Minimum-Cost Load-Balancing Partitions

Paz Carmi* Matthew J. Katz†

Abstract

We consider the problem of balancing the load among m service-providing facilities, while keeping the total cost low. Let R be the underlying demand region, and let $p_1, \ldots, p_m$ be m points representing m facilities. We consider the following problem. Divide R into m subregions $R_1, \ldots, R_m$, each of area $area(R)/m$, such that region R_i is served by facility p_i, and the average distance between a point q in R and the facility that serves q is minimal. We present constant-factor approximation algorithms for this problem.

1 Introduction

Given m facilities we would like to balance the load among these facilities, while keeping the total cost low. In other words, let R denote the region in the plane that must be served by the facilities. We would like to divide R into m subregions $R_1, \ldots, R_m$, each of area $area(R)/m$, such that all requests initiated by clients in R_i are served by the i-th facility, and the total cost is minimal.

For example, if the facilities are fire stations. Then, on the one hand, we would like the average distance between a point q in R and the fire station that is responsible for q to be as small as possible, and, on the other hand, we would like to balance the load among the fire stations.

We thus consider the following problem. Let R denote the underlying demand region, and let $p_1, \ldots, p_m$ be m points representing m facilities. Put $\mathcal{P} = \{p_1, \ldots, p_m\}$. One needs to divide R into m subregions $R_1, \ldots, R_m$, each of area $area(R)/m$, such that region R_i is associated with point p_i, and the total cost of the division is minimal. Given a division, the cost associated with facility p_i, $\mu(p_i)$, is the average distance between p_i and a point in R_i, and the total cost of the division is $\sum_i \mu(p_i)$.

Without the load-balancing requirement, one can simply compute the Voronoi diagram of $\mathcal{P}$ (restricted to R) and associate each facility with its Voronoi cell, in order to obtain an optimal solution. However, with the load-balancing requirement the problem becomes much more difficult.

In this paper we describe an algorithm that, under reasonable and natural assumptions on R and on the subregions, divides R into m subregions (each of area $area(R)/m$) and associates them with the facilities in $\mathcal{P}$. We call the division obtained GRID-MIN-SUM, and prove that GRID-MIN-SUM is a $(8 + \frac{3}{2}\sqrt{2\pi})$-approximation. That is, the total cost of GRID-MIN-SUM is at most $(8 + \frac{3}{2}\sqrt{2\pi})$ times the cost of an optimal division. The running time of our algorithm is $O(n^3)$. We also show how to reduce the running time to roughly $O(n^2)$ without increasing the approximation factor too much.

2 The Algorithm

Let us formulate the problem more precisely. We assume that the number m of facilities is equal to l^2, for some integer $l \geq 2$, and that each facility is represented by a point p_i in the plane. Put $\mathcal{P} = \{p_1, \ldots, p_m\}$. We also assume (for convenience only) that the underlying region R is a square. (This assumption is not necessary; all subsequent results hold for any rectangle R that can be divided into m squares of equal size. Thus the aspect ratio of R can be as large as m.) We study the following problem. Divide R into m regions $R_1, \ldots, R_m$, each of area $area(R)/m$, such that region R_i is associated with point p_i, and the total cost is minimal. Where the cost $\mu(p_i)$ associated with facility p_i is the average distance between p_i and a point in R_i, and the total cost of the division is $\sum_i \mu(p_i)$.

We describe a simple algorithm that divides R into m subregions and associates them with the points in $\mathcal{P}$. We call the division obtained GRID-MIN-SUM. We then prove that GRID-MIN-SUM is $(8+\frac{3}{2}\sqrt{2\pi})$-approximation, under the assumption that the subregions must be convex.

The algorithm divides R into $m = l^2$ squares of equal size, and associates the squares with the points in $\mathcal{P}$. Let $\mathcal{S}$ denote the set of squares $\sigma_1 \ldots \sigma_m$. Let $G = (\mathcal{P}, \mathcal{S}; E)$ be the complete bipartite graph with vertex sets $\mathcal{P}$ and $\mathcal{S}$. We associate weights with the edges in E. The weight of the edge (p, σ) is the average distance between p and the points in σ. The weight of an edge can be computed by integration in $O(1)$ time. We now

*Department of Computer Science, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel, carmip@cs.bgu.ac.il. Partially supported by grant no. 2000160 from the U.S.-Israel Binational Science Foundation, and by a Kreitman Foundation doctoral fellowship.

†Department of Computer Science, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel, carmip@cs.bgu.ac.il. Partially supported by grant no. 2000160 from the U.S.-Israel Binational Science Foundation.

associate the squares with the points by computing a minimum-weight matching in G, i.e., a matching for which the sum of the weights (of the m edges defining the matching) is minimal. Using the algorithm of Kuhn [4] this can be done in $O(n^3)$ time.

We next prove that the division that was obtained (i.e., GRID-MIN-SUM) is a constant-factor approximation.

2.1 GRID-MIN-SUM is a constant-factor approximation

Let opt denote an optimal division, i.e., a minimum-cost load-balancing partition of R into convex subregions, where region R_i is associated with point p_i, $i = 1, \ldots, m$. We use opt to obtain a new division, grid, that is also based on the squares in $\mathcal{S}$. We then show that grid is a $(8+\frac{3}{2}\sqrt{2\pi})$-approximation, immediately implying that GRID-MIN-SUM is also a $(8+\frac{3}{2}\sqrt{2\pi})$-approximation, since GRID-MIN-SUM is the best division among those based on the squares in $\mathcal{S}$.

Define a bipartite graph $G = (\mathcal{S}, R; E)$, where $\mathcal{R} = \{R_1, \ldots, R_m\}$ is the set of regions of opt, and there is an edge between $\sigma_i \in \mathcal{S}$ and $R_j \in \mathcal{R}$ if and only if $\sigma_i \cap R_j \neq \emptyset$. Hall's matching theorem [5] gives a necessary and sufficient condition for G to contain a perfect matching. According to Hall's matching theorem, G contains a perfect matching if and only if for any subset $\mathcal{S}'$ of $\mathcal{S}$ we have $|N(\mathcal{S}')| \geq |\mathcal{S}'|$, where $N(\mathcal{S}')$ is the set of regions in $\mathcal{R}$ that are connected by an edge to a square in $\mathcal{S}'$. However, this condition trivially holds in our case, since we need at least $|\mathcal{S}'|$ regions of $\mathcal{R}$ in order to cover a region of area $|\mathcal{S}'|area(R)/m$. We thus associate the squares in $\mathcal{S}$ with the points in $\mathcal{P}$ to obtain the division grid by computing any perfect matching in G (if square σ_j was matched to region R_i, then σ_j is associated with point p_i).

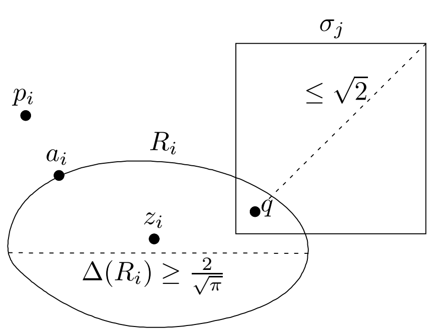

Figure 1: grid is a constant-factor approximation.

We now show that grid is a $(8+\frac{3}{2}\sqrt{2\pi})$-approximation, immediately implying that GRID-MIN-SUM is also a $(8+\frac{3}{2}\sqrt{2\pi})$-approximation, since GRID-MIN-SUM is the best division among those based on the squares in $\mathcal{S}$.

Let $p_i \in \mathcal{P}$ and let $\sigma_j \in \mathcal{S}$ be the square assigned to p_i by grid; see Figure 1. Let $q \in \sigma_j \cap R_i$, where R_i is the region assigned to p_i by opt. Assume w.l.o.g. that $area(\sigma_j) = area(R_i) = 1$, then the diameter of R_i, $\Delta(R_i)$, is at least $2/\sqrt{\pi}$. Let z_i be a point in the plane for which the average distance to the points in R_i is minimal. z_i is the *Fermat-Weber center* of R_i. Since R_i is convex, $d_{avg}(z_i, R_i) \geq \Delta(R_i)/7$ (See [3]), where $d_{avg}(z_i, R_i)$ is the average distance between z_i and the points in R_i, and also $d_{avg}(z_i, R_i) \geq 2/(3\sqrt{\pi})$. (The right side of the latter inequality is equal to the average distance between the center of a disc of radius $1/\sqrt{\pi}$ and the points of the disc.) Let $a_i \in R_i$ be the closest point to p_i (if $p_i \in R_i$, then $a_i = p_i$). Then the cost $\mu(p_i)$ of p_i in opt is, on the one hand, at least $d_{avg}(z_i, R_i)$, and, on the other hand, at least $||p_i a_i||$. As to $\mu(p_i)$ in grid we have $\mu(p_i) \leq ||p_i a_i|| + ||a_i q|| + \sqrt{2} \leq ||p_i a_i|| + \Delta(R_i) + \sqrt{2}$.

Now, if $||p_i a_i|| \geq d_{avg}(z_i, R_i)$, then using the second inequality for $\mu(p_i)$ in opt (and noticing that in this case $\Delta(R_i) \leq 7||p_i a_i||$) we obtain that

$$\frac{\mu_{\mathsf{grid}}(p_i)}{\mu_{\mathsf{opt}}(p_i)} \leq \frac{||p_i a_i|| + \Delta(R_i) + \sqrt{2}}{||p_i a_i||} \leq \frac{8||p_i a_i|| + \sqrt{2}}{||p_i a_i||} \leq$$

$$\leq 8 + \frac{\sqrt{2}}{d_{avg}(z_i, R_i)} \leq 8 + \frac{3}{2}\sqrt{2\pi} \ ,$$

and, if $||p_i a_i|| < d_{avg}(z_i, R_i)$, then using the first inequality for $\mu(p_i)$ in opt we obtain that

$$\frac{\mu_{\mathsf{grid}}(p_i)}{\mu_{\mathsf{opt}}(p_i)} \leq \frac{||p_i a_i|| + \Delta(R_i) + \sqrt{2}}{d_{avg}(z_i, R_i)} \leq$$

$$\leq \frac{8 d_{avg}(z_i, R_i) + \sqrt{2}}{d_{avg}(z_i, R_i)} \leq 8 + \frac{3}{2}\sqrt{2\pi} \ .$$

We conclude that in both cases the ratio between $\mu(p_i)$ in grid and $\mu(p_i)$ in opt is at most $8+\frac{3}{2}\sqrt{2\pi}$, for any $1 \leq i \leq m$, and therefore grid is a $(8+\frac{3}{2}\sqrt{2\pi})$-approximation. Finally, since the cost of GRID-MIN-SUM is at most the cost of grid, we conclude that GRID-MIN-SUM is a $(8 + \frac{3}{2}\sqrt{2\pi})$-approximation.

Theorem 1 *A division of R that is a $(8 + \frac{3}{2}\sqrt{2\pi})$-approximation can be computed in $O(n^3)$ time.*

2.2 Improving the running time

Consider the complete bipartite graph $G = (\mathcal{P}, S; E)$ in which we compute a minimum weight matching to obtain the division GRID-MIN-SUM. By modifying the definition of the weight of an edge $(p, \sigma) \in E$, we can both simplify the computation of the edge weights and reduce the running time of our algorithm to $O(n^{2+\epsilon})$, without increasing the approximation factor too much. We define the weight of (p, σ) to be the distance between p and the center of σ. Now, the graph G is actually the complete bipartite graph induced by two point sets in the plane, and we can apply to it the algorithm of Agarwal et al. [1] that computes a minimum weight matching in such graphs in time $O(n^{2+\epsilon})$ using $O(n^{1+\epsilon})$ space.

It remains to bound the approximation factor of the division GRID-MIN-SUM' that is obtained. We show that the cost of GRID-MIN-SUM' is at most the cost of GRID-MIN-SUM plus $m\sqrt{2}$, and therefore GRID-MIN-SUM' is a $(8+3\sqrt{2\pi})$-approximation (using the inequality $\mu(\mathsf{opt}) \geq 2m/(3\sqrt{\pi})$). Indeed, let M (resp., M') be the matching defining GRID-MIN-SUM (resp., GRID-MIN-SUM'). Also, for a point p_i and a square σ_j, let $q_j^i \in \sigma_j$ be the closest point to p_i in σ_j, and let o_j be the center of σ_j. Then

$$m\sqrt{2} + \sum_{(p_i,\sigma_j)\in M} d_{avg}(p_i,\sigma_j) \geq$$

$$\geq \frac{m\sqrt{2}}{2} + \sum_{(p_i,\sigma_j)\in M} (||p_i q_j^i|| + \frac{\sqrt{2}}{2}) \geq$$

$$\geq \frac{m\sqrt{2}}{2} + \sum_{(p_i,\sigma_j)\in M} ||p_i o_j|| \geq \frac{m\sqrt{2}}{2} + \sum_{(p_i,\sigma_j)\in M'} ||p_i o_j|| =$$

$$= \sum_{(p_i,\sigma_j)\in M'} (||p_i o_j|| + \frac{\sqrt{2}}{2}) \ .$$

But the first expression in the above sequence of inequalities is the cost of GRID-MIN-SUM plus $m\sqrt{2}$, and the last expression is greater or equal than the cost of GRID-MIN-SUM'.

Theorem 2 *A division of R that is a $(8+3\sqrt{2\pi})$-approximation can be computed in $O(n^{2+\epsilon})$ time.*

Proof. Using the inequality $\mu(\mathsf{opt}) \geq 2m/(3\sqrt{\pi})$,

$$\frac{\mu(\text{grid-min-sum}')}{\mu(\mathsf{opt})} \leq \frac{\mu(\text{grid-min-sum}) + m\sqrt{2}}{\mu(\mathsf{opt})} \leq$$

$$\leq (8 + \frac{3}{2}\sqrt{2\pi}) + \frac{m\sqrt{2}}{2m/(3\sqrt{\pi})} = 8 + 3\sqrt{2\pi} \ .$$

□

3 Concluding Remarks

Remark 1 In [2] Section 5, a generalization of the following related problem was studied. In this problem the cost associated with a facility p_i is the maximum distance between p_i and a point in its subregion, and the total cost of the division is the maximum over the costs of the facilities. There too a constant factor approximation algorithm is presented.

Remark 2 Very recently we managed to generalize the results of this paper to regions R that are convex and fat (rather than rectangular). This required an efficient algorithm for partitioning such a region R into m convex and fat subregions, each of area $area(R)/m$.

References

[1] P. K. Agarwal, A. Efrat and M. Sharir. Vertical decomposition of shallow levels in 3-dimensional arrangements and its applications. *SIAM Journal on Computing*, 29(3) (1999), 912–953.

[2] P. Carmi, S. Dolev, S. Har-Peled, M. J. Katz and M. Segal. Geographic quorum system approximations. *Algorithmica*, 41(4) (2005), 233–244.

[3] P. Carmi, S. Har-Peled and M. J. Katz, On the Fermat-Weber center of a convex object. *Comput. Geom. Theory Appl.*, to appear.

[4] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2 (1955), 83–97.

[5] D. B. West. *Introduction to Graph Theory*. Second edition, Prentice Hall, 2001.