UNIVERSITY OF CALIFORNIA, RIVERSIDE

BERKELEY · DAVIS · IRVINE · LOS ANGELES · RIVERSIDE · SAN DIEGO · SAN FRANCISCO   SANTA BARBARA · SANTA CRUZ

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING   RIVERSIDE, CALIFORNIA 92521-0304

June 12, 2021

Review of Mr. Paweł Schmidt's PhD Dissertation

Thank you for giving me an opportunity to provide an evaluation of Mr. Schmidt's PhD thesis.

In his thesis, Mr. Schmidt focusses on several online optimization problems: facility location, minimum-cost matchings, Steiner trees, and server problems. All these problems and their variants have been extensively studied in the online setting, starting in late 1980s and early 1990s, and with still a steady stream of papers reporting new results. Several powerful techniques for tackling these problems, that emerged over this time, include randomized rounding of linear-programming relaxations, primal-dual methods, and hierarchically well-separated trees. These approaches are particularly useful for designing randomized algorithms, and indeed, combinations of these techniques yielded significant improvements in randomized competitive ratios for some of these problems.

At the same time, our understanding of the deterministic case has generally lagged behind, and Mr. Schmidt's dissertation, that focusses mostly on deterministic algorithms, will help make some progress towards designing better deterministic algorithms and proving better lower bounds on the competitive ratio.

The dissertation has four parts, that are discussed below.

Facility location. Facility location problems are among the most extensively investigated problems in combinatorial optimization, especially in the area of approximation algorithms, both in offline and online settings. Most of the earlier work focussed on the metric case. In the non-metric case, as explained in the thesis, an upper bound of $O((\log|C|+\log|F|)\cdot(\log|C|+\log\log|F|))$ can be achieved by reducing the problem to the online set cover problem. (Here, F and C represent the set of facilities and clients, respectively.) In Chapter 2, the thesis improves this bound by giving a deterministic algorithm with ratio $O(\log|F|\cdot(\log|C|+\log\log|F|))$. This is significant, because it reduces the dependence of the ratio on the number of clients, which is typically much larger than the number of facilities, from quadratic to linear in $\log|C|$. The algorithm is based on rounding a fractional solution. This is a general technique that is now frequently used in approximation algorithms, although its application is by no means routine. This is especially true in the online setting, where it requires choosing a suitable linear programming formulation, and designing a strategy for updating, in an online fashion, the solution of the linear program, and for rounding the fractional solution to an integer solution.

Matchings. Various versions of online matching problems have been widely studied in the past, as they arise naturally in practical scenarios and involve interesting algorithmic challenges. In Chapter 3, the thesis addresses the minimum-cost perfect matching variant that also takes into account how long nodes need to wait for their mates. So there are actually two objectives: the matching cost and the delay cost, both to be minimized. This is captured by the objective function representing the sum of these costs. For this problem, based on the state-of-the-art, randomization appears to be essential in order to achieve a good approximation ratio. A randomized algorithm with ratio $O(\log n)$ is known, almost matching a known lower bound. But very little has been known about deterministic algorithms. Addressing this, the thesis provides an online deterministic algorithm with ratio $O(m^{2.46})$, where m is the number of requests. Competitive ratios dependent on the request sequence are not quite consistent with the spirit of competitive analysis (nor with the definition on page 5), but they still provide useful information about online approximability, showing that dependence on the distances and time delays can be avoided. This result also opens the possibility that bounds that depend only on n (the number of points) may be also possible, or perhaps even on the minimum of m and n. There is hope that the techniques and insights from this work will inspire further research and lead to even better algorithms.

Steiner trees. Another example of classical combinatorial optimization problems is the Steiner tree problem. It has multiple variants, one of which is the online Steiner tree leasing problem, considered in Chapter 4. The thesis provides the first online deterministic algorithm for this problem that works on arbitrary metrics. Its competitive ratio is $O(L \cdot \log k)$, where L is the number of lease types and k is the number of terminals. One nice feature of this algorithm is that the competitive ratio of this algorithm is independent of the metric space, and, when k is sufficiently small compared to n and L, this ratio is even better than that of the known $O(\log L \cdot \log n)$-competitive randomized algorithm. The analysis is interesting, as it involves a novel use of hierarchically well-separated trees – somewhat unusual in a deterministic setting.

Generalized server problems. The last part of the thesis, Chapter 5, presents some results about the generalized k-server problem. The k-server problem is the "flagship" problem of competitive analysis. Posed originally in 1987 and still not fully solved, it continues to be a topic of intense study. Its generalized variant, where each server moves in its own metric space, is even harder, and very little is known about its online approximability in arbitrary metrics. This chapter focuses on randomized algorithms for this problem in the uniform metric space. The first result improves the randomized competitive ratio from $O(k^3 \log k)$ to $O(k^2 \log k)$, using an elegant analysis via the so-called Hydra game. This chapter also includes a lower bound of $\Omega(k)$ (improving $\Omega(k/\log k)$). The proof, also quite elegant, is based on an interesting idea of using a large set of configurations that resembles an error correcting code.

The dissertation is very nicely written. It is well organized, contains just about right amount of background information, and the technical content is presented in rigorous, yet lucid fashion. Some minor comments and suggestions about the writing are attached at the end of this report.

Conclusion. In summary, Mr. Schmidt's work addresses algorithmic problems that are of central importance in online discrete optimization. The presented results are novel, technically challenging, they require new insights and ideas, and advance the state-of-the-art in the field by tightening the gaps between the lower and upper bounds on the competitive ratios for these problems. With this in mind, I believe that his thesis meets and exceeds the requirements for a PhD Dissertation.

Sincerely,

Marek Chrobak
Professor of Computer Science
University of California, Riverside

Minor comments.

p2: “using of online” $\Rightarrow$ “using online”

p2: “a fractional solutions” $\Rightarrow$ “fractional solutions”

p3: “performance of deterministic algorithm” $\Rightarrow$ “performance of a deterministic algorithm” (There are multiple places in the thesis with questionable choice of English articles.)

p4: “While the most of deterministic” $\Rightarrow$ “While most of deterministic”

p4: “a good approximate” $\Rightarrow$ “a good approximation”

p5: “independent of the online part of the input”. Explain what it means? The order of quantifiers already makes β independent of $\mathcal{I}$.

p6: “problem [Shmoo], which have” $\Rightarrow$ “problem [Shmoo], that has”

p8: “that are functions of k” $\Rightarrow$“functions of k only”?. Also on page 68.

p9: “””has not drawn much attention”. It is not hard to show that it’s NP-hard, even if all k spaces are the same uniform space.

p23: “whether a given client”. It’s a little unclear what client this refers to. I assume it’s c, mentioned at the beginning of Section 2.3.

p38: “path and cycles” $\Rightarrow$ “paths and cycles”.

p44: “cost ... contribute” $\Rightarrow$ “cost ... contributes”

p49: “$O(\log^2 k)$-deterministic” $\Rightarrow$ “a deterministic $O(\log^2 k)$-competitive”

p50: “follows by the online Steiner tree problem”. Follows by a problem?

p52: “for any time t and any edge”. Also, for any ℓ?

p53: “leasing the edge (u, v)”. Another example of questionable use of articles. In this case, it should definitely be “an edge”.

p69: Lemma 5.4. Notation clash: earlier q_i’s were used for coordinates of q_i. Also, the term “compatible” is not defined with respect to vectors with wildcards. (Given the common interpretation of wildcards, some readers might mistakenly think that wildcard symbols match any point in M.)