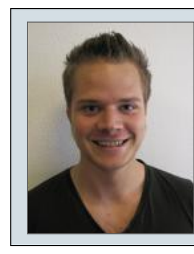
MADALGO
CENTER FOR MASSIVE DATA ALGORITHMICS

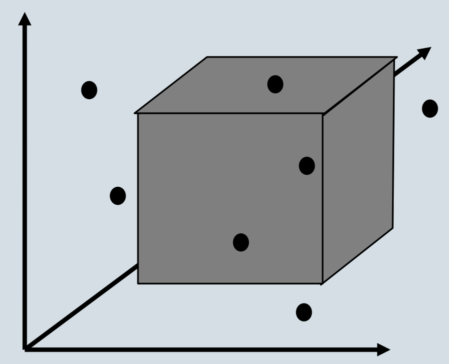


Kasper Dalgaard Larsen
Aarhus University

Orthogonal Range Reporting: Optimal 3-d Structures and Query Lower Bounds

Problem

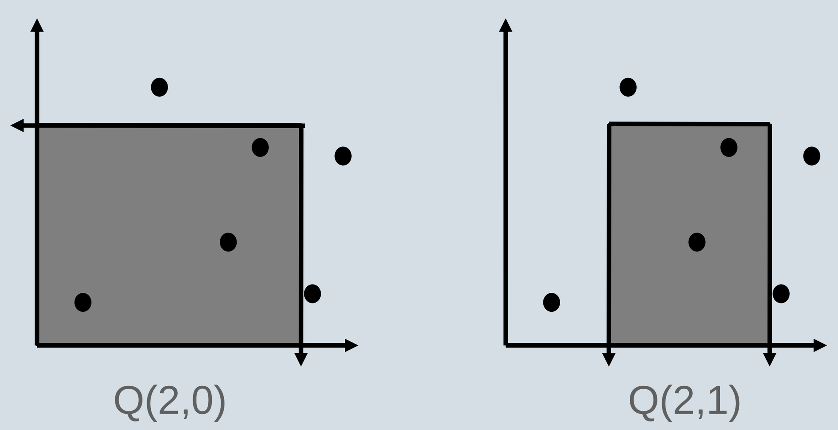
Input: n points in d dimensional space.



Support reporting the t points in a query box.

Variations

Query range may be unbounded in some dimensions:

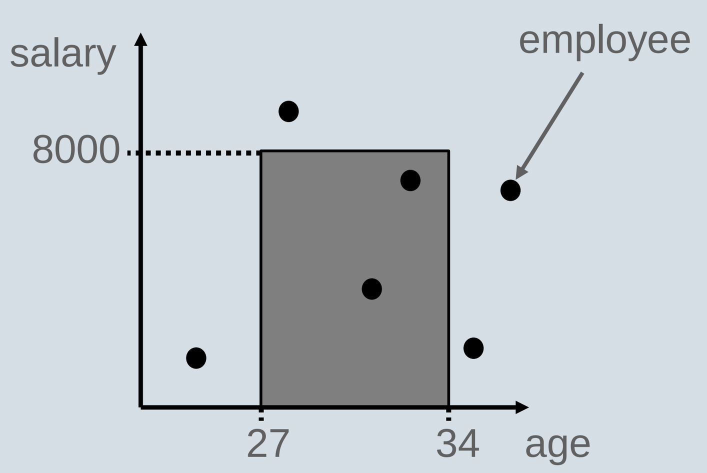


Q(d,k): d dimensional points, query ranges are finite in k dimensions.

Why important: more efficient solutions.

Motivation

Database queries:



"Give me all employees of age between 27 and 34, earning at most 8000 per month."

Results

The best previous result:

Type	Query	Space
Q(2,1)	**$\lg n+t$**	**n**
Q(2,2)	**$\lg n+t$**	**$n \lg n/\lg\lg n$**
Q(3,0)	**$\lg n+t$**	**n**
Q(3,1)	**$\lg n+t$**	**$n \lg n/\lg\lg n$**
Q(3,2)	**$\lg n+t$**	$n (\lg n/\lg\lg n)^2$
Q(3,2)	$\lg^2 n/\lg\lg n+t$	**$n \lg n/\lg\lg n$**
Q(3,3)	**$\lg n+t$**	$n (\lg n/\lg\lg n)^3$
Q(3,3)	$\lg^2 n/\lg\lg n+t$	**$n (\lg n/\lg\lg n)^2$**
Q(d,d)	$\lg n(\lg n/\lg\lg n)^{d-3}+t$	$n (\lg n/\lg\lg n)^d$
Q(d,d)	$\lg n(\lg n/\lg\lg n)^{d-2}+t$	**$n (\lg n/\lg\lg n)^{d-1}$**

Bounds in **bold** are optimal.
Space lower bound $\Omega(n(\lg n/\lg\lg n)^{d-1})$.
No query lower bound beyond $\Omega(\lg n+t)$.

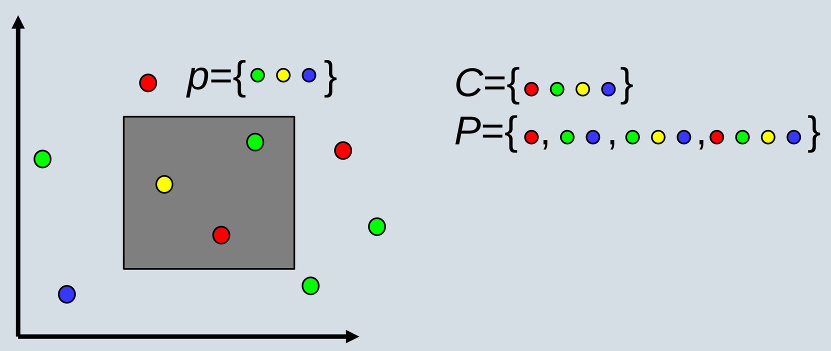
We achieve:

Type	Query	Space
Q(2,1)	**$\lg n+t$**	**n**
Q(2,2)	**$\lg n+t$**	**$n \lg n/\lg\lg n$**
Q(3,0)	**$\lg n+t$**	**n**
Q(3,1)	**$\lg n+t$**	**$n \lg n/\lg\lg n$**
Q(3,2)	**$\lg n+t$**	**$n \lg n/\lg\lg n$**
Q(3,3)	**$\lg n+t$**	**$n (\lg n/\lg\lg n)^2$**
Q(d,d)	$\lg n(\lg n/\lg\lg n)^{d-3}+t$	**$n (\lg n/\lg\lg n)^{d-1}$**

Bounds in **bold** are optimal.
We also show the query time lower bound of $\Omega((\lg n/\lg\lg n)^{d/2-1}+t)$.
Optimal solution for all 3-d variations.

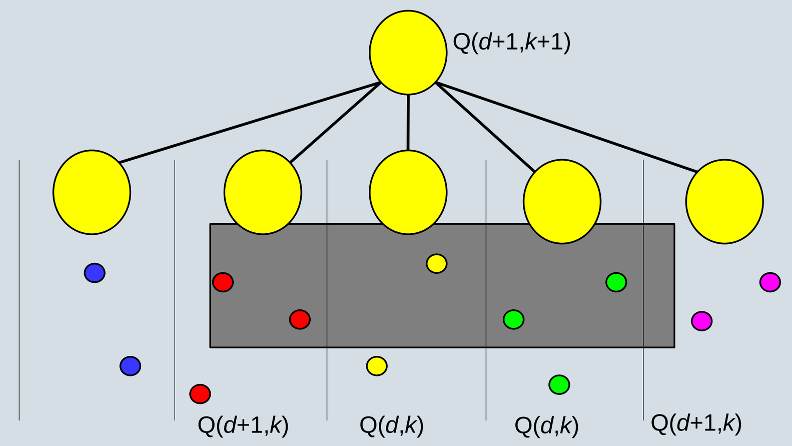
Concurrent Range Reporting

We solve harder variants of orthogonal range reporting. We call these concurrent Q(d,k) (CQ(d,k)).



Points have colors from set of colors C.
Query: Q(d,k) box q + set of colors p from a set $P \subseteq 2^C$.
Must report points in q with color in p.

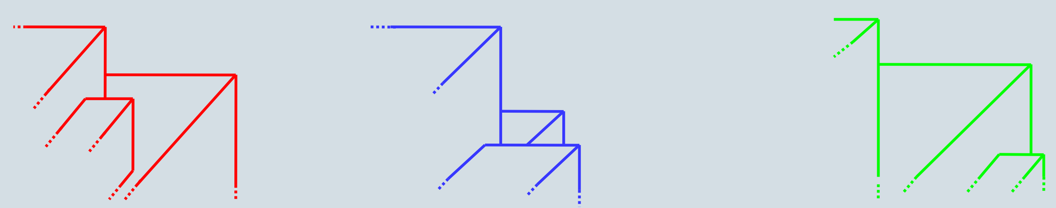
Solution for CQ(d,k) gives solution for Q(d,k+1) (and Q(d+1,k+1)) by paying $\lg n/\lg\lg n$ in space (and query time).



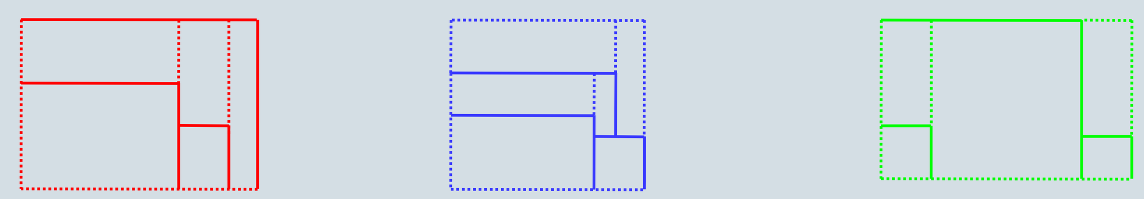
Range tree with fanout $\lg^\varepsilon n$.
Q(d+1,k+1) becomes CQ(d,k) with colors {• •}.
Height of tree $\lg n/\lg\lg n$.

Solving Concurrent Q(3,0)

$\lg n$-shallow cuttings: $O(n/\lg n)$ Q(3,0) boxes s.t. queries outputting < $\lg n$ points are inside at least one box. Each box contains $O(\lg n)$ input points.



Build $\lg n$-shallow cuttings for every color.
Store Q(3,0) structure inside each box, and for every color.



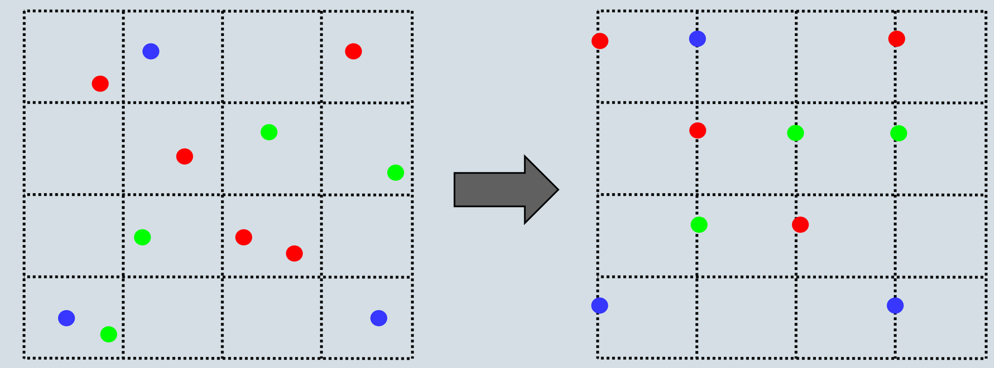
Given query q, find boxes containing it using 2-d point enclosure in time $O(\lg n+|C|)$.

For every color:

- If q inside a box b, query Q(3,0) structure inside b.
- Otherwise, query Q(3,0) structure on all points.

Solving Concurrent Q(3,2)

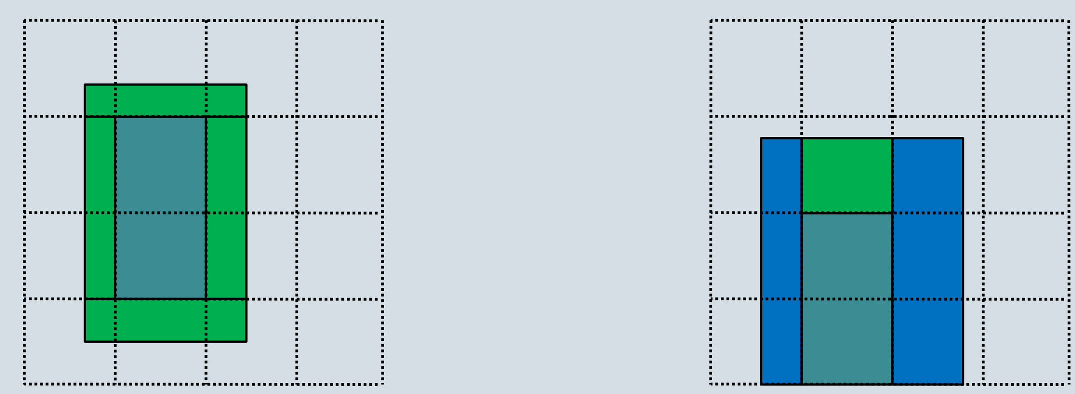
Query unbounded in z-dimension. Build grid on x and y dimensions.



The grid has size w by w where $w=\sqrt{n/|C||P|\lg^2 n}$.
n/w points in each horizontal and vertical slab.
Points snapped to grid.

Structures stored:

- CQ(3,0) structure on points in every slab.
- For every p in P, a Q(3,2) structure on points with least z coordinate from every color in every grid cell.
- Linked list for every color in every grid cell with points sorted by z coordinate.
- Recursive structure inside each slab.



Queries decompose into four CQ(3,1) queries and a grid query.
In recursive steps, they become one recursive CQ(3,1) query, two CQ(3,0) queries and one grid query.

MADALGO – Center for Massive Data Algorithmics, a Center of the Danish National Research Foundation