IKEBANA

Reducing Selectivity Dimensions with Minimal Impact on Plan Bouquet

Adarsh Patil – CSA
Pradeep Bansal – SSA

Outline

Introduction

- Statistical selectivity estimation of predicates researched for several decades
  - outdated
- Bouquet Based approach for query processing
  - Future for Query Processing
- Drawbacks of Plan Bouquet
  - Assumes all dimensions as ESS
  - No techniques available yet for dimension reduction
  - Number of plans increase exponentially as ESS increases
  - High compile time overhead
- Proposed new approach for dimension reduction with bounds - IKEABANA

Challenges

- Research based project
- No existing literature
- Exponential search space
- Institutive approaches suffer difficultly in establishing sub-optimality bounds

Problem Framework

- One dimension reduction at a time in the given n-dimension ESS
- *"HyperPlane based reduction"*
- Minimum and Maximum Costs for the Ikebana bouquet in reduced ESS

  C'_{min} and C'_{max}

- Minimum and Maximum Costs for the original Bouquet are

  C_{min} and C_{max}

- $C_{max} << C'_{max}$ AND $C_{min} < C'_{min}$ due suboptimal plans

- PIC is sliced into m slices $m = log_r \left\lceil \frac{C'_{max}}{C_{min}} \right\rceil$

Cost of Ikebana Bouquet

$$C_I(q_{k'}) = cost(IC_1) + ... + cost(IC_{k'}) = \frac{a(r^{k'} - 1)}{r - 1}$$

Cost of Oracle

$$C^*(q_k) = ar^{k-2}$$

Problem Framework contd..

- Theorem – Ikebana Bouquet sub-optimality w.r.t Oracle

$$MSO_I \leq \frac{\rho r^{\delta+2}}{r-1}$$

- **PROOF**

-> From previous result we get $SO_I \leq \dfrac{\frac{a(r^{k'}-1)}{r-1}}{ar^{k-2}}$

-> Execution contour of Ikebana Bouquet is higher than that of Oracle

-> Substitute k' = k + δ where δ >= 0

$$SO_I \leq \frac{a(r^{k+\delta}-1)}{(r-1)(ar^{k-2})} = \frac{r^{\delta+2}}{r-1} - \frac{r^{2-k}}{r-1} \leq \frac{r^{\delta+2}}{r-1}$$

Problem Framework contd..

- Since the previous expression is independent of k we get

$$MSO_I \leq \frac{\rho r^{\delta+2}}{r-1}$$

where

$$\rho = \begin{cases} |\text{ reduced POSP }| & \text{for 2 dimension} \\ \text{number of plans on densest contour for higher dimension} & \end{cases}$$

- Construct the iso-cost contour as previously with r = 2

- Special case – Sub-optimality w.r.t Plan Bouquet on original ESS

$$MSO_I \leq 4\rho 2^{\delta}$$

- NOTE : For δ = 0 we get back original bouquet sub-opt

Solution characteristics

- Minimal impact on final bouquet performance
- Plans and their budgets should cover entire original ESS
- Overlap factor should be minimized
- Increase in budgets for plans should be as less as possible
- Generic approach independent of specificities of SQL like data type, conjunction and join condition
- If possible, reduce number of plans to be executed along with dimension

Algorithm

Algorithm 1 Ikebana Algorithm

Algo-Ikebana(planCost,dimension,resolution)
for each dimension d
 for each hyperplane h in d
  for all the points p modulo points on h
   find the optimal plan , from the set of plans on h, at point p
  find maxDiffPair = (bestCost in reduced ESS,optCost)
  such that (bestCost in reduced ESS-optCost) is max $\forall$ points
  find (min,max) cost for all plans on the hyperplane h
  find ρ , k' and k using maxDiff , using r = 2, in reduced ESS find $\delta = k' - k$
  calculate MSO_h for the hyperplane h given by (7)
  choose the hyperplane with the least MSO_h return (h^*,MSO_h^*,(set of plans on h^*,min,max))

Experiments

- TPC-H Q5
  - 3D ESS
  - 20 x 20 x 20 sampling grid
  - 50 POSP discovered by DB Optimizer
  - Executes in 5.3 seconds [8 core / 64GB machine]

- Exploits parallelism of Dimensions in separate threads (i.e. 3 threads)

Table 1: Ikebana Bouquet for reduced dimension with budgets for TPC-H Q5

Reducing dimension 0
Use selectivity row 5 with MSO 24.0

Here's the Plan Bouquet
Plan Number : Min Cost : Max Cost : Overlap Factor
0 : 32047.57154 : 72584.8408 : 1.0
33 : 143221.48723 : 271056.00391 : 28.0
2 : 39315.74284 : 149520.57702 : 9.0
35 : 233741.21581 : 594256.69581 : 2.0
32 : 163600.859204 : 383326.749728 : 58.0
34 : 39316.88784 : 272311.44922 : 2.0
19 : 87052.302142 : 248789.973949 : 2.0
24 : 94099.625582 : 251011.807459 : 5.0
27 : 143187.97099 : 143196.36005 : 60.0
28 : 39316.01034 : 149520.73702 : 38.0
30 : 39316.25034 : 195476.54914 : 14.0

Reducing dimension 1
Use selectivity row 17 with MSO 20.0

Here's the Plan Bouquet
Plan Number : Min Cost : Max Cost : Overlap Factor
32 : 163600.859204 : 251555.591209 : 55.0
2 : 32514.30034 : 150061.09922 : 2.0
3 : 81292.75778 : 87647.38038 : 3.0
4 : 81334.47808 : 87647.58788 : 13.0
6 : 81392.754 : 87648.12038 : 6.0
8 : 81501.09053 : 87656.48038 : 5.0
41 : 234086.10581 : 464448.93956 : 4.0
42 : 286990.602609 : 330700.914017 : 27.0
43 : 315186.65456 : 498885.98956 : 103.0
34 : 34713.21409 : 272311.44922 : 2.0
35 : 233741.21581 : 295951.06331 : 5.0
24 : 94054.096832 : 251011.807459 : 2.0
28 : 32786.86034 : 173532.76672 : 15.0
30 : 33062.96784 : 195476.54914 : 7.0

Reducing dimension 2
Use selectivity row 3 with MSO 32.0

Here's the Plan Bouquet
Plan Number : Min Cost : Max Cost : Overlap Factor
0 : 32047.57154 : 72582.7633 : 2.0
2 : 39315.74284 : 150061.09922 : 9.0
3 : 81292.75778 : 87688.14913 : 1.0
44 : 32344.90764 : 72583.0533 : 16.0
34 : 39316.88784 : 272311.44922 : 2.0
46 : 32654.04994 : 72583.5583 : 11.0
15 : 87051.727142 : 248841.642699 : 4.0
35 : 233741.21581 : 594256.69581 : 2.0
24 : 94099.625582 : 384929.793478 : 6.0
47 : 34450.56499 : 72584.56205 : 21.0
28 : 39316.01034 : 173532.76672 : 40.0
30 : 39316.25034 : 195476.54914 : 14.0

Experiments – TPC-H, Q5

Figure 3: Min/Max MSO bounds for Reducing dimension 2 on TPC-H Q5

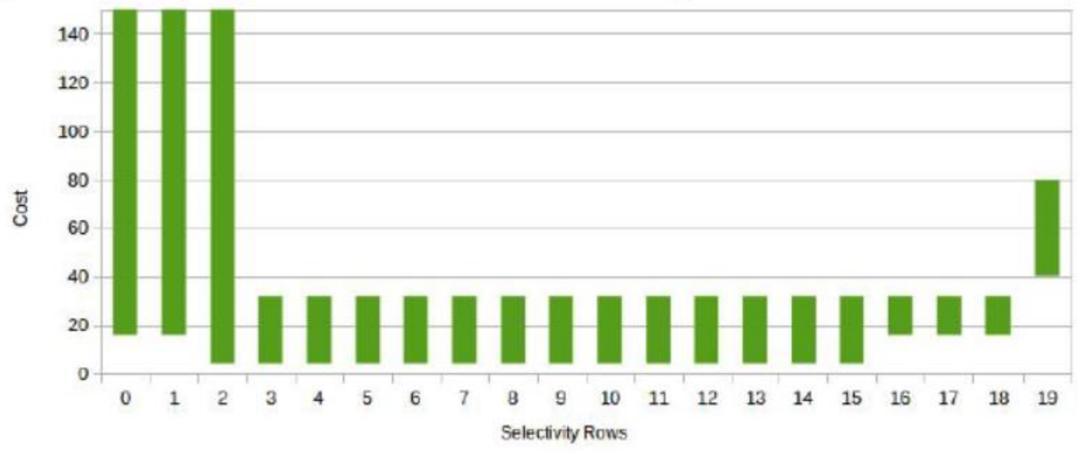


Figure 2: Min/Max MSO bounds for Reducing dimension 0 on TPC-H Q5

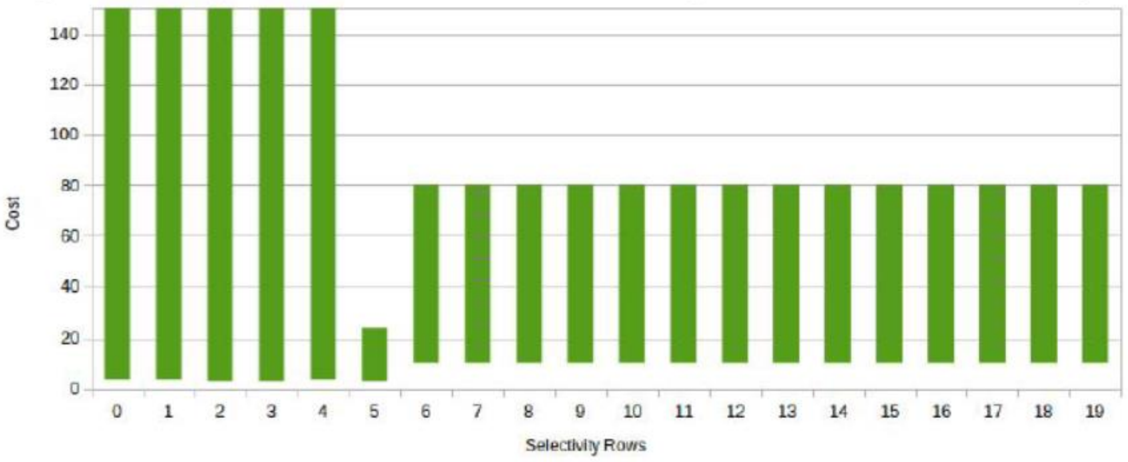
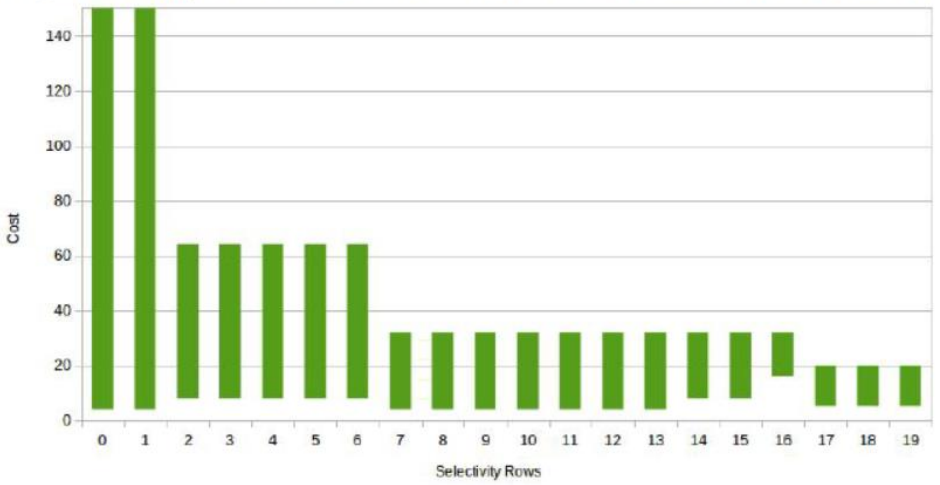
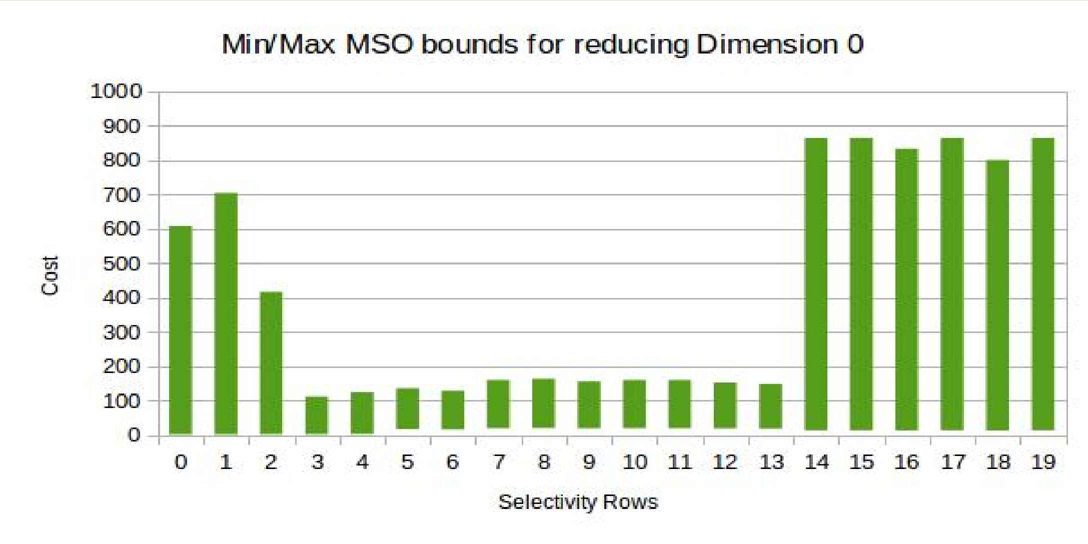
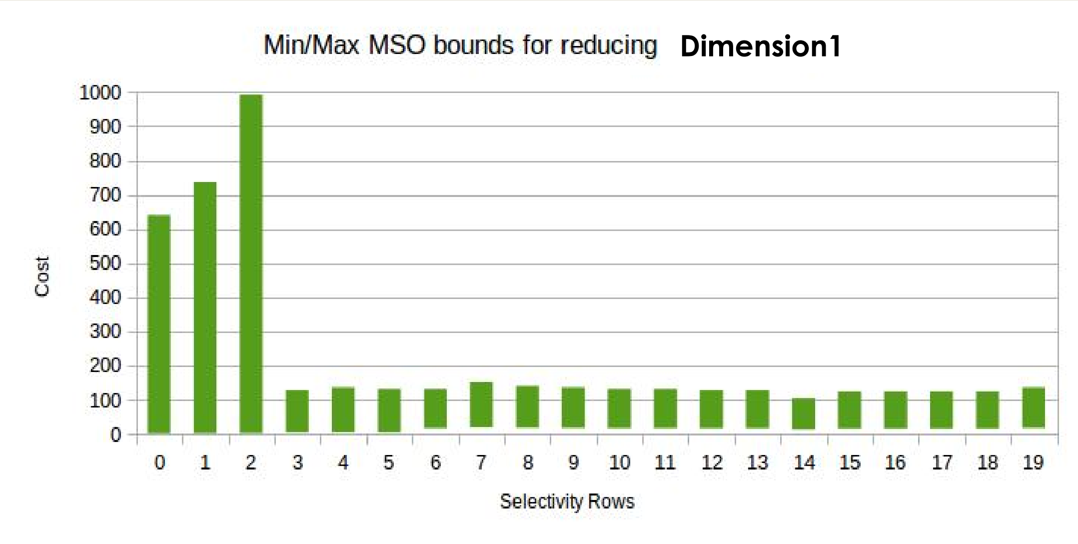


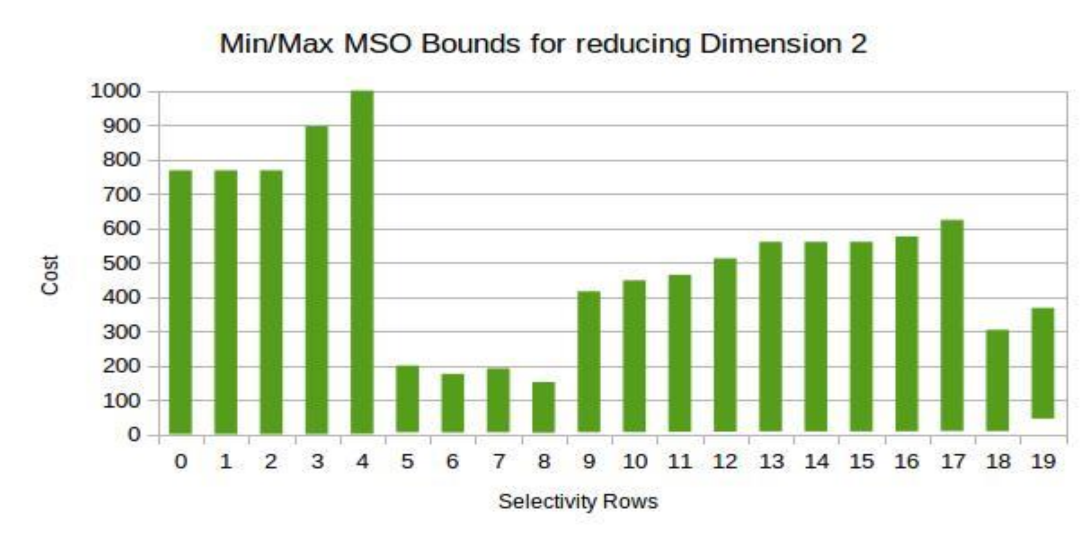
Figure 1: Min/Max MSO bounds for Reducing dimension 1 on TPC-H Q5

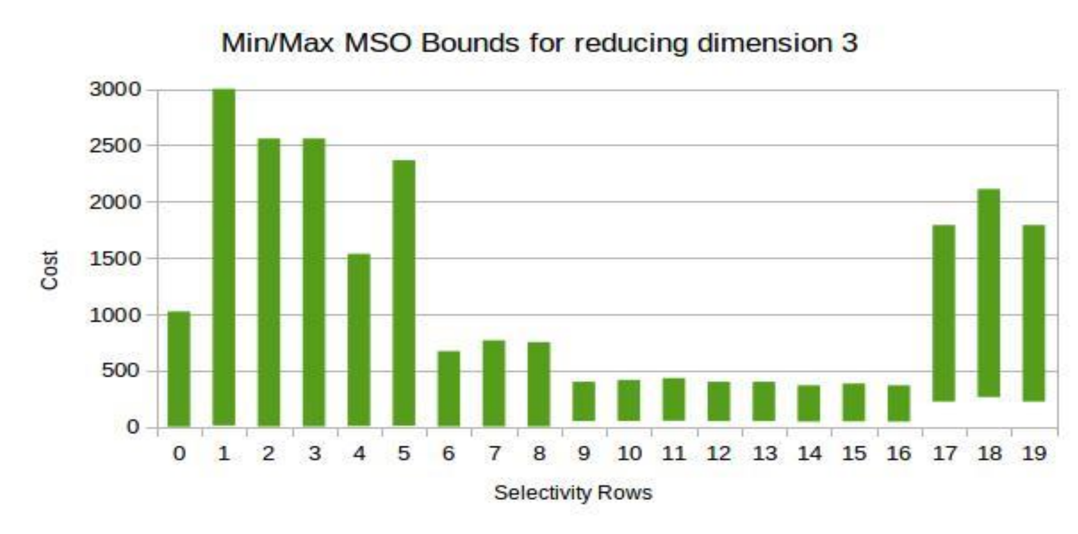


Experiments – TPC-H, Q8(4D, plans 324, resolution 20)









Conclusion

- Successfully quantified the impact of reducing dimensions
- Established concrete bounds on sub-optimality induced due to dimension reduction
- Improves bouquet performance by reducing run time due to reduction in plan density and also in ESS dimensions
- Reduces compile time cost for Plan Bouquet as the number of dimensions to explore are reduced.

Future Work

- Iterative vs combinatorial approach for dimension reduction
- Impact of overlap factor
- Technique using less number of FPC calls

ARIGATŌ

THANK YOU…