Algorithms and Data Structures:
Range Searching and Combinatorial Discrepancy

Kasper Green Larsen

Aarhus University

MADALGO

CENTER FOR MASSIVE DATA ALGORITHMICS

Danmarks Grundforskningsfond
Danish National Research Foundation

AARHUS UNIVERSITY

MADALGO – Who are we?

Faculty

Gerth

Lars

Post Docs

Elad

Brody

Lap Kei

Qin

Nodari

MADALGO – Who are we?

Administration


Ellen


Else

Programmer


Thor

PhD students


Casper


Jesper


Freek


Kasper


Kostas


Pooya


Mark


Jesper


Jakob


Morten


Lasse

This Talk

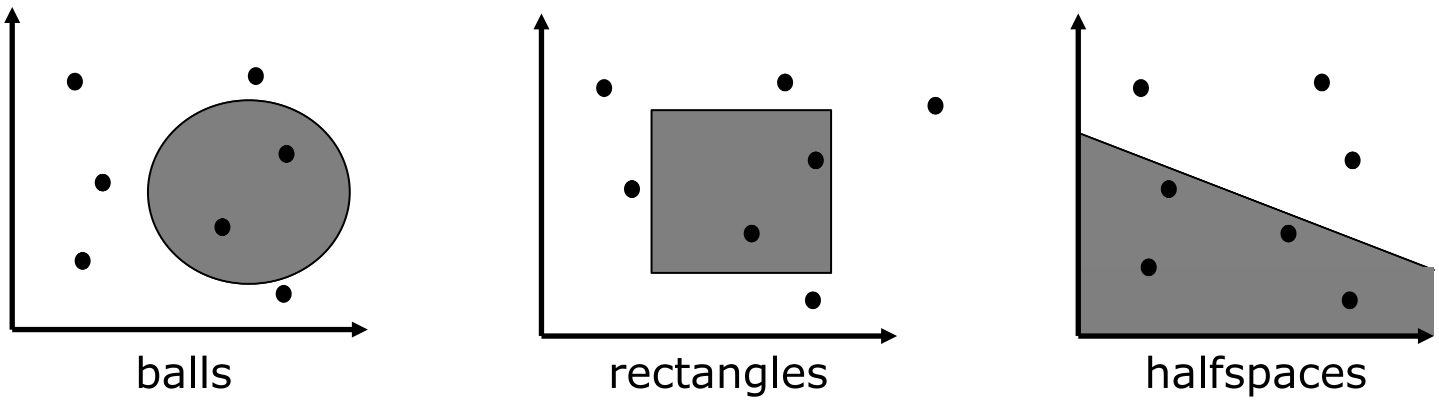
- Following submitted paper:

On Range Searching in the Group Model and Combinatorial Discrepancy.

Kasper Green Larsen.

Submitted to FOCS'11: 52nd IEEE Symposium on Foundations of Computer Science.

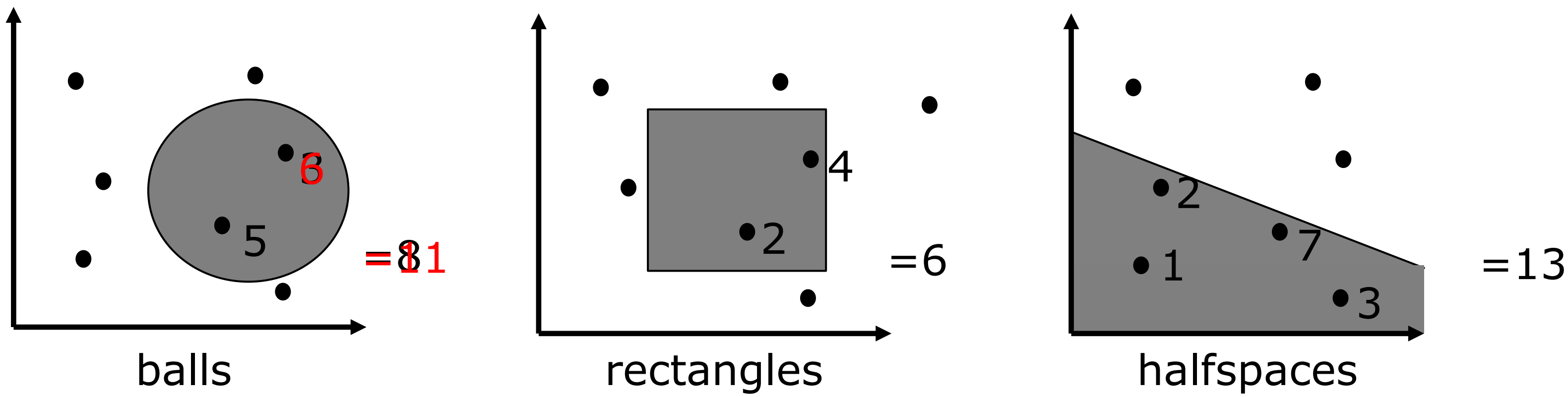
Range Searching



- Given n points.
- Build structure that aggregates information about the points inside a query range efficiently.

- Information Aggregated: Counting, Reporting, Emptiness, Sum of weights,…

Range Searching in the Group Model

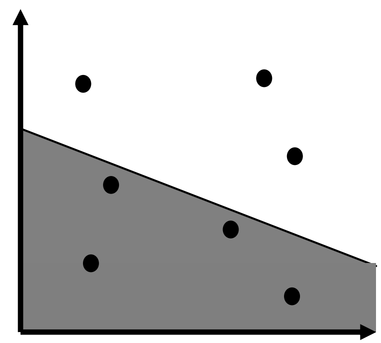


- Weights: The n points have a weight from a group.
- Query: Compute group sum of points in query range.
- Updates: Support changing the weight of a point.

- Data Structure: Collection of precomputed group sums. Answer query by adding and subtracting. Updates by re-evaluating.

Previous Results - Halfspaces

- Huge gap between upper and lower bounds:
  - Best upper bound:
    - Query time $t_q = O(n^{1/2})$
    - Update time $t_u = O(\lg\lg n)$
  - Highest lower bound for *any* problem:
    - $t_q = \Omega((\lg n/(\lg\lg n + \lg t_u))^2)$

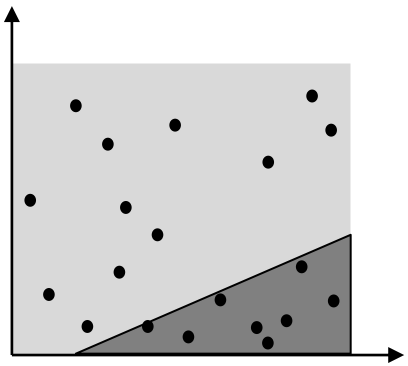


- The model was defined back in 1982!

- We prove:
  - $t_q t_u = \Omega(n^{1/2}/\lg n)$!

Discrepancy Theory - Halfspaces

- Discrepancy Theory:
  - Goal: Place n points in [0,1]x[0,1] as uniform as possible.



area: 0.2
Vol(r)*n: 4
$|r \cap P|$: 8
disc(P)≥4

  - Discrepancy: Measures uniformity, i.e.

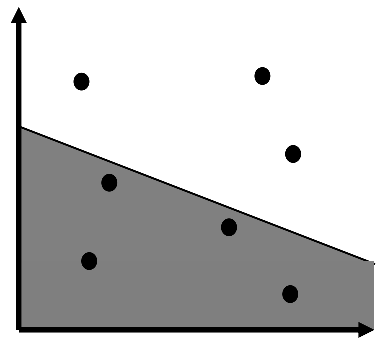
$$\text{disc}(P)=\max_r|\text{Vol}(r)*n-|r \cap P||$$

$$\text{disc}=\min_P \text{disc}(P)$$

- Example.

Discrepancy Theory - Halfspaces

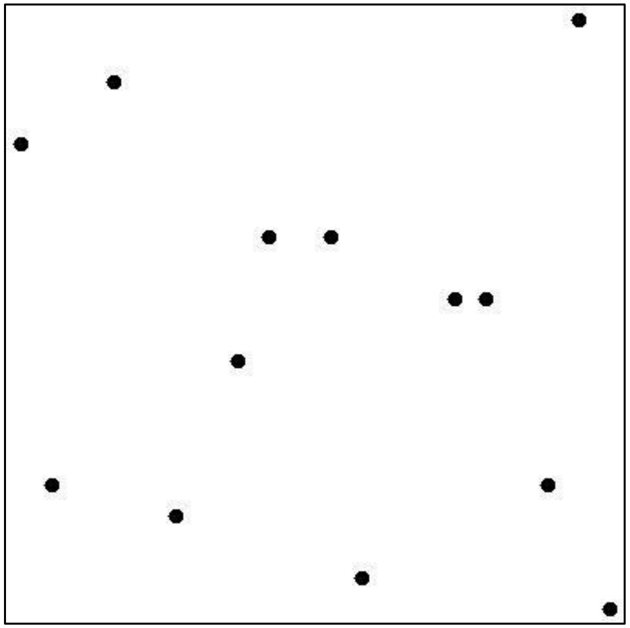
- Discrepancy Upper Bound results:
  - Halfspaces $O(n^{1/4})$.
- Discrepancy Lower Bounds well-understood:
  - Halfspaces $\Omega(n^{1/4})$.



- Our Result:
  - Intimate connection between the two fields:
    - $t_q t_u = \Omega(\text{disc}^2/\lg n)$.
- Implications:
  - $t_q t_u = \Omega(n^{1/2}/\lg n)$!
  - $\lg n \lg\lg n$ factor from upper bound!

Key Idea

- How we establish the connection:
  - Use efficient data structure to place points very uniformly.



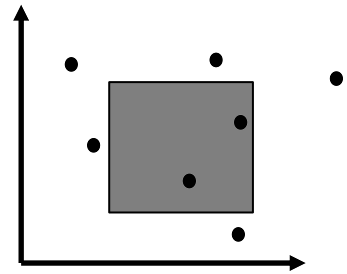
- To place *n* very uniform points: Place >>n points on grid.
  - Repeat until *n* points left:
    - Use structure to color red/blue very evenly. Remove red.

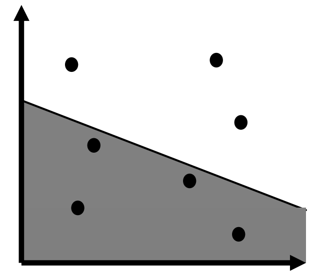
Key Idea

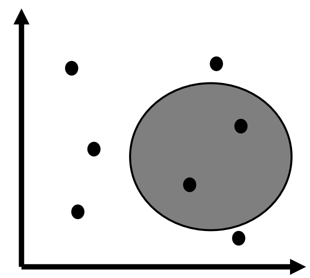
- We show:
  - If data has query time t_q and update time t_u then discrepancy of final point set is:
    - $O((t_q t_u \lg n)^{1/2})$
- We know that the discrepancy is at least disc, thus:
  - $t_q t_u = \Omega(\text{disc}^2/\lg n)$

Summary

- Our Result:
  - Intimate connection between the two fields:
    - $t_q t_u = \Omega(\text{disc}^2/\lg n)$.
- Immediate Implications:
  - Halfspaces $t_q t_u = \Omega(n^{1/2}/\lg n)$.
    - $\lg n \lg\lg n$ factor from best upper bound!
  - Rectangles $t_q t_u = \Omega(\lg^{\varepsilon} n)$.
  - Balls $t_q t_u = \Omega(n^{1/2}/\lg n)$.
- Also implications for discrepancy theory:
  - Rectangles disc=$O(\lg^{5/2} n)$.
  - Established using textbook range searching solutions.







Thank You