Parametric Query Optimization

by
Varakur Ganga Vara Prasad

Department of Computer Science and Engineering

Indian Institute of Technology
Kanpur

Motivation

- Queries may contain predicates with unknown selectivities at compile time.
- Actual value can be estimated only at runtime.
- Example:

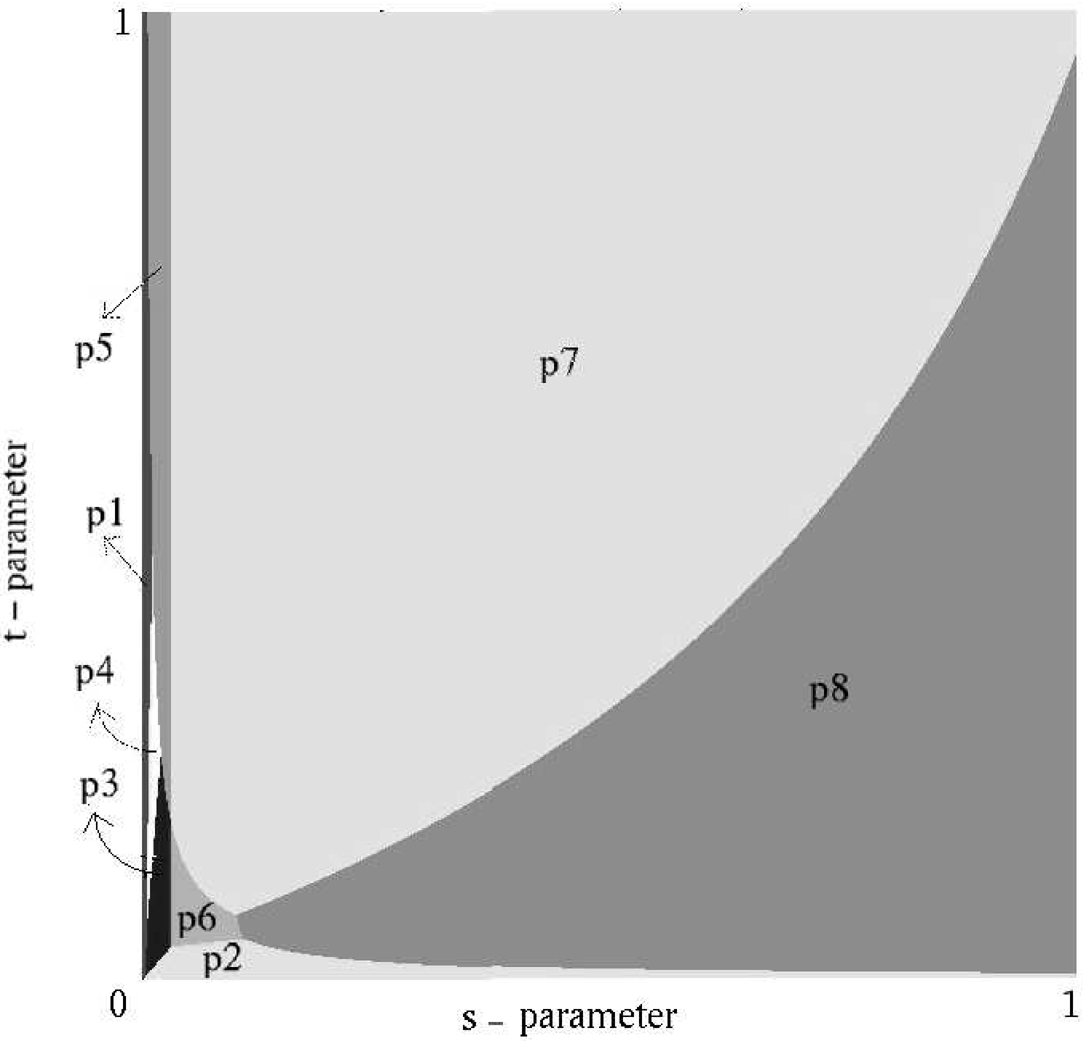
  select $*$
  from Emp, $Dept$
  where $Emp.DeptNo = Dept.DeptNo$
  and $Emp.DeptNo < ?$
  and $Dept.Assets < ??$
  order by $Dept.MgrId$

- Selectivity estimable only at run time.
- What is the Optimal Plan ?

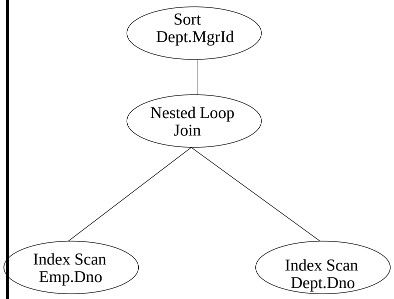
Typical Query Parameters

- Selectivity of predicates (s, t, . . .)
- Memory Available at runtime.

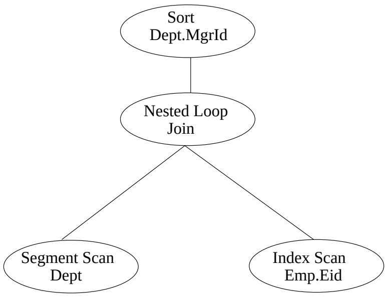
Regions of Optimality



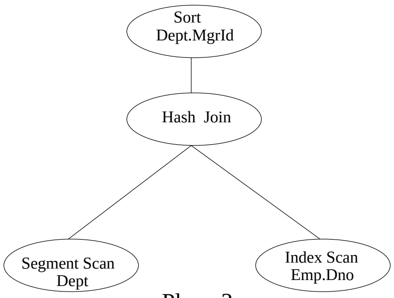
Parametric Optimal Plans



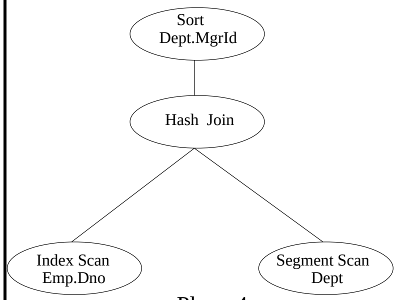
Plan p1



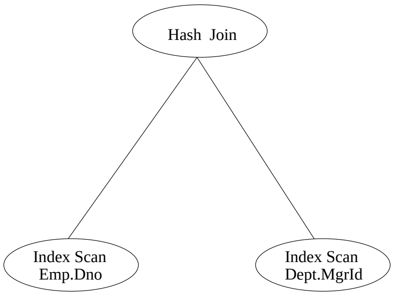
Plan p2



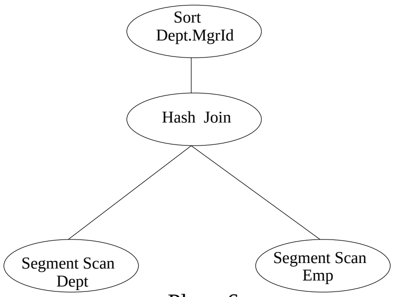
Plan p3



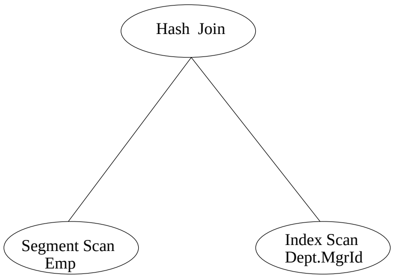
Plan p4



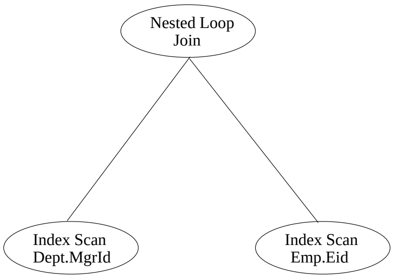
Plan p5



Plan p6



Plan p7



Plan p8

Problem Formulation

- Parametric Query Optimization.
- n-dimensional admissible parameter space.
- Single plan may not be optimal.
- Parmetric Optimal Plans ($POPs$).

$$P = \{p \,|\, p \text{ is optimal for some point in the admissible parameter space.}\}$$

- Each plan $p \in P$ will be optimal for a set of points in the parameter space.
- Region of Optimality $R(p)$ of a plan p.

$$R(p) = \{\mathbf{u}(s_1, s_2, \ldots, s_n) \,|\, p \text{ is optimal at } \mathbf{u}\}$$

- PQOP: Given a query, find all $POPs$ and their regions of optimality.

Classification of Cost functions

- Let p be a plan, s and t be parameters(e.g., selectivity).
- Linear 1 parameter cost function (1DL)

$$C(p,\ s) = a(p) + b(p).s$$

- Linear 2 parameter cost function (2DL)

$$C(p,\ s,\ t) = a(p) + b(p).s + c(p).t$$

- Linear n parameter cost function (nDL)

$$C(p, s_1, s_2, \ldots, s_n) = x_0 + x_1.s_1 + \ldots + x_n.s_n$$

Non-linear Cost functions

- Non-linear cost functions appear naturally.
- Non-linear 2 parameter cost function (2DNL)

$$C(p,\, s,\, t) = a(p) + b(p).s + c(p).t + d(p).s.t$$

Assumptions

- Uniform Distribution of column values. (Selinger + 79)
- Independent Predicates.

Parametric Optimality

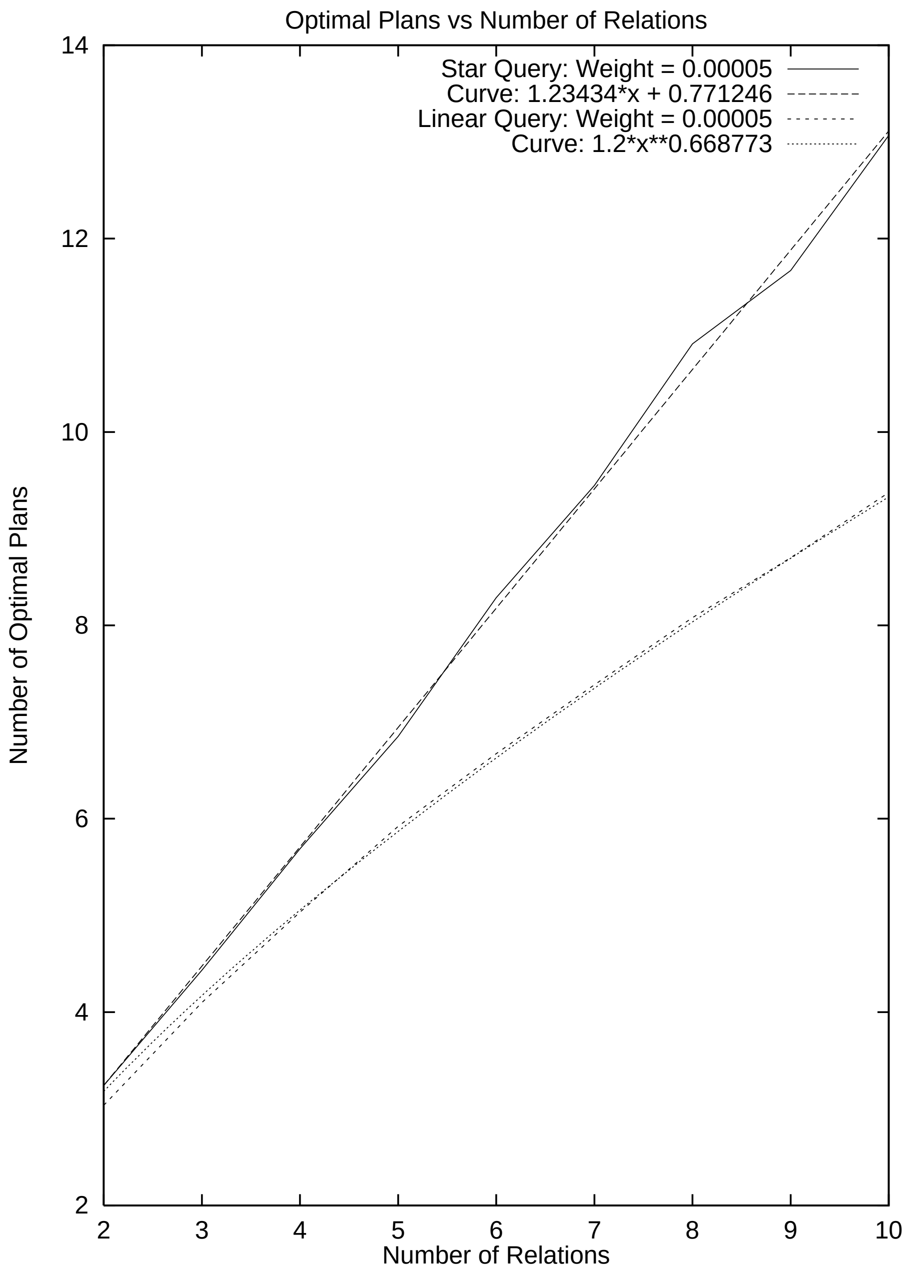
- Suppose cost function is linear in n parameters.
- Convexity Property holds.
- If plan p is optimal at points $\mathbf{u}$ and $\mathbf{v}$ of the parameter space, then p is optimal in the line segment joining them.
- **Corollary:** Regions of optimality are convex polyhedrons.
- PQOP: Find vertices/edges/faces/... etc., of the optimality regions of every parametric optimal plan.
- Convexity fails for Non-linear cost functions.

References

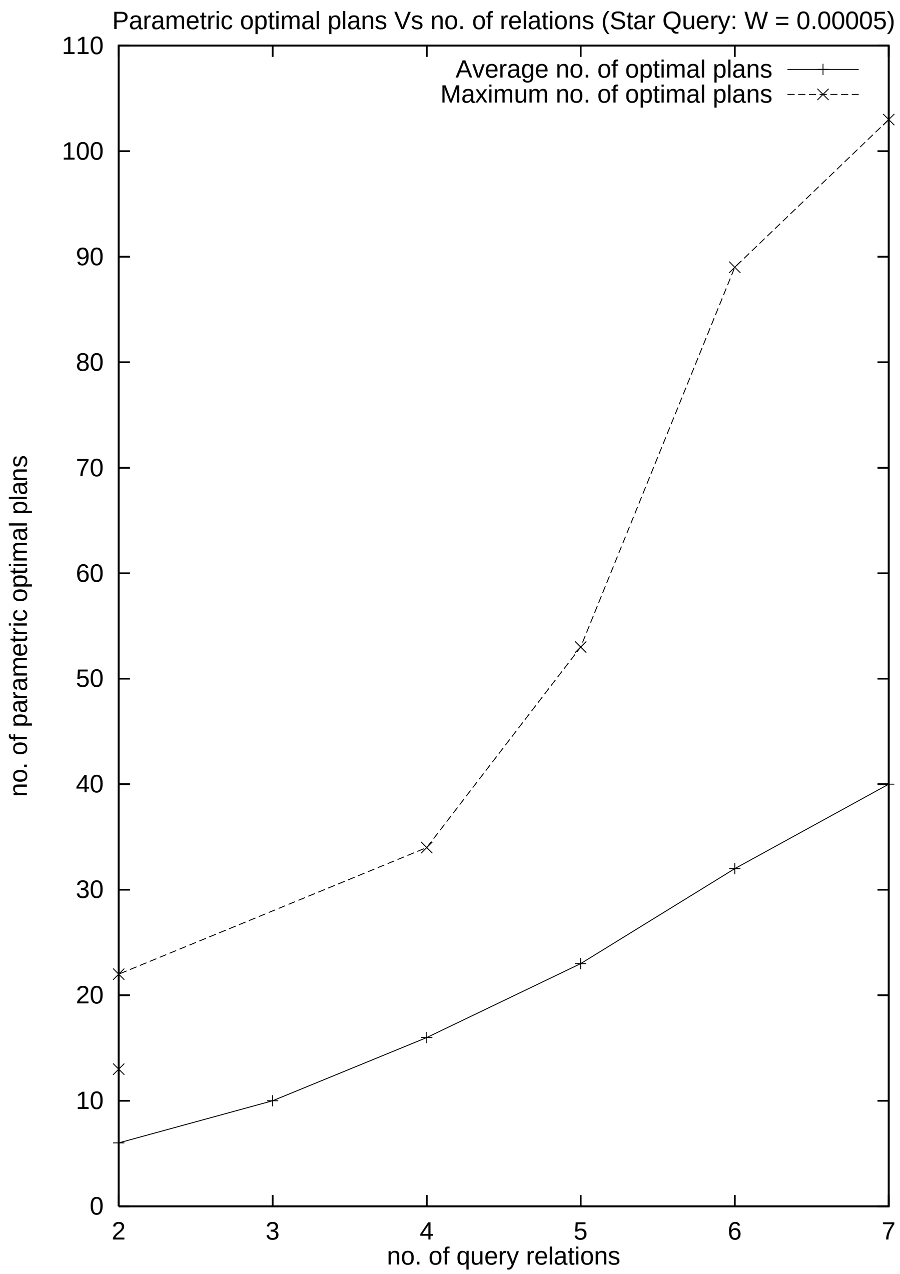
- Problem formulation, algorithms for 1DL, 2DL. Design and Analysis of Parametric Query Optimization Algorithms. VLDB'98, pp 228-238.
- 1DL experiments : Anjali Betawadkar, M.Tech Thesis '99, CSE, IIT, Kanpur.
- 2DNL Theory and experiments : Varakur Prasad, M.Tech Thesis '99, CSE, IIT, Kanpur.
- 2DNL Non-Geometric and Sampling Approaches : S.V. Uma Maheswara Rao, M.Tech Thesis '99, CSE, IIT, Kanpur.
- http://www.cse.iitk.ac.in/research/mtech-theses.html

Experimental Measurements: Number of Parametric Optimal Plans

One Parameter Queries (1DL)



Two Parameter Queries (2DNL)



Ongoing Work

- Memory as a parameter [Vasudev]
- Generalization to nDL and nDNL cost functions.
- Consolidation of Sampling Approach.
- Approximate Parametric Query Optimization.