INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

WEIGHTED SUM OF ORDER p AND MINISUM LOCATION MODELS

By
HALİT ÜSTER

A Dissertation
Submitted to the School of Graduate Studies in
Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (1999) McMaster University
(Management Science/Systems) Hamilton, Ontario

TITLE: Weighted Sum of Order p and Minisum Location Models

AUTHOR: Halit Üster
M.A. Hacettepe University
B.Sc. Middle East Technical University

SUPERVISORY COMMITTEE: Prof. Robert F. Love (Chairman)
Prof. George O. Wesolowsky
Prof. Mahmut Parlar

NUMBER OF PAGES: xii, 210

Abstract

The Weighted Sum of Order p is a norm (ℓ_{bp}-norm) that we adopt to estimate distances in a given transportation network. A distance function is used to transform the point coordinate differences of two points into an estimate of the travel distance between them. Distance functions are employed in applications such as location and location-allocation models, transportation and facility layout problems, distribution management software, computational geometry, distance calculations in Geographical Information Systems, accuracy validation of actual road network distance data, service cost quotations, and random problem generation to test algorithms.

In this dissertation, we investigate the superiority of the ℓ_{bp}-norm in distance estimation accuracy over the well-known weighted ℓ_p distance. We present computational procedures for determining the parameters of the ℓ_{bp}-norm for a given transportation network and we provide theoretical and empirical results indicating its higher accuracy. In order to determine the best parameter values of the ℓ_{bp}-norm and to measure its accuracy, we utilize an estimation errors function, the sum of Squared Deviations (SD), as the goodness-of-fit criterion. We develop certain properties of the ℓ_{bp}-norm and the SD function, and using these properties we produce a computational procedure for determining the parameters of the ℓ_{bp}-norm. We apply the procedure in seventeen geographical regions and find that the ℓ_{bp}-norm models the distances with a higher accuracy than its closest competitor, the weighted ℓ_p-norm.

A new method is devised to calculate the confidence intervals for estimated distances. Using this method, the confidence intervals for estimated actual distances are developed for the ℓ_p-norm and ℓ_{bp}-norm. Our empirical study in the seventeen geographical regions indicates that better confidence intervals for the unknown actual distances are obtained with the ℓ_{bp}-norm than the ℓ_p-norm.

A distance function constitutes an important part of the objective function in continuous location models. A minisum continuous location model is concerned with the determination of one or more new facility locations in a region so that the total transportation cost between the demand points and the new facilities and also between the new facilities is minimized. Total transportation cost is given by the sum of distances weighted by their corresponding demands. Since the model should represent the real situation as accurately as possible, the accuracy of the distance function employed plays a crucial role in terms of the validity and the applicability of the locational decisions. Therefore, we incorporate the new distance function in single-facility and multi-facility continuous location models and develop generalized iterative solution procedures and fixed point optimality conditions. We also investigate the convergence properties of the iterative procedure when it is applied to the single-facility minisum location models. In order to terminate the iterative procedure used to solve the location problem a bounding method is required. We consider a method which involves the solution of a rectangular distance location problem in

each iteration, and provide its generalizations to the approximated ℓ_p and the ℓ_{bp} distance location problems.

Finally, we develop the lagrangian and the conjugate dual formulations of the most general ℓ_p-norm multi-facility minisum location model considering both linear and distance constraints, and also generalize our results to the ℓ_{bp}-norm location models.

Acknowledgements

I am particularly grateful to my supervisor, Professor Robert F. Love, who generously spent his time with me, contributed tremendously to my knowledge in the subject, and acted as a facilitator. I would like to thank Professors George O. Wesolowsky and Mahmut Parlar for their careful review of the manuscript and useful comments. I am also indebted to Natural Sciences and Engineering Research Council of Canada, McMaster University and Michael G. DeGroote School of Business for providing financial assistance during the course of this study.

This thesis is dedicated to my family for their encouragement to learn and share more, and support throughout the years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The minisum location problem has been attracting researchers from diverse fields for about three hundred years. The original problem, proposed by Fermat (1601-1665), asked for a point in continuous space which minimized the sum of Euclidean distances to three fixed points. Cavallieri (1598-1647) and Torricelli (1608-1647) are also given credit in the literature with proposing the original version of the problem. An interesting historical review can be found in (Wesolowsky, 1993). Weber (1909) generalized the problem by introducing unequal positive weights to the fixed points and hence the objective becomes the minimization of the weighted sum of Euclidean distances (Weber, 1909). Mainly in the last sixty years, several extensions and generalized forms of the Weber problem have been introduced and analyzed. A comprehensive review of continuous location theory is found in (Love et al., 1988) and (Plastria, 1995).

A single facility unconstrained minisum location model in the 2-dimensional

Euclidean plane ($\Re^2$) is stated as follows:

$$\min\ S(\mathbf{x}) = \sum_{j=1}^{n} w_j\, d(\mathbf{x}, \mathbf{a}_j), \tag{1.1}$$

where n is the number of fixed facilities; $\mathbf{a}_j = (a_{j1}, a_{j2})$, $j = 1, \ldots, n$ are the fixed facility locations; $\mathbf{x} = (x_1, x_2)$ is the sought after location of the new facility; $w_j > 0$, $j = 1, \ldots, n$ is the weight (demand) associated with the fixed facility j; and $d(\mathbf{u}, \mathbf{v})$ is some distance function used to calculate the distance between any two points $\mathbf{u}, \mathbf{v} \in \Re^2$.

Using similar notation, the multi-facility unconstrained minisum location model is given by

$$\min\ M(\mathbf{X}) = \sum_{i=1}^{m}\sum_{j=1}^{n} w_{1ij}\, d(\mathbf{x}_i, \mathbf{a}_j) + \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} w_{2ir}\, d(\mathbf{x}_i, \mathbf{x}_r), \tag{1.2}$$

where m is the number of new facilities; w_{1ij} converts the distance between new and existing facilities i and j into a cost, where $i = 1, \ldots, m$, $j = 1, \ldots, n$; and w_{2ir} converts the distance between two new facilities i and r into a cost, where $i = 1, \ldots, m-1$, $r = i+1, \ldots, m$.

As it is readily seen in formulations (1.1) and (1.2), distance predicting functions are employed to construct the objective function of a continuous location model. Since the model should represent the real situation as closely as possible, the accuracy of the distance predicting function employed plays a crucial role in terms of the validity and the applicability of the locational decisions. Love and Morris (1972; 1979) present

several distance predicting functions which are mostly norms weighted by an inflation factor to account for hills, bends and other forms of "noise" in the transportation network. A significant conclusion of their study is that an empirical distance function should be tailored to a given region whenever a premium is placed on accuracy. This result is based on statistical analyses showing that the weighted ℓ_p-norm outperforms both the weighted Euclidean and the weighted rectangular norms.

The other ingredient in the models (1.1) and (1.2) is demand. The representation of demand is also an area of research in location theory and distribution system design. A discussion on this topic is given by Drezner (1995). The method described by Drezner uses a *distance correction approach* where the distance between a demand point and a facility is modified in such a way that it represents the average distance between the facility and all points in the sub-area replaced by the demand point. Other studies on average distance functions include Vaughan (1984), Stone (1991), Koshizuka and Kurita (1991) and Miller (1996). Note that in all these studies on calculating average distances between geometrical entities in a region, the existence of a good continuous distance function is assumed to start with.

In Section 1.1 we present a review of the literature, and in Section 1.2 we define the weighted sum of order p and give its useful properties. We conclude this chapter with the objectives and the outline of this dissertation.

1.1 Literature

In this section, we present a review of the relevant literature on distance predicting functions, goodness-of-fit criteria used in distance modelling and continuous location models.

1.1.1 Distance Predicting Functions

Besides being a major component of location models, distance predicting functions are utilized in several applications. Some of these uses are discussed here. For validating the accuracy of actual road network distance data, distance predicting functions can be used as suggested by Ginsburgh and Hansen (1974). To determine the optimal mix of trunking and tramping of a truck transportation network for the movement of finished goods and raw materials among national distribution centers, regional depots, and producers, a distance predicting function was utilized by Westwood (1977) to obtain estimates of the travel distances between possible links in the network. In some distribution problems for which only the demands and the general location of customers are known, a distance predicting function may be employed to calculate a predicted travel distance between the depot and the general area (Eilon et al., 1971).

Distance predicting functions are used in models that determine the response time of emergency vehicles to calls such as the model proposed by Kolesar et al. (1975) for calculating the response time of fire engines to fires.

Klein (1988) suggests that distance predicting functions which reflect the nature

of a geographical region's road network should be used for constructing Voronoi diagrams of the region. A Voronoi diagram subdivides a region into a number of subregions with each subregion being formed around a point belonging to a set of points. For example, the set of points may be the region's police stations, fire halls, or hospitals. Once the location of a query point is determined, the appropriate point of the set is notified to respond to the call by looking at the Voronoi diagram.

Distance predicting functions are being utilized in the truck dispatching software packages Roadnet (Roadnet-Technologies, 1993) and TruckStops2 (MicroAnalytics, 1993) since they are much more efficient and comprehensive to use in practice than attempting to assemble large files of distance data.

As Star and Estes (1990) state, distance measurements are of value in many geographic circumstances. Some of these circumstances are planning an irrigation channel between a pond and a field, locating a site for a fire tower in a forest, and calculating the distances among different geographical regions. To calculate distance measurements, a distance predicting function may be incorporated into a Geographical Information System (GIS).

Distance functions are also used in cluster analysis. Murray and Estivill-Castro examine the effects of the distance function choice on forming cluster regions (1998).

Norms are usually employed as the basis for distance predicting functions in continuous location models. The main reason for this lies in the basic properties of a

norm which makes it well-suited for distance predictions. Any function $h : \Re^N \rightarrow \Re^1$ satisfying the following properties is a norm (Rockafellar, 1970):

1. $h(\mathbf{x}) \geq 0, \ \forall \mathbf{x} \neq \mathbf{0}$, (nonnegativity)

2. $h(\mathbf{x}) = 0 \Leftrightarrow \mathbf{x} = 0$, (stationarity)

3. $h(\lambda \mathbf{x}) = \lambda h(\mathbf{x}), \ \forall \mathbf{x}, \forall \lambda > 0$, (homogeneity)

4. $h(-\mathbf{x}) = h(\mathbf{x}), \ \forall \mathbf{x}$, (symmetry)

5. $h(\mathbf{x}_1 + \mathbf{x}_2) \leq h(\mathbf{x}_1) + h(\mathbf{x}_2), \ \forall \mathbf{x}_1, \ \forall \mathbf{x}_2$ (triangle inequality).

The characteristics nonnegativity, stationarity, triangle inequality, and symmetry are typical for actual distances. However, it should be noted that for applications involving, for example, one-way streets (Drezner and Wesolowsky, 1995) or movement up and downhill (Hodgson et al., 1987), the symmetry property does not hold for obvious reasons. Since norms are convex functions (Rockafellar, 1970), incorporating a norm in the objective function of a continuous location problem provides the useful property of convexity in the optimization model. Additionally, norms as distance predicting functions can be employed for generating random problem instances in discrete location or location-allocation models since they satisfy the triangle inequality by nature.

Love and Morris (1972; 1979; 1988) introduced several *round norms* as distance predicting functions. Among them, the weighted ℓ_p-norm ($k\ell_p$-norm) has proved to

be relatively easy to fit to a geographical region with excellent predictive properties. Let $\mathbf{x} = (x_1, x_2)$, and $\mathbf{y} = (y_1, y_2)$ be any two points in the plane. Then the $k\ell_p$-norm is given by

$$k\ell_p(\mathbf{x}, \mathbf{y}) = k\ (|x_1 - y_1|^p + |x_2 - y_2|^p)^{1/p}, \quad k > 0, \quad p \geq 1. \tag{1.3}$$

The weighted Euclidean and rectangular distance functions are special cases of the weighted ℓ_p-norm where $p = 2$ and $p = 1$, respectively.

Ward and Wendell (1980; 1985) have introduced the concept of utilizing *block norms* as distance predictors. The difference between the round and block norms can be observed in the shape of their respective unit-balls. While the unit-ball of a round norm does not contain any flat spots, the contours of a block norm's unit-ball are polygons (Thisse et al., 1984). In mathematical terms, this fact can be stated as follows: For all distinct points $\mathbf{x}_1$, $\mathbf{x}_2$ on the unit contour of norm h, and $\lambda \in (0, 1)$ we have $h(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) < 1$ if and only if h is a round norm. The statement holds for a block norm if the strict inequality sign is replaced by the "$\leq$" sign.

A study by Love and Walker (1994) shows that although marginally better results can be obtained by using a block norm with eight parameters than by using the two parameter $k\ell_p$-norm, the computational cost of fitting the block norms can be prohibitive. The authors also observe that increasing the number of parameters of a block norm does not ensure that it becomes more accurate than the weighted ℓ_p-norm. Additionally, the original studies by Love and Morris (1972; 1979) show that

the $k\ell_p$-norm usually gives much superior results compared to other simpler norms such as the weighted Euclidean or weighted rectangular norms.

Love and Morris (1972) introduce the concept of axis rotation in their study on the road network in Milwaukee, Wisconsin. Brimberg, Love and Walker (1995) investigate this concept in detail and conclude that a reference axis rotation chosen to align with the underlying pattern of the transportation network improves the accuracy of distance predictions. Huriot and Perreur (1973) also discuss axis rotation and apply it in a study of the nine largest Swiss cities. Incorporating the axis rotation angle θ in the weighted ℓ_p-norm model (1.3) we obtain

$$k\ell_{p,\theta}(\mathbf{x}', \mathbf{y}') = k\left(|x'_1 - y'_1|^p + |x'_2 - y'_2|^p\right)^{1/p}, \quad k > 0, \quad p \geq 1 \tag{1.4}$$

where

$$\mathbf{x}' = (x'_1, x'_2), \quad \mathbf{y}' = (y'_1, y'_2), \quad \theta \in [0, \pi/2] \quad \text{and}$$

$$\begin{pmatrix} x'_1 & x'_2 \\ y'_1 & y'_2 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

In general, each transportation network has its own characteristic underlying road pattern. Therefore, if we require highly accurate distance predictions, we should know the best parameter values of a chosen distance function for the area under consideration. In order to determine the best parameter values of the distance predicting function for a region and also to measure its accuracy, we need a goodness-of-fit criterion. There are several possible goodness-of-fit criteria. We next present a review of this literature.

1.1.2 Goodness-of-Fit Criteria

The optimum parameter values of the distance predicting function are determined so that a criterion value is minimized. The criterion usually measures the aggregate amount of error generated by a particular distance function with known parameter values. Three criteria have been used to fit the parameters of a distance function to a set of data representing a region of interest: Sum of Absolute Deviations (AD_f); Sum of Squared Deviations (SD_f); Sum of Normalized Absolute Deviations (NAD_f). Let $d_f(\mathbf{a}_i, \mathbf{a}_j)$ be the predicted distance between points $\mathbf{a}_i$ and $\mathbf{a}_j$ by using the predicting function f. $A(\mathbf{a}_i, \mathbf{a}_j)$ is the actual distance between $\mathbf{a}_i$ and $\mathbf{a}_j$, and n is the number of points in the data set. Then the mathematical expressions for the goodness-of-fit criteria are as follows:

$$AD_f = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} |d_f(\mathbf{a}_i, \mathbf{a}_j) - A(\mathbf{a}_i, \mathbf{a}_j)| \tag{1.5}$$

$$SD_f = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{(d_f(\mathbf{a}_i, \mathbf{a}_j) - A(\mathbf{a}_i, \mathbf{a}_j))^2}{A(\mathbf{a}_i, \mathbf{a}_j)} \tag{1.6}$$

$$NAD_f = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{|d_f(\mathbf{a}_i, \mathbf{a}_j) - A(\mathbf{a}_i, \mathbf{a}_j)|}{A(\mathbf{a}_i, \mathbf{a}_j)} \tag{1.7}$$

The AD_f and SD_f criteria were introduced by Love and Morris (1972), and the NAD_f criterion by Love and Walker (1991). Several applications of the criteria are given by Berens (1988), Berens and Körling (1985), Brimberg, Dowling and Love (1996), Brimberg, Love and Walker (1995), Love and Morris (1972; 1979), Love, Walker and Tiku (1995), and Ward and Wendell (1980; 1985). In addition, AD_f and SD_f have

been used by Love and Morris (1972; 1979) to develop tests for statistically comparing the accuracy of different distance predicting functions.

The first criterion, AD_f, is the minimization of the sum of absolute deviations. Since the terms in AD_f are not the weighted ones but only the absolute errors for each pair, it has been described as a criterion which should estimate long distances more accurately than short distances. The second criterion, SD_f, is the minimization of the sum of squared deviations where each squared error term is weighted by $1/A(\mathbf{a}_i, \mathbf{a}_j)$. Squared deviations and the division by actual distance provide the criterion with certain desirable statistical properties as suggested by Love and Morris (1972; 1979). However, the assumption has still been made that the difference in the accuracy of predictions involving long and short distances in a region will favor the long distances (Berens, 1988; Berens and Körling, 1985; Love and Morris, 1972; Love and Morris, 1979; Ward and Wendell, 1980; Ward and Wendell, 1985). The last criterion, NAD_f , is relatively new in the literature. It has been utilized by Love and Walker (1991); Brimberg, Love, Walker (1995); Love, Walker and Tiku (1995); and Brimberg, Dowling and Love (1996). With the NAD_f criterion, a sum of normalized absolute deviations is minimized and the basic premise is that equal accuracy in predicting long and short distances in a region will result. Normalization is realized by dividing the absolute deviation by the actual distance between each pair. In this way both long and short distances are treated on the same relative basis.

Besides their above-mentioned structures, the three criteria also differ from each other by the computational procedures performed to determine the optimal parameter values of the distance predicting function. The computational procedures for fitting the AD_f and the SD_f criteria are given by Brimberg and Love (1991). For the NAD_f criterion Love and Walker (1991) show that the computational procedure is identical to that of AD_f. In general, the best θ and p values are determined by using an incremental search procedure and a four-stage incremental search procedure, respectively. In order to find the best k value some properties of the criteria are used. It is known that AD_f is a convex function of k, and SD_f is a strictly convex function of k (Brimberg and Love, 1991). NAD_f was shown to be a convex function of k by Love and Walker (1991). Therefore, when using the AD_f and NAD_f criteria it is necessary to employ an algorithm to find the optimal k for a given (θ, p) pair. The optimal k for the SD_f criteria is calculated with a simple closed-form formula derived by Brimberg and Love (1991). Using a closed-form formula to find the best value of parameter k makes the application of the SD_f criterion computationally more efficient than employing either the AD_f or the NAD_f criterion.

In order to model the parameters of the $k\ell_p$-norm Love and Walker (1993) collected sample data from seventeen geographical regions. The sample data for each geographical region included 15 point locations based on random selection of *point coordinates* on the map. These 15 points provided 105 actual road travel distances

to be modelled by the distance predicting function $k\ell_p$ using each criterion. The actual distance data and point coordinates from the seventeen geographical regions are presented in Love and Walker (1993). The empirical parameter values for the $k\ell_p$-norm and the corresponding minimum criterion values for seventeen geographical regions are computed by Love and Walker (1994) for the AD_f and SD_f criteria, and by Love and Walker (1991) for the NAD_f criterion.

1.1.3 Minisum Location Models

The early formulations of the single-facility minisum location problem in continuous space (1.1) adopted Euclidean distances. In a mathematical appendix to Weber's book (1909), G. Pick proposed a solution method using a mechanical analogue device called the Varignon Frame. Using a board with holes drilled in it to represent the locations of the fixed facilities, a string is passed through each hole and a weight corresponding to the transportation cost or quantity to be shipped is attached to the lower end of each string. On the top of the board, all the strings are tied together forming a common knot. When the system is released, under the assumption of no friction, it will be balanced at such a position that the common knot is at the optimum facility location. The one-point iteration method introduced by Weiszfeld (1937) is the most popular technique to solve this problem. Although the Weiszfeld procedure was first published in 1937, it remained virtually unknown for many years. The method was then rediscovered independently by Miehle (1958), Kuhn and Kuenne (1962)

and Cooper (1963). The Weiszfeld iterative procedure simply utilizes the convexity of the objective function, and thus calculates the first order derivatives. Since it is impossible to express the unknown variables in the form of equations, the first order derivatives cannot be solved directly. Instead an iteration function is obtained by using these derivatives. However, the first order derivatives are not defined at the existing facility locations. This poses a problem in the iterations of the Weiszfeld procedure if one of the existing facilities is actually an optimal solution. Love (1969) introduced a method of "fitted functions" to solve the difficulty associated with these discontinuities in the derivatives. On the other hand, using a hyperbolic approximation in the objective function also eliminates this problem (Wesolowsky and Love, 1972; Morris and Verdini, 1979). For squared Euclidean distances the minisum model is solved directly by using the first order derivatives. The optimal solution for this case corresponds to the center of gravity of the system. Considering the Varignon Frame again, if the strings attached to weights were fixed at the holes, i.e., not tied together, then the solution is at a point in which the system is hung to keep the board perfectly horizontal, assuming that the board is weight-less. If rectangular distances are used, the minisum problem becomes very easy to solve. The objective function can be decomposed into two independent one-dimensional problems and then each is solved by finding the median point of the weighted fixed facility locations. The Weiszfeld procedure has been generalized to solve the minisum

problem with ℓ_p distances. Several properties for that instance of the problem can be found in (Love et al., 1988).

If the number of facilities to be located is two or more with known interactions (weights) between them, we obtain the multi-facility minisum location problem (1.2). When rectangular distances are assumed, a linear programming approach is used after decomposing the problem as in the single facility case (Wersan et al., 1962; Wesolowsky and Love, 1971). Another method which is a modified edge descent procedure is given by Juel and Love in (1976). For the general ℓ_p distance problem, the Weiszfeld procedure is generalized for the single-facility case by Morris and Verdini (1979).

In terms of characterizing the optimal solution the following localization results constitute an important part of the minisum location research. In single-facility location models with Euclidean distances on $\Re^N$ the optimal solution must be in the convex hull of the existing facilities (Kuhn, 1973). For the Euclidean distance multi-facility location problem the same holds true for all the new facilities only for problems occurring on $\Re^2$ (Cabot and Francis, 1972). When ℓ_p distances are used on $\Re^2$ with $p > 1$, then again the optimal solutions to single- and multi-facility problems must be within the convex hull of existing facilities (Juel and Love, 1983). Brimberg and Love (1995a) extended the localization results of Juel and Love to round norms. In general, for any single norm on $\Re^2$, at least one optimal solution belongs to the

convex hull of existing facilities (Wendell and Hurter, 1985; Hansen et al., 1980). In higher dimensional spaces ($N > 2$) this generalized localization property holds only if Euclidean distance is used (Plastria, 1984). In rectangular distance problems on $\Re^2$ at least one optimal solution belongs to the smaller rectangular hull of the existing facilities, and furthermore, attention can be restricted to the intersection points contained in this hull (Love and Morris, 1975a).

1.2 Weighted Sum of Order p

A Weighted Sum of Order p is a generalization of the well-known ℓ_p-norm used in predicting distances in a transportation network. With the weighted sum of order p one introduces non-symmetric distance irregularities or costs along the axis directions.

A weighted sum of order p is defined as follows (Hardy et al., 1952, Section 2.10):

$$T(\mathbf{y};\mathbf{b},p) = \left[\sum_{i=1}^{K} b_i y_i^p\right]^{1/p}, \quad p \neq 0, \tag{1.8}$$

where

$$\begin{aligned} \mathbf{y} &= (y_1,\ldots,y_K), \quad y_i > 0, \quad i = 1,\ldots,K, \\ \mathbf{b} &= (b_1,\ldots,b_K), \quad b_i > 0, \quad i = 1,\ldots,K, \end{aligned}$$

The vector $\mathbf{b}$ and the scalar p can be considered as a set of parameter values. If all the weights are equal to one, then T becomes the well-known ordinary sum of order p. Note that the function $T(\mathbf{y};\mathbf{b},p)$ has the form of a generalized ℓ_p distance which

is given in 2-dimensional form as follows:

$$\ell_{bp}(\mathbf{x},\mathbf{y}) = [b_1\,|x_1 - y_1|^p + b_2\,|x_2 - y_2|^p]^{1/p},\ \ b_1, b_2 > 0,\ \ p \geq 1, \tag{1.9}$$

where $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2) \in \Re^2$ and $\ell_{bp}(\mathbf{x},\mathbf{y})$ estimates the distance between points $\mathbf{x}$ and $\mathbf{y}$. The weights b_1 and b_2 can be used to represent non-symmetric distance irregularities along the axis directions in a location model. This is in effect adding a single parameter to the $k\ell_p$-norm since one of the two weights in the weighted sum of order p function replaces the k in the $k\ell_p$-norm. $\ell_{bp}(\mathbf{x},\mathbf{y})$ can also be expressed in a similar way to the weighted ℓ_p-norm with a common stretch factor as follows:

$$\ell_{b'p}(\mathbf{x},\mathbf{y}) = b_1'\,[|x_1 - y_1|^p + b_2'\,|x_2 - y_2|^p]^{1/p},\ \ b_1', b_2' > 0,\ \ p \geq 1, \tag{1.10}$$

where $b_1' = b_1^{1/p}$ and $b_2' = b_2/b_1$. In our work, we will use the form given in (1.9).

Besides its expected higher accuracy in predicting distances in a geographical region, the weighted sum of order p also seems well-suited to some specific applications. Foulds and Hamacher (1990) present an example in the field of robotics when the movement is restricted to two directions only, generated by one motor each. In such an environment and under the assumption that the motors are not allowed to work simultaneously and their constant speeds are equal, the ℓ_1-norm can effectively be used to determine the minimal time necessary to move from one position in the plane to another. If the assumption of equal speeds of motors is relaxed, then the ℓ_{bp}-norm with $p = 1$ and b_1, b_2 proportional to the speeds in the corresponding

directions is a suitable function to determine the required minimal travel time. Similar examples are harbour cranes used to move containers around and plotters used to plot engineering drawings. Love and Dowling (1985) examine the effect of "doubling back" by fitting the weighted ℓ_p-norm to a sample of layout patterns which are basically rectangular. They find that while the best value of p stays close to 1, the extra travel distance caused by doubling back is captured by the parameter k. With the weighted ℓ_p-norm it is assumed that the effect of doubling back is equal in both horizontal and vertical travel directions. This may not be the case in some types of layouts with parallel bays along either the horizontal or vertical axes.

As noted above, the introduction of reference axis rotation brings more accuracy in predicting distances (Brimberg et al., 1995). The reference axes are rotated so that they correspond to the underlying pattern of the road network. Therefore, we include axis rotation, θ, in the above distance prediction model as follows:

$$\ell_{bp\theta}(\mathbf{x}', \mathbf{y}') = \left[\sum_{i=1}^{2} b_i \, |x'_i - y'_i|^p \right]^{1/p}, \quad b_1, b_2 > 0, \quad p \geq 1 \tag{1.11}$$

where

$$\mathbf{x}' = (x'_1, x'_2), \quad \mathbf{y}' = (y'_1, y'_2), \quad \theta \in [0, \pi/2] \quad \text{and}$$

$$\begin{pmatrix} x'_1 & x'_2 \\ y'_1 & y'_2 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{1.12}$$

In the following section we give some properties of the $\ell_{bp}(\mathbf{x})$.

1.2.1 Properties of $\ell_{bp}(\mathbf{x})$

In this section, we first show that an equivalent ℓ_p-norm can be obtained from $\ell_{bp}(\mathbf{x})$ via proper scaling of $\mathbf{x}$, provided that all the parameters, b_1, b_2 and p are known. Therefore, $\ell_{bp}(\mathbf{x})$ is also a norm. Then, we review the properties of the ℓ_{bp}-norm as a function of its parameter p, and examine its behaviour as a function of its parameters b_1 and b_2.

Property 1.2.1 *An equivalent ℓ_p-norm can be obtained from the $\ell_{bp}(\mathbf{x}, \mathbf{y})$ by scaling the horizontal and vertical components of $\mathbf{x}$ and $\mathbf{y}$ by $b_1^{1/p}$ and $b_2^{1/p}$, respectively.*

Proof We rewrite the ℓ_{bp}-norm (1.9) as follows:

$$\ell_{bp}(\mathbf{z}) = [(b_1^{1/p}\,|z_1|)^p + (b_2^{1/p}\,|z_2|)^p]^{1/p},$$

$$\text{where } \mathbf{z} = (z_1, z_2) \text{ and } z_t = |x_t - y_t|,\ t = 1, 2.$$

Taking $|z_t'| = {b_t}^{1/p}\,|z_t|,\ t = 1, 2$, we obtain

$$\ell_{bp}(\mathbf{z}) = (|z_1'|^p + |z_2'|^p)^{1/p}.$$

Notice that $\ell_{bp}(\mathbf{z})$ is in the form of the ℓ_p-norm, i.e., we have

$$\ell_{bp}(\mathbf{z}) = \ell_p(\mathbf{z}'),$$

and the result follows. □

The next property is useful in the context of using $\ell_{bp}(\mathbf{x})$ in continuous location models. It readily follows from Property 1.2.1.

Property 1.2.2 *The $\ell_{bp}(\mathbf{x})$ function is a norm and thus a convex function of* $\mathbf{x}$, *provided that* $p \geq 1$.

We now state a corollary given in Brimberg and Love (1995b) and a property regarding the behaviour of the ℓ_{bp}-norm with respect to its parameters b_1 and b_2. These findings are important when the use of the ℓ_{bp}-norm for predicting distances is considered.

Corollary 1.2.1 *The ℓ_{bp}-norm is a decreasing function of $p > 0$ for given weights b_1, b_2 and all positive $|x_i - y_i|$, if and only if b_1, $b_2 \geq 1$. Furthermore, the ℓ_{bp}-norm is strictly convex in p under these conditions.*

In predicting distances we would expect to see the coefficients b_1 and b_2 greater than one (Brimberg, 1989). Utilizing the asymptotic behaviour of the ℓ_{bp}-norm for this case (Brimberg and Love, 1995b) we observe the following limiting cases:

$$\lim_{p \to \infty} \ell_{bp} = \ell_\infty, \quad \lim_{p \to 0^+} \ell_{bp} = \infty$$

Property 1.2.3 *The ℓ_{bp}-norm is an increasing concave function of its parameters b_1 and b_2, provided that $p \geq 1$ and all $|x_i - y_i|$ are nonzero.*

Proof In order to check its convexity properties we need to calculate the following terms for the Hessian H_{b_1,b_2}

$$\frac{\partial \ell_{bp}}{\partial b_1} = \frac{1}{p}[b_1|x_1 - y_1|^p + b_2|x_2 - y_2|^p]^{\frac{1}{p}-1} |x_1 - y_1|^p,$$

$$\begin{aligned}
\frac{\partial \ell_{bp}}{\partial b_2} &= \frac{1}{p}[b_1|x_1-y_1|^p + b_2|x_2-y_2|^p]^{\frac{1}{p}-1}\ |x_2-y_2|^p, \\
\frac{\partial^2 \ell_{bp}}{\partial {b_1}^2} &= \frac{1-p}{p^2}[b_1|x_1-y_1|^p + b_2|x_2-y_2|^p]^{\frac{1}{p}-2}\ |x_1-y_1|^{2p}, \\
\frac{\partial^2 \ell_{bp}}{\partial {b_2}^2} &= \frac{1-p}{p^2}[b_1|x_1-y_1|^p + b_2|x_2-y_2|^p]^{\frac{1}{p}-2}\ |x_2-y_2|^{2p}, \\
\frac{\partial^2 \ell_{bp}}{\partial b_1 \partial b_2} = \frac{\partial^2 \ell_{bp}}{\partial b_2 \partial b_1} &= \frac{1-p}{p^2}[b_1|x_1-y_1|^p + b_2|x_2-y_2|^p]^{\frac{1}{p}-2}\ |x_1-y_1|^p\ |x_2-y_2|^p,
\end{aligned}$$

and

$$|H_{b_1,b_2}| = \frac{\partial^2 \ell_{bp}}{\partial {b_1}^2}\frac{\partial^2 \ell_{bp}}{\partial {b_2}^2} - \left(\frac{\partial^2 \ell_{bp}}{\partial b_1 \partial b_2}\right)^2.$$

If $p \geq 1$ then $\partial^2 \ell_{bp}/\partial {b_1}^2 \leq 0$, $\partial^2 \ell_{bp}/\partial {b_2}^2 \leq 0$ and $|H_{b_1,b_2}| = 0$.

We conclude that, under the conditions $p \geq 1$ and $|x_i - y_i| \neq 0$, the ℓ_{bp}-norm is concave in its parameters b_1 and b_2. Furthermore, since $\partial \ell_{bp}/\partial b_1 > 0$ and $\partial \ell_{bp}/\partial b_2 > 0$ under these conditions, the ℓ_{bp}-norm is also an increasing function of parameters b_1 and b_2. □

1.3 Objectives and Outline

In the preceding sections we provided a general view of distance predicting functions and continuous location models. We discussed the importance of distance functions as a component of continuous location models, as well as several other applications. We also introduced the distance function of our interest, the weighted sum of order p (ℓ_{bp}-norm), and provided some of its mathematical properties.

Although it has been suggested by Brimberg and Love (1995b), the use of the

weighted sum of order p as distance predicting function has not been studied in the literature. Since it is a generalized form of the weighted ℓ_p-norm, the weighted sum of order p should provide greater accuracy in estimating distances on a transportation network. Therefore, in the first part of this dissertation, our objective is to investigate certain properties and the performance of the weighted sum of order p when it is employed as a distance predicting function. The following first three chapters are devoted to this purpose, and can be outlined as follows:

In Chapter 2, we define the directional bias function of the ℓ_{bp}-norm. The properties of the directional bias function and the unit balls for the ℓ_{bp}-norm are of theoretical and practical interest. We investigate several properties of the directional bias function and compare them with the properties of the ℓ_p-norm's directional bias function and the unit balls. We find that the ℓ_{bp}-norm is better at capturing the nonlinearity in a transportation network than the weighted ℓ_p-norm. It is also shown that, in contrast to the weighted ℓ_p-norm, where the optimal parameter p value is confined to the interval $[1, 2]$, for the ℓ_{bp}-norm the parameter p can have an optimal value greater than 2.

In Chapter 3, we derive some mathematical properties of the goodness-of-fit criterion SD_f involving the ℓ_{bp}-norm as the distance function. (In the following chapters we denote SD_f by SD for simplicity). These properties are used to develop computational procedures to determine the best-fitting parameter values of

the ℓ_{bp}-norm for a transportation network. We apply the new norm to seventeen geographical regions and find significant improvements in the accuracy of distance estimations over the weighted ℓ_p-norm.

In Chapter 4, we analyze the statistical properties of estimation errors generated by the ℓ_{bp}-norm and the SD criterion in the seventeen geographical regions. We use these properties to develop a new error-related random variable which is homoscedastic and normally distributed. We develop confidence intervals for unknown actual distances by using the new random variable, and give some example confidence interval calculations. We also provide a comparison of the distance predicting accuracy of the weighted ℓ_p-norm and ℓ_{bp}-norm based on their confidence intervals for the seventeen geographical regions, and empirically show that better confidence intervals are obtained with the ℓ_{bp}-norm.

In the second part of the dissertation, we focus on the issues regarding continuous location models. In Chapter 5, we discuss two approaches to solve the ℓ_{bp}-norm single- and multi-facility minisum location models, and provide modified Weiszfeld procedures for the solution of these problems. We develop and examine the properties of a generalized bounding method for the Weiszfeld procedure used to solve the hyperbolically approximated ℓ_p-norm single- and multi-facility minisum location problems. The bounding method involves the solution of a rectangular distance location problem at each iteration. We also generalize the bounding method to the

ℓ_{bp}-norm location problems. Next, we derive the fixed point optimality condition for the ℓ_{bp}-norm single-facility problem by utilizing the well-known conditions for the ℓ_p-norm problem. Lastly, we investigate the convergence properties of the Weiszfeld procedure when it is applied to the approximated ℓ_p-norm single-facility location problem where $p > 2$. The convergence of the iterative procedure for $p \in [1, 2]$ is shown by Morris (1981) for the approximated ℓ_p-norm problem and by Brimberg and Love (1992; 1993b) for the unapproximated ℓ_p-norm problem. We show that the convergence for $p > 2$ can be obtained by introducing a step size factor to the iterative procedure. We also provide some numerical test results.

In Chapter 6, we consider the ℓ_p-norm multi-facility minisum location problem with linear and distance constraints, and develop the lagrangian and conjugate dual formulations for this problem. The model that we consider represents the most general location model in which the dual formulations are not found in the literature. We find that, because of its linear objective function the lagrangian dual is more useful for this case. However, for the ℓ_p-norm, we show that the conjugate dual problem can be reduced to a problem identical to the lagrangian dual problem. We also provide the dual formulations of the same problem with the ℓ_{bp}-norm and give a numerical example for this case.

In Chapter 7, we summarize our conclusions and discuss possible directions for future research.

Chapter 2

Directional Bias of the ℓ_{bp}-norm

In this chapter we investigate some useful properties of the directional bias function of the ℓ_{bp}-norm. The way that the ℓ_{bp}-norm captures the rectangularity and the nonlinearity inherent in a transportation network is particularly important in the process of distance modelling and later in interpreting the parameters b_1, b_2 and p. Using these properties we compare the ℓ_{bp}-norm and the well-known weighted ℓ_p-norm in terms of their ability to explain the underlying pattern in a transportation network and the procedures used to identify the best parameter values of each norm. Traditionally, the directional bias of norms is illustrated and compared by means of the unit balls associated with them. For this reason we examine the unit ball of the ℓ_{bp}-norm as well as its directional bias function.

In the next section we define the directional bias function for the ℓ_{bp}-norm and present its properties. The last section is devoted to the implications of the results on distance modelling procedures.

2.1 Properties of the Directional Bias Function

A directional bias function for any norm h on $\Re^2$ is defined as follows (Brimberg and Love, 1993a):

$$r(\theta) = \frac{h(\mathbf{x})}{\ell_2(\mathbf{x})}, \quad \mathbf{x} \neq 0, \quad \theta = \arctan\left(\frac{x_2}{x_1}\right). \tag{2.1}$$

Let $\mathbf{u}$ and $\mathbf{v}$ be two points in $\Re^2$ such that $h(\mathbf{u}) > h(\mathbf{v})$ and the same Euclidean distance is to be covered in both directions, i.e., $\ell_2(\mathbf{u}) = \ell_2(\mathbf{v})$. Then obviously $r(\theta_u) > r(\theta_v)$ and we say that the difficulty of travel in the θ_u direction is greater than the difficulty of travel in the θ_v direction. Employing the ℓ_{bp}-norm as a distance predicting function we define the directional bias function as

$$\begin{aligned} r_{b_1,b_2,p}(\theta) = \frac{\ell_{bp}(\mathbf{x})}{\ell_2(\mathbf{x})} &= \frac{(b_1|x_1|^p + b_2|x_2|^p)^{1/p}}{\ell_2(\mathbf{x})} \\ &= (b_1|\cos\theta|^p + b_2|\sin\theta|^p)^{1/p} \end{aligned} \tag{2.2}$$

where θ is the angle specifying the vector $\mathbf{x} \in \Re^2$. We adopt the notation $r(\theta)$ to replace $r_{b_1,b_2,p}(\theta)$ in the rest of the chapter.

Property 2.1.1 *The directional bias is the same for two θ values 90° apart where the b_1 and b_2 values have been exchanged, p being the same.*

Proof

$$\begin{aligned} r\left(\theta + \frac{\pi}{2}\right) &= \left(b_1\left|\cos\left(\theta + \frac{\pi}{2}\right)\right|^p + b_2\left|\sin\left(\theta + \frac{\pi}{2}\right)\right|^p\right)^{1/p} \\ &= (b_1|-\sin\theta|^p + b_2|\cos\theta|^p)^{1/p} \\ &= (b_2|\cos\theta|^p + b_1|\sin\theta|^p)^{1/p} \end{aligned}$$

so if we switch the b_1 and b_2 values, we obtain the same directional bias when θ is changed by $\pi/2$ with p being the same. □

Property 2.1.2 *$r(\theta)$ is periodic with period π.*

Proof

$$\begin{aligned} r(\theta+\pi) &= (b_1|\cos(\theta+\pi)|^p + b_2|\sin(\theta+\pi)|^p)^{1/p} \\ &= (b_1|-\cos\theta|^p + b_2|-\sin\theta|^p)^{1/p} \\ &= (b_1|\cos\theta|^p + b_2|\sin\theta|^p)^{1/p} \\ &= r(\theta) \end{aligned}$$

and the result follows. □

Property 2.1.3 *For any real w*

$$r\left(\frac{\pi}{2}-w\right) = r\left(\frac{\pi}{2}+w\right)$$

i.e. $r(\theta)$ is the mirror image of itself about the line $\theta = \pi/2$.

Proof It follows from observing the equalities

$$\left|\sin\left(\frac{\pi}{2}-w\right)\right| = \left|\sin\left(\frac{\pi}{2}+w\right)\right| \quad \text{and} \quad \left|\cos\left(\frac{\pi}{2}-w\right)\right| = \left|\cos\left(\frac{\pi}{2}+w\right)\right|.$$

Thus we need to consider θ only in the interval $[0, \pi/2]$. □

Property 2.1.4 *$r(\theta)$ is differentiable and continuous for $\theta \in [0, \pi/2]$.*

Proof We check the equality of the right- and left-hand limits of $dr(\theta'+\epsilon)/\,d\theta$ as $\epsilon \to 0, \quad \theta' \in [0, \pi/2]$

$$\left.\frac{dr(\theta)}{d\theta}\right|_{\theta=\theta'+\epsilon} = \frac{\sin 2(\theta'+\epsilon)(b_2 \sin^{p-2}(\theta'+\epsilon) - b_1 \cos^{p-2}(\theta'+\epsilon))}{2(b_1 \cos^p(\theta'+\epsilon) + b_2 \sin^p(\theta'+\epsilon))^{(p-1)/p}}.$$

Hence

$$\lim_{\epsilon \to 0^+} \left.\frac{dr(\theta)}{d\theta}\right|_{\theta=\theta'+\epsilon} = \frac{\sin 2\theta'(b_2 \sin^{p-2}\theta' - b_1 \cos^{p-2}\theta')}{2(r(\theta'))^{p-1}},$$

and

$$\lim_{\epsilon \to 0^-} \left.\frac{dr(\theta)}{d\theta}\right|_{\theta=\theta'+\epsilon} = \frac{\sin 2\theta'(b_2 \sin^{p-2}\theta' - b_1 \cos^{p-2}\theta')}{2(r(\theta'))^{p-1}}.$$

The equality of these limits shows that $r(\theta)$ is differentiable for $\theta \in [0, \pi/2]$. The continuity of $r(\theta)$ follows from the Theorems 4.7 and 4.9 given by Rudin (1976). □

In order to further explore the shape of the $r(\theta)$ function (Properties 2.1.5, 2.1.6, 2.1.7, 2.1.8) we use the first- and the second-order derivatives. We next derive these derivatives. Note that $\sin(\theta), \cos(\theta) \geq 0$ for $\theta \in [0, \pi/2]$.

$$\begin{aligned}\frac{dr(\theta)}{d\theta} &= \frac{1}{(r(\theta))^{p-1}}(b_2 \cos\theta(\sin\theta)^{p-1} - b_1 \sin\theta(\cos\theta)^{p-1}) \\ &= \frac{\sin 2\theta}{2(r(\theta))^{p-1}}(b_2(\sin\theta)^{p-2} - b_1(\cos\theta)^{p-2}). \qquad (2.3)\end{aligned}$$

Observe that $dr(\theta)/d\theta = 0$ for $\theta = 0$ and $\theta = \pi/2$, and also for $\theta = \pi/4$ if $b_1 = b_2$.

$$\begin{aligned}\frac{d^2r(\theta)}{d\theta^2} &= (r(\theta))^{1-2p}\frac{4}{(\sin 2\theta)^2}\left(-4b_1b_2\frac{(\sin 2\theta)^p}{2^p} + b_1b_2p(\sin\theta)^4\frac{(\sin 2\theta)^p}{2^p}\right) \\ &\quad -{b_1}^2(\cos\theta)^{2p}\frac{(\sin 2\theta)^2}{4} + 2b_1b_2p\frac{(\sin 2\theta)^p}{2^p}\frac{(\sin 2\theta)^2}{4}\end{aligned}$$

$$
\begin{aligned}
& -b_1 b_2 (\sin\theta)^4 \frac{(\sin 2\theta)^2}{2^p} - {b_2}^2 (\sin\theta)^{2p} \frac{(\sin 2\theta)^2}{4} \\
& +b_1 b_2 p (\cos\theta)^4 \frac{(\sin 2\theta)^p}{2^p} - b_1 b_2 (\cos\theta)^4 \frac{(\sin 2\theta)^p}{2^p} \\
= \; & (r(\theta))^{1-2p} \left(\frac{(\sin 2\theta)^p}{2^{p-1}} b_1 b_2 (p-2) - {b_1}^2 (\cos\theta)^{2p} - {b_2}^2 (\sin\theta)^{2p} \right. \\
& \left. + \left(\frac{(\sin 2\theta)}{2} \right)^{p-2} b_1 b_2 (p-1)((\sin\theta)^4 + (\cos\theta)^4) \right)
\end{aligned}
$$

After some further rearrangements, we find that

$$
\frac{d^2 r(\theta)}{d\theta^2} = -r(\theta) + b_1 b_2 (p-1)(r(\theta))^{1-2p} \left(\frac{\sin 2\theta}{2} \right)^{p-2} \tag{2.4}
$$

Next we give some properties of the $r(\theta)$ function by using the above derivatives. These properties are related to the shape of $r(\theta)$ for $\theta \in [0, \pi/2]$. The properties depend on the relative values of the parameters b_1, b_2 and p. There are four cases:

1. $1 \leq p < 2$, $b_1 < b_2$ (Property 2.1.5)
2. $1 \leq p < 2$, $b_1 > b_2$ (Property 2.1.6)
3. $p > 2$, $b_1 < b_2$ (Property 2.1.7)
4. $p > 2$, $b_1 > b_2$ (Property 2.1.8)

For all four cases, the stationary point of $r(\theta)$, θ^*, is defined for $\theta \in [0, \pi/2]$ as follows:

$$
\theta^* = \begin{cases} \arctan(b_1/b_2)^{\frac{1}{p-2}} & \text{if } p \geq 1, \; p \neq 2, \\ 0 \text{ or } \pi/2 & \text{if } p = 2. \end{cases} \tag{2.5}
$$

Notice that θ^* is a decreasing function of p when $b_1 > b_2$ and an increasing function of p when $b_1 < b_2$.

Property 2.1.5 *If* $1 \leq p < 2$, *and* $b_1 < b_2$, *then*

a. $r(\theta)$ *increases for* $\theta \in (0, \theta^*)$, *and decreases for* $\theta \in (\theta^*, \pi/2)$, *where* $\theta^* \in [\pi/4, \pi/2]$.

b. $r(\theta)$ *has two inflection points,* $\tilde{\theta}_1$, $\tilde{\theta}_2$ *where* $\tilde{\theta}_1 \in [0, \theta^*]$ *and* $\tilde{\theta}_2 \in [\theta^*, \pi/2]$.

Proof

a. First notice that

$$\left.\frac{dr(\theta)}{d\theta}\right|_{\theta=0} = \left.\frac{dr(\theta)}{d\theta}\right|_{\theta=\frac{\pi}{2}} = 0.$$

For $\theta < \theta^*$ we have $\theta < \arctan(b_1/b_2)^{1/(p-2)}$ which implies that $b_2 (\sin\theta)^{p-2} - b_1 (\cos\theta)^{p-2} > 0$. Thus we obtain $dr(\theta)/d\theta > 0$, $\theta \in (0, \theta^*)$. Similarly $dr(\theta)/d\theta < 0$, for $\theta \in (\theta^*, \pi/2)$. Furthermore, since θ^* is the stationary point we have $dr(\theta^*)/d\theta = 0$.

Finally we observe that for $b_1 < b_2$ and $1 \leq p < 2$, $\tan\theta^* = (b_2/b_1)^{1/(2-p)} > 1$ which implies that $\theta^* \in [\pi/4, \pi/2]$.

b. First we consider the inflection point $\tilde{\theta}_1$. Since $1 \leq p < 2$ we see that

$$\lim_{\theta \to 0+} \frac{d^2 r(\theta)}{d\theta^2} = +\infty.$$

Furthermore, considering Property 2.1.4 and part (a) above it follows that

$$\left.\frac{d^2 r(\theta)}{d\theta^2}\right|_{\theta=\theta^*} < 0.$$

Hence we can state that $\exists\, \theta = \tilde{\theta}_1 \ \ni \ d^2r(\theta)/d\theta^2 = 0$, where $\theta \in [0, \theta^*]$, $1 \leq p < 2$ and $b_1 < b_2$.

Next we consider the second inflection point $\tilde{\theta}_2$. It is readily known that

$$\left.\frac{d^2 r(\theta)}{d\theta^2}\right|_{\theta=\theta^*} < 0.$$

Furthermore, it can easily be verified that

$$\lim_{\theta \to \pi/2} \frac{d^2 r(\theta)}{d\theta^2} = +\infty.$$

Thus we conclude that $\exists\, \theta = \tilde{\theta}_2 \ \ni \ d^2r(\theta)/d\theta^2 = 0$, where $\theta \in [\theta^*, \pi/2]$, $1 \leq p < 2$ and $b_1 < b_2$. □

Property 2.1.6 *If* $1 \leq p < 2$, *and* $b_1 > b_2$, *then*

a. $r(\theta)$ *increases for* $\theta \in (0, \theta^*)$, *and decreases for* $\theta \in (\theta^*, \pi/2)$, *where* $\theta^* \in [0, \pi/4]$.

b. $r(\theta)$ *has two inflection points,* $\tilde{\theta}_1$, $\tilde{\theta}_2$ *where* $\tilde{\theta}_1 \in [0, \theta^*]$ *and* $\tilde{\theta}_2 \in [\theta^*, \pi/2]$.

Proof

a. Similar to Property 2.1.5.a we have $dr(\theta)/d\theta > 0$, where $\theta \in (0, \theta^*)$, $dr(\theta)/d\theta < 0$, for $\theta \in (\theta^*, \pi/2)$, and $dr(\theta^*)/d\theta = 0$. However, in this case since $b_1 > b_2$, we have $\tan\theta^* = (b_2/b_1)^{1/(2-p)} < 1$ which implies that $\theta^* \in [0, \pi/4]$.

b. We can use the same approach that was used in Property 2.1.5 to examine the inflection points because we still have the case where $1 \leq p < 2$. On the other hand, we can make use of Properties 2.1.1 and 2.1.3. In this case it is not necessary to

use the derivatives, but instead we utilize the relations between functions $r(\theta)$ with $1 \leq p < 2$, $b_1 < b_2$, and $r(\theta)$ with $1 \leq p < 2$, $b_1 > b_2$. □

Property 2.1.7 *If $p > 2$, and $b_1 < b_2$, then*

a. *$r(\theta)$ decreases for $\theta \in (0, \theta^*)$, and increases for $\theta \in (\theta^*, \pi/2)$, where $\theta^* \in [0, \pi/4]$.*

b. *$r(\theta)$ has two inflection points, $\tilde{\theta}_1$, $\tilde{\theta}_2$ where $\tilde{\theta}_1 \in [0, \theta^*]$ and $\tilde{\theta}_2 \in [\theta^*, \pi/2]$.*

Proof

a. We again make use of the first derivative of $r(\theta)$. Notice that

$$\left.\frac{dr(\theta)}{d\theta}\right|_{\theta=0} = \left.\frac{dr(\theta)}{d\theta}\right|_{\theta=\frac{\pi}{2}} = 0.$$

For $\theta < \theta^*$ we have $\theta < \arctan(b_1/b_2)^{1/(p-2)}$ which implies that $b_2 (\sin\theta)^{p-2} - b_1 (\cos\theta)^{p-2} > 0$. Thus we obtain $dr(\theta)/d\theta > 0$, $\theta \in (0, \theta^*)$. Similarly for $\theta \in (\theta^*, \pi/2)$, $dr(\theta)/d\theta < 0$. Also note that since θ^* is the stationary point we have $dr(\theta^*)/d\theta = 0$.

Finally we observe that for $b_1 < b_2$ and $p > 2$, $\tan\theta^* = (b_1/b_2)^{1/(p-2)} > 1$ implying that $\theta^* \in [0, \pi/4]$.

b. First consider the inflection point $\tilde{\theta}_1$. It can easily be verified that

$$\left.\frac{d^2 r(\theta)}{d\theta^2}\right|_{\theta=0} = -b_1^{1/p} < 0.$$

Furthermore, it follows from Property 2.1.4 and the first part of this property that

$$\left.\frac{d^2 r(\theta)}{d\theta^2}\right|_{\theta=\theta^*} > 0.$$

Thus we conclude that $\exists\, \theta = \tilde{\theta}_1 \ \ni \ d^2r(\theta)/d\theta^2 = 0$, where $\theta \in [0, \theta^*]$, $p > 2$ and $b_1 < b_2$.

Next we consider the second inflection point $\tilde{\theta}_2$ which is in the interval $[\theta^*, \pi/2]$. It is already known that

$$\left.\frac{d^2 r(\theta)}{d\theta^2}\right|_{\theta=\theta^*} > 0.$$

Moreover, it can be shown that

$$\left.\frac{d^2 r(\theta)}{d\theta^2}\right|_{\theta=\pi/2} = -{b_2}^{1/p},$$

and therefore the second derivative of $r(\theta)$ at $\theta = \pi/2$ is negative. Thus we conclude that $\exists\, \theta = \tilde{\theta}_2 \ \ni \ d^2r(\theta)/d\theta^2 = 0$, where $\theta \in [\theta^*, \pi/2]$, $p > 2$ and $b_1 < b_2$. □

Property 2.1.8 *If $p > 2$, and $b_1 > b_2$, then*

a. *$r(\theta)$ decreases for $\theta \in (0, \theta^*)$, and increases for $\theta \in (\theta^*, \pi/2)$, where $\theta^* \in [\pi/4, \pi/2]$.*

b. *$r(\theta)$ has two inflection points, $\tilde{\theta}_1$, $\tilde{\theta}_2$ where $\tilde{\theta}_1 \in [0, \theta^*]$ and $\tilde{\theta}_2 \in [\theta^*, \pi/2]$.*

Proof

a. As in Property 2.1.7.a we have $dr(\theta)/d\theta < 0$, where $\theta \in (0, \theta^*)$, $dr(\theta)/d\theta > 0$, for $\theta \in (\theta^*, \pi/2)$, and $dr(\theta^*)/d\theta = 0$. However, in this case where $b_1 > b_2$, we have $\tan\theta^* = (b_1/b_2)^{1/(p-2)} > 1$ showing that $\theta^* \in [\pi/4, \pi/2]$.

b. Similar to Property 2.1.6.b, we can use two approaches to examine the inflection points. We can employ the same approach that we have used for Property 2.1.7 since

we still have the case where $p > 2$. Secondly, we can make use of Property 2.1.1 and Property 2.1.3. In this case instead of using the derivatives we utilize the relations between functions $r(\theta)$ with $p > 2$, $b_1 < b_2$ and $r(\theta)$ with $p > 2$, $b_1 > b_2$. □

We have already mentioned that depending on the relative values of the parameters b_1, b_2, and p there are four possible shapes of the $r(\theta)$ function. In Properties 2.1.5 to 2.1.8 we have identified these shapes.

Observe that the boundary values of $r(\theta)$ are

$$r(0) = b_1^{1/p} \quad \text{and} \quad r(\pi/2) = b_2^{1/p}, \tag{2.6}$$

and for $p \geq 1$, $p \neq 2$, $r(\theta)$ evaluated at the stationary point θ^* is given by

$$r(\theta^*) = \left[b_1 \left(1 + (b_1/b_2)^{2/(p-2)}\right)^{-1/2p} + b_2 \left(\frac{(b_1/b_2)^{1/(p-2)}}{\left(1 + (b_1/b_2)^{2/(p-2)}\right)^{1/2}} \right)^p \right]^{1/p}. \tag{2.7}$$

For $p = 2$, using the θ^* values given in (2.5) we see that $r(\theta^*)$ is obtained by the boundary values (2.6). The expressions (2.6) and (2.7) directly follow from substituting the values of θ in $r(\theta)$.

We give the following property without an explicit proof. It follows from the definition of $r(\theta)$ and Corollary 1.2.1.

Property 2.1.9 *Let b_1 and b_2 be given parameter values and $p > 1$. Then $r(\theta)$ is a decreasing function of p for any fixed $\theta \in [0, \pi/2]$.*

Note that, in contrast to the directional bias function of the ℓ_p-norm, this property is also valid at the boundaries, i.e., $\theta = 0$ and $\theta = \pi/2$. As a direct consequence of

this result we state the following property.

Property 2.1.10 *Consider two $r(\theta)$ functions, $r_1(\theta)$, $r_2(\theta)$, with a set of given b_1 and b_2 values and the parameter p, such that $1 \leq p < 2$ for $r_1(\theta)$ and $p > 2$ for $r_2(\theta)$. Then we have*

$$r_1(\theta) \geq r_2(\theta), \quad \theta \in \left[0, \frac{\pi}{2}\right].$$

Using (2.6) we see that the equality above holds only at the boundary $\theta = 0$ ($\theta = \pi/2$) if the corresponding weight b_1 (b_2) is equal to 1, in which case $r_1(0) = r_2(0) = 1$ ($r_1(\pi/2) = r_2(\pi/2) = 1$).

We note the following limiting cases related to the stationary point $(\theta^*, r(\theta^*))$. Particularly, we are interested in the behaviour of θ^* and $r(\theta^*)$ when $p \to 2$ and also $b_1 \to b$, $b_2 \to b$ or $(b_1/b_2) \to 1$. We present the limiting cases in four groups along with the conditions examined in Properties 2.1.5 to 2.1.8.

1. $1 \leq p < 2, \quad b_1 < b_2$: (Property 2.1.5)

$$\lim_{p \to 2^-} \theta^* = \frac{\pi}{2}, \quad \lim_{p \to 1} \theta^* = \arctan\left(\frac{b_2}{b_1}\right), \quad \lim_{\frac{b_1}{b_2} \to 1^-} \theta^* = \frac{\pi}{4}, \quad \lim_{p \to 2^-} r(\theta^*) = {b_2}^{1/p}.$$

2. $1 \leq p < 2, \quad b_1 > b_2$: (Property 2.1.6)

$$\lim_{p \to 2^-} \theta^* = 0, \quad \lim_{p \to 1} \theta^* = \arctan\left(\frac{b_2}{b_1}\right), \quad \lim_{\frac{b_1}{b_2} \to 1^+} \theta^* = \frac{\pi}{4}, \quad \lim_{p \to 2^-} r(\theta^*) = {b_1}^{1/p}.$$

3. $p > 2, \quad b_1 < b_2$: (Property 2.1.7)

$$\lim_{p \to 2^+} \theta^* = 0, \quad \lim_{p \to +\infty} \theta^* = \frac{\pi}{4}, \quad \lim_{\frac{b_1}{b_2} \to 1^-} \theta^* = \frac{\pi}{4}, \quad \lim_{p \to 2^+} r(\theta^*) = {b_1}^{1/p}.$$

4. $p > 2, \quad b_1 > b_2$: (Property 2.1.8)

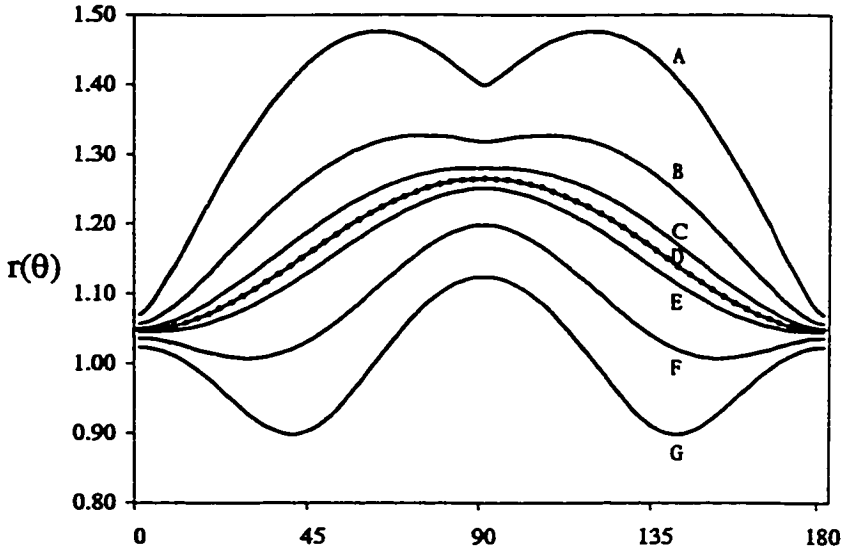
$$\lim_{p\to 2^+} \theta^* = \frac{\pi}{2}, \quad \lim_{p\to+\infty} \theta^* = \frac{\pi}{4}, \quad \lim_{\frac{b_1}{b_2}\to 1^+} \theta^* = \frac{\pi}{4}, \quad \lim_{p\to 2^+} r(\theta^*) = {b_2}^{1/p}.$$

Regardless of the case, we observe that

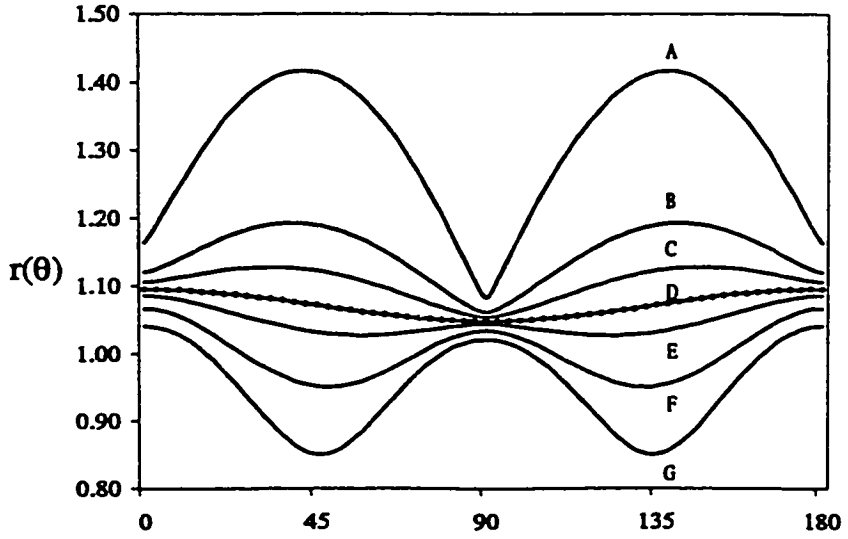
$$\lim_{b_1\to b}\lim_{b_2\to b} r(\theta^*) = b^{\frac{1}{p}}\, 2^{\frac{1}{p}-\frac{1}{2}}.$$

The plots of $r(\theta)$ with different parameter values b_1, b_2 and p are given in Figure 2.1. For comparison purposes, the plot of $r(\theta)$ for the ℓ_p-norm is also given in Figure 2.2. It can easily be seen from these plots that the ℓ_{bp}-norm models directional bias in a different way than the ℓ_p-norm. The difference will be discussed in detail in the following section.

The unit ball of the ℓ_{bp}-norm is given in Figure 2.3. For clarity, we show two sets of unit ball plots corresponding to Properties 2.1.5 to 2.1.8. We also include the unit ball plots of the ℓ_p-norm for different p values in Figure 2.4. One immediate observation is that, similar to the ℓ_p-norm, the unit balls do not contain flat spots for $p > 1$, and thus, the ℓ_{bp}-norm is also a round norm. However, in contrast to the unit ball of the ℓ_p-norm, the x_1- and x_2-axis intercepts are not equal to 1, but are given by ${b_1}^{-1/p}$ and ${b_2}^{-1/p}$, respectively. Consider the case where $b_1 \neq b_2$ and $p = 2$. Then the unit ball actually becomes an ellipse with the equation $(b_1x_1^2 + b_2x_2^2)^{1/2} = 1$. For values of p decreasing from two to one this elliptic shape of the unit ball shrinks and ultimately becomes a diamond-shaped ball for $p = 1$. Conversely, for values of p

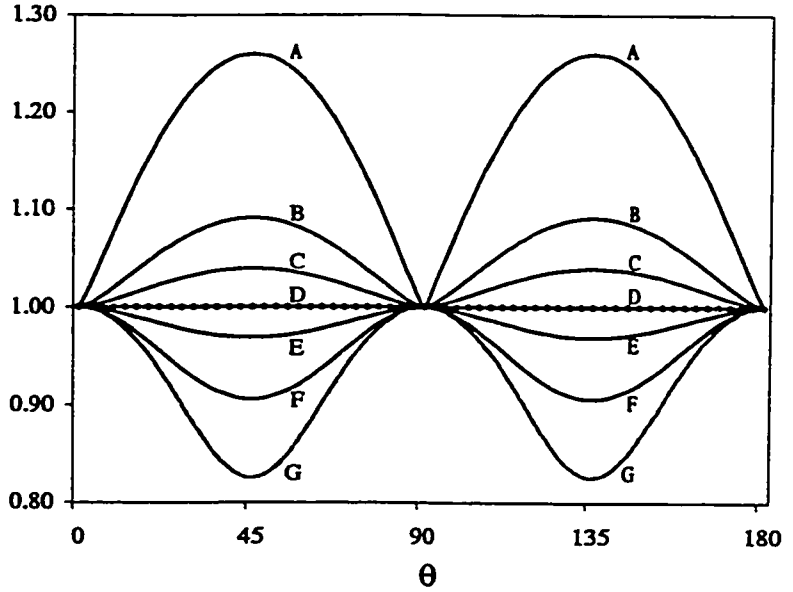


$b_1 = 1.1, b_2 = 1.6,$
$A: p = 1.4,\ B: p = 1.7,\ C: p = 1.9,\ D: p = 2.0,$
$E: p = 2.1,\ F: p = 2.6,\ G: p = 4.0$



$b_1 = 1.2, b_2 = 1.1,$
$A: p = 1.2,\ B: p = 1.6,\ C: p = 1.8,\ D: p = 2.0,$
$E: p = 2.2,\ F: p = 2.8,\ G: p = 4.5$

Figure 2.1: $r(\theta)$ plots for the ℓ_{bp}-norm



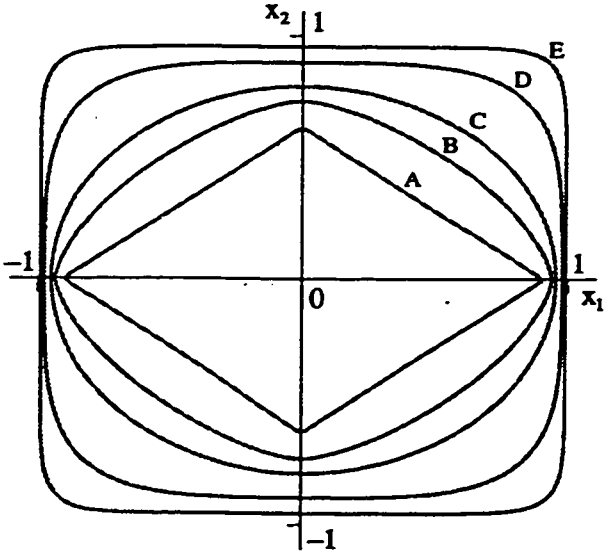
$A: p = 1.2,\ B: p = 1.6,\ C: p = 1.8,\ D: p = 2.0,$
$E: p = 2.2,\ F: p = 2.8,\ G: p = 4.5$

Figure 2.2: $r(\theta)$ plots for the ℓ_p-norm

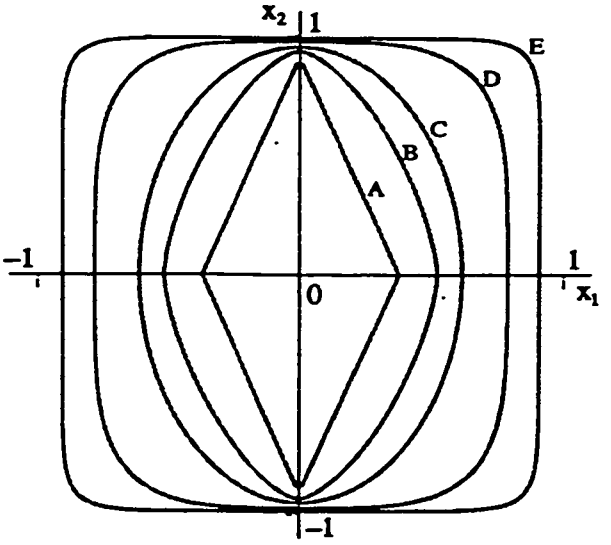
increasing from two to infinity the unit ball expands and converges to a unit square. That is, it becomes symmetric with respect to orthogonal coordinate axes. In a sense we can say that a relatively high degree of rectangularity offsets the directional nonlinearity captured by the parameters b_1 and b_2. Note also that since $b_1,\ b_2 > 1$, all unit balls for $p > 1$ will be enclosed in this unit square.

2.2 Implications for Distance Prediction

The directional bias function, $r(\theta)$, provides some practical insights which are useful in modelling distances. Distance modelling in a region involves determining the best parameter values of the distance predicting function, e.g. b_1, b_2 and p in the ℓ_{bp}-norm,

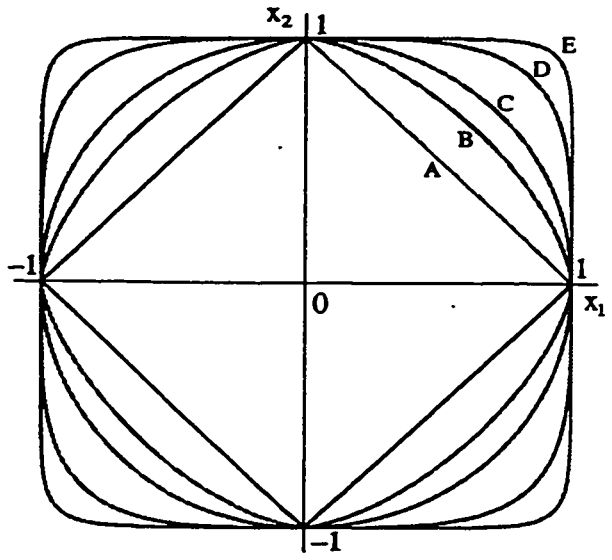


$b_1 = 1.1, b_2 = 1.6,$
$A: p = 1.0, \quad B: p = 1.5, \quad C: p = 2.0, \quad D: p = 4.0, \quad E: p = 10.0$



$b_1 = 2.6, b_2 = 1.1,$
$A: p = 1.0, \quad B: p = 1.5, \quad C: p = 2.0, \quad D: p = 4, \quad E: p = 10.0$

Figure 2.3: Plots of the Unit Ball for the ℓ_{bp}-norm



$A:\ p = 1.0,\ \ B:\ p = 1.5,\ \ C:\ p = 2.0,\ \ D:\ p = 4.0,\ \ E:\ p = 10.0$

Figure 2.4: Plots of the Unit Ball for the ℓ_p-norm

so that a prediction errors related goodness-of-fit criterion value is minimized. For the weighted ℓ_p-norm, the computational procedures to determine the best parameter values are given by Brimberg and Love (1991) and for the ℓ_{bp}-norm in Chapter 3. In both procedures, for the parameter p it is necessary to conduct a search over a safe range of values that includes the optimal value of p for the transportation network. Brimberg and Love (1993a) analyze the directional bias function of the ℓ_p-norm, $r_p(\theta)$, in detail and conclude that any $\ell_q(\mathbf{x})$, $q > 2$, can be accurately approximated by a corresponding norm $\sigma\ell_p(\mathbf{x}')$, $1 < p < 2$, where σ is a scaling factor and $\mathbf{x}'$ gives the coordinates of $\mathbf{x}$ after a 45° rotation of the reference axis. Therefore, for all practical purposes, the estimation of actual distances by an ℓ_p-norm with $p > 2$ need never be

considered, since the same degree of accuracy can be obtained with a value of p in the interval $(1, 2)$.

We define the direction of greatest (least) difficulty θ_G (θ_L) as the value of θ which maximizes (minimizes) the $r(\theta)$ function. Let x and y be two points in $\Re^2$ separated by a straight line segment $\mathcal{S}$ of fixed length $\ell_2(x-y)$. Then $\ell_{bp}(x-y)$ is maximized if $\mathcal{S}$ is parallel to θ_G and minimized if $\mathcal{S}$ is parallel to θ_L. It is shown by Brimberg and Love (1993a) that for the ℓ_p-norm, which is a special form of the ℓ_{bp}-norm, $\theta_G = \pi/4$, $\theta_L = 0,\ \pi/2$ for $1 \leq p < 2$ and $\theta_G = 0,\ \pi/2$, $\theta_L = \pi/4$ for $p > 2$. For the ℓ_{bp}-norm, inspecting the graph of $r(\theta)$ (Figure 2.1) and the unit ball (Figure 2.3), we see that when $1 \leq p < 2$, $\theta_L = 0$ for $b_1 < b_2$ and $\theta_L = \pi/2$ for $b_1 > b_2$, whereas the direction of greatest difficulty θ_G is such that $\theta_G \in (\arctan(b_2/b_1), \pi/2)$ for $b_1 < b_2$ and $\theta_G \in (0, \arctan(b_2/b_1))$ for $b_1 > b_2$. For $p > 2$ the situation is somewhat similar but in the opposite sense. In particular, the direction of least difficulty θ_L is such that $\theta_L \in (0, \pi/4)$ for $b_1 < b_2$ and $\theta_L \in (\pi/4, \pi/2)$ for $b_1 > b_2$, while the direction of greatest difficulty is at $\theta_G = \pi/2$ for $b_1 < b_2$ and $\theta_G = 0$ for $b_1 > b_2$. As a result of this non-fixed θ_G for $1 \leq p < 2$ and θ_L for $p > 2$, a phase change in the directions of greatest and least difficulty does not occur in the strict sense that it occurs for the ℓ_p-norm at $\theta = \pi/4$. In other words, we are not in general guaranteed that θ_G for $1 \leq p < 2$ and θ_L for $p > 2$ coincide as in the ℓ_p-norm where both are equal to $\pi/4$. A summary of these observations are given in Tables 2.1 and 2.2. Note that, if $p = 2$

	$1 \leq p < 2$		$p > 2$		$p = 2$	
	$b_1 < b_2$	$b_1 > b_2$	$b_1 < b_2$	$b_1 > b_2$	$b_1 < b_2$	$b_1 > b_2$
θ_L	0	$\pi/2$	$(0, \pi/4)$	$(\pi/4, \pi/2)$	0	$\pi/2$
θ_G	$(\arctan(b_2/b_1), \pi/2)$	$(0, \arctan(b_2/b_1))$	$\pi/2$	0	$\pi/2$	0

Table 2.1: Directions of Least and Greatest Difficulty for the ℓ_{bp}-norm

	$1 \leq p < 2$	$p > 2$
θ_L	0 and $\pi/2$	$\pi/4$
θ_G	$\pi/4$	0 and $\pi/2$

Table 2.2: Directions of Least and Greatest Difficulty for the ℓ_p-norm

then there does not exist a specific θ_L and θ_G for the ℓ_p-norm.

Suppose that we have a transportation network where an ℓ_{bp}-norm with $1 \leq p < 2$ can be closely approximated by an ℓ_{bq}-norm where $q > 2$ and the coordinate axes are rotated. Then while modelling distances in this network we have two minimum criterion values for $\theta \in [0, \pi/2]$, one with $1 \leq p < 2$ and the other with $p > 2$. We will call these minimum criterion values 'bottoms' in the $SD(\theta)$ graphs. In the examples presented at the end of this section we use the "Sum of Squared Deviations" SD as the criterion. Therefore, the bottoms correspond to SD_{I} and SD_{II} in Table 2.3. The $SD(\theta)$ graphs for the example regions are found in the Appendix A. If the weighted ℓ_p-norm is used, then the above mentioned close approximation is always possible and therefore the bottoms always occur 45° apart with the same minimum criterion values. This is a direct consequence of the exact phase change of 45° observed in the $r(\theta)$ graph given in Figure 2.2. However, for the ℓ_{bp}-norm, which is a generalization of the ℓ_p-norm, this approximation does not generally

hold. Whether an ℓ_{bq}-norm, $q > 2$, can be accurately approximated by another ℓ_{bp}-norm, where $1 < p < 2$, depends on the level of rectangularity and nonlinearity inherent in a transportation network. However similar to the weighted ℓ_p-norm, the ℓ_{bp}-norm attempts to identify two inherent characteristics of a transportation network: rectangularity which is mostly associated with the parameter p, and nonlinearity which is associated with the parameters b_1 and b_2. For example, a grid system or "Manhattan metric" type of road network has rectangularity properties and is captured by the parameter p ($p \approx 1$). On the other hand, a river or a mountain range or other irregularities will produce nonlinearity effects and will be explained by the parameters b_1 and b_2.

Before proceeding to the discussion on this difference between the weighted ℓ_p-norm and the ℓ_{bp}-norm, we define an *indicator of directional nonlinearity, τ, at an axis rotation* as follows:

$$\tau = \max\left\{\frac{b_1}{b_2}, \frac{b_2}{b_1}\right\}, \tag{2.8}$$

where b_1 and b_2 are the best parameter values for an axis rotation θ. The corresponding τ values for SD_{I} and SD_{II} are denoted by τ_1 and τ_2, respectively. We can employ $\Delta\tau = |\tau_1 - \tau_2|$ as an indicator for the *existence of directional nonlinearity in the transportation network*. While a high value of $\Delta\tau$ indicates the existence of directional nonlinearity, a low value presents evidence for uniform nonlinearity. For example, consider a perfectly rectangular transportation network, aligned with

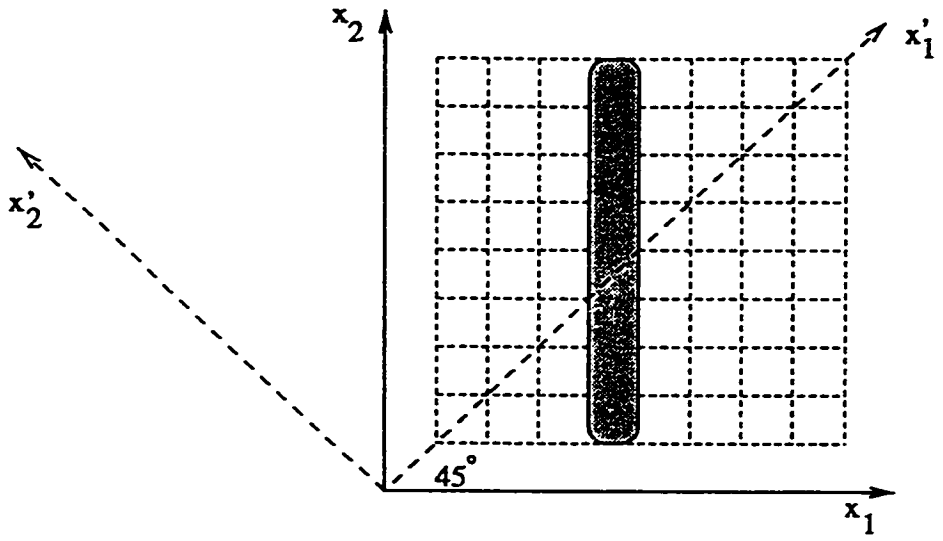


Figure 2.5: Example Road Pattern

the conventional coordinate axis, with an irregularity, say a mountain range, in the vertical direction as shown in Figure 2.5. It is clear that both nonlinearity and rectangularity are highly pronounced characteristics of the underlying pattern. When fitting distances in such a transportation network we expect to see a bottom in the $SD(\theta)$ graph at $\theta = 0$ with the parameter p value very close to 1. The parameter b_1 will be relatively high compared to b_2 showing the high level of nonlinearity or difficulty of travel in x_1 direction. The value of τ_1 will be high indicating the directional nonlinearity captured at this axis rotation. The second bottom occurs after 45° of axis rotation and corresponds to a p value much greater than 2. This large value of p still captures the rectangularity accurately in the network (Brimberg and Love, 1993a). Since the irregularity has the same effect on travel in both directions after the axis rotation, the b_1 and b_2 values will be very close. Relatively high values of

both b_1 and b_2 will still indicate the high level of nonlinearity. However,the low value of τ_2, close to 1, will show an existence of uniform nonlinearity at this axis rotation. The resulting high value of $\Delta\tau$ will indicate the existence of a predominant direction of nonlinearity in this particular transportation network. When $\theta = 0$, the axes are in perfect alignment with both rectangularity and directional nonlinearity inherent in the network, and the level of rectangularity, nonlinearity and the existence of directional nonlinearity are modelled accurately. After the axis rotation we will have the worst alignment with the underlying rectangularity and directional nonlinearity. When $\theta = 45°$, the distance model still captures the level of rectangularity and nonlinearity. However, τ_2 does not indicate the existing directional nonlinearity in the transportation network. As a result, we can not expect to obtain the same SD_{I} and SD_{II} values for such a network; SD_{I} will be much lower than SD_{II}. The road pattern in Sydney is an example for this case. Conversely, if we have uniform nonlinearity over all the transportation network, then τ will be insensitive to axis rotation so that $\Delta\tau$ will be very close to zero as in the case of Toronto (Table 2.3).

Returning to our discussion on the existence of two bottoms in the $SD(\theta)$ graph, we first consider only the boundaries $\theta = 0, \pi/2$. It follows from Figure 2.1 and Property 2.1.1 that if the criterion SD attains its minimum at an axes rotation θ, then the same minimum criterion value must be obtained at $(\theta+\pi/2)$ where b_1 and b_2 values are exchanged, p being the same. We call this minimum value of the criterion

occurring in the $SD(\theta)$ graph the first bottom. Secondly we consider the cases in which another minimum criterion value, the second bottom, is attained in the interval $(0, \pi/2)$. We argue that the existence of such a second bottom in the $SD(\theta)$ graph depends on the existence of directional nonlinearity and the level of rectangularity in the transportation network. If the underlying pattern is highly Euclidean, i.e. $p \approx 2$, then all four limiting cases suggest that a possible second bottom is not likely to occur. In this case the θ_L's approach the boundaries making the phase change of $\pi/2$ discussed above more likely to occur. The best parameter p value can then occur either in the interval $(1, 2)$ or $(2, +\infty)$. We can say that in such a case the directional nonlinearity clearly dominates the rectangularity in the network. On the other hand, if there exists a less pronounced directional nonlinearity, i.e. $\Delta\tau \approx 0$, or a high rectangularity, i.e. $p \to 1$ or $p \to +\infty$, then θ_G for $1 \leq p < 2$ and θ_L for $p > 2$ move towards $\pi/4$ or equivalently the unit ball actually converges to a shape similar to the unit ball of the weighted ℓ_p-norm. Hence this case actually resembles the weighted ℓ_p-norm and a second bottom is likely to occur in the $SD(\theta)$ graph. This time the rectangularity dominates the directional nonlinearity in the network. The second bottom which occurs as a result of approximately a 45° phase change on the $r(\theta)$ graph is about 45° apart from the first bottom.

These observations lead us to the conclusion that while modelling distances by using the ℓ_{bp}-norm we may not always be assured that there exists a good fit with

a parameter p value in the interval $(1, 2)$. Therefore we can not limit our search for the optimal p value of the norm to this interval. However, we can still impose an upper bound on the search range of p. Suppose that the underlying pattern has a dominant rectangularity. Then we know that there exists two sets of best parameter values where one of them has a p value in the interval $(1, 2)$. Therefore we have an upper bound of 2 for the search range of p in this case. Now suppose that there exists a predominant direction of nonlinearity in the transportation network. Because of the possibility of having an optimal parameter p value greater than 2 for such a case, we have to consider a larger search range, $(1, \bar{p})$ where $\bar{p} > 2$, in distance modelling algorithms. This upper bound must be high enough to capture the rectangularity in the transportation network so that if the optimal p is above this limit there will be a corresponding optimal fit with a p value close to 1. Although the choice of this upper bound for the search range of p is left to the analyst's preliminary inspection of the transportation network analyzed, our empirical work on seventeen geographic regions presented in Chapter 3 reveals that use of a search range $[1, 4]$ will always obtain the best parameter p value.

As an example, we comment on the optimum parameter values for three of the above mentioned seventeen regions, Great Britain, Sydney and Toronto. We are mainly interested in the interpretation of the optimal parameter values which are given here for convenience in Table 2.3. The $SD(\theta)$ graphs are given in the Appendix A.

REGION	SD_{I} (First Bottom) θ p, b_1, b_2 τ_1	SD_{II} (Second Bottom) θ p, b_1, b_2 τ_2	ΔSD $\Delta\tau$
Great Britain	172.34 0° 2.0352, 1.1116, 1.3925 1.2527	–	–
Sydney	1.0984 8° 1.5571, 1.3675, 1.1521 1.1870	1.3550 54° 3.0007, 2.1130, 2.1541 1.0195	23.26% 0.1675
Toronto	5.0557 42° 5.2215, 4.1261, 4.2427 1.0282	5.0619 87° 1.2009, 1.0581, 1.0275 1.0298	0.12% 0.0016

Table 2.3: Example Parameter Values

Great Britain is a case where the optimal p is greater than 2. Since p is very close to 2 the road network is highly Euclidean. The value of τ_1 (1.2527) represents the high directional nonlinearity modelled at the axis rotation $\theta = 0$. There exists only one minimum criterion value occurring in the $[0, \pi/2]$ interval. In Sydney's case we observe two bottoms, and there is a considerable gap between them, 23.36%. This is because the region is fairly rectangular ($p = 1.5571$) but at the same time there exists directional nonlinearity. However, neither is dominant. The directional nonlinearity in the transportation network is evident from the significant value of $\Delta\tau$ (0.1675). The first bottom better represents the existing directional nonlinearity with its higher τ_1 value and thus provides a lower SD criterion value than the second bottom. For Toronto, there exists a low level of directional nonlinearity with $\tau_1 = 1.0282$ and $\Delta\tau = 0.0016$, and the network is highly rectangular. The rectangularity in the road

pattern clearly dominates the directional nonlinearity. Hence, we see the two bottoms 45° apart and there is a negligible amount of gap (0.12%) between the criterion values at the bottoms.

Chapter 3

Distance Prediction With the ℓ_{bp}-norm

This chapter is concerned with the application of weighted sum of order p to predicting distances in a transportation network. To model distances for a given geographical region we need a goodness-of-fir criterion. In Section 1.1.2, we have given the mostly used three goodness-of-fit criteria AD, SD and NAD in the literature. Love and Üster (1996) have conducted a detailed study on the statistical comparison of these criteria by using the weighted ℓ_p-norm with axis rotation. The comparisons were carried out using the estimation error distributions for seventeen geographical regions. The authors considered two estimation error related random variables, $e(\mathbf{a}_i, \mathbf{a}_j)$ and $|e(\mathbf{a}_i, \mathbf{a}_j)|/A(\mathbf{a}_i, \mathbf{a}_j)$, where $e(\mathbf{a}_i, \mathbf{a}_j) = A(\mathbf{a}_i, \mathbf{a}_j) - k\ell_p(\mathbf{a}_i, \mathbf{a}_j)$, to statistically compare the criteria in several aspects such as the homoscedasticity and the expected value of $e(\mathbf{a}_i, \mathbf{a}_j)$, the homoscedasticity of $|e(\mathbf{a}_i, \mathbf{a}_j)|/A(\mathbf{a}_i, \mathbf{a}_j)$, and the accuracy in predicting long and short distances. The only statistically significant difference found is that, for

the SD criterion, the $e(\mathbf{a}_i, \mathbf{a}_j)$ have an expected value of zero without any exceptions in all the regions. Furthermore, we know that SD is strictly convex in parameter k when used with the $k\ell_p$-norm, and thus the best value of parameter k when fitting the weighted ℓ_p-norm is given by the closed form formula (Brimberg and Love, 1991). Since the weighted ℓ_p-norm is a special case of $\ell_{bp}(\mathbf{x})$, we obtain similar computational efficiencies when finding the best b_1 and b_2 values in modelling distances with the ℓ_{bp}-norm. For these reasons we will use the SD criterion in our work.

This chapter is organized as follows: In Section 3.1 we examine the behaviour of the criterion SD as a function of b_1, b_2 and p. Using these properties, in Section 3.2 we provide the computational procedures to find the best parameter values of b_1, b_2, p and the axis rotation θ for a given geographical region. In Section 3.3 we empirically compare the ℓ_{bp}-norm and the weighted ℓ_p-norm regarding their accuracy in predicting distances in seventeen geographical regions.

3.1 Properties of the SD Function

In this section we derive some useful properties of the SD function. These properties will later be utilized in designing computational procedures to determine the best parameter values of the ℓ_{bp}-norm for a given transportation network. First we examine the convexity of SD in terms of its parameters b_1 and b_2.

Property 3.1.1 *The criterion SD is convex function of its parameters b_1 and b_2*

provided that $p \in [1, 2]$.

Proof Suppose there are n points in our sample set where each point is given by its coordinates (a_{i1}, a_{i2}), $i = 1 \ldots n$, and the actual distance for each pair of points $(\mathbf{a}_i, \mathbf{a}_j)$ is denoted by A_{ij}. Then we have

$$SD = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{A_{ij}} \Big(A_{ij} - (b_1 |a_{i1} - a_{j1}|^p + b_2 |a_{i2} - a_{j2}|^p)^{1/p} \Big)^2.$$

Consider one of the terms in this sum, denote it by h_{ij} and also denote $\ell_{bp}(\mathbf{a}_i, \mathbf{a}_j)$ by ℓ_{bp}. We examine the convexity of h_{ij} in the parameters b_1 and b_2. The following terms are calculated for the Hessian

$$H_{b_1,b_2} = \begin{bmatrix} \frac{\partial^2 h_{ij}}{\partial b_1^2} & \frac{\partial^2 h_{ij}}{\partial b_1 \partial b_2} \\ \\ \frac{\partial^2 h_{ij}}{\partial b_2 \partial b_1} & \frac{\partial^2 h_{ij}}{\partial b_2^2} \end{bmatrix}.$$

$$\begin{aligned}
\frac{\partial h_{ij}}{\partial b_1} &= -2 \frac{(A_{ij} - \ell_{bp}) |a_{i1} - a_{j1}|^p}{A_{ij}\, p\, {\ell_{bp}}^{p-1}}, \\
\frac{\partial h_{ij}}{\partial b_2} &= -2 \frac{(A_{ij} - \ell_{bp}) |a_{i2} - a_{j2}|^p}{A_{ij}\, p\, {\ell_{bp}}^{p-1}}, \\
\frac{\partial^2 h_{ij}}{\partial b_1^2} &= 2 \frac{|a_{i1} - a_{j1}|^{2p}}{A_{ij}\, p^2\, {\ell_{bp}}^{2p-1}} (\ell_{bp} + (p-1)(A_{ij} - \ell_{bp})), \text{ and} \\
\frac{\partial^2 h_{ij}}{\partial b_2^2} &= 2 \frac{|a_{i2} - a_{j2}|^{2p}}{A_{ij}\, p^2\, {\ell_{bp}}^{2p-1}} (\ell_{bp} + (p-1)(A_{ij} - \ell_{bp})).
\end{aligned}$$

Furthermore,

$$\frac{\partial^2 h_{ij}}{\partial b_1 \partial b_2} = 2 \frac{|a_{i1} - a_{j1}|^p |a_{i2} - a_{j2}|^p}{A_{ij}\, p^2\, {\ell_{bp}}^{2p-1}} (\ell_{bp} + (p-1)(A_{ij} - \ell_{bp})),$$

which is also equal to $\partial^2 h_{ij}/\partial b_2 \partial b_1$. It can be verified that $|H_{b_1,b_2}| = 0$.

The second derivatives of h_{ij} w.r.t. b_1 and b_2 have a common term given by

$$M_{ij} = (\ell_{bp} + (p-1)(A_{ij} - \ell_{bp})).$$

The sign of M_{ij} must be found in order to see if the diagonal entries in the Hessian are nonnegative. There are two cases to consider:

(i) For the b_1 , b_2 values in which the ℓ_{bp}-norm under-predicts the actual distance A_{ij}, i.e. $A_{ij} \geq \ell_{bp}(\mathrm{a}_i, \mathrm{a}_j)$, M_{ij} is always nonnegative.

(ii) For the b_1 , b_2 values in which the ℓ_{bp}-norm over-predicts the actual distance A_{ij}, i.e. $A_{ij} \leq \ell_{bp}(\mathrm{a}_i, \mathrm{a}_j)$, we proceed as follows;

In order to see if M_{ij} is nonnegative we need to check if the inequality

$$\ell_{bp} > (p-1)|A_{ij} - \ell_{bp}|$$

holds. Rewriting this inequality we have

$$\frac{1}{p-1} > \left|\frac{A_{ij}}{\ell_{bp}} - 1\right|$$

or equivalently by using $A_{ij} \leq \ell_{bp}$ and rearranging , we obtain

$$\frac{A_{ij}}{\ell_{bp}} > \frac{2-p}{1-p} \quad \text{where} \quad \frac{A_{ij}}{\ell_{bp}} \leq 1.$$

The graph corresponding to the right hand side of this inequality is given in Figure 3.1. The shaded region in this graph represents the area that the above inequality holds,

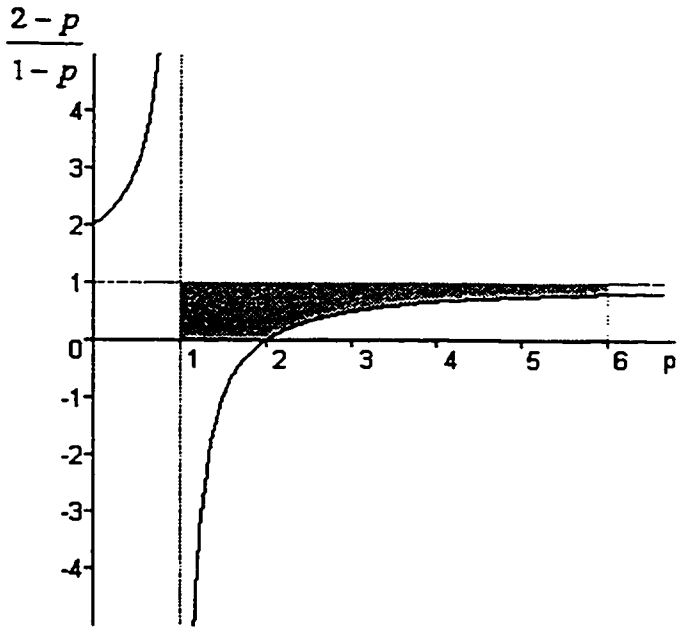


Figure 3.1: Condition on p

i.e. M_{ij} is nonnegative. It is apparent from the graph that we must consider the values for $p \in [1, \infty)$. Furthermore, for $p \in [1, 2]$ the above inequality always holds and therefore we have

$$\frac{\partial^2 h_{ij}}{\partial b_1^2} > 0 \quad \text{and} \quad \frac{\partial^2 h_{ij}}{\partial b_2^2} > 0, \quad \text{where} \quad 1 \leq p \leq 2.$$

Hence, h_{ij} is convex in b_1 and b_2 for $p \in [1, 2]$. The convexity of SD follows from the fact that a sum of convex functions is also a convex function. □

For $p \in (2, +\infty)$ we can argue that in practice there is a good chance for the M_{ij} to be nonnegative. First, it is readily seen in Chapter 2 that although it is possible to have an optimal p value greater than 2 with the ℓ_{bp}-norm this occurs with a p value close to 2, i.e. in a highly Euclidean transportation network. A minimum SD

criterion value with a large optimum value of p is closely approximated by another SD criterion value associated with a parameter p value less than 2 with the coordinate axes rotated. Secondly, for optimal values of the parameters, p, b_1 and b_2, we expect good predictions such that the ratio A_{ij}/ℓ_{bp} is close to 1. Third, since the ℓ_{bp}-norm is a decreasing function of p (Corollary 1.2.1), as p increases, ℓ_{bp} will decrease and given that $A_{ij}/\ell_{bp} \leq 1$, where A_{ij} is constant, the A_{ij}/ℓ_{bp} ratio will become closer to 1. It will be more likely to fall into the shaded region. Therefore, for all practical purposes we can assume that SD is a convex function of its parameters b_1 and b_2 provided that $p \geq 1$. Next we examine the behaviour of SD in terms of its parameter p. The following property is originally given for the SD criterion when it is used with the weighted ℓ_p-norm by Brimberg and Love (1991). It readily applies to the ℓ_{bp}-norm case as shown.

Property 3.1.2 *Consider any term h_{ij} in the sum SD as a function of p in the open interval $(0, +\infty)$. There are two cases:*

(i) *If $A_{ij} \leq \max\{|a_{i1} - a_{j1}|, |a_{i2} - a_{j2}|\} = \ell_\infty(a_i, a_j)$ then h_{ij} is a decreasing strictly convex function of p with a minimum approached asymptotically as $p \to +\infty$.*

(ii) *If $A_{ij} > \ell_\infty(a_i, a_j)$ then h_{ij} is a unimodal function of p with minimum at p^*. Furthermore, h_{ij} is strictly convex over the interval $0 < p \leq \mu$ and strictly concave for $\mu \leq p$ where μ is the inflection point such that*

$$\frac{\partial^2 h_{ij}(b_1, b_2, \mu)}{\partial p^2} = 0, \quad \mu > p^*.$$

Proof Let A and ℓ_{bp} be the actual and the predicted distances for the pair of points (a_i, a_j), respectively. Then

$$\begin{aligned}
\frac{\partial \ell_{bp}}{\partial p} &= \ell_{bp} K(p) \quad \text{where} \quad K(p) = -\frac{\ln(\ell_{bp})}{p} + \frac{\sum_{k=1}^{2} b_k |a_{ik} - a_{jk}|^p \ln(|a_{ik} - a_{jk}|)}{p(\ell_{bp})^p}, \\
\frac{\partial^2 \ell_{bp}}{\partial p^2} &= \ell_{bp} \left(K^2(p) + \frac{\partial K(p)}{\partial p} \right), \\
\frac{\partial h_{ij}}{\partial p} &= -\frac{2}{A}(A - \ell_{bp}) \frac{\partial \ell_{bp}}{\partial p}, \text{ and} \\
\frac{\partial^2 h_{ij}}{\partial p^2} &= \frac{2}{A} \left[\left(\frac{\partial \ell_{bp}}{\partial p} \right)^2 - (A - \ell_{bp}) \frac{\partial^2 \ell_{bp}}{\partial p^2} \right].
\end{aligned}$$

Also note that from Corollary 1.2.1 we have $\partial \ell_{bp}/\partial p < 0$ and $\partial^2 \ell_{bp}/\partial p^2 > 0$ for $p \in (0, +\infty)$.

(i) If $A \leq \ell_\infty$, we should have $A - \ell_{bp} < A - \ell_\infty \leq 0$, so that $\partial h_{ij}/\partial p < 0$ and $\partial^2 h_{ij}/\partial p^2 > 0$ for $p \in (0, +\infty)$. Clearly then, h_{ij} is a decreasing strictly convex function of p with the minimum $(A - \ell_\infty)^2/A$ approached asymptotically as $p \to +\infty$.

(ii) We next consider the case where $A > \ell_\infty$. In order to see the unimodality of h_{ij} recall that $A = \ell_{bp^*}$ and $\partial \ell_{bp}/\partial p < 0$ for $p \in (0, +\infty)$. Thus for $p < p^*$ we have $A - \ell_{bp} < 0$ which implies that $\partial h_{ij}/\partial p < 0$, and for $p > p^*$ we have $A - \ell_{bp} > 0$ implying that $\partial h_{ij}/\partial p > 0$. Therefore h_{ij} is a unimodal function of p with minimum at p^*. Also utilizing Corollary 1.2.1 we note that h_{ij} has a positive vertical asymptote at $p = 0$ and a horizontal asymptote approached from below as $p \to +\infty$. Hence h_{ij} must have an inflection point μ such that $\partial^2 h_{ij}(b_1, b_2, \mu)/\partial p^2 = 0$. We prove the rest of this case in two parts.

We first consider the $p < p^*$ case where the actual distances are over-predicted. It is readily seen that for $A < \ell_{bp^*}$ we have $\partial^2 h_{ij}/\partial p^2 > 0$. Therefore if $p < p^*$ then h_{ij} is a decreasing strictly convex positive function.

Secondly, we consider the case $p > p^*$ where the actual distances are under-predicted, i.e., $A > \ell_{bp^*}$. Substituting $\partial \ell_{bp}/\partial p$ and $\partial^2 \ell_{bp}/\partial p^2$ into the equation for $\partial^2 h_{ij}/\partial p^2$ and equating to zero we obtain

$$\frac{\partial^2 h_{ij}}{\partial p^2} = \frac{2}{A}\left[(\ell_{bp})^2 K^2(p) - (A - \ell_{bp})\ell_{bp}\left(K^2(p) + \frac{\partial K(p)}{\partial p}\right)\right] = 0,$$

and by rearranging we have

$$\frac{A}{\ell_{bp}} = \frac{2K^2(p) + (\partial K(p)/\partial p)}{K^2(p) + (\partial K(p)/\partial p)}.$$

The inflection point μ must solve this equation. Since $\partial^2 \ell_{bp}/\partial p^2 > 0$ we have

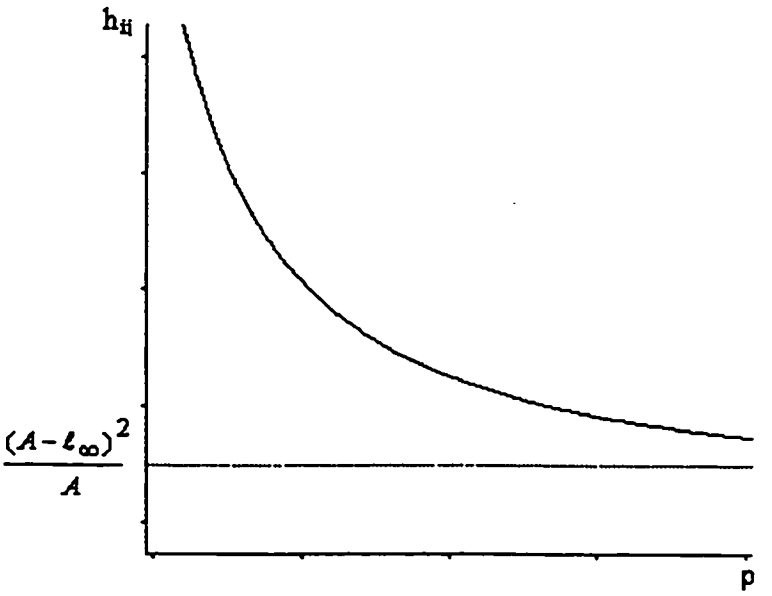
$$2K^2(p) + \frac{\partial K(p)}{\partial p} > K^2(p) + \frac{\partial K(p)}{\partial p} > 0$$

so that

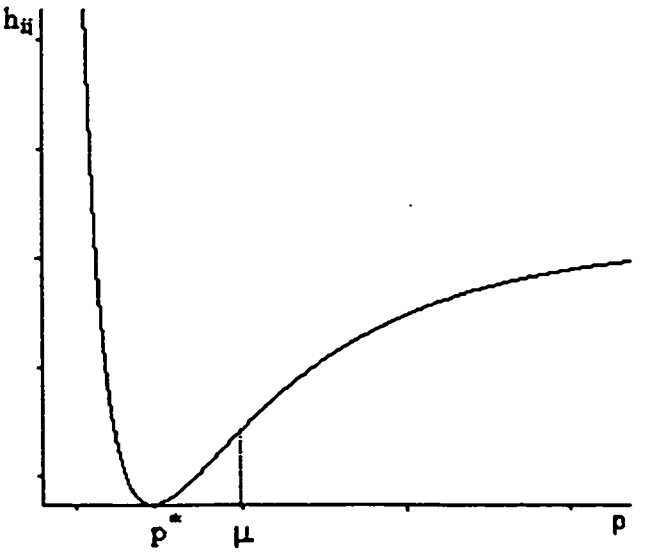
$$\frac{2K^2(p) + (\partial K(p)/\partial p)}{K^2(p) + (\partial K(p)/\partial p)} > 1, \quad \forall p > 0.$$

Hence $A > \ell_{b\mu}$. Finally since $A = \ell_{bp^*}$ and ℓ_{bp} is a decreasing function of p we must have $\mu > p^*$. The general shapes of h_{ij} for varying p are given in Figure 3.2 for both cases. □

Property 3.1.2 shows that h_{ij} is neither a convex nor concave function of p. Since the criterion SD is the sum of terms h_{ij} we conclude that SD is neither convex



i. $A \leq \ell_\infty$



ii. $A > \ell_\infty$

Figure 3.2: Shape of h_{ij}

nor concave in p. Therefore, in order to determine the best parameter value of p we have to conduct a numerical search over a safe range. The lower bound for such a search range is clearly 1. The upper bound, $\bar{p}$, is chosen by considering the level of rectangularity and nonlinearity in the transportation network. Chapter 2 reveals that if there exists only one optimal p value greater than 2 this happens in transportation networks which are nearly Euclidean, i.e. p is close to 2, or there exists a high directional nonlinearity, i.e. the b_1 and b_2 values are distinct. If the network has a high level of rectangularity and less pronounced directional nonlinearity then there exists two minimum SD criterion values. One occurs at a p value less than 2 and the other has p value greater than 2 which occurs after rotating the coordinate axes. Therefore, an appropriate upper bound, $\bar{p}$, for the search range of p can be chosen by making a preliminary inspection of the underlying pattern in the network so that the minimum SD criterion value is obtained.

3.2 A Computational Procedure

In order to calculate the SD criterion's values and to determine the empirical parameters of the ℓ_{bp}-norm, a *Search-Descent Algorithm*, similar to the one given by Brimberg and Love (1991), was developed. The outline of the computational procedure is given in Display 3.1.

The best θ and p values were determined by using an incremental *search* procedure and a four-stage incremental search procedure, respectively. To calculate

the best θ value, a two-degree incremental search was conducted on the interval $[0, 90°]$. We chose this interval because the directional bias function $r(\theta)$ of the ℓ_{bp}-norm indicates that the SD function is periodic with a period 90° (Property 2.1.2). Therefore a best-fit rotation angle must be encountered in the interval $[0, 90°]$. For each value of θ, parameters b_1, b_2 and p were determined. Once the best θ value, θ_b, was found for the interval, solutions were calculated for $\theta_b - 1$ and $\theta_b + 1$. The best θ value was then chosen from the solutions for $\{\theta_b - 1, \theta_b, \theta_b + 1\}$.

- Search for θ over the interval $[0, 90°]$;
  - Two-degree incremental search for θ_b,
  - Choose from $\{\theta_b - 1, \theta_b, \theta_b + 1\}$.
- Search for p over the interval $[1, \bar{p}]$;
  - $\Delta_1 = 0.1$ increments for p_1,
  - Fine-tune to four decimal places with increments $\Delta_i = 10^{-i}$ over the interval $p_{i-1} \mp 10^{1-i}, i = 2, 3, 4$.
- Descent Algorithm for the best b_1 and b_2
- Check convexity for $p > 2$.

Display 3.1: Outline of the Search-Descent Algorithm

For each value of θ, an initial p value was calculated by conducting a 0.1 increment search on the interval $[1, \bar{p}]$ where $\bar{p} > 2$. Once a p value was found, it

was fine-tuned to four decimal places of accuracy. This was accomplished by carrying out three further incremental searches. Each search used an interval centered around the best solution from the previous incremental search with a width equal to twice the increment of the previous search. The increment for the search was one-tenth the size of the increment from the previous incremental search. For example, if the best p value for the first search was 1.6, then the interval width for the second search was $[1.50, 1.70]$ with the increment being 0.01. If the best p value was 1.56 this time, then the interval width for the third search was $[1.550, 1.570]$ with the increment being 0.001. We note that, since the SD function is neither convex nor concave in the parameter p, this search procedure does not necessarily find the optimal p. However, similar to the ℓ_p-norm case given by by Brimberg and Love (1991), SD is a smooth function of p, and we conduct a relatively fine grid-search. Additionally, as will be soon in the next section, when compared to the ℓ_p-norm, we obtain better SD values with the ℓ_{bp}-norm. This fact also provides evidence that the search procedure generates fairly close values of p to its optimum.

The best b_1 and b_2 values were calculated by developing a *descent* algorithm given Display 3.2. It is readily seen that if $p \in [1, 2]$ then the SD criterion is convex in its parameters b_1 and b_2. Furthermore, for $p > 2$, if the nonnegativity of M_{ij} is satisfied for each pair of points in the sample, then the SD criterion is again convex in b_1 and b_2. On the other hand, although SD is convex, it is difficult to solve for

optimum b_1, b_2 values by using the first order equations. Therefore we employ the descent algorithm which minimizes $SD(b_1, b_2)$ for given θ and p values.

Input: Sample Data (a_{i1}, a_{i2}), $i = 1, \ldots, n$;
A_{ij}, $i = 1, \ldots, n-1$, $j = 1, \ldots, n$.
Parameters θ and p
Termination Criteria ϵ_λ, ϵ_d.
Output: Optimum b_1, b_2 and SD criterion value.
Initialize Set $b_1^{(1)}$, $b_2^{(1)}$. Assign a large number to $SD^{(0)}$.
Iteration $k = 1$

Compute $SD^{(1)}(b_1^{(1)}, b_2^{(1)})$
while $(SD^{(k-1)} - SD^{(k)})/SD^{(k)} > \epsilon_d$ do begin
Compute Normalized Gradients $(d_1^{(1)}, d_2^{(1)})$
Find $\lambda^{(k)}$ such that
$\min_{\lambda^{(k)} \geq 0} SD\lambda = SD(b_1^{(k)} - \lambda^{(k)} d_1^{(k)}, b_2^{(k)} - \lambda^{(k)} d_2^{(k)})$
by using Golden Section Method with ϵ_λ
Compute new parameters
$b_1^{(k+1)} = b_1^{(k)} - \lambda^{(k)} d_1^{(k)}$
$b_2^{(k+1)} = b_2^{(k)} - \lambda^{(k)} d_2^{(k)}$
Evaluate $SD^{(k+1)}(b_1^{(k+1)}, b_2^{(k+1)})$
$k = k + 1$
end.

Display 3.2: Pseudo-code for Descent Algorithm

Note that the normalized gradient vector $\mathbf{d} = (d_1, d_2)$ is given by

$$\mathbf{d} = \frac{(\nabla SD(b_1), \nabla SD(b_2))}{\| (\nabla SD(b_1), \nabla SD(b_2)) \|} \tag{3.1}$$

where

$$\nabla SD(b_1) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{\partial h_{ij}}{\partial b_1} \quad \text{and} \quad \nabla SD(b_2) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{\partial h_{ij}}{\partial b_2} \tag{3.2}$$

Considering the insignificance of the time required to evaluate the SD we have implemented a Golden Section Method to calculate the step size λ. Clearly SD is a unimodal function of (b_1, b_2) for $p \in [1,2]$. For $p \in (2, \bar{p}]$ the convexity of $SD(b_1, b_2)$ relies on the nonnegativity of M_{ij}. Therefore when $p > 2$ the Search-Descent Algorithm also checks if the nonnegativity of M_{ij} is satisfied at each iteration point $(p, b_1^{[k]}, b_1^{[k]})$ for each pair of points (a_i, a_j) contributing an error term in the SD. We note that, minimization of SD in the four parameters of the distance function (b_1, b_2, p, and θ) is in general a nonconvex optimization problem. Hence, the values found by using the search-descent algorithm may not be optimal. A simulated annealing algorithm may be used to obtain the best-fit values of these parameters simultaneously. This technique provides a probability such that the found solution is optimal (Van Laarhoven and Aarts, 1987).

To determine best-fit b_1 and b_2 values we have tested Quasi-Newton (Broyden-Fletcher-Goldfarb-Shanno; BFGS) and conjugate gradient methods with quadratic-fit line-search as alternatives to the descent algorithms. The details of these methods are given in (Polak, 1971). We have used the implementations provided by Press et al. (1996). These computational procedures performed very similar to the steepest descent algorithm given above. However, we prefer the steepest descent algorithm because of its easy implementation.

3.3 Application Results

In order to model the parameter values of the ℓ_{bp}-norm we used the sample data from seventeen geographical regions presented by Love and Walker (Love and Walker, 1993). The sample data for each geographical region includes 15 point locations based on random selection of *point coordinates* on the map and 105 actual distances corresponding to these points. We applied the Search-Descent Algorithm by using the termination criteria $\epsilon_\lambda = \epsilon_d = 0.001$. The search range for parameter p was taken as $[1, 6]$. The reason for choosing a large search interval for p is to find both sets of parameters, if two sets exist, with minimum SD criterion values.

We present the best parameter values for the ℓ_{bp}-norm applied to seventeen regions in Table 3.1. The corresponding SD criterion values and the percent difference between the criterion values of two ℓ_{bp}-norm best fits, SD_{I} and SD_{II}, are reported in Table 3.2. Furthermore, the $SD(\theta)$ plots for the regions can be found in the Appendix A. The plots of the b_1/b_2 and the p values for the best parameter values are given in Figures 3.3 and 3.4, respectively. The corresponding b_1/b_2 values for the regions can be found in Table 3.3. $\Delta\tau$, which indicates the existence of dominant directional nonlinearity in a transportation network, is also reported for the regions in Table 3.3. $\Delta\tau$ is defined as $|\tau_{\text{I}} - \tau_{\text{II}}|$ where $\tau = \max\{b_1/b_2, b_2/b_1\}$ and τ_{I}, τ_{II} are for SD_{I} and SD_{II}, respectively. $\Delta\tau$ is also utilized to explain the difference between SD_{I} and SD_{II} as discussed in Chapter 2.

		MIN I				MIN II			
No.	Region	θ	b_1	b_2	p	θ	b_1	b_2	p
1	Australia	42	1.4434	1.4443	2.1894	86	1.1945	1.2987	1.7996
2	BC Province	22	2.5828	2.4312	2.7555	66	1.3415	1.4859	1.5635
3	Canada	45	1.9325	1.3500	2.2317	–	–	–	–
4	France	76	1.1529	1.0819	1.8576	–	–	–	–
5	Great Britain	0	1.1116	1.3925	2.0352	–	–	–	–
6	NY State	40	1.5710	1.5630	2.6097	86	1.1303	1.1285	1.5841
7	Pennsylvania	10	1.0751	1.1629	1.6589	44	1.2250	1.4812	2.3892
8	United States	0	1.1358	1.1748	1.6956	42	1.4202	1.4729	2.4485
9	Brussels	3	1.2224	1.2020	2.1787	47	1.0696	1.1177	1.7969
10	London Central	26	1.2366	1.6162	2.2708	–	–	–	–
11	London North	14	1.0930	1.2953	1.7901	52	1.3707	1.4841	2.5324
12	Los Angeles	0	1.1011	1.3799	1.7750	54	1.7276	1.5204	2.7420
13	NY City	6	1.1673	1.2439	1.7539	52	1.5099	1.4367	2.3776
14	Paris	40	1.3634	1.1440	2.2734	–	–	–	–
15	Sydney	8	1.3675	1.1521	1.5571	54	2.1130	2.1541	3.0007
16	Tokyo	13	1.3465	1.3226	2.2249	57	1.1988	1.1670	1.8260
17	Toronto	42	4.1261	4.2427	5.2215	87	1.0581	1.0275	1.2009

Table 3.1: The ℓ_{bp}-norm Parameter Values for Sample Regions

No.	Region	SD_{I}	SD_{II}	$\Delta\%$
1	Australia	1158.05	1117.89	3.59
2	BC Province	1023.19	1015.41	0.77
3	Canada	497.01	–	–
4	France	78.87	–	–
5	Great Britain	172.34	–	–
6	NY State	164.10	159.80	2.69
7	Pennsylvania	104.17	95.11	9.53
8	United States	336.54	348.08	3.43
9	Brussels	3.60	3.47	3.77
10	London Central	15.01	–	–
11	London North	1.36	1.71	26.11
12	Los Angeles	13.18	14.93	13.31
13	NY City	13.29	13.38	0.70
14	Paris	5.84	–	–
15	Sydney	1.10	1.36	23.36
16	Tokyo	2.28	2.28	0.00
17	Toronto	5.06	5.06	0.12

Table 3.2: Best Criterion Values

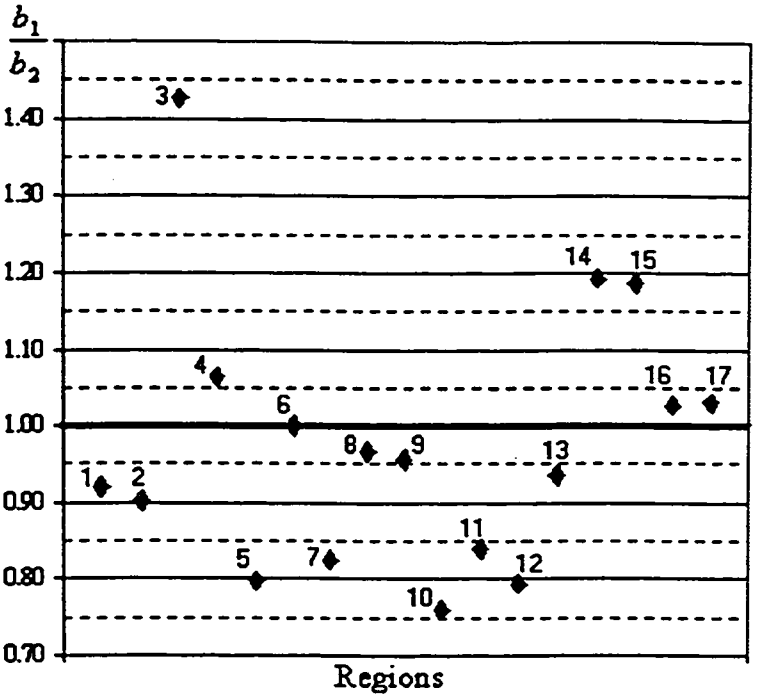


Figure 3.3: Best-fit b_1/b_2 Plots

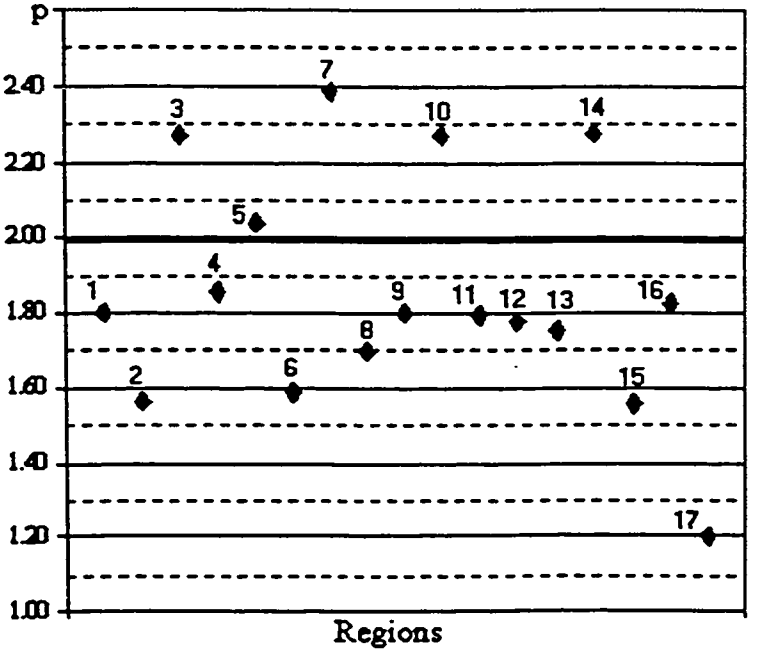


Figure 3.4: Best-Fit p Plots

No.	Region	MIN I b_1/b_2	MIN II b_1/b_2	$\Delta\tau$
1	Australia	0.9994	0.9198	0.0866
2	BC Province	1.0624	0.9028	0.0453
3	Canada	1.4315	–	–
4	France	1.0656	–	–
5	Great Britain	0.7983	–	–
6	NY State	1.0051	1.0016	0.0035
7	Pennsylvania	0.9245	0.8270	0.1275
8	United States	0.9668	0.9642	0.0028
9	Brussels	1.0170	0.9570	0.0280
10	London Central	0.7651	–	–
11	London North	0.8438	0.9236	0.1024
12	Los Angeles	0.7980	1.1363	0.1169
13	NY City	0.9384	1.0510	0.0147
14	Paris	1.1918	–	–
15	Sydney	1.1870	0.9809	0.1675
16	Tokyo	1.0181	1.0272	0.0092
17	Toronto	0.9725	1.0298	0.0016

Table 3.3: b_1/b_2 and $\Delta\tau$ Values

For comparison purposes in Table 3.4 we list the minimum SD criterion values for both the ℓ_{bp}-norm (from Table 3.1) and the weighted ℓ_p-norm where $p \in [1, 2]$ and the axes are rotated properly (Brimberg et al., 1995). Observe that except in New York State where SD has the same value for both norms, the ℓ_{bp}-norm always gives a lower SD criterion value implying that better predictions are obtained. However, the gain in accuracy varies among the regions. The variation is attributed to the underlying pattern of the transportation network. Conclusions can be summarized as follows:

1. In Table 3.1 we observe that for regions 1, 2, 6, 8, 9, 13, 16, and 17 the difference

No.	Region	SD for ℓ_{bp}	SD for $k\ell_p$	$\Delta(SD)\%$
1	Australia	1117.89	1163.59	3.93
2	BC Province	1015.41	1038.72	2.24
3	Canada	496.88	565.61	12.15
4	France	78.86	84.72	6.92
5	Great Britain	172.34	219.42	21.46
6	NY State	159.80	159.80	0.00
7	Pennsylvania	95.11	107.06	11.16
8	United States	336.53	342.68	1.79
9	Brussels	3.47	3.55	2.25
10	London Central	15.01	16.53	9.20
11	London North	1.36	1.78	23.60
12	Los Angeles	13.17	15.50	15.03
13	NY City	13.29	13.58	2.14
14	Paris	5.84	6.52	10.43
15	Sydney	1.10	1.35	18.52
16	Tokyo	2.28	2.30	0.87
17	Toronto	5.06	5.10	0.78

Table 3.4: Comparison of ℓ_{bp}-norm and $k\ell_p$-norm

between the two minimum SD values for the ℓ_{bp}-norm, SD_{I} and SD_{II}, is quite low. A close inspection of Figures 3.3 and 3.4 reveals that the above eight regions have relatively high rectangularity, i.e. p values are not close to 2, and have low directional nonlinearity, i.e. the b_1/b_2 values are close to 1. Low $\Delta\tau$ values show that the nonlinearity in these networks does not have a particular direction. We can say that under these conditions the ℓ_{bp}-norm converges to the weighted ℓ_p-norm. Therefore, similar to the weighted ℓ_p-norm, we have two minimum SD values such that one minimum has an optimum p value where $p \in (1, 2)$ and the other has an optimum p value where $p \in (2, +\infty)$. Moreover, if the b_1/b_2 ratio equals 1 regardless of the axis rotation, we would see these two

minimums exactly 45° apart. In the regions mentioned above, since the b_1 and b_2 values are not exactly equal, there is a minor deviation from 45°. In Table 3.4, we see that the gained accuracy over the weighted ℓ_p-norm is low for these regions. This can again be explained by the similarity of the two norms under the mentioned rectangularity and nonlinearity conditions.

2. In regions 4 and 5 the SD function attains its minimum only once. In Figure 3.4 we see that these regions, France and Great Britain, are highly Euclidean, i.e. p is close to 2. Since the ℓ_2-norm is invariant under axis rotation, in Euclidean transportation networks the variation in $SD(\theta)$ attributed to p is low. Therefore the variation in $SD(\theta)$ is influenced more by the nonlinearity in the transportation network. The study of the directional bias function shows that in such networks $SD(\theta)$ has two bottoms 90° apart. Therefore, for regions 4 and 5, only one minimum is encountered in the $[0, 90°]$ interval. Moreover, since the ℓ_{bp}-norm models the nonlinearity in the transportation network explicitly, it produces more accurate estimations of actual distances. We obtain 6.92% and 21.46% improvements over the best SD values of the weighted ℓ_p-norm for France and Great Britain, respectively.

3. In regions 3, 7, 10, 11, 12, 14, and 15, $SD(\theta)$ either has only one minimum or two minimums with considerably different criterion values.

   i. For regions 3, 10, and 14, in which there is only one minimum SD, the

b_1/b_2 values are far away from 1, implying the existence of dominant directional nonlinearity. Although these regions are relatively rectangular the nonlinearity is clearly a dominating characteristic of the regions. Therefore when the ℓ_{bp}-norm is fitted we observe only one minimum $SD(\theta)$ value in the $[0, 90°]$ interval for these three regions. For Canada, London Central and Paris, using the ℓ_{bp}-norm provides significant improvements in the SD criterion values ; 12.15%, 9.20% and 10.43%, respectively.

ii. For the rest of the regions, in which there are two minimum SD values, it is difficult to differentiate a dominating characteristic, i.e., nonlinearity or rectangularity. While the road networks are highly rectangular, relatively high values of $\Delta\tau$ indicates the existence of directional nonlinearity in the transportation networks. In these regions one of the minimum SD criteria gives the optimum parameter values for the region. The corresponding axis rotation provides the best alignment of the coordinate axes with the underlying directional nonlinearity and rectangularity. The percent difference between the two minimum SD criterion values is quite high; 9.53%, 26.11%, 13.31% and 23.36% for regions 7, 11, 12 and 15, respectively. However, the inherent nonlinearity in the regions is well captured by the ℓ_{bp}-norm. As a consequence we see, once again, considerably better distance predictions. For these regions, Pennsylvania,

London North, Los Angeles and Sydney, the improvements over the SD values given by the weighted ℓ_p-norm are 11.16%, 23.60%, 15.03% and 18.52%, respectively.

As noted earlier, for $p > 2$ the convexity of the SD criterion depends on the nonnegativity of the M_{ij} (shaded region in Figure 3.1). In our Search-Descent Algorithm we check if this condition is satisfied at any iteration with $p > 2$. Note that since we have 15 locations in a sample for a region, each iteration involves 105 violation checks. We present the results in Table 3.5. We report the percentage of

No.	Region	Violations Percentage	Maximum Violations
1	Australia	0.16	3
2	BC Province	0.32	7
3	Canada	0.27	6
4	France	0.00	0
5	Great Britain	0.00	0
6	NY State	0.05	4
7	Pennsylvania	0.23	5
8	United States	0.00	0
9	Brussels	0.00	1
10	London Central	0.38	2
11	London North	0.00	0
12	Los Angeles	0.56	4
13	NY City	0.00	1
14	Paris	0.51	4
15	Sydney	0.01	1
16	Tokyo	0.00	0
17	Toronto	0.17	8

Table 3.5: Convexity Check Results

total number of violations in total number of checks performed in the Search-Descent

Algorithm and the maximum number of violations to the condition observed at any iteration. For example, for Canada there exists an iteration in which for 6 pairs of locations out of 105 the condition is violated. We observe that percentages of violations are very low, below 1% in all regions. For Brussels and NY City, although the maximum number of violations in an iteration is 1, the percentage is very close to zero, and for France, Great Britain, United States, London North and Tokyo, no violations are observed. We note that no violations are observed in the vicinity of the parameter values reported in Table 3.1. We also note that the maximum number of violations are observed for iterations with $p \geq 5$. This is an expected result considering the narrowing shaded region for higher values of p in Figure 3.1. Finally the results given in Table 3.5 provide additional evidence concerning the validity of the assumption that for all practical purposes the SD is a convex function in its parameters b_1 and b_2 for $p \geq 1$.

Chapter 4

Calculation of Confidence Intervals

A confidence interval is calculated by using the estimated distance and it provides the analyst a range in which the actual distance lies with a predetermined level of expectation. Furthermore, the range provides insight about the accuracy of the estimation and the performance of the distance function employed. Another application of confidence intervals for road distance is found in the verification of road distance data. Ginsburgh and Hansen (1974) describe an ad hoc range into which the estimated distance must fall to be acceptable. However, the authors do not provide any analytical justification for the range.

In the literature there are two papers which discuss the statistical properties of errors and confidence intervals for estimated distances. Love, Walker and Tiku (1995) use the weighted ℓ_p-norm ($k\ell_p$) as the distance predicting function and the sum of Normalized Deviations (NAD) as the goodness-of-fit criterion. The authors consider two types of prediction errors: $e(\mathbf{a}_i, \mathbf{a}_j)$ and $e(\mathbf{a}_i, \mathbf{a}_j)/A(\mathbf{a}_i, \mathbf{a}_j)$,

where $e(\mathbf{a}_i, \mathbf{a}_j) = A(\mathbf{a}_i, \mathbf{a}_j) - k\ell_p(\mathbf{a}_i, \mathbf{a}_j)$, and investigate statistical properties for the seventeen geographical regions listed in Chapter 3. By inspecting the sample Pearson coefficients, skewness and kurtosis, they conclude that both types of error distributions are non-normal. On the other hand, while the scatter of points in the $e(\mathbf{a}_i, \mathbf{a}_j)$ versus $A(\mathbf{a}_i, \mathbf{a}_j)$ plots for the regions are contained in a diverging funnel that was symmetric about zero, the scatter of points in the $e(\mathbf{a}_i, \mathbf{a}_j)/A(\mathbf{a}_i, \mathbf{a}_j)$ versus $A(\mathbf{a}_i, \mathbf{a}_j)$ plots are contained within a narrow band which was also symmetric about zero. This observation suggested that the $e(\mathbf{a}_i, \mathbf{a}_j)$ distributions are heteroscedastic but the $e(\mathbf{a}_i, \mathbf{a}_j)/A(\mathbf{a}_i, \mathbf{a}_j)$ distributions are homoscedastic. Using these empirical results it was assumed that $\mu_e = 0$ and $\sigma_e^2 = \sigma^2 A(\mathbf{a}_i, \mathbf{a}_j)^2$ where μ_e is the mean of the $e(\mathbf{a}_i, \mathbf{a}_j)$ distribution, and σ_e^2 and σ^2 are the variances of the $e(\mathbf{a}_i, \mathbf{a}_j)$ and $e(\mathbf{a}_i, \mathbf{a}_j)/A(\mathbf{a}_i, \mathbf{a}_j)$ distributions, respectively. Based on these assumptions, the authors devise the following confidence interval for an unknown distance $A(\mathbf{a}_i, \mathbf{a}_j)$: $c_1 < A(\mathbf{a}_i, \mathbf{a}_j) < c_2$, where $c_1 = (z_1 s + 1)\ k\ell_p(\mathbf{a}_i, \mathbf{a}_j)$ and $c_2 = (z_2 s + 1)\ k\ell_p(\mathbf{a}_i, \mathbf{a}_j)$. The predicted distance $k\ell_p$ and s^2 are assumed to be the unbiased estimators of $A(\mathbf{a}_i, \mathbf{a}_j)$ and σ^2, respectively. The values of z_1 and z_2 are found by using the sample Pearson coefficients of the $e(\mathbf{a}_i, \mathbf{a}_j)$ distribution from the tables provided by Johnson, Nixon and Amos (1963). We should note that although the authors assume that $e(\mathbf{a}_i, \mathbf{a}_j)/A(\mathbf{a}_i, \mathbf{a}_j)$ are homoscedastic by inspecting the scatter plots, a study by Love and Üster (1996) reveals that, except in Australia, BC Province, France and Great

Britain, these distributions are heteroscedastic at the 5% significance level. The scatter of points in the $e(\mathbf{a}_i, \mathbf{a}_j)/A(\mathbf{a}_i, \mathbf{a}_j)$ versus $A(\mathbf{a}_i, \mathbf{a}_j)$ plots form a converging funnel.

In another study Brimberg, Dowling and Love (1994) consider the weighted one-two norm as the distance predicting function. The weighted one-two norm is given by the expression $h(\mathbf{u}, \mathbf{v}; \beta_0, \beta_1) = \beta_0 \ell_2(\mathbf{u}', \mathbf{v}') + \beta_1 \ell_2(\mathbf{u}', \mathbf{v}')$, where $\mathbf{u}, \mathbf{v} \in \Re^2$, $\beta_0, \beta_1 \geq 0$ and $\mathbf{u}', \mathbf{v}'$ are as defined in (1.12). The authors apply the weighted one-two norm model by using the regression R-square value as the goodness-of-fit criterion in Toronto and Ontario. They also consider two types of errors: $e(\mathbf{a}_i, \mathbf{a}_j)$ and $e(\mathbf{a}_i, \mathbf{a}_j)/\ell_2(\mathbf{a}_i, \mathbf{a}_j)$, where $e(\mathbf{a}_i, \mathbf{a}_j) = A(\mathbf{a}_i, \mathbf{a}_j) - h(\mathbf{a}_i, \mathbf{a}_j; \beta_0, \beta_1)$. Both error distributions are assumed to have a zero mean. To test the normality and homoscedasticity of these distributions the authors use normal probability plots and the non-parametric Kolmogorov-Smirnov test, respectively. It is found that the $e(\mathbf{a}_i, \mathbf{a}_j)$ distributions are non-normal and heteroscedastic in both regions. Additionally, the $e(\mathbf{a}_i, \mathbf{a}_j)/\ell_2(\mathbf{a}_i, \mathbf{a}_j)$ distributions are heteroscedastic in both regions and normally distributed in Toronto, but are non-normally distributed in Ontario. Next, the authors analyze the outliers in the $e(\mathbf{a}_i, \mathbf{a}_j)/\ell_2(\mathbf{a}_i, \mathbf{a}_j)$ distributions and find eight unusual observations for Ontario and five unusual observations for Toronto. After excluding the outliers in Ontario and fitting the distance function again, the $e(\mathbf{a}_i, \mathbf{a}_j)$ distribution stays non-normal and heteroscedastic but the $e(\mathbf{a}_i, \mathbf{a}_j)/\ell_2(\mathbf{a}_i, \mathbf{a}_j)$

distribution becomes normal and homoscedastic. For Toronto, excluding only two outliers provides normality and homoscedasticity for the $e(\mathbf{a}_i, \mathbf{a}_j)/\ell_2(\mathbf{a}_i, \mathbf{a}_j)$ distribution. Based on these results, the confidence intervals are devised by using the $e(\mathbf{a}_i, \mathbf{a}_j)/\ell_2(\mathbf{a}_i, \mathbf{a}_j)$ variable excluding outliers. The confidence interval is $c_1 < A(\mathbf{a}_i, \mathbf{a}_j) < c_2$, where $c_1 = h(\mathbf{a}_i, \mathbf{a}_j, \beta_0, \beta_1) - z\, s\, \ell_2(\mathbf{a}_i, \mathbf{a}_j)$ and $c_2 = h(\mathbf{a}_i, \mathbf{a}_j, \beta_0, \beta_1) + z\, s\, \ell_2(\mathbf{a}_i, \mathbf{a}_j)$. The value of $\mp z$ is found from normal probability tables by using $e(\mathbf{a}_i, \mathbf{a}_j)/\ell_2(\mathbf{a}_i, \mathbf{a}_j)$ distribution.

In Section 4.1, we investigate the statistical properties of the estimation errors generated by the ℓ_{bp}-norm and the SD criterion in the seventeen geographical regions. Then, using these properties we develop an error related random variable which is both homoscedastic and normally distributed at the 5% significance level. Since excluding outliers from a region's sample would weaken the sample's representation of the whole region we also aim to obtain this new random variable without excluding the outliers. In Section 4.2, we develop confidence intervals for unknown actual distances by using the new random variable. In Section 4.3 we give some example confidence interval calculations, and compare the the confidence interval calculation method developed in Section 4.2 with the method provided by Love, Walker and Tiku (1995). In Section 4.4, we provide a comparison of the distance predicting accuracy of the weighted ℓ_p-norm and ℓ_{bp}-norm based on their confidence intervals.

4.1 Statistical Properties of Estimation Errors

We define the relationship between the fitted distance and the actual distance as

$$A(\mathbf{x}_i, \mathbf{x}_j) = \ell_{bp}(\mathbf{x}_i, \mathbf{x}_j) + \varepsilon(\mathbf{x}_i, \mathbf{x}_j), \tag{4.1}$$

where $A(\mathbf{x}_i, \mathbf{x}_j)$ is the actual distance between any two points $\mathbf{x}_i$ and $\mathbf{x}_j$, $\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)$ is the predicted distance, and $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ is the related prediction error term. From a random sample of points taken from a geographical region, the empirical distance predicting function parameters are calculated. A computational procedure for finding the parameters of the ℓ_{bp}-norm is given in Chapter 3. Substituting these parameters and the point coordinates into the empirical distance predicting function, an estimate of the actual distance, $\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)$, is obtained. The error $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ for any pair of points may arise from point coordinate measurements errors, inaccurate instrument calibrations, and road network peculiarities that are not captured by the distance model. Since it is a purely random part of the actual distance $A(x_i, x_j)$, $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ is a continuous random variable. We assume that the errors for different pairs of points in a region are independent, i.e., the error of $\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)$ about $A(\mathbf{x}_i, \mathbf{x}_j)$ is not related to the error of $\ell_{bp}(\mathbf{x}_k, \mathbf{x}_l)$ about $A(\mathbf{x}_k, \mathbf{x}_l)$ for any four points $\mathbf{x}_i$, $\mathbf{x}_j$, $\mathbf{x}_k$, $\mathbf{x}_l$ in a geographical region.

In order to empirically examine the statistical properties of the $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ distributions for seventeen regions we use the same sample set that we used in fitting the parameters of the ℓ_{bp}-norm in Chapter 3. To calculate the predicted distances we

use the parameter values which are given by the lower SD values in Table 3.1. These values are reported here in Table 4.1. For our statistical tests we use the statistical analysis package SPSS® (1997).

No.	Region	θ	b_1	b_2	p	SD
1	Australia	86	1.1945	1.2987	1.7996	1117.89
2	BC Province	66	1.3415	1.4859	1.5635	1015.41
3	Canada	45	1.9325	1.3500	2.2317	497.01
4	France	76	1.1529	1.0819	1.8576	78.87
5	Great Britain	0	1.1116	1.3925	2.0352	172.34
6	NY State	86	1.1303	1.1285	1.5841	159.80
7	Pennsylvania	44	1.2250	1.4812	2.3892	95.11
8	United States	0	1.1358	1.1748	1.6956	336.54
9	Brussels	47	1.0696	1.1177	1.7969	3.47
10	London Central	26	1.2366	1.6162	2.2708	15.01
11	London North	14	1.0930	1.2953	1.7901	1.36
12	Los Angeles	0	1.1011	1.3799	1.7750	13.18
13	NY City	6	1.1673	1.2439	1.7539	13.29
14	Paris	40	1.3634	1.1440	2.2734	5.84
15	Sydney	8	1.3675	1.1521	1.5571	1.10
16	Tokyo	57	1.1988	1.1670	1.8260	2.28
17	Toronto	87	1.0581	1.0275	1.2009	5.06

Table 4.1: Best Parameter Values for the ℓ_{bp}-norm

To test the normality of the $\epsilon(\mathbf{x}_i, \mathbf{x}_j)$ distributions we apply the Kolmogorov-Smirnov test with the Lilliefors correction. The details of this test are given by Lilliefors (1967) and Dallal and Wilkinson (1986). The Lilliefors test is a modification of the Kolmogorov-Smirnov test that examines for normality when means and variances are not known, but must be estimated from the data, and it is based on the largest absolute difference between the observed and the expected cumulative distributions. The p-values of the normality tests for the seventeen regions are

reported in Table 4.2 (Note that the parameter p of a distance norm and the p-value of a statistical test are different. We use italics "p" to refer to a distance norm's parameter, and "p-value" for a statistical test's significance level). The p-

No.	Region	p-value	No.	Urban Center	p-value
1	Australia	0.000	9	Brussels	0.077
2	BC Province	0.012	10	London City	0.000
3	Canada	0.064	11	London North	0.200
4	France	0.051	12	Los Angeles	0.200
5	Great Britain	0.000	13	NY City	0.000
6	NY State	0.000	14	Paris	0.009
7	Pennsylvania	0.108	15	Sydney	0.200
8	United States	0.038	16	Tokyo	0.036
			17	Toronto	0.025

Table 4.2: Normality Tests for $\epsilon(\mathbf{x}_i, \mathbf{x}_j)$

values indicate that for most of the cases (10 regions) the $\epsilon(\mathbf{x}_i, \mathbf{x}_j)$ distributions are not normally distributed at the 5% significance level. Although the p-values for seven regions are greater than 5%, four of them are not quite convincing. There is significant evidence that $\epsilon(\mathbf{x}_i, \mathbf{x}_j)$ is normally distributed only in regions 11, 12 and 15.

Besides the Kolmogorov-Smirnov test , normal probability plots and histograms are examined. The related graphs are given in Appendices B and C. On normal probability plots, a linear relation is expected between the observed cumulative probabilities and the expected cumulative probabilities for a sample distribution to be from a normally distributed population. The histograms are expected to have a symmetric bell-shaped appearance with no violations at the tails. The normal probability plots and histograms also confirm that in general $\epsilon(\mathbf{x}_i, \mathbf{x}_j)$ is not normally

distributed.

To examine the homoscedasticity (Wesolowsky, 1976) of $\epsilon(\mathbf{x}_i, \mathbf{x}_j)$, the sample sets of 105 pairs are divided into three groups (short, medium and long) after they are ordered in their increasing order of predicted distances. In order to clarify what is meant by these three groups, Table 4.3 is constructed. In Table 4.3 the means of

No.	Region	Short(S)	Medium(M)	Long(L)	M/S	L/S
1	Australia	1115.74	2116.67	3500.00	1.90	3.14
2	BC Province	298.36	610.78	993.80	2.05	3.33
3	Canada	761.23	2215.23	4262.76	2.91	5.60
4	France	258.16	494.65	724.61	1.92	2.81
5	Great Britain	191.12	389.75	698.02	2.04	3.65
6	NY State	122.83	252.10	413.81	2.05	3.37
7	Pennsylvania	105.01	205.29	367.82	1.95	3.50
8	United States	1080.69	2135.41	3581.48	1.98	3.31
9	Brussels	5.79	10.68	16.37	1.85	2.83
10	London City	4.15	7.75	12.61	1.87	3.04
11	London North	3.55	6.58	11.10	1.85	3.12
12	Los Angeles	9.24	16.82	25.99	1.82	2.81
13	NY State	10.17	18.14	27.74	1.78	2.73
14	Paris	3.26	6.41	9.76	1.97	3.00
15	Sydney	1.22	2.19	3.36	1.80	2.75
16	Tokyo	4.23	7.22	11.22	1.71	2.65
17	Toronto	8.82	15.57	25.90	1.76	2.93

Table 4.3: The Means of the Predicted Distance Groups

short, medium and long predicted distance distributions, and also the ratio of medium to short and long to short predicted distance means are listed. The ratios are similar for all regions except Canada.

To test the homoscedasticity of $\epsilon(\mathbf{x}_i, \mathbf{x}_j)$ we apply Levene test to three groups of distributions. A p-value less than 5% indicates that there is at least one pair

of groups with a significantly different variance. The Levene test is a powerful test when the data come from continuous, but not necessarily normal, distributions (Kotz and Johnson, 1989). The p-values for the tests are reported in Table 4.4. The p-values, based on a 5% significance level, suggest that $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ are heteroscedastic in ten of the regions. The homoscedasticity is observed only in the $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ of seven urban centers. In addition to the Levene test we also inspect the scatter plots of

No.	Region	p-value	No.	Urban Center	p-value
1	Australia	0.000	9	Brussels	0.030
2	BC Province	0.006	10	London City	0.097
3	Canada	0.000	11	London North	0.021
4	France	0.004	12	Los Angeles	0.746
5	Great Britain	0.006	13	NY City	0.514
6	NY State	0.000	14	Paris	0.059
7	Pennsylvania	0.008	15	Sydney	0.123
8	United States	0.008	16	Tokyo	0.464
			17	Toronto	0.059

Table 4.4: The Levene Tests for $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$

errors for our samples. These scatter plots of $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ versus $\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)$ are given in Appendix D. We observe that, as expected, as the predicted distance gets larger the amount of prediction error becomes larger. In other words, the scatter plots of $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ versus $\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)$ form a diverging funnel.

To test whether the mean of $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$, μ_ε, is zero we employ the Student t-test. Although this test assumes that the data are normally distributed, it is fairly robust to departures from normality. The p-values for the test are given in Table 4.5. The p-values suggest that $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$ has zero mean at the 5% significance level in all seventeen

regions. This fact is also observed in the scatter plots where points are evenly scattered about zero.

No.	Region	p-value	No.	Urban Center	p-value
1	Australia	0.498	9	Brussels	0.585
2	BC Province	0.237	10	London City	0.146
3	Canada	0.626	11	London North	0.642
4	France	0.695	12	Los Angeles	0.262
5	Great Britain	0.544	13	NY City	0.418
6	NY State	0.501	14	Paris	0.315
7	Pennsylvania	0.512	15	Sydney	0.485
8	United States	0.706	16	Tokyo	0.564
			17	Toronto	0.570

Table 4.5: The Student t-test for $\mu_\epsilon = 0$

In general, we can say that $\epsilon(\mathbf{x}_i, \mathbf{x}_j)$ is non-normal, heteroscedastic, and its mean is equal to zero at the 5% significance level. In order to develop a confidence interval for an unknown actual distance we need a random variable which is *homoscedastic*. Therefore we need to use a transformation of $\epsilon(\mathbf{x}_i, \mathbf{x}_j)$. For that purpose, we define the following transformed random variable:

$$\epsilon_t(\mathbf{x}_i, \mathbf{x}_j) = \frac{\epsilon(\mathbf{x}_i, \mathbf{x}_j)}{(\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j))^{1/t}}, \quad t \geq 1, \tag{4.2}$$

where parameter t is to be determined for a given transportation network so that the corresponding $\epsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ is homoscedastic and normally distributed.

Before proceeding with our analysis for finding the best t value for a transportation network, we will first assume that $t = 1$ and investigate if the normalized error $\epsilon_1(\mathbf{x}_i, \mathbf{x}_j)$ is homoscedastic. This special form of $\epsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ is very

similar to the random variable used by Love, Walker and Tiku (1995). In order to test homoscedasticity we apply the Levene test in the same way that we used it for $\epsilon(\mathbf{x}_i, \mathbf{x}_j)$. In Table 4.6 we report the p-values. The scatter plots of $\epsilon_1(\mathbf{x}_i, \mathbf{x}_j)$ with respect to increasing values of the ℓ_{bp} distance are given in Appendix E. The p-values show that, except in Australia, BC Province, Great Britain, New York State and Paris, the $\epsilon_1(\mathbf{x}_i, \mathbf{x}_j)$ is heteroscedastic at the 5% significance level. Furthermore, for BC Province and Paris the p-values are very close to 5%. Inspecting the scatter plots in Appendix E we see that the scatter of points form a converging funnel, i.e., in general the percentage error is smaller for relatively long distances and larger for relatively short distances.

No.	Region	p-value	No.	Urban Center	p-value
1	Australia	0.5218	9	Brussels	0.0287
2	BC Province	0.0723	10	London City	0.0006
3	Canada	0.0000	11	London North	0.0004
4	France	0.0099	12	Los Angeles	0.0006
5	Great Britain	0.1659	13	NY City	0.0001
6	NY State	0.4847	14	Paris	0.0574
7	Pennsylvania	0.0153	15	Sydney	0.0155
8	United States	0.0026	16	Tokyo	0.0042
			17	Toronto	0.0200

Table 4.6: The Levene Tests for $\epsilon_1(\mathbf{x}_i, \mathbf{x}_j)$

We now turn our attention to the determination of the best t for a given network. For that purpose we generate samples of $\epsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ for $t = 1, \ldots, 3$ with 0.1 increments. In Table 4.7 we list the ranges of t in which the $\epsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ are homoscedastic and normally distributed. A "✓" represents statistically significant test results at the

No.	Region	t Range	t	Homoscedasticity	Normality
1	Australia	1.0–2.3	1.0	✓	X
2	BC Province	1.4–2.6	2.0	✓	✓
3	Canada	1.7–3.0	2.0	✓	✓
4	France	1.5–2.1	2.0	✓	✓
5	Great Britain	1.0–3.0	1.0	✓	X
6	NY State	1.0	1.0	✓	✓
7	Pennsylvania	1.2–3.0	2.0	✓	✓
8	United States	1.4–3.0	2.0	✓	✓
9	Brussels	1.5–3.0	2.0	✓	✓
10	London City	1.4–3.0	2.0	✓	X
11	London North	1.4–3.0	2.0	✓	✓
12	Los Angeles	1.8–3.0	2.0	✓	✓
13	NY City	1.7–3.0	2.0	✓	X
14	Paris	1.3–2.8	2.0	✓	✓
15	Sydney	1.2–3.0	2.0	✓	✓
16	Tokyo	1.6–3.0	2.0	✓	X
17	Toronto	1.2–3.0	2.0	✓	X

Table 4.7: Homoscedasticity and Normality for Ranges of t

5% significance level whereas an "X" represents otherwise. We note that in the United States and London City homoscedasticity could not be obtained for $t \in [1, 3]$. Therefore we exclude the actual distances which correspond to the outliers in the samples of $\varepsilon_1(\mathbf{x}_i, \mathbf{x}_j)$ for these regions and model the ℓ_{bp}-norm with the new reduced sample sets. The new samples provided well-behaved distributions for both regions. The number of outliers in the United States case was five and in the London City case it was nine.

In Table 4.7 we observe that homoscedasticity is obtained for $t = 2$ in all regions, except in the NY City case. However, if $t = 1$, then both homoscedasticity and normality are obtained for NY City. Based on our samples we devise the following rule

of thumb that can be used in obtaining a transformed random variable for confidence interval calculation purposes: $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ with $t = 1$ is tested for homoscedasticity and normality. Table 4.7 indicates that both conditions are not very likely to occur in a particular region. If not, then $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ with $t = 2$ is tested for homoscedasticity and normality. Based on our samples it is very likely that a well-behaved distribution will be obtained. In the case that neither $t = 1$ nor $t = 2$ provides a homoscedastic and normal distribution then other values of t should be considered. However, if all fails, then the outliers should be excluded and the same checks should be performed with $t = 1$ and $t = 2$. The t values determined by using this rule of thumb for seventeen geographical regions are also reported in Table 4.7. We note that, for purposes of confidence interval calculations, homoscedasticity is more important than normality. In order to obtain a reliable confidence interval the distribution must have a constant variance. Otherwise, one can easily obtain a confidence interval with a lower or higher level of expectation than the predetermined value, say 95%. Therefore, if only a homoscedastic but non-normal $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ is obtained before excluding the outliers, this distribution can be efficiently used to calculate the confidence interval as shown in the next section. In other words, in order not to weaken the sample's representation of the whole region, excluding outliers should be seen as a last resort in the course of obtaining a well-behaved error distribution.

4.2 Development of Confidence Intervals

In this section we develop the confidence interval for an estimated distance. For that purpose we use the transformed random variable $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ where $t \geq 1$.

Let μ_{ε_t} and $\sigma^2_{\varepsilon_t}$ represent the mean and the variance of $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$, herein denoted ε_t. Also let (e_{t1}, e_{t2}) be the $100(1-\alpha)\%$ confidence interval for $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ where $0 < \alpha < 1$. We have

$$\Pr(e_{t1} < \varepsilon_t < e_{t2}) = 1 - \alpha, \quad 0 < \alpha < 1. \tag{4.3}$$

In our development we will assume that the confidence interval is symmetric, i.e., $\Pr(\varepsilon_t < e_{t1}) = \Pr(\varepsilon_t > e_{t2}) = \alpha/2$. The standardized values of e_{t1} and e_{t2} are given by z_{t1} and z_{t2} where

$$z_{t1} = \frac{e_{t1} - \mu_{\varepsilon_t}}{\sigma_{\varepsilon_t}} \quad \text{and} \quad z_{t2} = \frac{e_{t2} - \mu_{\varepsilon_t}}{\sigma_{\varepsilon_t}}. \tag{4.4}$$

Rearranging (4.4) and substituting into (4.3) we have

$$\Pr(\mu_{\varepsilon_t} + z_{t1}\,\sigma_{\varepsilon_t} < \varepsilon_t < \mu_{\varepsilon_t} - z_{t2}\,\sigma_{\varepsilon_t}) = 1 - \alpha, \tag{4.5}$$

or equivalently

$$\Pr(z_{t1}\,\sigma_{\varepsilon_t} < \varepsilon_t - \mu_{\varepsilon_t} < z_{t2}\,\sigma_{\varepsilon_t}) = 1 - \alpha. \tag{4.6}$$

We assume that $\mu_{\varepsilon_t} = 0$ and the sample size $n = 105$ is large enough for the sample standard deviation to be a close approximation of σ_{ε_t}. (4.6) can be written as

$$\Pr(z_{t1}\,\sigma_{\varepsilon_t} < \varepsilon_t < z_{t2}\,\sigma_{\varepsilon_t}) = 1 - \alpha$$

$$\Pr(z_{t1}\,\sigma_{\epsilon_t} < \frac{A(\mathbf{x}_i,\mathbf{x}_j) - \ell_{bp}(\mathbf{x}_i,\mathbf{x}_j)}{(\ell_{bp}(\mathbf{x}_i,\mathbf{x}_j))^{1/t}} < z_{t2}\,\sigma_{\epsilon_t}) = 1-\alpha. \tag{4.7}$$

By rearranging 4.7 we find the confidence interval for an unknown actual distance between any two points $\mathbf{x}_i$ and $\mathbf{x}_j$ as

$$\Pr(c_{t1} < A(\mathbf{x}_i,\mathbf{x}_j) < c_{t2}) = 1-\alpha, \quad 0<\alpha<1. \tag{4.8}$$

where

$$\begin{aligned} c_{t1} &= \ell_{bp}(\mathbf{x}_i,\mathbf{x}_j)\,\left[1 + z_{t1}\,\sigma_{\epsilon_t}\,(\ell_{bp}(\mathbf{x}_i,\mathbf{x}_j))^{(1/t)-1}\right], \\ c_{t2} &= \ell_{bp}(\mathbf{x}_i,\mathbf{x}_j)\,\left[1 + z_{t2}\,\sigma_{\epsilon_t}\,(\ell_{bp}(\mathbf{x}_i,\mathbf{x}_j))^{(1/t)-1}\right]. \end{aligned} \tag{4.9}$$

If $\epsilon_t(\mathbf{x}_i,\mathbf{x}_j)$ is normally distributed, then the standardized values z_{t1} and z_{t2} are found in the standard normal distribution table. Since we use symmetry, the values of z_{t1} and z_{t2} are of equal magnitude but opposite sign. For example, if a 95% confidence interval is considered, then the standardized values are $z_t = \mp 1.96$. However, if the distribution is not known, then the standardized values z_{t1} and z_{t2} are found in the tables provided by Johnson, Nixon and Amos (1963). These tables utilize the Pearson coefficients (skewness and kurtosis) of the distribution and list the standardized values for several values of $\alpha/2$. For large samples ($n \geq 100$) the sample Pearson coefficients can be used as unbiased estimates of the population Pearson coefficients (Stuart and Ord, 1987).

4.3 Some Examples

In this section, we demonstrate the use of (4.9) by constructing the confidence intervals for six unknown actual distances. We use the regions and the point locations used in constructing example confidence intervals with the weighted ℓ_p-norm by Love, Walker and Tiku (1995). The regions, point locations (in centimeters), standard deviation of sample $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ distributions, and the standardized values z_{t1} and z_{t2} are given in Table 4.8. For four of the regions the $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ are normally distributed

Region	t	$\mathbf{x}_i$	$\mathbf{x}_j$	σ_{ε_t}	z_{t1}	z_{t2}
Australia	1	(15.50,18.65)	(43.00,33,90)	0.0786	-1.8258	2.1689
Canada	2	(106.65,24.50)	(19.10,13.10)	2.1869	-1.9600	1.9600
United States	2	(4.50,28.20)	(41.00,16.20)	1.8343	-1.9600	1.9600
Brussels	2	(19.30,25.55)	(38.30,7.35)	0.1862	-1.9600	1.9600
London North	2	(10.30,2.60)	(38.05,24.60)	0.1145	-1.9600	1.9600
Toronto	2	(54.45,41.85)	(49.05,17.30)	0.2182	-2.1149	1.8739

Table 4.8: Data for Example Confidence Interval Calculations

(Table 4.7). Therefore, to construct 95% confidence intervals the values of z_{t1} and z_{t2} are taken as -1.96 and 1.96, respectively. However, for Australia and Toronto, $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ is not normally distributed. Thus, to determine the standardized values we have to use the tables provided by Johnson, Nixon and Amos (1963). For Australia the sample Pearson coefficients are $\sqrt{b_1} = 0.4800$ and $b_2 = 4.7524$, and for Toronto they are $\sqrt{b_1} = -0.4086$ and $b_2 = 5.6495$. Using the interpolation technique as described by Johnson, Nixon and Amos we calculate the standardized values given in Table 4.8. In Table 4.9 we report the ℓ_{bp} distances (in kilometers) and the corresponding 95%

confidence intervals.,

Region	$\ell_{bp}(x_i, x_j)$	(c_{t1}, c_{t2})
Australia	3028.47	(2593.86,3544.75)
Canada	3130.36	(2890.54,3370.17)
United States	3174.07	(2971.56,3376.57)
Brussels	8.80	(7.72,9.88)
London North	6.64	(6.06,7.22)
Toronto	10.92	(9.40,12.27)

Table 4.9: Example Confidence Intervals (kilometers)

To compare the confidence interval calculation method developed in the previous section with the Love-Walker-Tiku method, we first calculate the confidence intervals (c_1, c_2) for the ℓ_{bp} distances by using the method and the parameters given by Love, Walker and Tiku (1995). Then we compare the range of these confidence intervals with the range of confidence interval calculated by our method. These comparisons are presented in Table 4.10. We observe that, except in Australia and Brussels, the

Region	$\ell_{bp}(x_i, x_j)$	(c_1, c_2)	$c_2 - c_1$	$c_{t2} - c_{t1}$
Australia	3028.47	(2565.39,3499.99)	934.60	950.89
Canada	3130.36	(2597.32,3484.85)	887.52	479.63
United States	3174.07	(2942.72,3474.45)	531.73	405.01
Brussels	8.80	(7.85,9.95)	2.10	2.16
London North	6.64	(5.92,7.38)	1.46	1.16
Toronto	10.92	(9.25,12.17)	2.92	2.88

Table 4.10: Comparison of Example Confidence Intervals

confidence intervals obtained by the method developed here are smaller than the ones obtained by the Love-Walker-Tiku method. Therefore we next would like to explore this difference.

Let $\mathcal{L}$ be a predicted distance, $\mathcal{I}$ and $\mathcal{I}_t$ be the width of the confidence intervals calculated by the Love-Walker-Tiku method and our method, respectively. Then the expressions for $\mathcal{I}$ and $\mathcal{I}_t$ are

$$\mathcal{I} = \mathcal{L}(1+z_2 s) - \mathcal{L}(1+z_1 s) \tag{4.10}$$

$$\mathcal{I}_t = \mathcal{L}(1+z_{t2}\sigma_{\epsilon_t}\mathcal{L}^{(1/t)-1}) - \mathcal{L}(1+z_{t1}\sigma_{\epsilon_t}\mathcal{L}^{(1/t)-1}) \tag{4.11}$$

It can easily be verified that $\mathcal{I}$ and $\mathcal{I}_t$ are increasing functions of $\mathcal{L}$ and for $t > 1$

$$\begin{aligned} &\mathcal{I} < \mathcal{I}_t \quad \text{for} \quad \mathcal{L} < \mathcal{L}_c \\ &\mathcal{I} = \mathcal{I}_t \quad \text{for} \quad \mathcal{L} = \mathcal{L}_c \\ &\mathcal{I} > \mathcal{I}_t \quad \text{for} \quad \mathcal{L} > \mathcal{L}_c \end{aligned} \tag{4.12}$$

where

$$\mathcal{L}_c = \left(\frac{\sigma_{\epsilon_t}(z_{t2}-z_{t1})}{s(z_2-z_1)}\right)^{t/(t-1)}, \tag{4.13}$$

and for $t = 1$

$$\begin{aligned} &\mathcal{I} < \mathcal{I}_t \quad \text{for} \quad \mathcal{R} > 1 \\ &\mathcal{I} = \mathcal{I}_t \quad \text{for} \quad \mathcal{R} = 1 \\ &\mathcal{I} > \mathcal{I}_t \quad \text{for} \quad \mathcal{R} < 1 \end{aligned} \tag{4.14}$$

where

$$\mathcal{R} = \frac{\sigma_{\epsilon_t}(z_{t2}-z_{t1})}{s(z_2-z_1)}. \tag{4.15}$$

(4.12) implies that, for $t > 1$, $\mathcal{I}$ and $\mathcal{I}_t$ are equal for a critical $\mathcal{L}$ value given by (4.13). This fact is observed in the $\mathcal{I}$ and $\mathcal{I}_t$ graphs which are presented in Figure 4.1 for the six example regions. For values of $\mathcal{L}$ greater than $\mathcal{L}_c$ the range of the confidence interval obtained by the Love-Walker-Tiku method is larger than the one calculated by our method and for values of $\mathcal{L}$ less than $\mathcal{L}_c$ the converse holds. Also note that, the difference between the intervals becomes larger, favouring $\mathcal{I}_t$, as the predicted distance increases. For example in Toronto, for the predicted distance of 10.92 kms. the difference between the two intervals is 0.04 (2.92 vs. 2.88 in Table 4.10). If the predicted distance was 30 kms., the difference would be 3.31 (8.01 vs. 4.70). On the other hand, (4.14) implies that, for $t = 1$, $\mathcal{I}$ and $\mathcal{I}_t$ are equal if $\mathcal{R}$ is equal to 1. If $\mathcal{R}$ is less than 1, then the confidence interval calculated by the Love-Walker-Tiku method is larger than the one calculated by our method for any distance $\mathcal{L}$. The converse holds for values of $\mathcal{R}$ greater than 1.

The values of t, ℓ_{bp} distances, $\mathcal{L}_c$ and $\mathcal{R}$ that we use for comparisons are given in Table 4.11. Observe that for regions with $t = 2$ the predicted distances are longer than

Region	t	$\ell_{bp}(\mathbf{x}_i, \mathbf{x}_j)$	$\mathcal{L}_c$ or $\mathcal{R}$
Australia	1	3028.47	1.017
Canada	2	3130.36	913.54
United States	2	3174.07	1842.33
Brussels	2	8.80	9.44
London North	2	6.64	4.16
Toronto	2	10.92	10.61

Table 4.11: Example $\mathcal{L}_c$ and $\mathcal{R}$ Values

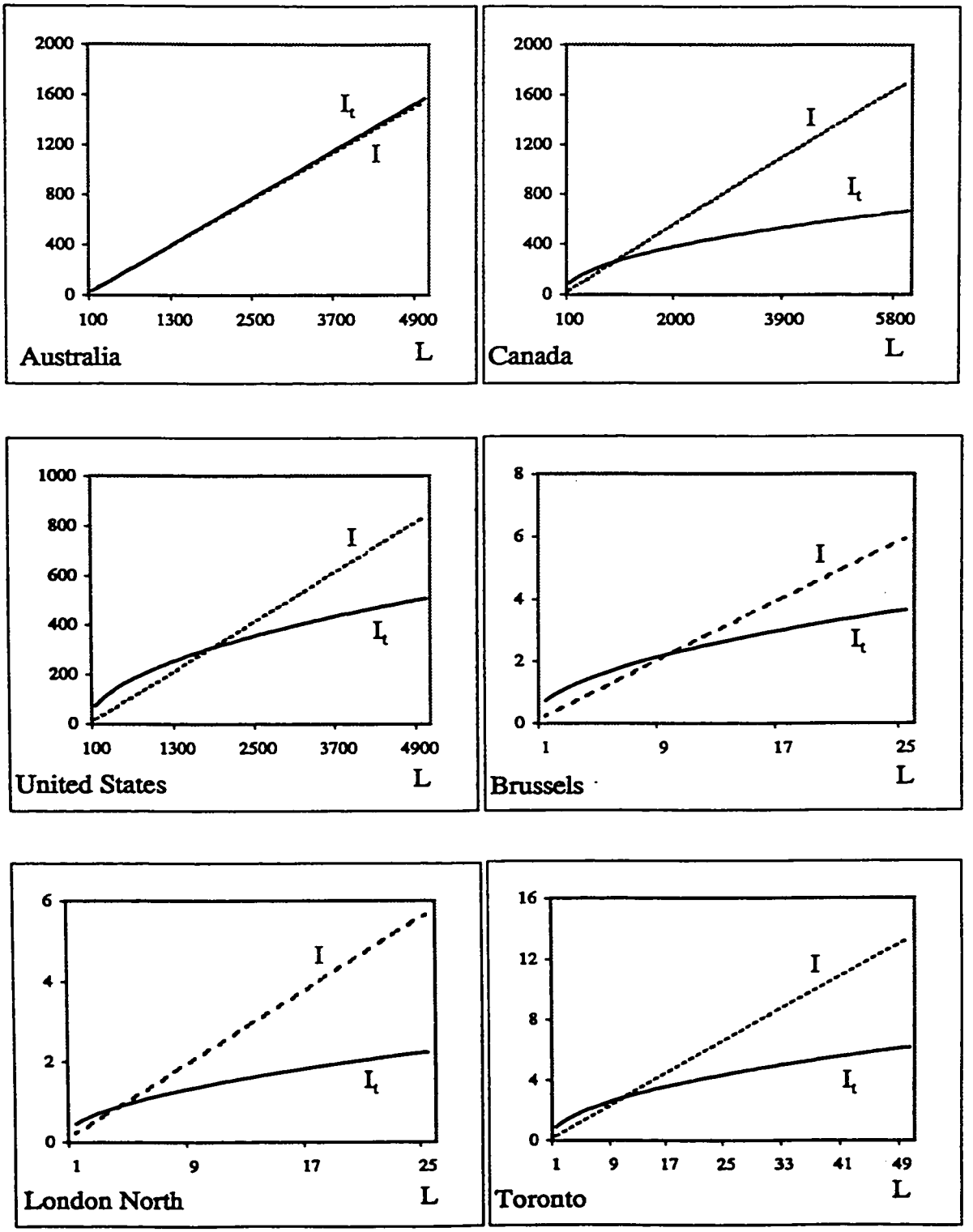


Figure 4.1: Example Confidence Interval Comparisons

their corresponding critical values $\mathcal{L}_c$, except in Brussels in which case the predicted distance is shorter than $\mathcal{L}_c$. Therefore, as it is already given in Table 4.10, we obtain smaller confidence intervals for four of the regions, and a larger interval for Brussels. For Australia, where $t = 1$, the $\mathcal{R}$ value is slightly greater than 1, i.e., $\mathcal{I}$ and $\mathcal{I}_t$ values are very close, $\mathcal{I}$ being smaller. This is because for Australia t is equal to 1 and thus $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ is indeed quite similar to the transformed random variable used in the Love-Walker-Tiku method. Furthermore, for Australia the weighted ℓ_{bp}-norm performs similar to the weighted ℓ_p-norm because of the low directional nonlinearity and high rectangularity in the transportation network as discussed in Section 3.3.

In Table 4.12 the ranges of predicted distances that represent short, medium and long distance groups for example regions are listed. Notice that, except for

Region	Short	Medium	Long
Australia	438–1532	1539–2689	2699–4609
Canada	204–1265	1266–3225	3288–5844
United States	165–1661	1674–2717	2720–4841
Brussels	2.7– 8.4	8.6 –13.2	13.7–22.9
London North	1.5– 5.0	5.1 – 8.1	8.1 –16.7
Toronto	3.9–11.6	11.8–20.2	20.2–43.8

Table 4.12: Ranges of the Predicted Distance Groups

Australia, $\mathcal{L}_c$ values are close to the cut-off point between short and medium distance groups. This observation suggests that the confidence intervals calculated by the Love-Walker-Tiku method are generally different than what they should be at the 95% level. They are smaller for relatively short distances and larger for relatively long

distances. This is indeed an expected result because, as we have mentioned earlier, except in Australia, BC Province, France and Great Britain, the scatter of points in the $e(\mathbf{a}_i, \mathbf{a}_j)/A(\mathbf{a}_i, \mathbf{a}_j)$ versus $A(\mathbf{a}_i, \mathbf{a}_j)$ plots form a converging funnel (Love and Üster, 1996). In other words, the random variable used in developing confidence intervals is not homoscedastic, but has a higher standard deviation for relatively short distances and a lower standard deviation for relatively long distances. The percentage error generated in predicting long distances is generally lower than the percentage error generated in predicting short distances. We have observed the same phenomenon in predicting distances with the ℓ_{bp}-norm, i.e., $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ with $t = 1$ is generally heteroscedastic. However, the confidence interval calculation method developed in this chapter accounts for this phenomenon and provides confidence intervals at the intended level of expectation.

4.4 Comparison of Distance Functions

In this section we compare the 95% confidence intervals for the weighted ℓ_p and the ℓ_{bp} distances modelled using SD as the goodness-of-fit criterion. This, in turn, provides a comparison of accuracy in distance prediction with the weighted ℓ_p-norm and the ℓ_{bp}-norm. We will use the transformed random variable $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ for that purpose, and adopt the notation t' and $\varepsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$ for the weighted ℓ_p-norm related distributions. To generate the sample $\varepsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$ distributions for each region we use the best parameter values of the weighted ℓ_p-norm, k, p and θ, where $p \in (1, 2)$. These

empirical parameter values for seventeen geographical regions are computed by Love and Walker (1994) for the SD criterion and are presented here in Table 4.13.

No.	Region	θ	k	p
1	Australia	0	1.1460	1.8585
2	BC Province	68	1.2495	1.5609
3	Canada	83	1.1715	1.4849
4	France	70	1.0609	1.8430
5	Great Britain	40	1.1095	1.7895
6	NY State	86	1.0794	1.5823
7	Pennsylvania	4	1.0611	1.6244
8	United States	0	1.0792	1.6641
9	Brussels	47	1.0549	1.8180
10	London City	72	1.1182	1.9241
11	London North	11	1.0599	1.6456
12	Los Angeles	2	1.0721	1.5734
13	NY City	6	1.1069	1.7340
14	Paris	86	1.0613	1.8189
15	Sydney	8	1.1266	1.4719
16	Tokyo	58	1.0963	1.8252
17	Toronto	88	1.0279	1.1863

Table 4.13: Best Parameter Values for the Weighted ℓ_p-norm

We first find the best t' values which provide homoscedasticity of $\epsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$. As in the ℓ_{bp}-norm case we test the homoscedasticity and the normality of sample $\epsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$ distributions for values of t' ranging from 1.0 to 3.0 with increments 0.1. The summary of these test results, including the t' values that are chosen by the rule of thumb devised in Section 4.1, is given in Table 4.14. Note that, for $t' \in [1, 3]$, homoscedasticity and normality could not be obtained for the $\epsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$ distributions of the United States and London City. This was also the case for the sample $\epsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ distributions of the same regions. Therefore, for comparison purposes, we will not

consider these two regions. For the rest of the regions, inspecting t and t' columns

No.	Region	t' Range	t'	Homoscedasticity	Normality
1	Australia	1.0–1.9	1.0	✓	X
2	BC Province	1.2–3.0	2.0	✓	✓
3	Canada	2.1–3.0	2.1	✓	✓
4	France	1.4–2.1	2.0	✓	X
5	Great Britain	1.1–3.0	2.0	✓	X
6	NY State	1.0	1.0	✓	✓
7	Pennsylvania	1.1–3.0	2.0	✓	✓
8	United States	1.0–3.0	-	X	X
9	Brussels	1.3–3.0	2.0	✓	✓
10	London City	1.0–3.0	-	X	X
11	London North	1.3–3.0	2.0	✓	✓
12	Los Angeles	1.7–3.0	2.0	✓	✓
13	NY City	1.6–3.0	2.0	✓	X
14	Paris	1.0–3.0	1.0	✓	✓
15	Sydney	1.7–3.0	2.0	✓	✓
16	Tokyo	1.6–3.0	2.0	✓	X
17	Toronto	1.2–3.0	2.0	✓	X

Table 4.14: Homoscedasticity and Normality for Ranges of t'

in Tables 4.7 and 4.14 we see that the chosen values of t and t' are identical, except in regions Canada, Great Britain and Paris. In order to obtain a unified comparison scheme we would like to have equal t and t' values, if possible. Thus, without violating the homoscedasticity requirement of the distributions, we can take $t = 2.1$ for Canada, $t = 2$ for Great Britain, and $t' = 2$ for Paris. The final t and t' values, sample standard deviations, and standardized values for the $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ and $\varepsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$ distributions are reported in Table 4.15. The columns σ'_{ε_t}, z'_{t1} and z'_{t2} represent the sample standard deviation and standardized values related to the sample $\varepsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$ distributions. For the error distributions in which normality can not

be obtained we calculate the standardized values by interpolation using the sample Pearson coefficients of $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ and $\varepsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$, and the biometrika tables of Johnson, Nixon and Amos (1963). We note that the skewness of the sample $\varepsilon_t(\mathbf{x}_i, \mathbf{x}_j)$ and $\varepsilon_{t'}(\mathbf{x}_i, \mathbf{x}_j)$ distributions for Tokyo were 2.3707 and 2.4243, respectively. These values are out of the range of skewness values in the biometrika tables where the range is $(0.0, 2.0)$. Thus, the values of t and t' are taken as 3.3 so that the standardized values can be calculated and both distributions are still homoscedastic.

No	Region	t', t	σ'_{ε_t}	z'_{t1}	z'_{t2}	σ_{ε_t}	z_{t1}	z_{t2}
1	Australia	1.0	0.0796	-1.7625	2.2002	0.0786	-1.8258	2.1689
2	BC Province	2.0	3.2817	-1.9600	1.9600	3.2441	-1.9600	1.9600
3	Canada	2.1	2.7559	-1.9600	1.9600	2.6074	-1.9600	1.9600
4	France	2.0	0.9215	-1.4853	2.3741	0.8884	-1.9600	1.9600
5	Great Britain	2.0	1.5024	-1.4092	2.4208	1.3483	-1.2548	2.5245
6	NY State	1.0	0.0756	-1.9600	1.9600	0.0756	-1.9600	1.9600
7	Pennsylvania	2.0	1.0135	-1.9600	1.9600	0.9569	-1.9600	1.9600
9	Brussels	2.0	0.1883	-1.9600	1.9600	0.1862	-1.9600	1.9600
11	London North	2.0	0.1315	-1.9600	1.9600	0.1146	-1.9600	1.9600
12	Los Angeles	2.0	0.3676	-1.9600	1.9600	0.3356	-1.9600	1.9600
13	NY City	2.0	0.3818	-1.2668	2.4895	0.3806	-1.1612	2.5634
14	Paris	2.0	0.2433	-1.9600	1.9600	0.2319	-1.9600	1.9600
15	Sydney	2.0	0.1182	-1.9600	1.9600	0.1050	-1.9600	1.9600
16	Tokyo	3.3	0.2236	-1.2693	2.5147	0.2231	-1.2452	2.5274
17	Toronto	2.0	0.2191	-2.1218	1.8648	0.2182	-2.1149	1.8739

Table 4.15: Data for Confidence Interval Comparisons

To compare the two distance functions we will again utilize the range of the confidence intervals. Let $\mathcal{L}$ be any predicted distance as before, and $\mathcal{I}_t$, $\mathcal{I}'_t$ be the ranges of the confidence intervals calculated for an ℓ_{bp} and a weighted ℓ_p distance,

respectively. Then using (4.9) $\mathcal{I}_t$ and $\mathcal{I}'_t$ for the same predicted distance are given by

$$\mathcal{I}_t = \mathcal{L}(1 + z_{t2}\sigma_{\epsilon_t}\mathcal{L}^{(1/t)-1}) - \mathcal{L}(1 + z_{t1}\sigma_{\epsilon_t}\mathcal{L}^{(1/t)-1}) \tag{4.16}$$

$$\mathcal{I}'_t = \mathcal{L}(1 + z'_{t2}\sigma'_{\epsilon_t}\mathcal{L}^{(1/t')-1}) - \mathcal{L}(1 + z'_{t1}\sigma'_{\epsilon_t}\mathcal{L}^{(1/t')-1}). \tag{4.17}$$

It readily follows that the intervals $\mathcal{I}_t$ and $\mathcal{I}'_t$ are increasing functions of $\mathcal{L}$ and for $t = t'$ we have

$$\begin{aligned} \mathcal{I}_t < \mathcal{I}'_t \quad &\text{if} \quad \mathcal{T} < 1 \\ \mathcal{I}_t = \mathcal{I}'_t \quad &\text{if} \quad \mathcal{T} = 1 \\ \mathcal{I}_t > \mathcal{I}'_t \quad &\text{if} \quad \mathcal{T} > 1 \end{aligned} \tag{4.18}$$

where

$$\mathcal{T} = \frac{\sigma_{\epsilon_t}(z_{t2} - z_{t1})}{\sigma'_{\epsilon_t}(z'_{t2} - z'_{t1})}. \tag{4.19}$$

The values of $\mathcal{T}$ for 15 regions used in the comparisons are calculated and reported in Table 4.16. Observe that, except in NY State, The $\mathcal{T}$ values are always less than 1, indicating that the confidence intervals calculated for the ℓ_{bp} distances are smaller indicating the superiority of the ℓ_{bp}-norm over the weighted ℓ_p-norm. The proximity of the $\mathcal{T}$ value to 1 determines the level of gained prediction accuracy by using the ℓ_{bp}-norm versus the weighted ℓ_p-norm. The gain in prediction accuracy increases as the $\mathcal{T}$ value becomes smaller. In Chapter 3 we have compared the performance of the ℓ_{bp}-norm and the weighted ℓ_p-norm in predicting distances by using the percent decrease in the SD value of the weighted ℓ_p-norm (Table 3.4). Since the ℓ_{bp}-norm

No.	Region	$\mathcal{T}$	No.	Urban Center	$\mathcal{T}$
1	Australia	0.9954	9	Brussels	0.9888
2	BC Province	0.9885	11	London North	0.8715
3	Canada	0.9461	12	Los Angeles	0.9129
4	France	0.9792	13	NY City	0.9884
5	Great Britain	0.8856	14	Paris	0.9531
6	NY State	1.0000	15	Sydney	0.8883
7	Pennsylvania	0.9442	16	Tokyo	0.9948
			17	Toronto	0.9964

Table 4.16: $\mathcal{T}$ Values

models the directional non-linearity in a transportation network explicitly, we have observed that the gain increases as the directional non-linearity increases. In Table 3.4 we see that a low gain in prediction accuracy (below 7%) is obtained for regions 1, 2, 4, 6, 9, 13, 16 and 17, a moderate gain ($10 - 12\%$) is obtained for regions 3, 7 and 14, and a high gain (above 15%) is obtained for regions 5, 11, 12 and 15. A close inspection of Table 4.16 reveals that the same groups of regions are formed based on the proximity of $\mathcal{T}$ values to unity.

Chapter 5

Minisum Location Models

We have previously defined the single- and multi-facility minisum location models in their general forms by formulations (1.1) and (1.2), respectively. As is readily seen in these formulations, distance predicting functions are an important part of the objective function of a continuous location model. Since the model should represent the real situation as closely as possible, the accuracy of the distance predicting function employed plays a crucial role in terms of the validity and the applicability of the locational decisions. Since we obtain better distance representations with the ℓ_{bp}-norm as presented in the previous chapters, in this chapter we are specifically interested in the solution of these location models when the distances $d(\mathbf{u}, \mathbf{v})$ are given by the ℓ_{bp}-norm, i.e., $\ell_{bp}(\mathbf{u}, \mathbf{v})$.

A single facility minisum location problem (SFMLP) in two dimensional Euclidean plane ($\Re^2$) is stated as follows:

$$\min\ S(\mathbf{x}) = \sum_{j=1}^{n} w_j\, \ell_{bp}(\mathbf{x}, \mathbf{a}_j), \tag{5.1}$$

where n is the number of fixed facilities; $\mathbf{a}_j = (a_{j1}, a_{j2})$, $j = 1, \ldots, n$ are the fixed facility locations; $\mathbf{x} = (x_1, x_2)$ is the sought-after location of the new facility; $w_j > 0$, $j = 1, \ldots, n$ is the weight (demand) associated with fixed facility j.

Using similar notation, a multi-facility minisum location problem (MFMLP) is given by

$$\min\ M(\mathbf{X}) = \sum_{i=1}^{m}\sum_{j=1}^{n} w_{1ij}\ \ell_{bp}(\mathbf{x}_i, \mathbf{a}_j) + \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} w_{2ir}\ \ell_{bp}(\mathbf{x}_i, \mathbf{x}_r). \tag{5.2}$$

where m is the number of new facilities; w_{1ij} converts the distance between new and existing facilities i and j into a cost, where $i = 1, \ldots, m$, $j = 1, \ldots, n$; and w_{2ir} converts the distance between two new facilities i and r into a cost, where $i = 1, \ldots, m-1$, $r = i+1, \ldots, m$.

As mentioned earlier, the family of ℓ_p-norm is generally used as the distance function in facility location models. The ℓ_2-norm (Euclidean distance) and the ℓ_1-norm (rectangular distance) are two well-studied special members of the ℓ_p-norm family. The SFMLP with the ℓ_p-norm is given by

$$\min\ PS(\mathbf{x}) = \sum_{j=1}^{n} w_j\ \ell_p(\mathbf{x}, \mathbf{a}_j), \tag{5.3}$$

and the MFMLP by

$$\min\ PM(\mathbf{X}) = \sum_{i=1}^{m}\sum_{j=1}^{n} w_{1ij}\ \ell_p(\mathbf{x}_i, \mathbf{a}_j) + \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} w_{2ir}\ \ell_p(\mathbf{x}_i, \mathbf{x}_r), \tag{5.4}$$

The notation used in (5.1) and (5.2) readily applies to (5.3) and (5.4), respectively. The difference between the two sets of formulations appears in distance representation.

General Solution Procedure with the ℓ_p-norm

In order to solve the single- and multi-facility location models (5.3) and (5.4) a one-point iterative procedure is used (Morris and Verdini, 1979). This iterative procedure is a generalization of the Weiszfeld procedure which is originally devised for the Euclidean distance facility location problems. The generalized Weiszfeld iterative procedure depends upon the convexity of the ℓ_p-norm, and thus utilizes the first-order necessary and sufficient conditions. Since it is impossible to express the unknown variables (new facility locations) in the form of equations, the first order derivatives can not be solved directly. Instead, an iteration function is obtained by using these derivatives. Note that the first order derivatives of $PS(\mathbf{x})$ and $PM(\mathbf{X})$ are not differentiable at the existing facility locations. Therefore, in order to avoid the problem caused by these discontinuities in the derivatives, the iterative procedure is devised by using a hyperbolic approximation of the ℓ_p-norm for actual computations.

A bound or a stopping rule is required to terminate the iterative procedure. There are several bounding methods examined in the literature. Among those, the rectangular bound, originally devised for single-facility Euclidean distance problems by Drezner (1984) and extended to unapproximated ℓ_p-norm single- and multi-facility problems by Love and Dowling (1989a; 1989b), is shown to be superior. The rectangular bound involves the solution of a rectangular distance location problem at each iteration of the Weiszfeld procedure.

In some cases, the optimal facility locations coincide with the existing facility locations. Thus, the existing facility locations are examined for optimality before applying the Weiszfeld procedure, and if an existing facility location is optimal then the rest of the solution procedure is not needed. This check is rather straightforward for the single-facility case and performed by using the fixed point optimality conditions (Juel and Love, 1981a). For the multi-facility case, first it is determined if any subset of new facilities coincide at optimality (Juel and Love, 1980; Juel, 1983; Plastria, 1992). If it is found that *all* the new facilities coincide, then the problem is solved as a single-facility problem, i.e., the fixed point optimality conditions developed for the single-facility case is applied. We note that, the need for the fixed point optimality conditions is originated from eliminating the problems caused by the discontinuities in the derivatives.

Approaches to Solve the ℓ_{bp}-norm Location Models

There are two approaches to solve the location problems (5.1) and (5.2).

1. ***Using the procedures for the ℓ_p-norm:***

   Property 1.2.1 suggests that, after a scaling based modification in a location model's setting, the Weiszfeld procedures developed for the ℓ_p-norm single- and multi-facility location problems are readily applicable to the ℓ_{bp}-norm location problems. The existing facility locations $\mathbf{a}_j$, $j = 1, \ldots, n$, in an ℓ_{bp}-norm location problem are first scaled in the corresponding directions by using the scale factors

$b_1^{1/p}$ and $b_2^{1/p}$. The location problem with the new setting is solved by using the procedures developed for the ℓ_p-norm location model and the solution is then rescaled using the scale factors $b_1^{-1/p}$ and $b_2^{-1/p}$ for corresponding coordinates.

2. ***Using modified procedures for the ℓ_{bp}-norm:***

   Making a direct use of Property 1.2.1, we can obtain modifications of the Weiszfeld procedures, the bounding methods, and the fixed point optimality condition for the ℓ_{bp}-norm location problems from the results developed for the ℓ_p-norm problems. This is done by including the scaling factors $b_1^{1/p}$ and $b_2^{1/p}$ in the expressions developed for the ℓ_p-norm location models.

The first approach is rather straightforward and utilizes the existing solution procedures for the SFMLP and MFMLP with the ℓ_p-norm. It additionally involves scaling and rescaling computations. The second approach uses the specialized procedures for the ℓ_{bp}-norm location models.

This chapter is organized as follows: In Section 5.1, we review the Weiszfeld procedure for the solution of the ℓ_p-norm single- and multi-facility location problems, and provide the generalizations to the ℓ_{bp}-norm location models (5.1) and (5.2). In Section 5.2, we first develop bounding methods for the Weiszfeld procedures used for the single- and multi-facility location problems with the *approximated* ℓ_p-norm. We also provide the generalizations of the results to the approximated ℓ_{bp}-norm location models. In Section 5.3, we first review the fixed point optimality conditions developed

for the ℓ_p-norm single-facility location problem. Then, we give its generalization to the ℓ_{bp}-norm location problem. In Section 5.4, we analyze the convergence properties of the Weiszfeld procedure when it is applied to the single-facility minisum location problems.

5.1 Modified Weiszfeld Procedure

In this section, we review the iterative procedures for the ℓ_p-norm single- and multi-facility location problems and provide their generalizations to ℓ_{bp}-norm location models.

5.1.1 Procedure for SFMLP with the ℓ_p-norm

Since the $\ell_p(\mathbf{x})$ is a norm, and thus a convex function, it readily follows that problem (5.3) is a convex optimization problem. Furthermore, if the fixed facility locations are non-collinear, then $\ell_{bp}(\mathbf{x})$ is strictly convex. Therefore, assuming the optimal solution $\mathbf{x}^*$ is a differentiable point of $PS(\mathbf{x})$, the first-order necessary and sufficient conditions require that

$$\frac{\partial PS(\mathbf{x}^*)}{\partial x_t} = 0, \quad t = 1, 2. \tag{5.5}$$

Evaluating the partial derivatives in (5.5) we have

$$\sum_{j=1}^{n} w_j \operatorname{sign}(x_t^* - a_{jt}) \frac{|x_t^* - a_{jt}|^{p-1}}{[\ell_p(\mathbf{x}^*, \mathbf{a}_j)]^{p-1}} = 0, \quad t = 1, 2. \tag{5.6}$$

Noting that $(x_t - a_{jt}) = \text{sign}(x_t - a_{jt})\,|x_t - a_{jt}|$, (5.6) can be rewritten as

$$\sum_{j=1}^{n} w_j\,(x_t^* - a_{jt})\,\frac{|x_t^* - a_{jt}|^{p-2}}{[\ell_p(\mathbf{x}^*, \mathbf{a}_j)]^{p-1}} = 0, \quad t = 1, 2. \tag{5.7}$$

Simplifying (5.7) and solving for x_t^* we have

$$x_t^* = \frac{\sum_{j=1}^{n} w_j\,|x_t^* - a_{jt}|^{p-2}\,[\ell_p(\mathbf{x}^*, \mathbf{a}_j)]^{1-p}\,a_{jt}}{\sum_{j=1}^{n} w_j\,|x_t^* - a_{jt}|^{p-2}\,[\ell_p(\mathbf{x}^*, \mathbf{a}_j)]^{1-p}}, \quad t = 1, 2. \tag{5.8}$$

Using (5.8) the one-point iteration scheme is devised as follows

$$x_t^{k+1} = \frac{\sum_{j=1}^{n} w_j\,|x_t^k - a_{jt}|^{p-2}\,[\ell_p(\mathbf{x}^k, \mathbf{a}_j)]^{1-p}\,a_{jt}}{\sum_{j=1}^{n} w_j\,|x_t^k - a_{jt}|^{p-2}\,[\ell_p(\mathbf{x}^k, \mathbf{a}_j)]^{1-p}}, \quad t = 1, 2 \tag{5.9}$$

where k represents the iteration number.

The iteration function (5.9) poses two main difficulties depending on the value of the parameter p in the application:

i. If $p < 2$, then x_t^{k+1} is undefined along the hyperplanes $|x_t^k - a_{jt}| = 0$ where $j = 1, \ldots, n$, and $t = 1, 2$.

ii. If $p \geq 2$, then x_t^{k+1} is undefined at the existing facility locations a_j, $j = 1, \ldots, n$.

In order to eliminate the obvious difficulty caused by the discontinuities in the derivatives an approximation of the ℓ_p-norm is used in the objective function $PS(\mathbf{x})$. The use of an approximation is discussed for rectangular distances by Wesolowsky and Love (1972) and for Euclidean and rectangular distances by Eyster et al. (1973). Similar approximations are given for the ℓ_p-norm by Love and Morris (1975b), and

Morris and Verdini (1979). Verdini (1976) shows that the approximation given by Eyster et al. (for Euclidean distance case) and Love and Morris is not appropriate when the Weiszfeld procedure is used for the ℓ_p distances problem with $p \geq 1$. Therefore, the approximation given here follows the one given by Morris and Verdini. We next present this approximation of $\ell_p(\mathbf{x})$, denoted by $\tilde{\ell}_p(\mathbf{x})$, and review its properties and the resulting iterative procedure.

Approximating Function for the ℓ_p-norm

Using the hyperbolic approximation of the ℓ_p-norm the approximated distance $\tilde{\ell}_p$ between any two points $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$ is given by

$$\tilde{\ell}_p(\mathbf{x}, \mathbf{y}) = \left[((x_1 - y_1)^2 + \epsilon)^{p/2} + ((x_2 - y_2)^2 + \epsilon)^{p/2}\right]^{1/p}, \quad p \geq 1, \ \epsilon > 0. \tag{5.10}$$

Notice that the approximation to the ℓ_p distance is not a norm; it lacks the stationarity property, i.e., $\tilde{\ell}_p(\mathbf{0}) \neq 0$. However, it is still a convex function of $\mathbf{x}$ as shown by Morris and Verdini (1979).

Iterative Procedure with Hyperbolic Approximation

Rewriting the problem (5.3) with the hyperbolic approximating distance function $\tilde{\ell}_p(x)$, we have

$$\min \ \widetilde{PS}(\mathbf{x}) = \sum_{j=1}^{n} w_j \ \left[((x_1 - a_{j1})^2 + \epsilon)^{p/2} + ((x_2 - a_{j2})^2 + \epsilon)^{p/2}\right]^{1/p}, \tag{5.11}$$

$$\text{where} \quad p \geq 1, \ \ \epsilon > 0, \ \ \text{and} \ \ w_j \geq 0, \ \ j = 1, \ldots, n.$$

Clearly the function $\widetilde{PS}(\mathbf{x})$, being a sum of n strictly convex functions, is a strictly convex function in $\mathbf{x} = (x_1, x_2)$. Therefore, again using the first-order necessary and

sufficient conditions, and following the same steps given in (5.6), (5.7), (5.8), (5.9) the modified iteration function is found as

$$x_t^{k+1} = \frac{\sum_{j=1}^{n} w_j \left((x_t^k - a_{jt})^2 + \epsilon\right)^{\frac{p}{2}-1} \left(\tilde{\ell}_p(\mathbf{x}^k, \mathbf{a}_j)\right)^{1-p} a_{jt}}{\sum_{j=1}^{n} w_j \left((x_t^k - a_{jt})^2 + \epsilon\right)^{\frac{p}{2}-1} \left(\tilde{\ell}_p(\mathbf{x}^k, \mathbf{a}_j)\right)^{1-p}}, \quad t = 1, 2. \tag{5.12}$$

5.1.2 Generalization to SFMLP with the ℓ_{bp}-norm

We employ the following hyperbolic approximation of the ℓ_{bp}-norm. Using the notation given in (1.9)

$$\tilde{\ell}_{bp}(\mathbf{x}, \mathbf{y}) = \left[b_1 \left((x_1 - y_1)^2 + \epsilon\right)^{p/2} + b_2 \left((x_2 - y_2)^2 + \epsilon\right)^{p/2}\right]^{1/p}, \tag{5.13}$$

$$\text{where} \quad b_1, b_2 > 0, \; p \geq 1, \; \epsilon > 0.$$

Similar to Property 1.2.1 given for the unapproximated case, we can write

$$\tilde{\ell}_{bp}(\mathbf{z}) = \tilde{\ell}_p(\mathbf{z}') \tag{5.14}$$

where $\mathbf{z} = \mathbf{x} - \mathbf{y}$, $z_1' = b_1^{1/p} z_1$, $z_2' = b_2^{1/p} z_2$ and denoting the small quantity associated with the ℓ_p-norm by ϵ', $\epsilon' = \min\{b_1^{2/p}\epsilon, \; b_2^{2/p}\epsilon\}$. As suggested by the relation (5.14), replacing the fixed facility locations a_{jt} with $b_t^{1/p} a_{jt}$ for $t = 1, 2$, $j = 1, \ldots, n$, unknown facility locations x_t with $b_t^{1/p} x_t$ for $t = 1, 2$, and ϵ with $\min\{b_1^{2/p}\epsilon, \; b_2^{2/p}\epsilon\}$ in (5.12), and simplifying, we obtain the modified iteration function for an ℓ_{bp}-norm SFMLP as

$$x_t^{k+1} = \frac{\sum_{j=1}^{n} w_j \left((x_t^k - a_{jt})^2 + \epsilon\right)^{\frac{p}{2}-1} \left(\tilde{\ell}_{bp}(\mathbf{x}^k, \mathbf{a}_j)\right)^{1-p} a_{jt}}{\sum_{j=1}^{n} w_j \left((x_t^k - a_{jt})^2 + \epsilon\right)^{\frac{p}{2}-1} \left(\tilde{\ell}_{bp}(\mathbf{x}^k, \mathbf{a}_j)\right)^{1-p}}, \quad t = 1, 2. \tag{5.15}$$

5.1.3 Procedure for MFMLP with the ℓ_p-norm

We consider the formulation given in (5.4). The convexity of $PM(\mathbf{X})$ again follows from the fact that $\ell_p(x)$ is a norm. When the partial derivatives with respect to $\mathbf{X}$ are calculated for the first part of $PM(\mathbf{X})$, the same type of singularities given for SFMLP will be introduced. The partial derivatives for the second part, which is the transportation cost based on the interactions between new facilities, will introduce new singularities. In particular, the partial derivatives may be discontinuous at any point in the solution space since any two new facilities may coincide at any point during the iterative procedure, making the derivatives undefined. Therefore, we again adopt the hyperbolic approximating function of the ℓ_p-norm. The approximation to problem (5.4) becomes

$$\begin{aligned}\min \widetilde{PM}(\mathbf{X}) \;=\; & \sum_{i=1}^{m}\sum_{j=1}^{n} w_{1ij}\left[((x_{i1}-a_{j1})^2+\epsilon)^{p/2}+((x_{i2}-a_{j2})^2+\epsilon)^{p/2}\right]^{1/p} \\ & + \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} w_{2ir}\left[((x_{i1}-x_{r1})^2+\epsilon)^{p/2}+((x_{i2}-x_{r2})^2+\epsilon)^{p/2}\right]^{1/p}\end{aligned} \tag{5.16}$$

where $p \geq 1$, $\epsilon > 0$, and $w_{1ij}, w_{2ir} \geq 0$. It is already known that the first sum in $\widetilde{PM}(\mathbf{X})$ is a convex function. The following property, originally given by Morris and Verdini (1979), relates to the convexity of the second sum in $\widetilde{PM}(\mathbf{X})$.

Property 5.1.1 *$\tilde{\ell}_p(\mathbf{x},\mathbf{y})$ is convex in $\mathbf{x}$ and $\mathbf{y}$, provided that $p \geq 1$.*

A Weiszfeld procedure for the solution of problem (5.16) is devised in a similar fashion to the single facility case. Note that, in order to deal with a well-formulated problem we assume that all new facilities are chained (Cabot and Francis, 1972). New facility i is chained if there exists a positive w_{1ij} where j is any existing facility or if there exists a positive w_{2ir} (or w_{2ri}) where r is any *chained* new facility. Taking the partial derivatives of $\widetilde{PM}(\mathbf{X})$, the first-order necessary and sufficient conditions are given as follows:

$$\sum_{j=1}^{n} w_{1ij}\,(x_{it}^{*} - a_{jt})\,\frac{\left((x_{it}^{*}-a_{jt})^2+\epsilon\right)^{\frac{p-2}{2}}}{\left(\tilde{\ell}_p(\mathbf{x}_i^{*},\mathbf{a}_j)\right)^{p-1}} \tag{5.17}$$

$$+\sum_{r=1}^{m} w_{2ir}\,(x_{it}^{*} - x_{rt})\,\frac{\left((x_{it}^{*}-x_{rt})^2+\epsilon\right)^{\frac{p-2}{2}}}{\left(\tilde{\ell}_p(\mathbf{x}_i^{*},\mathbf{x}_r)\right)^{p-1}} = 0, \quad i=1,\ldots,m, \quad t=1,2.$$

Rearranging and simplifying, the iterative procedure is devised as

$$x_{it}^{k+1} = \frac{NF_{it}^{k} + NS_{it}^{k}}{DF_{it}^{k} + DS_{it}^{k}} \quad r=1,\ldots,m, \quad t=1,2, \tag{5.18}$$

where

$$NF_{it}^{k} = \sum_{j=1}^{n} w_{1ij}\left((x_{it}^{k}-a_{jt})^2+\epsilon\right)^{\frac{p-2}{2}}\left(\tilde{\ell}_p(\mathbf{x}_i^{k},\mathbf{a}_j)\right)^{1-p} a_{jt},$$

$$DF_{it}^{k} = \sum_{j=1}^{n} w_{1ij}\left((x_{it}^{k}-a_{jt})^2+\epsilon\right)^{\frac{p-2}{2}}\left(\tilde{\ell}_p(\mathbf{x}_i^{k},\mathbf{a}_j)\right)^{1-p},$$

$$NS_{it}^{k} = \sum_{r=1}^{m} w_{2}\left((x_{it}^{k}-x_{rt}^{k})^2+\epsilon\right)^{\frac{p-2}{2}}\left(\tilde{\ell}_p(\mathbf{x}_i^{k},\mathbf{x}_r^{k})\right)^{1-p} x_{rt}^{k},$$

$$DS_{it}^{k} = \sum_{r=1}^{m} w_{2}\left((x_{it}^{k}-x_{rt}^{k})^2+\epsilon\right)^{\frac{p-2}{2}}\left(\tilde{\ell}_p(\mathbf{x}_i^{k},\mathbf{x}_r^{k})\right)^{1-p},$$

$$\text{and } w_2 = \begin{cases} w_{2ir} & \text{if } i<r \\ w_{2ri} & \text{if } r<i \end{cases}$$

5.1.4 Generalization to MFMLP with the ℓ_{bp}-norm

Using the relation (5.14) a generalized iterative procedure for the solution of ℓ_{bp}-norm MFMLP can now be given in a similar fashion to the single facility location problem.

$$x_{it}^{k+1} = \frac{NF_{it}^k + NS_{it}^k}{DF_{it}^k + DS_{it}^k} \quad r = 1, \ldots, m, \quad t = 1, 2, \tag{5.19}$$

where

$$\begin{aligned}
NF_{it}^k &= \sum_{j=1}^{n} w_{1ij} \left((x_{it}^k - a_{jt})^2 + \epsilon \right)^{\frac{p-2}{2}} \left(\tilde{\ell}_{bp}(\mathbf{x}_i^k, \mathbf{a}_j) \right)^{1-p} a_{jt}, \\
DF_{it}^k &= \sum_{j=1}^{n} w_{1ij} \left((x_{it}^k - a_{jt})^2 + \epsilon \right)^{\frac{p-2}{2}} \left(\tilde{\ell}_{bp}(\mathbf{x}_i^k, \mathbf{a}_j) \right)^{1-p}, \\
NS_{it}^k &= \sum_{r=1}^{m} w_2 \left((x_{it}^k - x_{rt}^k)^2 + \epsilon \right)^{\frac{p-2}{2}} \left(\tilde{\ell}_{bp}(\mathbf{x}_i^k, \mathbf{x}_r^k) \right)^{1-p} x_{rt}^k, \\
DS_{it}^k &= \sum_{r=1}^{m} w_2 \left((x_{it}^k - x_{rt}^k)^2 + \epsilon \right)^{\frac{p-2}{2}} \left(\tilde{\ell}_{bp}(\mathbf{x}_i^k, \mathbf{x}_r^k) \right)^{1-p}, \\
\text{and } w_2 &= \begin{cases} w_{2ir} & \text{if } i < r \\ w_{2ri} & \text{if } r < i \end{cases}
\end{aligned}$$

5.2 Bounding Method

We need a bounding method (or a stopping rule) for the iterative procedure described in Section 5.1. Several bounding methods have been proposed for single- and multi-facility continuous location models. Love and Yeong (1981) and Juel (1984) developed stopping rules that can be used with ℓ_p distance continuous location models. Drezner (1984) introduced a stopping rule, which involves the solution of a rectangular distance location problem, for the single-facility Euclidean distance location problem.

Dowling and Love (1986) provided the extension of Drezner's rectangular bound to multi-facility Euclidean distance models. Wendell and Peterson (1984) utilized the dual of the ℓ_p distance minisum location problem to calculate a lower bound for the primal objective function. Finally, Love and Dowling extended the rectangular bound to the single-facility ℓ_p distance model (Love and Dowling, 1989b) and the multi-facility ℓ_p distance model (Love and Dowling, 1989a). Both extensions involve the unapproximated ℓ_p distances. Among those methods, it has been shown by Love and Dowling (1986; 1989a; 1989b) that the rectangular bound is more efficient than the others. Based on this result, in this section we develop rectangular bounding methods for the single- and multi-facility minisum location models with the approximated ℓ_p-norm. The derivation of the rectangular bounds presented here follows the approach given for the unapproximated ℓ_p-norm case by Love and Dowling (1989a; 1989b). We additionally show that, at optimality, the solution to the bound problem coincides with the solution to the original location problem. We also provide the generalizations of the bounding methods to the approximated ℓ_{bp}-norm minisum location models.

5.2.1 Bound for SFMLP with the $\tilde{\ell}_p$-norm

A rectangular bound for the iterative procedure (5.12) can be obtained by using the Hölder Inequality given by

$$\sum_{i=1}^{N} |\alpha_i \beta_i| \leq \left(\sum_{i=1}^{N} |\alpha_i|^p \right)^{1/p} \left(\sum_{i=1}^{N} |\beta_i|^q \right)^{1/q}$$

where α and β are N-dimensional vectors, $p > 1$, and $1/p + 1/q = 1$. Taking $N = 2$ for the planar location model and letting

$$\alpha_1 = ((x_1 - a_{j1})^2 + \epsilon)^{1/2}, \quad \beta_1 = ((x_1^k - a_{j1})^2 + \epsilon)^{(p-1)/2},$$
$$\alpha_2 = ((x_2 - a_{j2})^2 + \epsilon)^{1/2}, \quad \text{and} \quad \beta_2 = ((x_2^k - a_{j2})^2 + \epsilon)^{(p-1)/2},$$

we obtain

$$\begin{aligned} &((x_1 - a_{j1})^2 + \epsilon)^{1/2} \, ((x_1^k - a_{j1})^2 + \epsilon)^{(p-1)/2} \\ &\qquad + \; ((x_2 - a_{j2})^2 + \epsilon)^{1/2} \, ((x_2^k - a_{j2})^2 + \epsilon)^{(p-1)/2} \\ &\leq \left((((x_1 - a_{j1})^2 + \epsilon)^{1/2})^p + (((x_2 - a_{j2})^2 + \epsilon)^{1/2})^p \right)^{\frac{1}{p}} \\ &\qquad \left((((x_1^k - a_{j1})^2 + \epsilon)^{(p-1)/2})^q + (((x_2^k - a_{j2})^2 + \epsilon)^{(p-1)/2})^q \right)^{\frac{1}{q}}. \end{aligned}$$

Rearranging terms we have

$$\begin{aligned} &\tilde{\ell}_p(\mathbf{x}, \mathbf{a}_j) \left(((x_1^k - a_{j1})^2 + \epsilon)^{p/2} + ((x_2^k - a_{j2})^2 + \epsilon)^{p/2} \right)^{\frac{1}{q}} \\ &\qquad \geq \quad ((x_1 - a_{j1})^2 + \epsilon)^{1/2} \, ((x_1^k - a_{j1})^2 + \epsilon)^{(p-1)/2} \\ &\qquad + \; ((x_2 - a_{j2})^2 + \epsilon)^{1/2} \, ((x_2^k - a_{j2}) + \epsilon)^{(p-1)/2}. \end{aligned}$$

Rewriting the second term on the left-hand side, we obtain

$$\begin{aligned} &\tilde{\ell}_p(\mathbf{x}, \mathbf{a}_j) \, (\tilde{\ell}_p(\mathbf{x}^k, \mathbf{a}_j))^{p-1} \\ &\qquad \geq \quad ((x_1 - a_{j1})^2 + \epsilon)^{1/2} \, ((x_1^k - a_{j1})^2 + \epsilon)^{(p-1)/2} \\ &\qquad + \; ((x_2 - a_{j2})^2 + \epsilon)^{1/2} \, ((x_2^k - a_{j2}) + \epsilon)^{(p-1)/2}. \end{aligned}$$

In order to obtain the cost function of the minisum model we multiply both sides by w_j and sum for $j = 1, \ldots, n$. Thus, we have

$$\begin{aligned}
\widetilde{PS}(\mathbf{x}) &= \sum_{j=1}^{n} w_j \, \tilde{\ell}_p(\mathbf{x}, \mathbf{a}_j) \\
&\geq \sum_{j=1}^{n} w_j \, \frac{((x_1 - a_{j1})^2 + \epsilon)^{1/2} \, ((x_1^k - a_{j1})^2 + \epsilon)^{(p-1)/2}}{(\bar{\ell}_p(\mathbf{x}^k, \mathbf{a}_j))^{p-1}} \\
&+ \sum_{j=1}^{n} w_j \, \frac{((x_2 - a_{j2})^2 + \epsilon)^{1/2} \, ((x_2^k - a_{j2})^2 + \epsilon)^{(p-1)/2}}{(\bar{\ell}_p(\mathbf{x}^k, \mathbf{a}_j))^{p-1}}.
\end{aligned}$$

Minimizing both sides of inequality over $\mathbf{x}$ gives

$$\begin{aligned}
\widetilde{PS}(\mathbf{x}^*) \geq \min_{\mathbf{x}} \Bigg\{ & \sum_{j=1}^{n} w_j \, \frac{((x_1 - a_{j1})^2 + \epsilon)^{1/2} \, ((x_1^k - a_{j1})^2 + \epsilon)^{(p-1)/2}}{(\bar{\ell}_p(\mathbf{x}^k, \mathbf{a}_j))^{p-1}} \\
& + \sum_{j=1}^{n} w_j \, \frac{((x_2 - a_{j2})^2 + \epsilon)^{1/2} \, ((x_2^k - a_{j2})^2 + \epsilon)^{(p-1)/2}}{(\bar{\ell}_p(\mathbf{x}^k, \mathbf{a}_j))^{p-1}} \Bigg\}.
\end{aligned}$$

By letting $\epsilon \to 0$ we simplify the terms $((x_1 - a_{j1})^2 + \epsilon)^{1/2}$ and $((x_2 - a_{j2})^2 + \epsilon)^{1/2}$ as $|x_1 - a_{j1}|$ and $|x_2 - a_{j2}|$, respectively. Thus, the bound as a rectangular distance problem, $\widetilde{PSB}^k(\mathbf{x})$, is found as

$$\widetilde{PSB}^k(\mathbf{x}) = \min_{x_1} \sum_{j=1}^{n} u_j \, |x_1 - a_{j1}| + \min_{x_2} \sum_{j=1}^{n} v_j \, |x_2 - a_{j2}|, \tag{5.20}$$

where

$$\begin{aligned}
u_j &= w_j \, \frac{((x_1^k - a_{j1})^2 + \epsilon)^{(p-1)/2}}{(\bar{\ell}_p(\mathbf{x}^k, \mathbf{a}_j))^{p-1}} \quad \text{and} \\
v_j &= w_j \, \frac{((x_2^k - a_{j2})^2 + \epsilon)^{(p-1)/2}}{(\bar{\ell}_p(\mathbf{x}^k, \mathbf{a}_j))^{p-1}}, \quad j = 1, \ldots, n.
\end{aligned}$$

For notational convenience we denote the solution of a rectangular distance location problem by $\mathbf{x}^R$, and thus the bound at an iteration k is given by $\widetilde{PSB}^k(\mathbf{x}^{R^*})$.

Let $\widetilde{PS}_j(\mathbf{x})$, $j = 1, \ldots, n$, denote the terms in $\widetilde{PS}(\mathbf{x})$. Then the first derivatives of $\widetilde{PS}_j(\mathbf{x})$ with respect to x_1 and x_2 are

$$\frac{\partial \widetilde{PS}_j(\mathbf{x})}{\partial x_1} = w_j \frac{((x_1 - a_{j1})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_p(\mathbf{x}, \mathbf{a}_j))^{p-1}} \left(\frac{(x_1 - a_{j1})}{((x_1 - a_{j1})^2 + \epsilon)^{1/2}} \right) \quad \text{and}$$

$$\frac{\partial \widetilde{PS}_j(\mathbf{x})}{\partial x_2} = w_j \frac{((x_2 - a_{j2})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_p(\mathbf{x}, \mathbf{a}_j))^{p-1}} \left(\frac{(x_2 - a_{j2})}{((x_2 - a_{j2})^2 + \epsilon)^{1/2}} \right), \quad j = 1, \ldots, n.$$

By letting $\epsilon \to 0$ as above, and using the the equality $(x_t - a_{jt}) = \text{sign}(x_t - a_{jt})\ |x_t - a_{jt}|$, for $t = 1, 2$, we can simplify the last terms and obtain

$$\frac{\partial \widetilde{PS}_j(\mathbf{x})}{\partial x_1} = w_j \frac{((x_1 - a_{j1})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_p(\mathbf{x}, \mathbf{a}_j))^{p-1}} \text{sign}(x_1 - a_{j1}) \quad \text{and}$$

$$\frac{\partial \widetilde{PS}_j(\mathbf{x})}{\partial x_2} = w_j \frac{((x_2 - a_{j2})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_p(\mathbf{x}, \mathbf{a}_j))^{p-1}} \text{sign}(x_2 - a_{j2}), \quad j = 1, \ldots, n.$$

Thus, u_j and v_j can be rewritten as

$$u_j = \left| \frac{\partial \widetilde{PS}_j(\mathbf{x}^k)}{\partial x_1} \right| \quad \text{and} \quad v_j = \left| \frac{\partial \widetilde{PS}_j(\mathbf{x}^k)}{\partial x_2} \right|, \quad j = 1, \ldots, n. \tag{5.21}$$

Property 5.2.1 *Let $\mathcal{S}^{\mathcal{R}}$ be the set of all the points which are optimal for the rectangular bound problem. Then, $\mathbf{x}^* \in \mathcal{S}^{\mathcal{R}}$*

Proof We order the existing facility locations $\mathbf{a}_j$, $j = 1, \ldots, n$ in their coordinates in the x_1- and x_2-directions as follows:

$$a_{j1}^{[1]} < \cdots < a_{j1}^{[\mu]} < a_{j1}^{[\mu+1]} < \cdots < a_{j1}^{[n]},$$

and

$$a_{j2}^{[1]} < \cdots < a_{j2}^{[\nu]} < a_{j2}^{[\nu+1]} < \cdots < a_{j2}^{[n]}$$

where the bracketed superscripts denote the ordering and ties are broken arbitrarily. Let $\mathcal{H}^C$ denote the convex hull of the existing facility locations. It is well-known that $\mathbf{x}^* \in \mathcal{H}^C$ (Brimberg and Love, 1995a). There are four cases to consider:

Case 1: Suppose that $\mathbf{x}^* \in \mathcal{H}^I$ where $\mathcal{H}^I = \mathcal{H}^C \cap \mathcal{H}^R$ and $\mathcal{H}^R$ is a rectangular hull defined by the edges $x_1 = a_{j1}^{[\mu]}$, $x_1 = a_{j1}^{[\mu+1]}$, $x_2 = a_{j2}^{[\nu]}$, $x_2 = a_{j2}^{[\nu+1]}$. First order necessary and sufficient conditions for the SFMLP state that

$$\sum_{j=1}^{n} \frac{\partial \widetilde{PS}_j(\mathbf{x}^*)}{\partial x_1} = 0 \quad \text{and} \quad \sum_{j=1}^{n} \frac{\partial \widetilde{PS}_j(\mathbf{x}^*)}{\partial x_2} = 0.$$

Since

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_1} > 0 \quad \text{for} \quad \sigma \leq \mu,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_1} < 0 \quad \text{for} \quad \sigma \geq \mu + 1,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_2} > 0 \quad \text{for} \quad \sigma \leq \nu,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_2} < 0 \quad \text{for} \quad \sigma \geq \nu + 1,$$

we must have

$$\sum_{\sigma \leq \mu} \left| \frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_1} \right| = \sum_{\sigma \geq \mu+1} \left| \frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_1} \right| \tag{5.22}$$

$$\sum_{\sigma \leq \nu} \left| \frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_2} \right| = \sum_{\sigma \geq \nu+1} \left| \frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_2} \right| \tag{5.23}$$

Using (5.21), (5.22) and (5.23) can be rewritten as

$$\sum_{\sigma \leq \mu} u_j^{[\sigma]} = \sum_{\sigma \geq \mu+1} u_j^{[\sigma]}, \tag{5.24}$$

$$\sum_{\sigma \leq \nu} v_j^{[\sigma]} = \sum_{\sigma \geq \nu+1} v_j^{[\sigma]}, \tag{5.25}$$

respectively. Observe that (5.24) and (5.25) imply the optimality conditions for the rectangular bound problem. *Any* point in $\mathcal{H}^{\mathcal{I}}$ is an optimal solution to the rectangular bound problem, i.e., $\mathcal{S}^{\mathcal{R}} = \mathcal{H}^{\mathcal{I}}$ and thus $\mathbf{x}^* \in \mathcal{S}^{\mathcal{R}}$.

Case 2: Suppose that $\mathbf{x}^* \in \mathcal{H}^{\mathcal{I}}$ where $\mathcal{H}^{\mathcal{I}} = \mathcal{H}^{\mathcal{C}} \cap \mathcal{H}^{\mathcal{L}}$ and $\mathcal{H}^{\mathcal{L}}$ is a line segment on the hyperplane $x_1 = a_{j1}^{[\mu]}$ between $x_2 = a_{j2}^{[\nu]}$ and $x_2 = a_{j2}^{[\nu+1]}$. Using first order necessary and sufficient conditions for the SFMLP, (5.21) and the relations

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_1} > 0 \quad \text{for} \quad \sigma \leq \mu - 1,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_1} = 0 \quad \text{for} \quad \sigma = \mu,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_1} < 0 \quad \text{for} \quad \sigma \geq \mu + 1,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_2} > 0 \quad \text{for} \quad \sigma \leq \nu,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_2} < 0 \quad \text{for} \quad \sigma \geq \nu + 1,$$

we must have

$$\sum_{\sigma \leq \mu-1} u_j^{[\sigma]} = \sum_{\sigma \geq \mu+1} u_j^{[\sigma]}, \tag{5.26}$$

$$\sum_{\sigma \leq \nu} v_j^{[\sigma]} = \sum_{\sigma \geq \nu+1} v_j^{[\sigma]}. \tag{5.27}$$

Observe that (5.26) and (5.27) imply the optimality conditions for the rectangular bound problem. We have $\mathcal{S}^{\mathcal{R}} = \mathcal{H}^{\mathcal{EI}}$ where $\mathcal{H}^{\mathcal{EI}} = \mathcal{H}^{\mathcal{C}} \cap \mathcal{H}^{\mathcal{ER}}$ and $\mathcal{H}^{\mathcal{ER}}$ is a rectangular hull defined by the edges $x_1 = a_{j1}^{[\mu-1]}$, $x_1 = a_{j1}^{[\mu+1]}$, $x_2 = a_{j2}^{[\nu]}$, and $x_2 = a_{j2}^{[\nu+1]}$. Obviously, $\mathcal{H}^{\mathcal{I}} \subset \mathcal{H}^{\mathcal{EI}}$ and thus, $\mathbf{x}^* \in \mathcal{S}^{\mathcal{R}}$.

Case 3: Similar to Case 2, suppose that $\mathbf{x}^* \in \mathcal{H}^{\mathcal{I}}$ where $\mathcal{H}^{\mathcal{I}} = \mathcal{H}^{\mathcal{C}} \cap \mathcal{H}^{\mathcal{L}}$ and $\mathcal{H}^{\mathcal{L}}$ is a line segment on the hyperplane $x_2 = a_{j2}^{[\nu]}$ between $x_1 = a_{j1}^{[\mu]}$ and $x_1 = a_{j1}^{[\mu+1]}$ Using first order necessary and sufficient conditions for the SFMLP, (5.21) and the relations

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_1} > 0 \quad \text{for} \quad \sigma \leq \mu,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_1} < 0 \quad \text{for} \quad \sigma \geq \mu + 1,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_2} > 0 \quad \text{for} \quad \sigma \leq \nu - 1,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_2} = 0 \quad \text{for} \quad \sigma = \nu,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_2} < 0 \quad \text{for} \quad \sigma \geq \nu + 1,$$

we must have

$$\sum_{\sigma\leq\mu} u_j^{[\sigma]} = \sum_{\sigma\geq\mu+1} u_j^{[\sigma]}, \tag{5.28}$$

$$\sum_{\sigma\leq\nu-1} v_j^{[\sigma]} = \sum_{\sigma\geq\nu+1} v_j^{[\sigma]}. \tag{5.29}$$

Observe that (5.28) and (5.29) imply the optimality conditions for the rectangular bound problem. We have $\mathcal{S}^{\mathcal{R}} = \mathcal{H}^{\mathcal{EI}}$ where $\mathcal{H}^{\mathcal{EI}} = \mathcal{H}^{\mathcal{C}} \cap \mathcal{H}^{\mathcal{ER}}$ and $\mathcal{H}^{\mathcal{ER}}$ is a rectangular hull defined by the edges $x_1 = a_{j1}^{[\mu]}$, $x_1 = a_{j1}^{[\mu+1]}$, $x_2 = a_{j2}^{[\nu-1]}$, and $x_2 = a_{j2}^{[\nu+1]}$. Obviously, $\mathcal{H}^{\mathcal{I}} \subset \mathcal{H}^{\mathcal{EI}}$ and thus, $\mathbf{x}^* \in \mathcal{S}^{\mathcal{R}}$.

Case 4: Suppose that $\mathbf{x}^*$ is the intersection point of hyperplanes $x_1 = a_{j1}^{[\mu]}$ and $x_2 = a_{j2}^{[\nu]}$. Using first order necessary and sufficient conditions for the SFMLP, (5.21) and the relations

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_1} > 0 \quad \text{for} \quad \sigma \leq \mu - 1,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_1} = 0 \quad \text{for} \quad \sigma = \mu,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_1} < 0 \quad \text{for} \quad \sigma \geq \mu + 1,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_2} > 0 \quad \text{for} \quad \sigma \leq \nu - 1,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_2} = 0 \quad \text{for} \quad \sigma = \nu,$$

$$\frac{\partial \widetilde{PS}_j^{[\sigma]}(\mathbf{x}^*)}{\partial x_2} < 0 \quad \text{for} \quad \sigma \geq \nu + 1,$$

we must have

$$\sum_{\sigma \leq \mu-1} u_j^{[\sigma]} = \sum_{\sigma \geq \mu+1} u_j^{[\sigma]}, \tag{5.30}$$

$$\sum_{\sigma \leq \nu-1} v_j^{[\sigma]} = \sum_{\sigma \geq \nu+1} v_j^{[\sigma]}. \tag{5.31}$$

Observe that (5.30) and (5.31) imply the optimality conditions for the rectangular bound problem. We have $\mathcal{S}^{\mathcal{R}} = \mathcal{H}^{\mathcal{EI}}$ where $\mathcal{H}^{\mathcal{EI}} = \mathcal{H}^{\mathcal{C}} \cap \mathcal{H}^{\mathcal{ER}}$ and $\mathcal{H}^{\mathcal{ER}}$ is a rectangular hull defined by the edges $x_1 = a_{j1}^{[\mu-1]}$, $x_1 = a_{j1}^{[\mu+1]}$, $x_2 = a_{j2}^{[\nu-1]}$, and $x_2 = a_{j2}^{[\nu+1]}$. Obviously, $\mathcal{H}^{\mathcal{I}} \subset \mathcal{H}^{\mathcal{EI}}$ and thus, $\mathbf{x}^* \in \mathcal{S}^{\mathcal{R}}$. □

Property 5.2.2 *The rectangular bound converges to* $\widetilde{PS}(\mathbf{x}^*)$ *for all* $p > 1$, *i.e.,* $\lim_{k \to \infty} \widetilde{PSB}^k(\mathbf{x}^*) = \widetilde{PS}(\mathbf{x}^*)$

Proof We first rewrite (5.20) at optimality in its original form by introducing $\epsilon > 0$

$$\widetilde{PSB}^k(\mathbf{x}^{R^*}) = \sum_{j=1}^{n} w_j \frac{((x_1^{R^*} - a_{j1})^2 + \epsilon)^{1/2} \, ((x_1^k - a_{j1})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_p(\mathbf{x}^k, \mathbf{a}_j))^{p-1}}$$

$$+ \sum_{j=1}^{n} w_j \frac{((x_2^{R^*} - a_{j2})^2 + \epsilon)^{1/2} \, ((x_2^k - a_{j2})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_p(\mathbf{x}^k, \mathbf{a}_j))^{p-1}}.$$

By Property 5.2.1 we have

$$\widetilde{PSB}^k(\mathbf{x}^*) = \sum_{j=1}^{n} w_j \frac{((x_1^* - a_{j1})^2 + \epsilon)^{1/2} \, ((x_1^k - a_{j1})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_p(\mathbf{x}^k, \mathbf{a}_j))^{p-1}}$$

$$+ \sum_{j=1}^{n} w_j \frac{((x_2^* - a_{j2})^2 + \epsilon)^{1/2} \, ((x_2^k - a_{j2})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_p(\mathbf{x}^k, \mathbf{a}_j))^{p-1}}.$$

Assuming that $\lim_{k\to\infty} \mathbf{x}^k = \mathbf{x}^*$, we obtain

$$\begin{aligned}\lim_{k\to\infty} \widetilde{PSB}^k(\mathbf{x}^*) &= \sum_{j=1}^{n} w_j \frac{((x_1^* - a_{j1})^2 + \epsilon)^{1/2}\,((x_1^* - a_{j1})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_p(\mathbf{x}^*, \mathbf{a}_j))^{p-1}} \\ &+ \sum_{j=1}^{n} w_j \frac{((x_2^* - a_{j2})^2 + \epsilon)^{1/2}\,((x_2^* - a_{j2})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_p(\mathbf{x}^*, \mathbf{a}_j))^{p-1}},\end{aligned}$$

and simplifying

$$\begin{aligned}\lim_{k\to\infty} \widetilde{PSB}^k(\mathbf{x}^*) &= \sum_{j=1}^{n} w_j \frac{((x_1^* - a_{j1})^2 + \epsilon)^{p/2} + ((x_2^* - a_{j2})^2 + \epsilon)^{p/2}}{(\tilde{\ell}_p(\mathbf{x}^*, \mathbf{a}_j))^{p-1}} \\ &= \sum_{j=1}^{n} w_j \frac{(\tilde{\ell}_p(\mathbf{x}^*, \mathbf{a}_j))^{p}}{(\tilde{\ell}_p(\mathbf{x}^*, \mathbf{a}_j))^{p-1}} \\ &= \tilde{S}(\mathbf{x}^*)\end{aligned}$$

□

5.2.2 Generalization to SFMLP with the $\tilde{\ell}_{bp}$-norm

To obtain the generalization of the rectangular bound (5.20) for the ℓ_{bp}-norm SFMLP we utilize the Property 1.2.1 and the relation (5.14). Introducing the scale factors $b_1^{1/p}$ and $b_2^{1/p}$ accordingly and simplifying, the bound for the SFMLP with the ℓ_{bp} distances, $\tilde{SB}^k(\mathbf{x}^R)$, is found as

$$\tilde{SB}^k(\mathbf{x}^R) = \min_{x_1^R} \sum_{j=1}^{n} u_j\,|x_1^R - a_{j1}| + \min_{x_2^R} \sum_{j=1}^{n} v_j\,|x_2^R - a_{j2}|, \tag{5.32}$$

where

$$u_j = w_j \frac{b_1\,((x_1^k - a_{j1})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_{bp}(\mathbf{x}^k, \mathbf{a}_j))^{p-1}} \quad \text{and}$$

$$v_j = w_j \frac{b_2\,((x_2^k - a_{j2})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_{bp}(\mathbf{x}^k, \mathbf{a}_j))^{p-1}}, \quad j = 1, \ldots, n.$$

Note that the Properties 5.2.1 and 5.2.2 readily apply to this case.

5.2.3 Bound for MFMLP with the $\tilde{\ell}_p$-norm

We develop a rectangular bound for the iterative procedure (5.18). The Hölder Inequality is used in the same way as in the single-facility case. Considering the first part of the objective function, let

$$\alpha_1 = ((x_{i1} - a_{j1})^2 + \epsilon)^{1/2}, \quad \beta_1 = ((x_{i1}^k - a_{j1})^2 + \epsilon)^{(p-1)/2},$$
$$\alpha_2 = ((x_{i2} - a_{j2})^2 + \epsilon)^{1/2}, \quad \text{and} \quad \beta_2 = ((x_{i2}^k - a_{j2})^2 + \epsilon)^{(p-1)/2},$$

Then it follows that

$$\begin{aligned}\tilde{\ell}_p(\mathbf{x}_i, \mathbf{a}_j) \left(((x_{i1}^k - a_{j1})^2 + \epsilon)^{p/2} + ((x_{i2}^k - a_{j2})^2 + \epsilon)^{p/2}\right)^{\frac{1}{q}} \\ \geq \quad ((x_{i1} - a_{j1})^2 + \epsilon)^{1/2} \, ((x_{i1}^k - a_{j1})^2 + \epsilon)^{(p-1)/2} \\ + \quad ((x_{i2} - a_{j2})^2 + \epsilon)^{1/2} \, ((x_{i2}^k - a_{j2}) + \epsilon)^{(p-1)/2}.\end{aligned}$$

and using similar steps as these for single-facility case we obtain

$$\sum_{i=1}^{m}\sum_{j=1}^{n} w_{1ij} \, \tilde{\ell}_p(\mathbf{x}_i, \mathbf{a}_j) \geq \sum_{i=1}^{m}\sum_{j=1}^{n} u_{1ij} \, |x_{i1} - a_{j1}| + \sum_{i=1}^{m}\sum_{j=1}^{n} v_{1ij} \, |x_{i2} - a_{j2}|, \qquad (5.33)$$

where

$$u_{1ij} = w_{1ij} \frac{((x_{i1}^k - a_{j1})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_p(\mathbf{x}_i^k, \mathbf{a}_j))^{p-1}} \quad \text{and}$$

$$v_{1ij} = w_{1ij} \frac{((x_{i2}^k - a_{j2})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_p(\mathbf{x}_i^k, \mathbf{a}_j))^{p-1}}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n.$$

For the second part of the objective function we define

$$\alpha_1 = ((x_{i1} - x_{r1})^2 + \epsilon)^{1/2}, \quad \beta_1 = ((x_{i1}^k - x_{r1}^k)^2 + \epsilon)^{(p-1)/2},$$
$$\alpha_2 = ((x_{i2} - x_{r2})^2 + \epsilon)^{1/2}, \quad \text{and} \quad \beta_2 = ((x_{i2}^k - x_{r2}^k)^2 + \epsilon)^{(p-1)/2},$$

and using the Hölder Inequality it follows that

$$\sum_{i=1}^{m-1}\sum_{r=i+1}^{m} w_{2ir}\,\bar{\ell}_p(\mathbf{x}_i,\mathbf{x}_r) \geq \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} u_{2ir}\,|x_{i1}-x_{r1}| + \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} v_{2ir}\,|x_{i2}-x_{r2}|, \quad (5.34)$$

where

$$u_{2ir} = w_{2ir}\,\frac{((x_{i1}^k - x_{r1}^k)^2+\epsilon)^{(p-1)/2}}{(\bar{\ell}_p(\mathbf{x}_i^k,\mathbf{x}_r^k))^{p-1}} \quad \text{and}$$

$$v_{2ir} = w_{2ir}\,\frac{((x_{i2}^k - x_{r2}^k)^2+\epsilon)^{(p-1)/2}}{(\bar{\ell}_p(\mathbf{x}_i^k,\mathbf{x}_r^k))^{p-1}}, \quad i=1,\ldots,m-1, \quad r=i+1,\ldots,m.$$

Combining the two results, (5.33) and (5.34), we obtain

$$\begin{aligned}\widetilde{PM}(\mathbf{X}) \geq & \sum_{i=1}^{m}\sum_{j=1}^{n} u_{1ij}\,|x_{i1}-a_{j1}| + v_{1ij}\,|x_{i2}-a_{j2}| \\ & + \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} u_{2ir}\,|x_{i1}-x_{r1}| + v_{2ir}\,|x_{i2}-x_{r2}|. \end{aligned} \quad (5.35)$$

Minimizing both sides of (5.35) and defining $\mathbf{X}^R$ as in the SFMLP case, we have

$$\begin{aligned}\widetilde{PM}(\mathbf{X}^*) \geq \min_{\mathbf{X}^R}\Bigg\{ & \sum_{i=1}^{m}\sum_{j=1}^{n} u_{1ij}\,|x_{i1}^R-a_{j1}| + v_{1ij}\,|x_{i2}^R-a_{j2}| \\ & + \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} u_{2ir}\,|x_{i1}^R-x_{r1}| + v_{2ir}\,|x_{i2}^R-x_{r2}|\Bigg\}. \end{aligned}$$

Let $\widetilde{PM}_{1ij}(\mathbf{X})$ and $\widetilde{PM}_{2ir}(\mathbf{X})$ denote the terms of the first and second sum in (5.16), respectively. Then, similar to the SFMLP case, the created weights are given by

$$u_{1ij} = \left|\frac{\partial \widetilde{PM}_{1ij}(\mathbf{X}^k)}{\partial x_{i1}}\right|, \qquad v_{1ij} = \left|\frac{\partial \widetilde{PM}_{1ij}(\mathbf{X}^k)}{\partial x_{i2}}\right|, \quad i=1,\ldots,m,\ j=1,\ldots,n.$$

$$u_{2ir} = \left|\frac{\partial \widetilde{PM}_{2ir}(\mathbf{X}^k)}{\partial x_{i1}}\right|, \qquad v_{2ir} = \left|\frac{\partial \widetilde{PM}_{2ir}(\mathbf{X}^k)}{\partial x_{i2}}\right|, \quad i=1,\ldots,m-1,\ r=i+1,\ldots,m.$$

If we denote the right hand side of the inequality (5.35) by $\widetilde{PMB}^k(\mathbf{X}^R)$, then a bound for the multi-facility location problem is obtained at any iteration k by solving a multi-facility rectangular distance location problem $\min_{\mathbf{X}^R} \widetilde{PMB}^k(\mathbf{X}^R)$. The convergence of this bound, i.e., $\lim_{k\to\infty} \widetilde{PMB}^k(\mathbf{X}^*) = \widetilde{PM}(\mathbf{X}^*)$ can be obtained as in the single facility case (Property 5.2.2).

5.2.4 Generalization to MFMLP with the $\tilde{\ell}_{bp}$-norm

Similar to the single-facility case, using the Property 1.2.1 and the relation (5.14), the rectangular bound $\tilde{MB}^k(\mathbf{X}^R)$ can be devised for the MFMLP with the ℓ_{bp}-norm as follows:

$$\tilde{MB}^k(\mathbf{X}^R) = \min_{\mathbf{X}^R} \left\{ \sum_{i=1}^{m}\sum_{j=1}^{n} u_{1ij}\,|x_{i1}^R - a_{j1}| + v_{1ij}\,|x_{i2}^R - a_{j2}| \right.$$
$$\left. + \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} u_{2ir}\,|x_{i1}^R - x_{r1}| + v_{2ir}\,|x_{i2}^R - x_{r2}| \right\}, \qquad (5.36)$$

where

$$u_{1ij} = w_{1ij}\,\frac{b_1\,((x_{i1}^k - a_{j1})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_{bp}(\mathbf{x}_i^k, \mathbf{a}_j))^{p-1}} \quad \text{and}$$

$$v_{1ij} = w_{1ij}\,\frac{b_2\,((x_{i2}^k - a_{j2})^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_{bp}(\mathbf{x}_i^k, \mathbf{a}_j))^{p-1}}, \text{ for } i = 1,\ldots,m,\ j = 1,\ldots,n,$$

$$u_{2ir} = w_{2ir}\,\frac{b_1\,((x_{i1}^k - x_{r1}^k)^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_{bp}(\mathbf{x}_i^k, \mathbf{x}_r^k))^{p-1}} \quad \text{and}$$

$$v_{2ir} = w_{2ir}\,\frac{b_2\,((x_{i2}^k - x_{r2}^k)^2 + \epsilon)^{(p-1)/2}}{(\tilde{\ell}_{bp}(\mathbf{x}_i^k, \mathbf{x}_r^k))^{p-1}}, \text{ for } i = 1,\ldots,m-1,\ r = i+1,\ldots,m.$$

5.3 Fixed Point Optimality Condition

In the discussion of the Weiszfeld procedure for SFMLP we have mentioned the difficulties caused by the discontinuities in the derivatives of $PS(\mathbf{x})$. In particular, the ℓ_p-norm, and thus $PS(\mathbf{x})$, is not differentiable at the existing facility locations. In order to resolve this difficulty, we mentioned the use of an approximated distance function in the previous section. In this section we give a generalization of the condition for an existing facility to be the optimal solution. This condition is originally developed for the ℓ_p-norm by Juel and Love (1981a). The generalization again follows from using the scaling factors given in Property 1.2.1. Each existing facility location is checked by the condition and, if an optimal solution is found, it is not necessary to use the Weiszfeld procedure.

It is shown by Juel and Love (1981a) that a fixed facility location is optimum for a SFMLP with the ℓ_p-norm if and only if

$$\sum_{t=1}^{2} \left| \sum_{i \neq j} w_i \frac{|a_{jt} - a_{it}|^p}{(a_{jt} - a_{it})} \left(\ell_p(\mathbf{a}_j, \mathbf{a}_i)\right)^{1-p} \right|^{p/(p-1)} \leq w_j^{p/(p-1)}, \quad j = 1, \ldots, n. \tag{5.37}$$

To obtain the condition for the SFMLP with the ℓ_{bp}-norm we substitute the terms $(a_{jt} - a_{it})$ with $b_t^{1/p}(a_{jt} - a_{it})$, $t = 1, 2$, as suggested by the Property 1.2.1. Thus, we have

$$\sum_{t=1}^{2} \left| \sum_{i \neq j} b_t^{-1/p} w_i \frac{|a_{jt} - a_{it}|^p}{(a_{jt} - a_{it})} \left(\ell_{bp}(\mathbf{a}_j, \mathbf{a}_i)\right)^{1-p} \right|^{p/(p-1)} \leq w_j^{p/(p-1)}, \quad j = 1, \ldots, n. \tag{5.38}$$

If (5.38) holds for the jth fixed point, a_j, then the optimal solution of the location

problem with the ℓ_{bp} distances is at a_j. Note that in order for the conditions (5.37) and (5.38) to be defined, each pair of fixed facilities must be non-collinear in the axis directions. This is not a restricting assumption, because in such a case, the coordinates of a fixed facility can be perturbed by a very small amount without affecting the results significantly.

5.4 Convergence of the Weiszfeld Procedure

In this section we examine the convergence properties of the modified Weiszfeld procedure for SFMLP. For the minisum location problems with the approximated ℓ_p-norm global convergence is shown by Morris (1981) for the single-facility case, and by Frenk and Kleijn (1994) for the multi-facility case. For the unapproximated ℓ_p-norm single-facility problem Brimberg and Love (1992) prove local convergence of the Weiszfeld procedure. The authors also prove global convergence for the same problem (Brimberg and Love, 1993b). Note that, all of these convergence results apply for values of p in the interval $[1, 2]$. This is not a restricting interval since for a given transportation network the optimal value of p for the ℓ_p-norm can always be found in this interval. The convergence of the modified Weiszfeld procedure with the ℓ_{bp}-norm can be shown similar to the ℓ_p-norm case for $p \in [1, 2]$. However, as shown in Chapter 2 when the ℓ_{bp}-norm is fitted to a region, it is possible to obtain an optimal p value greater than 2, i.e., the parameter p is not necessarily confined to the interval $[1, 2]$. Therefore, for the minisum location problems with the ℓ_{bp}-norm, convergence

properties for values of p greater than 2 are of interest.

We have already shown that the Weiszfeld procedure used for the ℓ_p-norm SFMLP is applicable to the ℓ_{bp}-norm SFMLP after some modification. Therefore, our specific interest in this section is to analyze the convergence properties of the Weiszfeld procedure when it is used to solve the ℓ_p-norm SFMLP where $p > 2$.

We rewrite the iteration function (5.12) as follows:

$$x_t^{k+1} = x_t^k - \frac{1}{\Theta_t^k} \frac{\partial \widetilde{PS}(\mathbf{x}^k)}{\partial x_t}, \quad t = 1, 2, \tag{5.39}$$

where

$$\Theta_t^k = \sum_{j=1}^{n} w_j \left((x_t^k - a_{jt})^2 + \epsilon\right)^{\frac{p}{2}-1} \left(\bar{\ell}_p(\mathbf{x}^k, \mathbf{a}_j)\right)^{1-p}, \quad t = 1, 2. \tag{5.40}$$

Thus the Weiszfeld procedure is indeed a steepest-descent procedure with a varying step size $1/\Theta_t^k$ at each iteration. It is well-known that when the procedure is applied to the ℓ_p-norm SFMLP with $p > 2$, an iterate may overshoot (Kuhn, 1973). In other words, the descent property of the objective function may not be guaranteed. The descent property is stated as $\widetilde{PS}(\mathbf{x}^{k+1}) < \widetilde{PS}(\mathbf{x}^k)$. In order to remedy this problem the iteration function (5.39) can be modified in several ways: introduction of a factor that would change the step size; changing the direction of descent; changing both step size and the direction of descent. A recent book by Bertsekas (1995, Chapter 1) includes a comprehensive review of these approaches.

Our primary concern is the convergence of the iterative procedure rather then the speed of its convergence. On the other hand, we already know that

the Weiszfeld procedure performs the iterations by moving in the steepest-descent direction. Therefore, we choose to explore the first alternative. For that purpose, we aim to find a *good approximation* of the step size factor Ω to be introduced in (5.39) as follows:

$$x_t^{k+1} = x_t^k - \Omega \frac{1}{\Theta_t^k} \frac{\partial PS(\mathbf{x}^k)}{\partial x_t}, \quad t = 1, 2. \tag{5.41}$$

Since the iterations are performed in the steepest descent direction, the existence of a step size that ensures the descent property is readily known (Ortega and Rheinboldt, 1970, page 243). In order to find an approximate step size factor we compare the Weiszfeld procedure with a modified Newton method. This approach was first suggested by Harris (1978) in order to speed up the Weiszfeld iterative procedure when it is used for the Euclidean distance single-facility minisum location problem. It is later used in the context of solving minisum location problems where the distances are given by the powers of Euclidean distances (Chen, 1984a), and functions of Euclidean distances (Chen, 1984b). For the former case where the distances are given by $\ell_2^n(\mathbf{x})$, Chen was able to obtain rapid convergence by using the Weiszfeld procedure with a step size factor $2/n$. If the step size factor is not considered, then the convergence is ensured only for values of $n \in [1, 3]$ (Cooper, 1968). However, with the inclusion of the step size factor $2/n$ convergence is obtained for values of n up to 100 (Chen, 1984a).

If the Newton method is used to solve the SFMLP, then the iteration function

is given by

$$x_t^{k+1} = x_t^k - \mathrm{H}_{\mathbf{x}^k}^{-1} \frac{\partial \widetilde{PS}(\mathbf{x}^k)}{\partial x_t}, \quad t = 1, 2, \tag{5.42}$$

where the Hessian $H_{\mathbf{x}}$ is

$$H_{\mathbf{x}} = \begin{bmatrix} \frac{\partial^2 \widetilde{PS}(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 \widetilde{PS}(\mathbf{x})}{\partial x_1 \partial x_2} \\ \\ \frac{\partial^2 \widetilde{PS}(\mathbf{x}^k)}{\partial x_2 \partial x_1} & \frac{\partial^2 \widetilde{PS}(\mathbf{x})}{\partial x_2^2} \end{bmatrix}.$$

Following a similar analysis given by Chen(1984a), we devise a step size factor Ω by using modified Newton iterations (5.42). In order to prevent the possibility of oscillation in the iterative process we consider only the the diagonal elements of the Hessian as suggested by Harris (1978) and Chen (1984a). This also greatly simplifies the Newton method, and ensures the semidefiniteness of the Hessian. Defining $\mathbf{s}_j = \mathbf{x} - \mathbf{a}_j$, the iteration function (5.42) becomes

$$x_t^{k+1} = x_t^k - \left(\frac{\partial^2 \widetilde{PS}(\mathbf{x})}{\partial x_t^2} \right)^{-1} \frac{\partial \widetilde{PS}(\mathbf{x}^k)}{\partial x_t}, \quad t = 1, 2, \tag{5.43}$$

where

$$\frac{\partial^2 \widetilde{PS}(\mathbf{x})}{\partial x_1^2} = \sum_{j=1}^{n} w_j \left(s_{j1}^2 + \epsilon\right)^{\frac{p}{2}-2} \left(\tilde{\ell}_p(\mathbf{s}_j)\right)^{1-2p} \left[(p-1)\, s_{j1}^2 \left(s_{j2}^2 + \epsilon\right)^{\frac{p}{2}} + \epsilon \left(\tilde{\ell}_p(\mathbf{s}_j)\right)^p\right], \tag{5.44}$$

and

$$\frac{\partial^2 \widetilde{PS}(\mathbf{x})}{\partial x_2^2} = \sum_{j=1}^{n} w_j \left(s_{j2}^2 + \epsilon\right)^{\frac{p}{2}-2} \left(\tilde{\ell}_p(\mathbf{s}_j)\right)^{1-2p} \left[(p-1)\, s_{j2}^2 \left(s_{j1}^2 + \epsilon\right)^{\frac{p}{2}} + \epsilon \left(\tilde{\ell}_p(\mathbf{s}_j)\right)^p\right] \tag{5.45}$$

Consider first the diagonal entry (5.44). We approximate (5.44) by replacing the term (x_1^2) with $(x_1^2+\epsilon)$ and by deleting the term $(\epsilon\bar{\ell}_p(\mathbf{x},\mathbf{a}_j))$ for small $\epsilon > 0$. Thus, we have

$$\frac{\partial^2\widetilde{PS}(\mathbf{x})}{\partial x_1^2} \approx \sum_{j=1}^{n} w_j\,(s_{j1}^2+\epsilon)^{\frac{p}{2}-2}\,(\bar{\ell}_p(\mathbf{s}_j))^{1-2p}(p-1)\,(s_{j1}^2+\epsilon)\,(s_{j2}^2+\epsilon)^{\frac{p}{2}}, \tag{5.46}$$

and with some further arrangement we obtain

$$\frac{\partial^2\widetilde{PS}(\mathbf{x})}{\partial x_1^2} \approx \sum_{j=1}^{n} w_j\,(s_{j1}^2+\epsilon)^{\frac{p}{2}-1}\,(\bar{\ell}_p(\mathbf{s}_j))^{1-p}(p-1)\,\mathcal{H}_j, \tag{5.47}$$

where

$$\mathcal{H}_j = \frac{(s_{j2}^2+\epsilon)^{\frac{p}{2}}}{(s_{j1}^2+\epsilon)^{\frac{p}{2}}+(s_{j2}^2+\epsilon)^{\frac{p}{2}}}.$$

$\mathcal{H}_j$ can be approximated by letting $\epsilon \to 0$. Denoting the approximate value by $\tilde{\mathcal{H}}_j$ and with some further rearrangement, we have

$$\tilde{\mathcal{H}}_j = \frac{|s_{j2}|^p}{|s_{j1}|^p+|s_{j2}|^p} = \frac{1}{1+|\cot(\phi_j)|^p}$$

where ϕ_j specifies the approximate value of the angle between the horizontal axis and the line connecting $\mathbf{x}$ and $\mathbf{a}_j$. If the existing facility locations are uniformly distributed over the region of interest, then ϕ_j can be taken as uniformly distributed over $0 \le \phi < 2\pi$. It can easily be verified that the average value of $\tilde{\mathcal{H}}_j$ is 1/2. Replacing $\mathcal{H}_j$ with 1/2 in (5.47) we have

$$\frac{\partial^2\widetilde{PS}(\mathbf{x})}{\partial x_1^2} \approx \sum_{j=1}^{n} w_j\,(s_{j1}^2+\epsilon)^{\frac{p}{2}-1}\,(\bar{\ell}_p(\mathbf{s}_j))^{1-p}\frac{(p-1)}{2}, \tag{5.48}$$

equivalently, using (5.40)

$$\frac{\partial^2 \widetilde{PS}(\mathbf{x})}{\partial x_1^2} \approx \frac{(p-1)}{2}\Theta_1^k. \tag{5.49}$$

Carrying similar steps to (5.46) – (5.48), it can be verified that

$$\frac{\partial^2 \widetilde{PS}(\mathbf{x})}{\partial x_2^2} \approx \frac{(p-1)}{2}\Theta_2^k. \tag{5.50}$$

However for this case, instead of $\mathcal{H}_j$, we will have $\mathcal{V}_j$ where

$$\tilde{\mathcal{V}}_j = \frac{|s_{j1}|^p}{|s_{j1}|^p + |s_{j2}|^p} = \frac{1}{1 + |\tan(\phi_j)|^p}.$$

The average value of $\tilde{\mathcal{V}}_j$ is also 1/2. (5.49) and (5.50) together with (5.43) suggest that the step size factor Ω can be taken as $2/(p-1)$, i.e., we have

$$x_t^{k+1} = x_t^k - \frac{2}{p-1}\frac{1}{\Theta_t^k}\frac{\partial PS(\mathbf{x}^k)}{\partial x_t}, \quad t = 1, 2, \tag{5.51}$$

Note that, for $2 < p \leq 3$, $1 < 2/(p-1) \leq 2$, i.e., the new step size used in the Weiszfeld procedure is greater than the original step size. This may cause an iterate to fall outside the convex hull of the existing facility locations. In order to avoid this difficulty, we suggest that for $p \in [2, 3]$ a smaller step size factor $2/p$ should be used. For $p > 3$, the step size can be taken as $2/(p-1)$. Note that, the step size is a decreasing function of p, suggesting that, as the p value increases we need to use a smaller step size factor in the Weiszfeld procedure.

In order to see the effect of introducing a step size factor Ω into the Weiszfeld iterative procedure we conducted some numerical tests. For that purpose, we first

generated six SFMLP's with random existing facility locations, three with unit weights and three with random weights. The number of existing facility locations in each group was 10, 25 and 50. Then, we ran the iterative procedure (5.41) with incremental values of Ω from 0.10 to 2.0. We employed the rectangular bounding method given in Section 5.2, and as the stopping criterion we used 0.01% difference between the bound value and the objective function value, or 300 as the maximum number of iterations, whichever is reached first. As the initial point of iterations we used the center of gravity location. The results of six test problems are given in Tables 5.1–5.6. In these tables an "X" represents non-convergence up to 300 iterations. The pattern of the step sizes that provide convergence is evident. As the value of p increases the step size factor decreases, a pattern which resembles the function $2/(p-1)$. As can be observed in Tables 5.1–5.6, for each value of p used in the tests a step size factor of $2/(p-1)$ provides convergence.

To test the performance of the iteration function (5.51) further, we used a pathological example given by Brimberg and Love (1993b). The example is concerned with locating a single facility with respect to four existing facilities located at points $\mathbf{a}_1 = (0,0)$, $\mathbf{a}_2 = (0,10)$, $\mathbf{a}_3 = (10,10)$ and $\mathbf{a}_4 = (10,0)$, with weights $w_1 = 2$, $w_2 = 2$, $w_3 = 1$ and $w_4 = 1$. The authors start the iterations at $\mathbf{x} = (0,9)$ and observe that the iterates oscillate for $p > 2$. Our test results are given in Table 5.7. As can be observed in this table, the use of a step size factor $2/(p-1)$ provides convergence.

Ω	p																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2.00	10	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.80	11	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.60	13	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.40	15	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.20	18	28	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.00	22	10	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.95	23	10	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.90	25	11	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.85	26	12	23	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.80	28	13	14	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.75	30	14	9	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.70	32	15	10	32	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.65	35	16	11	15	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.60	38	18	12	9	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.55	42	20	13	10	12	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.50	46	22	15	11	9	8	X	X	X	X	X	X	X	X	X	X	X	X	X
0.45	52	25	17	13	11	9	23	X	X	X	X	X	X	X	X	X	X	X	X
0.40	58	28	19	15	12	11	10	17	X	X	X	X	X	X	X	X	X	X	X
0.35	67	33	22	17	15	13	11	11	14	20	28	X	X	X	X	X	X	X	X
0.30	78	39	27	21	17	15	14	13	12	12	11	11	11	16	22	26	28	16	X
0.25	94	47	32	26	22	19	17	16	15	14	14	13	13	13	13	13	13	13	13
0.20	118	59	41	33	28	24	22	21	19	19	18	17	17	17	16	16	16	16	16
0.15	159	80	56	44	38	34	31	28	27	26	25	24	23	23	23	22	22	22	21
0.10	239	121	85	68	58	52	47	44	42	40	38	37	36	36	35	34	34	33	33

Table 5.1: Modified Weiszfeld Iterations: 10 Facilities, Unit Weights

Ω	p																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2.00	3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.80	4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.60	4	21	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.40	5	11	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.20	6	6	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.00	8	4	13	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.95	8	3	10	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.90	8	3	9	42	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.85	9	2	7	24	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.80	10	4	6	16	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.75	10	5	5	12	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.70	11	6	4	9	26	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.65	12	7	4	7	16	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.60	13	8	6	5	11	28	X	X	X	X	X	X	X	X	X	X	X	X	X
0.55	14	10	7	5	8	16	X	X	X	X	X	X	X	X	X	X	X	X	X
0.50	16	11	9	7	6	10	18	X	X	X	X	X	X	X	X	X	X	X	X
0.45	18	13	10	8	7	6	10	18	X	X	X	X	X	X	X	X	X	X	X
0.40	21	15	12	10	8	7	6	9	14	X	X	X	X	X	X	X	X	X	X
0.35	24	17	14	12	10	9	7	6	6	8	13	20	X	X	X	X	X	X	X
0.30	29	21	17	15	13	11	9	7	6	7	8	10	11	13	16	19	18	X	X
0.25	35	25	21	18	16	14	12	9	5	8	8	8	8	9	9	10	12	14	17
0.20	45	32	27	24	21	18	15	12	6	9	11	11	11	11	11	11	11	11	11
0.15	60	44	37	33	29	25	22	18	11	13	15	15	15	15	15	15	15	15	15
0.10	92	68	58	50	45	39	34	28	20	19	23	24	24	24	24	24	24	24	25

Table 5.2: Modified Weiszfeld Iterations: 10 Facilities, Random Weights

Ω	p																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2.00	3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.80	2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.60	2	12	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.40	2	6	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.20	2	5	13	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.00	3	6	6	13	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.95	3	6	7	9	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.90	3	6	7	7	17	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.85	3	7	7	7	11	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.80	4	7	8	8	8	17	X	X	X	X	X	X	X	X	X	X	X	X	X
0.75	4	8	8	8	8	9	21	X	X	X	X	X	X	X	X	X	X	X	X
0.70	4	8	9	8	8	8	9	19	X	X	X	X	X	X	X	X	X	X	X
0.65	5	9	9	9	9	8	8	7	7	X	X	X	X	X	X	X	X	X	X
0.60	5	10	10	10	10	9	9	8	10	12	15	17	19	X	X	X	X	X	X
0.55	5	11	11	11	10	10	10	9	9	8	10	11	11	12	11	11	10	9	8
0.50	6	12	12	12	12	11	11	10	9	9	8	8	7	7	7	7	8	8	7
0.45	7	13	14	13	13	12	12	11	10	10	9	8	8	7	7	7	6	6	6
0.40	8	15	16	15	14	14	13	12	11	10	9	8	7	7	6	6	6	5	5
0.35	9	17	18	17	17	16	15	13	12	10	8	6	4	7	8	9	9	9	9
0.30	10	20	21	20	19	18	17	16	14	12	9	5	6	11	12	13	13	13	13
0.25	13	24	26	24	23	22	20	18	16	13	8	6	12	14	15	15	15	15	14
0.20	16	30	32	31	29	27	25	23	19	15	9	12	19	21	22	22	21	21	21
0.15	21	40	43	41	39	36	34	30	25	17	13	22	28	30	30	30	30	29	29
0.10	32	61	65	62	58	55	50	44	35	21	21	39	45	47	48	48	47	46	45

Table 5.3: Modified Weiszfeld Iterations: 25 Facilities, Unit Weights

Ω	p																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2.00	3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.80	2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.60	2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.40	2	11	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.20	2	5	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.00	3	5	11	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.95	3	5	8	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.90	3	5	6	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.85	4	5	6	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.80	4	6	7	7	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.75	4	6	7	8	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.70	5	7	8	8	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.65	5	7	8	9	15	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.60	6	8	9	10	10	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.55	6	9	10	11	11	12	X	X	X	X	X	X	X	X	X	X	X	X	X
0.50	7	10	11	12	12	12	13	15	X	X	X	X	X	X	X	X	X	X	X
0.45	8	11	13	13	14	13	13	13	13	13	13	13	13	X	X	X	X	X	X
0.40	9	12	14	15	15	15	14	13	12	11	11	11	11	11	12	12	12	12	12
0.35	10	14	16	17	17	17	16	14	12	10	9	7	7	7	7	7	8	8	9
0.30	12	17	19	20	20	19	18	16	14	12	10	8	8	7	7	7	7	7	8
0.25	15	20	23	24	24	23	21	18	15	12	10	8	7	7	8	10	11	12	12
0.20	19	25	29	31	30	29	26	22	18	13	7	6	6	5	5	5	5	6	6
0.15	25	34	39	41	41	38	34	29	22	13	9	9	9	9	9	9	8	8	7
0.10	39	52	59	62	61	57	50	41	29	15	15	16	18	23	25	24	23	21	17

Table 5.4: Modified Weiszfeld Iterations: 25 Facilities, Random Weights

Ω	p																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2.00	3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.80	4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.60	4	9	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.40	5	5	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.20	6	3	7	18	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.00	7	4	4	6	10	19	X	X	X	X	X	X	X	X	X	X	X	X	X
0.95	8	5	3	5	8	12	X	X	X	X	X	X	X	X	X	X	X	X	X
0.90	8	5	3	5	6	9	16	X	X	X	X	X	X	X	X	X	X	X	X
0.85	9	5	2	5	5	6	10	22	X	X	X	X	X	X	X	X	X	X	X
0.80	10	6	2	5	6	6	8	12	X	X	X	X	X	X	X	X	X	X	X
0.75	10	6	3	5	6	7	7	8	15	X	X	X	X	X	X	X	X	X	X
0.70	11	6	3	6	7	7	7	7	8	18	X	X	X	X	X	X	X	X	X
0.65	12	7	3	6	7	8	8	8	8	8	18	X	X	X	X	X	X	X	X
0.60	13	8	4	7	8	9	9	9	9	9	9	9	X	X	X	X	X	X	X
0.55	15	8	4	8	9	9	10	10	10	10	10	10	13	X	X	X	X	X	X
0.50	16	9	5	9	10	10	11	11	11	11	11	11	11	13	X	X	X	X	X
0.45	18	10	5	10	11	12	12	12	12	12	12	12	12	12	13	21	X	X	X
0.40	20	12	6	11	13	13	14	14	14	14	14	14	14	14	14	14	14	21	X
0.35	23	14	7	13	15	15	16	16	16	16	16	16	16	16	16	16	15	15	16
0.30	27	16	8	15	17	18	19	19	19	19	19	19	19	18	18	18	18	18	18
0.25	33	19	10	18	21	22	23	23	23	23	23	23	23	22	22	22	22	22	22
0.20	42	24	13	23	26	28	29	29	29	29	29	29	28	28	28	28	28	28	28
0.15	56	32	18	31	35	37	38	39	39	39	39	38	38	38	38	38	37	37	37
0.10	84	49	27	46	53	56	58	59	59	59	59	58	58	57	57	57	56	56	56

Table 5.5: Modified Weiszfeld Iterations: 50 Facilities, Unit Weights

Ω	p																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2.00	4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.80	4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.60	4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.40	6	14	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.20	7	5	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.00	9	7	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.95	10	7	13	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.90	10	8	9	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.85	11	8	8	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.80	12	9	7	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.75	13	10	8	17	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.70	14	11	9	11	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.65	15	11	9	8	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.60	17	12	10	9	12	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.55	18	14	11	10	9	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.50	20	15	13	12	11	11	X	X	X	X	X	X	X	X	X	X	X	X	X
0.45	23	17	14	13	13	12	12	X	X	X	X	X	X	X	X	X	X	X	X
0.40	26	20	17	15	15	14	14	14	X	X	X	X	X	X	X	X	X	X	X
0.35	30	23	20	18	17	16	16	16	16	21	X	X	X	X	X	X	X	X	X
0.30	35	27	23	21	20	19	19	19	19	19	19	23	X	X	X	X	X	X	X
0.25	42	32	28	26	25	24	24	24	25	25	26	26	27	27	27	X	X	X	X
0.20	53	41	36	33	31	31	30	31	31	31	32	32	33	33	34	34	34	35	35
0.15	72	56	48	45	43	42	41	42	42	43	44	45	46	47	48	49	50	51	52
0.10	108	84	73	68	65	63	63	64	65	66	67	69	70	72	73	74	75	76	77

Table 5.6: Modified Weiszfeld Iterations: 50 Facilities, Random Weights

Ω	p																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2.00	11	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.80	12	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.60	14	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.40	16	22	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.20	19	12	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.00	24	10	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.95	25	9	19	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.90	27	6	15	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.85	29	9	12	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.80	31	9	10	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.75	33	9	8	19	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.70	36	9	7	15	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.65	39	10	7	10	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.60	42	12	8	8	13	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.55	46	15	9	7	12	X	X	X	X	X	X	X	X	X	X	X	X	X	X
0.50	51	18	9	7	8	11	X	X	X	X	X	X	X	X	X	X	X	X	X
0.45	57	21	12	5	7	11	X	X	X	X	X	X	X	X	X	X	X	X	X
0.40	64	25	15	11	8	9	13	X	X	X	X	X	X	X	X	X	X	X	X
0.35	74	30	19	14	11	8	9	12	14	X	X	X	X	X	X	X	X	X	X
0.30	87	37	24	18	14	11	9	10	13	17	X	X	X	X	X	X	X	X	X
0.25	105	47	31	23	19	15	13	11	10	11	13	17	X	X	X	X	X	X	X
0.20	132	61	42	32	26	22	19	16	15	13	12	12	13	16	19	X	X	X	X
0.15	176	85	59	47	40	34	30	27	24	21	20	18	17	16	15	14	16	17	18
0.10	266	132	94	76	66	59	54	50	48	45	43	42	40	39	38	36	31	28	27

Table 5.7: Modified Weiszfeld Iterations: Example in Brimberg and Love (1993)

Chapter 6

Duality in Constrained Location Models

A constrained multi-facility minisum location problem (MFMLP) in two dimensional Euclidean plane ($\Re^2$) is stated as follows:

$$\begin{aligned} \min \quad M(\mathbf{X}) &= \sum_{i=1}^{m}\sum_{j=1}^{n} w_{1ij}\, d(\mathbf{x}_i, \mathbf{a}_j) + \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} w_{2ir}\, d(\mathbf{x}_i, \mathbf{x}_r) \qquad (6.1) \\ \text{s.t.} \quad C_k(\mathbf{X}) &\leq c_k, \quad k = 1, \ldots, T \end{aligned}$$

where n is the number of fixed facilities and $\mathbf{a}_j = (a_{j1}, a_{j2})$, $j = 1, \ldots, n$ are the fixed facility locations; m is the number of new facilities and $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$ are their unknown locations, $\mathbf{x}_i = (x_{i1}, x_{i2})$, $i = 1, \ldots, m$; $d(\mathbf{u}, \mathbf{v})$ is some distance function (usually a norm) used to calculate the distance between any two points $\mathbf{u}, \mathbf{v} \in \Re^2$; w_{1ij} converts the distance between new and existing facilities i and j into a cost, where $i = 1, \ldots, m$, $j = 1, \ldots, n$; and w_{2ir} converts the distance between two new facilities i and r into a cost, where $i = 1, \ldots, m-1$, $r = i+1, \ldots, m$. The constraints

define a convex and nonempty set $C = \{\mathbf{X} : C_k(\mathbf{X}) \leq c_k, k = 1, \ldots, T\}$.

The solution of the dual formulation of a location problem provides a lower bound on the objective function of the primal problem (Wendell and Peterson, 1984; Dowling and Love, 1987). Such a lower bound (or stopping criterion) is useful when the primal is solved by Weiszfeld's one-point iteration algorithm or its generalizations (Weiszfeld, 1937; Kuhn, 1973; Morris and Verdini, 1979). The dual formulation also provides a theoretical framework in devising procedures for solving location-allocation problems. Love and Juel (1982) used this approach to develop efficient heuristic algorithms for location-allocation problems. Furthermore, the dual formulation avoids the difficulty arising from the discontinuities in the derivatives of the primal objective function (Love and Morris, 1975b).

The dual formulation for the various minisum location models are found in the literature. The dual for the original problem, known as the Fermat problem where there exists three fixed facilities with unit weights, was published by Fasbender (1846). Kuhn and Kuenne (1962) provided the dual for the single-facility Euclidean distance problem. Later, Bellman (1965) and Witzgall (1964) developed the dual formulations for this problem by using quasilinearization and conjugate function theory, respectively. Kuhn (1967) presented a comprehensive treatment of the dual formulations for the single-facility Euclidean distance minisum location problems. The dual for the multi-facility Euclidean distance problem was given by Francis and

Cabot (1972). Love and Kraemer (1973) developed a generalized dual for this problem by considering linear constraints, and also provided a decomposition solution method. The dual formulation using quasilinearization and a decomposition solution method for the linearly constrained hyperbolically approximated ℓ_p distance multi-facility location model were given by Love (1974). Juel and Love (1981b) developed the dual for this problem with generalized distances using conjugate function theory. The dual formulation for the general norm single-facility location problem with linear and distance constraints was given by Juel and Love (1987). Idrissi et al. (1989) consider the dual of a multi-facility location problem with linear and rectangular distance constraints and provide a solution procedure using a decomposition method. Michelot (1993), in a review of the mathematical properties of continuous location models, provides the conjugate dual formulation of a multi-facility location problem, and a linearly constrained single-facility location problem. The author then examines the optimality conditions for these problems and provides a proximal mapping algorithm for the solution. A proximal mapping algorithm can be regarded as a special implementation of the gradient projection method (Bertsekas, 1995). Scott et al. (1995) give a summary of conjugate duality in facility location problems with Euclidean and ℓ_p distances.

This chapter is organized as follows: In Section 6.1, we develop a generalization of the dual formulation given by Juel and Love (1987) for the single-facility location

problem to the ℓ_p-norm multi-facility location problem. We develop both the lagrangian and the conjugate dual problems by considering two types of constraints on the location of new facilities given by (6.1). *Linear* constraints define a feasible region where the new facilities must be located, and *distance* constraints where each constraint imposes an upper bound on the distance between a fixed facility and a new facility, or between two new facilities. In section 6.2, we compare the two dual formulations and show that the conjugate dual problem can be converted to the lagrangian dual problem which has a linear objective function. In Section 6.3 we generalize our results for the ℓ_p-norm MFMLP to the ℓ_{bp}-norm MFMLP. In Section 6.4, we give a numerical example on formulating and solving the dual of an ℓ_{bp}-norm constrained MFMLP.

6.1 Dual of the ℓ_p-norm Location Model

In this section we develop the dual formulation for the following problem:

$$\min \quad PM(\mathbf{X}) = \sum_{i=1}^{m}\sum_{j=1}^{n} w_{1ij}\, \ell_p(\mathbf{x}_i, \mathbf{a}_j) + \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} w_{2ir}\, \ell_p(\mathbf{x}_i, \mathbf{x}_r) \qquad (6.2)$$

$$\text{s.t.} \quad \ell_p(\mathbf{x}_i, \mathbf{a}_j) \le d_{1ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n,$$

$$\ell_p(\mathbf{x}_i, \mathbf{x}_r) \le d_{2ir}, \quad i = 1, \ldots, m-1, \quad r = i+1, \ldots, m,$$

$$G_{ik}(\mathbf{x}_i) \le r_{ik}, \quad i = 1, \ldots, m, \quad k = 1, \ldots, K$$

where $m.K$ inequalities in the constraint set define the feasible region for the new facility locations, and d_{1ij} and d_{2ir} are the upper bounds for the ℓ_p distance between a new facility i and a fixed facility j, and between the new facilities i and r, respectively. It can easily be verified that the constraint set is convex. Also note that the linear constraints $G_{ik}(\mathbf{x}_i)$ can be expressed as $g_{ik1}x_{i1} + g_{ik2}x_{i2}$ where g_{ik1} and g_{ik2} are scalars.

6.1.1 Lagrangian Dual with Quasilinearization

Introducing the Lagrange multipliers $\boldsymbol{\mu_1} = (\mu_{111}, \ldots, \mu_{11n}, \ldots \mu_{1mn})$, $\boldsymbol{\mu_1} \geq \mathbf{0}$, $\boldsymbol{\mu_2} = (\mu_{212}, \ldots, \mu_{21m}, \ldots \mu_{2(m-1)m})$, $\boldsymbol{\mu_2} \geq \mathbf{0}$, and $\boldsymbol{\lambda} = (\lambda_{11}, \ldots, \lambda_{1K}, \ldots \lambda_{mK})$, $\boldsymbol{\lambda} \geq \mathbf{0}$, we obtain the lagrangian dual formulation for problem (6.2) as follows:

$$\begin{aligned} \max_{\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \boldsymbol{\lambda}} \min_{\mathbf{X}} \quad & LPM(\mathbf{X}, \boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \boldsymbol{\lambda}) \\ \text{s.t.} \quad & \boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \boldsymbol{\lambda} \geq 0. \end{aligned} \tag{6.3}$$

where

$$\begin{aligned} LPM(\mathbf{X}, \boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \boldsymbol{\lambda}) \;=\; & \sum_{i=1}^{m}\sum_{j=1}^{n} w_{1ij}\, \ell_p(\mathbf{x}_i, \mathbf{a}_j) + \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} w_{2ir}\, \ell_p(\mathbf{x}_i, \mathbf{x}_r) \\ & + \sum_{i=1}^{m}\sum_{j=1}^{n} \mu_{1ij}\, (\ell_p(\mathbf{x}_i, \mathbf{a}_j) - d_{1ij}) \\ & + \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} \mu_{2ir}\, (\ell_p(\mathbf{x}_i, \mathbf{x}_r) - d_{2ir}) \\ & + \sum_{i=1}^{m}\sum_{k=1}^{K} \lambda_{ik}\, (G_{ik}(\mathbf{x}_i) - r_{ik}). \end{aligned} \tag{6.4}$$

Rearranging (6.4), we have

$$\begin{aligned} LPM(\mathbf{X}, \boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \boldsymbol{\lambda}) &= \sum_{i=1}^{m}\sum_{j=1}^{n} PMD_{1ij}(\mathbf{x}_i, \mathbf{a}_j) \\ &+ \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} PMD_{2ir}(\mathbf{x}_i, \mathbf{x}_r) \\ &- \sum_{i=1}^{m}\sum_{j=1}^{n} \mu_{1ij}\, d_{1ij} - \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} \mu_{2ir}\, d_{2ir} \\ &+ \sum_{i=1}^{m}\sum_{k=1}^{K} \lambda_{ik}\,(G_{ik}(\mathbf{x}_i) - r_{ik}). \end{aligned} \tag{6.5}$$

where

$$PMD_{1ij}(\mathbf{x}_i, \mathbf{a}_j) = (w_{1ij} + \mu_{1ij})\, \ell_p(\mathbf{x}_i, \mathbf{a}_j) \quad \text{and}$$

$$PMD_{2ir}(\mathbf{x}_i, \mathbf{x}_r) = (w_{2ir} + \mu_{2ir})\, \ell_p(\mathbf{x}_i, \mathbf{x}_r).$$

Define the vectors

$$\begin{aligned} \boldsymbol{\chi}_{1ij} &= ((x_{i1} - a_{j1}), (x_{i2} - a_{j2})), \\ \boldsymbol{\theta}_{ij} &= (\theta_{1ij}, \theta_{2ij}) \qquad \text{for} \quad i = 1, \ldots, m,\ j = 1, \ldots, n, \quad \text{and} \\ \boldsymbol{\chi}_{2ir} &= ((x_{i1} - x_{r1}), (x_{i2} - x_{r2})), \\ \boldsymbol{\pi}_{ir} &= (\pi_{1ir}, \pi_{2ir}) \qquad \text{for} \quad i = 1, \ldots, m-1,\ r = i+1, \ldots, m. \end{aligned}$$

We quasilinearize the terms $PMD_{1ij}(\mathbf{x}_i, \mathbf{a}_j)$ and $PMD_{2ir}(\mathbf{x}_i, \mathbf{x}_r)$ as described by Love (1974). Utilizing the Hölder inequality with the above vector definitions, the

quasilinearized forms are given by

$$\begin{aligned} PMD_{1ij}(\mathbf{x}_i, \mathbf{a}_j) &= \max_{\boldsymbol{\theta}_{ij}} \quad CQ_{1ij} = \langle \mathcal{X}_{1ij}, \boldsymbol{\theta}_{ij} \rangle \\ &\quad \text{s.t.} \;\; \ell_q(\boldsymbol{\theta}_{ij}) = w_{1ij} + \mu_{1ij}, \end{aligned} \tag{6.6}$$

and

$$\begin{aligned} PMD_{2ir}(\mathbf{x}_i, \mathbf{x}_r) &= \max_{\boldsymbol{\pi}_{ir}} \quad CQ_{2ir} = \langle \mathcal{X}_{2ir}, \boldsymbol{\pi}_{ir} \rangle \\ &\quad \text{s.t.} \;\; \ell_q(\boldsymbol{\pi}_{ir}) = w_{2ir} + \mu_{2ir}. \end{aligned} \tag{6.7}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product, and $1/p + 1/q = 1,\ p > 1$.

Consider the maximization problem (6.6). After rearranging and expanding, we have

$$\begin{aligned} \max\ CQ_{1ij} &= (x_{i1} - a_{j1})\theta_{1ij} + (x_{i2} - a_{j2})\theta_{2ij} \\ \text{s.t.} &\quad |\theta_{1ij}|^q + |\theta_{2ij}|^q = (w_{1ij} + \mu_{1ij})^q. \end{aligned} \tag{6.8}$$

In order to find the θ_{1ij} and θ_{2ij} values that solve (6.8) we write the lagrangian function where ψ_{ij} is the Lagrange multiplier as

$$\begin{aligned} LCQ_{1ij}(\psi_{ij}, \boldsymbol{\theta}_{ij}) &= (x_{i1} - a_{j1})\theta_{1ij} + (x_{i2} - a_{j2})\theta_{2ij} \\ &\quad - \psi_{ij}(|\theta_{1ij}|^q + |\theta_{2ij}|^q - (w_{1ij} + \mu_{1ij})^q). \end{aligned} \tag{6.9}$$

Assuming that $(x_{it} - a_{jt})$ and θ_{tij}, $t = 1, 2$, have the same sign,and noting the equality $(x_{it} - a_{jt}) = \text{sign}(x_{it} - a_{jt})\ |x_{it} - a_{jt}|$, for $t = 1, 2$, we obtain the following first-order

necessary and sufficient conditions for maximizing (6.9):

$$\frac{\partial LCQ_{1ij}}{\partial \theta_{1ij}} = (x_{i1} - a_{j1}) - q\,\psi_{ij}\,\theta_{1ij}^{q-1}\,\mathrm{sign}(x_{i1} - a_{j1}) = 0, \tag{6.10}$$

$$\frac{\partial LCQ_{1ij}}{\partial \theta_{2ij}} = (x_{i2} - a_{j2}) - q\,\psi_{ij}\,\theta_{2ij}^{q-1}\,\mathrm{sign}(x_{i2} - a_{j2}) = 0, \tag{6.11}$$

$$\frac{\partial LCQ_{1ij}}{\partial \psi_{ij}} = |\theta_{1ij}|^q + |\theta_{2ij}|^q - (w_{1ij} + \mu_{1ij})^q = 0. \tag{6.12}$$

Solving (6.10), (6.11), and (6.12) simultaneously, we have the following optimal values for θ_{1ij} and θ_{2ij}:

$$\theta_{1ij}^* = (w_{1ij} + \mu_{1ij})\,\mathrm{sign}(x_{i1} - a_{j1}) \left(\frac{|x_{i1} - a_{j1}|}{\ell_p(\mathbf{x}_i, \mathbf{a}_j)}\right)^{p-1}, \tag{6.13}$$

$$\theta_{2ij}^* = (w_{1ij} + \mu_{1ij})\,\mathrm{sign}(x_{i2} - a_{j2}) \left(\frac{|x_{i2} - a_{j2}|}{\ell_p(\mathbf{x}_i, \mathbf{a}_j)}\right)^{p-1}. \tag{6.14}$$

Carrying out a similar analysis for (6.7) we obtain

$$\pi_{1ir}^* = (w_{2ir} + \mu_{2ir})\,\mathrm{sign}(x_{i1} - x_{r1}) \left(\frac{|x_{i1} - x_{r1}|}{\ell_p(\mathbf{x}_i, \mathbf{x}_r)}\right)^{p-1}, \tag{6.15}$$

$$\pi_{2ir}^* = (w_{2ir} + \mu_{2ir})\,\mathrm{sign}(x_{i2} - x_{r2}) \left(\frac{|x_{i2} - x_{r2}|}{\ell_p(\mathbf{x}_i, \mathbf{x}_r)}\right)^{p-1}. \tag{6.16}$$

Observe that the expressions (6.13) and (6.14) can be rewritten as

$$x_{i1} = a_{j1} + \frac{(\ell_p(\mathbf{x}_i, \mathbf{a}_j))^{p-1}}{(w_{1ij} + \mu_{1ij})\,|x_{i1} - a_{j1}|^{p-2}}\,\theta_{1ij}^*, \tag{6.17}$$

$$x_{i2} = a_{j2} + \frac{(\ell_p(\mathbf{x}_i, \mathbf{a}_j))^{p-1}}{(w_{1ij} + \mu_{1ij})\,|x_{i2} - a_{j2}|^{p-2}}\,\theta_{2ij}^*, \tag{6.18}$$

and rewriting the expressions (6.15) and (6.16), we have

$$x_{i1} = x_{r1} + \frac{(\ell_p(\mathbf{x}_i, \mathbf{x}_r))^{p-1}}{(w_{2ir} + \mu_{2ir})\,|x_{i1} - x_{r1}|^{p-2}}\,\pi_{1ir}^*, \tag{6.19}$$

$$x_{i2} = x_{r2} + \frac{(\ell_p(\mathbf{x}_i, \mathbf{x}_r))^{p-1}}{(w_{2ir} + \mu_{2ir})\,|x_{i2} - x_{r2}|^{p-2}}\,\pi^*_{2ir}. \tag{6.20}$$

The optimal values of θ_{1ij} and θ_{2ij} given by (6.13) and (6.14) satisfy the equality constraint in (6.6), provided that at least one of θ^*_{1ij} and θ^*_{2ij} is nonzero in the solution of the dual problem. If both are zero in the dual solution, then it is implied by the expressions (6.17) and (6.18) that $\mathbf{x}^*_i = \mathbf{a}_j$.

On the other hand, observe that the dual variables θ^*_{tij} given by (6.13) and (6.14) can also be expressed as $(w_{1ij} + \mu_{1ij})\partial\ell_p(\mathbf{x}_i, \mathbf{a}_j)/\partial x_{it}$, $t = 1, 2$. It can easily be verified that if the hyperbolic approximation of the ℓ_p-norm (5.10) is used while solving the primal problem (to deal with the discontinuities in the derivatives), then the θ^*_{tij} are given by $(w_{1ij} + \mu_{1ij})\partial\bar{\ell}_p(\mathbf{x}_i, \mathbf{a}_j)/\partial x_{it}$, $t = 1, 2$. It follows that, in the case that the optimal solution to the primal problem is at an existing facility location, i.e., $\mathbf{x}^*_i = \mathbf{a}_j$, the dual variables at optimality should be equal to zero. If $\mathbf{x}^*_i \neq \mathbf{a}_j$, then the dual variables θ^*_{1ij} and θ^*_{2ij} will have nonzero values such that the equality constraint in (6.6) is satisfied.

Using (6.15), (6.16), (6.19) and (6.20), a similar relation applies to the dual variables π_{1ir}, π_{2ir} and the primal variables $\mathbf{x}_i$ and $\mathbf{x}_r$. Thus, we state the following property.

Property 6.1.1 *Optimality for the new facility location $\mathbf{x}_i$ is achieved at a fixed facility location $\mathbf{a}_j$ if and only if the corresponding dual variables θ^*_{1ij} and θ^*_{2ij} are zero. Similarly, the locations of two new facilities $\mathbf{x}_i$ and $\mathbf{x}_r$ coincide at optimality if*

and only if the corresponding dual variables π^*_{1ir} *and* π^*_{2ir} *are zero.*

Using Property 6.1.1 we convert the equality constraints into inequality ($\leq$) constraints in (6.6) and (6.7). Note that, as mentioned above, if the coincidence conditions do not occur, then the equalities will hold .

Let $\Omega = (\theta_{11}, \ldots, \theta_{mn}, \pi_{12}, \ldots, \pi_{(m-1)m})$ where $\theta_{ij} = (\theta_{1ij}, \theta_{2ij})$ for $i = 1, \ldots, m$, $j = 1, \ldots, n$, and $\pi_{ir} = (\pi_{1ir}, \pi_{2ir})$ for $i = 1, \ldots, m-1$, $r = i+1, \ldots, m$. Using (6.6), (6.7), the lagrangian dual problem (6.5) is rewritten as

$$\begin{aligned} \max_{\mu_1, \mu_2, \lambda, \Omega} \min_{\mathbf{X}} QLPM(\mathbf{X}, \mu_1, \mu_2, \lambda, \Omega) \;=\; & \sum_{i=1}^{m} \sum_{j=1}^{n} \langle \mathcal{X}_{1ij}, \theta_{ij} \rangle + \sum_{i=1}^{m-1} \sum_{r=i+1}^{m} \langle \mathcal{X}_{2ir}, \pi_{ir} \rangle \\ & - \sum_{i=1}^{m} \sum_{j=1}^{n} \mu_{1ij}\, d_{1ij} - \sum_{i=1}^{m-1} \sum_{r=i+1}^{m} \mu_{2ir}\, d_{2ir} \\ & + \sum_{i=1}^{m} \sum_{k=1}^{K} \lambda_{ik} \left(G_{ik}(\mathbf{x}_i) - r_{ik} \right) \end{aligned} \tag{6.21}$$

$$\begin{aligned} \text{s.t.} \quad & \ell_q(\theta_{ij}) \leq w_{1ij} + \mu_{1ij}, \;\; i = 1, \ldots, m, \; j = 1, \ldots, n, \\ & \ell_q(\pi_{ir}) \leq w_{2ir} + \mu_{2ir}, \;\; i = 1, \ldots, m-1, \; r = i+1, \ldots, m, \\ & \mu_1, \mu_2, \lambda \geq 0. \end{aligned}$$

Expanding and rearranging $QLPM(\mathbf{X}, \mu_1, \mu_2, \lambda, \Omega)$ we have

$$\begin{aligned} QLPM(\mathbf{X}, \mu_1, \mu_2, \lambda, \Omega) \;=\; & \sum_{i=1}^{m} x_{i1} \left[\sum_{j=1}^{n} \theta_{1ij} + \sum_{r=i+1}^{m} \pi_{1ir} - \sum_{r=1}^{i-1} \pi_{1ri} + \sum_{k=1}^{K} \lambda_{ik}\, g_{ik1} \right] \\ & + \sum_{i=1}^{m} x_{i2} \left[\sum_{j=1}^{n} \theta_{2ij} + \sum_{r=i+1}^{m} \pi_{2ir} - \sum_{r=1}^{i-1} \pi_{2ri} + \sum_{k=1}^{K} \lambda_{ik}\, g_{ik2} \right] \end{aligned}$$

$$- \sum_{i=1}^{m}\sum_{j=1}^{n}(a_{j1}\,\theta_{1ij} + a_{j2}\,\theta_{2ij}) - \sum_{i=1}^{m}\sum_{k=1}^{K}\lambda_{ik}\,r_{ik}$$

$$- \sum_{i=1}^{m}\sum_{j=1}^{n}\mu_{1ij}\,d_{1ij} - \sum_{i=1}^{m-1}\sum_{r=i+1}^{m}\mu_{2ir}\,d_{2ir}.$$

We calculate the first-order necessary and sufficient conditions for minimizing $QLPM(\mathbf{X}, \mu_1, \mu_2, \lambda, \Omega)$ over $\mathbf{X}$ as

$$\frac{\partial QLPM}{\partial x_{i1}} = \sum_{j=1}^{n}\theta_{1ij} + \sum_{r=i+1}^{m}\pi_{1ir} - \sum_{r=1}^{i-1}\pi_{1ri} + \sum_{k=1}^{K}\lambda_{ik}g_{ik1} = 0, \quad \text{for } i = 1,\ldots,m, \quad (6.22)$$

and

$$\frac{\partial QLPM}{\partial x_{i2}} = \sum_{j=1}^{n}\theta_{2ij} + \sum_{r=i+1}^{m}\pi_{2ir} - \sum_{r=1}^{i-1}\pi_{2ri} + \sum_{k=1}^{K}\lambda_{ik}g_{ik2} = 0, \quad \text{for } i = 1,\ldots,m. \quad (6.23)$$

Modifying $QLPM(\mathbf{X}, \mu_1, \mu_2, \lambda, \Omega)$ by using (6.22) and (6.23) and introducing the implied constraints in (6.21), we obtain the quasilinearized lagrangian dual problem as

$$\begin{aligned}\max\ DQPM(\mu_1, \mu_2, \lambda, \Omega) \;=\; & -\sum_{i=1}^{m}\sum_{j=1}^{n}(a_{j1}\,\theta_{1ij} + a_{j2}\,\theta_{2ij}) - \sum_{i=1}^{m}\sum_{k=1}^{K}\lambda_{ik}\,r_{ik} \\ & - \sum_{i=1}^{m}\sum_{j=1}^{n}d_{1ij}\,\mu_{1ij} - \sum_{i=1}^{m-1}\sum_{r=i+1}^{m}d_{2ir}\,\mu_{2ir} \qquad (6.24)\end{aligned}$$

$$\text{s.t.} \quad \sum_{j=1}^{n}\theta_{1ij} + \sum_{r=i+1}^{m}\pi_{1ir} - \sum_{r=1}^{i-1}\pi_{1ri} + \sum_{k=1}^{K}\lambda_{ik}\,g_{ik1} = 0,\ i = 1,\ldots,m,$$

$$\sum_{j=1}^{n}\theta_{2ij} + \sum_{r=i+1}^{m}\pi_{2ir} - \sum_{r=1}^{i-1}\pi_{2ri} + \sum_{k=1}^{K}\lambda_{ik}\,g_{ik2} = 0,\ i = 1,\ldots,m,$$

$$(|\theta_{1ij}|^q + |\theta_{2ij}|^q)^{1/q} \leq w_{1ij} + \mu_{1ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n,$$

$$(|\pi_{1ir}|^q + |\pi_{2ir}|^q)^{1/q} \leq w_{2ir} + \mu_{2ir}, \quad i = 1, \ldots, m-1, \; r = i+1, \ldots, m,$$

$$\lambda_{ik} \geq 0, \quad i = 1, \ldots, m, \quad k = 1, \ldots, K,$$

$$\mu_{1ij} \geq 0, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n,$$

$$\mu_{2ir} \geq 0, \quad i = 1, \ldots, m-1, \quad r = i+1, \ldots, m.$$

6.1.2 Conjugate Dual

Define the functions $\mathcal{F}_{ik}(\mathbf{x}_i)$ for $i = 1, \ldots, m$, $k = 1, \ldots, K$, $\mathcal{D}_{1ij}(\mathbf{x}_i)$ for $i = 1, \ldots, m$, $j = 1, \ldots, n$, and $\mathcal{D}_{2ir}(\mathbf{x}_i - \mathbf{x}_r)$ for $i = 1, \ldots, m-1$, $r = i+1, \ldots, m$, as follows:

$$\mathcal{F}_{ik}(\mathbf{x}_i) = \begin{cases} 0 & \text{if } G_{ik}(\mathbf{x}_i) \leq r_{ik} \\ \infty & \text{otherwise.} \end{cases} \tag{6.25}$$

$$\mathcal{D}_{1ij}(\mathbf{x}_i) = \begin{cases} 0 & \text{if } \ell_p(\mathbf{x}_i, \mathbf{a}_j) \leq d_{1ij} \\ \infty & \text{otherwise} \end{cases} \tag{6.26}$$

$$\mathcal{D}_{2ir}(\mathbf{x}_i - \mathbf{x}_r) = \begin{cases} 0 & \text{if } \ell_p(\mathbf{x}_i, \mathbf{x}_r) \leq d_{2ir} \\ \infty & \text{otherwise} \end{cases} \tag{6.27}$$

Then rewriting problem (6.2) we have

$$\begin{aligned} \min \; CPM(\mathbf{X}) &= PM(\mathbf{X}) + \sum_{i=1}^{m} \sum_{j=1}^{n} \mathcal{D}_{1ij}(\mathbf{x}_i) \\ &+ \sum_{i=1}^{m-1} \sum_{r=i+1}^{m} \mathcal{D}_{2ir}(\mathbf{x}_i - \mathbf{x}_r) + \sum_{i=1}^{m} \sum_{k=1}^{K} \mathcal{F}_{ik}(\mathbf{x}_i) \end{aligned} \tag{6.28}$$

Let $PM_{1ij}(\mathbf{x}_i, \mathbf{a}_j) = w_{1ij}\,\ell_p(\mathbf{x}_i, \mathbf{a}_j)$, $\mathcal{U}_{ij} = (u_{1ij}, u_{2ij})$ for $i = 1, \ldots, m$, $j = 1, \ldots, n$, and $PM_{2ir}(\mathbf{x}_i, \mathbf{x}_r) = w_{2ir}\,\ell_p(\mathbf{x}_i, \mathbf{x}_r)$, $\mathcal{V}_{ir} = (v_{1ir}, v_{2ir})$ for $i = 1, \ldots, m-1$, $r = i+1, \ldots, m$. Using the elements of conjugate functions as given by Witzgall (1964), it can be shown that the conjugate functions of the components of $PM(\mathbf{X})$ are

$$PM^c_{1ij}(\mathcal{U}_{ij}) = \begin{cases} \langle \mathbf{a}_j, \mathcal{U}_{ij} \rangle & \text{if } \ell^\circ_p(\mathcal{U}_{ij}) \le w_{1ij} \\ \infty & \text{otherwise} \end{cases} \tag{6.29}$$

$$PM^c_{2ir}(\mathcal{V}_{ir}) = \begin{cases} 0 & \text{if } \ell^\circ_p(\mathcal{V}_{ir}) \le w_{2ir} \\ \infty & \text{otherwise} \end{cases} \tag{6.30}$$

where c denotes the conjugate, and $\ell^\circ_p(\mathbf{x})$ is the polar function of $\ell_p(\mathbf{x})$, given by $\ell_q(\mathbf{y})$ where $(1/p) + (1/q) = 1$, and $p > 1$. On the other hand, by the definition of a conjugate function (Rockafellar, 1970, Section 12) we have

$$\mathcal{F}^c_{ik}(\mathbf{z}_{ik}) = \sup_{\mathbf{x}_i} \{\langle \mathbf{z}_{ik}, \mathbf{x}_i \rangle - \mathcal{F}_{ik}(\mathbf{x}_i)\},$$

or equivalently, considering the finite values of $\mathcal{F}_{ik}(\mathbf{x}_i)$

$$\mathcal{F}^c_{ik}(\mathbf{z}_{ik}) = \sup_{\mathbf{x}_i} \{\langle \mathbf{z}_{ik}, \mathbf{x}_i \rangle \,:\, \langle \mathbf{g}_{ik}, \mathbf{x}_i \rangle \le r_{ik}\},$$

and by the duality theorem of linear programming

$$\mathcal{F}^c_{ik}(\mathbf{z}_{ik}) = \inf_{\lambda_{ik}} \{\lambda_{ik}\, r_{ik} \,:\, \lambda_{ik}\, \mathbf{g}_{ik} = \mathbf{z}_{ik},\ \lambda_{ik} \ge 0\} \tag{6.31}$$

where $\mathbf{g}_{ik} = (g_{ik1}, g_{ik2})$, $i = 1, \ldots, m$, $k = 1, \ldots, K$, are as defined before.

We also need to determine the conjugate functions of $\mathcal{D}_{1ij}(\mathbf{x}_i)$ and $\mathcal{D}_{2ir}(\mathbf{x}_i - \mathbf{x}_r)$. It follows from the definition of the conjugate of a convex function that

$$\mathcal{D}^c_{1ij}(\boldsymbol{\beta}_{ij}) = \sup_{\mathbf{x}_i} \{\langle \boldsymbol{\beta}_{ij}, \mathbf{x}_i \rangle - \mathcal{D}_{1ij}(\mathbf{x}_i)\} \quad \text{where} \quad \boldsymbol{\beta}_{ij} = (\beta_{1ij}, \beta_{2ij}). \tag{6.32}$$

We consider only the finite values of $\mathcal{D}_{1ij}(\mathbf{x}_i)$, and thus (6.32) can be written as

$$\mathcal{D}^c_{1ij}(\boldsymbol{\beta}_{ij}) = \sup_{\mathbf{x}_i} \{\langle \boldsymbol{\beta}_{ij}, \mathbf{x}_i \rangle \ : \ \ell_p(\mathbf{x}_i, \mathbf{a}_j) \leq d_{1ij}\}.$$

Using the equality $\mathbf{x}_i = \mathbf{x}_i - \mathbf{a}_j + \mathbf{a}_j$ we have

$$\mathcal{D}^c_{1ij}(\boldsymbol{\beta}_{ij}) = \sup_{\mathbf{x}_i} \{\langle \boldsymbol{\beta}_{ij}, (\mathbf{x}_i - \mathbf{a}_j) \rangle + \langle \boldsymbol{\beta}_{ij}, \mathbf{a}_j \rangle \ : \ \ell_p(\mathbf{x}_i, \mathbf{a}_j) \leq d_{1ij}\},$$

or equivalently

$$\mathcal{D}^c_{1ij}(\boldsymbol{\beta}_{ij}) = \langle \boldsymbol{\beta}_{ij}, \mathbf{a}_j \rangle + \sup_{\mathbf{x}_i} \{\langle \boldsymbol{\beta}_{ij}, (\mathbf{x}_i - \mathbf{a}_j) \rangle \ : \ \ell_p(\mathbf{x}_i, \mathbf{a}_j) \leq d_{1ij}\}.$$

Using the definition of the polar of a convex function, we obtain

$$\mathcal{D}^c_{1ij}(\boldsymbol{\beta}_{ij}) = \langle \boldsymbol{\beta}_{ij}, \mathbf{a}_j \rangle + d_{1ij}\, \ell^o_p(\boldsymbol{\beta}_{ij}). \tag{6.33}$$

Similar to (6.32), the conjugate of $\mathcal{D}_{2ir}(\mathbf{x}_i - \mathbf{x}_r)$ is given by

$$\mathcal{D}^c_{2ir}(\boldsymbol{\gamma}_{ir}) = \sup_{\mathbf{x}_i} \{\langle \boldsymbol{\gamma}_{ij}, (\mathbf{x}_i - \mathbf{x}_r) \rangle - \mathcal{D}_{2ir}(\mathbf{x}_i - \mathbf{x}_r)\} \quad \text{where} \quad \boldsymbol{\gamma}_{ir} = (\gamma_{1ir}, \gamma_{2ir}). \tag{6.34}$$

Again considering the finite values of $\mathcal{D}_{2ir}(\mathbf{x}_i - \mathbf{x}_r)$ and using the definition of a polar function, we have

$$\begin{aligned} \mathcal{D}^c_{2ir}(\boldsymbol{\gamma}_{ir}) &= \sup_{\mathbf{x}_i} \{\langle \boldsymbol{\gamma}_{ir}, (\mathbf{x}_i - \mathbf{x}_r) \rangle \ : \ \ell_p(\mathbf{x}_i, \mathbf{x}_r) \leq d_{2ir}\} \\ &= d_{2ir}\, \ell^o_p(\boldsymbol{\gamma}_{ir}). \end{aligned} \tag{6.35}$$

In order to devise the dual formulation we use the following properties of conjugate functions (Witzgall, 1964, Section 14).

Property 6.1.2 *Let $h_i(x)$, $i = 1, \ldots, n$, be closed convex functions. $h_i(x)$ is closed if $h_i^{cc} = h_i$. Define $H(x) = \sum_{i=1}^{n} h_i(x)$. Then we have*

$$H^c(0) = \min \left\{ \sum_{i=1}^{n} h_i^c(y_i) \; : \; \sum_{i=1}^{n} y_i = 0 \right\}.$$

Property 6.1.3 *Let $H(x)$ be a closed convex function. Then*

$$H^c(0) = -\min \; H(x).$$

It can easily be verified that each component of $PM(\mathbf{X})$ is a closed convex function. Using the equations (6.29), (6.30), (6.31), (6.33), (6.35) and Property 6.1.2 we obtain the conjugate of $CPM(\mathbf{X})$ in (6.28) evaluated at zero as

$$\begin{aligned} CPM^c(0) = \min \Bigg\{ & \sum_{i=1}^{m} \sum_{j=1}^{n} \langle \mathbf{a}_j, \mathcal{U}_{ij} \rangle + \sum_{i=1}^{m} \sum_{j=1}^{n} \mathcal{D}_{1ij}^c(\beta_{ij}) + \sum_{i=1}^{m-1} \sum_{r=i+1}^{m} \mathcal{D}_{2ir}^c(\gamma_{ir}) \\ & + \sum_{i=1}^{m} \sum_{k=1}^{K} \lambda_{ik} r_{ik} \; : \quad \sum_{j=1}^{n} \mathcal{U}_{ij} + \sum_{r=i+1}^{m} \mathcal{V}_{ir} - \sum_{r=1}^{i-1} \mathcal{V}_{ri} + \sum_{j=1}^{n} \beta_{ij} + \sum_{r=i+1}^{m} \gamma_{ir} \\ & - \sum_{r=1}^{i-1} \gamma_{ri} + \sum_{k=1}^{K} \mathbf{z}_{ik} = 0, \;\; i = 1, \ldots, m; \;\; \ell_p^c(\mathcal{U}_{ij}) \leq w_{1ij}, \;\; i = 1, \ldots, m, \\ & j = 1, \ldots, n; \;\; \ell_p^c(\mathcal{V}_{ir}) \leq w_{2ir}, \;\; i = 1, \ldots, m-1, \; r = i+1, \ldots, m; \\ & \lambda_{ik} \geq 0, \;\; \lambda_{ik} \mathbf{g}_{ik} = \mathbf{z}_{ik}, \;\; i = 1, \ldots, m, \;\; k = 1, \ldots, K. \Bigg\} \end{aligned}$$

Let $\mathbf{U} = (\mathcal{U}_{11}, \ldots, \mathcal{U}_{mn}, \mathcal{V}_{12}, \ldots, \mathcal{V}_{(m-1)m})$ where $\mathcal{U}_{ij} = (u_{1ij}, u_{2ij})$ for $i = 1, \ldots, m$, $j = 1, \ldots, n$ and $\mathcal{V}_{ir} = (v_{1ir}, v_{2ir})$ for $i = 1, \ldots, m-1$, $r = i+1, \ldots, m$, $\mathbf{B} = (\beta_{11}, \ldots, \beta_{mn}, \gamma_{12}, \ldots, \gamma_{(m-1)m})$ where $\beta_{ij} = (\beta_{1ij}, \beta_{2ij})$ for $i = 1, \ldots, m$, $j =$

$1,\ldots,n$, and $\gamma_{ir} = (\gamma_{1ir}, \gamma_{2ir})$ for $i = 1,\ldots,m-1,\ r = i+1,\ldots,m$. Then, applying Property 6.1.3, expanding and rearranging we obtain the dual formulation of (6.2) as

$$\max \quad DCPM(\lambda, \mathrm{U}, \mathrm{B}) = -\sum_{i=1}^{m}\sum_{j=1}^{n}(a_{j1}\,u_{1ij} + a_{j2}\,u_{2ij}) - \sum_{i=1}^{m}\sum_{j=1}^{n}(a_{j1}\,\beta_{1ij} + a_{j2}\,\beta_{2ij})$$

$$- \sum_{i=1}^{m}\sum_{j=1}^{n} d_{1ij}\,\ell_p^{\circ}(\beta_{ij}) - \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} d_{2ir}\,\ell_p^{\circ}(\gamma_{ir}) - \sum_{i=1}^{m}\sum_{k=1}^{K}\lambda_{ik}\,r_{ik} \tag{6.36}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} u_{1ij} + \sum_{r=i+1}^{m} v_{1ir} - \sum_{r=1}^{i-1} v_{1ri} + \sum_{j=1}^{n}\beta_{1ij}$$

$$+ \sum_{r=i+1}^{m}\gamma_{1ir} - \sum_{r=1}^{i-1}\gamma_{1ri} + \sum_{k=1}^{K}\lambda_{ik}\,g_{ik1} = 0,\ \ i = 1,\ldots,m,$$

$$\sum_{j=1}^{n} u_{2ij} + \sum_{r=i+1}^{m} v_{2ir} - \sum_{r=1}^{i-1} v_{2ri} + \sum_{j=1}^{n}\beta_{2ij}$$

$$+ \sum_{r=i+1}^{m}\gamma_{2ir} - \sum_{r=1}^{i-1}\gamma_{2ri} + \sum_{k=1}^{K}\lambda_{ik}\,g_{ik2} = 0,\ \ i = 1,\ldots,m,$$

$$\ell_p^{\circ}(u_{1ij}, u_{2ij}) \le w_{1ij},\ \ i = 1,\ldots,m,\ j = 1,\ldots,n,$$

$$\ell_p^{\circ}(v_{1ir}, v_{2ir}) \le w_{2ir},\ \ i = 1,\ldots,m-1,\ r = i+1,\ldots,m,$$

$$\lambda_{ik} \ge 0,\ \ i = 1,\ldots,m,\ \ k = 1,\ldots,K.$$

The polar of $\ell_p(\mathrm{x})$, $\ell_p^{\circ}(\mathrm{x})$, is given by $\ell_q(\mathrm{y})$ where $(1/p) + (1/q) = 1$ and $p > 1$. Incorporating this polar function in (6.36), we have

$$\max \quad DCPM(\lambda, \mathrm{U}, \mathrm{B}) \ = \ -\sum_{i=1}^{m}\sum_{j=1}^{n}(a_{j1}\,u_{1ij} + a_{j2}\,u_{2ij})$$

$$- \sum_{i=1}^{m} \sum_{j=1}^{n} (a_{j1}\,\beta_{1ij} + a_{j2}\,\beta_{2ij}) - \sum_{i=1}^{m} \sum_{j=1}^{n} d_{1ij}\,(|\beta_{1ij}|^q + |\beta_{2ij}|^q)^{1/q}$$

$$- \sum_{i=1}^{m-1} \sum_{r=i+1}^{m} d_{2ir}\,(|\gamma_{1ir}|^q + |\gamma_{2ir}|^q)^{1/q} - \sum_{i=1}^{m} \sum_{k=1}^{K} \lambda_{ik}\,r_{ik} \qquad (6.37)$$

s.t.
$$\sum_{j=1}^{n} u_{1ij} + \sum_{r=i+1}^{m} v_{1ir} - \sum_{r=1}^{i-1} v_{1ri} + \sum_{j=1}^{n} \beta_{1ij}$$

$$+ \sum_{r=i+1}^{m} \gamma_{1ir} - \sum_{r=1}^{i-1} \gamma_{1ri} + \sum_{k=1}^{K} \lambda_{ik}\,g_{ik1} = 0, \quad i = 1, \ldots, m,$$

$$\sum_{j=1}^{n} u_{2ij} + \sum_{r=i+1}^{m} v_{2ir} - \sum_{r=1}^{i-1} v_{2ri} + \sum_{j=1}^{n} \beta_{2ij}$$

$$+ \sum_{r=i+1}^{m} \gamma_{2ir} - \sum_{r=1}^{i-1} \gamma_{2ri} + \sum_{k=1}^{K} \lambda_{ik}\,g_{ik2} = 0, \quad i = 1, \ldots, m,$$

$$(|u_{1ij}|^q + |u_{2ij}|^q)^{1/q} \le w_{1ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n,$$

$$(|v_{1ir}|^q + |v_{2ir}|^q)^{1/q} \le w_{2ir}, \quad i = 1, \ldots, m-1, \; r = i+1, \ldots, m,$$

$$\lambda_{ik} \ge 0, \quad i = 1, \ldots, m, \quad k = 1, \ldots, K.$$

6.2 Comparison of the Dual Formulations

Comparing the dual formulations (6.24) and (6.37), we immediately observe that the conjugate dual problem (6.37) involves a nonlinear objective function. The nonlinearity is introduced by the polar functions of $\ell_p(\beta_{ij})$ and $\ell_p(\gamma_{ij})$. Since the polar function of an ℓ_p-norm is also a norm, the problem with nondifferentiability still exists in the conjugate dual formulation. In this section, we examine the

relationship between the dual variables in (6.24) and (6.37), and eventually show that the conjugate dual problem can be converted to the lagrangian dual problem, which has a linear objective function and smaller number of variables. For that purpose, we return to the development of the lagrangian dual problem, and employ a lagrangian function arranged slightly different as given in (6.38). In the quasilinearization of this lagrangian function, we define the dual variables in such a way that they eventually resemble the dual variables defined in developing the conjugate dual problem.

Rewriting the lagrangian function (6.4) we have

$$\begin{aligned} LPM(\mathbf{X}, \mu_1, \mu_2, \lambda) &= \sum_{i=1}^{m}\sum_{j=1}^{n} w_{1ij}\, \ell_p(\mathbf{x}_i, \mathbf{a}_j) + \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} w_{2ir}\, \ell_p(\mathbf{x}_i, \mathbf{x}_r) \\ &+ \sum_{i=1}^{m}\sum_{j=1}^{n} \mu_{1ij}\, \ell_p(\mathbf{x}_i, \mathbf{a}_j) + \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} \mu_{2ir}\, \ell_p(\mathbf{x}_i, \mathbf{x}_r) \\ &- \sum_{i=1}^{m}\sum_{j=1}^{n} \mu_{1ij}\, d_{1ij} - \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} \mu_{2ir}\, d_{2ir} \\ &+ \sum_{i=1}^{m}\sum_{k=1}^{K} \lambda_{ik}\, (G_{ik}(\mathbf{x}_i) - r_{ik}). \end{aligned} \tag{6.38}$$

Let $PM_{1ij}(\mathbf{x}_i, \mathbf{a}_j) = w_{1ij}\, \ell_p(\mathbf{x}_i, \mathbf{a}_j)$, $D_{1ij}(\mathbf{x}_i, \mathbf{a}_j) = \mu_{1ij}\, \ell_p(\mathbf{x}_i, \mathbf{a}_j)$ for $i = 1, \ldots, m$, $j = 1, \ldots, n$, and $PM_{2ir}(\mathbf{x}_i, \mathbf{x}_r) = w_{2ir}\, \ell_p(\mathbf{x}_i, \mathbf{x}_r)$, $D_{2ir}(\mathbf{x}_i, \mathbf{x}_r) = \mu_{2ir}\, \ell_p(\mathbf{x}_i, \mathbf{x}_r)$, for $i = 1, \ldots, m-1$, $r = i+1, \ldots, m$. We first quasilinearize the terms $PM_{1ij}(\mathbf{x}_i, \mathbf{a}_j)$ and $PM_{2ir}(\mathbf{x}_i, \mathbf{x}_r)$. Defining the vectors

$$\mathcal{X}_{1ij} = ((x_{i1} - a_{j1}), (x_{i2} - a_{j2})),$$

$$\mathcal{U}_{ij} = (u_{1ij}, u_{2ij}) \qquad \text{for} \quad i = 1, \ldots, m,\ j = 1, \ldots, n,$$

$$\mathcal{X}_{2ir} = ((x_{i1} - x_{r1}), (x_{i2} - x_{r2})),$$

$$\mathcal{V}_{ir} = (v_{1ir}, v_{2ir}) \qquad \text{for} \quad i = 1, \ldots, m-1,\ r = i+1, \ldots, m,$$

and using the Hölder inequality the quasilinearized forms are given by

$$PM_{1ij}(\mathbf{x}_i, \mathbf{a}_j) = \max_{\mathcal{U}_{ij}} \ Q_{1ij} = \langle \mathcal{X}_{1ij}, \mathcal{U}_{ij} \rangle \qquad (6.39)$$
$$\text{s.t.} \ \ell_q(\mathcal{U}_{ij}) = w_{1ij},$$

$$PM_{2ir}(\mathbf{x}_i, \mathbf{x}_r) = \max_{\mathcal{V}_{ir}} \ Q_{2ir} = \langle \mathcal{X}_{2ir}, \mathcal{V}_{ir} \rangle \qquad (6.40)$$
$$\text{s.t.} \ \ell_q(\mathcal{V}_{ir}) = w_{2ir},$$

where $1/p + 1/q = 1$ and $p > 1$.

Consider the problem (6.39), applying a similar analysis given in (6.8) – (6.20), we obtain

$$u^*_{1ij} = w_{1ij} \operatorname{sign}(x_{i1} - a_{j1}) \left(\frac{|x_{i1} - a_{j1}|}{\ell_p(\mathbf{x}_i, \mathbf{a}_j)} \right)^{p-1}, \qquad (6.41)$$

$$u^*_{2ij} = w_{1ij} \operatorname{sign}(x_{i2} - a_{j2}) \left(\frac{|x_{i2} - a_{j2}|}{\ell_p(\mathbf{x}_i, \mathbf{a}_j)} \right)^{p-1} \qquad (6.42)$$

and for the maximization problem (6.40), we have

$$v^*_{1ir} = w_{2ir} \operatorname{sign}(x_{i1} - x_{r1}) \left(\frac{|x_{i1} - x_{r1}|}{\ell_p(\mathbf{x}_i, \mathbf{x}_r)} \right)^{p-1}, \qquad (6.43)$$

$$v^*_{2ir} = w_{2ir} \operatorname{sign}(x_{i2} - x_{r2}) \left(\frac{|x_{i2} - x_{r2}|}{\ell_p(\mathbf{x}_i, \mathbf{x}_r)} \right)^{p-1}. \qquad (6.44)$$

Thus, analogous to Property 6.1.1, we have the following property.

Property 6.2.1 *Optimality for the new facility location* $\mathbf{x}_i$ *is achieved at a fixed facility location* $\mathbf{a}_j$ *if and only if the corresponding dual variables* u^*_{1ij} *and* u^*_{2ij} *are zero. Similarly, the locations of two new facilities* $\mathbf{x}_i$ *and* $\mathbf{x}_r$ *coincide at optimality if and only if the corresponding dual variables* v^*_{1ir} *and* v^*_{2ir} *are zero. If these coincidence conditions are not met, the equalities will hold in the constraints of (6.39) and (6.40).*

Using Property 6.2.1 we take the equality constraints in (6.39) and (6.40) as inequality ($\leq$) constraints .

We quasilinearize the nonlinear component $D_{1ij}(\mathbf{x}_i, \mathbf{a}_j)$ similar to $PM_{1ij}(\mathbf{x}_i, \mathbf{a}_j)$, and $D_{2ir}(\mathbf{x}_i, \mathbf{x}_r)$ similar to $PM_{2ir}(\mathbf{x}_i, \mathbf{x}_r)$. Defining the vectors $\beta_{ij} = (\beta_{1ij}, \beta_{2ij})$ for $i = 1, \ldots, m$, $j = 1, \ldots, n$, and $\gamma_{ir} = (\gamma_{1ir}, \gamma_{2ir})$ for $i = 1, \ldots, m-1$, $r = i+1, \ldots, m$, we have

$$D_{1ij}(\mathbf{x}_i, \mathbf{a}_j) = \max_{\beta_{ij}} \langle \mathcal{X}_{1ij}, \beta_{ij} \rangle \quad \text{s.t. } \ell_q(\beta_{ij}) = \mu_{1ij}, \tag{6.45}$$

and

$$D_{2ir}(\mathbf{x}_i, \mathbf{x}_r) = \max_{\gamma_{ir}} \langle \mathcal{X}_{2ir}, \gamma_{ir} \rangle \quad \text{s.t. } \ell_q(\gamma_{ir}) = \mu_{2ir}. \tag{6.46}$$

Again, following the steps similar to (6.8) – (6.20), it can easily be verified that

$$\beta^*_{1ij} = \mu_{1ij} (x_{i1} - a_{j1}) \left(\frac{|x_{i1} - a_{j1}|}{\ell_p(\mathbf{x}_i, \mathbf{a}_j)} \right)^{p-1}, \tag{6.47}$$

$$\beta^*_{2ij} = \mu_{1ij}\,(x_{i2} - a_{j2}) \left(\frac{|x_{i2} - a_{j2}|}{\ell_p(\mathbf{x}_i, \mathbf{a}_j)}\right)^{p-1}. \tag{6.48}$$

$$\gamma^*_{1ir} = \mu_{2ir}\,(x_{i1} - x_{r1}) \left(\frac{|x_{i1} - x_{r1}|}{\ell_p(\mathbf{x}_i, \mathbf{x}_r)}\right)^{p-1}, \tag{6.49}$$

$$\gamma^*_{2ir} = \mu_{2ir}\,(x_{i2} - x_{r2}) \left(\frac{|x_{i2} - x_{r2}|}{\ell_p(\mathbf{x}_i, \mathbf{x}_r)}\right)^{p-1}. \tag{6.50}$$

The following property states the relationship between $\mathcal{U}^*_{ij}$ and β^*_{ij}, and between $\mathcal{V}^*_{ir}$ and γ^*_{ir}. It immediately follows from comparing the relations (6.41) – (6.44) and (6.47) – (6.50), respectively.

Property 6.2.2 *The vectors β^*_{ij} and $\mathcal{U}^*_{ij}$, and γ^*_{ir} and $\mathcal{V}^*_{ir}$ are proportional in such a way that*

$$\beta^*_{ij} = \frac{\mu_{1ij}}{w_{1ij}}\mathcal{U}^*_{ij}, \qquad \gamma^*_{ir} = \frac{\mu_{2ir}}{w_{2ir}}\mathcal{V}^*_{ir}.$$

As a direct consequence of Property 6.2.2 and Property 6.2.1 we state the following property.

Property 6.2.3 *Optimality for the new facility location $\mathbf{x}_i$ is achieved at a fixed facility location $\mathbf{a}_j$ if and only if the corresponding dual variables β^*_{1ij} and β^*_{2ij} are zero. Similarly, two new facilities $\mathbf{x}_i$ and $\mathbf{x}_r$ coincide at optimality if and only if the corresponding dual variables γ^*_{1ir} and γ^*_{2ir} are zero. If these coincidence conditions are not met, the equalities will hold in the constraints of (6.45) and (6.46).*

Property 6.2.3 suggests that the equality constraints in problems (6.45) and (6.46) should be taken as inequalities ($\leq$).

The following property immediately follows from inspecting the expressions (6.41), (6.47) with (6.13), (6.42), (6.48) with (6.14), (6.43), (6.49) with (6.15), and (6.44), (6.50) with (6.16).

Property 6.2.4 $\theta^*_{ij} = \mathcal{U}^*_{ij} + \beta^*_{ij}$ *for* $i = 1, \ldots, m,\ j = 1, \ldots, n,$ *and* $\pi^*_{ir} = \mathcal{V}^*_{ir} + \gamma^*_{ir}$ *for* $i = 1, \ldots, m-1,\ r = i+1, \ldots, m.$

In order to convert the conjugate dual problem to the lagrangian dual problem we proceed as follows: Define μ_{1ij} and μ_{2ir} such that Property 6.2.3 holds and

$$\mu_{1ij} = (|\beta_{1ij}|^q + |\beta_{2ij}|^q)^{1/q} \quad \text{for } i = 1, \ldots, m,\ j = 1, \ldots, n, \tag{6.51}$$

$$\mu_{2ir} = (|\gamma_{1ir}|^q + |\gamma_{2ir}|^q)^{1/q} \quad \text{for } i = 1, \ldots, m-1,\ r = i+1, \ldots, m, \tag{6.52}$$

Also define $\theta_{ij} = (\theta_{1ij}, \theta_{2ij})$ and $\pi_{ir} = (\pi_{1ir}, \pi_{2ir})$ such that

$$\theta_{ij} = \mathcal{U}_{ij} + \beta_{ij} \quad \text{for } i = 1, \ldots, m,\ j = 1, \ldots, n, \tag{6.53}$$

$$\pi_{ir} = \mathcal{V}_{ir} + \gamma_{ir} \quad \text{for } i = 1, \ldots, m-1,\ r = i+1, \ldots, m, \tag{6.54}$$

as suggested by the Property 6.2.4. We assume that μ_{1ij}, μ_{2ir}, β_{ij}, γ_{ir}, θ_{ij}, $\mathcal{U}_{ij}$, π_{ir} and $\mathcal{V}_{ir}$ are as defined in the analysis regarding the lagrangian dual problem. Then, using (6.51), (6.52), (6.53) and (6.54), the conjugate dual problem (6.37) becomes

$$\begin{aligned} \max \ DCPM(\mu_1, \mu_2, \lambda, \Omega) &= -\sum_{i=1}^{m}\sum_{j=1}^{n}(a_{j1}\,\theta_{1ij} + a_{j2}\,\theta_{2ij}) \\ &\quad - \sum_{i=1}^{m}\sum_{j=1}^{n}\mu_{1ij}\,d_{1ij} - \sum_{i=1}^{m-1}\sum_{r=i+1}^{m}\mu_{2ir}\,d_{2ir} - \sum_{i=1}^{m}\sum_{k=1}^{K}\lambda_{ik}\,r_{ik} \end{aligned} \tag{6.55}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} \theta_{1ij} + \sum_{r=i+1}^{m} \pi_{1ir} - \sum_{r=1}^{i-1} \pi_{1ri} + \sum_{k=1}^{K} \lambda_{ik}\, g_{ik1} = 0, \;\; i = 1, \ldots, m,$$

$$\sum_{j=1}^{n} \mu_{2ij} + \sum_{r=i+1}^{m} \pi_{2ir} - \sum_{r=1}^{i-1} \pi_{2ri} + \sum_{k=1}^{K} \lambda_{ik}\, g_{ik2} = 0, \;\; i = 1, \ldots, m,$$

$$\mu_{1ij} \geq 0, \;\; i = 1, \ldots, m, \;\; j = 1, \ldots, n.$$

$$\mu_{2ir} \geq 0, \;\; i = 1, \ldots, m-1, \;\; r = i+1, \ldots, m.$$

$$\lambda_{ik} \geq 0, \;\; i = 1, \ldots, m, \;\; k = 1, \ldots, K.$$

$$(|u_{1ij}|^q + |u_{2ij}|^q)^{1/q} \leq w_{1ij}, \;\; i = 1, \ldots, m, \; j = 1, \ldots, n, \tag{6.56}$$

$$(|v_{1ir}|^q + |v_{2ir}|^q)^{1/q} \leq w_{2ir}, \;\; i = 1, \ldots, m-1, \; r = i+1, \ldots, m, \tag{6.57}$$

$$(|\beta_{1ij}|^q + |\beta_{2ij}|^q)^{1/q} \leq \mu_{1ij}, \;\; i = 1, \ldots, m, \; j = 1, \ldots, n, \tag{6.58}$$

$$(|\gamma_{1ir}|^q + |\gamma_{2ir}|^q)^{1/q} \leq \mu_{2ir}, \;\; i = 1, \ldots, m-1, \; r = i+1, \ldots, m, \tag{6.59}$$

Observe that (6.55) differs from the lagrangian dual problem (6.24) only in the last four constraints (6.56)–(6.59). We already know from Properties 6.2.1 and 6.2.3 that, in the optimal solution, these constraints either will be satisfied as equalities or the left hand sides will be equal to zero. Furthermore, because of the proportionality of dual variables (Property 6.2.2), we must have

$$(|u_{1ij}|^q + |u_{2ij}|^q)^{1/q} = 0 \;\; \Leftrightarrow \;\; (|\beta_{1ij}|^q + |\beta_{2ij}|^q)^{1/q} = 0,$$

$$(|u_{1ij}|^q + |u_{2ij}|^q)^{1/q} = w_{1ij} \;\; \Leftrightarrow \;\; (|\beta_{1ij}|^q + |\beta_{2ij}|^q)^{1/q} = \mu_{1ij},$$

$$(|v_{1ir}|^q + |v_{2ir}|^q)^{1/q} = 0 \;\; \Leftrightarrow \;\; (|\gamma_{1ir}|^q + |\gamma_{2ir}|^q)^{1/q} = 0,$$

$$(|v_{1ir}|^q + |v_{2ir}|^q)^{1/q} = w_{2ir} \quad \Leftrightarrow \quad (|\gamma_{1ir}|^q + |\gamma_{2ir}|^q)^{1/q} = \mu_{2ir}.$$

Thus, we can combine (6.56) and (6.58) as

$$(|u_{1ij}|^q + |u_{2ij}|^q)^{1/q} + (|\beta_{1ij}|^q + |\beta_{2ij}|^q)^{1/q} \leq w_{1ij} + \mu_{1ij}, \tag{6.60}$$

(6.57) and (6.59) as

$$(|v_{1ir}|^q + |v_{2ir}|^q)^{1/q} + (|\gamma_{1ir}|^q + |\gamma_{2ir}|^q)^{1/q} \leq w_{2ir} + \mu_{2ir}. \tag{6.61}$$

Using the Minkowski Inequality and the proportionality of the dual variables (Property 6.2.2), (6.60) and (6.61) become

$$(|u_{1ij} + \beta_{1ij}|^q + |u_{2ij} + \beta_{2ij}|^q)^{1/q} \leq w_{1ij} + \mu_{1ij}, \tag{6.62}$$

$$(|v_{1ir} + \gamma_{1ir}|^q + |v_{2ir} + \gamma_{2ir}|^q)^{1/q} \leq w_{2ir} + \mu_{2ir}, \tag{6.63}$$

respectively. By the definitions (6.53) and (6.54), we obtain

$$(|\theta_{1ij}|^q + |\theta_{2ij}|^q)^{1/q} \leq w_{1ij} + \mu_{1ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n, \tag{6.64}$$

$$(|\pi_{1ir}|^q + |\pi_{2ir}|^q)^{1/q} \leq w_{2ir} + \mu_{2ir}, \quad i = 1, \ldots, m-1, \; r = i+1, \ldots, m \tag{6.65}$$

Thus, the four constraints (6.56)–(6.59) are reduced to two constraints as (6.64) and (6.65) in which the Property 6.1.1 is applicable. Specifically, the constraints (6.64) and (6.65) either will be satisfied as equalities or the left hand sides will be equal to zero. Rewriting the dual formulation with the new constraints, (6.55) becomes

$$\max \; DCPM(\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \boldsymbol{\lambda}, \boldsymbol{\Omega}) \; = \; -\sum_{i=1}^{m} \sum_{j=1}^{n} (a_{j1}\, \theta_{1ij} + a_{j2}\, \theta_{2ij})$$

$$- \sum_{i=1}^{m} \sum_{j=1}^{n} \mu_{1ij}\, d_{1ij} - \sum_{i=1}^{m-1} \sum_{r=i+1}^{m} \mu_{2ir}\, d_{2ir} - \sum_{i=1}^{m} \sum_{k=1}^{K} \lambda_{ik}\, r_{ik} \quad (6.66)$$

s.t. $$\sum_{j=1}^{n} \theta_{1ij} + \sum_{r=i+1}^{m} \pi_{1ir} - \sum_{r=1}^{i-1} \pi_{1ri} + \sum_{k=1}^{K} \lambda_{ik}\, g_{ik1} = 0, \quad i = 1, \ldots, m,$$

$$\sum_{j=1}^{n} \mu_{2ij} + \sum_{r=i+1}^{m} \pi_{2ir} - \sum_{r=1}^{i-1} \pi_{2ri} + \sum_{k=1}^{K} \lambda_{ik}\, g_{ik2} = 0, \quad i = 1, \ldots, m,$$

$$(|\theta_{1ij}|^q + |\theta_{2ij}|^q)^{1/q} \leq w_{1ij} + \mu_{1ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n,$$

$$(|\pi_{1ir}|^q + |\pi_{2ir}|^q)^{1/q} \leq w_{2ir} + \mu_{2ir}, \quad i = 1, \ldots, m-1, \; r = i+1, \ldots, m,$$

$$\mu_{1ij} \geq 0, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n.$$

$$\mu_{2ir} \geq 0, \quad i = 1, \ldots, m-1, \quad r = i+1, \ldots, m.$$

$$\lambda_{ik} \geq 0, \quad i = 1, \ldots, m, \quad k = 1, \ldots, K.$$

Observe that (6.66) is equivalent to the formulation given in (6.24). Therefore, the lagrangian dual and the conjugate dual are essentially the same formulations. However, solving the lagrangian dual is preferable because of its linear objective function and less number of variables.

6.3 Dual of the ℓ_{bp}-norm Location Model

We develop the dual formulation for the following problem:

$$\min \quad M(\mathbf{X}) = \sum_{i=1}^{m} \sum_{j=1}^{n} w_{1ij}\, \ell_{bp}(\mathbf{x}_i, \mathbf{a}_j) + \sum_{i=1}^{m-1} \sum_{r=i+1}^{m} w_{2ir}\, \ell_{bp}(\mathbf{x}_i, \mathbf{x}_r) \quad (6.67)$$

$$\text{s.t.} \quad \ell_{bp}(\mathbf{x}_i, \mathbf{a}_j) \leq d_{1ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n,$$

$$\ell_{bp}(\mathbf{x}_i, \mathbf{x}_r) \leq d_{2ir}, \quad i = 1, \ldots, m-1, \quad r = i+1, \ldots, m,$$

$$G_{ik}(\mathbf{x}_i) \leq r_{ik}, \quad i = 1, \ldots, m, \quad k = 1, \ldots, K$$

where all the variables are as defined for the problem (6.2). The dual formulations of the constrained multi-facility location problem with the ℓ_{bp}-norm can be obtained from the dual formulations (6.24) and (6.36).

By Property 1.2.1, the existing facility locations should be changed as

$$(a_{j1}, a_{j2}) \mapsto (b_1^{1/p} a_{j1}, b_2^{1/p} a_{j2}). \tag{6.68}$$

A linear constraint $g_{ik1}x_{i1} + g_{ik2}x_{i2} \leq r_{ik}$ in an ℓ_{bp}-norm problem can be transformed to a linear constraint in the equivalent ℓ_p-norm problem as $g'_{ik1}\,x'_{i1} + g'_{ik2}\,x'_{i2} \leq r_{ik}$ where $g'_{ik1} = b_1^{-1/p} g_{ik1}$, $g'_{ik2} = b_2^{-1/p} g_{ik2}$, and x'_{i1}, x'_{i2} are the transformed coordinates as given in Property 1.2.1. This transformation readily follows from rewriting the linear constraint as $b_1^{-1/p} g_{ik1}\,(b_1^{1/p} x_{i1}) + b_2^{-1/p} g_{ik2}\,(b_2^{1/p} x_{i2}) \leq r_{ik}$. Thus, in the dual formulations we take

$$(g_{ik1}, g_{ik2}) \mapsto (b_1^{-1/p} g_{ik1}, b_2^{-1/p} g_{ik2}). \tag{6.69}$$

To define the transformations of dual variables, θ_{ij}, π_{ir}, $\mathcal{U}_{ij}$, $\mathcal{V}_{ir}$, β_{ij} and γ_{ir}, we first observe that, for $t = 1, 2$,

$$\theta_{tij} = \frac{\partial PMD_{1ij}(\mathbf{x}_i, \mathbf{a}_j)}{\partial x_{it}}, \tag{6.70}$$

$$\pi_{tir} = \frac{\partial PMD_{2ir}(\mathbf{x}_i, \mathbf{x}_r)}{\partial x_{it}}, \tag{6.71}$$

$$u_{tij} = \frac{\partial PM_{1ij}(\mathbf{x}_i, \mathbf{a}_j)}{\partial x_{it}}, \tag{6.72}$$

$$v_{tir} = \frac{\partial PM_{2ir}(\mathbf{x}_i, \mathbf{x}_r)}{\partial x_{it}}, \tag{6.73}$$

$$\beta_{tij} = \frac{\partial D_{1ij}(\mathbf{x}_i, \mathbf{a}_j)}{\partial x_{it}}, \tag{6.74}$$

$$\gamma_{tir} = \frac{\partial D_{2ir}(\mathbf{x}_i, \mathbf{x}_r)}{\partial x_{it}}. \tag{6.75}$$

Let $f(\mathbf{z}) = W\ell_{bp}(\mathbf{z})$, $\mathbf{z} = (z_1, z_2)$, and $f'(\mathbf{z}') = W\ell_p(\mathbf{z}')$, $\mathbf{z}' = (b_1^{1/p} z_1, b_2^{1/p} z_2)$, i.e., $f'(\mathbf{z}')$ is obtained from $f(\mathbf{z})$ by using the corresponding scale factors in the horizontal and vertical directions as suggested by Property 1.2.1. Then, for $t = 1, 2$, the first order derivatives of $f(\mathbf{z})$ and $f'(\mathbf{z}')$ are given by

$$\frac{\partial f(\mathbf{z})}{\partial z_t} = b_t W \operatorname{sign}(z_t) \left(\frac{|z_t|}{\ell_{bp}(\mathbf{z})}\right)^{p-1}, \tag{6.76}$$

$$\frac{\partial f'(\mathbf{z}')}{\partial z'_t} = W \operatorname{sign}(z'_t) \left(\frac{|z'_t|}{\ell_p(\mathbf{z}')}\right)^{p-1}. \tag{6.77}$$

Rewriting (6.77) by substituting $\mathbf{z}' = (b_1^{1/p} z_1, b_2^{1/p} z_2)$ and rearranging, we have

$$\frac{\partial f'(\mathbf{z}')}{\partial z'_t} = W \operatorname{sign}(z_t)\, b_t^{(p-1)/p} \left(\frac{|z_t|}{\ell_{bp}(\mathbf{z})}\right)^{p-1}. \tag{6.78}$$

Comparing (6.76) and (6.78), it immediately follows that

$$\frac{\partial f'(\mathbf{z}')}{\partial z'_t} = b_t^{-1/p} \frac{\partial f(\mathbf{z})}{\partial z_t}. \tag{6.79}$$

Observe that, when the ℓ_{bp}-norm is considered, the functions $PMD_{1ij}(\mathbf{x}_i, \mathbf{a}_j)$, $PMD_{2ir}(\mathbf{x}_i, \mathbf{x}_r)$, $PM_{1ij}(\mathbf{x}_i, \mathbf{a}_j)$, $PM_{2ir}(\mathbf{x}_i, \mathbf{x}_r)$, $D_{1ij}(\mathbf{x}_i, \mathbf{a}_j)$ and $D_{2ir}(\mathbf{x}_i, \mathbf{x}_r)$ are in the

general form of $f(\mathbf{z}) = W\ell_{bp}(\mathbf{z})$. Thus, using the relations (6.70)–(6.75) we have the following transformations for the dual variables:

$$(\theta_{1ij}, \theta_{2ij}) \mapsto (b_1^{-1/p}\theta_{1ij}, b_2^{-1/p}\theta_{2ij}) \tag{6.80}$$

$$(\pi_{1ir}, \pi_{2ir}) \mapsto (b_1^{-1/p}\pi_{1ir}, b_2^{-1/p}\pi_{2ir}) \tag{6.81}$$

$$(\pi_{1ri}, \pi_{2ri}) \mapsto (b_1^{-1/p}\pi_{1ri}, b_2^{-1/p}\pi_{2ri}) \tag{6.82}$$

$$(u_{1ij}, u_{2ij}) \mapsto (b_1^{-1/p}u_{1ij}, b_2^{-1/p}u_{2ij}) \tag{6.83}$$

$$(v_{1ir}, v_{2ir}) \mapsto (b_1^{-1/p}v_{1ir}, b_2^{-1/p}v_{2ir}) \tag{6.84}$$

$$(v_{1ri}, v_{2ri}) \mapsto (b_1^{-1/p}v_{1ri}, b_2^{-1/p}v_{2ri}) \tag{6.85}$$

$$(\beta_{1ij}, \beta_{2ij}) \mapsto (b_1^{-1/p}\beta_{1ij}, b_2^{-1/p}\beta_{2ij}) \tag{6.86}$$

$$(\gamma_{1ir}, \gamma_{2ir}) \mapsto (b_1^{-1/p}\gamma_{1ir}, b_2^{-1/p}\gamma_{2ir}) \tag{6.87}$$

$$(\gamma_{1ri}, \gamma_{2ri}) \mapsto (b_1^{-1/p}\gamma_{1ri}, b_2^{-1/p}\gamma_{2ri}) \tag{6.88}$$

This method of obtaining the dual formulation for the ℓ_{bp}-norm problem from the ℓ_p-norm problem is similar to the second approach described in the minisum location models context on page 102. Note that, the first approach described there is not applicable for the dual formulation of the *constrained* location models. This is mainly because of the existence of distance constraints. The bound values (d_{1ij} and d_{2ir}) stated for the ℓ_{bp} distance problem needs to be transformed to the ℓ_p distance setting of the problem. The transformation can successfully be done only if the stretch factors b_1 and b_2 are equal. Since, in general, the b_1 and b_2 values are not equal for

an ℓ_{bp}-norm, we have to use a dedicated dual formulation of (6.67).

6.3.1 Lagrangian Dual

Using (6.68), (6.69), (6.80), (6.81) and (6.82) in (6.24), simplifying and rearranging, we obtain the lagrangian dual problem of (6.67) as follows:

$$\begin{aligned}\max \quad DQM(\mu_1,\mu_2,\lambda,\Omega) \;=\; & -\sum_{i=1}^{m}\sum_{j=1}^{n}(a_{j1}\,\theta_{1ij}+a_{j2}\,\theta_{2ij})-\sum_{i=1}^{m}\sum_{k=1}^{K}\lambda_{ik}\,r_{ik} \\ & -\sum_{i=1}^{m}\sum_{j=1}^{n}d_{1ij}\,\mu_{1ij}-\sum_{i=1}^{m-1}\sum_{r=i+1}^{m}d_{2ir}\,\mu_{2ir} \qquad (6.89)\end{aligned}$$

$$\text{s.t.} \quad \sum_{j=1}^{n}\theta_{1ij}+\sum_{r=i+1}^{m}\pi_{1ir}-\sum_{r=1}^{i-1}\pi_{1ri}+\sum_{k=1}^{K}\lambda_{ik}\,g_{ik1}=0,\; i=1,\ldots,m,$$

$$\sum_{j=1}^{n}\theta_{2ij}+\sum_{r=i+1}^{m}\pi_{2ir}-\sum_{r=1}^{i-1}\pi_{2ri}+\sum_{k=1}^{K}\lambda_{ik}\,g_{ik2}=0,\; i=1,\ldots,m,$$

$$(b_1^{1-q}|\theta_{1ij}|^q+b_2^{1-q}|\theta_{2ij}|^q)^{1/q}\le w_{1ij}+\mu_{1ij},\;\; i=1,\ldots,m,\; j=1,\ldots,n,$$

$$(b_1^{1-q}|\pi_{1ir}|^q+b_2^{1-q}|\pi_{2ir}|^q)^{1/q}\le w_{2ir}+\mu_{2ir},\;\; i=1,\ldots,m-1,\; r=i+1,\ldots,m,$$

$$\lambda_{ik}\ge 0,\;\; i=1,\ldots,m,\;\; k=1,\ldots,K,$$

$$\mu_{1ij}\ge 0,\;\; i=1,\ldots,m,\;\; j=1,\ldots,n,$$

$$\mu_{2ir}\ge 0,\;\; i=1,\ldots,m-1,\;\; r=i+1,\ldots,m.$$

6.3.2 Conjugate Dual

Making the substitutions (6.68), (6.69) and (6.83) – (6.88) in the ℓ_p-norm conjugate dual formulation (6.37), simplifying and rearranging, we obtain the conjugate dual

for the ℓ_{bp}-norm problem (6.67) as

$$\begin{aligned}
\max \quad DCM(\boldsymbol{\lambda}, \mathrm{U}, \mathrm{B}) \;=\; & -\sum_{i=1}^{m}\sum_{j=1}^{n}(a_{j1}\,u_{1ij} + a_{j2}\,u_{2ij}) \\
& - \sum_{i=1}^{m}\sum_{j=1}^{n}(a_{j1}\,\beta_{1ij} + a_{j2}\,\beta_{2ij}) - \sum_{i=1}^{m}\sum_{j=1}^{n} d_{1ij}\,(b_1^{1-q}|\beta_{1ij}|^q + b_2^{1-q}|\beta_{2ij}|^q)^{1/q} \\
& - \sum_{i=1}^{m-1}\sum_{r=i+1}^{m} d_{2ir}\,(b_1^{1-q}|\gamma_{1ir}|^q + b_2^{1-q}|\gamma_{2ir}|^q)^{1/q} - \sum_{i=1}^{m}\sum_{k=1}^{K}\lambda_{ik}\,r_{ik} \qquad (6.90)
\end{aligned}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} u_{1ij} + \sum_{r=i+1}^{m} v_{1ir} - \sum_{r=1}^{i-1} v_{1ri} + \sum_{j=1}^{n} \beta_{1ij} + \sum_{r=i+1}^{m} \gamma_{1ir} - \sum_{r=1}^{i-1} \gamma_{1ri} + \sum_{k=1}^{K} \lambda_{ik}\, g_{ik1} = 0, \quad i = 1, \ldots, m,$$

$$\sum_{j=1}^{n} u_{2ij} + \sum_{r=i+1}^{m} v_{2ir} - \sum_{r=1}^{i-1} v_{2ri} + \sum_{j=1}^{n} \beta_{2ij} + \sum_{r=i+1}^{m} \gamma_{2ir} - \sum_{r=1}^{i-1} \gamma_{2ri} + \sum_{k=1}^{K} \lambda_{ik}\, g_{ik2} = 0, \quad i = 1, \ldots, m,$$

$$(b_1^{1-q}\,|u_{1ij}|^q + b_2^{1-q}\,|u_{2ij}|^q)^{1/q} \le w_{1ij}, \quad i = 1, \ldots, m, \; j = 1, \ldots, n,$$

$$(b_1^{1-q}\,|v_{1ir}|^q + b_2^{1-q}\,|v_{2ir}|^q)^{1/q} \le w_{2ir}, \quad i = 1, \ldots, m-1, \; r = i+1, \ldots, m,$$

$$\lambda_{ik} \ge 0, \quad i = 1, \ldots, m, \quad k = 1, \ldots, K.$$

Inspecting, for example, the second to last constraint in the dual formulations (6.90), (6.36) and (6.37) suggests that the polar function of $\ell_{bp}(\mathbf{x}) = (b_1|x_1|^p + b_2|x_2|^p)^{1/p}$ is given by $\ell^{\circ}_{bp}(\mathbf{y}) = (b_1^{1-q}|y_1|^q + b_2^{1-q}|y_2|^q)^{1/q}$ where $(1/p) + (1/q) = 1$ and $p \ge 1$. Next,

we find the polar of the ℓ_{bp}-norm using conjugate function theory to see the validity of this observation.

The Polar of the ℓ_{bp}-norm

A discussion on the polars of convex functions can be found in (Rockafellar, 1970, Section 15). The polar of a norm $h(\mathbf{x})$ on $\Re^m$, denoted by $h^\circ(\mathbf{y})$, is defined as follows:

$$\begin{aligned} h^\circ(\mathbf{y}) &= \max\left\{\frac{\langle \mathbf{x}, \mathbf{y}\rangle}{h(\mathbf{x})} : \mathbf{x} \neq 0\right\} \\ &= \max\{\langle \mathbf{x}, \mathbf{y}\rangle : h(\mathbf{x}) \leq 1\}, \end{aligned} \tag{6.91}$$

where $\langle \mathbf{x}, \mathbf{y}\rangle$ is the inner product of m-vectors $\mathbf{x}$ and $\mathbf{y}$. In order to determine the polar of $\ell_{bp}(\mathbf{x})$ we use conjugate function theory (Rockafellar, 1970, Section 12). The conjugate of a convex function $f(\mathbf{x})$ on $\Re^m$, denoted by $f^c(\mathbf{y})$, is defined by

$$f^c(\mathbf{y}) = \sup_{\mathbf{x}}\{\langle \mathbf{x}, \mathbf{y}\rangle - f(\mathbf{x})\} \tag{6.92}$$

Note that since the extrema are actually attained we will use max instead of sup. Let $\mathbf{Z}^\alpha$ denote an $m \times n$ matrix with components $|Z_{ij}|^\alpha$ for $i = 1, \ldots, m$, $j = 1, \ldots, n$ where Z_{ij} for $i = 1, \ldots, m$, $j = 1, \ldots, n$ are the components of $\mathbf{Z}$ and $\alpha \in \Re$. Then, for any p, $p > 1$, we define the function

$$f(\mathbf{x}) = \frac{1}{p}\,\langle \mathbf{x}^{p/2}, \mathbf{B}\,\mathbf{x}^{p/2}\rangle \tag{6.93}$$

where $\mathbf{x} = (x_1, x_2)$ and $\mathbf{B}$ is a 2×2 diagonal matrix with positive entries $B_{11} = b_1$ and $B_{22} = b_2$. It can be shown that, for $p > 1$, $f(\mathbf{x})$ is a convex function. Then the

conjugate function of $f(\mathbf{x})$ is given by

$$f^c(\mathbf{y}) = \max_{\mathbf{x}} \left\{ \langle \mathbf{x}, \mathbf{y} \rangle \; - \; \frac{1}{p} \langle \mathbf{x}^{p/2}, \mathbf{B}\, \mathbf{x}^{p/2} \rangle \right\}. \tag{6.94}$$

Using the first-order necessary and sufficient conditions one can show that the maximum of $\{\langle \mathbf{x}, \mathbf{y} \rangle - (1/p)\, \langle \mathbf{x}^{p/2}, \mathbf{B}\; \mathbf{x}^{p/2} \rangle\}$ is attained at $\mathbf{x}^* = \mathbf{B}^{1/(1-p)}\, \mathbf{y}^{1/(p-1)}$. Substituting $\mathbf{x}^*$ into (6.94) we obtain the conjugate function of $f(\mathbf{x})$ as

$$\begin{aligned} f^c(\mathbf{y}) \;\; &= \;\; \frac{1}{q}\, \langle \mathbf{y}^{q/2}, \mathbf{B}^{1-q}\, \mathbf{y}^{q/2} \rangle \\ \text{where} \;\; &\frac{1}{p} + \frac{1}{q} = 1, \;\; p > 1. \end{aligned} \tag{6.95}$$

It is readily known that the function $\ell_{bp}(\mathbf{x})$ is a norm. Furthermore, the expression (6.93) can be rewritten as $f(\mathbf{x}) = (1/p)\, (\ell_{bp}(\mathbf{x}))^p$. Then, using Corollary 15.3.1 given by Rockafellar (1970, page 135), the polar of the function $(p\, f(\mathbf{x}))^{1/p}$ is given by $(q\, f^c(\mathbf{y}))^{1/q}$. Observing that $\ell_{bp}(\mathbf{x}) = (pf(\mathbf{x}))^{1/p}$, we find the polar of the ℓ_{bp}-norm, denoted by $\ell^o_{bp}(\mathbf{y})$, as follows:

$$\begin{aligned} \ell^o_{bp}(\mathbf{y}) \;\; &= \;\; (b_1^{1-q}\, |y_1|^q + b_2^{1-q} |y_2|^q)^{1/q} \\ \text{where} \;\; &\frac{1}{p} + \frac{1}{q} = 1, \;\; p > 1. \end{aligned} \tag{6.96}$$

This polar agrees with our observation above. Therefore, when the conjugate dual problem with the ℓ_p-norm is generalized to the conjugate dual with the ℓ_{bp}-norm, the polar function of the ℓ_p-norm that appears in (6.36) readily generalizes to the polar function of the ℓ_{bp}-norm.

6.4 Solving the Dual Problem - An Example

The dual formulations can be solved by using the decomposition method as described by Love and Kraemer (1973). On the other hand, standard nonlinear programming software can also be used for solving the dual problem as shown by Juel and Love (1981b). In this case, the lagrange multipliers associated with the equality constraints provide the optimal solution for the primal problem. Note that, the non-differentiability problem associated with the norm constraints is solved as suggested by Juel and Love (1981b). A simple transformation, which eliminates the power term in the norm expression, is applied to the norm constraints.

In what follows we give an example to illustrate the lagrangian dual formulation for an ℓ_{bp}-norm location problem. The general form of the dual problem is already given in (6.89). The parameters of the example distance function are taken as $b_1 = 1.2$, $b_2 = 1.5$ and $p = 1.8$.

We consider the problem of locating two new facilities ($m = 2$) with respect to four existing facilities ($n = 4$). Existing facility locations and the weights are given in Table 6.1. The weight between the new facilities is $w_{212} = 8$. The same allowable

j	a_{j1}	a_{j2}	w_{11j}	w_{12j}
1	0	0	8	18
2	7	24	15	8
3	20	28	14	7
4	15	2	9	16

Table 6.1: Existing Facility Locations and Weights

i	k	g_{ik1}	g_{ik2}	r_{ik}
1	1	-1.5	1.5	10
1	2	0.4	-1	10
2	1	-1.5	1.5	-8
2	2	0.4	-1	-8

Table 6.2: Coefficients in Linear Constraints

region is defined for both new facilities by the following two linear constraints ($K = 2$): $-1.5x_1 + 1.5x2 \leq 10$ and $0.4x_1 - x_2 \leq -8$. The corresponding coefficients that are used in the dual problem are summarized in Table 6.2. Additionally, we consider two existing facility related distance constraints, one for each new facility, and a distance constraint for the new facilities. These are $\ell_{bp}(\mathbf{x}_1, \mathbf{a}_3) \leq 10$, $\ell_{bp}(\mathbf{x}_2, \mathbf{a}_4) \leq 15$ and $\ell_{bp}(\mathbf{x}_1, \mathbf{x}_2) \leq 9$, i.e., we have $d_{113} = 10$, $d_{124} = 15$ and $d_{212} = 9$.

The lagrangian dual formulation is given in Display 6.1. We have solved the dual using the nonlinear programming SOLVER© (by Frontline Systems, Inc. and Optimal Methods, Inc.) included in the commercial spreadsheet program Microsoft Excel. The Solver uses the Generalized Reduced Gradient (GRG2) optimization code for the constrained nonlinear programming problems (Lasdon et al., 1978; Fylstra et al., 1998). The lagrange multipliers associated with the equality constraints (in the order that they appear in dual formulation) are 16.26, 14.79, 20.97, 13.97. Thus, the optimal solution is $\mathbf{x}_1^* = (16.26, 20.97)$, $\mathbf{x}_2^* = (14.79, 13.97)$ and the total cost is 1804.20. The same solution is also found by solving the primal problem with the same software package.

$$
\begin{aligned}
\max \quad & -7\theta_{112} - 24\theta_{212} - 20\theta_{113} - 28\theta_{213} - 15\theta_{114} - 2\theta_{214} \\
& -2\theta_{122} - 24\theta_{222} - 20\theta_{123} - 28\theta_{223} - 15\theta_{124} - 2\theta_{224} \\
& -10\lambda_{11} + 8\lambda_{12} - 10\lambda_{21} + 8\lambda_{22} \\
& -10\mu_{113} - 15\mu_{124} - 9\mu_{212}
\end{aligned}
$$

$$
\begin{aligned}
\text{s.t.} \quad & \theta_{111} + \theta_{112} + \theta_{113} + \theta_{114} + \pi_{112} - 1.5\lambda_{11} + 0.4\lambda_{12} = 0 \\
& \theta_{121} + \theta_{122} + \theta_{123} + \theta_{124} - \pi_{112} - 1.5\lambda_{21} + 0.4\lambda_{22} = 0 \\
& \theta_{211} + \theta_{212} + \theta_{213} + \theta_{214} + \pi_{212} + 1.5\lambda_{11} - \lambda_{12} = 0 \\
& \theta_{221} + \theta_{222} + \theta_{223} + \theta_{224} - \pi_{212} + 1.5\lambda_{21} - \lambda_{22} = 0 \\
& 1.2^{-1.25}|\theta_{111}|^{2.25} + 1.5^{-1.25}|\theta_{211}|^{2.25} \leq 8^{2.25} \\
& 1.2^{-1.25}|\theta_{112}|^{2.25} + 1.5^{-1.25}|\theta_{212}|^{2.25} \leq 15^{2.25} \\
& 1.2^{-1.25}|\theta_{113}|^{2.25} + 1.5^{-1.25}|\theta_{213}|^{2.25} \leq (14 + \mu_{113})^{2.25} \\
& 1.2^{-1.25}|\theta_{114}|^{2.25} + 1.5^{-1.25}|\theta_{214}|^{2.25} \leq 9^{2.25} \\
& 1.2^{-1.25}|\theta_{121}|^{2.25} + 1.5^{-1.25}|\theta_{221}|^{2.25} \leq 18^{2.25} \\
& 1.2^{-1.25}|\theta_{122}|^{2.25} + 1.5^{-1.25}|\theta_{222}|^{2.25} \leq 8^{2.25} \\
& 1.2^{-1.25}|\theta_{123}|^{2.25} + 1.5^{-1.25}|\theta_{223}|^{2.25} \leq 7^{2.25} \\
& 1.2^{-1.25}|\theta_{124}|^{2.25} + 1.5^{-1.25}|\theta_{224}|^{2.25} \leq (16 + \mu_{124})^{2.25} \\
& 1.2^{-1.25}|\pi_{112}|^{2.25} + 1.5^{-1.25}|\pi_{212}|^{2.25} \leq (8 + \mu_{212})^{2.25} \\
& \lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}, \mu_{113}, \mu_{124}, \mu_{212} \geq 0.
\end{aligned}
$$

Display 6.1: Example Lagrangian Dual Problem

Chapter 7

Conclusions and Future Research

This dissertation is concerned with the use of a weighted sum of order p as a distance predicting function, the calculation of the confidence intervals for the estimated distances and the single- and multi-facility minisum continuous location models.

A weighted sum of order p is a three parameter function. It is well-known that, when distances are modelled in a particular transportation network, the coordinate system should be rotated in such a way that it aligns with the underlying pattern of the road network. Considering this axis rotation also as a parameter, a distance modelling with the ℓ_{bp}-norm involves the determination of four parameters for a given region.

We showed that a weighted sum of order p is a norm, the ℓ_{bp}-norm, for values of p greater than or equal to 1, and derived some its useful properties with respect to its parameters.

The properties of the directional bias function and the unit balls for the ℓ_{bp}-norm

are of theoretical and practical interest. We examined these properties in detail and compared them with the properties of the ℓ_p-norm's directional bias function and the unit balls. We found that the ℓ_{bp}-norm is better at capturing the nonlinearity in a transportation network than the weighted ℓ_p-norm. This is due to having two stretch factors b_1 and b_2 with the ℓ_{bp}-norm where each factor accounts for the nonlinearity in an axis direction. We also observed that, in contrast to the weighted ℓ_p-norm, where the optimal parameter p value is confined to the interval $[1, 2]$, for the ℓ_{bp}-norm the parameter p can have an optimal value greater than 2. These results were also confirmed by empirical tests.

In order to model distances in a given transportation network a goodness-of-fit criterion is used. For that purpose, we employed a sum of squared deviations (SD) criterion, and examined its properties in the parameters b_1, b_2 and p. We found that the SD function is neither convex nor concave in p, but convex in b_1 and b_2, provided that $p \in [1, 2]$. Using these properties we presented an efficient computational procedure to determine the best values of the parameters b_1, b_2 and p for a given region. The algorithm also determines the best axis rotation for the region's transportation network. We applied the computational procedure to seventeen geographical regions. Comparing the estimation errors produced by the weighted ℓ_p-norm and the ℓ_{bp}-norm we observed that the inclusion of non-symmetric distance irregularities b_1 and b_2 into the model provides a better description of the

distances in a region. We have obtained improvements in the goodness-of-fit criterion value for all the seventeen geographical regions, except in New York State where there exists virtually no directional nonlinearity in its road network. As the level of directional nonlinearity decreases ($b_1 \approx b_2$) the ℓ_{bp}-norm behaves similar to the weighted ℓ_p-norm, indicating that the improvement in SD is relatively low. As the level of directional nonlinearity increases, the ℓ_{bp}-norm provides more pronounced improvements in distance estimations. Additionally, although it is shown that the SD criterion is not necessarily convex in b_1 and b_2 for values of p greater than 2, our empirical tests revealed that convexity is always obtained in a large neighbourhood of the best b_1 and b_2 values.

To develop the confidence intervals for estimated distances we first analyzed the statistical properties of estimation errors in the sample geographical regions. Based on the empirical results we devised a transformed error related random variable which is homoscedastic and, in most of the cases, normally distributed. Using this new random variable we developed the confidence interval calculation method for an unknown actual distance. The confidence intervals provide the means to compare the performance of a distance predicting function. Therefore, using the new method of confidence interval calculation we compared the weighted ℓ_p-norm and the ℓ_{bp}-norm. The ℓ_{bp}-norm provided tighter intervals in all the sample regions, except, again, in New York State. In general, the comparison of the two norms using the confidence

intervals agreed with the comparison using the SD criterion. The new method also indicated the importance of using homoscedastic distributions in calculating confidence intervals.

In the rest of the dissertation we focused on the minisum location models. We have obtained some new results regarding the minisum location models with the ℓ_p-norm. since a model should represent the real situation as accurately as possible, the accuracy of the distance function employed plays a crucial role in terms of the validity and the applicability of the locational decisions. Therefore, having obtained better distance representations with the ℓ_{bp}-norm, we incorporated it into the single- and multi-facility minisum continuous location models. We presented two approaches to solve the ℓ_{bp}-norm location problems. The first approach involved the use of ℓ_p-norm results with some modifications in the problem setting. The second approach required the use of special algorithms for the ℓ_{bp}-norm location models. For that purpose, we developed generalizations of the Weiszfeld procedure for these problems and provided the fixed-point optimality condition for the single facility case. The modified Weiszfeld procedure is basically an iterative steepest-descent algorithm with a pre-determined step size. Therefore, to terminate the iterative procedure a stopping rule or a bound for the best objective function value is required. We provided generalizations of a bounding method known as the rectangular bound in the continuous facility location literature. The rectangular bound at an iteration is obtained by solving a rectangular

distance location problem. The bound problem involves locating the same number of facilities in the original problem with respect to the existing facility locations with newly created weights. We first generalized the bounding method to the approximated ℓ_p-norm location problems and then to the approximated ℓ_{bp}-norm models. We have also shown that, at optimality, the rectangular bound problem has multiple solutions which include the solution to the original location problem. We have also studied convergence properties of the Weiszfeld procedure when it is used to solve the ℓ_p distance (or ℓ_{bp} distance) single-facility location models where $p > 2$. We have found that Weiszfeld's iterative scheme has converged with the introduction of the step size factor $2/p$ for $2 \leq p \leq 3$ and $2/(p-1)$ for $p > 3$. This result is especially useful for the ℓ_{bp}-norm location models because, as shown in Chapter 2, the best value of p for a geographical region can be greater than 2.

We have developed both lagrangian and conjugate dual formulations for the most general constrained multi-facility minisum problems with the ℓ_p-norm by considering both linear and distance constraints in the model. Linear constraints define a feasible region for the new facilities and distance constraints impose upper bounds on the distances between any pair of facilities. We have observed that the lagrangian dual formulation is preferable in applications because of its linear objective function. Nevertheless, a detailed comparison of the two formulations revealed that the conjugate dual can indeed be transformed to a simpler formulation identical to the

lagrangian dual. We next incorporated the new distance function ℓ_{bp}-norm into the same location model and provided generalizations of the dual formulation. We have also presented a numerical example to setup and solve a lagrangian dual formulation.

Our future research will focus on mainly two issues regarding the use of the ℓ_{bp}-norm in distance estimations and the continuous location models in general.

First, recall that the minimization of the SD criterion in parameters b_1, b_2 and p is in general a nonconvex optimization problem. We have tested a BFGS type Quasi-Newton method for this three variable problem and obtained the same results at each axis rotation in $[0°, 90°]$ and in all the sample regions, regardless of several different initial points. This presents some evidence on the possibility that the SD is actually a unimodal function of b_1, b_2 and p.

Secondly, we plan to study some further localization results using the relationship between the optimal solutions to the location problems and the rectangular bound problems (Property 5.2.1).

Bibliography

BELLMAN, R. 1965. An Application of Dynamic Programming to Location-Allocation Problems, *SIAM Review*, **7(1)**, 126–128

BERENS, W. 1988. The Suitability of the Weighted ℓ_p-norm in Estimating Actual Road Distances, *European Journal of Operational Research*, **34**, 39–43

BERENS, W. AND KÖRLING, F. J. 1985. Estimating Road Distances by Mathematical Functions, *European Journal of Operational Research*, **21**, 54–56

BERTSEKAS, D. M. 1995. *Nonlinear Programming.* Athena Scientific. Belmont, Massachusetts

BRIMBERG, J. 1989. Properties of Distance Functions and Minisum Location Problems. *Ph.D. thesis.* McMaster University. Hamilton, Canada

BRIMBERG, J., DOWLING, P. D., AND LOVE, R. F. 1994. The Weighted One-Two Norm Distance Model: Empirical Validation and Confidence Interval Estimation, *Location Science*, **2(2)**, 91–100

BRIMBERG, J., DOWLING, P. D., AND LOVE, R. F. 1996. Estimating the Parameters of the Weighted ℓ_p-norm by Linear Regression, *IIE Transactions*, **28**,

363–367

BRIMBERG, J. AND LOVE, R. F. 1991. Estimating Travel Distances by the Weighted ℓ_p-norm, *Naval Research Logistics*, **38**, 241–259

BRIMBERG, J. AND LOVE, R. F. 1992. Local Convergence in a Generalized Fermat-Weber Problem, *Annals of Operations Research*, **40**, 33–66

BRIMBERG, J. AND LOVE, R. F. 1993a. Directional Bias of the ℓ_p-norm, *European Journal of Operational Research*, **67**, 287–294

BRIMBERG, J. AND LOVE, R. F. 1993b. Global Convergence of a Generalized Iterative Procedure for the Minisum Location Problem with ℓ_p Distances, *Operations Research*, **41(6)**, 1153–1163

BRIMBERG, J. AND LOVE, R. F. 1995a. Generalized Hull Properties for Location Problems, *IIE Transactions*, **27**, 226–232

BRIMBERG, J. AND LOVE, R. F. 1995b. Properties of Ordinary and Weighted Sums of Order p Used for Distance Estimation, *Recherche Opérationnelle*, **29(1)**, 59–72

BRIMBERG, J., LOVE, R. F., AND WALKER, J. H. 1995. The Effect of Axis Rotation on Distance Estimation, *European Journal of Operational Research*, **80**, 357–364

CABOT, A. AND FRANCIS, R. L. 1972. Properties of a Multi-facility Location Problem Involving Euclidean Distances, *Naval Research Logistics Quarterly*, **19**, 335–353

CHEN, R. 1984a. Location Problems with Costs Being sums of Powers of Euclidean Distances, *Computers and Operations Research*, **11**, 285–294

CHEN, R. 1984b. Solution of Location problems with Radial Cost Functions, *Computers and Mathematics with Applications*, **10**, 87–94

COOPER, L. 1963. Location-Allocation problems, *Operations Research*, **11(3)**, 37–52

COOPER, L. 1968. An Extension to the Generalized Weber Problem, *Journal of Regional Science*, **8**, 181–197

DALLAL, G. E. AND WILKINSON, L. 1986. An Analytic Approximation to the Distribution of Lilliefors's Test Statistic for Normality, *The American Statistician*, **40(4)**, 294–296

DOWLING, P. D. AND LOVE, R. F. 1986. Bounding Methods for Facilities Location Algorithms, *Naval Research Logistics Quarterly*, **33**, 775–787

DOWLING, P. D. AND LOVE, R. F. 1987. An Evaluation of the Dual as a Lower Bound in Facilities Location Problems, *IIE Transactions*, **19(2)**, 160–166

DREZNER, Z. 1984. The Planar Two-center and Two-median Problems, *Transportation Science*, **18**, 351–361

DREZNER, Z. 1995. Replacing discrete demand with continuous demand. in Z. Drezner (ed.), *Facility Location, A Survey of Applications and Methods*. Springer-Verlag. New York

DREZNER, Z. AND WESOLOWSKY, G. O. 1995. Location on a One-way Rectilinear

Grid, *Journal of the Operational Research Society*, **46**, 735–746

EILON, S., WATSON-GANDY, C. D. T., AND CHRISTOFIDES, N. 1971. *Distribution Management: Mathematical Modeling and Practical Analysis.* Hafner Publishing Company. New York

EYSTER, J. W., A., W. J., AND WIERWILLE, W. W. 1973. On solving Multifacility Location Problems Using a Hyperboloid Approximation Procedure, *AIIE Transactions*, **5**, 1–6

FASBENDER, E. 1846. *Journal für die Reine und Angewandte Mathematik*, **30**, 230–231

FOULDS, L. R. AND HAMACHER, H. W. 1990. Optimal Bin Location and Sequencing in Printed Circuit Board Assembly. Technical Report 181. University of Kaiserslauteren

FRENK, J. B. G. AND KLEIJN, M. J. 1994. On the Miehle's Algorithm and the Perturbed ℓ_p-distance Multifacility Location Problem, *Studies in Locational Analysis*, **7**, 61–75

FYLSTRA, D., LASDON, L., WATSON, W., AND WAREN, A. 1998. Design and Use of the Microsoft Excel Solver, *Interfaces*, **28**, 29–55

GINSBURGH, V. AND HANSEN, P. 1974. Procedures for the Reduction of Errors in Road Network Data, *Operational Research Quarterly*, **25**, 321–322

HANSEN, P., PERREUR, J., AND THISSE, J. F. 1980. Location Theory, Dominance

and Convexity: Some Further Results, *Operations Research*, **28**, 1241–1250

HARDY, G. H., LITTLEWOOD, J. E., AND PÓLYA, G. 1952. *Inequalities*. Cambridge University Press. Cambridge

HARRIS, B. 1978. Speeding up Iterative Algorithms- the Generalized Weber Problem, *Journal of Regional Science*, **16**, 411–413

HODGSON, M., WONG, R., AND HONSAKER, J. 1987. The p-centroid Problem on an Inclined Plane, *Operations Research*, **35**, 221–233

HURIOT, J. M. AND PERREUR, J. 1973. On the Weber Problem with Rectangular Distance: A Comment, *Management Science*, **20**, 418–419

IDRISSI, H., LEFEBVRE, O., AND MICHELOT, C. 1989. Duality for Constrained Multifacility Location Problems with Mixed Norms and Applications, *Annals of Operations Research*, **18**, 71–92

JOHNSON, N. L., NIXON, E., AND AMOS, D. E. 1963. Tables of Percentage Points of Pearson Curves for Given $\sqrt{\beta_1}$ and β_2 Expressed in Standard Measure, *Biometrika*, **50**, 459–498

JUEL, H. 1983. Coincident Optima for Two-Facility Weber Problems, *Transportation Science*, **17**, 110–113

JUEL, H. 1984. On a Rational Stopping Rule for Facilities Location Algorithms, *Naval Research Logistics Quarterly*, **9**, 9–11

JUEL, H. AND LOVE, R. F. 1976. An Efficient Computational Procedure

for Solving Multi-facility Rectilinear Facilities Location Problem, *Operational Research Quarterly*, **27**, 697–703

JUEL, H. AND LOVE, R. F. 1980. Sufficient Conditions for Optimal Facility Locations To Coincide, *Transportation Science*, **14**, 125–129

JUEL, H. AND LOVE, R. F. 1981a. Fixed Point Optimality Criteria for the Location Problem with Arbitrary Norms, *Journal of the Operational Research Society*, **32**, 891–897

JUEL, H. AND LOVE, R. F. 1981b. On the Dual of the Linearly Constrained Multifacility Location Problem with Arbitrary Norms, *Transportation Science*, **15(4)**, 329–337

JUEL, H. AND LOVE, R. F. 1983. Hull properties in Location Problems, *European Journal of Operational Research*, **12**, 262–265

JUEL, H. AND LOVE, R. F. 1987. Duality in Constrained Location Problems, *Operations Research Letters*, **6(6)**, 281–284

KLEIN, R. 1988. in *Graph-Theoretic Concepts in Computer Science*. 434–441. Amsterdam

KOLESAR, P., WALKER, W., AND HAUSNER, J. 1975. Determining the Relation Between Fire Engine Travel Times and Travel Distances in New York City, *Operations Research*, **23**, 614–627

KOSHIZUKA, T. AND O. KURITA, O. 1991. Approximate formulas of average

distances associated with regions and their applications to location problems, *Mathematical Programming*, **52**, 99–123

KOTZ, S. AND JOHNSON, N. L. (eds.) 1989. *Encyclopedia of Statistical Sciences.* Vol. 4. John-Wiley and Sons. New York

KUHN, H. W. 1967. On a pair of dual nonlinear programs. in J. Abadie (ed.), *Nonlinear Programming.* John Wiley and Sons. New York

KUHN, H. W. 1973. A Note on Fermat's Problem, *Mathematical Programming*, **4**, 98–107

KUHN, H. W. AND KUENNE, R. E. 1962. An Efficient Algorithm for the Numerical Solution of the Generalized Weber Problem in Spatial Economics, *Journal of Regional Science*, **4(2)**, 21–34

LASDON, L., WAREN, A. D., JAIN, A., AND RATNER, M. 1978. Design and Testing of Generalized Reduced Gradient Code for Nonlinear Programming, *ACM Transactions on Mathematical Software*, **4**, 34–49

LILLIEFORS, H. W. 1967. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown, *American Statistical Association Journal*, 399–402

LOVE, R. F. 1969. Locating Facilities in Three-dimensional space by Convex Programming, *Naval Research Logistics Quarterly*, **16**, 503–516

LOVE, R. F. 1974. The Dual of a Hyperbolic Approximation to the Generalized Constrained Multi-Facility Location Problem with ℓ_p Distances, *Management*

Science, **21(1)**, 22–33

LOVE, R. F. AND DOWLING, P. D. 1985. Optimal Weighted ℓ_p-norm Parameters for Facilities Layout Distance Characteristics, *Management Science*, **31**, 200–206

LOVE, R. F. AND DOWLING, P. D. 1989a. A Generalized Bounding Method for Multi-facility Location Models, *Operations Research*, **37**, 653–657

LOVE, R. F. AND DOWLING, P. D. 1989b. A New Bounding Method for Single Facility Location Models, *Annals of Operations Research*, **18**, 103–112

LOVE, R. F. AND JUEL, H. 1982. Properties and Solution Methods for Large Location-Allocation problems, *Journal of the Operational Research Society*, **33(5)**, 443–452

LOVE, R. F. AND KRAEMER, S. A. 1973. A Dual Decomposition Method for Minimizing Transportation Costs in Multifacility Location Problems, *Transportation Science*, **7(4)**, 297–316

LOVE, R. F. AND MORRIS, J. G. 1972. Modelling Inter-City Road Distances by Mathematical Models, *Operational Research Quarterly*, **23**, 61–71

LOVE, R. F. AND MORRIS, J. G. 1975a. A Computation Procedure for the Exact Solution of Location-Allocation Problems with Rectangular Distances, *Naval Research Logistics Quarterly*, **22**, 441–453

LOVE, R. F. AND MORRIS, J. G. 1975b. Solving Constrained Multi-facility Location Problems Involving ℓ_p Distances Using Convex Programming, *Operations*

Research, **23(3)**, 581–587

LOVE, R. F. AND MORRIS, J. G. 1979. Mathematical Models of Road Travel Distances, *Management Science*, **25**, 130–139

LOVE, R. F. AND MORRIS, J. G. 1988. On Estimating Road Distances by Mathematical Functions, *European Journal of Operational Research*, **36**, 251–253

LOVE, R. F., MORRIS, J. G., AND WESOLOWSKY, G. O. 1988. *Facilities Location – Models and Methods.* North-Holland. NewYork

LOVE, R. F. AND ÜSTER, H. 1996. Comparison of the Properties and the Performance of the Criteria Used to Evaluate the Accuracy of Distance Predicting Functions. Technical Report 419. McMaster University, Faculty of Business. Hamilton, Ontario, Canada

LOVE, R. F. AND WALKER, J. H. 1991. A New Criterion for Evaluating the Accuracy of Distance Predicting Functions. Technical Report 370. McMaster University, Faculty of Business. Hamilton, Ontario, Canada

LOVE, R. F. AND WALKER, J. H. 1993. Distance Data for Eighteen Geographic Regions. Technical Report 383. McMaster University, Faculty of Business. Hamilton, Ontario, Canada

LOVE, R. F. AND WALKER, J. H. 1994. An Empirical Comparison of Block and Round Norms for Modeling Actual Distances, *Location Science*, **2(1)**, 21–43

LOVE, R. F., WALKER, J. H., AND TIKU, M. L. 1995. Confidence Intervals for

$\ell_{k,p,\theta}$ Distances, *Transportation Science*, **29**, 93–100

LOVE, R. F. AND YEONG, W. Y. 1981. A Stopping Rule for Facilities Location Algorithms, *AIIE Transactions*, **13**, 357–362

MICHELOT, C. 1993. The Mathematics of Continuous Location, *Studies in Locational Analysis*, **5**, 59–83

MICROANALYTICS 1993. TruckStops 2. Suite One, 2045 North 15th Street, Arlington, Virginia and 1986 Queen Street East, Toronto, Ontario

MIEHLE, W. 1958. Link-Length Minimization in Networks, *Operations Research*, **6(2)**, 232–243

MILLER, H. J. 1996. GIS and geometric representation in facility location problems, *International Journal of Geographical Information Systems*, **10(7)**, 791–816

MORRIS, J. G. 1981. Convergence of the Weiszfeld Algorithm for Weber Problems Using a Generalized "Distance" Function, *Operations Research*, **29**, 37–48

MORRIS, J. G. AND VERDINI, W. A. 1979. Minisum ℓ_p distance Location Problems Solved via a Perturbed Problem and Weiszfeld's Algorithm, *Operations Research*, **27**, 1180–1188

MURRAY, A. T. AND ESTIVILL-CASTRO, V. 1998. Cluster Discovery Techniques for Exploratory Spatial Data Analysis, *International Journal of Geographical Information Science*, **12(5)**, 431–443

ORTEGA, J. M. AND RHEINBOLDT, W. C. 1970. *Iterative Solution of Nonlinear*

Equations in Several Variables. Academic Press. New York

PLASTRIA, F. 1984. Localization in Single Facility Location, *Journal of Operational Research*, **18**, 215–219

PLASTRIA, F. 1992. When Facilities Coincide: Exact Optimality Conditions in Multifacility Location, *Journal of Mathematical Analysis and Application*, **169**, 476–498

PLASTRIA, F. 1995. Continuous location problems. in Z. Drezner (ed.), *Facility Location, A Survey of Applications and Methods.* Springer-Verlag. New York

POLAK, E. 1971. *Computational Methods in Optimization.* Academic Press. New York

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T., AND FLANNERY, B. P. 1996. *Numerical Recipes in C.* Cambridge University Press. New York

ROADNET-TECHNOLOGIES 1993. ROADNET. 10540 York Rd., Huntvalley, Maryland

ROCKAFELLAR, R. T. 1970. *Convex Analysis.* Princeton University Press. New Jersey

RUDIN, W. 1976. *Principles of Mathematical Analysis.* McGraw-Hill, Inc.. New York

SCOTT, H. C., JEFFERSON, T. R., AND JORJANI, S. 1995. Conjugate duality in facility location. in Z. Drezner (ed.), *Facility Location, A Survey of Applications and Methods.* Springer-Verlag. New York

SPSS 1997. Release 8.0.0.. 444 N.Michigan Ave., Chicago, IL

STAR, J. AND ESTES, J. 1990. *Geographic Information Systems – An Introduction.* Prentice-Hall. New Jersey

STONE, R. E. 1991. Some Average Distance Results, *Transportation Science*, **25(1)**, 83–91

STUART, A. AND ORD, J. K. 1987. *Kendall's Advanced Theory of Statistics, Volume I - Distribution Theory.* Oxford University Press. New York

THISSE, J. F., WARD, J. E., AND WENDELL, R. E. 1984. Some Properties of Location Problems with Block and Round Norms, *Operations Research*, **32**, 1309–1327

VAN LAARHOVEN, P. J. M. AND AARTS, E. H. L. 1987. *Simulated Annealing, Theory and Applications.* Kluwer Academic Publishers. Boston

VAUGHAN, R. 1984. Approximate formulas for average distances associated with zones, *Transportation Science*, **18(3)**, 231–244

VERDINI, W. A. 1976. On Solving a Class of Location and Location-Allocation Problems with Extensions to Clustering. *Ph.D. thesis.* Kent State University. Graduate School of Business Administration

WARD, J. E. AND WENDELL, R. E. 1980. A New Norm for Measuring Distance Which Yields Linear Location Models, *Operations Research*, **28**, 836–844

WARD, J. E. AND WENDELL, R. E. 1985. Using Block Norms for Location Modeling, *Operations Research*, **33**, 1074–1090

WEBER, A. 1909. *Über den standort der Industrien (Theory of the Location and Industries).* University of Chicago Press. (Friedrich, C.J. Trans. 1929)

WEISZFELD, E. 1937. Sur le point lequel la somme des distances de n points donnés est minimum, *Tôhoku Mathematical Journal*, **43**, 355–386

WENDELL, R. E. AND HURTER, A. P. 1985. Location Theory, Dominance and Convexity, *Operations Research*, **21**, 1074–1090

WENDELL, R. E. AND PETERSON, E. L. 1984. A Dual Approach for Obtaining Lower Bounds to the Weber Problem, *Journal of Regional Science*, **24**, 219–228

WERSAN, S. J., QUON, J. E., AND CHARNES, A. 1962. Systems Analysis of Refuse Collection and Disposable Practices, *American Public Works Association, yearbook*, 195–211

WESOLOWSKY, G. O. 1976. *Multiple Regression and Analysis of Variance.* John Wiley and Sons. Toronto

WESOLOWSKY, G. O. 1993. The Weber Problem: History and Perspectives, *Location Science*, **1(1)**, 5–23

WESOLOWSKY, G. O. AND LOVE, R. F. 1971. The Optimal Location of New Facilities Using Rectangular Distances, *Operations Research*, **19**, 124–130

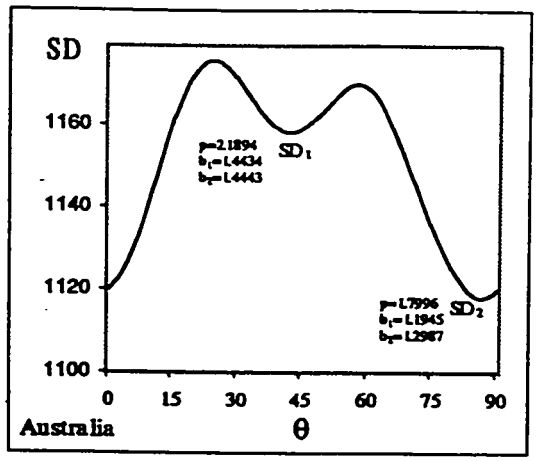
WESOLOWSKY, G. O. AND LOVE, R. F. 1972. A Nonlinear Approximation Method for Solving a Generalized Rectangular Distances Weber Problem, *Management Science*, **18**, 656–663

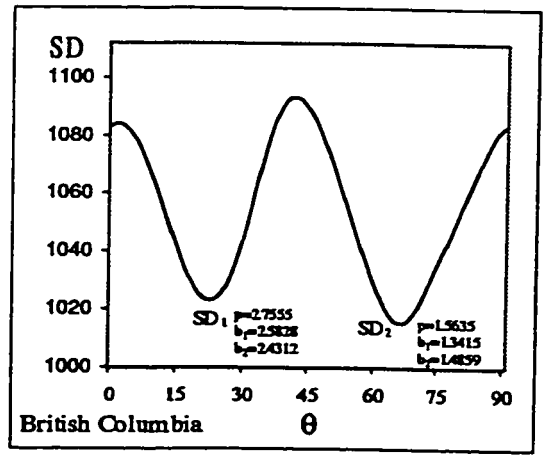
WESTWOOD, J. B. 1977. A Transport Planning Model for Primary Distribution, *Interfaces*, 8, 1–10

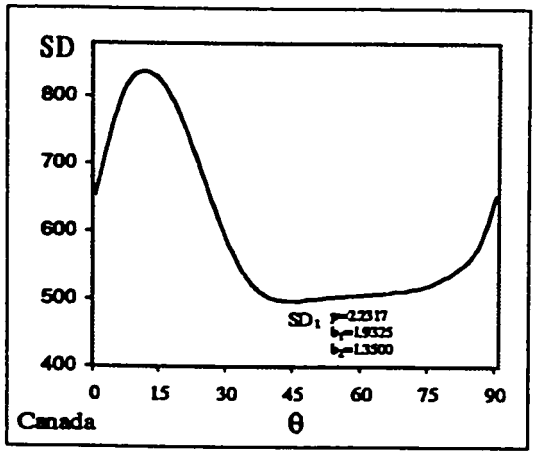
WITZGALL, C. 1964. Optimal Location of a Central Facility-Mathematical Models and Concepts. Technical Report 8388. National Bureau of Standards. U.S. Department of Commerce

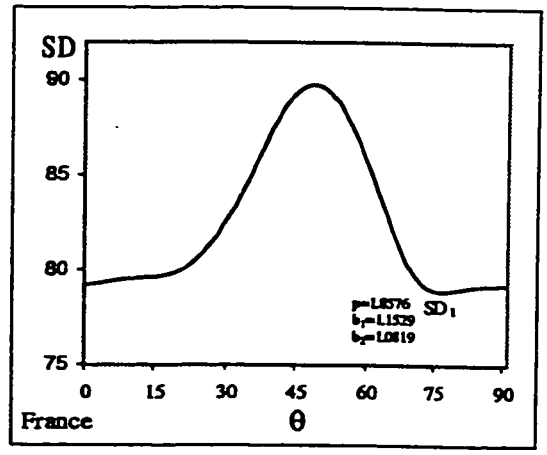
Appendix A

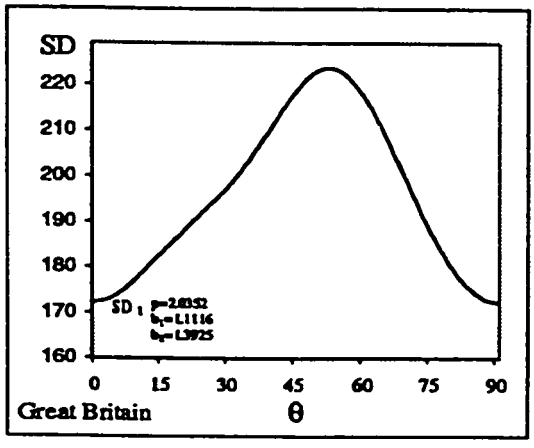
$SD(\theta)$ Plots

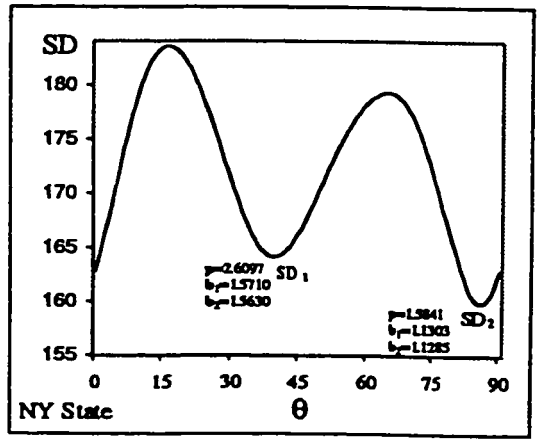

SD
1160
1140
1120
1100
0 15 30 45 60 75 90
θ
Australia
SD1
SD2

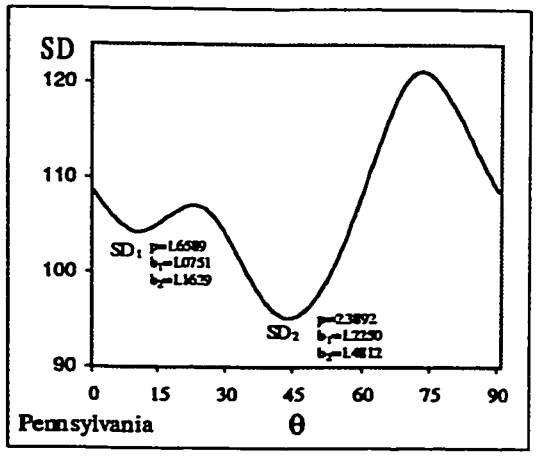

SD
1100
1080
1060
1040
1020
1000
0 15 30 45 60 75 90
θ
British Columbia
SD1
SD2

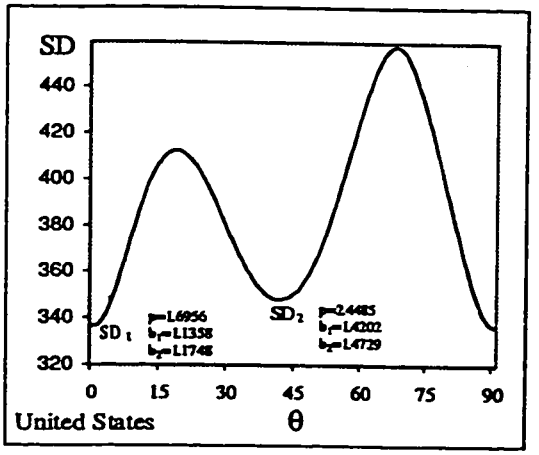

SD
800
700
600
500
400
0 15 30 45 60 75 90
θ
Canada
SD1

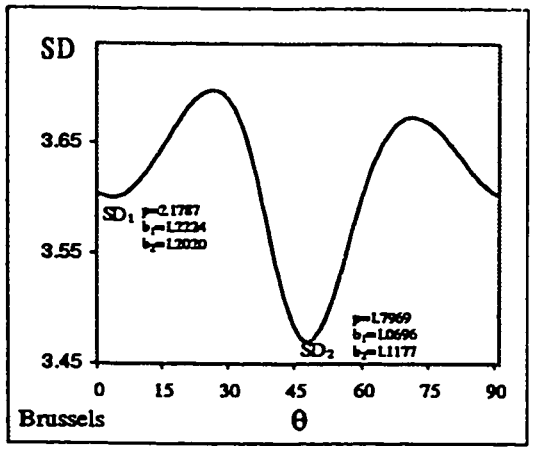

SD
90
85
80
75
0 15 30 45 60 75 90
θ
France
SD1

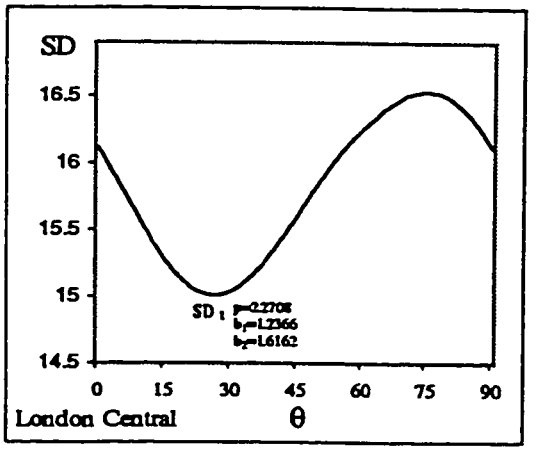

SD
220
210
200
190
180
170
160
0 15 30 45 60 75 90
θ
Great Britain
SD1

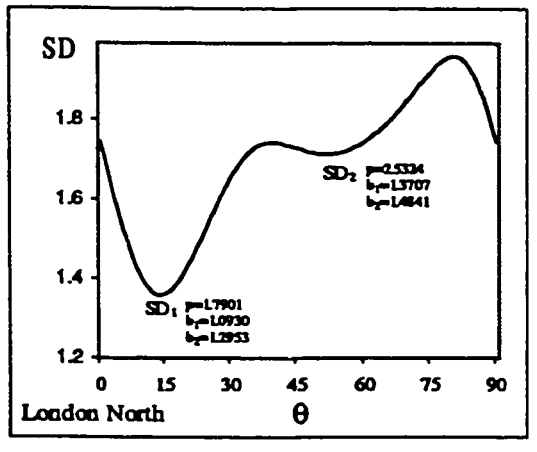

SD
180
175
170
165
160
155
0 15 30 45 60 75 90
θ
NY State
SD1
SD2

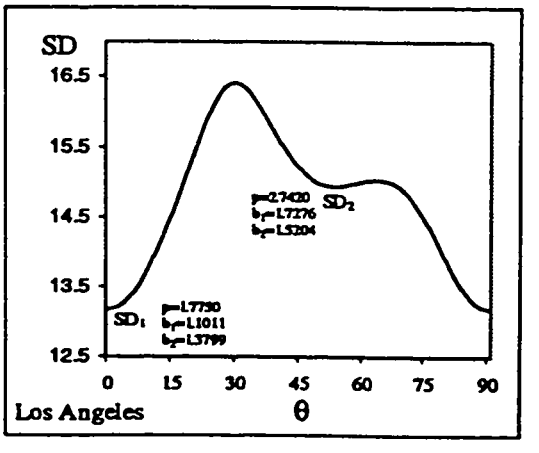

SD
120
110
100
90
0 15 30 45 60 75 90
θ
Pennsylvania

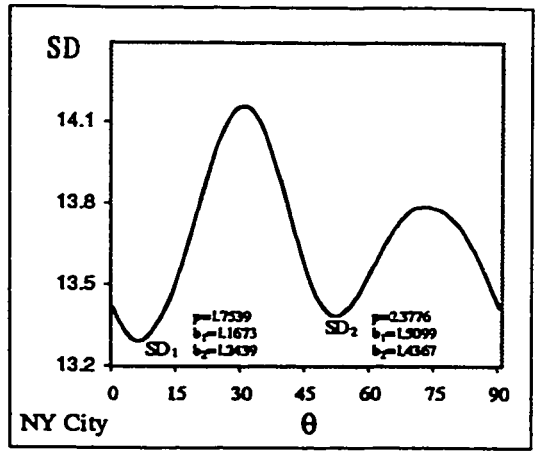

SD
440
420
400
380
360
340
320
0 15 30 45 60 75 90
θ
United States

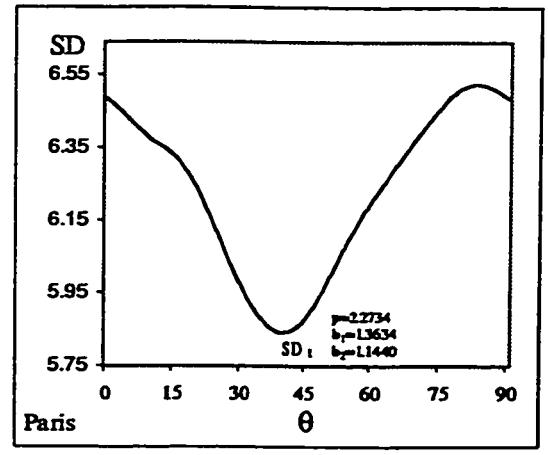

SD
3.65
3.55
3.45
0 15 30 45 60 75 90
θ
Brussels

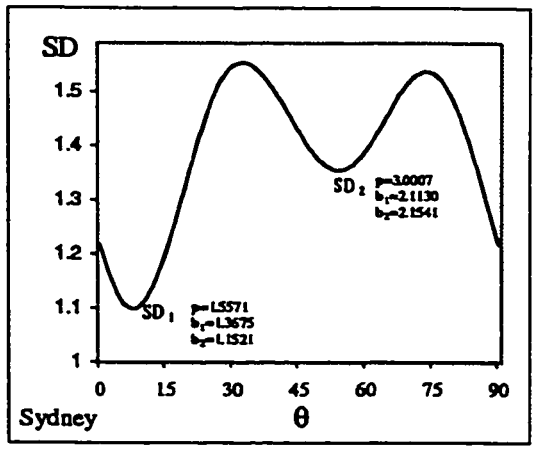

SD
16.5
16
15.5
15
14.5
0 15 30 45 60 75 90
θ
London Central

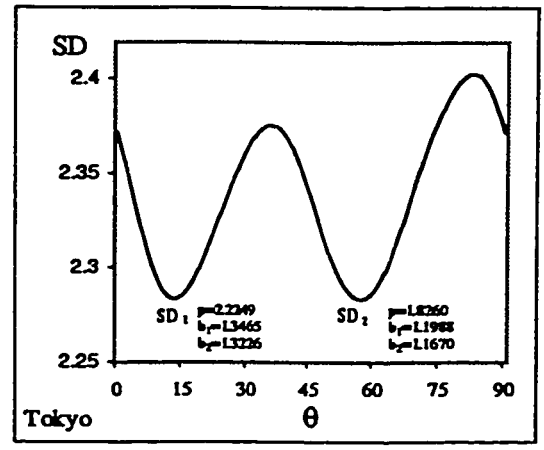

SD
1.8
1.6
1.4
1.2
0 15 30 45 60 75 90
θ
London North

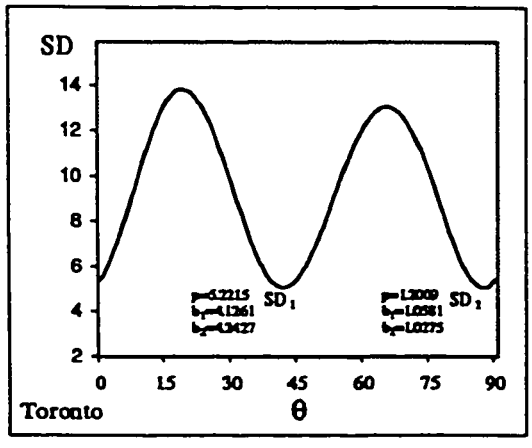

SD
16.5
15.5
14.5
13.5
12.5
0 15 30 45 60 75 90
θ
Los Angeles


SD
14.1
13.8
13.5
13.2
0 15 30 45 60 75 90
θ
NY City


SD
6.55
6.35
6.15
5.95
5.75
0 15 30 45 60 75 90
θ
Paris

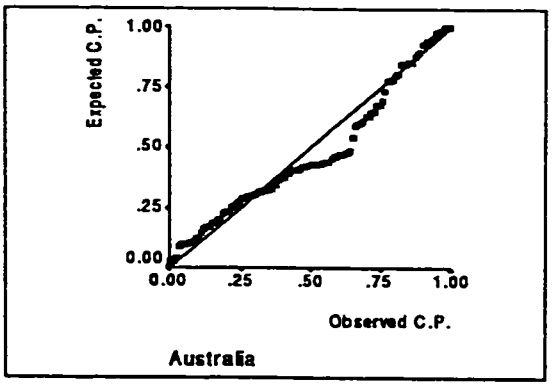

SD
1.5
1.4
1.3
1.2
1.1
1
0 15 30 45 60 75 90
θ
Sydney

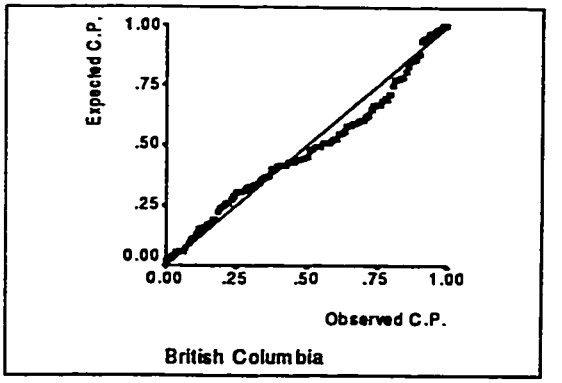

SD
2.4
2.35
2.3
2.25
0 15 30 45 60 75 90
θ
Tokyo

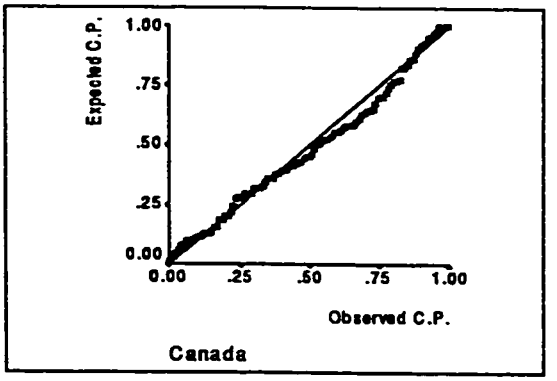

SD
14
12
10
8
6
4
2
0 15 30 45 60 75 90
θ
Toronto

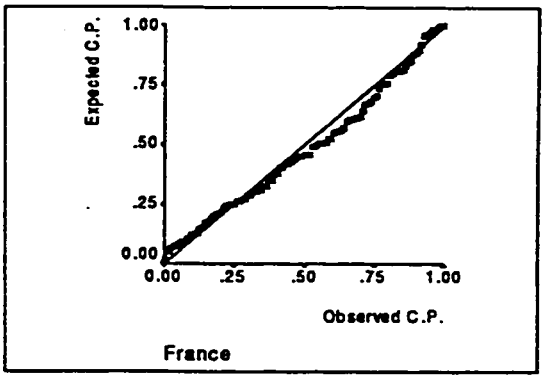
Appendix B

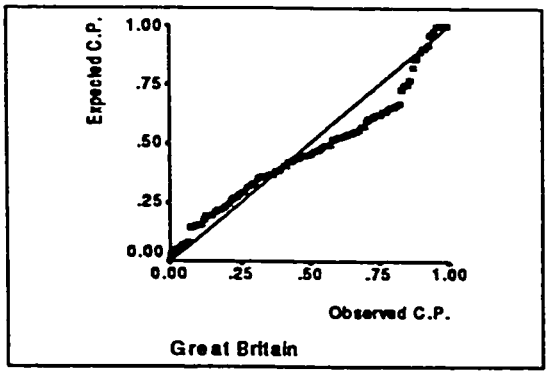
Normal Probability Plots of $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$

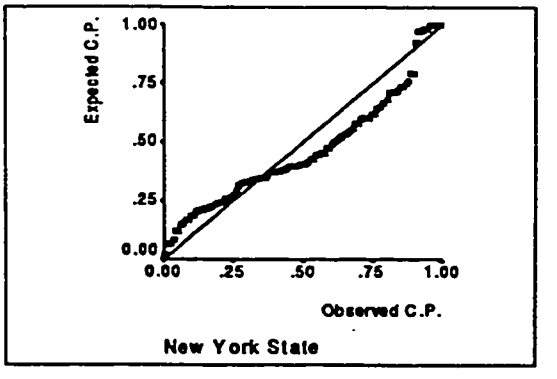

Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
Australia

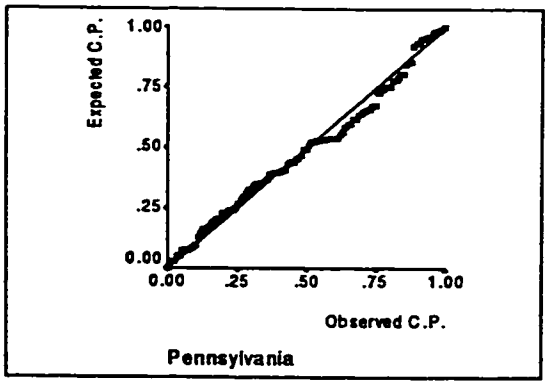

Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
British Columbia

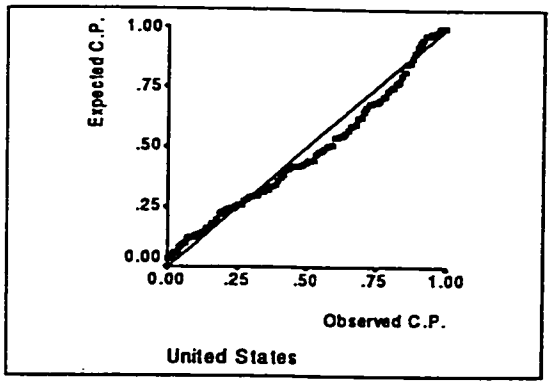

Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
Canada

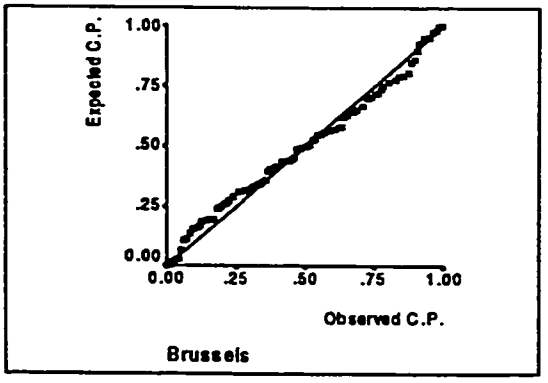

Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
France

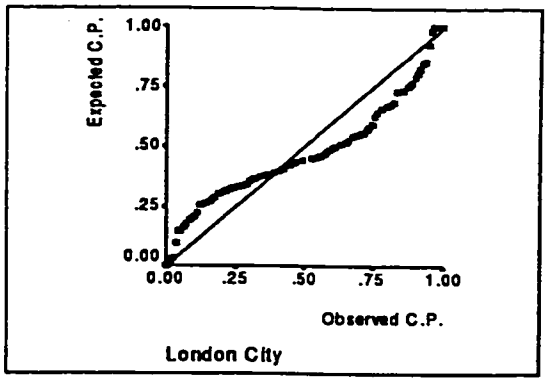

Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
Great Britain

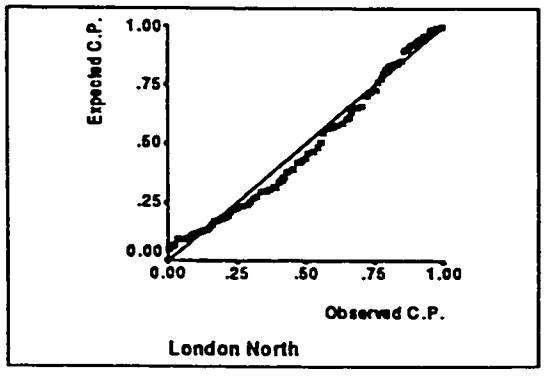

Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
New York State

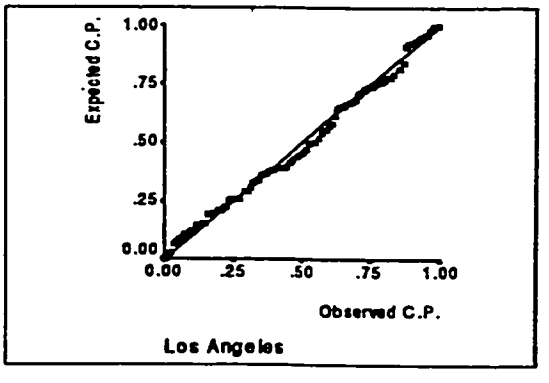

Expected C.P.
1.00
.75
.50
.25
0.00
0.00 .25 .50 .75 1.00
Observed C.P.
Pennsylvania

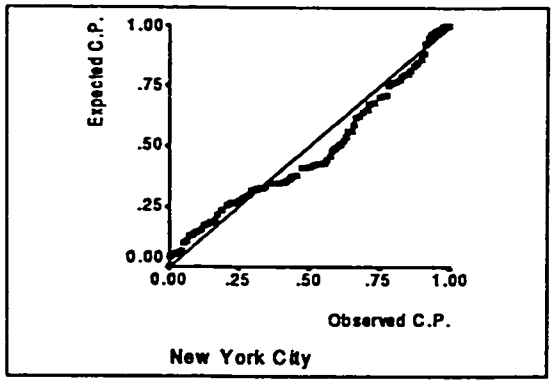

Expected C.P.
1.00
.75
.50
.25
0.00
0.00 .25 .50 .75 1.00
Observed C.P.
United States

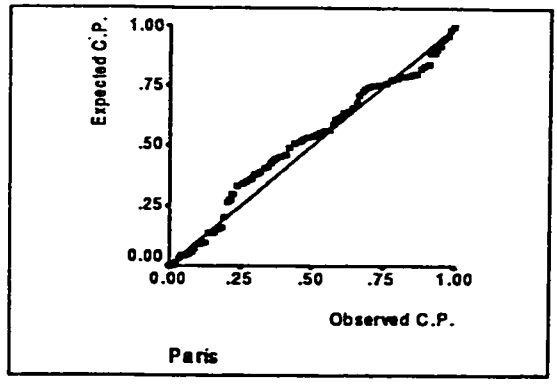

Expected C.P.
1.00
.75
.50
.25
0.00
0.00 .25 .50 .75 1.00
Observed C.P.
Brussels

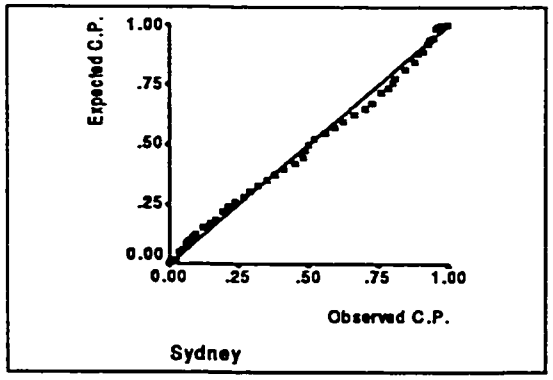

Expected C.P.
1.00
.75
.50
.25
0.00
0.00 .25 .50 .75 1.00
Observed C.P.
London City

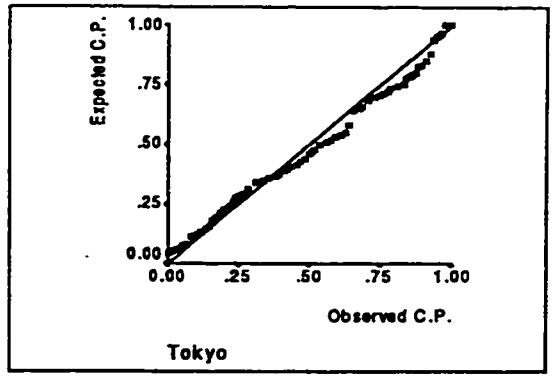

Expected C.P.
1.00
.75
.50
.25
0.00
0.00 .25 .50 .75 1.00
Observed C.P.
London North

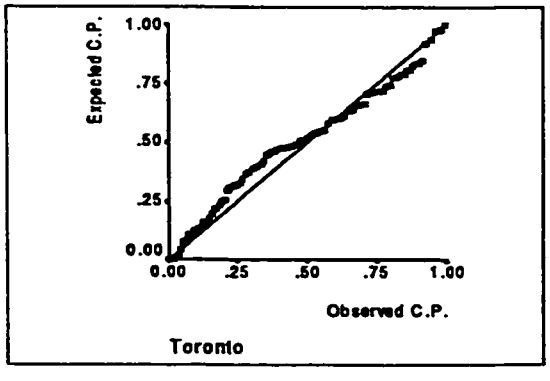

Expected C.P.
1.00
.75
.50
.25
0.00
0.00 .25 .50 .75 1.00
Observed C.P.
Los Angeles


Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
New York City


Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
Paris

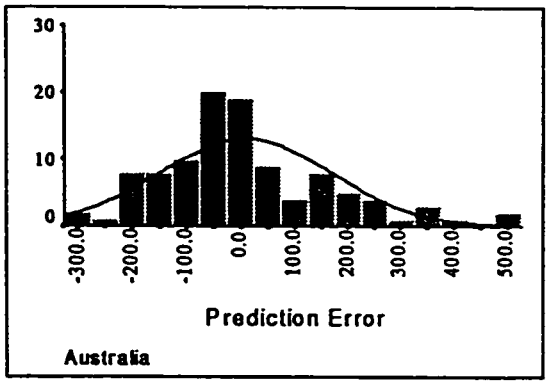

Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
Sydney

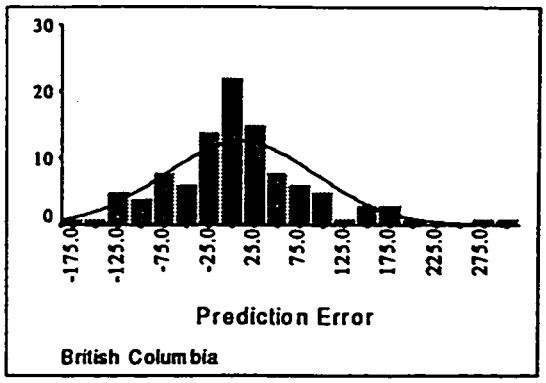

Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
Tokyo

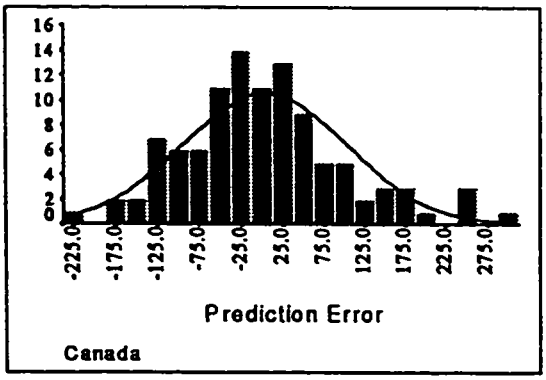

Expected C.P.
1.00
.75
.50
.25
0.00
0.00
.25
.50
.75
1.00
Observed C.P.
Toronto

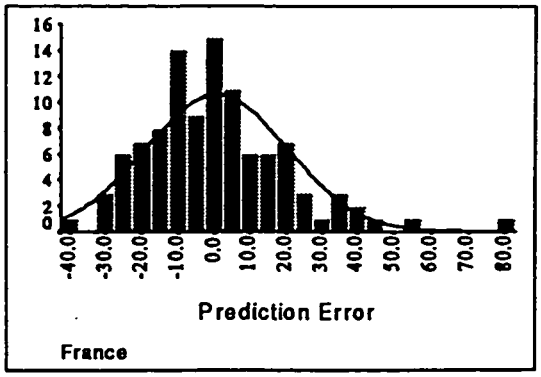
Appendix C

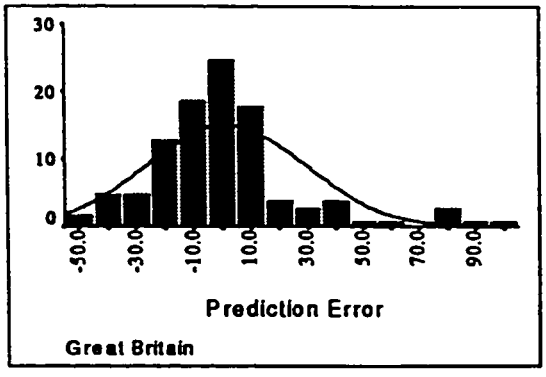
Histograms of $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$

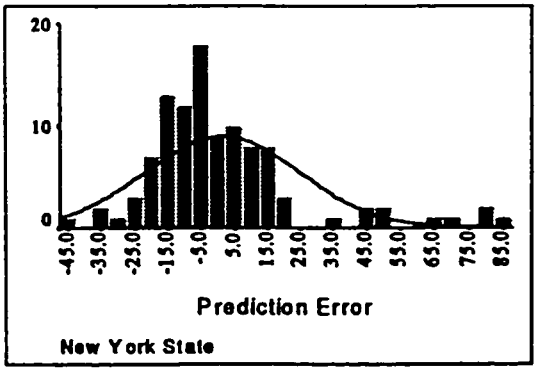

30
20
10
0
-300.0
-200.0
-100.0
0.0
100.0
200.0
300.0
400.0
500.0
Prediction Error
Australia

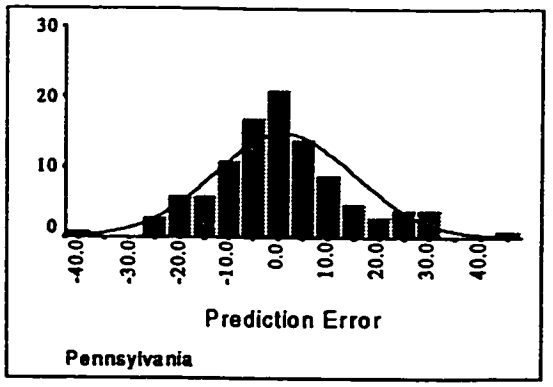

30
20
10
0
-175.0
-125.0
-75.0
-25.0
25.0
75.0
125.0
175.0
225.0
275.0
Prediction Error
British Columbia

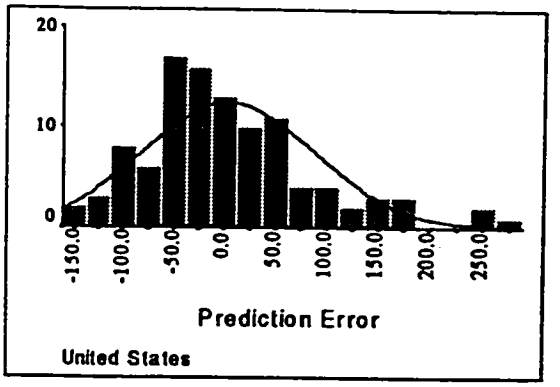

16
14
12
10
8
6
4
2
0
-225.0
-175.0
-125.0
-75.0
-25.0
25.0
75.0
125.0
175.0
225.0
275.0
Prediction Error
Canada

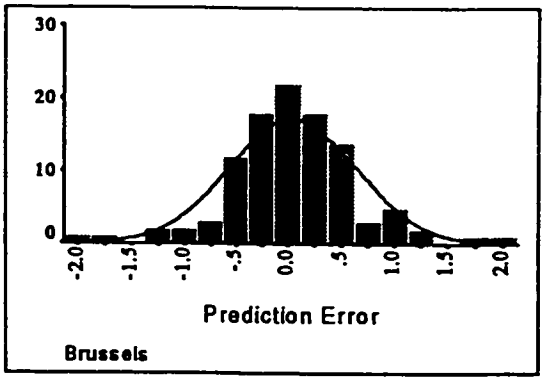

16
14
12
10
8
6
4
2
0
-40.0
-30.0
-20.0
-10.0
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
Prediction Error
France

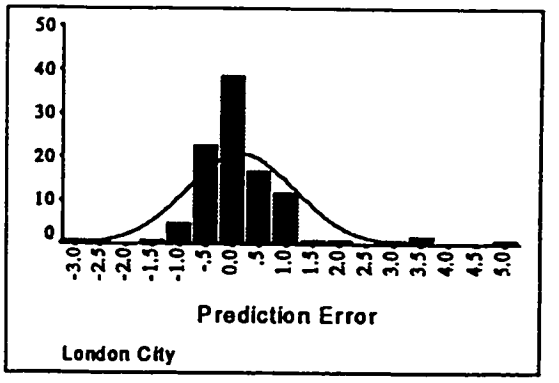

30
20
10
0
-50.0
-30.0
-10.0
10.0
30.0
50.0
70.0
90.0
Prediction Error
Great Britain

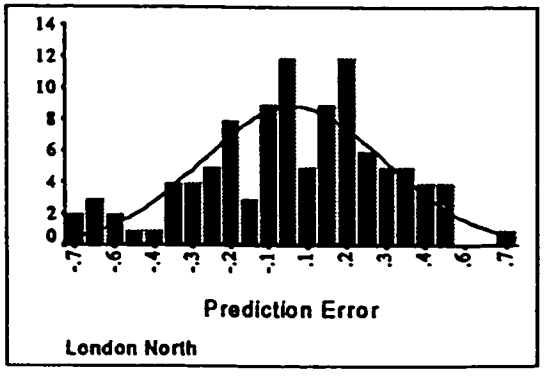

20
10
0
-45.0
-35.0
-25.0
-15.0
-5.0
5.0
15.0
25.0
35.0
45.0
55.0
65.0
75.0
85.0
Prediction Error
New York State

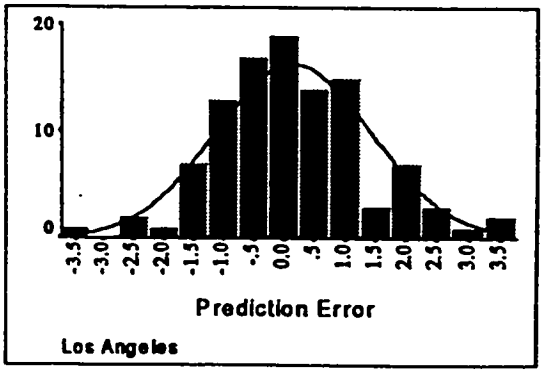

30
20
10
0
-40.0
-30.0
-20.0
-10.0
0.0
10.0
20.0
30.0
40.0
Prediction Error
Pennsylvania

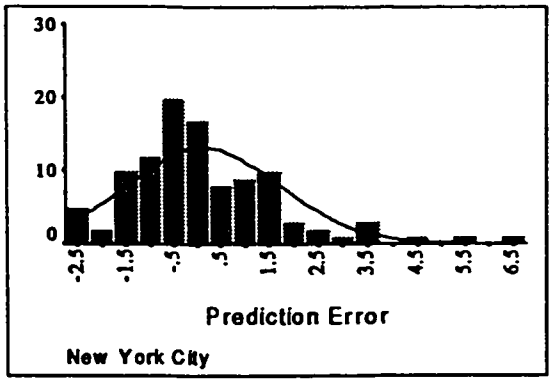

20
10
0
-150.0
-100.0
-50.0
0.0
50.0
100.0
150.0
200.0
250.0
Prediction Error
United States

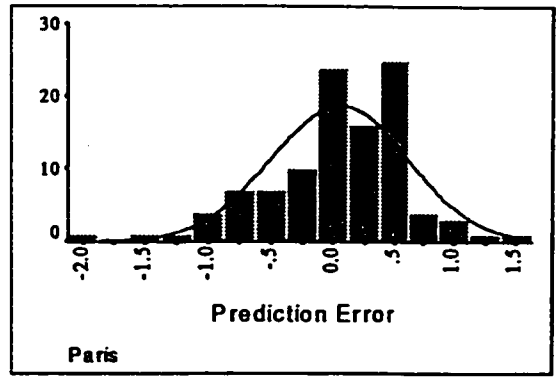

30
20
10
0
-2.0
-1.5
-1.0
-.5
0.0
.5
1.0
1.5
2.0
Prediction Error
Brussels

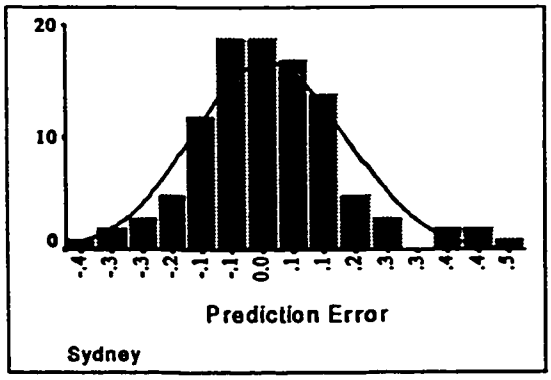

50
40
30
20
10
0
-3.0
-2.5
-2.0
-1.5
-1.0
-.5
0.0
.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Prediction Error
London City

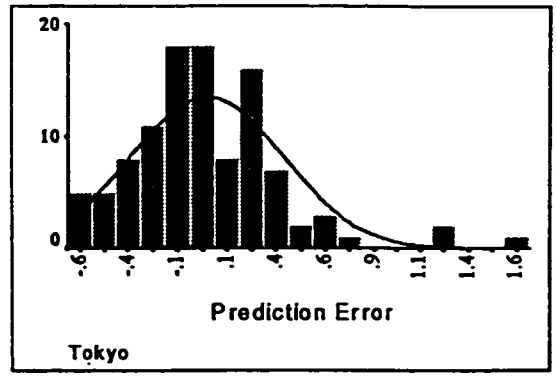

14
12
10
8
6
4
2
0
-.7
-.6
-.4
-.3
-.2
-.1
.1
.2
.3
.4
.6
.7
Prediction Error
London North

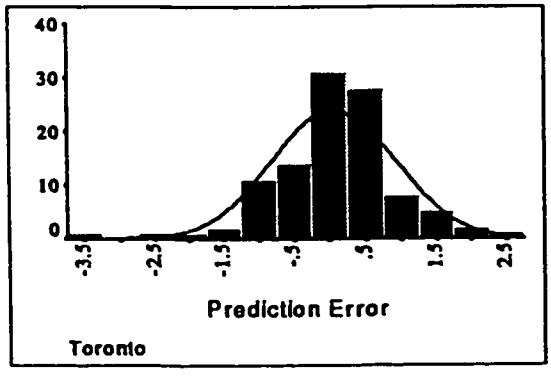

20
10
0
-3.5
-3.0
-2.5
-2.0
-1.5
-1.0
-.5
0.0
.5
1.0
1.5
2.0
2.5
3.0
3.5
Prediction Error
Los Angeles


30
20
10
0
-2.5
-1.5
-.5
.5
1.5
2.5
3.5
4.5
5.5
6.5
Prediction Error
New York City


30
20
10
0
-2.0
-1.5
-1.0
-.5
0.0
.5
1.0
1.5
Prediction Error
Paris

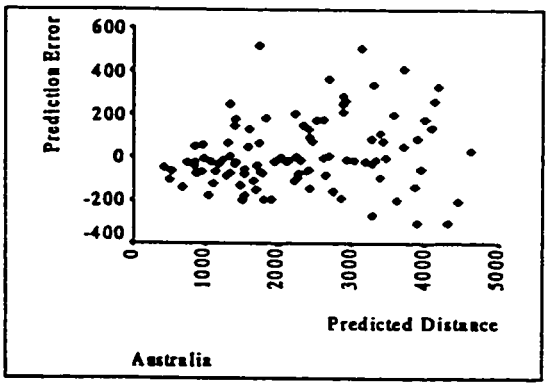

20
10
0
-.4
-.3
-.3
-.2
-.1
-.1
0.0
.1
.1
.2
.3
.3
.4
.4
.5
Prediction Error
Sydney

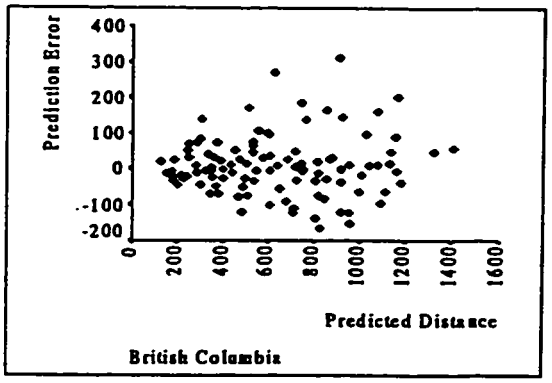

20
10
0
-.6
-.4
-.1
.1
.4
.6
.9
1.1
1.4
1.6
Prediction Error
Tokyo

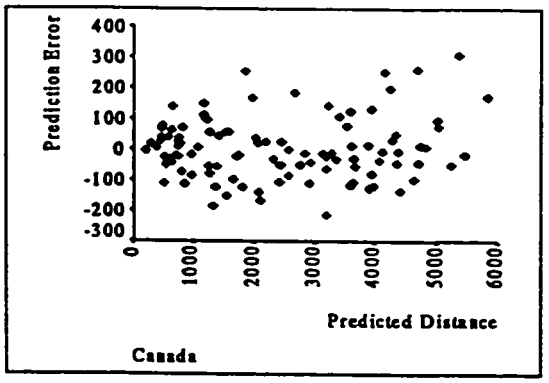

40
30
20
10
0
-3.5
-2.5
-1.5
-.5
.5
1.5
2.5
Prediction Error
Toronto

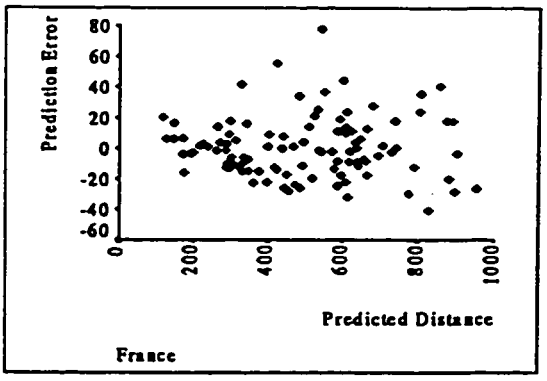
Appendix D

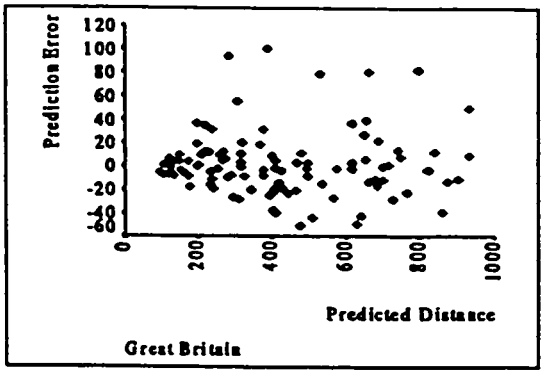
Scatter Plots of $\varepsilon(\mathbf{x}_i, \mathbf{x}_j)$

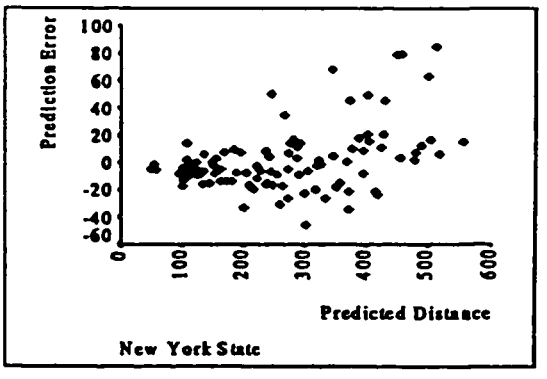

Prediction Error
600
400
200
0
-200
-400
0
1000
2000
3000
4000
5000
Predicted Distance
Australia

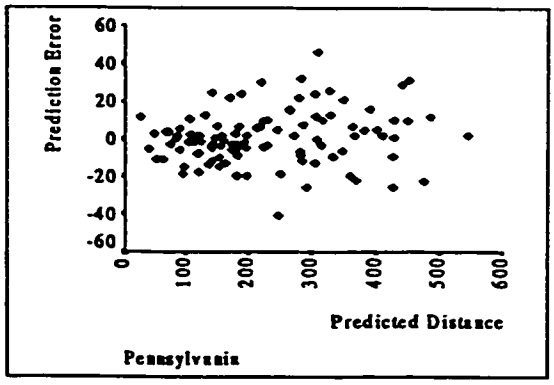

Prediction Error
400
300
200
100
0
-100
-200
0
200
400
600
800
1000
1200
1400
1600
Predicted Distance
British Columbia

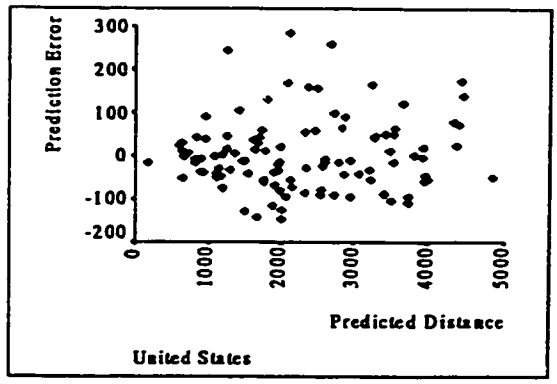

Prediction Error
400
300
200
100
0
-100
-200
-300
0
1000
2000
3000
4000
5000
6000
Predicted Distance
Canada

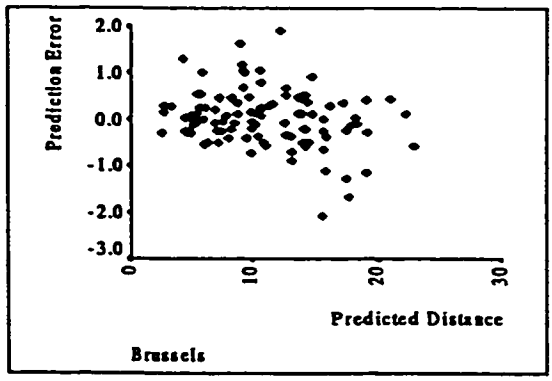

Prediction Error
80
60
40
20
0
-20
-40
-60
0
200
400
600
800
1000
Predicted Distance
France

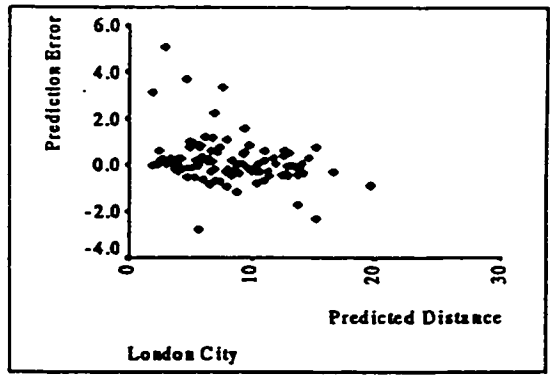

Prediction Error
120
100
80
60
40
20
0
-20
-40
-60
0
200
400
600
800
1000
Predicted Distance
Great Britain

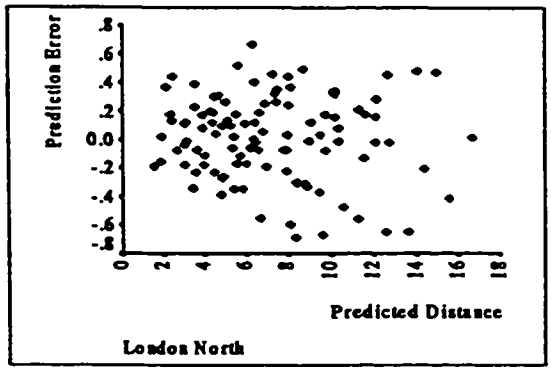

Prediction Error
100
80
60
40
20
0
-20
-40
-60
0
100
200
300
400
500
600
Predicted Distance
New York State

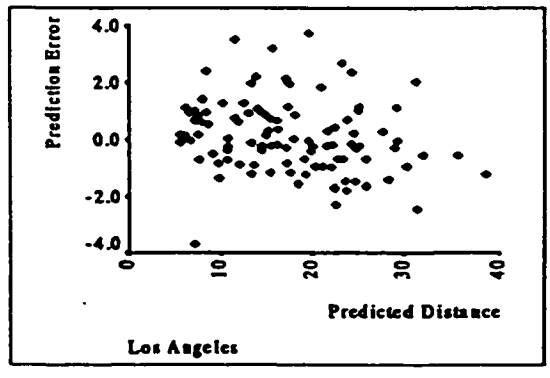

Prediction Error
60
40
20
0
-20
-40
-60
0
100
200
300
400
500
600
Predicted Distance
Pennsylvania

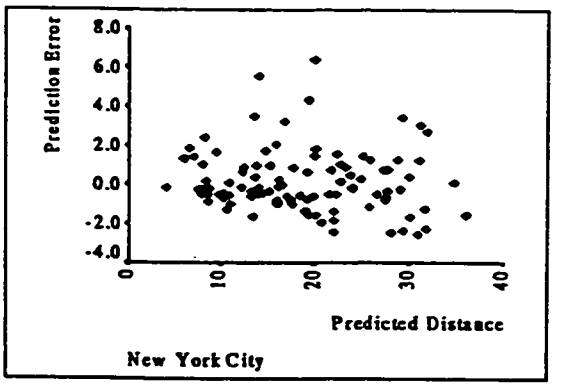

Prediction Error
300
200
100
0
-100
-200
0
1000
2000
3000
4000
5000
Predicted Distance
United States

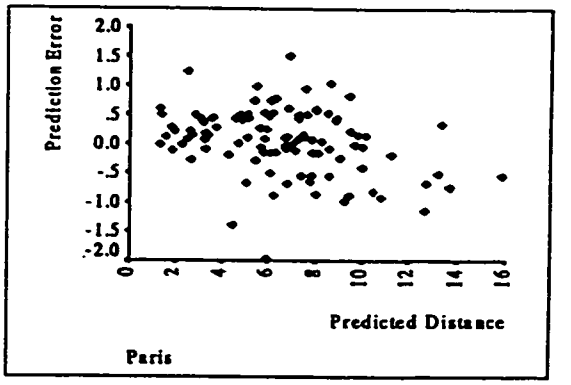

Prediction Error
2.0
1.0
0.0
-1.0
-2.0
-3.0
0
10
20
30
Predicted Distance
Brussels

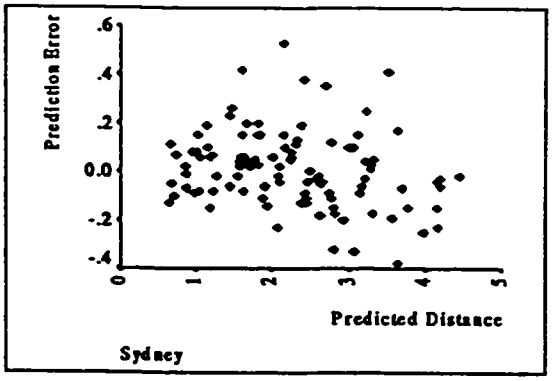

Prediction Error
6.0
4.0
2.0
0.0
-2.0
-4.0
0
10
20
30
Predicted Distance
London City

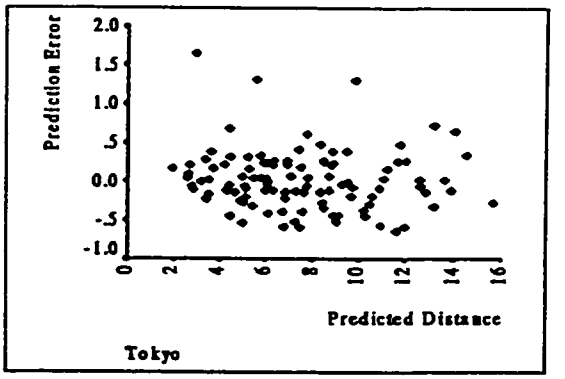

Prediction Error
.8
.6
.4
.2
0.0
-.2
-.4
-.6
-.8
0
2
4
6
8
10
12
14
16
18
Predicted Distance
London North

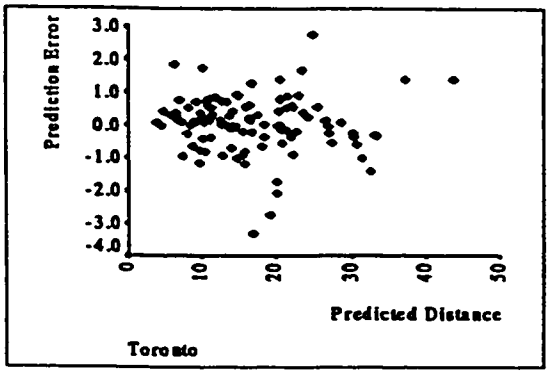

Prediction Error
4.0
2.0
0.0
-2.0
-4.0
0
10
20
30
40
Predicted Distance
Los Angeles


Prediction Error
8.0
6.0
4.0
2.0
0.0
-2.0
-4.0
0
10
20
30
40
Predicted Distance
New York City


Prediction Error
2.0
1.5
1.0
.5
0.0
-.5
-1.0
-1.5
-2.0
0
2
4
6
8
10
12
14
16
Predicted Distance
Paris

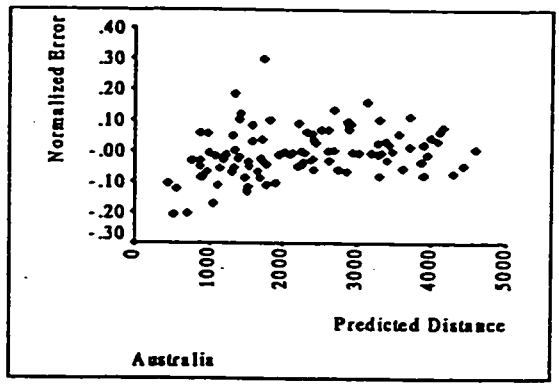

Prediction Error
.6
.4
.2
0.0
-.2
-.4
0
1
2
3
4
5
Predicted Distance
Sydney

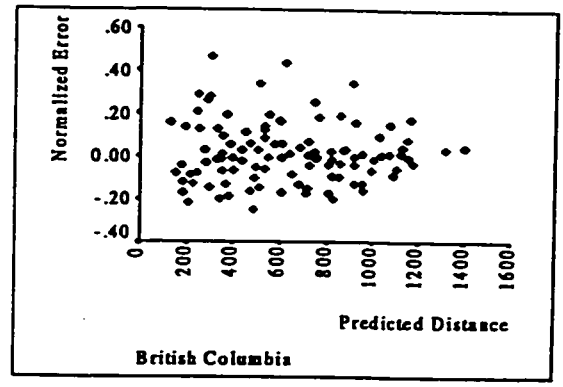

Prediction Error
2.0
1.5
1.0
.5
0.0
-.5
-1.0
0
2
4
6
8
10
12
14
16
Predicted Distance
Tokyo

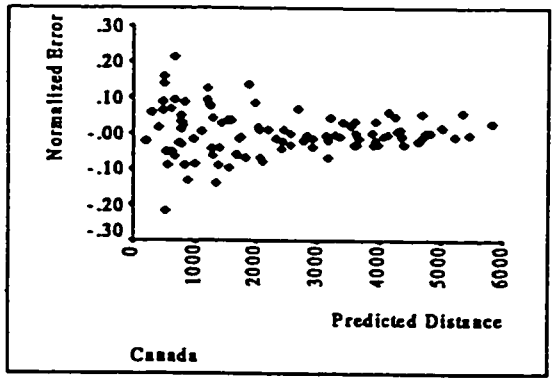

Prediction Error
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
0
10
20
30
40
50
Predicted Distance
Toronto

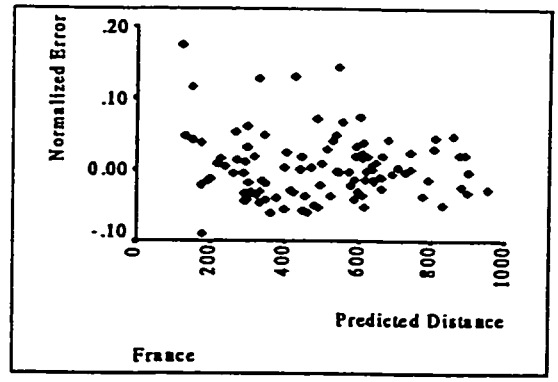
Appendix E

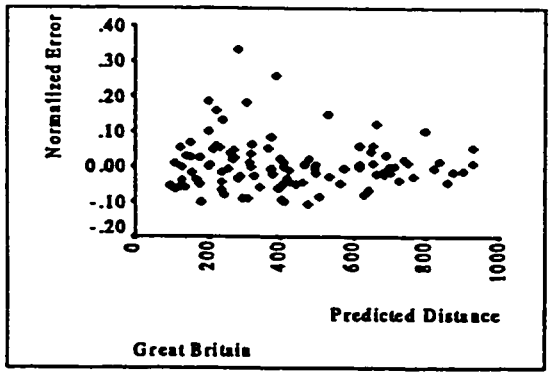
Scatter Plots of $\varepsilon_1(\mathbf{x}_i, \mathbf{x}_j)$

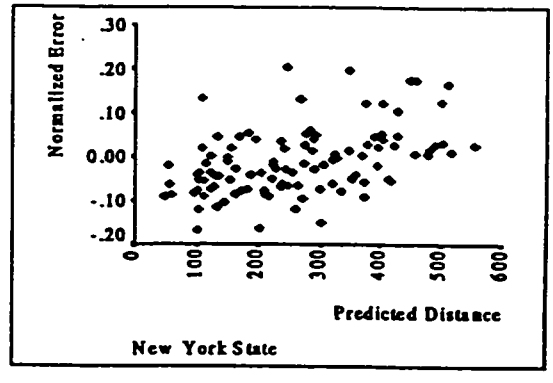

Normalized Error
.40
.30
.20
.10
-.00
-.10
-.20
-.30
0
1000
2000
3000
4000
5000
Predicted Distance
Australia

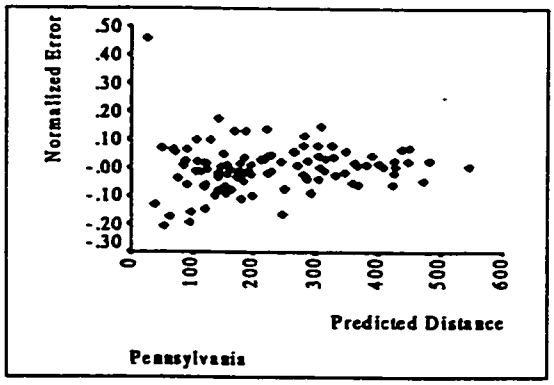

Normalized Error
.60
.40
.20
0.00
-.20
-.40
0
200
400
600
800
1000
1200
1400
1600
Predicted Distance
British Columbia

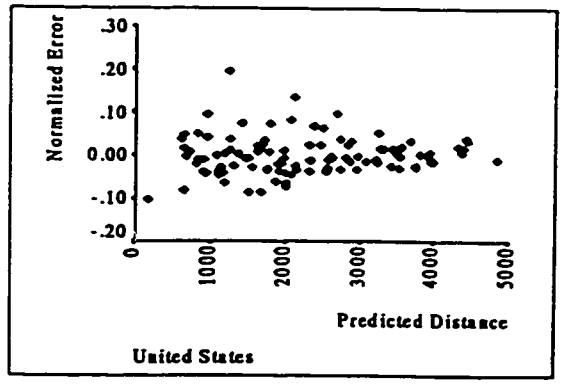

Normalized Error
.30
.20
.10
-.00
-.10
-.20
-.30
0
1000
2000
3000
4000
5000
6000
Predicted Distance
Canada

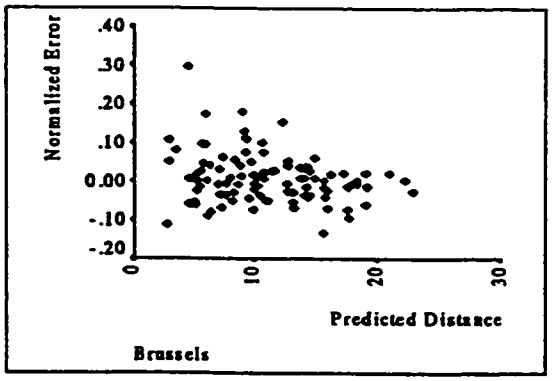

Normalized Error
.20
.10
0.00
-.10
0
200
400
600
800
1000
Predicted Distance
France

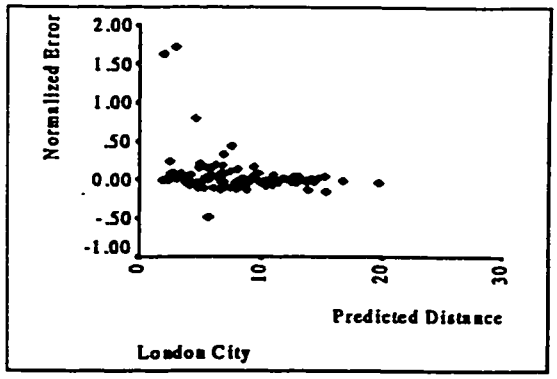

Normalized Error
.40
.30
.20
.10
0.00
-.10
-.20
0
200
400
600
800
1000
Predicted Distance
Great Britain

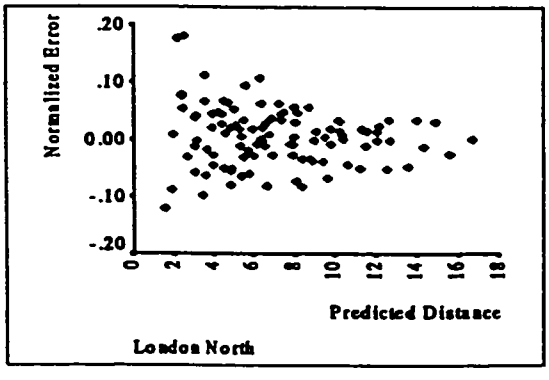

Normalized Error
.30
.20
.10
0.00
-.10
-.20
0
100
200
300
400
500
600
Predicted Distance
New York State

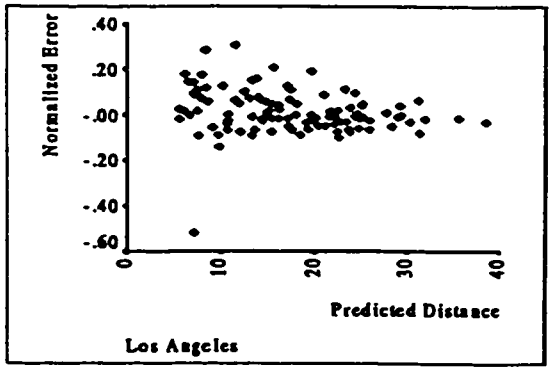

Normalized Error
.50
.40
.30
.20
.10
-.00
-.10
-.20
-.30
0
100
200
300
400
500
600
Predicted Distance
Pennsylvania

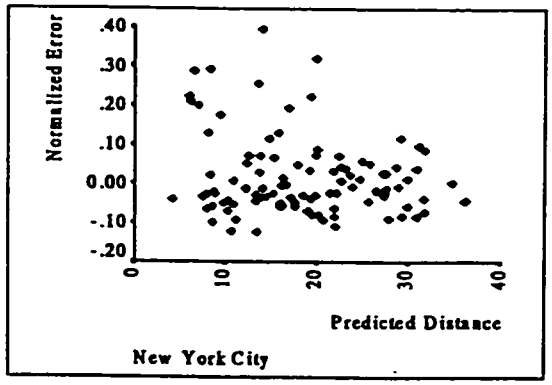

Normalized Error
.30
.20
.10
0.00
-.10
-.20
0
1000
2000
3000
4000
5000
Predicted Distance
United States

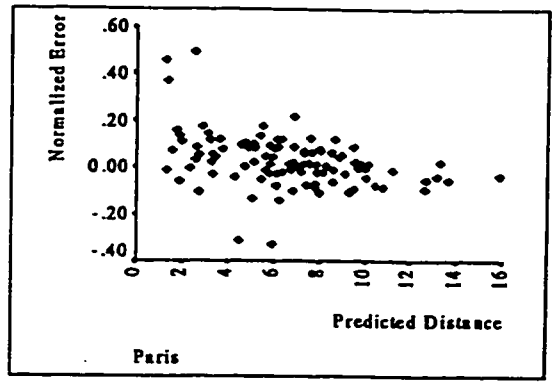

Normalized Error
.40
.30
.20
.10
0.00
-.10
-.20
0
10
20
30
Predicted Distance
Brussels

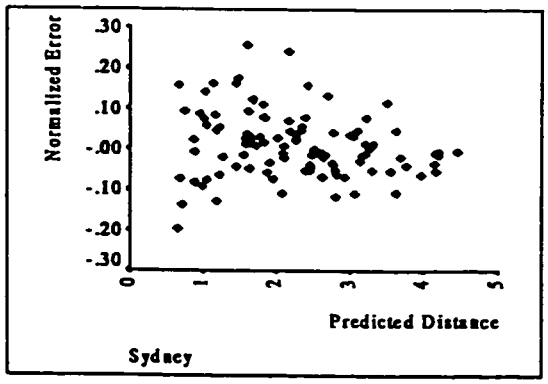

Normalized Error
2.00
1.50
1.00
.50
0.00
-.50
-1.00
0
10
20
30
Predicted Distance
London City

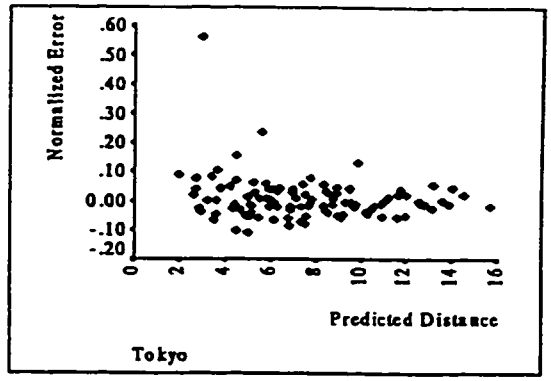

Normalized Error
.20
.10
0.00
-.10
-.20
0
2
4
6
8
10
12
14
16
18
Predicted Distance
London North

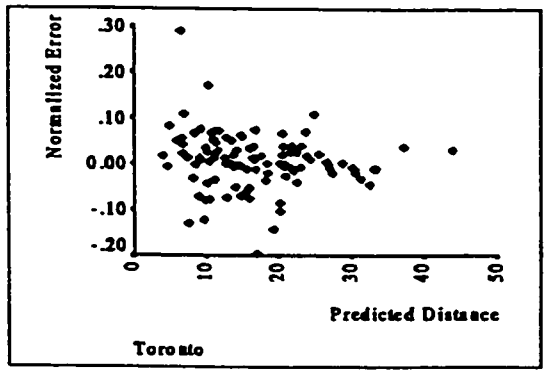

Normalized Error
.40
.20
-.00
-.20
-.40
-.60
0
10
20
30
40
Predicted Distance
Los Angeles


Normalized Error
.40
.30
.20
.10
0.00
-.10
-.20
0
10
20
30
40
Predicted Distance
New York City


Normalized Error
.60
.40
.20
0.00
-.20
-.40
0
2
4
6
8
10
12
14
16
Predicted Distance
Paris


Normalized Error
.30
.20
.10
-.00
-.10
-.20
-.30
0
1
2
3
4
5
Predicted Distance
Sydney


Normalized Error
.60
.50
.40
.30
.20
.10
0.00
-.10
-.20
0
2
4
6
8
10
12
14
16
Predicted Distance
Tokyo


Normalized Error
.30
.20
.10
0.00
-.10
-.20
0
10
20
30
40
50
Predicted Distance
Toronto