Orthogonal Range Searching

Sándor Kisfaludi-Bak

Computaional Geometry Summer semester 2020



• Intro and problem definition

• Intro and problem definition

• Kd trees

• Intro and problem definition

• Kd trees

• Range trees

• Intro and problem definition

• Kd trees

• Range trees

• Fractional cascading

• Intro and problem definition

• Kd trees

• Range trees

• Fractional cascading

• Priority search trees



Database of cakes.

Which cakes have

- sugar content [0.12,0.17]
- cocoa content [0.05,0.1]
- and sold btw. 3 and 4 tons last year?







Task: support such queries efficiently

Problem definition

Given n points in \mathbb{Z}^d or \mathbb{Q}^d ,

- 1. Preprocess them in O(n) time and space to
- 2. support orthogonal range queries in $poly(\log n) + O(k)$

3. on a Word RAM.

Problem definition

Given n points in \mathbb{Z}^d or \mathbb{Q}^d ,

- 1. Preprocess them in O(n) time and space to
- 2. support orthogonal range queries in $poly(\log n) + O(k)$

3. on a Word RAM.

output size

Problem definition

Given n points in \mathbb{Z}^d or \mathbb{Q}^d ,

- 1. Preprocess them in O(n) time and space to
- 2. support orthogonal range queries in $poly(\log n) + O(k)$

3. on a Word RAM.

output size

Static: preprocess and answer queries

Dynamic: update insertions and deletions in poly(log(n))

Query: [x, x']

Query: [x, x']

Option 1. Use sorted array: Binary search for xReport next until exceeds x'

Query: [x, x']

Option 1. Use sorted array: Binary search for xReport next until exceeds x'

Static only

Query: [x, x']

Option 1. Use sorted array: Binary search for xReport next until exceeds x'

Option 2. Binary search tree:



Static only

Query: [x, x']

Option 1. Use sorted array: Binary search for xReport next until exceeds x'

Option 2. Binary search tree:

Static only

P in the leaves Query: [4, 11]



inner vertex = largest value in left child's subtree

Answering a query

Binary search for Split(x, x')

Search for x, reporting right child subtrees Search for x', reporting left child subtrees



Answering a query

Binary search for Split(x, x')Search for x, reporting right child subtrees Search for x', reporting left child subtrees



Answering a query

Binary search for Split(x, x')Search for x, reporting right child subtrees Search for x', reporting left child subtrees



 $\mathsf{Space} = O(n)$

$$\mathsf{Preprocess} = O(n \log n)$$

Answering a query

Binary search for Split(x, x')Search for x, reporting right child subtrees Search for x', reporting left child subtrees



Space = O(n)Preprocess $= O(n \log n)$ Update $= O(\log n)$

Answering a query

Binary search for Split(x, x')Search for x, reporting right child subtrees Search for x', reporting left child subtrees



Space = O(n)Preprocess = $O(n \log n)$ Update = $O(\log n)$ Query = $O(\log n + k)$ Kd trees Bentley 1975

Assume distinct x- and y-coords

Assume distinct x- and y-coords



Assume distinct x- and y-coords



Assume distinct x- and y-coords



Assume distinct x- and y-coords



Assume distinct x- and y-coords



Assume distinct x- and y-coords



Kd tree anatomy

Space: O(n)Tree has $O(\log n)$ depth.

Preprocessing: use linear time median: $T(n) = 2T(n/2) + O(n) \Rightarrow O(n \log n)$

Kd tree anatomy

Space: O(n)Tree has $O(\log n)$ depth.

Preprocessing: use linear time median: $T(n) = 2T(n/2) + O(n) \Rightarrow O(n \log n)$



Each ℓ has a rectangular region

Kd tree anatomy

Space: O(n)Tree has $O(\log n)$ depth.

Preprocessing: use linear time median: $T(n) = 2T(n/2) + O(n) \Rightarrow O(n \log n)$



Each ℓ has a rectangular region

$$reg(\ell_1) = \mathbb{R}^2$$

Child regions of ℓ :
separated by ℓ

Querying a kd tree



Querying a kd tree



Report subtree if $reg(\ell) \subseteq Q$
Querying a kd tree



Report subtree if $reg(\ell) \subseteq Q$

If $reg(\ell)$ intersects ∂Q : check!

Querying a kd tree



Report subtree if $reg(\ell) \subseteq Q$

If $reg(\ell)$ intersects ∂Q : check!

Searching, reporting covered regions: $O(\log n + k)$

Querying a kd tree



Report subtree if $reg(\ell) \subseteq Q$

If $reg(\ell)$ intersects ∂Q : check!

Searching, reporting covered regions: $O(\log n + k)$

How many regions can intersect ∂Q ?

Regions interseted by the boundary

Claim. A vertical line intersects at most $O(\sqrt{n})$ node regions.



I(n): vert. line intersects this many regions in kd tree of size n

Regions interseted by the boundary

Claim. A vertical line intersects at most $O(\sqrt{n})$ node regions.



Regions interseted by the boundary

Claim. A vertical line intersects at most $O(\sqrt{n})$ node regions.



 $I(n) = 2 + 2I(n/4) = 2 + 2 \cdot (2 + 2 \cdot (\dots)) = 2 + 4 + \dots + 2^{\log_4 n}$

 $I(n) = 2 + 2I(n/4) = 2 + 2 \cdot (2 + 2 \cdot (\dots)) = 2 + 4 + \dots + 2^{\log_4 n}$ $\Rightarrow I(n) = O(\sqrt{n})$

$$I(n) = 2 + 2I(n/4) = 2 + 2 \cdot (2 + 2 \cdot (\dots)) = 2 + 4 + \dots + 2^{\log_4 n}$$
$$\Rightarrow I(n) = O(\sqrt{n})$$

Q has ≤ 2 vertical and ≤ 2 horizontal sides $\Rightarrow \partial Q$ intersects $O(\sqrt{n})$ regions

$$I(n) = 2 + 2I(n/4) = 2 + 2 \cdot (2 + 2 \cdot (\dots)) = 2 + 4 + \dots + 2^{\log_4 n}$$
$$\Rightarrow I(n) = O(\sqrt{n})$$

Q has ≤ 2 vertical and ≤ 2 horizontal sides $\Rightarrow \partial Q$ intersects $O(\sqrt{n})$ regions

Orthogonal range queries of a 2-dim kd tree take $O(\sqrt{n} + k)$ time.

Range trees

Binary search tree on x-xoordinates Each node has a new BST for y-coords of descendant leaves



T(v) Auxiliary tree

BST on descendants of v sorted for y-coord.

Binary search tree on x-xoordinates Each node has a new BST for y-coords of descendant leaves



T(v) Auxiliary tree

BST on descendants of v sorted for y-coord.

Construction:

• Presort on y-coord to array A

Construct(v, A):

- Construct T(v) using A
- Find median x, split A to A_{left}, A_{right}
- Construct($lc(v), A_{left}$), Construct($rc(v), A_{right}$)

Binary search tree on x-xoordinates Each node has a new BST for y-coords of descendant leaves



T(v) Auxiliary tree

BST on descendants of v sorted for y-coord.

O(n)

Construction:

- Presort on y-coord to array A $O(n \log n)$ Construct(v, A):
 - Construct T(v) using A
 - Find median x, split A to A_{left}, A_{right} O(n)
 - Construct($lc(v), A_{left}$), Construct($rc(v), A_{right}$)

Binary search tree on x-xoordinates Each node has a new BST for y-coords of descendant leaves



T(v) Auxiliary tree

 BST on descendants of vsorted for *y*-coord.

 $O(n \log n)$

 $O(n \log n)$

Construction:

• Presort on y-coord to array A

Construct(v, A):

- Construct T(v) using A
- O(n)• Find median x, split A to $A_{left}, A_{right} O(n)$
- Construct $(lc(v), A_{left})$, Construct $(rc(v), A_{right})$



Do 1-dim range query on x-coordiantes Find $O(\log n)$ reported subtree roots



Do 1-dim range query on x-coordiantes Find $O(\log n)$ reported subtree roots

Do 1-dim query on each of the $O(\log n)$ selected y-trees



Do 1-dim range query on x-coordiantes Find $O(\log n)$ reported subtree roots

Do 1-dim query on each of the $O(\log n)$ selected y-trees

Total query time:

$$\sum_{v} O(\log n + k_v) = O(\log^2 n + k)$$



Do 1-dim range query on x-coordiantes Find $O(\log n)$ reported subtree roots

Do 1-dim query on each of the $O(\log n)$ selected y-trees

Total query time:

$$\sum_{v} O(\log n + k_v) = O(\log^2 n + k)$$

In \mathbb{R}^d , it gives $O(\log^d n + k)$



Do 1-dim range query on x-coordiantes Find $O(\log n)$ reported subtree roots

Do 1-dim query on each of the $O(\log n)$ selected y-trees

Each level stores chunks of a (d-1)-dim range tree

Total query time:

$$\sum_{v} O(\log n + k_v) = O(\log^2 n + k)$$

In \mathbb{R}^d , it gives $O(\log^d n + k) \Big(\mathsf{But! space is up to } \Theta(n \log^{d-1} n) \Big)$

Composite numbers: $a, b \in \mathbb{R} \to (a|b)$ Sorted lexicographically.

Composite numbers: $a, b \in \mathbb{R} \rightarrow (a|b)$ Sorted lexicographically.

repaice p=(x,y) with $p^*=((x|y),(y|x))$ for each $p\in P$

 \Rightarrow all first and second coords are distinct

Composite numbers: $a, b \in \mathbb{R} \rightarrow (a|b)$ Sorted lexicographically.

repaice p=(x,y) with $p^*=((x|y),(y|x))$ for each $p\in P$

 \Rightarrow all first and second coords are distinct

Given query $Q = [x, x'] \times [y, y']$, replace by $Q^* = [(x|-\infty), (x'|\infty)] \times [(y|-\infty), (y'|\infty)]$

Composite numbers: $a, b \in \mathbb{R} \rightarrow (a|b)$ Sorted lexicographically.

repaice p=(x,y) with $p^*=((x|y),(y|x))$ for each $p\in P$

 \Rightarrow all first and second coords are distinct

Given query $Q = [x, x'] \times [y, y']$, replace by $Q^* = [(x|-\infty), (x'|\infty)] \times [(y|-\infty), (y'|\infty)]$

Observation.

$$p \in Q \Leftrightarrow p^* \in Q^*$$

(Special) Fractional cascading Lueker 1978, Willard 1978 Chazelle and Guibas 1986

Array and subarray, query [4, 45] \Rightarrow two binary searhces?

$$A \quad \boxed{2 \ 3 \ 5 \ 8 \ 13 \ 21 \ 34 \ 55 \ 89 \ 92 \ 95 \ 99}$$

$$B \ \ 2 \ \ 8 \ 13 \ 34 \ 55 \ 92 \ 95$$

Array and subarray, query [4, 45] \Rightarrow two binary searhces?

$$A \quad \boxed{2 \ 3 \ 5 \ 8 \ 13 \ 21 \ 34 \ 55 \ 89 \ 92 \ 95 \ 99}$$

$B \ \ 2 \ \ 8 \ \ 13 \ \ 34 \ \ 55 \ \ 92 \ \ 95$

Add pointer from A[x] to the smallest in B larger than A[x]

Array and subarray, query [4, 45] \Rightarrow two binary searhces?



Add pointer from A[x] to the smallest in B larger than A[x]

Array and subarray, query [4, 45] \Rightarrow two binary searhces?



Add pointer from A[x] to the smallest in B larger than A[x]Binary search A, use pointer of smallest reported

Idea:

Replace 2nd level BSTs for y-coords w/ sorted arrays A_v + 2 pointers per entry

Idea:

Replace 2nd level BSTs for y-coords w/ sorted arrays A_v + 2 pointers per entry

At node v, array A_v has:

• Points in decendant leafs of v sorted by y-coords

Idea:

Replace 2nd level BSTs for y-coords w/ sorted arrays A_v + 2 pointers per entry

- Points in decendant leafs of v sorted by y-coords
- Each entry $(a,b) \in A_v$ has:
 - Pointer to entry in $A_{lc(v)}$ with smallest y-coord $\geq b$
 - Pointer to entry in $A_{rc(v)}$ with smallest y-coord $\geq b$

Idea:

Replace 2nd level BSTs for y-coords w/ sorted arrays A_v + 2 pointers per entry

- Points in decendant leafs of v sorted by y-coords
- Each entry $(a,b) \in A_v$ has:
 - Pointer to entry in $A_{lc(v)}$ with smallest y-coord $\geq b$
 - Pointer to entry in $A_{rc(v)}$ with smallest y-coord $\geq b$

Idea:

Replace 2nd level BSTs for y-coords w/ sorted arrays A_v + 2 pointers per entry

- Points in decendant leafs of v sorted by y-coords
- Each entry $(a,b) \in A_v$ has:
 - Pointer to entry in $A_{lc(v)}$ with smallest y-coord $\geq b$
 - Pointer to entry in $A_{rc(v)}$ with smallest y-coord $\geq b$



Idea:

Replace 2nd level BSTs for y-coords w/ sorted arrays A_v + 2 pointers per entry

- Points in decendant leafs of v sorted by y-coords
- Each entry $(a,b) \in A_v$ has:
 - Pointer to entry in $A_{lc(v)}$ with smallest y-coord $\geq b$
 - Pointer to entry in $A_{rc(v)}$ with smallest y-coord $\geq b$



Idea:

Replace 2nd level BSTs for y-coords w/ sorted arrays A_v + 2 pointers per entry

- Points in decendant leafs of v sorted by y-coords
- Each entry $(a,b) \in A_v$ has:
 - Pointer to entry in $A_{lc(v)}$ with smallest y-coord $\geq b$
 - Pointer to entry in $A_{rc(v)}$ with smallest y-coord $\geq b$


Querying a layered range tree Query: $[x, x'] \times [y, y']$

Search top (x-)tree for x and x'. At v = Split(x, x'), binary search for y in $A_v \rightarrow A_v.find(y)$ Querying a layered range tree Query: $[x, x'] \times [y, y']$

Search top (x-)tree for x and x'. At v = Split(x, x'), binary search for y in $A_v \rightarrow A_v.find(y)$

Stepping to lc(v), follow left pointer from $A_v.find(y)$

Querying a layered range tree Query: $[x, x'] \times [y, y']$

Search top (x-)tree for x and x'. At v = Split(x, x'), binary search for y in $A_v \rightarrow A_v.find(y)$

Stepping to lc(v), follow left pointer from $A_v.find(y)$

 \Rightarrow Maintaining position in $A_{lc(v)}$ at each step takes O(1) time!

If w is root of a selected subtree: Reporting from T(w) can be done from A_w in $O(1 + k_w)$ time. Querying a layered range tree Query: $[x, x'] \times [y, y']$

Search top (x-)tree for x and x'. At v = Split(x, x'), binary search for y in $A_v \rightarrow A_v.find(y)$

Stepping to lc(v), follow left pointer from $A_v.find(y)$

 \Rightarrow Maintaining position in $A_{lc(v)}$ at each step takes O(1) time!

If w is root of a selected subtree: Reporting from T(w) can be done from A_w in $O(1 + k_w)$ time.

Total query time: $O(\log n + k)$

Priority search trees McCreight 1985

0

Construction:

• root r stores point with max y-coord $med_x(root) := median \ x$ -coordinate of $P \setminus \{r\}$

- root r stores point with max y-coord $med_x(root) := median \ x$ -coordinate of $P \setminus \{r\}$
- left subtree: points $p \in P \setminus \{r\}$ with $p_x \leq med_x(r)$



- root r stores point with max y-coord $med_x(root) := median \ x$ -coordinate of $P \setminus \{r\}$
- left subtree: points $p \in P \setminus \{r\}$ with $p_x \leq med_x(r)$
- right subtree: points $p \in P \setminus \{r\}$ with $p_x > med_x(r)$



- root r stores point with max y-coord $med_x(root) := median \ x$ -coordinate of $P \setminus \{r\}$
- left subtree: points $p \in P \setminus \{r\}$ with $p_x \leq med_x(r)$
- right subtree: points $p \in P \setminus \{r\}$ with $p_x > med_x(r)$



- root r stores point with max y-coord $med_x(root) := median \ x$ -coordinate of $P \setminus \{r\}$
- left subtree: points $p \in P \setminus \{r\}$ with $p_x \leq med_x(r)$
- right subtree: points $p \in P \setminus \{r\}$ with $p_x > med_x(r)$



- root r stores point with max y-coord $med_x(root) := median \ x$ -coordinate of $P \setminus \{r\}$
- left subtree: points $p \in P \setminus \{r\}$ with $p_x \leq med_x(r)$
- right subtree: points $p \in P \setminus \{r\}$ with $p_x > med_x(r)$



Answering 3-sided queries Wlog queries of the type $[x, x'] \times [y, \infty]$



Answering 3-sided queries Wlog queries of the type $[x, x'] \times [y, \infty]$

To answer the query from root = v:





Answering 3-sided queries Wlog queries of the type $[x, x'] \times [y, \infty]$

To answer the query from root = v:





+ check each node stored on the search paths!

Answering 3-sided queries Wlog queries of the type $[x, x'] \times [y, \infty]$

To answer the query from root = v:





+ check each node stored on the search paths!

Query time: $O(\log n + k)$

Current best data structures

Assume coordinates fit in machine words on w bits. Multiply all by $\log w \simeq \log \log n$.



Chan-Larsen-Pătrașcu

Chan–Tsakalidis



 \rightarrow intervals ending in Q: range searching on endpoints

 \rightarrow intervals "crossing" Q?



 \rightarrow intervals ending in Q: range searching on endpoints

 \rightarrow intervals "crossing" Q? if horizontal, they cross left boundary of Q



 \rightarrow intervals ending in Q: range searching on endpoints

 \rightarrow intervals "crossing" Q? if horizontal, they cross left boundary of Q

Given a vertical query segment, report all horizontal segments it is crossing.



 \rightarrow intervals ending in Q: range searching on endpoints

 \rightarrow intervals "crossing" Q? if horizontal, they cross left boundary of Q

Given a vertical query segment, report all horizontal segments it is crossing.



 \rightarrow intervals ending in Q: range searching on endpoints

 \rightarrow intervals "crossing" Q? if horizontal, they cross left boundary of Q

Given a vertical query segment, report all horizontal segments it is crossing.

 \Leftrightarrow in \mathbb{R}^1 , which intervals contain query point q?

Interval trees (Edelsbrunner 1980 McCreight 1980)

root has intervals intersected by ℓ sorted for left endpoints in a list (same for right)



Interval trees (Edelsbrunner 1980 McCreight 1980)

root has intervals intersected by ℓ sorted for left endpoints in a list (same for right)

Preprocess: $O(n \log n)$ Space: O(n)Query (\mathbb{R}^1): $O(\log n + k)$



Interval trees (Edelsbrunner 1980 McCreight 1980)

root has intervals intersected by ℓ sorted for left endpoints in a list (same for right)

Preprocess: $O(n \log n)$ Space: O(n)Query (\mathbb{R}^1): $O(\log n + k)$

If segment query: use priority search tree instead of list



