ELSEVIER

Available at www.ComputerScienceWeb.com
POWERED BY SCIENCE @ DIRECT•

Information Processing Letters 85 (2003) 117–122

Information Processing Letters

www.elsevier.com/locate/ipl

Minimizing the sum of the k largest functions in linear time

Wlodzimierz Ogryczak [a], Arie Tamir [b,*]

[a] *Institute of Control and Computation Engineering, Warsaw University of Technology, 00-665 Warsaw, Poland*
[b] *School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel*

Received 28 November 2001; received in revised form 29 July 2002

Communicated by M. Yamashita

Abstract

Given a collection of n functions defined on $\mathbb{R}^d$, and a polyhedral set $Q \subset \mathbb{R}^d$, we consider the problem of minimizing the sum of the k largest functions of the collection over Q. Specifically we focus on collections of linear functions and several classes of convex, piecewise linear functions which are defined by location models. We present simple linear programming formulations for these optimization models which give rise to linear time algorithms when the dimension d is fixed. Our results improve complexity bounds of several problems reported recently by Tamir [Discrete Appl. Math. 109 (2001) 293–307], Tokuyama [Proc. 33rd Annual ACM Symp. on Theory of Computing, 2001, pp. 75–84] and Kalcsics, Nickel, Puerto and Tamir [Oper. Res. Lett. 31 (1984) 114–127].

Keywords: Algorithms; Computational geometry; k-largest; Location; Rectilinear; k-centrum

1. Introduction

Given a collection $\{g_i(x)\}_{i=1}^n$ of n functions defined on $\mathbb{R}^d$, and a polyhedral set $Q \subset \mathbb{R}^d$, in this paper we consider the problem of minimizing the sum of the k largest functions of the collection over Q. Specifically we focus on collections of linear functions and several classes of convex, piecewise linear functions which are defined by location models. We present simple linear programming formulations for these optimization models which give rise to linear time algorithms when the dimension d is fixed. Tokuyama [22] has recently discussed the case where all the functions in the collection are linear, and Q is the intersection of p half spaces. Tokuyama presents a very complex (randomized) algorithm which finds the optimal solution in $O(p + n \log n)$ time when d is fixed. We show how to use our formulation to improve this bound. We obtain a deterministic $O(p + n)$ algorithm. In the location models that we consider, each function $g_i(x)$ represents a (weighted) distance of x from a given point $v^i \in \mathbb{R}^d$. The distance function is defined either by the l_1 or l_∞ norms. These problems are called the k-centrum location models. Again, when d is fixed our linear programming formulations lead to $O(p + n)$ algorithms which find optimal solutions.

In Section 2 we formally define the optimization model and present the linear programming formulation, leading to linear time algorithms. In the last section we discuss extensions to more general problems.

* Corresponding author.
E-mail addresses: ogryczak@ia.pw.edu.pl (W. Ogryczak), atamir@post.tau.ac.il (A. Tamir).

0020-0190/02/$ – see front matter
PII: S0020-0190(02)00370-8

2. A linear programming formulation

To facilitate the discussion we first introduce some notation. For any real number z define $(z)_+ = \max(z, 0)$. Let $y = (y_1, \dots, y_m)$ be a vector in $\mathbb{R}^m$. Define $\theta(y) = (\theta_1(y), \theta_2(y), \dots, \theta_m(y))$ to be the vector in $\mathbb{R}^m$, obtained by sorting the m components of y in nonincreasing order, i.e., $\theta_1(y) \geqslant \theta_2(y) \geqslant \dots \geqslant \theta_m(y)$. $\theta_i(y)$ will be referred to as the ith largest component of y. Finally, for $k = 1, \dots, m$, define $\Theta_k(y) = \sum_{i=1}^k \theta_i(y)$, the sum of the k largest components of y.

Lemma 1. *For any vector $y \in \mathbb{R}^m$ and $k = 1, \dots, m$,*

$$\Theta_k(y) = \frac{1}{m}\left(k\sum_{i=1}^m y_i + \min_{t\in\mathbb{R}} \sum_{i=1}^m \bigl[k(t-y_i)_+ + (m-k)(y_i-t)_+\bigr]\right).$$

Moreover, $t^* = \theta_k(y)$ is an optimizer of the above minimization problem.

Proof. Define $h(t) = \sum_{i=1}^m [k(t-y_i)_+ + (m-k)(y_i-t)_+]$. This function is piecewise linear and convex. It is easy to verify that the one sided derivatives at $t^* = \theta_k(y)$ are of opposite signs, and therefore t^* is a minimum point of $h(t)$. Substituting $t^* = \theta_k(y)$ in the expression for $h(t)$ we obtain

$$\begin{aligned} h\bigl(\theta_k(y)\bigr) &= k \sum_{i=k+1}^m \bigl(\theta_k(y) - \theta_i(y)\bigr) \\ &\quad - (m-k)\sum_{i=1}^k \bigl(\theta_k(y) - \theta_i(y)\bigr) \\ &= m\sum_{i=1}^k \theta_i(y) - k\sum_{i=1}^m \theta_i(y). \end{aligned}$$

Hence, $h(\theta_k(y)) = m\Theta_k(y) - k\sum_{i=1}^m y_i$, and therefore

$$\Theta_k(y) = \frac{1}{m}\left(k\sum_{i=1}^m y_i + \min_{t\in\mathbb{R}} h(t)\right). \qquad \square$$

It follows from the above lemma that $\Theta_k(y)$ can be represented as the solution value of the following linear program.

$$\Theta_k(y) = \min \frac{1}{m}\left(\sum_{i=1}^m \bigl[kd_i^- + (m-k)d_i^+ + ky_i\bigr]\right)$$

subject to

$$d_i^+ - d_i^- = y_i - t, \quad d_i^+, d_i^- \geqslant 0, \quad i = 1, \dots, m.$$

Substituting $d_i^- = d_i^+ - y_i + t$, we obtain

$$\Theta_k(y) = \min\left(kt + \sum_{i=1}^m d_i^+\right)$$

subject to

$$d_i^+ \geqslant y_i - t, \quad d_i^+ \geqslant 0, \quad i = 1, \dots, m.$$

Given the collection of functions, $\{g_i(x)\}_{i=1}^n$, and the polyhedral set $Q \subset \mathbb{R}^d$, defined in the introduction, let $g(x) = (g_1(x), \dots, g_n(x))$. The problem of minimizing $\Theta_k(g(x))$, the sum of the k largest functions of the collection over Q, can now be formulated as

$$\min\left(kt + \sum_{i=1}^n d_i^+\right)$$

subject to

$$d_i^+ \geqslant g_i(x) - t, \quad d_i^+ \geqslant 0, \quad i = 1, \dots, n,$$

$$x = (x_1, \dots, x_d) \in Q.$$

2.1. Minimizing the sum of the k largest linear functions

Consider the linear case. For $i = 1, \dots, n$, $g_i(x) = a^i x + b_i$, where $a^i = (a_1^i, \dots, a_d^i) \in \mathbb{R}^d$, and $b_i \in \mathbb{R}^1$. With the above notation the problem can be formulated as the following linear program:

$$\min\left(kt + \sum_{i=1}^n d_i^+\right)$$

subject to

$$d_i^+ \geqslant a^i x + b_i - t, \quad i = 1, \dots, n,$$

$$d_i^+ \geqslant 0, \quad i = 1, \dots, n,$$

$$x = (x_1, \dots, x_d) \in Q.$$

Note that this linear program has $n + d + 1$ variables, $d_1^+, \dots, d_n^+, x_1, \dots, x_d, t$, and $2n + p$ constraints. This formulation constitutes a special case of the class of linear programs defined as the duals of linear multiple-choice knapsack problems. Therefore, using the results in [24], when d is fixed, an optimal solution can be obtained in $O(p + n)$ time. (See also [13].) This bound improves upon the $O(p + n\log n)$ bound in [22] by a factor of $O(\log n)$.

2.2. Solving the rectilinear k-centrum location problem

For each pair of points $u = (u_1, \ldots, u_d)$, $v = (v_1, \ldots, v_d)$ in $\mathbb{R}^d$ let $d(u, v)$ denote the rectilinear distance between u and v,

$$d(u, v) = \sum_{j=1}^{d} |v_j - u_j|.$$

Given is a set $\{v^1, \ldots, v^n\}$ of n points in $\mathbb{R}^d$. Suppose that v^i, $i = 1, \ldots, n$, is associated with a nonnegative real weight w_i. For each point $x \in \mathbb{R}^d$ define the vector $D(x) \in \mathbb{R}^n$ by $D(x) = (w_1 d(x, v^1), \ldots, w_n d(x, v^n))$. For a given $k = 1, \ldots, n$, the single facility rectilinear k-centrum problem in $\mathbb{R}^d$ is to find a point $x \in \mathbb{R}^d$ minimizing the objective $f_k(x) = \sum_{i=1}^{k} \theta_i(D(x))$. (To the best of our knowledge the concept of a k-centrum was first defined by Slater [18] and Andreatta and Mason [1,2].) Note that the case $k = 1$ coincides with the classical (weighted) rectilinear 1-center problem in $\mathbb{R}^d$, while the case $k = n$ defines the classical (weighted) rectilinear 1-median problem, proposed by Hakimi [5,6]. It is well known that the last two problems can be formulated as linear programs. The center problem ($k = 1$) is formulated as,

$$\min t$$

subject to

$$t \geqslant w_i d(x, v^i), \quad i = 1, \ldots, n,$$

$$x = (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

To obtain a linear program we replace each one of the n nonlinear constraints $t \geqslant w_i d(x, v^i)$, by a set of 2^d linear constraints. For $i = 1, \ldots, n$, let $\Delta^i = (\delta^i_1, \ldots, \delta^i_d)$ be a vector all of whose components are equal to $+1$ or -1. Consider the set of 2^d linear constraints, $t \geqslant \sum_{j=1}^{d} \delta^i_j w_i (x_j - v^i_j)$, $\delta^i_j \in \{-1, 1\}$, $j = 1, \ldots, d$. This linear program has $d + 1$ variables, $t, x_1, \ldots, x_d$, and $2^d n$ constraints. Therefore, when d is fixed, it can be solved in $O(n)$ time by the algorithm of Megiddo [12].

Similarly, the median problem ($k = n$) is formulated as,

$$\min \sum_{i=1}^{n} z_i$$

subject to

$$z_i \geqslant \sum_{j=1}^{d} \delta^i_j w_i (x_j - v^i_j),$$

$$\delta^i_j \in \{-1, 1\}, \quad j = 1, \ldots, d, \ i = 1, \ldots, n,$$

$$x = (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

This linear program has $n + d$ variables, $z_1, \ldots, z_n$, $x_1, \ldots, x_d$, and $2^d n$ constraints. This formulation is also a special case of the class of linear programs discussed by Zemel [24]. Therefore, when d is fixed, an optimal solution can be obtained in $O(n)$ time. The above formulation of the median problem can be replaced by another, where the number of variables is $nd + d$ and the number of constraints is only $2dn$.

$$\min \sum_{i=1}^{n} \sum_{j=1}^{d} y_{i,j}$$

subject to

$$y_{i,j} \geqslant w_i (x_j - v^i_j), \quad j = 1, \ldots, d, \ i = 1, \ldots, n,$$

$$y_{i,j} \geqslant -w_i (x_j - v^i_j), \quad j = 1, \ldots, d, \ i = 1, \ldots, n,$$

$$x = (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

The latter compact formulation is also solvable in $O(n)$ time by the procedure of Zemel when d is fixed. In fact, it is easy to see from this formulation that the d-dimensional median problem is decomposable into d 1-dimensional problems. Therefore, it can be solved in $O(dn)$ time.

We note in passing that when d is fixed even the 1-centdian objective function, defined by Halpern [7–9] and Handler [10], as a convex combination of the center objective $f_1(x)$ and the median objective $f_n(x)$ can be solved in linear time (see [20]).

To the best of our knowledge for a general value of k, no linear time algorithms are reported in the literature even for $d = 1$. In [19] the case $d = 1$ is treated as a special case of a tree network. In particular, this one dimensional problem is solved in $O(n)$ time when k is fixed, and in $O(n \log n)$ time when k is variable. Subquadratic algorithms for any fixed d and variable k are given in [11]. For example, for $d = 2$ the algorithm there has $O(n \log^2 n)$ complexity.

Using the above results we can now formulate the rectilinear k-centrum problem in $\mathbb{R}^d$ as the following optimization problem:

$$\min\left(kt + \sum_{i=1}^{n} d_i^+\right)$$

subject to

$$d_i^+ \geqslant w_i d(x, v^i) - t, \quad d_i^+ \geqslant 0,\ i = 1, \ldots, n,$$

$$x = (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

As above, to obtain a linear program we replace each one of the n nonlinear constraints $d_i^+ \geqslant w_i d(x, v^i) - t$, by a set of 2^d linear constraints. The rectilinear k-centrum problem is now formulated as the linear programming problem,

$$\min\left(kt + \sum_{i=1}^{n} d_i^+\right)$$

subject to

$$d_i^+ + t \geqslant \sum_{j=1}^{d} \delta_j^i w_i \left(x_j - v_j^i\right),$$

$$\delta_j^i \in \{-1, 1\},\ j = 1, \ldots, d,\ i = 1, \ldots, n,$$

$$d_i^+ \geqslant 0, \quad i = 1, \ldots, n.$$

Note that the linear program has $n + d + 1$ variables, $d_1^+, \ldots, d_n^+, x_1, \ldots, x_d, t$, and $2^d n + n$ constraints. This formulation is again a special case of the class of linear programs defined as the duals of linear multiple-choice knapsack problems. Therefore, using the results of Zemel [24], when d is fixed, an optimal solution can be obtained in $O(n)$ time.

We note in passing that the results for the rectilinear problem can be extended to other polyhedral norms. For example, if we use the l_∞ norm, and let the distance between $u, v \in \mathbb{R}^d$ be defined by $d(u, v) = \max_{j=1,\ldots,d} |v_j - u_j|$, we get the following formulation for the respective k-centrum problem:

$$\min\left(kt + \sum_{i=1}^{n} d_i^+\right)$$

subject to

$$d_i^+ + t \geqslant w_i\left(x_j - v_j^i\right), \quad j = 1, \ldots, d,\ i = 1, \ldots, n,$$

$$d_i^+ + t \geqslant -w_i\left(x_j - v_j^i\right),$$

$$j = 1, \ldots, d,\ i = 1, \ldots, n.$$

3. Related problems and extensions

Recently, a new type of objective function in location modeling, called ordered median function, has been introduced and analyzed. See, for example, [15,17,14,4]. (This criterion was introduced already in [23] in the context of multi-criteria decision making.) In our context this objective function generalizes the k-centrum objective. We extend the formulation in Section 2 to the rectilinear ordered median problem. To define this general model we first need to introduce some notation. Given is a nonnegative vector $\lambda = (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$, satisfying $\lambda_1 \geqslant \cdots \geqslant \lambda_m$. For convenience define $\lambda_{m+1} = 0$. For each $y = (y_1, \ldots, y_m)$ define

$$\Lambda(y) = \sum_{i=1}^{m} \lambda_i \theta_i(y).$$

Although $\Lambda(y)$ may also be considered for arbitrary sequences of $(\lambda_1, \ldots, \lambda_m)$ [23], the monotonicity assumption $\lambda_1 \geqslant \cdots \geqslant \lambda_m$ is important whitin the location analysis context since it guarantees the so-called equitable properties of solutions [16]. (Note that if λ is the vector whose first k components are equal to 1 and the others are equal to 0, then $\Lambda(y) = \Theta_k(y)$.)

For each vector λ satisfying the above we have the following expression,

$$\Lambda(y) = \sum_{k=1}^{m} (\lambda_k - \lambda_{k+1}) \Theta_k(y).$$

Using the results in Section 2, due to the monotonicity assumption, $\lambda_1 \geqslant \cdots \geqslant \lambda_m \geqslant 0$, we can represent $\Lambda(y)$ as the solution value of the following linear programming:

$$\Lambda(y) = \min \sum_{k=1}^{m} (\lambda_k - \lambda_{k+1}) \left(kt_k + \sum_{i=1}^{m} d_{i,k}^+\right)$$

subject to

$$d_{i,k}^+ \geqslant y_i - t_k, \quad d_{i,k}^+ \geqslant 0,\ i = 1, \ldots, m,\ k = 1, \ldots, m.$$

The latter formulation can be used to generalize the results in Sections 2.1 and 2.2. For example, consider a generalization of the rectilinear k-centrum problem, called the *rectilinear ordered median problem*, defined as follows:

Find a point $x \in \mathbb{R}^d$ minimizing the objective function $\Lambda(D(x))$. (To use the above model set $m = n$.)

This problem can now be formulated as a linear program.

$$\min \sum_{k=1}^{n}(\lambda_k - \lambda_{k+1})\left(kt_k + \sum_{i=1}^{n} d_{i,k}^+\right)$$

subject to

$$d_{i,k}^+ + t_k \geqslant \sum_{j=1}^{d} \delta_j^i w_i\left(x_j - v_j^i\right),$$

$$\delta_j^i \in \{-1, 1\}, \quad j = 1, \ldots, d,\ i = 1, \ldots, n,\ k = 1, \ldots, n,$$

$$d_{i,k}^+ \geqslant 0, \quad i = 1, \ldots, n,\ k = 1, \ldots, n.$$

Note that this linear program has $n^2 + d + n$ variables, $\{d_{i,k}^+\}$, $i, k = 1, \ldots, n$, $x_1, \ldots, x_d$, $t_1, \ldots, t_n$, and $2^d n^2 + n^2$ constraints.

The best known algorithm to solve the above rectilinear ordered median in a fixed dimension d is the $O(n \log^{2d} n)$ procedure described in [11]. This procedure is actually a direct application of the general algorithm of Cohen and Megiddo [3]. At this point in time we do not yet know whether the above linear programming formulation, which involves a quadratic number of variables, can be used to improve the bounds reported in [11]. Nevertheless, for some special cases of the vector $\lambda = (\lambda_1, \ldots, \lambda_n)$, the above problem can still be solved in linear time. Specifically, if the components of λ can take on a constant, say q, number of distinct values, the above problem reduces to a linear program with $qn + d + n$ variables and $2^d qn + qn$ constraints. Therefore, it can be solved in $O(n)$ time when both, d and q are fixed. As an example, consider the centdian problem mentioned in Section 2.2. In this case $\lambda = (1, \mu, \ldots, \mu)$, where μ is a positive number bounded above by 1.

Finally, we address the solvability of the rectilinear ordered median problem when the dimension d is variable. (Note that the formulation given above has $2^d n^2 + n^2$ constraints.) Consider the alternative formulation:

$$\min \sum_{k=1}^{n}(\lambda_k - \lambda_{k+1})\left(kt_k + \sum_{i=1}^{n} d_{i,k}^+\right)$$

subject to

$$d_{i,k}^+ + t_k \geqslant \sum_{j=1}^{d} y_{i,j}, \quad i = 1, \ldots, n,\ k = 1, \ldots, n,$$

$$y_{i,j} \geqslant w_i\left(x_j - v_j^i\right), \quad j = 1, \ldots, d,\ i = 1, \ldots, n,$$

$$y_{i,j} \geqslant -w_i\left(x_j - v_j^i\right), \quad j = 1, \ldots, d,\ i = 1, \ldots, n,$$

$$d_{i,k}^+ \geqslant 0, \quad i = 1, \ldots, n,\ k = 1, \ldots, n.$$

This linear program has $n^2 + (n+1)d + n$ variables and $2n^2 + 2nd$ constraints. Therefore, the problem is polynomially solvable. Moreover, if $w_i = w$, $i = 1, \ldots, n$, for some positive constant w, it follows from [21] that the problem can be solved in strongly polynomial time.

References

[1] G. Andreatta, F.M. Mason, k-eccentricity and absolute k-centrum of a tree, European J. Oper. Res. 19 (1985) 114–117.

[2] G. Andreatta, F.M. Mason, Properties of the k-centrum in a network, Networks 15 (1985) 21–25.

[3] E. Cohen, N. Megiddo, Maximizing concave functions in fixed dimension, in: P.M. Pardalos (Ed.), Complexity in Numerical Optimization, World Scientific, Singapore, 1993, pp. 74–87.

[4] R.L. Francis, T.J. Lowe, A. Tamir, Aggregation error bounds for a class of location models, Oper. Res. 48 (2000) 294–307.

[5] S.L. Hakimi, Optimum locations of switching centers and the absolute centers and medians of a graph, Oper. Res. 12 (1964) 450–469.

[6] S.L. Hakimi, Optimum distribution of switching centers in a communication network and some related graph-theoretic problems, Oper. Res. 13 (1965) 462–475.

[7] J. Halpern, The location of a cent-dian convex combination on an undirected tree, J. Regional Sci. 16 (1976) 237–245.

[8] J. Halpern, Finding minimal center-median convex combination (cent-dian) of a graph, Management Sci. 24 (1978) 535–544.

[9] J. Halpern, Duality in the cent-dian of a graph, Oper. Res. 28 (1980) 722–735.

[10] G.Y. Handler, Medi-centers of a tree, Transportation Sci. 19 (1985) 246–260.

[11] J. Kalcsics, S. Nickel, J. Puerto, A. Tamir, Algorithmic results for ordered median problems, Oper. Res. Lett. 30 (2002) 149–158.

[12] N. Megiddo, Linear programming in linear time when the dimension is fixed, J. ACM 31 (1984) 114–127.

[13] N. Megiddo, A. Tamir, Linear time algorithms for some separable quadratic programming problems, Oper. Res. Lett. 13 (1993) 203–211.

[14] S. Nickel, J. Puerto, A unified approach to network location problems, Networks 34 (1999) 283–290.
[15] W. Ogryczak, On the lexicographic minimax approach to location-allocation problems, Discussion Paper IS-MG 94/22, Institute de Statistique, Mathematiques de la Gestion, Universite Libre de Bruxelles, 1994.
[16] W. Ogryczak, Inequality measures and equitable approaches to location problems, European J. Oper. Res. 122 (2000) 374–391.
[17] J. Puerto, F.R. Fernandez, The symmetrical single facility location problem, Technical Report, Prepublicacion de la Facultad de Matematicas, Universidad de Sevilla, 1995.
[18] P.J. Slater, Centers to centroids in a graph, J. Graph Theory 2 (1978) 209–222.
[19] A. Tamir, The k-centrum multi-facility location problem, Discrete Appl. Math. 109 (2001) 293–307.
[20] A. Tamir, D. Perez-Brito, J.A. Moreno-Perez, A polynomial algorithm for the p-centdian problem on a tree, Networks 32 (1998) 255–262.
[21] E. Tardos, A strongly polynomial algorithm to solve combinatorial linear programs, Oper. Res. 34 (1986) 250–256.
[22] T. Tokuyama, Minimax parametric optimization problems in multi-dimensional parametric searching, in: Proc. 33rd Annual ACM Symposium on Theory of Computing, 2001, pp. 75–84.
[23] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decision making, IEEE Trans. Systems, Man Cybernet. 18 (1988) 183–190.
[24] E. Zemel, An $O(n)$ algorithm for the linear multiple choice knapsack problem and related problems, Inform. Process. Lett. 18 (1984) 123–128.