International Workshop On Combinatorial Algorithms
IWOCA

Tight bounds for sorting a multiset?

Travis Gagie
University of Helsinki
travis.gagie@gmail.com

14 July 2009
updated on 23 March 2013

Comparison-based sorting is perhaps the most studied problem in computer science but there remain basic open questions about it. For example, how many ternary comparisons are needed to sort a multiset S of size n? By a ternary comparison, we mean one that can return $<$, $=$ or $>$; we count only comparisons between elements of the multiset, not those between data generated by the algorithm. Over thirty years ago, Munro and Spira [5] proved distribution-sensitive upper and lower bounds that differ by $\mathcal{O}(n \log \log \sigma)$, where σ is the number of distinct elements in S. Their bounds have been improved in a series of papers — summarized in the table below — and now the best known upper and lower bounds (of which we are aware) differ by a term linear in n (about $(1 + \log_2 e)n \approx 2.44n$ when $\sigma = o(n)$); in the table, $H = \sum_{i=1}^{\sigma} \frac{\mathrm{occ}(a_i, S)}{n} \log_2 \frac{n}{\mathrm{occ}(a_i, S)}$ is the entropy of the distribution of the elements in S, $a_1, \ldots, a_i$ are the distinct elements and $\mathrm{occ}\,(a_i, S)$ is the number times a_i occurs in S. Nevertheless, we are still unable to say exactly how many comparisons are needed as n goes to infinity. Can we prove bounds that differ by a term sublinear in n? (We recently proved such bounds for the special case of online stable sorting [3].)

	upper bound	lower bound
Munro and Raman [4]		$(H - \log_2 e)n + \mathcal{O}(\log n)$
Fischer [2]	$(H + 1)n - \sigma$	$(H - \log_2 H)n - \mathcal{O}(n)$
Dobkin and Munro [1]		$\left(H - n \log_2 \left(\log_2 n - \frac{\sum_i \mathrm{occ}(a_i, S) \log_2 \mathrm{occ}(a_i, S)}{n}\right)\right) n - \mathcal{O}(n)$
Munro and Spira [5]	$nH + \mathcal{O}(n)$	$nH - (n - \sigma) \log_2 \log_2 \sigma - \mathcal{O}(n)$

References

[1] D. P. Dobkin and J. I. Munro. Determining the mode. *Theoretical Computer Science*, 12:255–263, 1980.

[2] T. M. Fischer. On entropy decomposition methods and algorithm design. *Colloquia Mathematica Societatis János Bolyai*, 44:113–127, 1984.

[3] T. Gagie and Y. Nekrich. Tight bounds for online stable sorting. *Journal of Discrete Algorithms* 9(2):176–181, 2011.

[4] J. I. Munro and V. Raman. Sorting multisets and vectors in-place. In *Proceedings of the 2nd Workshop on Algorithms and Data Structures (WADS)*, pages 473–480, 1991.

[5] J. I. Munro and P. M. Spira. Sorting and searching in multisets. *SIAM Journal on Computing*, 5(1):1–8, 1976.