On the Importance of Idempotence

Sunil Arya* Theocharis Malamatos† David M. Mount‡

For FWCG 2005

Answering range queries is a problem of fundamental importance in spatial information retrieval and computational geometry. The objective is to store a set of n points P in $\mathbb{R}^d$, each associated with a weight, so that it is possible to count, or more generally to compute some function of the weights of the points lying inside a given query range. Range searching is among the most heavily studied problems, and many search structures have been proposed and analyzed [1, 7]. There is a spectrum of space-time tradeoffs. The most relevant work to ours involves halfspace range counting queries, which Matoušek [6] has shown can be answered in $n/m^{1/d}$ time from a data structure of space $O(m)$. Nearly matching lower bounds were given by by Brönnimann, Chazelle and Pach [4] (or BCP).

Given the relatively high complexity of range searching, it is natural to consider the problem in the context of approximation. We are given an approximation parameter $\varepsilon > 0$ and assume that ranges are bounded. Let η denote a range, and let $diam(\eta)$ denote its diameter. All the points that lie in the range must be counted, and any of the points that lie within distance $\varepsilon \cdot diam(\eta)$ of the range's boundary may be counted as well. Arya and Mount [3] showed that in any fixed dimension d with $O(n \log n)$ preprocessing time and $O(n)$ space, ε-approximate range queries for any bounded convex range can be answered in time $O(\log n + 1/\varepsilon^{d-1})$ [3]. Later, Chazelle, Liu, and Magen [5] considered approximate halfspace range and Euclidean ball searching in the high dimensional setting. Ignoring polylogarithmic factors, they showed that is possible to answer queries in $O(d/\varepsilon^2)$ time with $O(dn^{O(1/\varepsilon^2)})$ space.

In fixed dimensional spaces a natural goal is to achieve query times that are polylogarithmic in n while using space that is roughly linear in n. Throughout, we treat both n and ε as asymptotic quantities, and assume that $n \gg \varepsilon^{-1}$. We are concerned with the following very broad question: *What is the computational complexity of approximate range searching in spaces of constant dimension?* This line of thought raises a number of questions. What are the best ε-dependencies that can be achieved? How do various aspects of the problem formulation affect these dependencies? As mentioned above, in range searching we are computing some function of the weights of the points lying within a range. Such a function is commonly assumed to arise from a *faithful semigroup* over the domain of weights. We consider how semigroup properties affect the complexity of approximate (and exact) range searching.

A key issue in this regard seems to be idempotence. A semigroup is said to be *idempotent* if $x + x = x$ for all semigroup elements x. In contrast, if for all nonzero semigroup elements x and all natural numbers $k \geq 2$, the k-fold sum $x + \cdots + x$ is not equal to x then the semigroup is said to be *integral*. For example, $(\mathbb{R}, \min)$ and $(\{0, 1\}, \vee)$ are both idempotent semigroups, whereas $(\mathbb{R}, +)$ is integral. Idempotence is relevant because of the way that most range search algorithms work. At preprocessing time the algorithm implicitly computes the semigroup sum of a number of suitably chosen subsets of P, called *generators*. To answer a query η, the algorithm determines an (ideally small) set of generators whose union covers $P \cap \eta$, and then returns their total sum. If the semigroup is idempotent, these generators may overlap, but for integral semigroups they must be disjoint.

Because of the constraint of disjointness, one would expect that range searching over integral semigroups should be harder than for idempotent semigroups. It is remarkable, however, that for virtually all formulations of range searching, idempotence seems to be of no significant advantage. In their survey Agarwal and Erickson state, "Although in principle, storage schemes can exploit special properties of the semigroup, in practice, they never do. All known upper and lower bounds in the semigroup arithmetic model hold for all faithful semigroups." [1].

Our main result is that semigroup properties, idempotence in particular, do indeed make a difference in the complexity of both exact and approximate range searching. We show that for exact halfspace range searching, the assumption that the semigroup is integral allows us to prove a stronger lower bound in the arithmetic semigroup model than

*Department of Computer Science, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Supported by the Research Grants Council, Hong Kong, China (HKUST6184/04E). Email: arya@cs.ust.hk.

†Max-Planck-Institut für Informatik, Im Stadtwald, D-66123 Saarbrücken, Germany. Email: tmalamat@mpi-inf.mpg.de.

‡Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, Maryland 20742. Partially supported by the National Science Foundation under grant CCR-0098151. Email: mount@cs.umd.edu.

the one proved by BCP. For example, ignoring polylog factors, for the case of linear storage in the plane we provide a lower bound for integral semigroups of $\Omega^*(n^{2/5})$. This improves upon the existing lower bound of $\Omega^*(n^{1/3})$ [4], which holds for arbitrary semigroups. Our proof requires the assumption of *convex generators*, which states that for each generator G, we have $P \cap conv(G) = G$, where $conv(G)$ denotes the convex hull of G. (We conjecture that our bounds hold even without this assumption.)

Given the lengthy history of range searching, it is surprising that this fact has escaped notice until now. This may be because the lower bounds for exact integral range searching are only marginally better than the BCP bounds for arbitrary semigroups. We show, however, that the story is dramatically different for approximate range searching. We present lower bounds for approximate range searching for both types of semigroups. We also present nearly matching upper bounds for idempotent semigroups (and upper bounds for integral semigroups were given in [2]). These bounds show that, given $O(n)$ space, the advantages afforded by idempotence cause exponents in the ε-dependencies to be roughly *reduced by half*.

We consider space-time tradeoffs for this problem. Rather than expressing our space and time tradeoffs in the conventional manner, we adopt a notation that more clearly illustrates the dependencies on ε. Recall that n denotes the data size, and let m denote the space of the data structure. Let $\rho = m/n$ denote the *expansion ratio* of the data structure size over data size. We express query times as a ratio whose numerator gives the running time for the case of linear space ($m = O(n)$), and the denominator gives the *tradeoff rate*, that is, the rate with which query time decreases as a function of a multiplicative increase in space. Here is a summary of our results. We use the notation O^* and Ω^* to indicate the omission of polylogarithmic factors. All our lower bound results are for worst-case query time in the semigroup arithmetic model, assuming a faithful semigroup.

- We present a lower bound for exact halfspace range searching over any integral semigroup. Assuming convex generators, we show that the query time is at least $\Omega^*\left(n^{1-\frac{1}{d}-O\left(\frac{1}{d^2}\right)} \big/ \rho^{\frac{1}{d}+\frac{1}{d^2}}\right)$. By contrast, the BCP lower bound stated in this form is $\Omega^*\left(n^{1-\frac{2}{d}+O\left(\frac{1}{d^2}\right)} \big/ \rho^{\frac{1}{d}}\right)$.
- We present a lower bound for answering ε-approximate range queries for Euclidean balls over any arbitrary (including idempotent) semigroup. We show that the query time is at least $\Omega^*\left(\left(\frac{1}{\varepsilon}\right)^{\frac{d}{2}-1} \big/ \rho^{\frac{1}{2}-\frac{1}{2(d+1)}}\right)$.
- We present a lower bound for answering ε-approximate range queries for Euclidean balls over any integral semigroup under the convex generator assumption. We show that the query time is at least $\Omega^*\left(\left(\frac{1}{\varepsilon}\right)^{d-5} \big/ \rho^{1-\frac{4}{d}}\right)$.
- We present a data structure for answering ε-approximate range queries for Euclidean balls over any idempotent semigroup. We show that queries can be answered in time $O^*\left(\left(\frac{1}{\varepsilon}\right)^{\frac{d}{2}-\frac{1}{2d}} \big/ \rho^{\frac{1}{2}-\frac{1}{2d}}\right)$.

In [2] we showed that Euclidean ball range queries over integral semigroups can be answered with a space-time tradeoff of $O^*\left((1/\varepsilon)^{d-1} \big/ \rho^{1-\frac{1}{d}}\right)$. Our results are noteworthy for a number of reasons. First, they exhibit, for the first time to our knowledge, the impact of semigroup properties on the complexity of range searching. These are the first lower bound results for *any* approximate geometric retrieval problems. The existence of nearly matching upper bounds, throughout the range of space-time tradeoffs, suggests that we are close to resolving the computational complexity of both idempotent and integral approximate spherical range searching in the semigroup arithmetic model.

References

[1] P. K. Agarwal and J. Erickson. Geometric range searching and its relatives. In B. Chazelle, J. E. Goodman, and R. Pollack, editors, *Advances in Discrete and Computational Geometry*, volume 223 of *Contemporary Mathematics*, pages 1–56. American Mathematical Society, Providence, RI, 1999.

[2] S. Arya, T. Malamatos, and D. M. Mount. Space-time tradeoffs for approximate spherical range counting. In *Proc. 16th ACM-SIAM Sympos. Discrete Algorithms*, pages 535–544, 2005.

[3] S. Arya and D. M. Mount. Approximate range searching. *Computational Geometry: Theory and Applications*, 17:135–152, 2000.

[4] H. Brönnimann, B. Chazelle, and J. Pach. How hard is halfspace range searching. *Discrete Comput. Geom.*, 10:143–155, 1993.

[5] B. Chazelle, D. Liu, and A. Magen. Approximate range searching in higher dimension. In *Proc. 16th CCCG*, 2004.

[6] J. Matoušek. Range searching with efficient hierarchical cuttings. *Discrete Comput. Geom.*, 10(2):157–182, 1993.

[7] J. Matoušek. Geometric range searching. *ACM Comput. Surv.*, 26:421–461, 1994.