On the Average Performance of Orthogonal Range Search in Multidimensional Data Structures

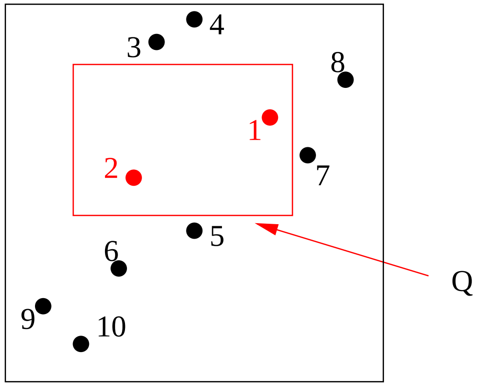
Amalia Duch and Conrado Martínez

Universitat Politècnica de Catalunya

ICALP'02. Málaga, Spain.

Orthogonal Range Search

Given a set of K-dimensional data points and a query hyperrectangle Q, the goal of an orthogonal range search is to report all the points in the collection which fall inside Q.



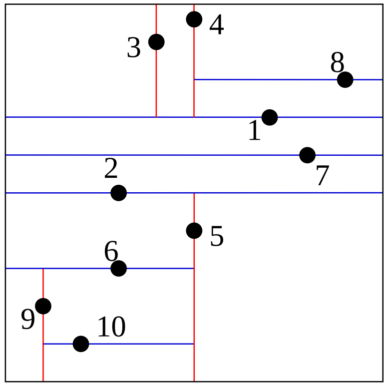
K=2

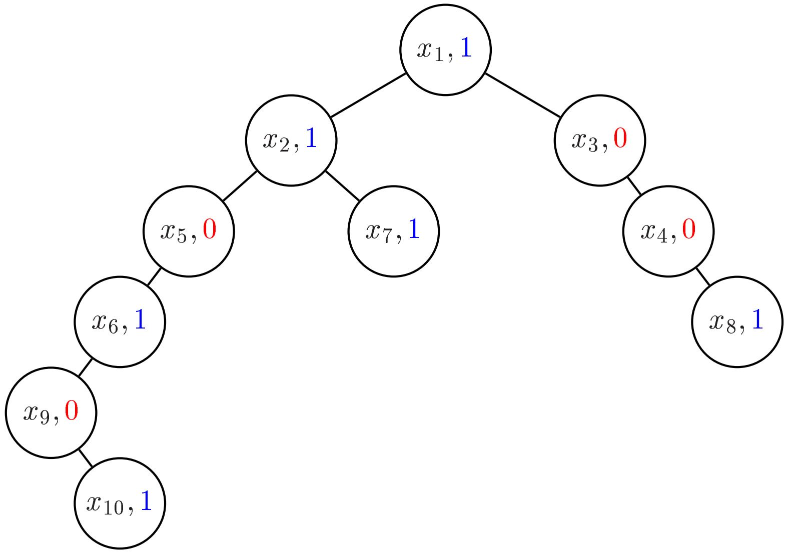
Data Structures

- Specific data structures: range trees, ...

- General data structures: K-d trees, quadtrees, quadtries, ...

Relaxed K-d Trees





The Algorithm

```
▷ Q = [ℓ_0, u_0] × ··· × [ℓ_{K-1}, u_{K-1}]
procedure range_search(Q : hyperrectangle, T : K-d tree)
    if T is empty then return
    else x ← root(T); j ← discriminant(T)
        if x_j ≥ ℓ_j then range_search(Q, left(T))
        if x_j ≤ u_j then range_search(Q, right(T))
        if x inside Q then report x
end
```

Previous Work

- [BS75] J. L. Bentley and D. F. Stanat, Analysis of Range Searches in Quadtrees.
- [Si79] Y. V. Silva-Filho, Average-case analysis of region search in balanced kd trees.
- [DJZ00] L. Devroye, J. Jabbour and C. Zamora-Cura, Squarish kd trees.
- [CDZ01] P. Chanzy, L. Devroye and C. Zamora-Cura, Analysis of range search for random kd trees.

The Randomness Model

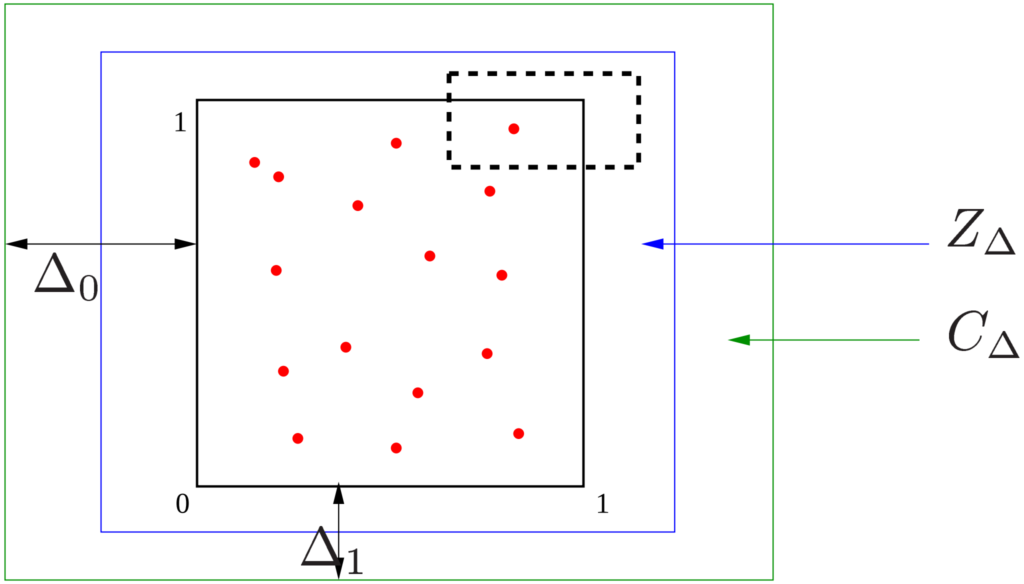
- Cost R_n = Number of visited nodes = Number of reported nodes + Overhead

- Random relaxed trees built by random insertions
  - Data points are independently drawn from some continuous distribution in $[0,1]^K$
  - Discriminants are independently and uniformly drawn from $\{0,\ldots,K-1\}$

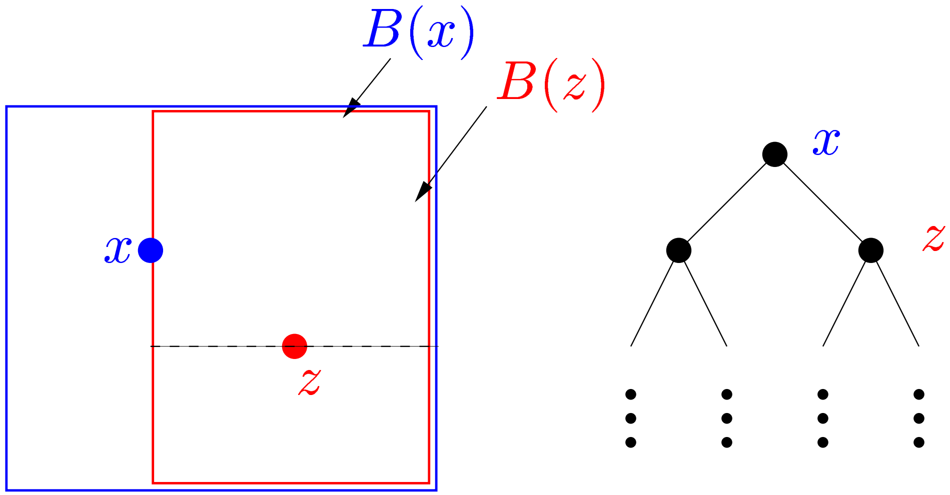
The Randomness Model (cont'd)

- Random queries have given lengths Δ_0, ... , Δ_{K-1}, with $0 < \Delta_i \leq 1/2$.
- The center of a random query is independently drawn from some continuous distribution in Z_Δ.
- Random queries may fall partially off $[0, 1]^K$

The Randomness Model (cont'd)



Bounding Rectangles



Lemma 1. *A point x with bounding rectangle $B(x)$ is visited by a range search with query hyperrectangle Q if and only if $B(x)$ intersects Q.*

Partial Matches

Given a query $q = (q_0, q_1, \ldots, q_{K-1})$, with $q_i \in [-\Delta_i, 1+\Delta_i] \cup \{*\}$ the goal of partial match is to report the points in the file that match the query, that is, all x in the file such that $x_i = q_i$ whenever $q_i \neq *$.

For a query $q = (q_0, q_1, \ldots, q_{K-1})$ its specification pattern is a bitstring $w = (w_0, w_1, \ldots, w_{K-1})$ such that $w_i = 0$ if $q_i = *$ and $w_i = 1$ if $q_i \neq *$.

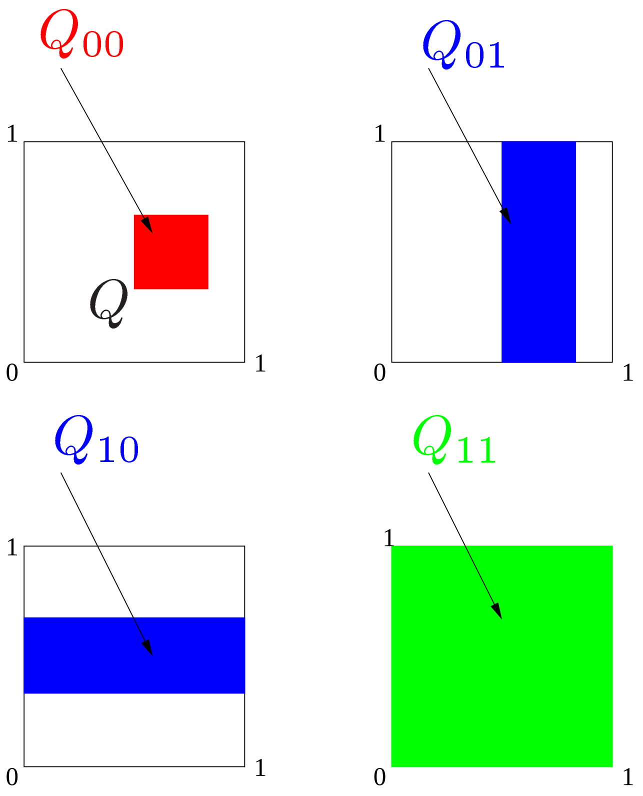
Ex: Find all points that satisfy the query $q = (*, 0.3, *, 0.1, 0.67)$. Here, $w(q) = (0, 1, 0, 0)$.

Partial Matches (cont'd)

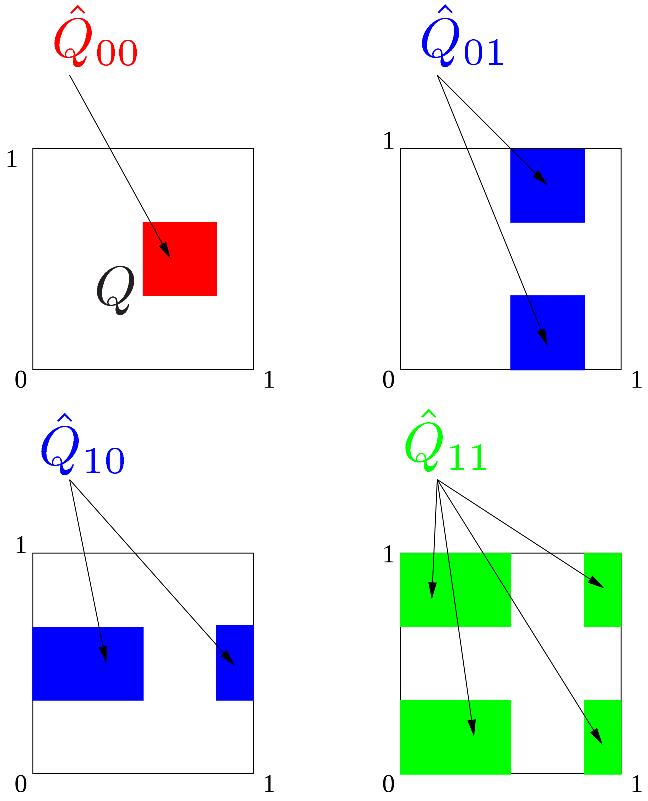
▷ $q = (q_0, q_1, \ldots, q_{K-1})$

procedure partial_match(q : pattern, T : K-d tree)
    if T is empty **then return**
    else $x \leftarrow$ root(T); $j \leftarrow$ discriminant(T)
        if $q_j = *$ **or** $x_j \geq q_j$ **then** partial_match(q, left(T))
        if $q_j = *$ **or** $x_j \leq q_j$ **then** partial_match(q, right(T))
        if x matches q **then** report x
end

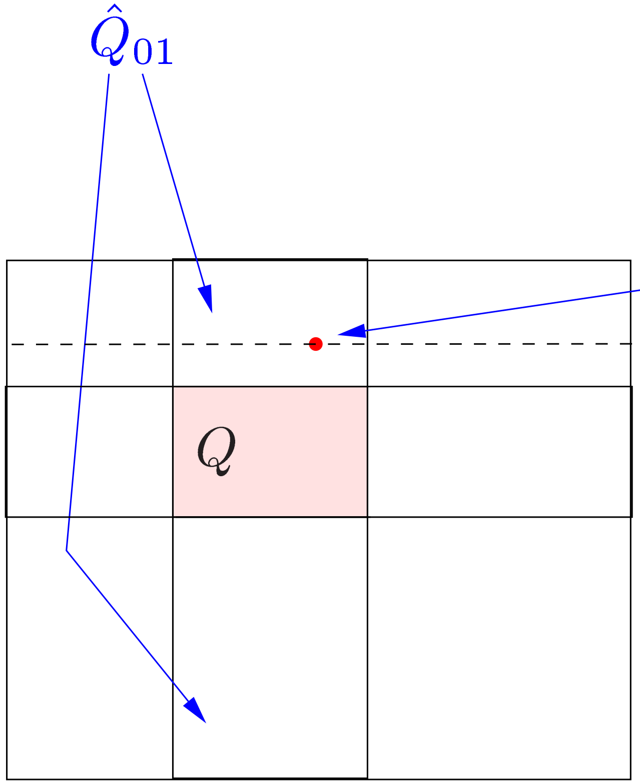
Slices & Proper Slices



Slices & Proper Slices (cont'd)

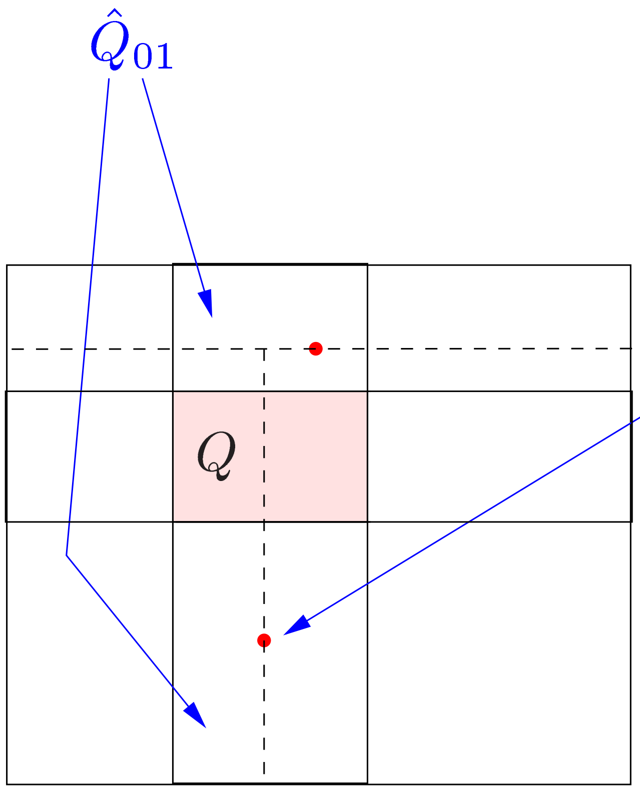


Slices & Proper Slices (cont'd)



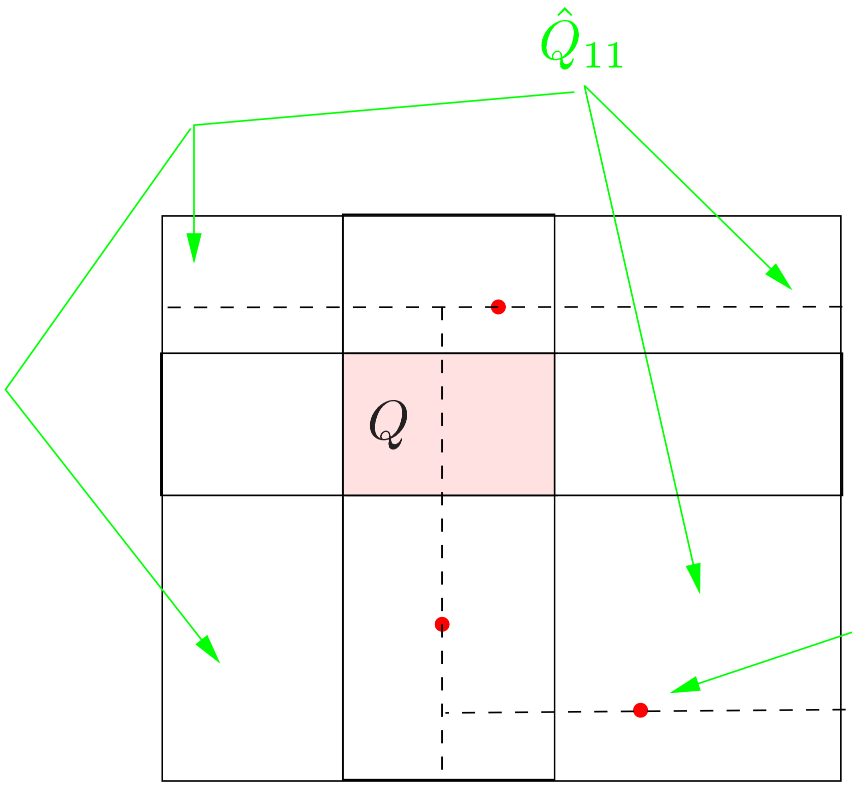
orthogonal range and partial match with $w = 01$ do the same here

Slices & Proper Slices (cont'd)



orthogonal range and partial match with $w = 01$ do the same here

Slices & Proper Slices (cont'd)



orthogonal range and partial match with $w = 11$ do the same here

Partial Match & Hyperplanes

$$H(y, w) = \{x \in C_\Delta \,|\, \forall i : w_i = 1 \implies x_i = y_i\}.$$

Notice $H(y, 00) = C_\Delta$, $H(y, 11) = \{y\}$.

Lemma 2. *A point x is visited by a partial match with query $q \equiv \langle y, w\rangle$ if and only if $H(y, w)$ intersects the bounding box $B(x)$ of x.*

Sliced Partial Matches

A sliced partial match with query $q \equiv \langle y, w \rangle$ is just a standard partial match where we only do count the points x visited by the partial match that are inside $\hat{Q}_w$.

An Upper Bound

Theorem 3. *Given a query Q with corners $v_0, v_1, \ldots, v_{2^K-1}$ and a K-d tree t, the cost $R(t)$ of the range search Q in the tree t satisfies*

$$R(t) \leq \sum_{0 \leq j < 2^k} \sum_{w \in (0+1)^k} P_w(t, v_j),$$

where $P_w(t, v_j)$ is the cost of a sliced partial match with query $q \equiv \langle v_j, w \rangle$.

An Upper Bound (cont'd)

Sketch. $B(x) \cap Q \neq \emptyset \wedge x \in \hat{Q}_w \implies B(x) \cap H(v,w) \neq \emptyset$for at least a corner v of Q