Average Instance Optimality Ratio

From Theory to Practice, and Back

Jérémy Barbay

jeremy@barbay.cl

[2014-10-27 Mon]@AofA

Based on work with Afshani and Chan [Afshani et al., 2017], Rojas [Rojas, 2018], Ochoa [Ochoa, 2019].

Outline

Outline

Quick Select/Sort/Deferred Data Structures

- `Quick Sort` [Hoare, 1961a]
- `Quick Select` [Hoare, 1961b]
- `Deferred Data Structure` [Karp et al., 1988]

Quick Select

7	6	5	10	11	1	12	13	2	4	14	3	15	0	9	8

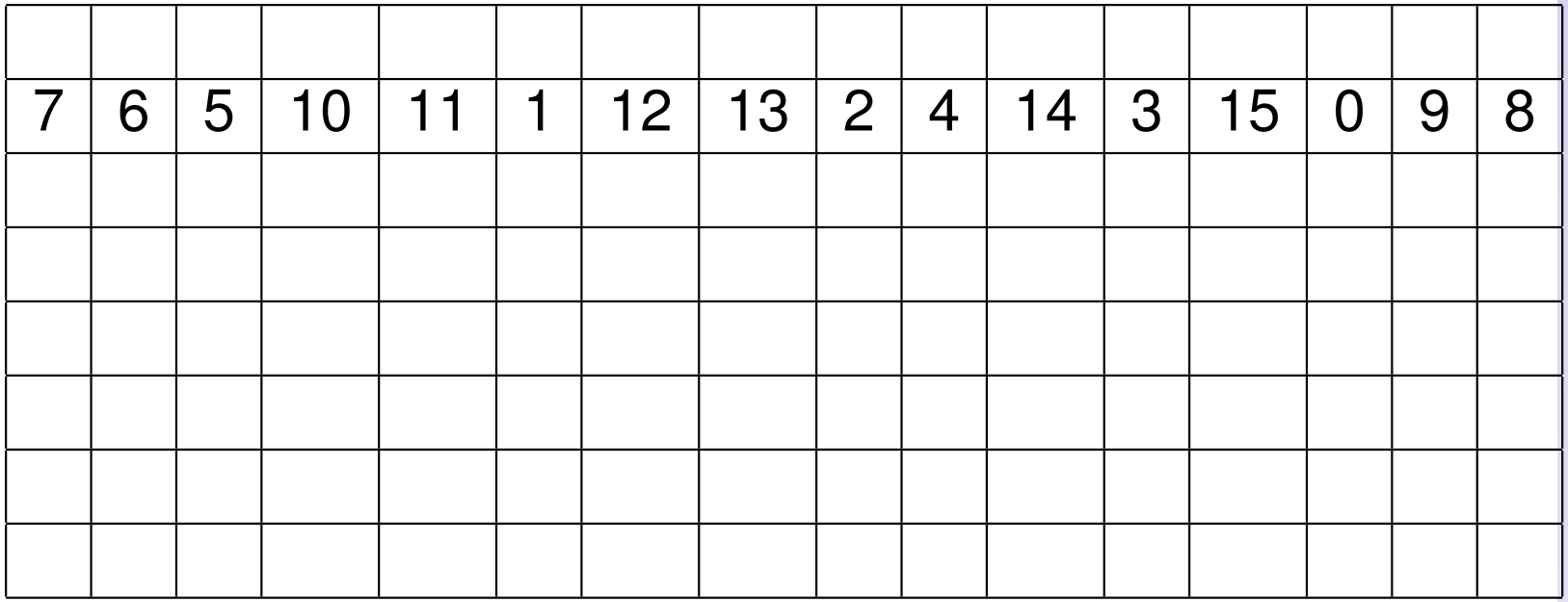
[Hoare, 1961a]

Quick Sort

7	6	5	10	11	1	12	13	2	4	14	3	15	0	9	8

[Hoare, 1961b]

Deferred Data Structure



[Karp et al., 1988]
[Barbay et al., 2016]

Outline

Competitive Ratio (Online)

Given an online algorithm A, and
the optimal offline algorithm OPT,
the Competitive Ratio of A is
the maximum of their ratio
over all possible input sequences S.

$$C(A) = \max_S \frac{C(A,S)}{C(OPT,S)}$$

(Same for data-structures supporting queries.)

[Sleator and Tarjan, 1985]
[Borodin and El-Yaniv, 1998]

Outline

Instance Optimality (Offline)

$$\forall I,\ C(A, I) \in O(C(OPT, I))$$

- [Kenyon, 1996]
  - BinPacking's random input order instance optimality
- [Fagin et al., 2001, Fagin et al., 2003]
  - Aggregation's instance optimality
  - Godel prize! [EATCS, 2014]
- [Afshani et al., 2009, Afshani et al., 2017]

$$\left\{ \begin{array}{c} \text{Maxima} \\ \text{Convex Hull} \end{array} \right\} \left\{ \begin{array}{c} \text{2d} \\ \text{3d} \end{array} \right\} \left\{ \begin{array}{c} \text{input order oblivious} \\ \text{random input order} \end{array} \right\} \text{instance optimality}$$

Note: can be written as $\max_I \frac{C(A,I)}{C(OPT,I)} \in O(1)$

Outline

Instance Optimality

- Instance Optimal:
  - AGGREGATION [Fagin et al., 2003]
  - MERGING SORTED ARRAYS [Demaine et al., 2000];
  - ...
- NOT Instance Optimal:
  - MAXIMA;
  - CONVEX HULL;
  - INTERSECTING SORTED ARRAYS;

    (if the number k of arrays is not a constant.)
  - ...

[Fagin et al., 2003]

(Input Order Oblivious) Instance Optimality

- Input Order Oblivious Instance Optimal:
  - AGGREGATION [Fagin et al., 2003]
  - MERGING SORTED ARRAYS [Demaine et al., 2000];
  - MAXIMA 2d/3d;
  - CONVEX HULL 2d/3d;
  - ...
- NOT Input Order Oblivious Instance Optimal:
  - INTERSECTING SORTED ARRAYS

    (if the number k of arrays is not a constant.)
  - ...

[Afshani et al., 2017]

Instance Optimality Ratio (the proposal)

$C(A) = \max_I \frac{C(A,I)}{C(OPT,I)}$; $C(Pb) = \min_A C(A)$

- AGGREGATION:
- MERGING MULTISETS:
- SORTING MULTISETS:
- MAXIMA/CONVEX HULL 2d/3d:
- INTERSECTING MULTISETS:
- ...

$$A \in \text{Instance Optimal} \Leftrightarrow C(A) \in O(1)$$

Instance Optimality Ratio (the proposal)

$C(A) = \max_I \frac{C(A,I)}{C(OPT,I)}$; $C(Pb) = \min_A C(A)$

- AGGREGATION: $O(1)$
- MERGING MULTISETS: $O(1)$
- SORTING MULTISETS:
- MAXIMA/CONVEX HULL 2d/3d:
- INTERSECTING MULTISETS:
- ...

$$A \in \text{Instance Optimal} \Leftrightarrow C(A) \in O(1)$$

Instance Optimality Ratio (the proposal)

$C(A) = \max_I \frac{C(A,I)}{C(OPT,I)}$; $C(Pb) = \min_A C(A)$

- AGGREGATION: $O(1)$
- MERGING MULTISETS: $O(1)$
- SORTING MULTISETS: $O(\lg n)$, $O(1)$
- MAXIMA/CONVEX HULL 2d/3d:
- INTERSECTING MULTISETS:
- ...

$$A \in \text{Instance Optimal} \Leftrightarrow C(A) \in O(1)$$

Instance Optimality Ratio (the proposal)

$C(A) = \max_I \frac{C(A,I)}{C(OPT,I)}$; $C(Pb) = \min_A C(A)$

- AGGREGATION: $O(1)$
- MERGING MULTISETS: $O(1)$
- SORTING MULTISETS: $O(\lg n)$, $O(1)$
- MAXIMA/CONVEX HULL 2d/3d: $O(C(\text{SORT}))$
- INTERSECTING MULTISETS:
- ...

$$A \in \text{Instance Optimal} \Leftrightarrow C(A) \in O(1)$$

Instance Optimality Ratio (the proposal)

$C(A) = \max_I \frac{C(A,I)}{C(OPT,I)}$; $C(Pb) = \min_A C(A)$

- AGGREGATION: $O(1)$
- MERGING MULTISETS: $O(1)$
- SORTING MULTISETS: $O(\lg n)$, $O(1)$
- MAXIMA/CONVEX HULL 2d/3d: $O(C(\text{SORT}))$
- INTERSECTING MULTISETS: $O(\lg k)$
- ...

$$A \in \text{Instance Optimal} \Leftrightarrow C(A) \in O(1)$$

Instance Optimality Ratio (the proposal)

$C(A) = \max_I \frac{C(A,I)}{C(OPT,I)}$; $C(Pb) = \min_A C(A)$

- AGGREGATION: $O(1)$
- MERGING MULTISETS: $O(1)$
- SORTING MULTISETS: $O(\lg n)$, $O(1)$
- MAXIMA/CONVEX HULL 2d/3d: $O(C(\text{SORT}))$
- INTERSECTING MULTISETS: $O(\lg k)$
- ...

$A \in$ Instance Optimal $\Leftrightarrow C(A) \in O(1)$

Outline

Computation of the Median

Solutions heavily based on median computations.
Various ways to compute the median.

Algorithm	Worst Case	Average
Median of Median	$O(n)$	$O(n)$
Quick Select	$O(n^2)$	$O(n)$
Median of $(2k+1)$	$O(n^2)$	$O(n)$

Their expected running time is well studied.

Why does using MEDIAN OF 5 yields better results than MEDIAN OF 3 or MEDIAN OF 7 in practice in a deferred data structure supporting `rank` and `select` queries on MULTISETS? [Barbay et al., 2013]

Average Instance Optimality Ratio (Randomized Algorithm)

Can we extend the average results on `Quicksort` to

- the computation of
  1. MAXIMA SETS or
  2. CONVEX HULL; or to
- the offline support of
  1. `rank` and `select` queries;
  2. `domination` queries;
  3. `convex hull membership` queries?

Outline

Average Competitive Ratio (Randomized Data Structure)

$$C(A) = \max_{I} \frac{C(A, I)}{C(OPT, I)}$$

- Can we extend the average results to Deferred Data Structures on
  1. Multi Sets supporting `rank` and `select` queries;
  2. Point Sets supporting `domination` queries;
  3. Point Sets supporting `convex hull membership` queries?

Outline

Other Models and Problems

- Synergistic solutions for MULTISETS, MAXIMA and CONVEX HULL [Barbay et al., 2017a, Barbay and Ochoa, 2018, Ochoa, 2019]
- OPTIMAL BOXES [Barbay et al., 2014]
- KLEE'S MEASURE [Barbay et al., 2015]
- DEPTH DISTRIBUTION [Barbay et al., 2017b]
- COVERAGE KERNEL [Barbay et al., 2018]

Outline

External Memory

Deferred Data Structure supporting `rank` and `select` in external memory with instance optimality ratio $O(1)$ in memory accesses [Barbay et al., 2016]

- Can we refine the constant factor? (probably meaningless)
- Can we prove the same for MAXIMA, for CONVEX HULL?

Outline

Higher Dimensions

- MAXIMA's computation in dimension $d > 2$

$$\in O\left(n \lg^{d-2} n - \sum n_i \lg^{d-2} n_i\right)$$

$$\subseteq O\left(n \lg^{d-2} h\right)$$

$$\subseteq O\left(n \lg^{d-2} n\right)$$

- How would it perform with `Median of Five`?
- Can the technique be extended to a deferred data structure supporting queries?

[Kirkpatrick and Seidel, 1985]
[Barbay and Rojas-Ledesma, 2017]

Summary

- INSTANCE OPTIMALITY RATIO:
  - INSTANCE OPTIMALITY $\approx$ COMPETITIVE RATIO.
  - INSTANCE OPTIMALITY fails to explain some results.
  - INSTANCE OPTIMALITY RATIO improves the situation.

- AVERAGE RATIO:
  - is likelly to further explain some experimental results:
  - computation performs **many** median approximations,
  - few erroneous ones do not matter in total.

References

Afshani, P., Barbay, J., and Chan, T. M. (2009).

Instance-optimal geometric algorithms.
In *Proceedings of the Annual IEEE Symposium on Foundations of Computer Science (**FOCS**)*, pages 129–138. IEEE Computer Society.

Afshani, P., Barbay, J., and Chan, T. M. (2017).

Instance-optimal geometric algorithms.
*Journal of the ACM (**JACM**)*, 64(1):3:1–3:38.

Barbay, J., Chan, T. M., Navarro, G., and Pérez-Lantero, P. (2014).

Maximum-weight planar boxes in $o(n^2)$ time (and better).
*Information Processing Letters (**IPL**)*, 114(8):437–445.

Barbay, J., Gupta, A., Jo, S., Rao, S. S., and Sorenson, J. (2013).

Theory and implementation of online multiselection algorithms.
In *Proceedings of the Annual European Symposium on Algorithms (**ESA**)*.

Barbay, J., Gupta, A., Satti, S. R., and Sorenson, J. (2016).

Near-optimal online multiselection in internal and external memory.
*Journal of Discrete Algorithms (**JDA**)*, 36:3–17.

Barbay, J. and Ochoa, C. (2018).

Synergistic computation of planar maxima and convex hull.
In *Proceedings of the Annual International Computing and Combinatorics Conference (**COCOON**)*.

Barbay, J., Ochoa, C., and Satti, S. R. (2017a).

Synergistic Solutions on MultiSets.
In Kärkkäinen, J., Radoszewski, J., and Rytter, W., editors, *Proceedings of the Annual Symposium on Combinatorial Pattern Matching (**CPM**)*, volume 78 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 31:1–31:14, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Barbay, J., Pérez-Lantero, P., and Rojas-Ledesma, J. (2015).