Exploiting Self-Canceling Demand Point Aggregation Error for Some Planar Rectilinear Median Location Problems

R.L. Francis,[1] T.J. Lowe,[2] M.B. Rayco,[3] A. Tamir[4]

[1] *Department of Industrial and Systems Engineering, University of Florida, Gainesville, Florida 32611*

[2] *Tippie College of Business, University of Iowa, Iowa City, Iowa 52242*

[3] *Department of Management Science, University of Miami, Coral Gables, Florida 33124*

[4] *Department of Statistics & Operations Research, School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel*

Received 18 July 2000; revised 22 January 2003; accepted 22 January 2003

DOI 10.1002/nav.10079

Abstract: When solving location problems in practice it is quite common to aggregate demand points into centroids. Solving a location problem with aggregated demand data is computationally easier, but the aggregation process introduces error. We develop theory and algorithms for certain types of centroid aggregations for rectilinear 1-median problems. The objective is to construct an aggregation that minimizes the maximum aggregation error. We focus on row-column aggregations, and make use of aggregation results for 1-median problems on the line to do aggregation for 1-median problems in the plane. The aggregations developed for the 1-median problem are then used to construct approximate n-median problems. We test the theory computationally on n-median problems ($n \geq 1$) using both randomly generated, as well as real, data. Every error measure we consider can be well approximated by some power function in the number of aggregate demand points. Each such function exhibits decreasing returns to scale.
© 20032003 Wiley Periodicals, Inc. Naval Research Logistics 50: 614–637, 2003.

1. INTRODUCTION

A location problem usually involves locating one or more facilities with respect to demand points, also called existing facilities. In urban modeling contexts, each private residence can be a demand point. Thus there can be millions of demand points to deal with. Demand point data may be readily available, available at some cost, or unavailable within the time and budget limitations imposed on solving the problem. Even if the data are readily available, it may be computationally impractical to make use of all of it.

Correspondence to: T.J. Lowe

Thus, it is a very common practice in location modeling, and other related geographic modeling areas, to aggregate demand points and solve the problem using the reduced data set. For example, if a postal code area (PCA) has 1000 distinct residences, we might suppose all 1000 residences are at the centroid of the PCA. Centroids (to be defined) are commonly used, for example, with geographic information systems and CD-ROM phone books. Some Bureau of the Census data is organized by centroids. Unless we state otherwise, centroid aggregations are the only ones we shall consider in this paper. The tax office location problem in a metropolitan area setting considered by Domich et al. [6] is a good example of using centroids for demand points when solving median sorts of problems. Another good example is the branch bank location problem discussed by Chelst, Schultz, and Sanghvi [4].

Aggregation often results in lower costs to obtain the demand point data. Certainly solving the smaller aggregated location problem is easier than solving the original problem. However, aggregation also introduces error into the model, due to inaccurate distance measures. Thus, there is a tradeoff to consider. A sensible strategy is to try to aggregate in such a way as to minimize the error, or to put some upper bound on the error while minimizing cost. Alternatively, the problem of minimizing the error can be viewed as a resource allocation problem; allocating aggregate demand points by choosing their number and placements. We shall adopt this latter point of view.

There is little agreement on how best to measure aggregation error, and numerous error criteria are available (Francis et al. [11]). To introduce the error criterion we advocate, let X and P denote collections of new facility locations, and demand points respectively. We let $f(X)$ be the cost function for the underlying location model, the cost if we choose X, given P. Let P' be the list of aggregate demand points, with P'_i the aggregate demand point replacing P_i. Let $g(X)$ denote the approximating cost function resulting from using P' instead of P in the original model. Thus $e(X) \equiv |f(X) - g(X)|$ is the (absolute) *error for* X, and $\max\{e(X):X\}$ is *the maximum (absolute) error over all values of* X. Geoffrion [16] gives theoretical arguments for using the maximum absolute error as an error measurement. He also points out how having an error bound (an upper bound on the maximum absolute error) can provide bounds on differences of values of optimal solutions to the true and approximating problems. This absolute error measurement is well accepted in the field of numerical analysis (Francis, Lowe, and Rayco [10]).

The purpose of this paper is to study, for median problems, the error associated with using centroids (defined below) as aggregate demand points. We develop theory for the 1-median problem with rectilinear distances. In particular we study the properties of an *aligned row-column (ARC) algorithm* that minimizes, over all aligned row–column aggregations, the maximal error for the 1-median problem with rectilinear distances. We then develop an algorithm, denoted as CRC, which uses the partitions defined by ARC as aggregate demand sets, but uses centroids of the sets as aggregation points. We test CRC on n-median problems with rectilinear distances. Our computational experience for some n-median problems, reported in Section 5, is encouraging.

2. THE AGGREGATION PROBLEM

2.1. Problem Formulation: Centroids and the 1-Median Problem

We now introduce the specific median problem we analyze. Let $I = \{1, \ldots, m\}$ denote the set of demand point indices, and $P = \{P_i: i \in I\}$ denote the set of (distinct) planar demand points. We let $X = (x_1, x_2)$ denote any new facility location, $P_i = (p_1^i, p_2^i)$, and denote by

$\|X - P_i\| = [|x_1 - p_1^i|^p + |x_2 - p_2^i|^p]^{1/p}$ the ℓ_p-distance between X and P_i, with $p \geq 1$. We let w_i denote a positive "weight" for each demand point i; typically w_i is proportional to the frequency of travel between the new facility and the demand point. Thus the planar 1-median problem is to find X^* to minimize

$$f(X) \equiv \sum \{w_i\|X - P_i\|: i \in I\}.$$

The n-median model is a generalization of the 1-median model. There are n new facilities ($n \geq 2$), and the travel distance between each demand point and the n new facilities is the distance between the demand point and a closest new facility.

Let $\{IP_1, \ldots, IP_q\}$ denote any partition of P into q nonempty (disjoint) subsets. For each $u = 1, \ldots, q$, let $W_u = \Sigma \{w_i: P_i \in IP_u\}$ denote the sum of weights of demand points in IP_u. The *centroid* C_u of IP_u is defined by $C_u \equiv \Sigma \{(w_i/W_u)P_i: P_i \in IP_u\}$. Also, define $f'(X) \equiv W_1\|X - C_1\| + \cdots + W_q\|X - C_q\|$. Note that f' is an approximating 1-median function defined using centroids as aggregate demand points. Thus f' is a *specific instance* of g. The absolute error function for f' is

$$e(X) \equiv |f(X) - f'(X)| \qquad \text{for all } X.$$

Francis and White [9] prove that $f'(X) \leq f(X)$ for all X, and all $p \geq 1$. We make extensive use of this result and restate it as:

CENTROID AGGREGATION LEMMA: For the 1-median model with the ℓ_p-distance, for all $p \geq 1$,

$$e(X) = f(X) - f'(X) \geq 0 \qquad \text{for all } X.$$

For 1-median models, we therefore can write $f(X) - f'(X) = e(X) = \Sigma \{e^i(X): i \in I\}$, with $e^i(X) = w_i\|X - P_i\| - w_i\|X - P_i'\|$ a "difference" error for demand point i. Each $e^i(X)$ can be negative or positive. The result is that there is often self-canceling error (negative values offset positive values). An alternative to our error measurement approach is worst-case error, as measured by an upper bound on the error (Francis, Lowe, and Tamir [12]). While it is applicable to many location models, this latter approach does not have the self-cancellation property.

Without loss of generality, we assume each IP_u has at least two points. Then it can be shown that when $p > 1$, the inequality in the lemma is strict except on the union of a finite collection of colinear line segments (possibly empty). Therefore the error is positive except on a set of measure/area zero. It is possible to use the lemma to obtain some insight into the effect of various degrees of *centroid aggregation* (abbreviated as CAG). One immediate observation is that a CAG of a CAG is itself a CAG of the original data. Suppose we have an original location model f, and use CAG to construct an approximating model $\hat{f}$. Next, we do aggregation of the aggregate demand points that define $\hat{f}$, resulting in another model f'. For example, let $C_u = \Sigma \{(w_i/W_u)P_i: P_i \in IP_u\}$ for $u = 1, 2$, and let $\hat{f}$ be the location model defined using C_1 and C_2. We could then compute $C' = (W_1/(W_1 + W_2)C_1 + (W_2/(W_1 + W_2)C_2 = \Sigma \{(w_i/(W_1 + W_2))P_i: P_i \in IP_1 \cup IP_2\}$. Doing this, we aggregate C_1 and C_2 into C' with weight $W_1 + W_2$ to obtain an approximating location model f'. From the lemma we have $f(X) \geq \hat{f}(X) \geq f'(X)$. The error for f' will therefore be no less than the error for $\hat{f}$. In the case when $p > 1$, the error for f' will be greater than the error for $\hat{f}$ almost everywhere. Thus further centroid aggregation causes more error.

The *relative error,* defined by $\text{rel}(X) = e(X)/f(X)$ for all X (perhaps expressed in percent), will also be of interest. We can view $\text{rel}(X)$ as a scaled version of $e(X)$.

Henceforth we focus on the case where $p = 1$, i.e., the rectilinear distance measure. Also, much of our theory is for the case with $n = 1$, i.e., the 1-median problem. Results for the 1-median problem lead to aggregation schemes for the rectilinear distance n-median problem.

2.2. Motivation for Focus of Study

A principle objective of our study is to obtain insights into aggregation schemes for solving the NP-Hard (Meggido and Supowit [23]) rectilinear-distance n-median problem by studying the 1-median problem. We note that although the 1-median problem is well solved *if* all of the data is available (Francis and White [9]), in many instances all data are *not* available. In fact, the demand data may be changing over time due to shifts in demographic data. In these situations, aggregation may be used to generate an approximation to the true underlying model. Thus, we seek qualitative insights into the problem that are not especially data dependent, and which apply to n-median problems with $n \geq 2$. We study in detail a particular type of aggregation scheme (row–column). This scheme partitions demand points into cells made up of rows and columns (with edges parallel to the x_1- and x_2-axes) of varying spacings, and then aggregates all original data points in each cell into the centroid of the cell. Thus, our approach is a heuristic approach for solving the underlying aggregation problem. At the end of subsection 2.3, we list some of the insights gained from our work. This list also provides an overview of our paper.

2.3. Related Literature and Insights from Results

Rogers et al. [32] give a review of basic aggregation ideas in optimization models. Early recognition of various errors created by demand point aggregation appears to have begun in the geography literature, with papers by Hillsman and Rhoda [19] and by Casillas [3]. This literature has been discussed in some detail by Francis et al. [11]. Plastria [28, 29] has studied centroid aggregation error for the planar 1-median problem. An important finding of his is an asymptotic result; for $q = 1$ centroid, $e(X)$ goes to zero as the distance between X and the centroid increases while X varies along a half-line with its end point at the centroid. For other work on demand point aggregation for various n-median models, see Erkut and Bozkaya [7], Murray and Gottsegen [26], and Zhao and Batta [34].

Francis and Lowe [8] showed how to compute error bounds on the maximum absolute error for both n-median and n-center problems. Their work was preceded to some extent by the work of Zemel [33], although the bounds he computed were not for purposes of aggregation. The error bound results in these two papers have recently been substantially generalized by Francis, Lowe, and Tamir [12]. They develop a methodology to compute aggregation error bounds for an entire class of location models, including many of the best-known models. Their work strongly suggests the need to exploit model structure to obtain aggregations with small error bounds.

The work most closely related to ours is by Francis, Lowe, and Rayco [10], abbreviated hereafter as FLR. For rectilinear distance n-median problems, they develop a means of minimizing the error bound of Francis and Lowe [8] over a class of row–column aggregations. They use medians, instead of centroids, as aggregate demand points. Also, we minimize the maximum error instead of an error bound. Therefore, we use a different objective and approach for doing aggregation. The FLR approach was specifically designed for n-median problems, while our new approach is derived from analysis of the 1-median problem. However, our new approach worked uniformly better on all problems we tried, including several n-median problems, where $n > 1$. For example, it gave a maximum relative error of less that 1%, using 400 aggregate demand points for a problem with 70,000 actual demand points.

A row–column aggregation, when rotated 45° with respect to the axes, provides an effective means of doing aggregation for rectilinear distance n-center problems. Various aspects of these ideas have been studied by Francis and Rayco [13], Rayco, Francis, and Lowe [30], and Rayco, Francis, and Tamir [31]. Again, an upper bound (often tight) on the maximum error is being minimized instead of the maximum error. Centers, instead of medians, are used as aggregate demand points. Andersson et al. [1] have experimented with how to adapt the demand point aggregations obtained through these various row–column approaches to the case where demand points all lie on a network.

Much of the work mentioned above has indicated, mostly based on computer experimentation, that the maximum error decreases at a decreasing rate as q increases. We believe that this decreasing returns to scale phenomenon has important implications for aggregation done in practice. Choosing a small number of aggregate demand points can cause a large error. Choosing a larger number can cause a small error, while an even larger choice will not decrease the error appreciably.

We conclude this section with a list of insights supported by what follows in our paper.

- Centroid aggregation always causes underestimation of the median cost function.
- For the 1-median problem on the line, and where all weights are identical and centroid aggregation is used, a contiguity property holds. That is, there is an aggregation that minimizes the maximal error such that the demand points aggregated into each centroid are contiguous. In addition, the maximal error (and maximal relative error) is always attained at a centroid.
- For planar problems, doing aggregation on each axis using "projected" planar demand point data leads to a good way (the row–column method) of doing aggregation.
- For the rectilinear 1-median model, the maximum error, as well as the maximum relative error, always occurs at a centroid or at a point with the property that each of its coordinates is a coordinate of some centroid (a total of q^2 points).
- Doing aggregation well for the 1-median problem helps with the n-median problem.
- The maximum error decreases at a decreasing rate as q increases and this error is proportional to $1/q$. All our computational experience, and some theoretical analysis, indicates that the maximum error (as well as other related error measures) can be represented quite well as a power function of q, of the form a/q^b, with $b \geq 1$.
- The maximum relative error does not seem to depend significantly upon the size of the region containing the demand points, but only on q, the number of centroids.

3. CENTROID AGGREGATION ERROR ANALYSIS

3.1. Row–Column Aggregation

We now motivate our row–column approach. For the rectilinear 1-median model $f(X)$, we would like to find a centroid aggregation approximation function $f'(X)$ such that the maximum error, $\max\{e(X)\colon X\}$, is minimized (where $e(X) = |f(X) - f'(X)|$). Note that each approximation function is defined by some partition of the demand set into q (nonempty) disjoint subsets. Even for $q = 2$, the total number of partitions is exponential in m, and therefore exhaustive enumeration is not tractable. Generally, little is known about the structure of an optimal partition. One might conjecture that partitioning the demand set into rectangular blocks is optimal, but this is not the case even for the 1-dimensional situation, i.e., when all the demand points are collinear. We conjecture that finding an optimal partition is NP-hard even in the 1-dimensional case. (We later show that the 1-dimensional case is polynomially solvable if all demand points are equally weighted.) We thus suggest concentrating only on some classes of partitions, and seeking a best partition in these classes.

Several classes of partitions have been considered in earlier studies of general partitioning problems of 2-dimensional arrays (grid graphs). The most common class used for partitioning grid graphs is the class of partitions defined by horizontal and vertical cuts only. A more general class is the class defined by allowing guillotine cuts. This latter class is then a subclass of all partitions of the 2-dimensional array into rectangles with the same orientation as the given array. (See Conti, Malucelli, Nicoloso, and Simeone [5] and Khanna, Muthukrishnan, and Skiena [20]). Unfortunately, as illustrated in the above references, several 2-dimensional (discrete) partitioning problems, defined by simple objective functions, are NP-hard, or at least very difficult to solve, even in the class defined by horizontal and vertical cuts.

The objective function we focus on, minimizing the maximum error, seems to be more involved and complicated than the criteria used in the literature cited above. While our problem appears to be a continuous problem, we prove (see Appendix, Theorem 11) that it can be reduced to a discrete problem. The maximum error is the maximum of the errors evaluated at the set of q^2 grid points defined by using vertical and horizontal cuts through each one of the centroids of the q subsets of the partition. Because of this reduction to a discrete problem, our problem seems closely related to the ones discussed in the previous paragraph.

Since we believe that our problem is computationally difficult even when we limit the search to horizontal and vertical cuts, we consider an aggregation heuristic, called the aligned row–column aggregation method. It is based on projecting the demand point data onto the two axes, and solving the two 1-dimensional problems optimally. The vertical cuts of the solution the heuristic provides are determined by the optimal solution of this problem projected on the x_1-axis, while the horizontal cuts are determined by the problem on the x_2-axis.

For the rectilinear-distance 1-median problem, called the *rectilinear problem* for short, we observe in this subsection that for certain types of aggregations, the aggregation error is separable into x_1 and x_2 error components. This separability allows aggregation on each axis using "projected" data from the planar problem. By taking the cross-products of these aggregations on the axes we can construct a planar aggregation, referred to for short as an ARC (aligned row–column) aggregation.

To motivate this row–column approach, imagine a plot of all the demand points with a grid superimposed. The grid has n_2 rows and n_1 columns [sometimes written as $n(2)$ and $n(1)$, respectively]. Spacings of rows, and of columns, need not be the same. Given a collection of demand points and weights, P_i, w_i, $i = 1, \ldots, m$, an *aligned row column* (ARC) *aggregation* with n_2 rows and n_1 columns is defined as follows.

1. Choose positive integers n_1 and n_2, numbers of columns and rows respectively.
2. Construct a smallest box B in the plane, with sides parallel to the axes, containing all demand points. Let $v1_0$ and $v1_{n(1)}$ ($h1_0$ and $h1_{n(2)}$) be smallest and largest x_1 coordinates (x_2 coordinates) respectively in B: $B = \{(x_1, x_2)\colon v1_0 \leq x_1 \leq v1_{n(1)},\ h1_0 \leq x_2 \leq h1_{n(2)}\}$.

Table 1. Separation of original and approximating location models, and error, into independent x_1 and x_2 parts.

Notation	Notation meaning	Name
$f^1(x_1)$	$\Sigma\ \{w_i\|x_1 - p_1^i\|: i \in I\}$	Original model: x_1 part
$f^2(x_2)$	$\Sigma\ \{w_i\|x_2 - p_2^i\|: i \in I\}$	Original model: x_2 part
$f(x_1, x_2)$	$f^1(x_1) + f^2(x_2)$	Original model
$g^1(x_1)$	$\Sigma\ \{W_t^1\|x_1 - c_1^t\|: t = 1, \ldots, n_1\}$	Approximating model: x_1 part
$g^2(x_2)$	$\Sigma\ \{W_s^2\|x_2 - c_2^s\|: s = 1, \ldots, n_2\}$	Approximating model: x_2 part
$g(x_1, x_2)$	$\Sigma\ \{W_{st}(\|x_1 - c_1^t\| + \|x_2 - c_2^s\|)$: $s = 1, \ldots, n_2, t = 1, \ldots, n_1\}$	Approximating model
$e^1(x_1)$	$f^1(x_1) - g^1(x_1)$	Error for x_1
$e^2(x_2)$	$f^2(x_2) - g^2(x_2)$	Error for x_2
$e(x_1, x_2)$	$e^1(x_1) + e^2(x_2)$	Total error

3. Construct any $n_1 - 1$ vertical lines (any $n_2 - 1$ horizontal lines) intersecting B from bottom to top (left to right) to partition B into n_1 columns (n_2 rows). Denote the x_1 coordinates of the vertical lines by $v1_1 < \cdots < v1_{n(1)-1}$ (x_2 coordinates of the horizontal lines by $h1_1 < \cdots < hy_{n(2)-1}$), respectively.
4. Define $\text{Col}_t = \{(x_1, x_2): (x_1, x_2) \in B,\ v1_{t-1} \le x_1 < v1_t\}$, $t = 1, \ldots, n_1 - 1$; $\text{Col}_{n(1)} = \{(x_1, x_2): (x_1, x_2) \in B,\ v1_{n(1)-1} \le x_1 \le v1_{n(1)}\}$.
5. Define $\text{Row}_s = \{(x_1, x_2): (x_1, x_2) \in B,\ h1_{s-1} \le x_2 < h1_s\}$, $s = 1, \ldots, n_2 - 1$; $\text{Row}_{n(2)} = \{(x_1, x_2): (x_1, x_2) \in B,\ h1_{n(2)-1} \le x_2 \le h1_{n(2)}\}$.
6. For each Col_t (each Row_s) containing demand points, define W_t^1 (W_s^2) to be the total weight of all demand points in Col_t (in Row_s).
7. For each Col_t (Row_s) containing demand points, define the centroid c_1^t of Col_t by $c_1^t = \Sigma\ \{w_i p_1^i / W_t^1: (p_1^i, p_2^i) \in \text{Col}_t\}$ (centroid c_2^s of Row_s by $c_2^s = \Sigma\ \{w_i p_2^i / W_s^2: (p_1^i, p_2^i) \in \text{Row}_s\}$).
8. For each Row_s and Col_t whose intersection contains demand points, define W_{st} to be the total weight of all demand points in both Row_s and Col_t.
9. For each s and t with $W_{st} > 0$, aggregate all demand points in Row_s and Col_t into (c_1^t, c_2^s), so that (c_1^t, c_2^s) has a total weight of W_{st}.
10. Define the approximating function g by $g(x_1, x_2) = \Sigma\ \{W_{st}(|x_1 - c_1^t| + |x_2 - c_2^s|): s = 1, \ldots, n_2, t = 1, \ldots, n_1\}$.

An ARC aggregation is *aligned* in the sense that all the aggregation points in a given row (column) have the same x_2 coordinate (x_1 coordinate). Note that an ARC aggregation is not necessarily a centroid aggregation since (c_1^t, c_2^s) may not be the centroid of the demand points in cell s, t. Another way to think of an ARC aggregation is that each cell s, t represents a city block. Line segments separating adjoining columns (rows) can be thought of as streets parallel to the x_2-axis (x_1-axis), as can the edges of the smallest enclosing box, B.

Table 1 introduces some needed notation and terminology. With reference to Table 1, for example, note that the original model, $f(x_1, x_2)$, is the sum of $f^1(x_1)$ and $f^2(x_2)$. We shall do aggregation on the line to replace the functions f^1 and f^2 by the approximating functions g^1 and g^2, respectively.

ARC LEMMA: Given any ARC aggregation, $e(x_1, x_2) = e^1(x_1) + e^2(x_2) \ge 0$ for all (x_1, x_2).

PROOF: We have

$$g(x_1, x_2) = \sum \{W_{st}(|x_1 - c_1^t| + |x_2 - c_2^s|): s = 1, \ldots, n_2, t = 1, \ldots, n_1\}$$

$$= \sum \{W_{st}|x_1 - c_1^t|: s = 1, \ldots, n_2, t = 1, \ldots, n_1\}$$

$$+ \sum \{W_{st}|x_2 - c_2^s|: s = 1, \ldots, n_2, t = 1, \ldots, n_1\}$$

$$= \sum \{W_t^1|x_1 - c_1^t|: t = 1, \ldots, n_1\} + \sum \{W_s^2|x_2 - c_2^s|: s = 1, \ldots, n_2\}$$

$$= g^1(x_1) + g^2(x_2).$$

But then $e(x_1, x_2) = f(x_1, x_2) - g(x_1, x_2) = f^1(x_1) + f^2(x_2) - g^1(x_1) - g^2(x_2) = e^1(x_1) + e^2(x_2)$. That $e(x_1, x_2) \geq 0$ for all (x_1, x_2) is a consequence of the Centroid Aggregation Lemma applied independently to $e^1(x_1)$ and $e^2(x_2)$. (This lemma applies since, for example, c_1^t is the centroid of all demand points in column t.)

Because of the ARC Lemma, for fixed n_2 and n_1 an ARC aggregation that minimizes maximum error over all possible ARC aggregations can be found by separately minimizing $\max\{e^1(x_1): x_1\}$ (with n_1 centroids) and $\max\{e^2(x_2): x_2\}$ (with n_2 centroids). The next subsection exploits the error separability stated in the lemma. The subsection is of theoretical interest in itself, and also provides the basis for our planar aggregation approach.

3.2. Centroid Aggregation: One-Median Problem on the Line

In this section we develop a simple way to compute the maximum error, and characterize where error is positive and zero. Whenever the error is zero, it is due to the self-cancellation effect. We obtain an easily computed upper bound on the maximum error. Assuming equally weighted demand points, we give a contiguity property for an optimal (minimizes maximum error) aggregation on the line. We obtain an upper bound on the relative error for the location problem on the line, and show it goes to zero quickly as the number of aggregate demand points increases. Finally, we give two algorithms of low computational order for doing contiguous DP aggregation on the line. One algorithm uses bisection search; the other uses dynamic programming. The former is simpler to implement, but the latter has a lower order. See the Appendix for the details of the latter.

3.2.1. Notation, General Partitions

For ease of exposition, we establish some notation unique to this section. All the results apply with obvious modifications to the functions defined in Table 1. We use the function f to represent either f^1 or f^2, and the function e to represent either e^1 or e^2. We also assume that there are given n demand points (DPs), $v_1 < \cdots < v_n$, on the real line, with positive (demand) weights, $w_1, \ldots, w_n$, respectively. Any of these weights may be the sum of several weights of the (projected) original location model. Also DP coordinates have been renamed and put into strictly increasing order.

The objective of the 1-median problem on the line is to find a point x on the line, minimizing $f(x) = \Sigma \{w_j|x - v_j|: j = 1, \ldots, n\}$. It is well-known that an optimal solution coincides with a weighted median of the DPs, and it can be found in $O(n)$ time. In demand point aggregation we partition the above set of DPs into p, $p \le n$, nonempty subsets, $V_1, V_2, \ldots, V_p$. We compute $W_u \equiv \Sigma \{w_j: v_j \in V_u\}$ for each subset V_u, and aggregate all the DPs of V_u, $u = 1, \ldots, p$, to c_u, the centroid of V_u, $c_u \equiv \Sigma \{(w_j/W_u)v_j: v_j \in V_u\}$. For each subset V_u of $V = \{v_1, \ldots, v_n\}$, let CH_u denote the convex hull of V_u, and let L_u be the length of the interval defining CH_u.

We approximate the original problem by a new 1-median problem. For each point x on the line, we let $f'(x)$ denote the sum of weighted distances of the above p centroids, $c_1, \ldots, c_p$, from x:

$$f'(x) \equiv \sum \{W_u|x - c_u|: u = 1, \ldots, p\}.$$

For each x, we define the (centroid) *aggregation error* at x, $e(x) \equiv f(x) - f'(x)$. We call the ratio $e(x)/f(x)$ the (centroid) *relative error* at x.

For each $u = 1, \ldots, p$, define $e_u(x)$ to be the error with respect to the demand points in V_u, i.e.,

$$e_u(x) = \sum \{w_j|v_j - x|: v_j \in V_u\} - \sum \{w_j|x - c_u|: v_j \in V_u\}$$

With the above notation we have $e(x) = \Sigma \{e_u(x): u = 1, \ldots, p\}$.

THEOREM 1: For $u = 1, \ldots, p$,

a. $e_u(x)$ is positive for any interior point of CH_u, and is zero otherwise.
b. $e_u(x) \le e_u(c_u) = \Sigma \{w_j|v_j - c_u|: v_j \in V_u\} \le (\frac{1}{2})L_u W_u$ for all x.

PROOF: To prove part a, the nonnegativity of $e_u(x)$ follows directly from the Centroid Aggregation Lemma, and uses the triangle inequality. From Minkowski's inequality, (see Hardy, Littlewood, and Polya [17]), we conclude that $e_u(x) = 0$ if and only if either $x - v_j \ge 0$, for all $v_j \in V_u$, or $x - v_j \le 0$, for all $v_j \in V_u$. Hence, $e_u(x)$ is positive if and only if x is an interior point of CH_u. The proof of part b appears in the Appendix.

We note that part a of the above theorem has been observed by Plastria [28]. Also note that part b implies that the maximum error of $e_u(x)$ occurs at the centroid c_u, and that there is an easily computable upper bound on $e_u(c_u)$. Furthermore, it can be shown that the upper bound is tight iff the total demand weight in V_u is equally distributed between the two endpoints of V_u.

3.2.2. Contiguous Partition Properties

We call a subset V_u of V *contiguous* if there exist indices k and t, $k \le t$ such that $V_u = \{v_k, v_{k+1}, \ldots, v_t\}$. We call a partition $V_1, V_2, \ldots, V_p$ of $\{v_1, \ldots, v_n\}$ *contiguous* if for any pair of distinct indices s, t, $1 \le s, t \le p$, the intersection of CH_s and CH_t is empty. Note that a partition is contiguous if and only if each subset of the partition is contiguous. Theorem 1 implies the following corollary. The corollary states where the error is zero, where it is positive, and that it is largest at some centroid, c_u.

COROLLARY 2: Let $V_1, \ldots, V_p$ be a contiguous partition of $\{v_1, \ldots, v_n\}$. Then $e(x) = e_u(x)$, for any x in CH_u, $u = 1, \ldots, p$; $e(x) = 0$ if x is not in the union of the intervals

$CH_1, \ldots, CH_p$, i.e., the error $e(x)$ is zero between adjacent intervals of the partition. Also, $\max\{e(x): x\} = \max\{e_u(c_u): u = 1, \ldots, p\}$.

In the remainder of this section, we consider only contiguous partitions. In addition to the practical aspects, the next theorem gives the main motivation for concentrating on contiguous partitions. We show, for the equally weighted case, that the minimum value (over all possible partitions of V) of the maximum error is attained when the partition is contiguous. To facilitate the discussion we introduce the following notation. For each partition $V_1, \ldots, V_p$, we let $E(x: V_1, \ldots, V_p)$ denote the error function $e(x)$ corresponding to the partition $V_1, \ldots, V_p$.

THEOREM 3: Suppose that $v_1 \le v_2 \le \cdots \le v_n$, and for $j = 1, \ldots, n$, the demand weight w_j that is associated with v_j is equal to 1. With p fixed, over all partitions of V into p subsets, $E(x: V_1, \ldots, V_p)$ is minimized on a contiguous partition.

PROOF: See the Appendix.

Theorem 3 is valid also for weighted problems, provided that for $j = 1, \ldots, n$, the demand w_j at a point v_j can be split between (at most) two adjacent subsets of the partition. The following example shows that the theorem is not true when weights are unequal and splitting is not allowed.

Demand points 1–4 are located on the line at 0, 10, 10.99, and 12, and have respective weights of 1000, 100, 1, and 100. The following table shows the maximum error for the optimal grouping and the three possible contiguous groupings. It can be seen that the only optimal partition into two nonempty subsets is obtained by the noncontiguous grouping, 1 & 3 and 2 & 4. The reason that a noncontiguous partition is optimal in this example is that the weight at 0 is very large relative to the other weights, in particular the weight at 10.99. Thus the centroid of the group 1 & 3 is very close (located at 0.011) to 0 and so the maximum error over the group 1 & 3 [$e(0.011) = 22.97$] is small compared to the maximum error (=200) over the group 2 & 4. However, when grouping the demand at 0 with the demand at 10, the centroid will not be as close to 0, and so the maximum error over the group 1, 2 will exceed 200.

Groupings	**1 & 3; 2 & 4**	**1; 2 & 3 & 4**	**1 & 2; 3 & 4**	**1 & 2 & 3; 4**
Max. error	200.000000	200.009950	1818.181818	1836.494096

Lemma 4 states an upper bound on the error that holds for all x.

LEMMA 4: Let $V_1, \ldots, V_p$ be a contiguous partition. For $u = 1, \ldots, p$, let $W_u = \Sigma\ \{w_j: v_j \in V_u\}$. Then

$$e(x) \le \text{Max}\{e_u(x): u = 1, \ldots, p\} \le (\tfrac{1}{2})\text{Max}\{W_u L_u: u = 1, \ldots, p\}, \text{ for all } x.$$

PROOF: The result follows from Theorem 1 and Corollary 2.

In order to help obtain an upper bound on the relative error, we now give a lower bound on $f(x)$ for all x. In what follows, let b and B be the values of the smallest and largest elements in $\{L_u: u = 1, \ldots, p\}$. Similarly let ω and Ω be the smallest and largest elements in $\{W_u: u = 1, \ldots, p\}$.

LEMMA 5: Let $V_1, \ldots, V_p$, be a contiguous partition. Then for any real x and for any $p \geq 3$, $f(x) \geq \omega(b/4)(p-1)^2$ if p is odd, and $f(x) \geq \omega(b/4)p(p-2)$ if p is even.

PROOF: See the Appendix.

The next corollary follows directly from Lemmas 4 and 5.

COROLLARY 6: Let $V_1, \ldots, V_p$ be a contiguous partition. For each real x, the relative error, $e(x)/f(x)$, satisfies

$$e(x)/f(x) \leq (2B\Omega)/(b\omega p(p-2)).$$

PROOF: From Lemma 4 and the definition of B and Ω, $e(x) \leq \frac{1}{2}\max\{W_u L_u: u = 1, \ldots, p\} \leq \frac{1}{2}B\Omega$. From Lemma 5, $f(x) \geq \omega(b/4)\min\{p(p-2), (p-1)^2\}$. Since $p(p-2) < (p-1)^2$, the result follows.

In the case of a uniform discrete demand distribution, where the set of demand points is partitioned into p intervals of equal length, i.e., $\omega = \Omega$ and $b = B$, the upper bound on the relative error is asymptotically tight (in the parameter p).

Next, we wish to find a contiguous partition into p subsets which minimizes the maximum aggregation error, i.e., minimizes $\max\{e_u(c_u): u = 1, \ldots, p\}$. To introduce the solution procedures, we first introduce a unifying formulation for min–max contiguous partitioning problems defined on the set $\{v_1, \ldots, v_n\}$. A *contiguous partition* can be defined by a set of $p - 1$ dividers, indices separating consecutive subsets. A *contiguous subset* V_u is defined by a pair of indices, say k, t, such that $V_u = \{v_k, v_{k+1}, \ldots, v_t\}$. We assume that, for each such pair $k \leq t$, there is a nonnegative real number $a[k, t]$, called the *value* of V_u. For example, for our problem of minimizing the maximum error, we define

$$a[k, t] = \sum \{w_j|v_j - c_u|: j = k, k+1, \ldots, t\},$$

where $c_u = \Sigma\{w_j v_j: j = k, \ldots, t\}/\Sigma\{w_j: j = k, \ldots, t\}$ is the centroid of V_u. Since with centroid aggregation, $f'(c_u) = 0$, we note that $e_u(c_u) = a[k, t]$, when $V_u = \{v_k, v_{k+1}, \ldots, v_t\}$. We first observe that for our problem of minimizing the maximum error of the median model, after an $O(n)$ preprocessing, $a[k, t]$ can be computed for any $k \leq t$ in $O(\log n)$ time. To show this, let $c[k, t]$ be the centroid of the contiguous set $v_k, \ldots, v_t$. Also, define $\theta_t = \Sigma\{w_j: j = 1, \ldots, t\}$, the sum of all weights from 1 to t, and $\Delta_t = \Sigma\{w_j v_j: j = 1, \ldots, t\}$. Note that it takes $O(n)$ time to compute all θ_t and Δ_t, $t = 1, \ldots, n$, since they can be computed recursively. But then for any $k \leq t$ note that

$$c[k, t] = (\Delta_t - \Delta_{k-1})/(\theta_t - \theta_{k-1}).$$

For fixed k and t, let j' be the largest j such that $v_j \leq c[k, t]$. Then we have

$$a[k, t] = (\theta_{j'} - \theta_{k-1})c[k, t] - (\Delta_{j'} - \Delta_{k-1}) + (\Delta_t - \Delta_{j'}) + (\theta_t - \theta_{j'})c[k, t].$$

Finding j' takes $O(\log n)$ effort, but $c[k, t]$ and $a[k, t]$ can be computed in constant time.

3.2.3. Algorithms for Contiguous Partitions

We now give a Bisection Method for solving our model. This method, as well as the dynamic programming approach given in the Appendix, depends on the following Monotonicity Property of the $a[k, t]$. The proof of this property appears in the Appendix.

Monotonicity Property. Suppose that $v_1 < v_2 < \cdots < v_n$, and $w_j > 0$, for $j = 1, \ldots, n$. Then $a[k+1, t] < a[k, t] < a[k, t+1]$, for $1 \leq k \leq t \leq n$.

Bisection Method. We describe a general bisection approach. This approach is similar to the scheme given in Megiddo and Tamir [24] and Manne and Sorevik [22]. To solve the model, we use binary search on a parameter r.

Given r, let $p(r)$ be the *minimum* number of consecutive subintervals in the partition, such that the value of each subinterval is at most r. Clearly, the optimal solution to our *minmax* problem is the smallest value of r such that $p(r) \leq p$. The function $p(r)$ is nonincreasing with r, so we can use a binary search to find the smallest value of r, such that $p(r) \leq p$. Note that we can either do (a) exact binary search on the set $\{a[k, t]\}$, $1 \leq k \leq t \leq n$, since the optimal value is one of the $O(n^2)$ values, or (b) view r as a real parameter and use bisection until the length of the remaining interval containing the optimal value is smaller than some prespecified precision level ε. In (a) we apply exact binary search on the set $\{a[k, t]\}$, we compute $p(r)$ for $O(\log n)$ values in $\{a[k, t]\}$. In (b), the search over values of r, we compute $p(r)$ for $O(\log (M/\varepsilon))$ times, where $M = a[1, n]$, the largest possible value.

Computation of $p(r)$. Finally, the computation of $p(r)$ [determining if $p(r) \leq p$] for a given value of r is done as follows:

We assume that the values $\{a[k, t]\}$ are given, or can be computed in constant time, after some preprocessing. Consider the sequence of DPs $v_1, v_2, \ldots, v_n$.

Step 0.
    Set $p(r) = 0$, and $i = 1$.
Step 1.
    Using binary search on the index set $\{i, \ldots, n\}$, let $\bar{j}$ the largest j such that $a[i, j] \leq r$.
    Add 1 to $p(r)$.
    If $p(r) > p$, stop, the optimal value is bigger than r.
    If $\bar{j} = n$, stop, $p(r) \leq p$, and the optimal value is at most r.
Step 2.
    Otherwise, set $i = \bar{j} + 1$, and return to Step 1.

The validity of the above procedure to determine whether $p(r) \leq p$ follows directly from the monotonicity property of $\{a[k, t]\}$. In the description given above the procedure terminates after p steps. At each step we perform a binary search where we need to compute the values of $O(\log n)$ terms of the sequence $\{a[i, j]\}, j = 1, \ldots, r$. Hence, if any such term can be computed in time T, the total complexity is $O(pT \log n)$. We have shown above that for our median problem $T = O(\log n)$, and therefore the complexity is $O(p \log^2 n)$. We also note in passing that if we replace the binary search of Step 1 by a successive evaluation of $a[i, i+1]$, $a[i, i+2], \ldots, a[i, \bar{j}]$, the overall complexity will be $O(n)$. This observation follows from the fact that all the centroids $c[i, i+1], c[i, i+2], \ldots, c[i, \bar{j}]$ and all the terms $a[i, i+1]$, $a[i, i+2], \ldots, a[i, \bar{j}]$ can be computed in $O(|\bar{j} - i|)$ time, by using the compact expressions for $c[k, t]$ and $a[k, t]$, given above. Binary search calls the above procedure repeatedly, doing bisection search on r to determine whether the optimal value, say ERR, is bigger than or equal to r. Since ERR is one of the terms in the set $\{a[k, t]\}$, we can utilize the monotonicity property to search efficiently over this set. In particular, if we apply the procedure in Megiddo et al. [25] [recalling that $T = O(\log n)$], we can find ERR in $O(n \log^2 n)$ time.

Alternatively, we can implement the idea behind the search routine in Megiddo and Tamir [24], used originally to solve the p-center problem on the line. For the sake of completeness we briefly describe this idea. There are p stages. In the first stage we search for ERR in the sequence $a[1, 1], a[1, 2], \ldots, a[1, n]$, using the above $O(p \log^2 n)$ procedure to determine if $p(r) \le p$. In $O(\log n)$ trials we identify an index, say j, such that $a[1, j] < \text{ERR} \le a[1, j+1]$. If ERR is strictly less than $a[1, j+1]$, then the first set of an optimal partition must consist of the points $\{v_1, \ldots, v_j\}$. Thus, in $O(p \log^3 n)$ time we identify the index j, and can proceed to the second stage where we now search over the sequence $a[j+1, j+1], \ldots, a[j+1, n]$, etc. Altogether there will be p stages, and therefore the total complexity [including the initial $O(n)$ preprocessing time] is $O(n + p^2 \log^3 n)$. Thus, if $n \log^2 n < n + p^2 \log^3 n$, the search procedure of Megiddo et al. [25] should be used. Otherwise, the procedure of Megiddo and Tamir [24] is preferred. We summarize our results as follows.

THEOREM 7: Suppose we are given a set of n demand points on the real line, and a positive integer p. Let ERR denote the minimum value of the maximum error of a centroid decomposition for the 1-median problem, over the set of all centroid decompositions defined by partitions into p contiguous subsets of demand points. For any positive ε, the bisection algorithm computes, in $O(n \log[a[1, n]/\varepsilon])$ time, a partition into p contiguous subsets with a maximum centroid decomposition error that is bounded above by ERR + ε. The exact bisection algorithm computes an optimal partition in $O(\min[n \log^2 n; n + p^2 \log^3 n])$ time. On the other hand, the dynamic programming algorithm (see Appendix) computes ERR and an optimal centroid decomposition of maximum error value ERR in $O(n \log n)$ time.

The above algorithms find a best centroid decomposition over the set of all such decompositions defined by partitions into contiguous subsets of demand points. The example following Theorem 3 shows that if demand weights are not identical and unsplittable, then an optimal centroid decomposition is not necessarily defined by a partition into contiguous subsets. (As far as we know, the complexity of finding an optimal centroid decomposition among the set of all, not necessarily contiguous, decompositions is still open for this weighted, unsplittable case.) As noted above, when demand weights are splittable there is an optimal decomposition defined by a partition into contiguous subsets. Indeed, if we assume that the demand weights are integer and splittable into integral parts, an optimal centroid decomposition can be obtained by a modified version of the above bisection algorithm. Specifically, if we let $w_{\max} = \max\{w_j : j = 1, \ldots, n\}$, then an optimal centroid decomposition into p subsets in the splittable case can be found in $O(n + p^2 \log n \log^2(nw_{\max}))$. For the sake of brevity we omit the details.

4. ERROR FOR THE PLANAR PROBLEM WITH RECTILINEAR DISTANCES

We now give a centroid row–column aggregation (CRC) algorithm that is shown to generate an error no larger than the error which would result by using aggregation points specified by ARC. Given the cells (intersections of rows and columns) provided by ARC, the individual cell centroids and weights redefine the approximating problem. We also establish a bound on the relative error. This bound leads to a decreasing returns to scale error phenomenon we consider practically important.

LEMMA 8: Given an ARC aggregation, for each Row_s and Col_t with demand points in their intersection having total weight W_{st}, define

$$(c_1^{st}, c_2^{st}) = \sum \{(w_i/W_{st})(p_1^i, p_2^i)\colon (p_1^i, p_2^i) \text{ in } \text{Row}_s \text{ and } \text{Col}_t\}.$$

(a) For all x_1, we have

$$W_t^1|x_1 - c_1^t| \leq \sum \{W_{st}|x_1 - c_1^{st}|\colon s = 1, \ldots, n_2\}.$$

(b) For all x_2, we have

$$W_s^2|x_2 - c_2^s| \leq \sum \{W_{st}|x_2 - c_2^{st}|\colon t = 1, \ldots, n_1\}.$$

PROOF: It is enough to prove part (a). Note that c_1^t is the centroid of the $\{c_1^{st}\}$, $s = 1, \ldots, n_2$. Thus the result follows from the Centroid Aggregation Lemma.

LEMMA 9: Given an ARC aggregation, let $g(x_1, x_2)$ be the approximating function defined by Table 1:

$$g(x_1, x_2) = \sum \{W_{st}(|x_1 - c_1^t| + |x_2 - c_2^s|)\colon s = 1, \ldots, n_2, t = 1, \ldots, n_1\}.$$

Suppose another approximating function f' is defined by

$$f'(x_1, x_2) = \sum \{W_{st}(|x_1 - c_1^{st}| + |x_2 - c_2^{st}|)\colon s = 1, \ldots, n_2, t = 1, \ldots, n_1\}.$$

For all (x_1, x_2) we have $g(x_1, x_2) \leq f'(x_1, x_2) \leq f(x_1, x_2)$. Therefore, the error in using f' is never more than the error in using g.

PROOF: Lemma 8 establishes $g(x_1, x_2) \leq f'(x_1, x_2)$. By the Centroid Aggregation Lemma, $f'(x_1, x_2) \leq f(x_1, x_2)$.

We now describe CRC. For a given choice of n_1 and n_2, we use the methodology of Section 3 to do independent demand point aggregation on the x_1 and x_2 axes, resulting in row spacings and column spacings. Centroids of the individual cells are used as aggregation points.

CRC Algorithm

Input: $P = \{(p_1^i, p_2^i)\colon i = 1, \ldots, m\}$, $\{w_i\colon i = 1, \ldots, m\}$

1. Choose positive integers n_1 and n_2.
2. Set up the functions f^1 and f^2 defined in Table 1. Rank the demand point x_1 and

x_2 coordinates in increasing order; add weights of demand points on the line with the same coordinate.

3. Find optimal contiguous partitions (see Section 3) of the x_1 coordinate demand points and x_2 coordinate demand points into n_1 and n_2 subsets respectively. Denote these two resulting minimal maximum errors by er_1 and er_2 respectively.
4. Define the approximating model $f'(X)$ by using the centroid of each cell having demand points as the aggregate demand point for all points in the cell (see Lemma 9). The weight of the centroid of each cell is the total weight of all demand points in the cell.

Output. An ARC aggregation and approximating location model $f'(X)$ with error at most $er_1 + er_2$.

Step 2 of the algorithm can be done in $O(m \log m)$. Step 4 is $O(n_2 n_1 + m)$; in the worst case, each cell and each demand point must be considered. Note n_2 and n_1 are typically small compared to m. Most of the effort occurs in step 3, and depends on which algorithm (Section 3 and the Appendix) is used.

Consider now the relative error $e(x_1, x_2)/f(x_1, x_2)$ generated by the function $g(x_1, x_2)$ defined in the ARC algorithm. Since $e(x_1, x_2) = e^1(x_1) + e^2(x_2)$ and $f(x_1, x_2) = f^1(x_1) + f^2(x_2)$ (see Table 1), we have $e(x_1, x_2)/f(x_1, x_2) \le e^1(x_1)/f^1(x_1) + e^2(x_2)/f^2(x_2)$, the sum of the relative errors for x_1 and x_2, respectively. If we have no other information about x_1 and x_2, we can use Corollary 6 to bound this relative error. With reference to the definition of an ARC aggregation, for every s and t, let L_t^1 and L_s^2 denote the width and height of Col_t and Row_s, respectively, with total weights W_t^1 and W_s^2 respectively. Define

$$\begin{aligned} b^1 &= \min\{L_t^1: t = 1, \ldots, n_1\} > 0, & b^2 &= \min\{L_s^2: s = 1, \ldots, n_2\} > 0, \\ B^1 &= \max\{L_t^1: t = 1, \ldots, n_1\}, & B^2 &= \max\{L_s^2: s = 1, \ldots, n_2\}, \\ \omega^1 &= \min\{W_t^1: t = 1, \ldots, n_1\} > 0, & \omega^2 &= \min\{W_s^2: s = 1, \ldots, n_2\} > 0, \\ \Omega^1 &= \max\{W_t^1: t = 1, \ldots, n_1\}, & \Omega^2 &= \max\{W_s^2: s = 1, \ldots, n_2\}. \end{aligned}$$

It now follows from Corollary 6, for any ARC aggregation with $n_1, n_2 \ge 3$, that

$$e(x_1, x_2)/f(x_1, x_2) \le (2B^1\Omega^1)/[b^1\omega^1 n_1(n_1 - 2)] + (2B^2\Omega^2)/[b^2\omega^2 n_2(n_2 - 2)], \text{ for all } (x_1, x_2).$$

Due to Lemma 9, this inequality is also true for any CRC aggregation with $n_1, n_2 \ge 3$. The latter displayed inequality suggests that the relative error decreases at a decreasing rate as n_1 and n_2 increase. This phenomenon has been noted in other related work (for example, see Francis et al. [11]) and is practically important.

For purposes of insight into this decreasing returns to scale error phenomenon, consider the idealized case where the demand points are continuously and uniformly distributed over a rectangle B. Denote the x_1 and x_2 dimensions of B by L^1 and L^2, respectively. If we have n_2 rows and n_1 columns, then it can be shown that the rows and columns generated by the ARC algorithm will have, respectively, widths of L^1/n_1 and heights of L^2/n_2. Note that each $W_t^1 = 1/n_1$, and each $W_s^2 = 1/n_2$. We know the maximum x_1 error is the largest of the $f^1(x_1)$ restricted to the demand points in some Col_t and evaluated at the x_1-coordinate centroid of Col_t (which is the midpoint of Col_t). The maximum x_1 error turns out to be $L^1/(4n_1^2)$. Likewise, the maximum x_2 error is $L^2/(4n_2^2)$. Thus the maximum error is $L^1/(4n_1^2) + L^2/(4n_2^2)$. Assuming

$A = L^1L^2$, and $q = n_1n_2$ is a constant, if we relax the integrality conditions on n_1 and n_2, we can find the values of n_1 and of n_2 that minimize this upper bound. We find that $n_1^* = (L^2/L^1)^{1/4}q^{1/2}$, $n_2^* = (L^1/L^2)^{1/4}q^{1/2}$ and the bound, for these values of n_1 and n_2, becomes $(0.5\sqrt{(A)})/q$. Thus, the maximum error is proportional to $\sqrt{A}$ and inversely proportional to q, so the error decreases at a decreasing rate as q increases. Computational experience indicates that this k/q phenomenon is robust.

Next, we consider the relative error for the case of uniformly distributed demand. From Corollary 6, using $\omega = \Omega$, $b = B = L^1/n_1$ or L^2/n_2, we conclude that an upper bound on the relative error is $2/(n_1(n_1 - 2)) + 2/(n_2(n_2 - 2))$. Clearly this upper bound goes to zero quickly as n_1 and n_2 increase. For example, if $n_1 = n_2 = 5$, 10 or 20, then the upper bound is 0.2 (20%), 0.05 (5%), and 0.0111 (1.11%), respectively. Note these bounds are independent of L^1 and L^2.

FLR consider a case similar to the one above. They obtain an approximate expression for an error bound (on the maximum error). Their error bound is $L^1/(4n_1) + L^2/(4n_2)$. By comparison, we have a bound on the maximum error of $L^1/(4n_1^2) + L^2/(4n_2^2)$. Certainly, this is an indicator of a much smaller bound. Note that if $q = n_1 \times n_2$ and $n_1 = n_2$, then these error measures are proportional to $1/\sqrt{q}$, and $1/q$, respectively. Our computational experience is consistent with these measures.

Consider another insight of interest. We can think of the partition of demand points provided by CRC as being a collection of city blocks, with cells corresponding to blocks. Then, using Corollary 2, we can conclude that if centroid aggregation is used at the city block level, the error is zero at each street intersection. In this case, there would be many places of interest for which there is no error.

5. COMPUTATIONAL EXPERIENCE

Much of the theory of CRC has been developed for the rectilinear distance 1-median problem. However, the aggregate demand points it provides can be used to define an approximating problem for the NP-hard rectilinear distance n-median problem. To test how well CRC worked for n-median problems, we used computational experimentation. All runs were made on a Unix Sun sparc station (OS 5.6); the program was coded in C++. The exact bisection method was used; execution times were relatively small. The elapsed time for the largest problem we solved (a real data set from Palm Beach County, Florida) was about 20 s. For this problem we used CRC with 30 rows and columns. Memory requirements are reasonable; these are basically linearly proportional to m, the number of demand points.

We used much the same computational testing method as in FLR. This approach facilitated making comparisons between our approach and theirs. Much of the following description of the method is taken from FLR. Our objectives were to study various error values and determine their dependence on the number of aggregate demand points, and to develop qualitative insights. Further, we wished to compare our approach with the previous row–column approach. In all our experimentation, we took every demand point weight to be $1/m$, with m the number of demand points (equivalent to taking every weight to be 1).

We studied a computer-generated problem we call the "central tendency" (CT) problem that defined the distribution of the demand points. Each marginal demand point density function of this distribution is a symmetric triangular distribution with a value of zero at endpoints of its interval of definition. This problem simulates demand point locations in an urban area with the highest population concentration in the middle of the area. The box B on which the distribution was defined had dimensions of 1000 by 1000.

Table 2. CRC computational experimentation with 25,000 demand points, based on the Central Tendency distribution, for 5-median problems.

$n_2 = n_1$	ame	sae	sme	sare	smre	avg. q
5	16.827	6.534	20.574	2.90%	9.62%	25
10	4.327	1.632	4.842	0.72%	2.20%	100
15	1.94	0.707	2.188	0.31%	1.00%	225
20	1.098	0.416	1.24	0.19%	0.56%	400
25	0.703	0.253	0.785	0.11%	0.35%	625
30	0.49	0.171	0.518	0.08%	0.24%	899.95

We now describe our central tendency experiments. We varied m in increments of 5000 between 5000 and 25,000, and considered 1, 3, and 5 as values of n. For each given m and n value we took $n_1 = n_2$ and varied n_1 in increments of 5 between 5 and 30, resulting in values of q ranging between 25 and 900. For given m, n, n_1, and n_2, we created and solved 20 central tendency problems generated with the Monte Carlo method. Define a *sample* to be a collection of choices, X, of n new facility locations. For each given problem, we randomly generated 100 samples. We sampled only from coordinates of demand points. For each X we computed the n-median function value $f(X)$, and the approximating n-median value $f'(X)$ using the aggregate demand points provided by CRC. From these function values for the sample we computed the absolute errors. Then, using all 100 samples, we computed the sample average error (sae) and sample maximum error (sme). Likewise, we computed the sample average relative error (sare) and sample maximum relative error (smre). The error value provided by CRC for each problem was averaged over 20 problems to give what we call the average maximum error (ame). Similarly, the sae, sme, sare, and smre values computed as above for each sample were then averaged over all 20 problems. Table 2 illustrates results of our experimentation for $n = 5$. In the table, each entry represents an average over 100 samples (of X) and 20 sets of demand points. Since demand points are randomly generated, it is possible that some cells may have no demand points. In such cases, we adjusted q accordingly and averaged over all 20 values of q to give the average q value reported. Such a case occurred as is illustrated below where one cell in one of the 30 by 30 problems was empty.

Table 2 clearly illustrates how the error measures rapidly decrease as $n_2 = n_1$ increases. For example, note that the smre value is less than 1% for $n_2 = n_1 \geq 15$. For the data of Table 2, Figure 1 shows a graph of how the sare and smre values from CRC vary versus (the average value of) q. In addition to plotting the sare and smre data, Figure 1 shows the result of using the Power Curve Fitting option in Excel to fit a power curve of the form aq^b to the smre graph, with an $R^2 \approx 1$. We shall sometimes refer to a power curve as aq^b and sometimes as a/q^b (with b changed accordingly), depending on which form is more convenient.

We have fitted such power curves for most of our experimentation. Much of the motivation for trying power curves comes from the formula $(\frac{1}{2}\sqrt{A})q^{-1}$ for uniformly distributed demand points discussed in Section 4. These power curves have R^2 values so close to 1.00 that they provide a very useful way of summarizing all the error graphs. Data on the power curves appears in the upper part of Table 3.

We compared CRC with the previous (FLR) row–column method (abbreviated as MRC to denote that cells were determined via a row–column procedure and that medians of demand points in each cell were used as aggregate points). The lower part of Table 3 reports on the results of our experimentation with MRC with central tendency distributed data and $n = 5$. (Results for $n = 1$ and 3 are available from the authors and were quite similar.) As with CRC, the numbers in the table corresponding to MRC are averages over 100 samples and 20 demand

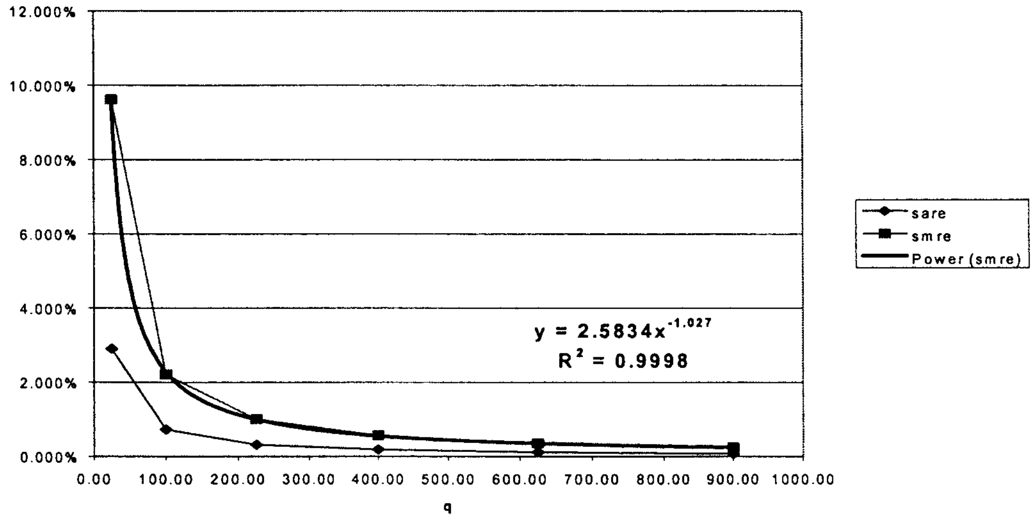

Figure 1. CRC sare and smre Values vs. q, $n = 5$, $m = 25{,}000$.

point sets. The $n_2 = n_1$ values were the same as for the CRC runs. MRC uses an error bound, an upper bound on the maximum error, which is valid for all n. It finds a row–column aggregation to minimize (heuristically) the error bound over a certain class of such aggregations. This bound is the sum of weighted distances between each demand point and the aggregate demand point that replaces it. We abbreviate the error bound as meb. We abbreviate the average of these error bounds over all the data sets as ameb.

We can see from Table 3 that each of the error measures for CRC is essentially of the form a_{CRC}/q, for some positive constant a_{CRC}. By contrast, the error bound measure (ameb) for MRC is approximately of the form $a_{MRC}/\sqrt{q}$ (see discussion at the end of Section 4). Each of the other measures for MRC is roughly of the form $a_{MRC}/q^{0.9}$. Comparing the corresponding values in Table 3 it is clear that CRC does a better job than MRC. In our experiments we found that as n increases, error measures mostly increase with n and MRC becomes more competitive with CRC. With $n = 1$ we found that the fitted formula for ame was $405.48/q^{0.9871}$. Note that the formula $(0.5\sqrt{(A)})/q$ would predict a maximum error of $500/q$ (although the demand points

Table 3. Error power curve fits, aq^b, with central tendency demand point data for both CRC and MRC, with $m = 25{,}000$ demand points, for $n = 5$.

CRC Central Tendency Error Function Power Fits, aq^b					
$n = 5$	ame	sae	sme	sare	smre
a	405.71	171.35	540.44	0.7589	2.5834
b	−0.9872	−1.0119	−1.018	−1.0116	−1.027
R^2	1.0000	0.9997	0.9997	0.9997	0.9998
MRC Central Tendency Error Function Power Fits, aq^b					
$n = 5$	ameb	sae	sme	sare	smre
a	419.67	202.8	538.52	0.9012	2.6784
b	−0.4978	−0.8917	−0.9428	−0.891	−0.9565
R^2	0.9997	0.9988	0.9997	0.9989	0.9999

are *not* uniformly distributed). All the above discussion is for $m = 25{,}000$, but results with smaller m values were much the same.

We also did some testing of problems we call the Gainesville (Gvl) and Palm Beach County (PBCo) problems. The experimental approach was the same as for the central tendency problems with the same $n_2 = n_1$ values, except that we used a sample size of 1000. We used $n = 3$, and had only one data set for each problem. These problems had m values of 11,993, and 69,960 respectively. The former problem is a computer-generated problem based on a map of Gainesville, Florida; it spaces hypothetical demand points equally along the major streets. For the PBCo problem, about 96% of the demand points are in the eastern third of the county, and all fall on the street network of the county. A more detailed description of these two problems can be found in FLR, as well as figures for each showing the demand points and an example of an (old) row–column aggregation. Detailed results of the experiments are available from the authors.

Many of the conclusions made for the central tendency runs apply to the Gvl and PBCo runs. The most notable difference is that error graphs now appear to be more of the form a/q^b with b in the range 1.06–1.14 for PBCo, and 1.27–1.42 for Gvl, whereas $b \approx 1$ for the CT runs. It is interesting to note also that the exponents for sae and sare were almost identical, as were those for the sme and smre curve fits. The ratio of the first to the second error curve is essentially constant. We found that CRC always outperformed MRC with respect to sae, sme, sare, and smre, although for large q values there was not much difference.

The formula $(0.5\sqrt{(A)})/q$ would predict $17.397/q$ and $128237.65/q$ for Gvl and PBCo, respectively. The CRC curve fits for ame were $31.73/q^{1.1823}$ and $111561/q^{1.1699}$, respectively. For $28 \le q \le 74$, each prediction exceeds the corresponding fit, while for $75 \le q \le 1000$, each prediction improves as q increases. The Gainesville prediction is the better of the two.

We found that CRC gave smre values less than 1% for both Gvl (0.819%) and PBCo (0.466%) with $n_2 = n_1 = 15$ and 20 respectively ($q = 225$ and 400, respectively). The PBCo problem has almost six times as many demand points as the Gvl problem. Assuming integer values of q, the use of the formulas $0.00819 = 3.3792/q^{1.1049}$ and $0.00466 = 7.3014/q^{1.3345}$ would give $q = 233$ and $q = 248$, respectively. The use of the formulas $0.01 = 3.3792/q^{1.1049}$ and $0.01 = 7.3014/q^{1.3345}$ would give $q = 50$ and $q = 140$, respectively, to achieve an error of 1%. In general, it seems striking how few aggregate demand points, as compared to actual demand points, are needed in order to achieve quite small errors.

6. CONCLUSIONS

While the theory of CRC is based on the 1-median problem, we believe CRC can be quite effective for doing n-median problem demand point aggregation, $n \ge 2$. Our computational testing found CRC to be uniformly better than the row–column method of FLR, which was specifically designed for n-median model aggregation. For example, with $n_2 = n_1 = 15$ or 20, we found the sample maximum relative error to be no more than 1% for problems with as many as 70,000 demand points. All the error measures we examined for CRC could be very well modeled by a power function in q of the form a/q^b, with $b \ge 1$. This power function nicely captured how the errors varied with q (the number of aggregate demand points). We found that q can be small, compared to m, and CRC will still provide a very good aggregation.

Assuming error behavior in q can be well modeled by an error power function, say $\mathrm{er}(q) = a/q^b$, with $b \ge 1$, it is interesting to investigate some of the implications. Since $\mathrm{er}(q)$ is strictly decreasing and continuous, it has an inverse function, denoted by $\mathrm{er}^{-1}(t)$. With $b' = 1/b$, $a' = (1/a)^{b'}$, we have $\mathrm{er}^{-1}(t) = a'/t^{b'}$. We interpret $\mathrm{er}^{-1}(t)$ as the number of aggregate demand points we need to obtain an error of t. If we want a value of q so that the error is at most t, $\text{er}(q) \leq t$, we need $q \geq \text{er}^{-1}(t)$. Further, because the inverse function is a power function that decreases at a decreasing rate, to achieve a very small error, we may need a relatively large number of aggregate demand points. Allowing only a slightly larger error might significantly decrease the number of aggregate demand points needed.

Francis and Lowe [8] speculated that a model like an economic order quantity model could be used to find a number of aggregate demand points to minimize the total cost of an aggregation. The cost might be the sum of an error cost, say α/q, and an aggregation cost, say β/q. If the power function expression for error proves to be robust, such a model may be possible, with perhaps the error costs and aggregation costs modeled somewhat more generally. Finally, note that if the error function is of the form a/q^b, it would only be necessary to make runs for two different q values in order to fit the function. The following question now becomes interesting. For what class of demand point distributions will the error curves be of this form? There appear to be promising opportunities for further research on this question.

While the question of how to use CRC with shortest-path network distances in a GIS context remains open, we are optimistic that an approach similar to that used by Andersson et al. [1] can be used. They adapted the (different) row–column approach of [10] for use with shortest-path network distances. Basically, their approach was to extract the subnetwork spanning each cell in the row–column aggregation, and then solve a location problem on the subnetwork to find an aggregate demand point for the cell. They found this approach worked well for n-median problems.

APPENDIX

We first show, as was asserted in Section 3.1, that, for any centroid aggregation, the maximum error and maximum relative error for the rectilinear 1-median occur at some grid point defined by the q^2 cross-products of all centroid coordinates.

LEMMA 10: Given a centroid aggregation for the rectilinear 1-median problem, with a partition of the demand set into q subsets, let $\{C_u = (c_1^u, c_2^u): u = 1, \ldots, q\}$ denote the respective set of centroids. Consider the cell partition of the plane into closed rectangular cells, defined by the q vertical lines $\{(x_1, x_2): x_1 = c_1^u\}$, $u = 1, \ldots, q$, and the q horizontal lines $\{(x_1, x_2): x_2 = c_2^u\}$, $u = 1, \ldots, q$. Then on each cell of the partition, the function $f'(x_1, x_2)$ is linear, the error function $e(x_1, x_2) = f(x_1, x_2) - f'(x_1, x_2)$ is convex, and the relative error function $e(x_1, x_2)/f(x_1, x_2)$ is quasiconvex.

PROOF: The linearity of $f'(x_1, x_2)$ over any cell follows directly from the definition of the cell partition. The function $f(x_1, x_2)$ is convex over the entire plane. Therefore, the error function $e(x_1, x_2)$ is convex over any cell. To prove the quasiconvexity of the relative error function on a given cell CE, we note that $\text{rel}(X) = e(X)/f(X) = 1 + (-f'(X)/f(X))$. But this is a constant term plus a nonpositive linear function divided by a nonzero convex function. This is sufficient (Avriel [2], page 156) to establish quasiconvexity. □

THEOREM 11: Consider a centroid aggregation for the rectilinear 1-median model, with a partition of the demand set into q subsets. Let $\{C_u = (c_1^u, c_2^u): u = 1, \ldots, q\}$ denote the respective set of centroids. Then the maximum error and the maximum relative error occur at a pair of points with the property that their x_1 coordinates are in the set $\{c_1^u: u = 1, \ldots, q\}$, and their x_2 coordinates are in the set $\{c_2^u: u = 1, \ldots, q\}$.

PROOF: From Lemma 10, we know that both the error and the relative error functions are quasiconvex over each cell. Therefore, the maximum error and the maximum relative error over any bounded cell CE are attained at one of the four extreme points of CE (see Mangasarian [21]). Clearly each such corner point has the property stated.

Consider any unbounded cell CE. We note that, for the rectilinear median problem, starting at a corner point of CE, the error function is monotone nonincreasing along each infinite edge incident to the corner point, and the relative error tends to zero along this edge. The supremum of a quasiconvex function over CE is equal to its supremum over the boundary of CE. Thus, we conclude from the above that the maximum error, and the maximum relative error, over CE occurs at one of the (at most) two corner points of CE. This completes the proof.

We now provide proofs of several results in Section 3.2. First, we consider Theorem 1, part b.

PROOF OF THEOREM 1, PART b: The first inequality follows directly from the triangle inequality, $|v_j - x| - |x - c_u| \leq |v_j - c_u|$. To prove the second inequality, suppose that $V_u = \{v_{j(1)}, \ldots, v_{j(t)}\}$, where $v_{j(1)} \leq v_j \leq v_{j(t)}$, for all v_j in V_u. Then $L_u = v_{j(t)} - v_{j(1)}$. Define the function

$$g(h_1, \ldots, h_t) = \sum \{w_{j(s)}|h_s - h'|: s = 1, \ldots, t\}$$

where

$$h' = \sum \{w_{j(s)}h_s: s = 1, \ldots, t\}/\sum \{w_{j(s)}: s = 1, \ldots, t\}.$$

Note that

$$\sum \{w_j|v_j - c_u|: v_j \in V_u\} = g(v_{j(1)}, \ldots, v_{j(t)}).$$

Now $g(h_1, \ldots, h_t)$ is a convex function, and therefore its maximum over the box defined by the constraints, $v_{j(1)} \leq h_s \leq v_{j(t)}$, $s = 1, \ldots, t$, is attained when $h_s \in \{v_{j(1)}, v_{j(t)}\}$, for all $s = 1, \ldots, t$. Let $(h_1^*, \ldots, h_t^*)$ be a maximum point, and let $I = \{s: h_s^* = v_{j(1)}\}$, and $J = \{s: h_s^* = v_{j(t)}\}$. Define $W(I) = \Sigma \{w_{j(s)}: s \in I\}$, $W(J) = \Sigma \{w_{j(s)}: s \in J\}$, and $W_u = \Sigma \{w_{j(s)}: s = 1, \ldots, t\}$. Then the maximum value of g in the above box is given by $2W(I)W(J)(v_{j(t)} - v_{j(1)})/W_u$. Since $W(I) + W(J) = W_u$, an upper bound on the maximum value of g is obtained when we set $W(I) = W(J) = W_u/2$. This proves the second inequality of the theorem.

PROOF OF THEOREM 3: To prove the theorem, let $V_1, \ldots, V_p$ be a noncontiguous partition of $V = \{v_1, \ldots, v_n\}$. It is sufficient to show that there exists a contiguous partition $V'_1, \ldots, V'_p$, such that $|V_u| = |V'_u|$, $u = 1, \ldots, p$, and $E(x: V_1, \ldots, V_p) \geq E(x: V'_1, \ldots, V'_p)$ for every real x.

Let $\{c_1, \ldots, c_p\}$ be the set of centroids corresponding to the subsets of the partition $V_1, \ldots, V_p$. Without loss of generality, suppose that $c_1 \leq c_2 \leq \cdots \leq c_p$. Due to the additivity of the error function, it is sufficient to prove the result under the assumption that the subset V_1 is not contiguous of the form $\{v_1, v_2, \ldots, v_m\}$, where $m = |V_1|$. (Otherwise, set $V'_1 = V_1$ and consider the partition of $\{v_{m+1}, \ldots, v_n\}$ defined by $V_2, \ldots, V_p$.)

We will show that we can perform a sequence of interchanges of elements of the subset V_1, and obtain a subset $V'_1 = \{v_1, \ldots, v_m\}$ while maintaining the above properties of the new partition.

Let v_t be the largest element in V_1. Since V_1 is not contiguous, there is a point v_s, which is not in V_1, and $v_s < v_t$. Suppose that v_s is in V_r. Consider the partition $V''_1, \ldots, V''_p$, obtained from $V_1, \ldots, V_p$, by interchanging v_t with v_s, i.e.,

$$V''_u = V_u, \qquad \text{for } u \neq 1, r, \qquad V''_1 = \{V_1 - \{v_t\}\} \cup \{v_s\}, \qquad V''_r = \{V_r - \{v_s\}\} \cup \{v_t\},$$

so that $|V_u| = |V''_u|$ for $u = 1, \ldots, p$. Define $z = v_t - v_s$. Since $v_t > v_s$, z is positive and so

$$c''_1 = c_1 - z/|V_1| < c_1 \text{ and } c''_r = c_r + z/|V_r| > c_r.$$

In particular, it follows that c''_1 is smaller than or equal to the centroids of all subsets V''_u, $u = 1, \ldots, p$. We then obtain, $\delta(x) \equiv E(x: V_1, \ldots, V_p) - E(x: V''_1, \ldots, V''_p) = |V_1|\,|x - c''_1| + |V_r||x - c''_r| - |V_1|\,|x - c_1| - |V_r|\,|x - c_r|$. The above difference, $\delta(x)$, is a continuous piecewise linear function with breakpoints at $\{c_1, c_r, c''_1, c''_r\}$. The function $\delta(x)$ is easily observed to be constant and zero outside the convex hull of $\{c_1, c_r, c''_1, c''_r\}$. It is straightforward to show that $\delta(x)$ is nonnegative at its four breakpoints, and thus it follows that $\delta(x)$ is nonnegative everywhere. This completes the proof of the theorem.

We now establish Lemma 5.

PROOF OF LEMMA 5: The minimum of $f(x)$ is attained at a weighted median of the set $V = \{v_1, \ldots, v_n\}$. Due to convexity of $f(x)$ it is sufficient to assume that x is in the interval $[v_1, v_n]$. Consider the case where x is a point in the kth interval, CH_k, and without loss of generality let $1 < k \leq (p + 1)/2$. Let $[A, B]$ denote the kth interval, and for simplicity let x be a real number between 0 and $b' = B - A$, where $x = 0$ refers to A, and $x = b'$ refers to B.

Now, for each interval j, $j \le k - 1$, the sum of weighted distances from the DPs in the interval to x is bounded below by $\omega[b(k - j - 1) + x]$.

For each interval j, $j \ge k + 1$, the respective lower bound is $\omega[b(j - k) - x]$. (Recall that $b' \ge b$, by definition of b.) Summing over all values of j, we get the following lower bound for the sum of weighted distances of *all* DPs from x:

$$\omega(b/2)[(k - 1)(k - 2) + (p - k + 1)(p - k)] + \omega x[(k - 1) - (p - k)].$$

By assumption, $k \le (p + 1)/2$, and so $[(k - 1) - (p - k)] \le 0$. But then the minimum of the above expression over all values of x, $0 \le x \le b$, is attained at $x = b$. Substituting $x = b$, we get the lower bound

$$g(k) = \omega(b/2)[k(k - 1) + (p - k)(p - k - 1)].$$

This bound obviously depends on k, the index of the interval assumed to contain the one-median, x.

To find a lower bound on $f(x)$ for all x, we want the integer minimizer of $g(k)$ in the range $k \le (p + 1)/2$. The real minimizer is $k' = p/2$. Thus, if p is even, $p/2$ is the minimum integer point, and the optimal value of g is $\omega(b/4)p(p - 2)$. If p is odd, by the convexity of $g(k)$ the integer minimum is found either at $k = (p + 1)/2$ or $k = (p - 1)/2$. Both of these values give $g(k) = \omega(b/4)(p - 1)^2$, and so the result follows.

We next provide a proof of the Monotonicity Property (see Section 3.2) of the set $\{a[k, t]\}$.

PROOF OF MONOTONICITY PROPERTY: Due to symmetry, it is sufficient to prove that for our median model $a[1, t] < a[1, t + 1]$, for $t = 2, \dots, n - 1$. Let j' be the largest index j such that $v_j \le c[1, t]$, and let j'' be the largest index j such that $v_j \le c[1, t + 1]$. Clearly, $c[1, t] < c[1, t + 1]$, and $1 \le j' \le j'' \le t$. From the definition of j' and j'' we now have $a[1, t + 1] - a[1, t] = (\Sigma\ \{w_j: j = 1, \dots, j'\})(c[1, t + 1] - c[1, t]) - (\Sigma\ \{w_j: j = j'' + 1, \dots, t\})(c[1, t + 1] - c[1, t]) + w_{t+1}(v_{t+1} - c[1, t + 1]) + (\Sigma\ \{w_j(c[1, t + 1] - v_j): j = j' + 1, \dots, j''\}) - (\Sigma\ \{w_j(v_j - c[1, t]): j = j' + 1, \dots, j''\})$, where if $j' = j''$; the last two terms in the sum are absent.

For $j \le j''$, $c[1, t + 1] - v_j \ge 0$, and therefore the third sum on the right-hand side of the above equation is nonnegative. Moreover, since $v_j \le c[1, t + 1]$ for $j \le j''$, the fourth sum on the right-hand side of the above equation is greater than or equal to $-(\Sigma\ \{w_j: j = j' + 1, \dots, j''\})(c[1, t + 1] - c[1, t])$. Thus $a[1, t + 1] - a[1, t] \ge (\Sigma\ \{w_j: j = 1, \dots, j'\})(c[1, t + 1] - c[1, t]) - (\Sigma\ \{w_j: j = j'' + 1, \dots, t\})(c[1, t + 1] - c[1, t]) - (\Sigma\ \{w_j: j = j' + 1, \dots, j''\})(c[1, t + 1] - c[1, t]) + w_{t+1}(v_{t+1} - c[1, t + 1]) = (\Sigma\ \{w_j: j = 1, \dots, j'\})(c[1, t + 1] - c[1, t]) - (\Sigma\ \{w_j: j = j' + 1, \dots, t\})(c[1, t + 1] - c[1, t]) + w_{t+1}(v_{t+1} - c[1, t + 1]) = 2(\Sigma\ \{w_j: j = 1, \dots, j'\})(c[1, t + 1] - c[1, t]) - (\Sigma\ \{w_j: j = 1, \dots, t\})(c[1, t + 1] - c[1, t]) + w_{t+1}(v_{t+1} - c[1, t + 1]) = 2(\Sigma\ \{w_j: j = 1, \dots, j'\})(c[1, t + 1] - c[1, t]) - (\Sigma\ \{w_j: j = 1, \dots, t + 1\})c[1, t + 1] + (\Sigma\ \{w_j: j = 1, \dots, t\})c[1, t] + w_{t+1}v_{t+1}$. Finally, from the definition of $c[1, t]$ and $c[1, t + 1]$ we have $(\Sigma\ \{w_j: j = 1, \dots, t + 1\})c[1, t + 1] = \Sigma\ \{w_j v_j: j = 1, \dots, t\} + w_{t+1}v_{t+1} = (\Sigma\ \{w_j: j = 1, \dots, t\})c[1, t] + w_{t+1}v_{t+1}$. Thus we conclude that $a[1, t + 1] - a[1, t] \ge 2(\Sigma\ \{w_j: j = 1, \dots, j'\})(c[1, t + 1] - c[1, t]) > 0$.

Finally, we give a dynamic programming method to solve the model given in Section 3.2.

Dynamic Programming Method. This method has been discussed in the literature extensively in other location contexts (see Hassin and Tamir [18]). For each index i, $i = 1, \dots, n$, and integer q, $q = 1, \dots, \min(n - i + 1, p)$, let $h(i, q)$ be the minimum of the maximum value of a subset in an optimal partitioning of the set of points $V^i = \{v_i, \dots, v_n\}$ into q nonempty consecutive (contiguous) subsets.

From the definition we have

$$h(i, 1) = a[i, n], \text{ and for } q = 2, \dots, \min(n - i + 1, p),$$

$$h(i, q) = \min\{\max(a[i, k], h(k + 1, q - 1)): k = i, \dots, n - q + 1\}.$$

The optimal solution value to the above partition problem is then given by $h(1, p)$.

Olstad and Manne [27] give an $O(pn)$ algorithm to solve the above model, under the following assumptions on the set of values $\{a[k, t]\}$:

i. $a[k, t]$ is positive for all $k \leq t$.
ii. $a[k + 1, t] < a[k, t] < a[k, t + 1]$, $k \leq t$.
iii. $a[k, k]$ can be computed in constant time.
iv. Given $a[k, t]$, we can calculate $a[k - 1, t]$, $a[k + 1, t]$, $a[k, t - 1]$ and $a[k, t + 1]$ in constant time.

Frederickson [14, 15] gives an $O(n)$ algorithm under **i–iii.** above as well as

v. $a[k, t]$ is computable in constant time for any pair $k \leq t$, after an $O(n)$ preprocessing algorithm.

Megiddo and Tamir [24] present an $O(p^2 \log^2 n)$ algorithm, assuming that $a[k, t]$ is computable in constant time for any pair $k \leq t$. (This sublinear bound is valid, for example, for the unweighted p-center problem on the line, where $a[k, t] = v_t - v_k$.) Manne and Sorevik [22] describe a bisection method based on a simple $O(n)$ feasibility test for finding an approximate solution which runs in $O(n \log(a[1, n]/\varepsilon))$, where ε is the desired precision.

As shown in Section 3.2 for our problem, after the initial $O(n)$ preprocessing $a[k, t]$ can be computed in $O(\log n)$ time for any pair, $k \leq t$. Due to the Monotonicity Property as well as the discussion in Section 3.2 of computing the $a[k, t]$, the set of values $\{a[k, t]\}$ clearly satisfy **i–iii.** Thus, in this case, an $O(n \log n)$ algorithm will follow from Frederickson's scheme. The algorithm of Megiddo and Tamir [24] will run in $O(n + p^2 \log^3 n)$ time, and will therefore dominate Frederickson's algorithm when p is relatively smaller than n, e.g., $p = O(\sqrt{n}/\log n)$. The algorithm of Olstad and Manne [27] will take $O(pn \log n)$ time.

ACKNOWLEDGMENT

We are pleased to acknowledge constructive referee comments. Their comments have led to an improved version. This research has been sponsored in part by the National Science Foundation, Grant 9908124.

REFERENCES

[1] G. Andersson, R.L. Francis, T. Normark, and M.B. Rayco, Aggregation method experimentation for large-scale network location problems, Location Sci 6 (1998), 25–39.
[2] M. Avriel, Nonlinear programming: Analysis and methods, Prentice Hall, Englewood Cliffs, NJ, 1976.
[3] P.A. Casillas, "Data aggregation and the p-median problem in continuous space," Spatial analysis and location allocation models, A. Ghosh and G. Rushton (Editors), Van Nostrand Reinhold, New York, 1987, pp. 327–244.
[4] K.R. Chelst, J.P. Schultz, and N. Sanghvi, Issues and decision aids for designing branch networks, J Retail Bank X(2) (Summer 1988), 5–17.
[5] F. Conti, F. Malucelli, S. Nicoloso, and B. Simeone, On a 2-dimensional equipartition problem, Eur J Oper Res 113 (1999), 215–231.
[6] P.D. Domich, K.L. Hoffman, R.H.F. Jackson, and M.A. McClain, Locating tax facilities: A graphics-based microcomputer optimization model, Manage Sci 37 (1991), 960–979.
[7] E. Erkut and B. Bozkaya, Analysis of aggregation errors for the p-median problem, Comput Oper Res 26 (1999), 1075–1096.
[8] R.L. Francis and T.J. Lowe, On worst-case aggregation analysis for network location problems, Ann Oper Res 40(4) (1992), 229–246.
[9] R.L. Francis and J.A. White, Facility layout and location, Prentice-Hall, Englewood Cliffs, NJ, 1974, inequality (7.21), p. 299; extensions in problem 7.25, p. 324.
[10] R.L. Francis, T.J. Lowe, and M.B. Rayco, Row–column aggregation for rectilinear p-median problems, Transportation Sci 30 (1996), 160–174.
[11] R.L. Francis, T.J. Lowe, G. Rushton, and M.B. Rayco, A synthesis of aggregation methods for multi-facility location problems: strategies for containing error, Geogr Anal 31 (1999), 67–87.
[12] R.L. Francis, T.J. Lowe, and A. Tamir, On aggregation error bounds for a class of location models, Oper Res 48 (2000), 294–307.
[13] R.L. Francis and M.B. Rayco, Asymptotically optimal aggregation for some unweighted p-center problems with rectilinear distances, Stud Locational Anal 10 (1996), 25–36.

[14] G.N. Frederickson, Optimal algorithms for partitioning trees and locating p-centers in trees, Technical Report CSD-TR-1029, Purdue University, West Lafayette, IN, 1991.
[15] G.N. Frederickson, Optimal algorithms for partitioning trees and locating p-centers in trees, Proc Second ACM-SIAM Symp Discrete Algorithms, San Francisco, 1991, pp. 168–177.
[16] A. Geoffrion, Objective function approximations in mathematical programming, Math Program 13 (1977), 23–37.
[17] G. Hardy, J.E. Littlewood, and G. Polya, Inequalities, Cambridge University Press, Cambridge, 1934.
[18] R. Hassin and A. Tamir, Improved complexity bounds for location problems on the real line, Oper Res Lett 10 (1991), 395–402.
[19] E. Hillsman and R. Rhoda, Errors in measuring distances from populations to service centers, Ann Regional Sci 12 (1978), 74–88.
[20] S. Khanna, S. Muthukrishnan, and S. Skiena, Efficient array partitioning, Proc 24th Int Coll Automata, Languages and Programming, 1997, pp. 616–626.
[21] O.L. Mangasarian, Nonlinear programming, McGraw-Hill, New York, 1969.
[22] F. Manne and T. Sorevik, Optimal partitioning of sequences, J Algorithms 19 (1995), 235–249.
[23] N. Megiddo and K. Supowit, On the complexity of some common geometric location problems, SIAM J Comput 13 (1984), 182–196.
[24] N. Megiddo and A. Tamir, An $O(p^2 \log^2 n)$ algorithm for the unweighted p-center problem on the line, Technical Report 1981, revised 1991, Department of Statistics, Tel Aviv University, Tel Aviv, Israel.
[25] N. Megiddo, A. Tamir, E. Zemel, and R. Chandrasekaran, An $O(n \log^2 n)$ algorithm for the kth longest path in a tree with applications to location problems, SIAM J Comput 10 (1981), 328–337.
[26] A.T. Murray and J.M. Gottsegen, The influence of data aggregation on the stability of p-median location model solutions, Geogr Anal 29 (1997), 200–213.
[27] B. Olstad and F. Manne, Efficient partitioning of sequences, IEEE Trans Comput 44 (1995), 1322–1326.
[28] F. Plastria, On the choice of aggregation points for continuous p-median problems: A case for the gravity centre, TOP 9 (2001), 217–242.
[29] F. Plastria, New error bounds in continuous minisum location for aggregation at the gravity centre, Stud Locational Anal 14 (2000), 101–119.
[30] M.B. Rayco, R.L. Francis, and T.J. Lowe, Error-bound driven demand point aggregation for the rectilinear distance p-center model, Location Sci 4 (1996), 213–236.
[31] M.B. Rayco, R.L. Francis, and A. Tamir, A p-center grid-positioning aggregation procedure, Comput Oper Res (Special Issue on Aggregation & Disaggregation for Operations Research) 26 (1999), 1113–1124.
[32] D.F. Rogers, R.D. Plante, R.T. Wong, and J.R. Evans, Aggregation and disaggregation techniques and methodology in optimization, Oper Res 39 (1991), 553–582.
[33] E. Zemel, Probabilistic analysis of geometric location problems, SIAM J Alg Discrete Methods 6 (1985), 189–200.
[34] P. Zhao and R. Batta, Analysis of centroid aggregation for the euclidean distance p-median problem, Eur J Oper Res 113 (1999), 147–168.