CPM 2024

A data structure for the maximum-sum segment problem with offsets

Yoshifumi Sakai
Tohoku University

Maximum-sum segment (MSS) problem

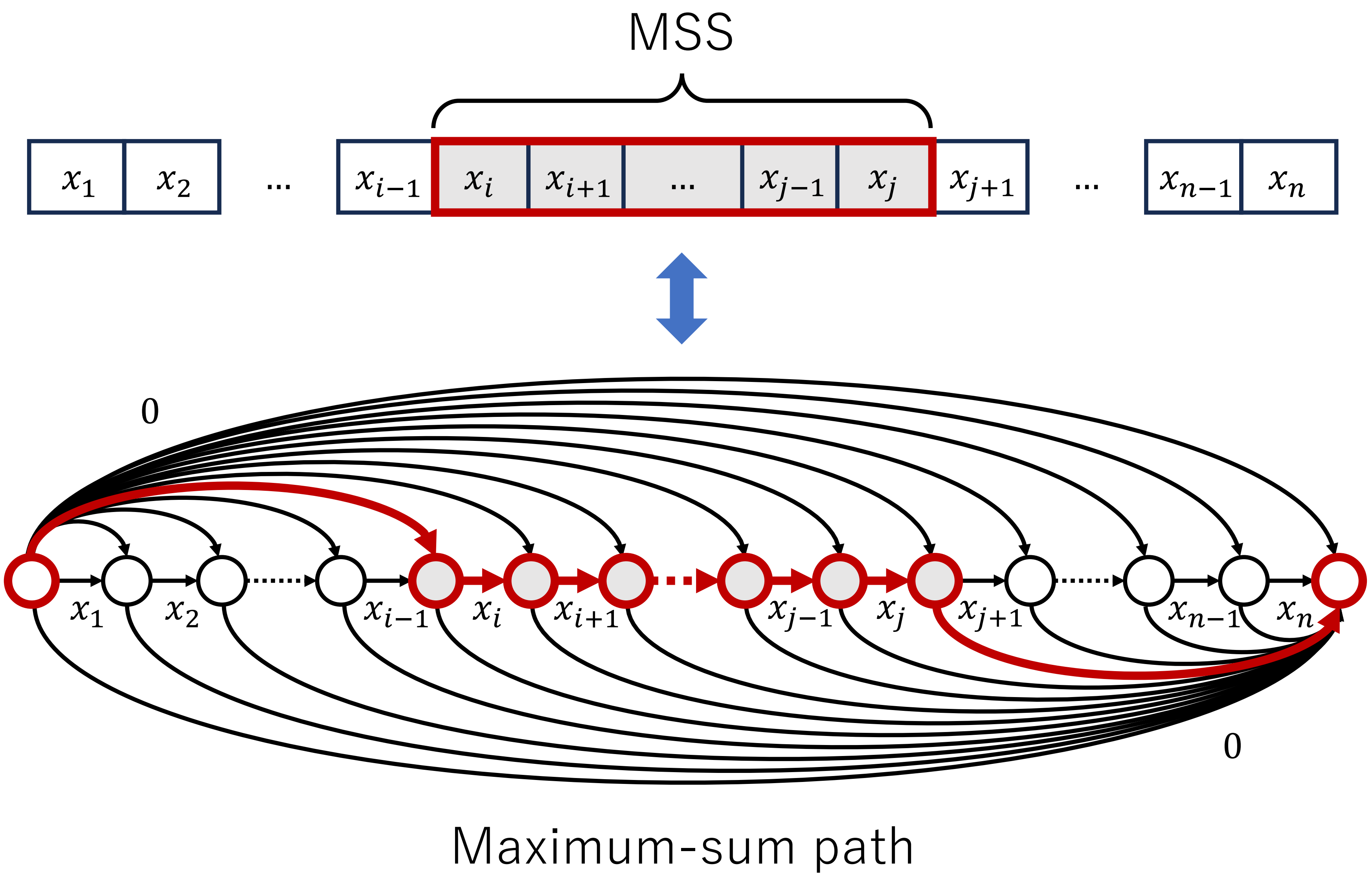
Which segment of a given numerical sequence maximizes the sum of its elements?

Example of application: Contiguous region highly conserved in homologous amino acid sequences

	L	V	V	L	T	C	L	L	C	V	A	F	T	Q	D	T
	L	L	V	Y	L	T	L	Y	C	A	A	F	V	V	G	S
Similarity	2	−1	2	−3	−3	−3	2	−3	7	−2	2	4	−2	−4	−3	−1

MSS

Kadane's linear-time algorithm (Bentley 1984)



Variants and related problems

- All maximal local MSSs (MLMSSs) (Ruzzo & Tompa 1999)
- Maximum-density segment (Chung & Lu 2005)
- Range MSS query (Chen & Chao 2007)
- Given number of non-overlapping segments that maximize the sum of their elements (Bengtsson & Chen 2007)
- Density constrained MSS (Cheng et al. 2009)
- Range position specific MLMSS query (Sakai 2018)
- MSS with uncertainty (Yu et al. 2021)

Offset-MSS problem

Given a numerical sequence X and a real number a, the offset-MSS problem is to find an MSS of X_a, where X_a is obtained by replacing each element x of X with $x - a$.

Example: MSSs of X_{12} and X_{13}

X	12	11	10	19	15	16	11	13	3	1	11	2	8	14	21	14	10	16	1	9	6
X_{12}	0	−1	−2	7	3	4	−1	1	−9	−11	−1	−10	−4	**2**	**9**	**2**	**−2**	**4**	−11	−3	−6
X_{13}	−1	−2	−3	**6**	**2**	**3**	−2	0	−10	−12	−2	−11	−5	1	8	1	−3	3	−12	−4	−7

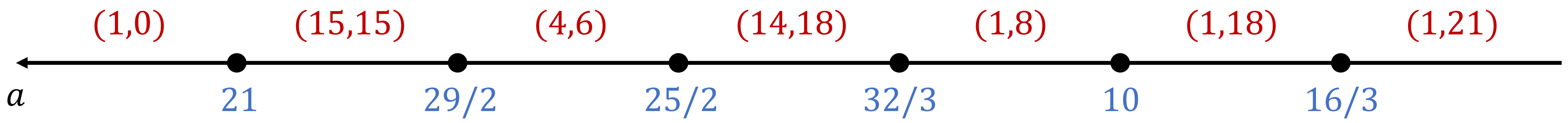
Data structure proposed

Number-line partition by a with the same MSS

- ✓ $O(n \log^2 n)$-time, $O(n)$-space constructible for any X of length n
- ✓ Supporting $O(\log n)$-time queries of an MSS of X_a for any a

Example:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
X	12	11	10	19	15	16	11	13	3	1	11	2	8	14	21	14	10	16	1	9	6



Notations

$X_a(i,j)$: Subsequence of X_a at position between i and j

$S_a(i,j)$: Sum of all elements in $X_a(i,j)$

$\alpha(i,j)$: Minimum a having some k with $i \leq k \leq j$ such that $S_a(i,k) = 0$ or $S_a(k,j) = 0$

$\kappa(i,j)$: Any k achieving $\alpha(i,j)$

Example:

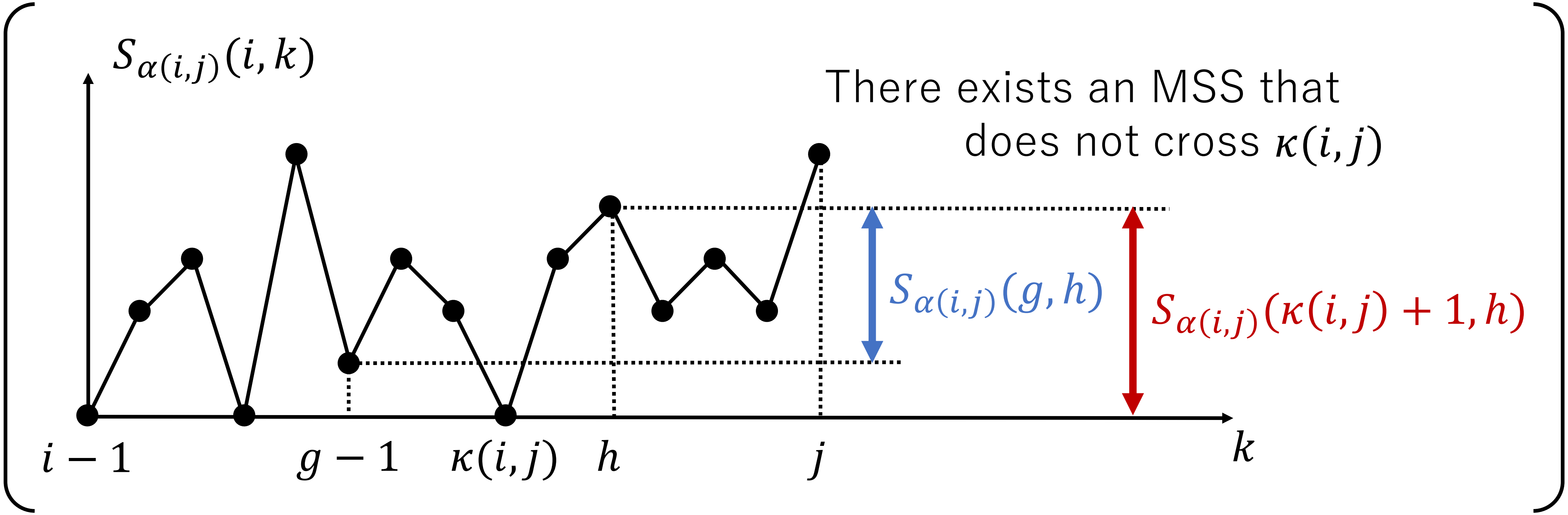
	4	5	6	7	8
$X_0(4,8) =$	19	15	16	11	13
($X_{12}(4,8) =$	7	3	4	−1	1)

→ $\alpha(4,8) = 12, \quad \kappa(4,8) = 7$

Basic lemma

If $a \leq \alpha(i,j)$, then (i,j) is an MSS of $X_a(i,j)$

If $a \geq \alpha(i,j)$, then an MSS of $X_a(i,j)$ can be found as any MSS of at least one of $X_a(i,\kappa(i,j)-1)$ and $X_a(\kappa(i,j)+1,j)$



Corollary of basic lemma

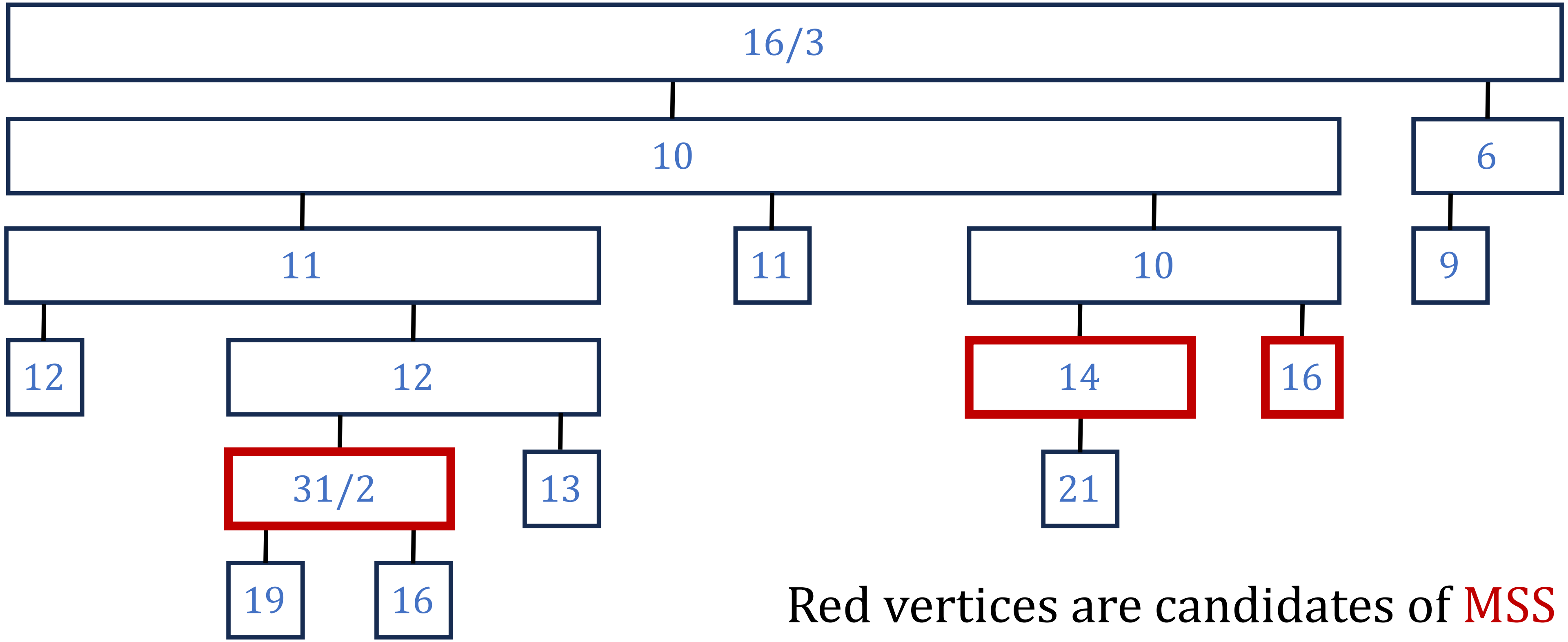
$\tau(i,j)$: Tree such that (i,j) is the root and any vertex (g,h) has $(g,\kappa(g,h)-1)$ and $(\kappa(g,h)+1,h)$ as its children, if $\alpha(g,h) \leq \alpha(i,j)$, and is a leaf, otherwise

T: Tree such that $(1,n)$ is the root and any vertex (i,j) has all leaves of $\tau(i,j)$ as its children, if $i \leq j$, and is a leaf, otherwise

⬇

Any vertex (g,h) of T maximizing $S_a(g,h)$ is an MSS of X_a

Example of T (leaves are omitted; labels indicate $\alpha(i,j)$)



Red vertices are candidates of MSS of X_{13}

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
X	12	11	10	19	15	16	11	13	3	1	11	2	8	14	21	14	10	16	1	9	6

Cor. of cor. of the basic lemma

(i_p, j_p): Any vertex (i, j) of T with $j - i + 1 = p$
that maximizes $S_0(i, j)$

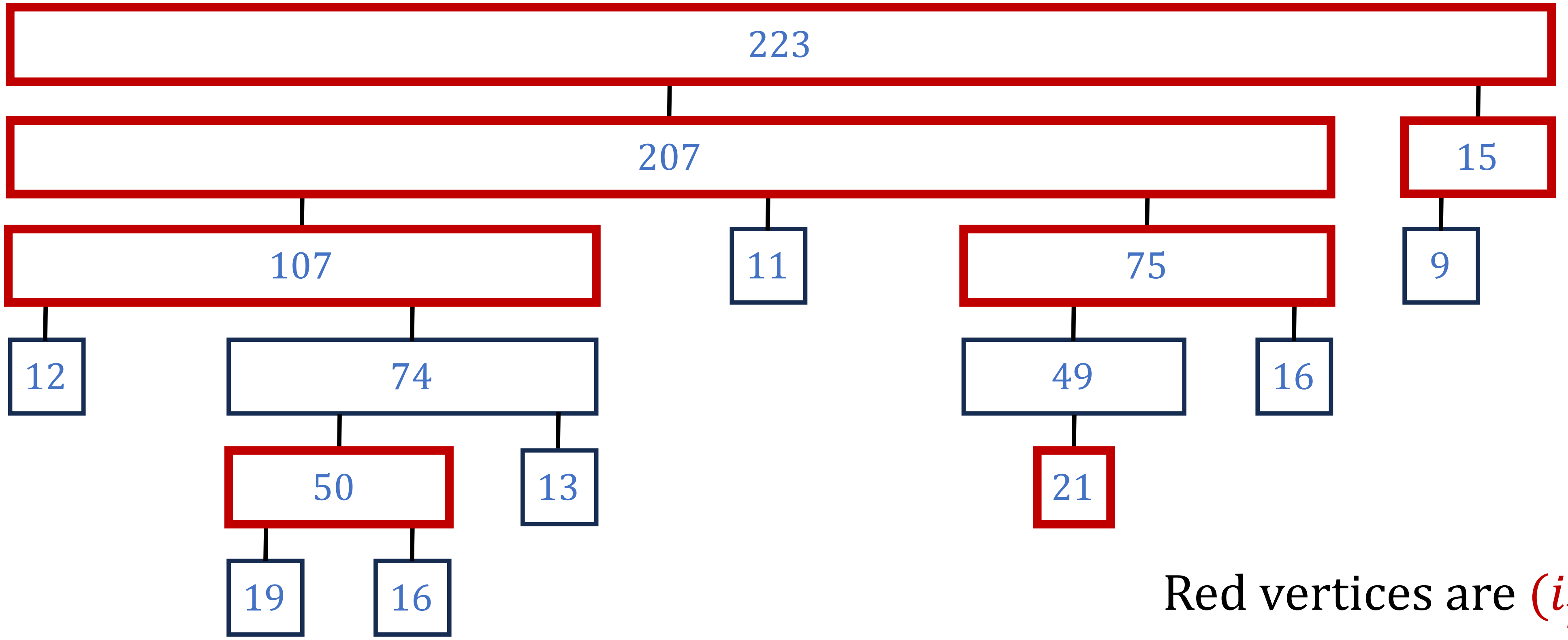
⬇

Any (i_p, j_p) maximizing $S_a(i_p, j_p)$ is an MSS of X_a

(For any vertex (g, h) of T with $h - g + 1 = p$ and any a,

$$S_a(g, h) = S_0(g, h) - ap \leq S_0(i_p, j_p) - ap = S_a(i_p, j_p)$$
)

Example of T (leaves are ommied; labels indicate $S_0(i,j)$)



Red vertices are (i_p, j_p)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
X	12	11	10	19	15	16	11	13	3	1	11	2	8	14	21	14	10	16	1	9	6

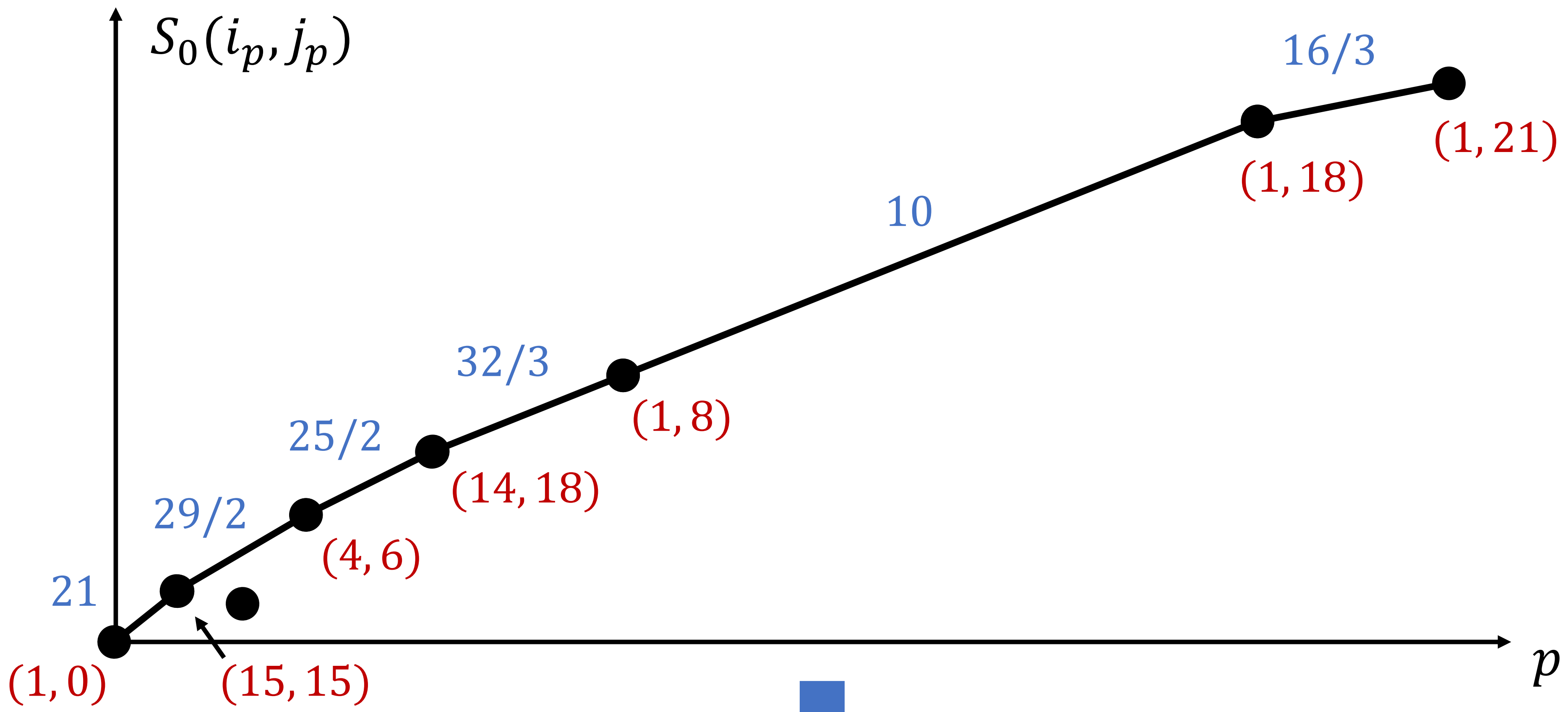
Cor. of cor. of cor. of the basic lemma

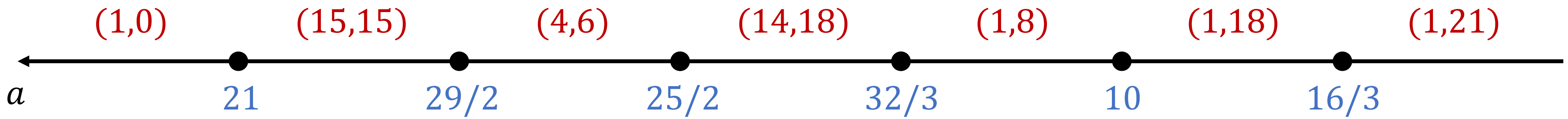
(i_r, j_r) is an MSS of X_a,
where the convex hull for all points $(p, S_0(i_p, j_p))$ is tangent to a straight line of slope a at vertex $(r, S_0(i_r, j_r))$

If $q \leq r$, then

$$S_a(i_r, j_r) \geq S_a(i_q, j_q) \Leftrightarrow a \leq \frac{S_0(i_r, j_r) - S_0(i_q, j_q)}{r - q}$$

Example of the convex hull





Algorithm to construct the data structure

1. Enumerate all vertices of T

   $O(nq)$ time using q-time queries of $\alpha(i, j)$ and $\kappa(i, j)$

2. Determine the convex hull for all points $(p, S_0(i_p, j_p))$

   $O(n)$ time

Redefinition of $\alpha(i,j)$ and $\kappa(i,j)$

$\alpha(i,j)$: Minimum a having some k with $i \leq k \leq j$
such that $S_a(i,k) = 0$ or $S_a(k,j) = 0$

⇕

$\delta(i,j)$: Density $S_0(i,j)/(j-i+1)$ of $X_0(i,j)$

$\kappa'(i,j)$: Any k with $i \leq k \leq j$ minimizing $\delta(i,k)$

$\kappa''(i,j)$: Any k with $i \leq k \leq j$ minimizing $\delta(k,j)$

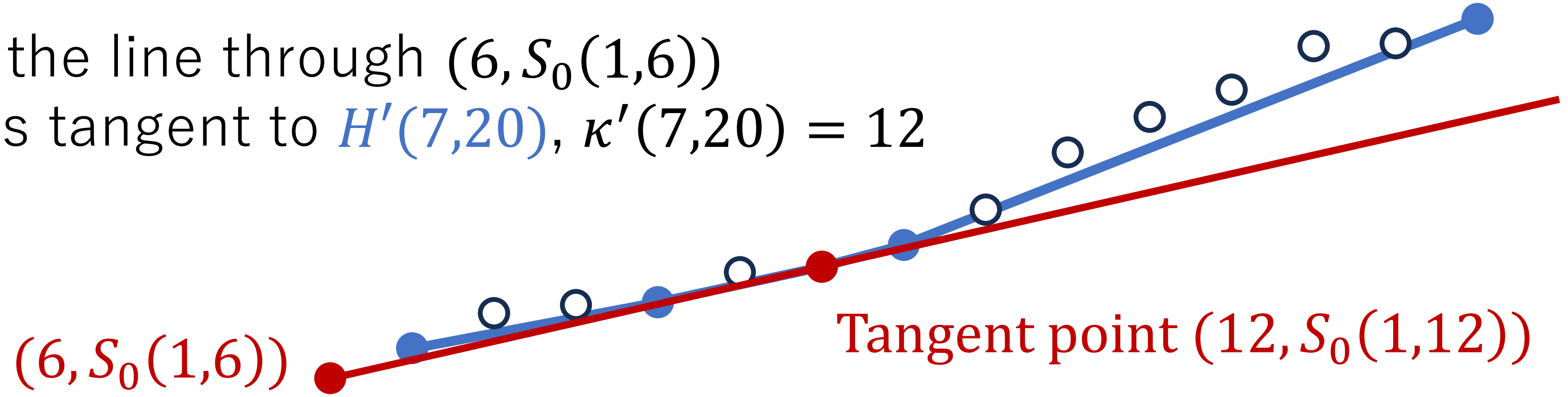
$\alpha(i,j)$: Minimum of $\delta(i,\kappa'(i,j))$ and $\delta(\kappa''(i,j),j)$

Basic idea for determining $\kappa'(i,j)$

$H'(i,j)$: Convex hull for points $(k, S_0(1,k))$ with $i \le k \le j$

If a line passing through point $(i-1, S_0(1, i-1))$ is tangent to $H'(i,j)$ at vertex $(k, S_0(1,k))$, then $\kappa'(i,j) = k$

Ex: Due to the line through $(6, S_0(1,6))$
that is tangent to $H'(7,20)$, $\kappa'(7,20) = 12$

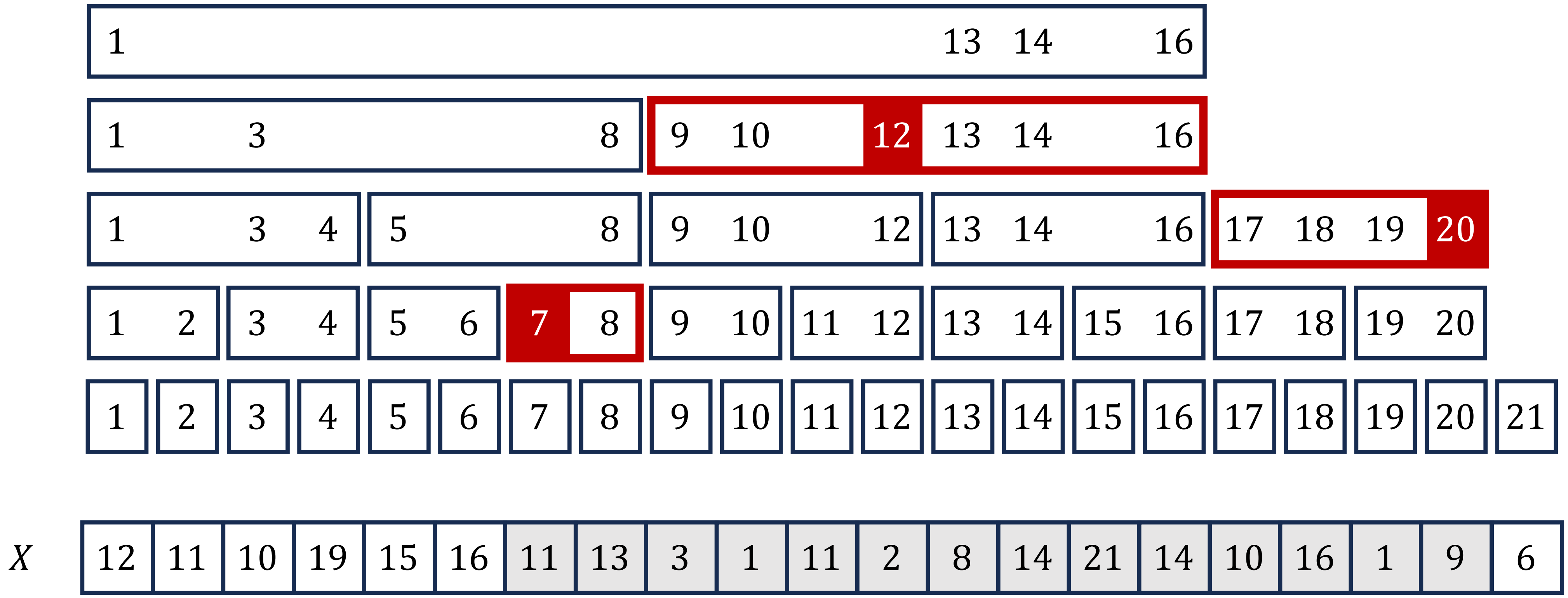


	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
X	12	11	10	19	15	16	11	13	3	1	11	2	8	14	21	14	10	16	1	9	6

$O(\log^2 n)$ query-time using $O(n \log n)$ space

Storing all $H'[l, m]$ with $1 \leq l \leq \log_2 n$ and $1 \leq m \leq n/2^l$,
where $H'[l, m]$ denotes $H'(2^l(m-1)+1, 2^l m)$

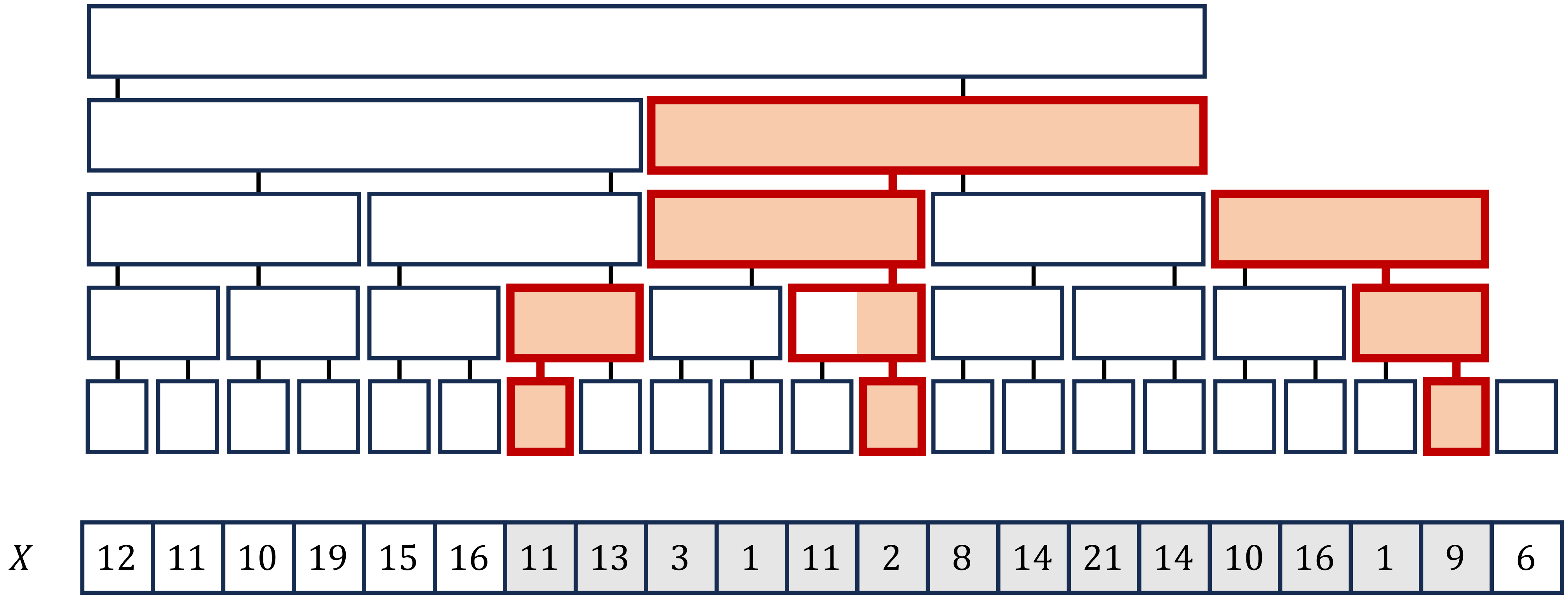
Ex:



Reduction of space to $O(n)$

Representing $H'[l, m]$
based on $H'[l-1, 2m-1]$ and $H'[l-1, 2m]$ recursively

Ex:



Conclusion

Given a numerical sequence X and a real number a,
the offset-MSS problem is to find an MSS of X_a

⬇

Number-line partition by a with the same MSS

- ✓ $O(n \log^2 n)$-time, $O(n)$-space constructible for any X of length n
- ✓ Supporting $O(\log n)$-time queries of an MSS of X_a for any a

$O(n)$-time constructible?