

Approximation Algorithms for Facility Location Problems and Other Supply Chain Problems*

Lehilton L. C. Pedrosa¹,
Flávio K. Miyazawa¹ (Supervisor), Maxim Sviridenko² (Co-Supervisor)

¹Institute of Computing – University of Campinas
Campinas, São Paulo

²Department of Computer Science – University of Warwick
Coventry, UK

1. Introduction

Approximation algorithms have fascinated an ever-greater community of researchers working on optimization and theory of computing. Whereas the so called NP-hard problems cannot be efficiently solved under the largely believed assumption that $P \neq NP$, the approximation algorithms provide a compromise between the accuracy of a solution, and the time used to compute it. The challenge of both practitioners and theorists is to obtain a solution as close as possible to the optimum, while maintaining the running time small. The study of approximation algorithms helps not only unveiling which problems can be solved efficiently, but also how well they can be solved.

In a combinatorial optimization problem, the objective is to minimize or maximize the value of a given function over some finite domain. In such a problem, one looks for an optimal solution, that is a feasible solution with maximum or minimum value. For example, one might be interested in finding a path between two locations with minimum length, or selling items to several clients according to the distribution that maximizes the profit. While finding the best solution may be difficult, or even not viable, an acceptable solution can be obtained in many cases. The trade-off between the solution quality and the algorithm's running time is normally reflected on the ratio between the values of generated and optimal solutions. For an *approximation algorithm*, this ratio has a proven bound, and the running time is always polynomial.

Many optimization problems appear in the decision-making process of industries of several areas, and thus obtaining fast near-to-optimal solutions is essential to one company's success, as it might represent significant economy, and overall costs reduction. *This thesis* investigated several optimization problems that appear in the management of an organization's supply chain and related activities. It comprises numerous tasks that extend from the configuration of a supply chain network, up to the packing, transportation, and inventory control of goods. A wide list of NP-hard problems were considered, that may be grouped according to their objective: *location*, *inventory management*, and *packing*. The class of location problems include, for example, the Facility Location Problem (FLP), and clustering problems, such as k -means or k -medians. For the inventory problems, examples are the One-Warehouse Multiple-Retailer Problem (OWMR), and the Joint Replenishment Problem (JRP). Finally, packing problems include the Circle Bin Packing and the Circle Strip Packing problems.

*This thesis has been partially supported by FAPESP grants 2010/20710-4, 2012/17634-0.

We gave approximation algorithms in each considered area, for both classical problems, such as metric FLP, and many other important variants, such as non-metric FLP. Also, several novel techniques for approximation algorithms were developed. While this work focused on the formal aspects of algorithms, the considered problems — and thus the corresponding obtained results — have meaningful interest in practice. Indeed, they appear in a wide range of applications. The FLP models network projects, such as the installation of cache and routers, or data grouping and server replication. Clustering problems, such as k -means, have diverse applications in image compression, distribution of resources, cellular biology, etc. Problems that consider exponents in Euclidean distance function find application in the power attribution of wireless networks. The JRP and the OWMR appear in the inventory management of many multi-product environments, etc.

Obtaining approximations is a challenging task that varies widely according to the problem being studied. For problems that have a rich literature in the approximation area, such as the classical Metric FLP, the k -means clustering, and the packing of rectangular items, the difficulty lays on improving state-of-art results (both approximations and bounds on the approximability), and solving standing open questions. For problems with just few or no works on approximation, as is the case of some inventory management problems, and the packing of non-rectangular items, the obstacle is the scarcity of known techniques to tackle the problem, and the lack of indications of a problem's hardness. Even for heavily studied problems, there is a lengthy list of relevant variants that are underexplored from the perspective of approximation algorithms. This thesis aimed at filling such gaps, and provided contributions of diverse natures: *new algorithms and constant approximations, improved analysis of approximation factors, novel techniques, lower bound on approximation factors, approximation-preserving reductions*, etc.

The following sections describe the main problems tackled in the thesis, and list the obtained results. The sections are organized according to the class of problems considered. Whilst this summary only describes the results informally, and is addressed to the general reader, precise problem definitions and formal result statements are available in the thesis text, as well as the full list of references. The final section lists publications and discusses the thesis's impact.

2. Location Problems

2.1. The Facility Location Problem

In the FLP, we are given sets of clients and facilities, and the objective is to open a subset of facilities, and assign clients to open facilities, such that the total opening and connection costs are minimized. For the general case, there is an $O(\log(n))$ -approximation. A very studied particular case is the so called metric variant, when the distance function satisfies the triangle inequality. For this variant, there is no approximation with factor better than 1.463, unless $\text{NP} \subseteq \text{DTIME}[n^{O(\log \log n)}]$, what is unlikely. For positive results, there is a long series of algorithms that culminates with the best known 1.488-approximation.

In the thesis, we addressed the squared metric distance function, that is not as restrictive as the triangle inequality, neither too general. Previously, the only known approximation had factor 9 [Jain and Vazirani 2001]. First, we gave a hardness result.

Theorem 1. *No $(2.04 - \epsilon)$ -approximation for Squared Metric FLP exists, unless $\text{P} = \text{NP}$.*

We also gave an approximation through a linear programming formulation (LP).

Theorem 2. *There is a 2.04-approximation for the Squared Metric FLP.*

Notice that this factor matches exactly the lower bound of the problem, and thus is the best possible factor unless $P = NP$. In fact, we gave best possible factors to the case of more general distances, when the triangle inequality terms are multiplied by a constant τ .

Theorem 3. *For $\tau \in (2.62, 31.02)$, the standard LP-rounding for the τ -relaxed FLP has a constant approximation. Moreover, the constant is the best possible, unless $P = NP$.*

Perhaps more important than obtaining tight factors was the study of the primal-dual algorithms of [Jain et al. 2003], that, although have slightly worse approximation factors, are very efficient when compared to the LP-rounding approach. The analysis of these algorithms used the so called factor-revealing LP's. Solving one such LP gives a bound z_k on the factor for instances of size k . The approximation factor, on the other hand, were hand-calculated by complicated and tedious proofs, that depended on many guessing steps, and demanded several pages of verification. We introduced the *upper bound factor-revealing program* (UPFRP), whose value x_t is solved by a computer and gives the approximation factor directly, for any t . Obtaining an UPFRP is more systematic, and does not require guessing steps as hand-made proofs.

Theorem 4. *For any fixed t , let x_t be value of the UPFRP obtained for FLP. Then x_t is a bound of the approximation factor of the algorithm of Jain et al. Moreover, x_t converges to the absolute approximation factor, as $t \rightarrow \infty$.*

The convergence means that the obtained factor may be tightened if we compute an UPFRP for larger t . We obtained tight analyses for both squared metric and metric cases. In particular, we improved the analysis of an algorithm by Jain et al., thus improving approximation factors for other algorithms that use the latter as a subroutine.

2.2. Continuous Facility Location Problem and k-Clustering

The Continuous FLP (ConFL) is a variant of FLP whose set of clients are points of the d -dimensional Euclidean space \mathbb{R}^d , and the set of facilities is the complete \mathbb{R}^d space. This is a clustering problem similar to k -means or k -medians, except that, instead of fixing a number k of clusters, there is a cost f for creating a cluster. The previously best known approximation factors were 1.861 and 9, when the distance functions are norms ℓ_2 and ℓ_2^2 , respectively. We showed that ConFL can be reduced to FLP, provided the existence of *centers sets*: sets that contain near-optimal centers for the considered distance function.

Theorem 5. *Assume there is an algorithm to generate center sets for a distance function c . If there is a β -approximation for Euclidean FLP with distance c , then there is a $(\beta + \varepsilon)$ -approximation for the ConFL with distance c .*

We combined this with known center sets to improve approximations for ConFL.

Theorem 6. *There is a $(1 + \varepsilon)$ -approximation for ℓ_2 -ConFL with fixed d , and a $(1.488 + \varepsilon)$ -approximation for arbitrary d . There is a $(2.04 + \varepsilon)$ -approximation for ℓ_2^2 -ConFL.*

We developed center sets for the ℓ_2^α norm, for any $\alpha \geq 1$. This can be used by applications that might set α as parameter, and obtain a refined control of the distance function. Combined with Theorem 3, we obtained approximations for the ℓ_2^α -ConFL.

Theorem 7. *There are constant approximations for the ℓ_2^α -ConFL, for every $\alpha \geq 1$.*

Moreover, the obtained center sets were used to obtain the first $(1 + \varepsilon)$ -approximation (PTAS) for the k -clustering problem with ℓ_2^α norm. This problem generalizes both the continuous k -medians and k -means, when $\alpha = 1$ and $\alpha = 2$, respectively.

Theorem 8. *For every $\alpha \geq 1$, there is a PTAS for the ℓ_2^α k -clustering.*

One possibility to obtain approximations for ConFL is using center sets, and solving k -medians for each $k = 1, \dots, n$. Next, we showed that such strategy cannot improve the $2.04 + \varepsilon$ factor for ℓ_2^2 -ConFL, as squared metric k -medians has a 3.943 lower bound.

Theorem 9. *There is no β -approximation algorithm for the Squared Metric k -medians with $\beta < 1 + 8/e \approx 3.943$, unless $P = NP$.*

3. Supply Chain Problems

3.1. Production and Distribution Problem

We considered a series of supply chain problems. The first was a common generalization of FLP and JRP, that is called the Production and Distribution Problem (PDP). In PDP, transportation and holding costs must be minimized in an integrated manner, what may represent considerable economy to using FLP and JRP separately. Namely, we are given sets of retailers and warehouses. Each retailers may face a demand at a given time, which must be satisfied by the items held in the inventory, and previously ordered from one warehouse. The objective is defining an inventory policy that minimizes total warehouse ordering, transportation and holding costs. No approximations were known before, and most related works deal only with heuristics, and integers formulations. We gave an approximation based on a new LP formulation, and the corresponding rounding algorithm.

Theorem 10. *There is a 2.77-approximation for the PDP.*

We also considered a more general, and thus harder, version of PDP. In this variant, in addition to warehouse ordering cost, each retailer order also incurs a setup cost. We studied this problem with additive holding costs, and gave a constant approximation.

Theorem 11. *There is a 5-approximation for the PDP with retailer ordering costs.*

3.2. The One-Warehouse Multiple-Retailer Problem

The OWMR is a classical generalization of JRP, when the warehouse may also hold items in the stock (whereas only retailers have inventory in JRP). There is a 1.8-approximation based on LP-rounding by [Levi et al. 2008]. This approximation, however, makes the unnatural assumption that warehouse and retailer holding cost functions are dependent, thus limiting the algorithm's application. We considered the problem when warehouse and retailers holding costs are independent. Namely, we gave constant approximation for the (more general) OWMR when holding costs need to satisfy only natural monotonicity and subadditivity properties.

Theorem 12. *There is a 5-approximation for the OWMR with subadditive holding costs.*

We remark that our algorithm is based on the primal-dual method. This answers positively an open question left by [Levi et al. 2008], who asked whether the primal-dual approach could be extended to work with OWMR. Moreover, the algorithm is based on a novel dual ascent technique that may find applications in other problems.

Theorem 13. *The OWMR admits an approximation via a new primal-dual approach with non-uniform dual ascent.*

3.3. Multilevel Joint Replenishment Problem

In JRP, a retailer must order items to the warehouse. In the 2-level JRP, a warehouse also order items from a producer, that is higher in the supply chain network. More generally, in the Multilevel JRP, we are given a supply chain tree, and each non-root node may order to its parent. This very natural generalization of the JRP had no constant approximation before. We showed that Multilevel JRP can be reduced to the Multistage Assembly Problem, that is the problem of assembling items of a given kind according to a tree of components, while minimizing the total cost of holding parts and assembling components.

Theorem 14. *It there is a β -approximation for Multistage Assembly Problem, then there is a $(\beta + \varepsilon)$ -approximation for Multilevel JRP.*

Using this reduction, we thus obtained an approximation for Multilevel JRP.

Theorem 15. *There is a $(2 + \varepsilon)$ -approximation for Multilevel JRP.*

4. Circle Packing Problems

We also considered the Circle Bin Packing Problem, whose objective is to pack a set of circles into the minimum number of unit square bins. Most of the works which give approximation guarantees for packing problems deal only with rectangular items or d -dimensional boxes. The packing problems involving circles are mainly considered through heuristics, or numerical methods, and, to our knowledge, there were no approximation algorithms for the Circle Bin Packing before. We give the first approximation for this problem. Indeed, we give an asymptotic PTAS for the problem.

Theorem 16. *There is an algorithm that, given a set of circles \mathcal{C} , generates a packing into no more than $(1 + \varepsilon)\text{OPT}(\mathcal{C}) + 2$ bins of unit width and height $1 + \gamma$, for any $\gamma > 0$, where $\text{OPT}(\mathcal{C})$ is the number of bins in an optimal solution. The algorithm is polynomial on $(1/\gamma)$, and on the size of \mathcal{C} .*

Notice that we used slightly augmented bins, since the solutions for circle bin packing involve solutions of algebraic equations. To obtain rational solutions, it is necessary to approximate the coordinates. It is currently widely open whether every circle bin packing instance admits a rational solution. Nevertheless, if the augmentation parameter γ is a constant ε , then it is easy to obtain a resource augmentation scheme, as follows.

Theorem 17. *There is a polynomial-time algorithm that, given a set of circles \mathcal{C} , generates a packing into no more than $\text{OPT}(\mathcal{C})$ bins of unit width and height $1 + \varepsilon$, for any constant $\varepsilon > 0$.*

Our algorithm can also be extended to the Circle Strip Packing, that is the problem of finding a packing of circles into a unit strip with minimum height.

Theorem 18. *There is a polynomial-time algorithm that, given a set of circles \mathcal{C} , generates a packing into a strip of unit width and height $(1 + \varepsilon)\text{OPT}(\mathcal{C}) + O(1)$, where $\text{OPT}(\mathcal{C})$ is the height of an optimal solution.*

5. Summary and Impact

Papers We submitted the results of this thesis to highly prestigious conferences and journals in the area of Approximation Algorithms and Theoretical Computer Science. In addition the next list, other three complete journal articles with thesis results have already been submitted, and are in the process of acceptance. The list of currently published papers follows:

- [Fernandes et al. 2012] *Conference APPROX-RANDOM 2012*: presents the Squared Metric Facility Location, and gives best possible approximation
- [Pedrosa and Sviridenko 2014] *Conference LATIN 2014*: presents the Production and Distribution Problem, and gives the first constant approximation
- [Miyazawa et al. 2014] *Conference ESA 2014*: presents the first approximation algorithms for the circle packing problems
- [Fernandes et al. 2014] *Journal Mathematical Programming*: details the upper bound factor revealing technique, the application to Metric and Squared Metric Facility Location Problems, and studies more general distance functions

Impact This thesis’s impact and relevance may be measured by the extensive set of results, that examined a large list of NP-hard problems with the broadest objectives. Also — and more importantly —, each problem was deeply studied, always with the goal of having a good understanding of a problem’s structure, hardness, and approximability. A notable example is the Squared Metric FLP, for which we provided a lower bound on the approximation factor, and showed an approximation that matched this bound. Furthermore, the developed techniques, that were initially used for problems considered here, incorporated into the set of tools of the approximation algorithms machinery, and may be used for other problems as well. For instance, in the very recent paper by [Wu et al. 2014], that improves the best known approximation for the metric k -medians, an upper bound factor revealing program is used as a standard available technique.

References

- Fernandes, C. G., Meira, L. A., Miyazawa, F. K., and Pedrosa, L. L. (2014). A systematic approach to bound factor-revealing LPs and its application to the metric and squared metric facility location problems. *Mathematical Programming*, pages 1–31. In press.
- Fernandes, C. G., Meira, L. A. A., Miyazawa, F. K., and Pedrosa, L. L. C. (2012). A Systematic Approach to Bound Factor Revealing LPs and Its Application to the Metric and Squared Metric Facility Location Problems. In *APPROX 2012*, pages 146–157.
- Jain, K., Mahdian, M., Markakis, E., Saberi, A., and Vazirani, V. V. (2003). Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. *Journal of the ACM*, 50(6):795–824.
- Jain, K. and Vazirani, V. V. (2001). Approximation algorithms for metric facility location and k -Median problems using the primal-dual schema and Lagrangian relaxation. *Journal of the ACM*, 48(2):274–296.
- Levi, R., Roundy, R. O., Shmoys, D. B., and Sviridenko, M. (2008). A Constant Approximation Algorithm for the One-Warehouse Multiretailer Problem. *Management Science*, 54(4):763–776.
- Miyazawa, F., Pedrosa, L., Schouery, R., Sviridenko, M., and Wakabayashi, Y. (2014). Polynomial-Time Approximation Schemes for Circle Packing Problems. In *ESA 2014*, pages 713–724.
- Pedrosa, L. L. C. and Sviridenko, M. (2014). Integrated Supply Chain Management via Randomized Rounding. In *LATIN 2014*, pages 562–573.
- Wu, C., Xu, D., Du, D., and Wang, Y. (2014). An improved approximation algorithm for k -median problem using a new factor-revealing LP. Arxiv Preprint.