Distributed Localization and Clustering Using Data Correlation and the Occam's Razor Principle

Alon Efrat Chris Gniady
University of Arizona

Pankaj K. Agarwal
Duke University

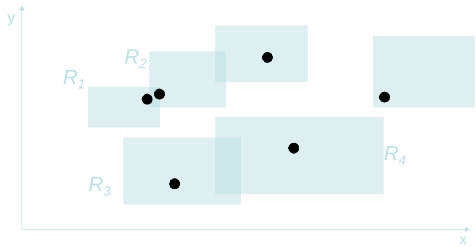
Joseph S. B. Mitchell
State University of New York at Stony Brook

Valentin Polishchuk
Helsinki Institute for Information Technology

Girishkumar R. Sabhnani
Metron Aviation Inc.

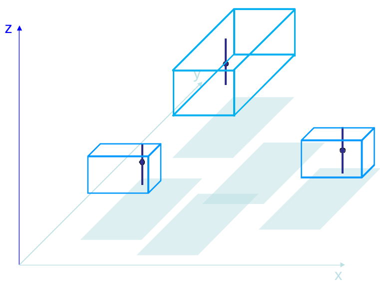
n sensors

- Location of s is not known exactly
  - region R_s in which s lies



Measurements

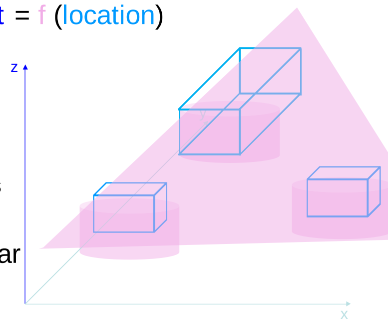
- z_s
  - error e_s



Regression

Measurement = f (location)
z = f (x,y)

f – consistent
(graph of) f
intersect all boxes
-- "simple"
(piecewise-)linear



Occam Razor's Principle

William of Ockham (ca. 1285-1349):

"Entities must not be multiplied beyond necessity"

In other words:

"Given several explanations to a phenomenon, choose the simplest one"

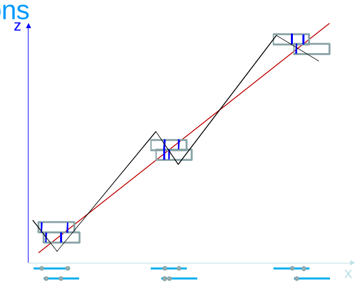
Regression: Know more about z

Our view: Know more also about (x,y)

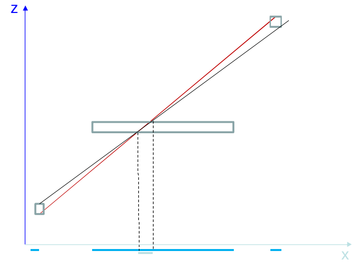
2D

Locations

Better locations



Improved localization

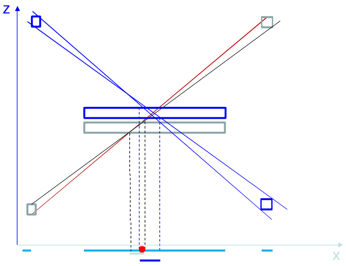


Readings from several days

Day 1

Day 2

Even better
localization



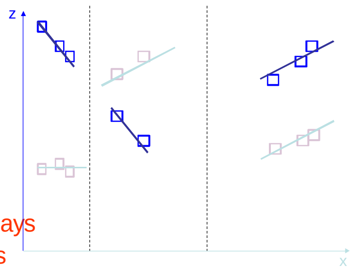
Clustering

Day 1

Day 2

Group sensors
within group:

Linear fit
on *every* day



Grouping
same on all days

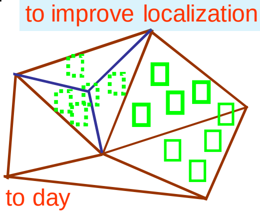
Linear functions
may be different from day to day

Distributed Localization and Clustering Using Data Correlation and the Occam's Razor Principle

3D

- Partition the plane into triangles
  (Occam: as few as possible)
- Sensors within a triangle:
  Linear fit
  on *every* day

At the same time:
Use linear functions
to improve localization



Grouping
same on all days

Linear functions
may be different from day to day

Related work

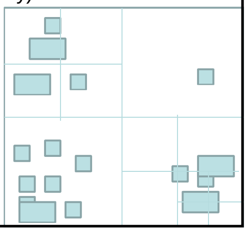
- Terrain simplification [Agarwal and Desikan'97, Agarwal and Suri'98]
  – NP-hard
  – log OPT - apx
- Angle-based localization [Basu, Gao, Mitchell, Sabhnani '06]
- SLAT, dictionary learning, GPS,…

Quad-Tree approach

- Basic problem (oracle):
  - Given a (sub)set S of sensors, find
    - locations p_s in R_s (one per sensor)
    - linear functions h^d (one per day)

  such that $| z^d_s - h^d(p_s) | \leq e_s$

  If impossible,
  split S into 4 quadrants
  recurse

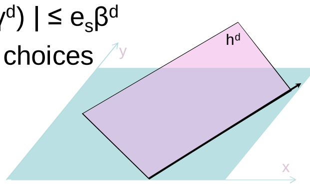


Basic problem

- Find p_s ,h^d s.t. $| z^d_s - h^d(p_s) | \leq e_s$
- $| z^d_s - (a^d x_s + b^d y_s + c^d) | \leq e_s$

  x_s , y_s , a^d, b^d, c^d – variables
- Reparametrizaion $\alpha^d = a^d/b^d$, $\beta^d = 1/b^d$, $\gamma^d = c^d/b^d$

  $| z^d_s \beta^d - (\alpha^d x_s + y_s + \gamma^d) | \leq e_s \beta^d$
- α^d : 32 fixed discrete choices
- LP $\rightarrow$ $\{p_s\}$ in R_s , $\{h^d\}$



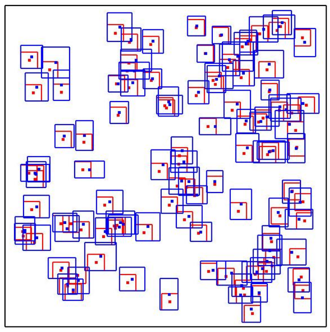
Randomized distributed algorithm

- Find (x_s,y_s) , β^d,γ^d such that
  $| z^d_s \beta^d - (\alpha^d x_s + y_s + \gamma^d) | \leq e_s \beta^d$, (x_s,y_s) in R_s
- Q_s = set of $(\beta^1, \gamma^1 , \beta^2, \gamma^2 ,\ldots,\beta^D ,\gamma^D)$ good for s
- LP is feasible iff $\bigcap Q_s$ is non-empty
- Sensors send to Q_s a leader
  - random set of D^2 sensors do
  - with high probability
    - Terminate in D log|S| rounds if LP feasible
    - LP is infeasible

[K. Clarkson,"Las Vegas algorithms for linear and integer programming when the dimension is small", *J. ACM*, 1995]

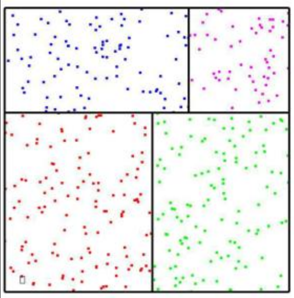
Localization

- 1 day
- 100 nodes
- R_s blue
- Red – improved



Clustering

True, original: 4 rooms



Output: color-coded clusters