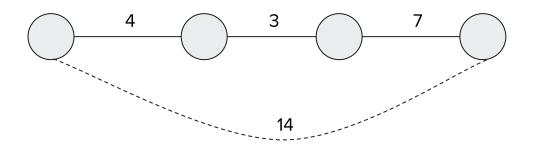
Subtask 2 (16%)

Notice that a "fake edge" will always connect two nodes already connected by edges with weights smaller than it.



Subtask 2 (16%)

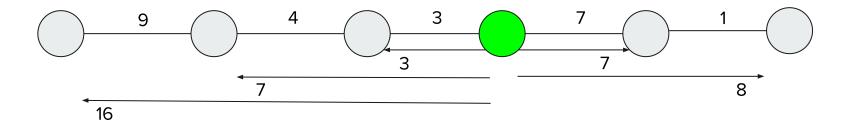
To eliminate "fake edges", we can run Kruskal, and the actual tree will be the Minimum Spanning Tree (MST) of the complete graph.



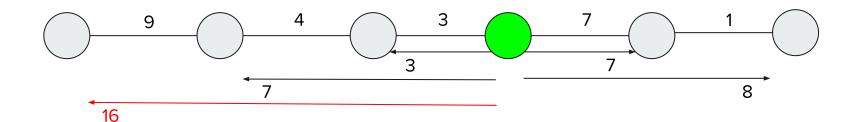
The tree is a line, Q = 12000.

(There exists solutions that only solve Subtask 3 but we won't go through them.)

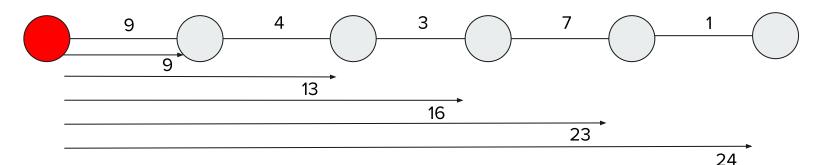
We take a random root node (node 0) and query its distance to all nodes.



Notice that the node furthest away must lie on one end of the line.



From this, we take the distances and sort them. An edge exists between two nodes which have adjacent distances in sorted order.

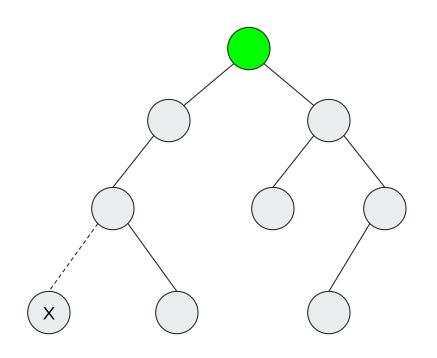


Full problem, Q = 6500.

First, we query the distance from a root node (node 0) to all other nodes. Then sort them in ascending order.

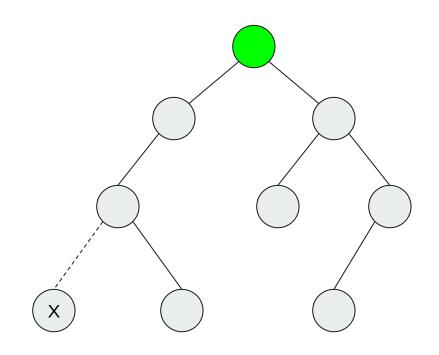
Full problem, Q = 6500.

Notice that a path to the root from a certain node X must only pass through nodes that have a **lower distance** to the root than node X.

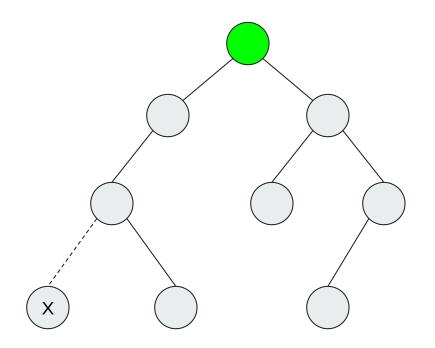


Full problem, Q = 6500.

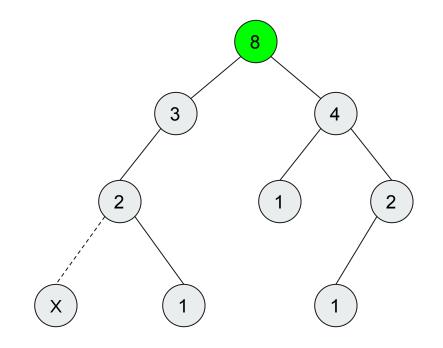
Therefore, we can process the nodes in sorted order, and connect them one by one into the "found tree".



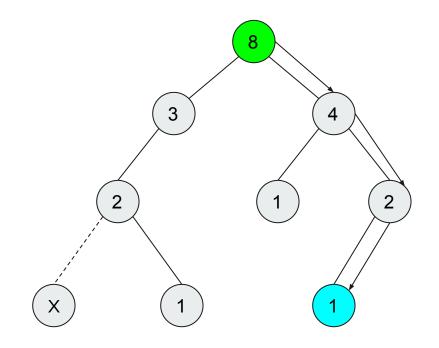
Let's consider the connecting of node X.



We first count the number of nodes in each subtree in the "found tree". (can be done with a single Depth-First Search)

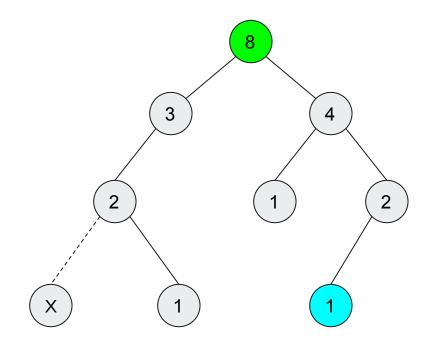


We start from the root. At each point we travel down to the child with the most children. (If tie, pick anyone.)



We query the distance from X to this leaf node.

The answer is 6.

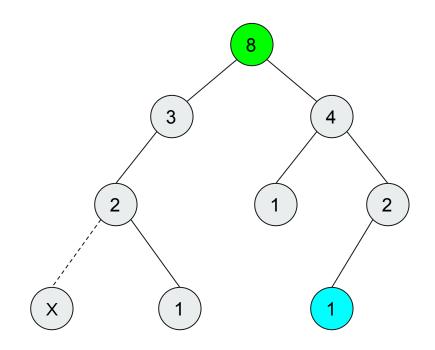


We have three distances now.

d(R, L) R = root

d(L, X) L = leaf

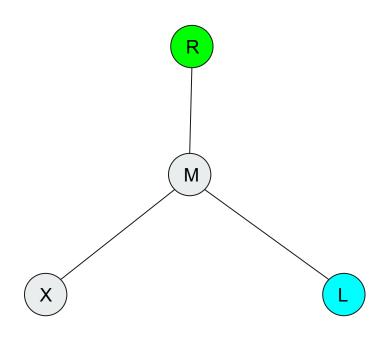
d(R, X) X = new node



From this we can find

d(M, R), d(M, X) and d(M, L)

where **M** is a centre connector node (it can end up being R, L or X)

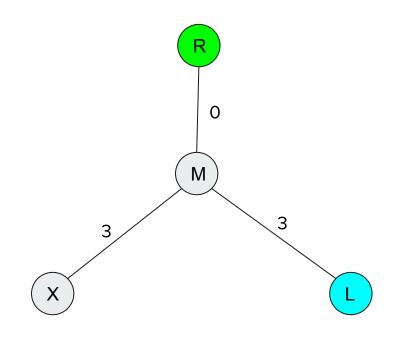


In this case, since:

$$d(R, L) = 3, d(R, X) = 3, d(X, L) = 6.$$

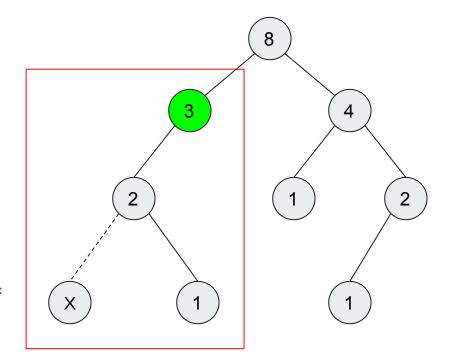
We can deduce:

$$d(R, M) = 0, d(M, L) = 3, d(M, X) = 3.$$

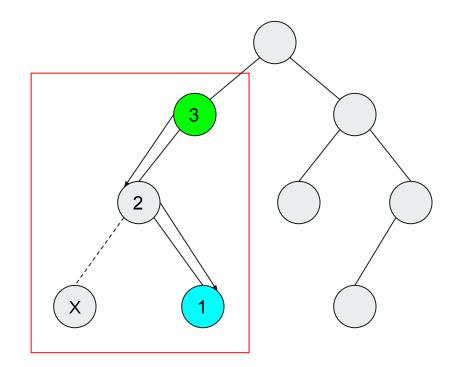


From this we can deduce X lies in this particular subtree, and the distance between the **new root** and X is 2 (since 3 - 1 = 2).

This is because we know d(M, R) = 0, so M (which is the branching point) is at the root.

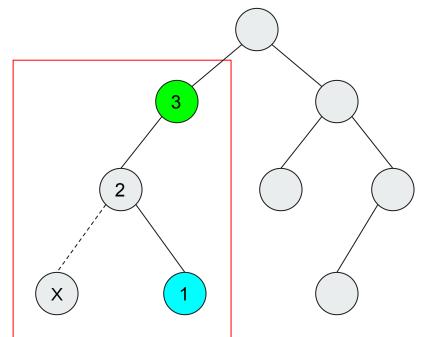


We can repeat this process.

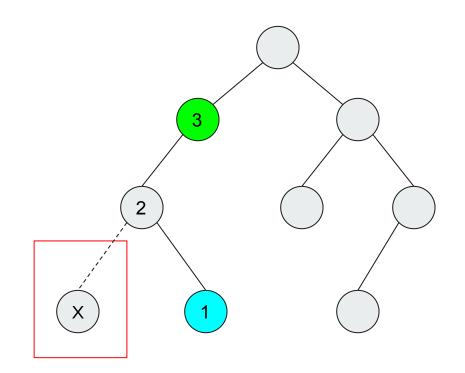


d(R, X) = 2 (known), d(R, L) = 2 (known), d(X, L) = 2 (costs 1 query).

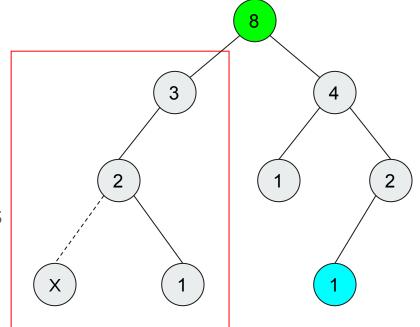
From this we can deduce that d(R, M) = d(M, X) = d(M, L) = 1.



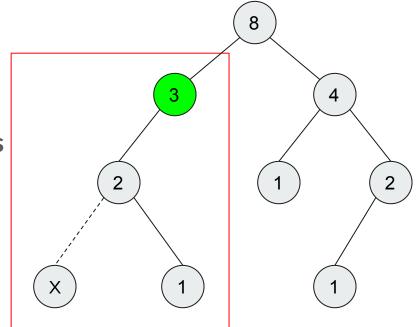
Now we can narrow down X to a single location.



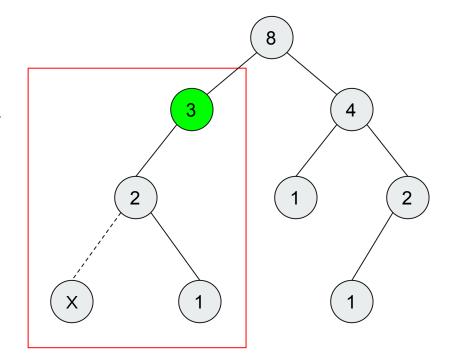
Each query will narrow down the search space for X to a single branch on the root -> leaf path (since we can find out d(R, M)). This is repeated until X is found exactly.



Since each time we travel down to the node with the most nodes in its subtree, we will prune out at least half of the nodes in every iteration (except for when the root has 3 children, then one-third).



Therefore, the overall complexity of queries is $N \log N$ with a low constant, which passes under the Q = 6500 limit.



Task 5: safety

Problem Author: Bernard Teo Zhi Yi

Abridged Problem

Given an array of **N** integers, find the minimum number of increment/decrement operations (on elements in the array) required such that all pairs of adjacent elements differ by no larger than **H**.

Abridged Problem

N = size of array

H = required maximum difference between adjacent elements

K = maximum possible height of any element

H = required maximum difference

K = maximum possible height

General Observation #1

Given two sequences A and B, both of length N, the minimum number of steps needed to get from A to B is

$$\sum_{i=1}^{N} |A[i] - B[i]|$$

where A[i] is the ith element of A, and B[i] is the ith element of B.

H = required maximum difference

K = maximum possible height

General Observation #2

If **H** > **K**, we can just set **H** to **K**.

Largest possible difference between two elements

= (largest possible height) – (smallest possible height)

≥ **K** - 0

= K

H = required maximum difference

K = maximum possible height

Subtask 1 (3%)

$$1 \le N \le 10, 0 \le H \le 4, K = 4$$

For each candidate sequence of length N (there are $(K+1)^N$ of them): If the sequence is safe, find the number of steps required

Output the minimum number of steps required amongst all the safe sequences.

Time complexity: O(K^NN)

Subtask 2 (4%)

$$1 \le N \le 14, 0 \le H \le 1, K = 4$$

H ≤ 1 ⇒ reduced branching factor

If S[i] is determined to be X, then S[i+1] can only be X-1, X, or X+1

It is faster to do a DFS that will only look through *safe* sequences

Subtask 2 (4%)

$$1 \le N \le 14, 0 \le H \le 1, K = 4$$

For each possible S[1] in between 0 and K (inclusive):

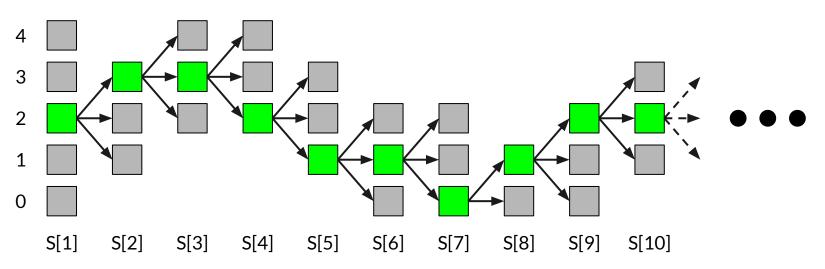
DFS to look through all possible *safe* sequences, finding the minimum number of stops required

finding the minimum number of steps required

H = required maximum difference

K = maximum possible height

Subtask 2 (4%)



Time complexity: $O((2H)^{N-1}NK)$

Subtask 3 (9%)

H = required maximum difference

K = maximum possible height

 $1 \le N \le 10, 0 \le H \le 2, K = 10^9$

Constraint for **K** is increased to 10⁹, but **N** and **H** remains small

⇒ Branching factor is still small, but how to choose starting points?

Subtask 3 (9%)

H = required maximum difference

K = maximum possible height

$$1 \le N \le 10, 0 \le H \le 2, K = 10^9$$

It can be shown that at least one stack need not be modified! (Try proving by contradiction. This is left as an exercise for the reader.)

Time complexity: $O((2H)^{N-1}N^2)$

H = required maximum difference

K = maximum possible height

Subtask 4 (5%)

$$1 \le N \le 2 \times 10^5$$
, $H = 0$, $K = 10^9$

The constraint **H** = 0 means that all stacks must have the same height after modification

⇒ the optimal height is the median height amongst all stacks

Time complexity: O(N log N)

H = required maximum difference

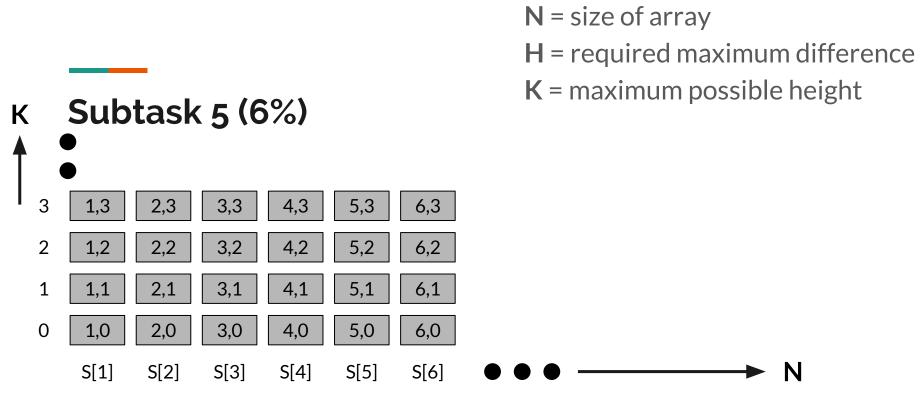
K = maximum possible height

$1 \le N \le 500$, $H = 10^9$, K = 400

Subtask 5 (6%)

Dynamic programming (on N and K dimensions)

best(\mathbf{n} , \mathbf{k}) = minimum number of operations required from stack 1 to stack \mathbf{n} such that the \mathbf{n}^{th} stack has height exactly \mathbf{k}



best(\mathbf{n} , \mathbf{k}) = minimum number of operations required from stack 1 to stack \mathbf{n} such that the \mathbf{n}^{th} stack has height exactly \mathbf{k}

Subtask 5 (6%)

H = required maximum difference

K = maximum possible height

 $1 \le N \le 500$, $H = 10^9$, K = 400

best(n, k) depends only best(n-1,?) (i.e. the previous column)

⇒ there is no cyclic dependency; dynamic programming is possible

H = required maximum difference

K = maximum possible height

Subtask 5 (6%)

$$1 \le N \le 500$$
, $H = 10^9$, $K = 400$

$$best(n,k) = \min_{k'=\max(k-H,0)}^{\min(k+H,K)} \{best(n-1,k')\} + |k-S[n]|$$

State: O(NK)

Transition: O(H)

Time complexity: O(NHK)

H = required maximum difference

K = maximum possible height

Subtask 6 (11%)

$$1 \le N \le 500$$
, $H = 10^9$, $K = 5000$

$$best(n,k) = \min_{k'=\max(k-H,0)}^{\min(k+H,K)} \{best(n-1,k')\} + |k-S[n]|$$

Use a balanced binary search tree to store best(n-1,?) within valid range – C++ std::multiset or Java TreeSet (careful with duplicates!)

Time complexity: O(NK log H)

H = required maximum difference

K = maximum possible height

Subtask 7 (11%)

 $1 \le N \le 5000$, $H = 10^9$, K = 5000

In the previous subtask, we have reduced the transition time from O(H) to O(log H). We can further reduce transition time to amortized O(1) using a double-ended queue.

The sliding range min query problem is well-known. (Proof omitted.)

Time complexity: O(NK)

H = required maximum difference

K = maximum possible height

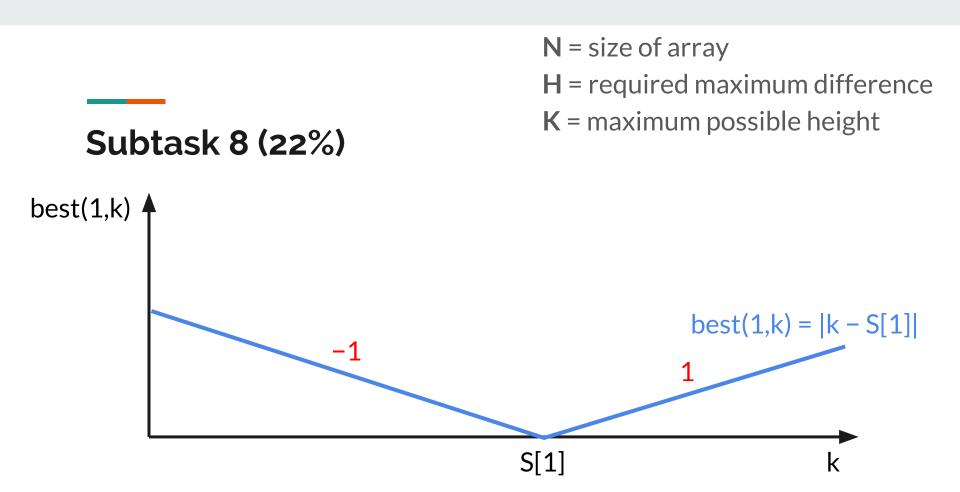
Subtask 8 (22%)

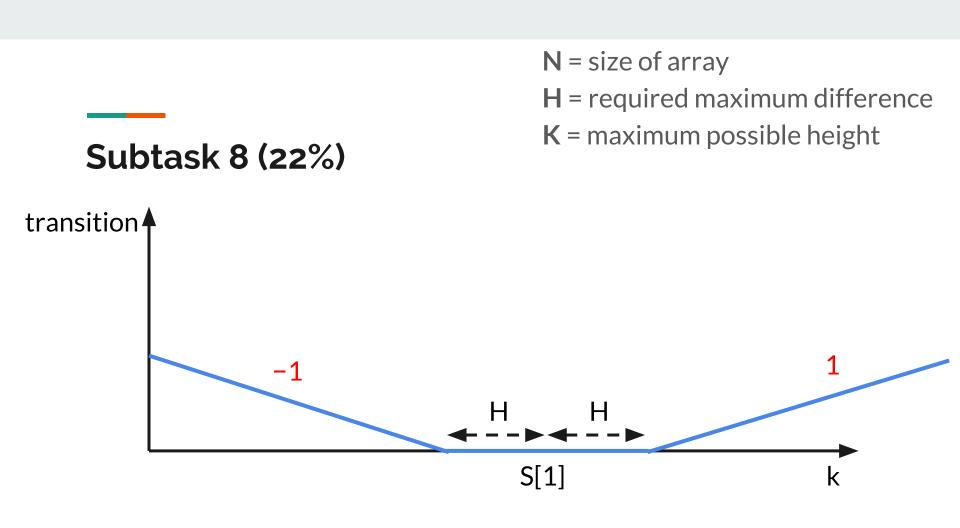
$$1 \le N \le 5000$$
, $H = 10^9$, $K = 10^9$

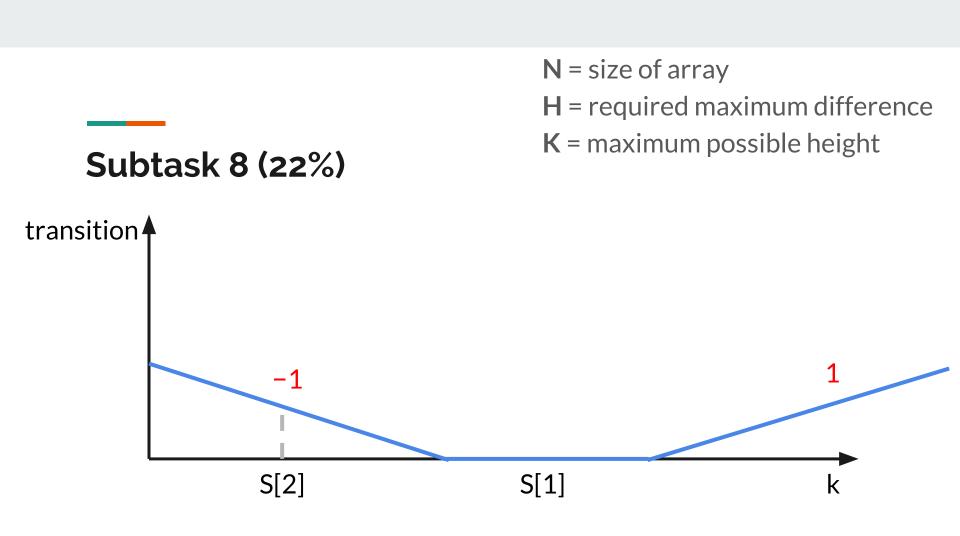
How to reduce the time taken beyond the DP state complexity?

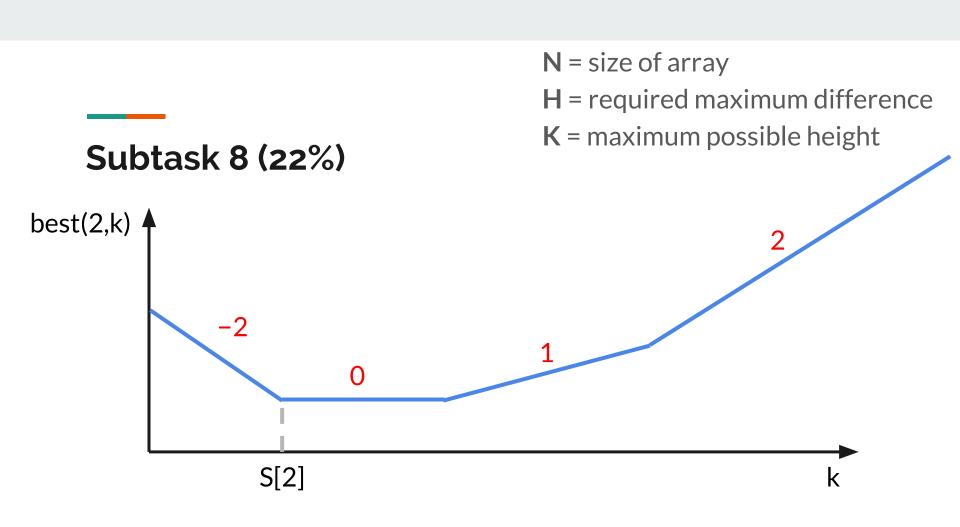
Can we avoid storing best(n, k) as is?

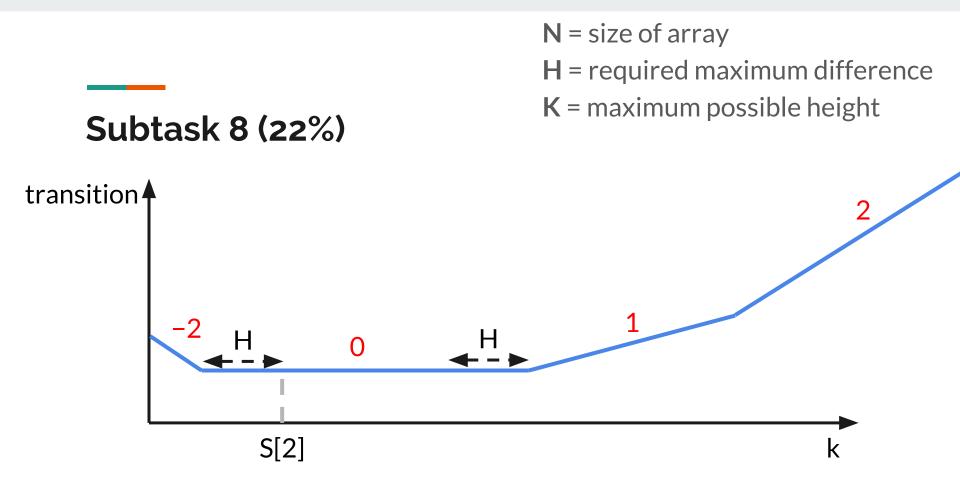
Can we quickly obtain the whole column best(\mathbf{n} ,?) from the previous column best(\mathbf{n} -1,?)?

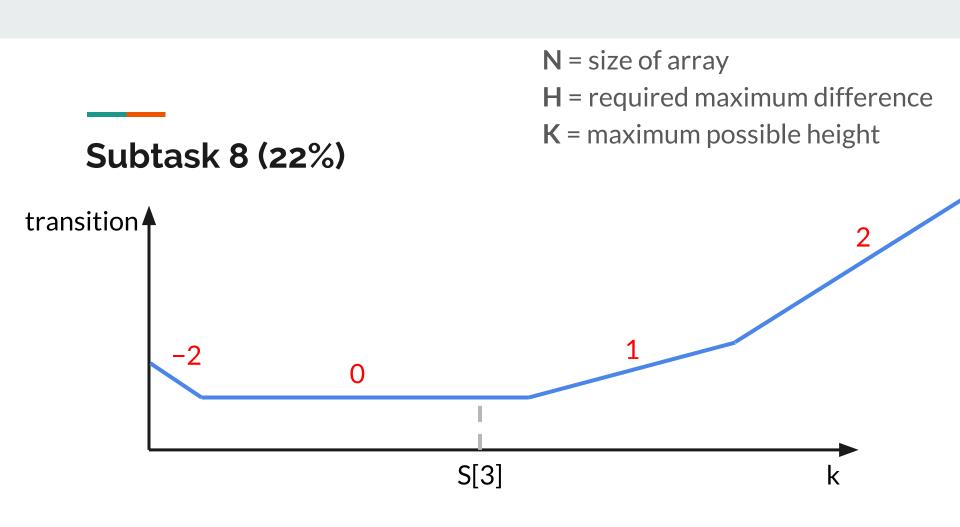


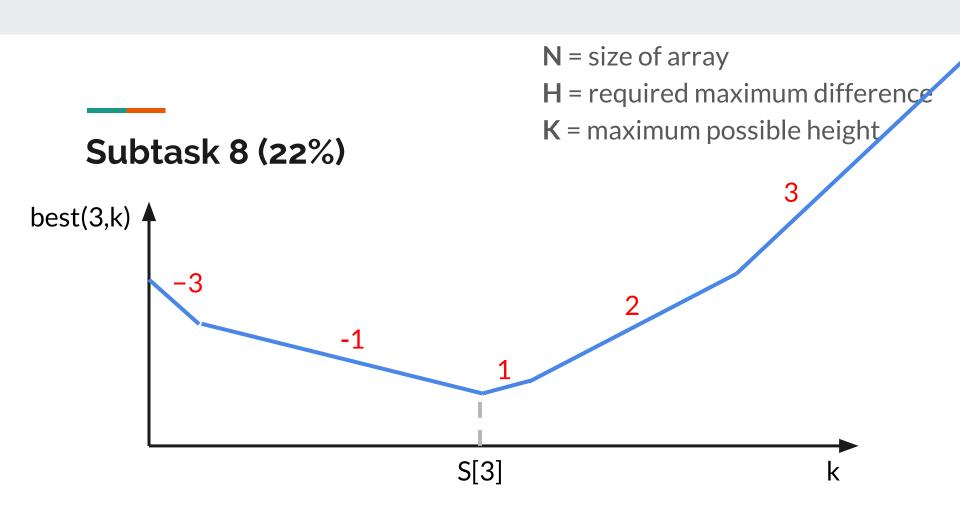












H = required maximum difference

K = maximum possible height

Subtask 8 (22%)

Notice that the number of points where the gradient changes is at most 2N

- \Rightarrow We can adjust these critical points in O(N)
- \Rightarrow Obtaining the whole column best(\mathbf{n} ,?) from the previous column best(\mathbf{n} -1,?) can be done in O(\mathbf{N})

Time complexity: $O(N^2)$

Subtask 9 (29%)

H = required maximum difference

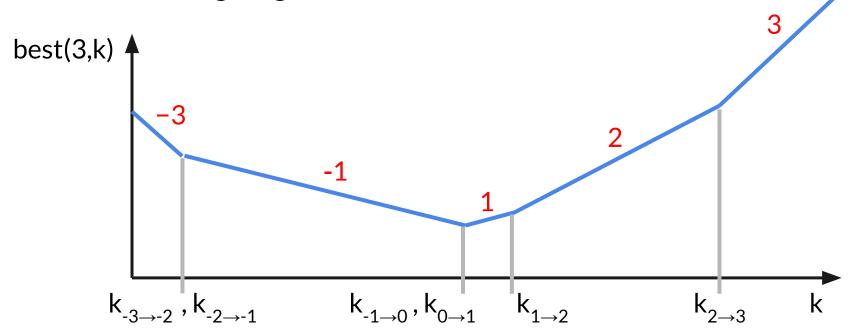
K = maximum possible height

$$1 \le N \le 2 \times 10^5$$
, $H = 10^9$, $K = 10^9$

Can we obtain the whole column best(\mathbf{n} ,?) from the previous column best($\mathbf{n}-1$,?) even more efficiently?

Subtask 9 (29%)

N = size of arrayH = required maximum differenceK = maximum possible height



H = required maximum difference

K = maximum possible height

Subtask 9 (29%)

Store the k-values of all the critical points in pair of priority queues or multisets (see task analysis for more details)

 \Rightarrow Obtaining the whole column best(\mathbf{n} ,?) from the previous column best(\mathbf{n} -1,?) can be done in O($\log \mathbf{N}$)

Time complexity: O(N log N)

End