A General Framework for Modeling and Processing Optimization Queries

Michael Gibas, Ning Zheng, Hakan Ferhatosmanoglu
Ohio State University

Optimization Queries – Examples without Constraints

- What is the closest restaurant to my current location?
- What is the highest ranked school according to my scoring criteria?
- Which patients have the highest AST/ALT ratio?
- Which coastal locations are most sensitive to environmental changes?

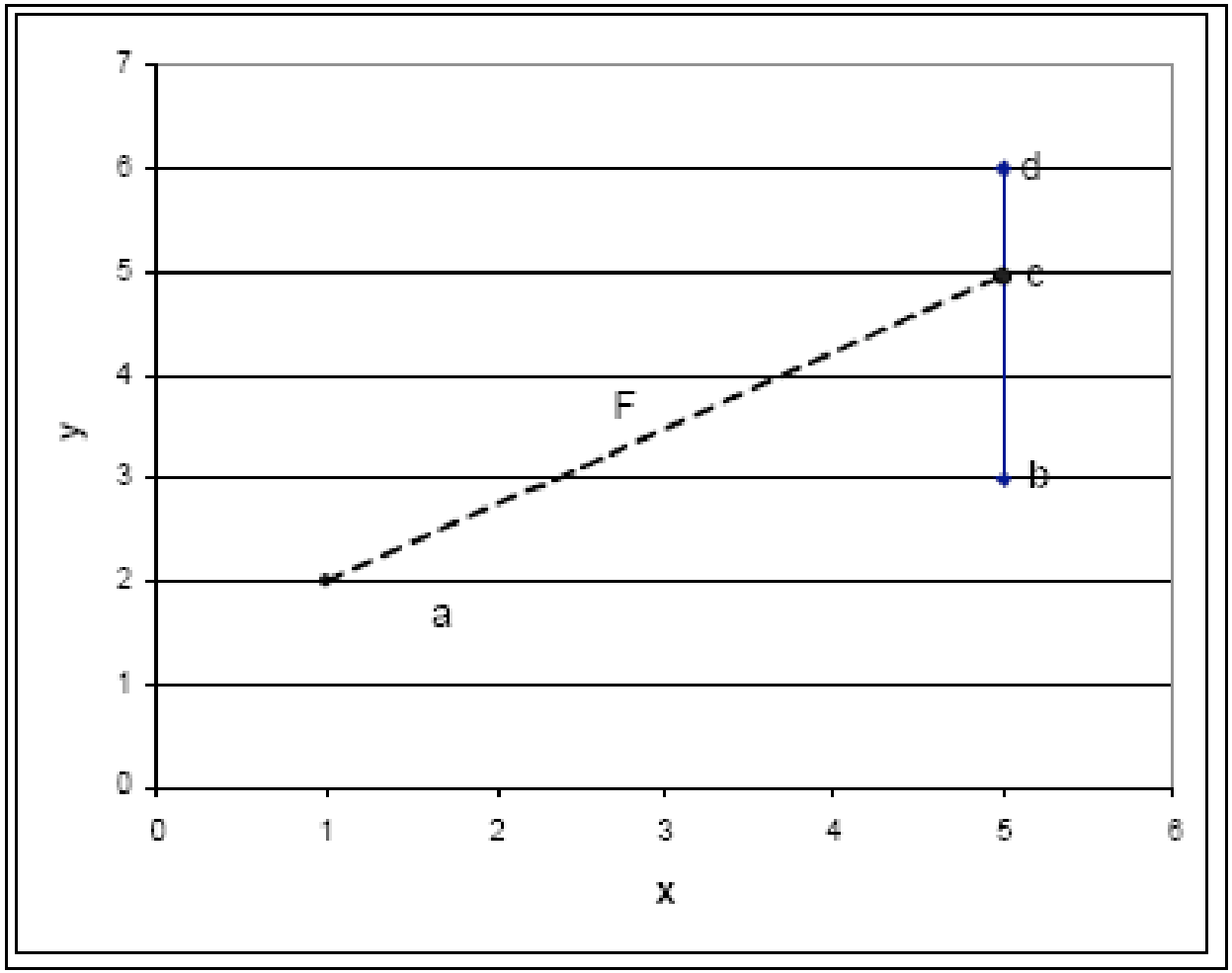
Optimization Queries – Examples with Constraints

- What is the closest restaurant to my current location which is inside “the ring”?
- What is the highest ranked school in Europe my scoring criteria?
- Which females, age 45-55 patients have the highest AST/ALT ratio?
- Which coastal locations on the Great Lakes are most sensitive to environmental changes?

Model Based Queries

DEFINITION 1 (MODEL-BASED OPTIMIZATION QUERY). *Given a relation* $Re(a_1, a_2, \ldots, a_n)$, *a model-based optimization query* Q *is given by (i) an objective function* $F(\vec{\theta}, A)$, *with optimization objective* o (min *or* max), *(ii) a set of constraints (possibly empty) specified by inequality constraints* $g_i(\vec{\alpha}, A) \leq 0, 1 \leq i \leq u$ *(if not empty) and equality constraints* $h_j(\vec{\beta}, A) = 0, 1 \leq j \leq v$ *(if not empty), and (iii) user/query adjustable objective parameters* $\vec{\theta}$, *constraint parameters* $\vec{\alpha}$ *and* $\vec{\beta}$, *and answer set size integer* k.

Sample Model Based Query



Model Based Queries - Summary

- Objective Function
- Optimization Objective (minimize or maximize)
- Constraints
- Adjustable parameters on functions and constraints
- k – number of objects to return

Convex Optimization Queries

DEFINITION 2 (CONVEX OPTIMIZATION (CP) QUERY). *A model-based optimization query Q is a Convex Optimization (CP) query if (i) $F(\vec{\theta}, A)$ is convex, (ii) $g_i(\vec{\alpha}, A), 1 \leq i \leq u$ (if not empty) are convex, and (iii) $h_j(\vec{\beta}, A), 1 \leq j \leq v$ (if not empty) are linear or affine.*

Convex Optimization Queries - Summary

- Significant subset of Model Based Optimization Queries
- Objective function is convex
- Constraints are convex
- Can be I/O-optimally processed

Query Types under Model

- (Un)Constrained-Weighted k Nearest Neighbor
- (Un)Constrained k Linear Optimization
- Range over Irregular Regions
- (Un)Constrained Arbitrary Convex Functions

Example – Euclidean Weighted Nearest Neighbor

- Objective function is to minimize weighted distance to the query point
- $WNN(a_1,a_2,\ldots a_n) = (w_1(a_1\text{-}a0_1)^2 + w_2(a_2\text{-}a0_2)^2 + \ldots + w_n(a_n\text{-}a0_n)^2)^{0.5}$
- Can be over arbitrary convex constraints for arbitrary k

Example – Linear Optimization Queries

- Objective function is to maximize a linear score
- $L(a_1,a_2,\ldots a_n) = w_1*a_1+w_2*a_2+\ldots w_n*a_n$
- Can be over arbitrary convex constraints for arbitrary k

Example – Range Queries

- Objective function is any constant
- Set k to n
- Use constraints to define ranges
- Can be used to model irregular ranges
  - e.g. $l \leq a_1+a_2 \leq u$

Goal

- Develop query processing framework to I/O-optimally solve:
  - Arbitrary convex function
  - Over arbitrary convex problem constraints
  - Using arbitrary access structure built over convex partitions

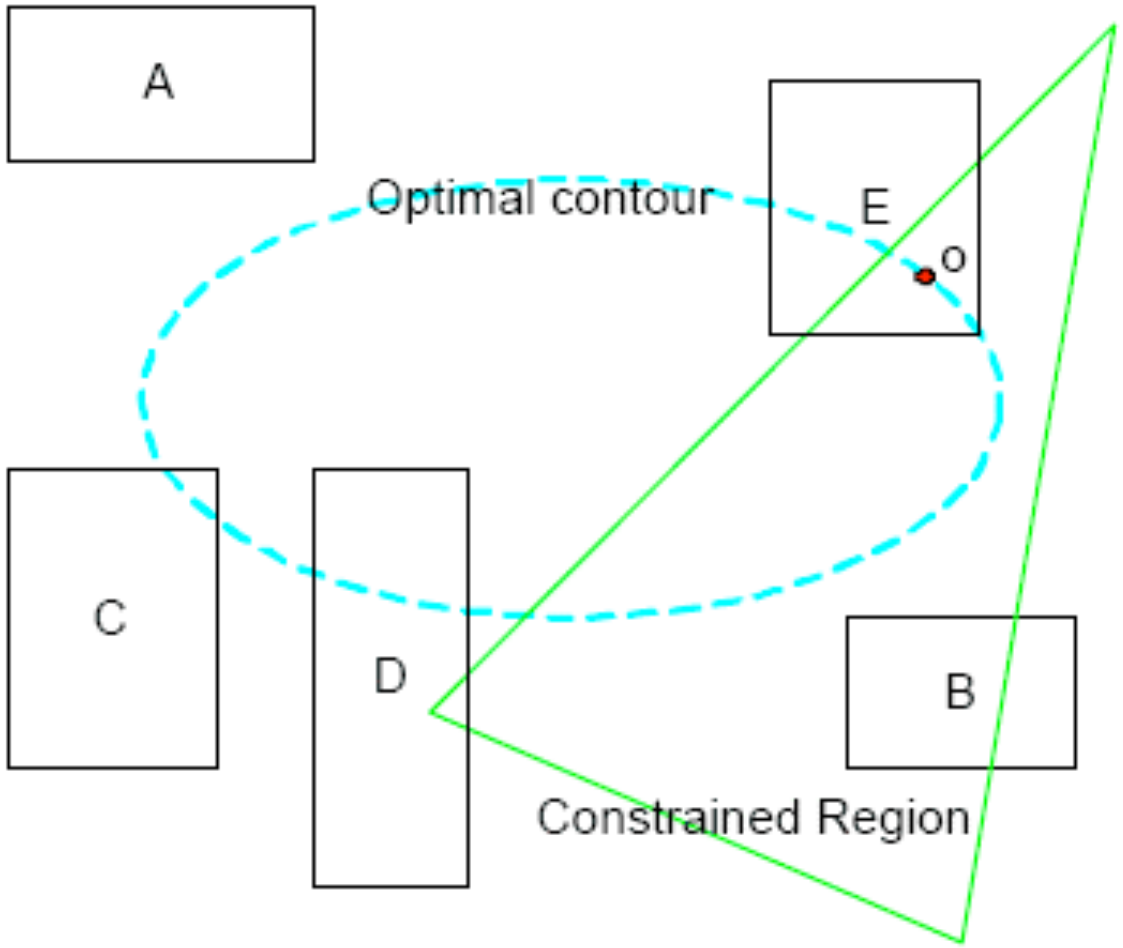
Approach

- Borrow Convex Optimization (CP) from Operations Research domain
- Find best possible answer in continuous space
- Begin searching in this partition, ordered by how promising the partitions are

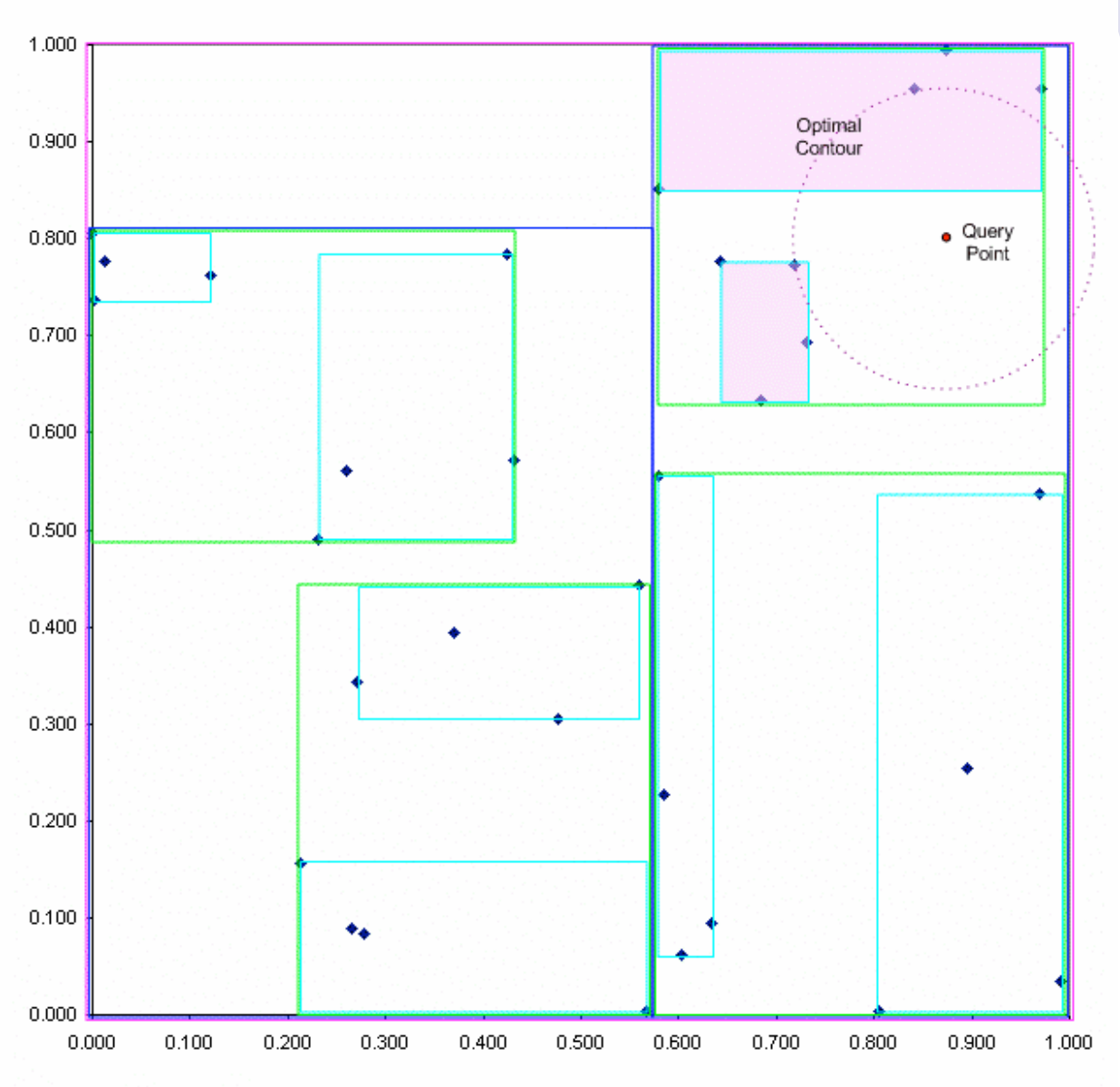
I/O Optimal Query Processing

- Solve CP problems as access structure is traversed
- Incorporate problem constraints and partition constraints to find optimal functional objective value for candidate partition
- Keep partitions ordered according to how promising they are
- Stop when partitions can not yield an optimal point

Proof of I/O Optimality

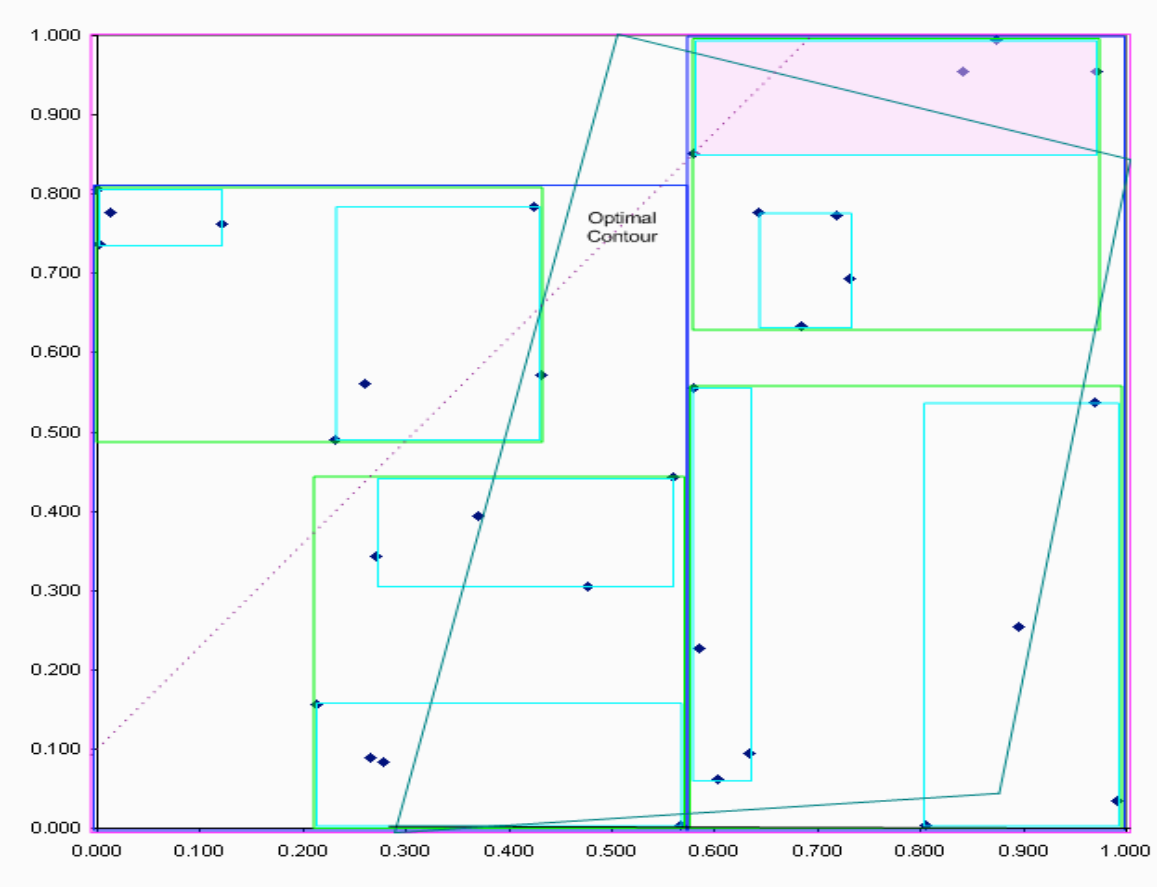


Example – Nearest Neighbor



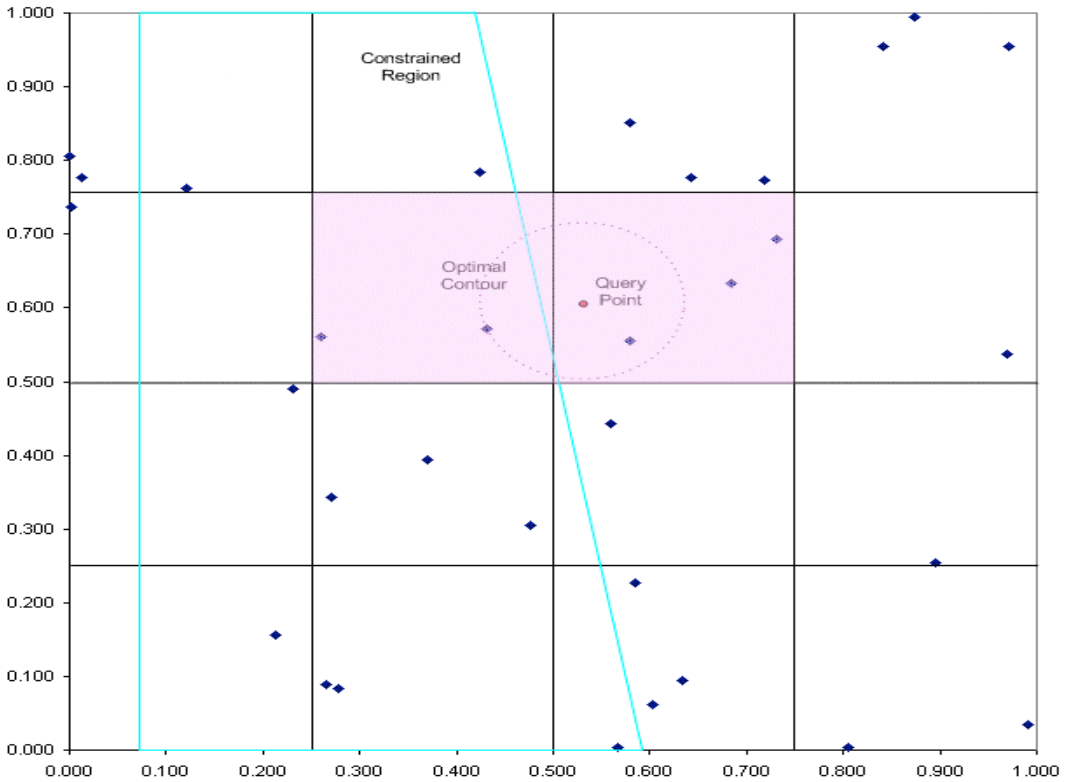
- Hierarchical Access Structure
- Only access partitions that intersect Optimal Contour

Example – Constrained Linear Optimization



- Maximize f=-6x+5y
- Within constrained area

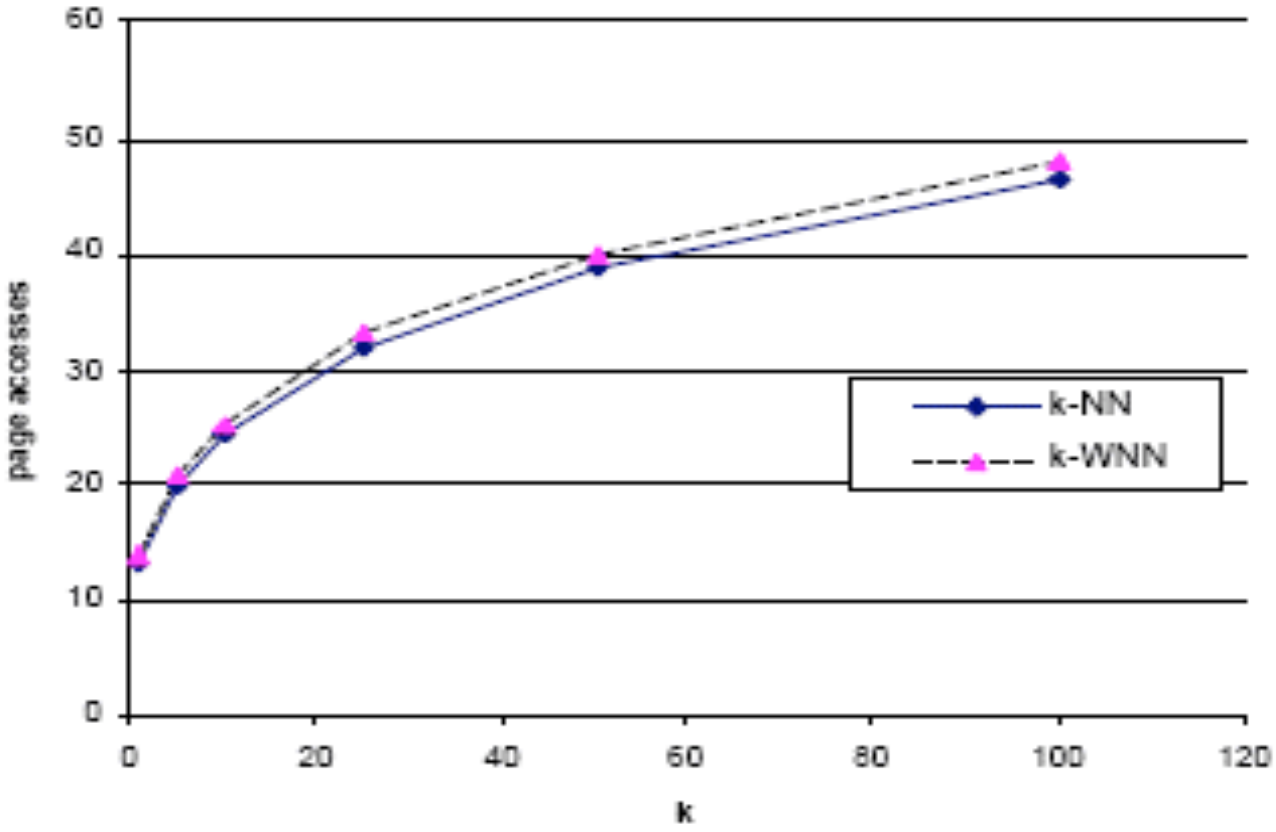
Example - Non-Hierarchical Constrained



- Non-hierarchical Structure
- NN-Query

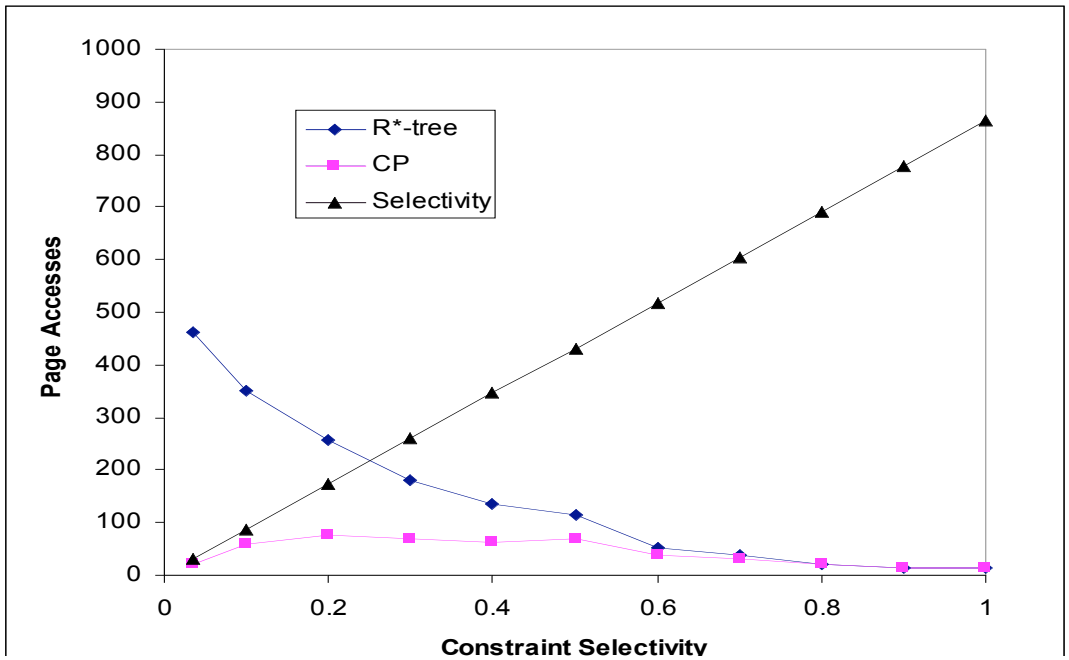
Experimental Results

k-NN and Weighted k-NN Queries



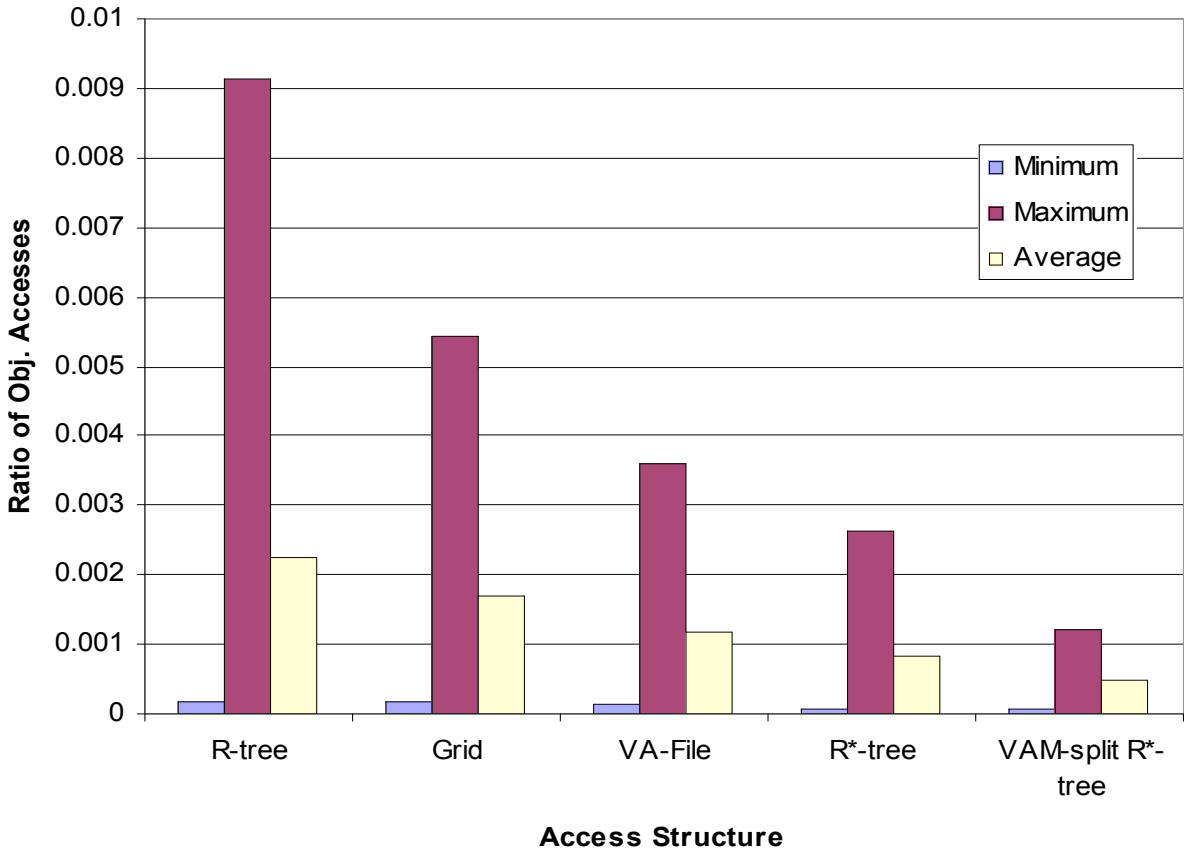
100k points, 8-D Color Histogram Data

Incorporating Constraints During Search



- NN-Query, Color Histogram
- Prune MBR's as they are discovered to be infeasible

Random Functions, Different Access Structures



5-D Uniform Random, 50k

Conclusions

- Handle Any Convex Function
- Incorporate Constraints During Access Structure Traversal
- A unified tool/algorithm for any type of optimization query
- Allows use of existing index types

Questions?