PROPERTIES OF DISTANCE FUNCTIONS AND MINISUM

LOCATION MODELS

PROPERTIES OF DISTANCE FUNCTIONS AND MINISUM LOCATION MODELS

By

JACK BRIMBERG, B.ENG., M.ENG., M.B.A.

A Thesis

Submitted to the School of Graduate Studies in Partial Fulfilment of the

Requirements for the Degree

Doctor of Philosophy

McMaster University

March 1989

DOCTOR OF PHILOSOPHY (1989) McMASTER UNIVERSITY

(Management Science/Systems) Hamilton, Ontario

TITLE: Properties of Distance Functions and Minisum Location Models

AUTHOR: Jack Brimberg, B.Eng. (McGill University)

M.Eng. (McGill University)

M.B.A. (University of Toronto)

SUPERVISOR: Professor R.F. Love

NUMBER OF PAGES: vii, 287

ABSTRACT

This study is divided into two main parts. The first section deals with mathematical properties of distance functions. The ℓ_p norm is analyzed as a function of its parameter p, leading to useful insights for fitting this distance measure to a transportation network. Properties of round norms are derived, which allow us later to generalize some well-known results. The properties of a norm raised to a power are also investigated, and these prove useful in our subsequent analysis of location problems with economies or dis-economies of scale. A positive linear combination of the Euclidean and rectangular distance measures, which we term the weighted one-two norm, is introduced. This distance function provides a linear regression model with interesting implications on the characterization of transportation networks. A directional bias function is defined, and examined in detail for the ℓ_p and weighted one-two norms.

In the second part of this study, several properties are derived for various forms of the continuous minisum location model. The Weiszfeld iterative solution procedure for the standard Weber problem with ℓ_p distances is also examined, and global and local convergence results obtained. These results are extended to the mixed-norm problem. In addition, optimality criteria are derived at non-differentiable points of the objective function.

ACKNOWLEDGEMENTS

I would like to thank Dr. Robert F. Love, my supervisor, whose encouragement, support and guidance have been invaluable throughout the development of this research. It has been a privilege and an honour to work under him.

To the rest of my supervisory committee, Dr. George O. Wesolowsky and Dr. N. Balakrishnan, I also wish to express my appreciation.

As well, I wish to thank Paul Behnke for his capable assistance in preparing the computer programs, and the ladies of the Engineering Word Processing Centre at McMaster University for the excellent typing of this manuscript and their patience in putting up with me.

To my father, Maurice Brimberg, most of all – thanks.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

Figure

CHAPTER 1

INTRODUCTION

Locational analysis deals with the formulation of location models and their solution. These models are mathematical representations of decision problems in which new facilities are to be situated. The term facility is used here in a generic sense to denote such diverse entities as warehouses in a geographical region, machines on a shop floor, and electronic components on a circuit board. Location models can be classified according to whether the set of possible sites is finite or infinite. The first category contains the discrete or network models. We shall be involved only with the latter category, known as continuous location models.

Whatever the practicalities of the real system at hand, the locational analysis invariably consists of the formulation of a mathematical model which is an optimization of some type, and a methodology to find a suitable solution to the model. Different criteria may be used for the optimization problem. In the minisum model, the objective is to minimize the total cost, defined as a sum of cost elements each of which is a function of some distance measure between two points (supply and demand centres). Alternatively, the objective might be to find a point which minimizes the maximum distance to a set of customers. This minimax criterion is popular in such cases as the location of emergency facilities (e.g., ambulance centers, fire stations, etc.) where service takes precedence over costs. A third, less-utilized criterion involves finding a point in a constrained region which maximizes the minimum distance to a set of customers (a maximin problem), as in the location of noxious facilities such as waste treatment plants.

Our aim in overview is to develop properties of various useful classes and functional forms of distance measures, and to examine the implications on certain important models and solution procedures. The location problems we shall look at will consist solely of different forms of the minisum model, with emphasis primarily on the location of a single new facility. The minimax and maximin problems will not be considered here. For a detailed exposition of these two topics, the interested reader is referred to Buchanan (1988).

We begin this chapter with a brief literature review of minisum location models, which is not intended to be a complete survey. Extensive lists of references on this subject and other general location problems can be found in the exhaustive bibliography by Domschke and Drexl (1985), in the selective reviews of Francis, McGinnis and White (1983), and Hansen, Peeters and Thisse (1983), and in the recent book by Love, Morris and Wesolowsky (1988). The purpose of our literature review is to provide a suitable background and motivation for the work in subsequent chapters. Next, we discuss some fundamental concepts pertaining to the distance functions employed in continuous location models, and the empirical work related to these functions. The empirical fitting of distance measures is carried out in order to improve the accuracy of the travel distances predicted by them in the system being modelled. The chapter finally ends with a summary of our objectives.

1.1 Minisum Models

The first known formulation of a location problem dates back to the early seventeenth century, when Fermat sought a point in the plane which minimized the sum of straight-line distances to three given points. This puzzle was worked on and extended over the centuries. For an interesting historical perspective, see Kuhn (1967). An important generalization known as the Weber, or Fermat-Weber, problem concerns the siting of a facility so as to minimize the sum of weighted distances to a set of fixed points. In a practical setting, these weighted distances represent cost components, and their sum gives the total cost. The fixed points have known locations, and are alternatively called demand points, demands, customers, destinations, existing facilities or vertices.

The Fermat-Weber problem and its extensions form a major part of continuous location theory, having received most of the attention of researchers in this field. A formulation of the basic problem in N-dimensional Euclidean space (R^N) is given below:

$$\text{minimize } W(x) = \sum_{i=1}^{n} w_i d(x, a_i) , \tag{1.1}$$

where $a_i=(a_{i1}, \ldots, a_{iN})^T$ is the known position of the i^{th} fixed point, $i=1, \ldots, n$; n equals the number of fixed points; $x=(x_1,\ldots,x_N)^T$ is the unknown position of the new facility; w_i is a positive weighting constant ($w_i>0$), which converts distance travelled between the facility and the i^{th} destination into a cost, for $i=1,\ldots,n$; and $d(y,z)$ is some function used to evaluate the distance between any two points $y,z \in R^N$. For the majority of practical applications, we have $N=2$; that is, the location problem occurs in the plane. Also note that the superscript T signifies the transpose operation, and we shall always deal with Euclidean spaces unless otherwise specified.

Problem (1.1) cannot be solved in closed form for general distance functions. An iterative numerical solution method was first proposed by Weiszfeld (1937), for d equal to the Euclidean (straight-line) distance on R^2. His technique remained in obscurity for several years, until its rediscovery by Miehle (1958), Kuhn and Kuenne (1962), and Cooper (1963). In a seminal paper by Kuhn (1973), global convergence of Weiszfeld's procedure is proven for the Euclidean case, provided that an iterate does not land on a fixed point. Furthermore, such an event is shown to occur only for a denumerable number of starting points. Thus, the probability that the algorithm will fail for an arbitrarily chosen starting point becomes negligible when high precision arithmetic is used. The iterative technique is generalized by

Morris and Verdini (1979) to ℓ_p distances on R^N. The ℓ_p function is a popular distance measure in location models, and will be introduced in the next section. Katz (1969) develops some convergence properties of the algorithm, adapted to the case where the cost components are general functions of Euclidean distances. The Weiszfeld procedure is easy to implement, and can be readily extended to generalizations of the Fermat-Weber problem. We shall have much more to say about this important method.

When rectangular (also called rectilinear, Manhattan or city-block) distances are used in problem (1.1), the coordinates of vector x become separable. The resulting N sub-problems, one for each x_i, can be solved quickly and exactly by hand or on the computer. The computational advantage of the rectangular measure over other distance functions extends to several variations of the Fermat-Weber problem, (e.g., see Wesolowsky, 1977). Thus, in addition to being the most appropriate measure for certain cases such as urban settings, the rectangular distance is often used as a first approximation in more complex location models.

Several modifications or generalizations of the basic model in (1.1) have been proposed, some of which are considered below.

a) Up till now the customers are represented as points in space. Witzgall (1964) formulates a two-dimensional model in which they can be either point or area demands. The latter should be considered when the number of customers in a specified region is sufficiently large that the demand here can be accurately approximated by a density function. Such a condition occurs for example with postal deliveries in a city or suburb. Love (1972) considers the case where demands are over rectangular regions and the distance measure is Euclidean; with rectangular distances this problem can be solved exactly (Wesolowsky and Love, 1971a). Drezner and Wesolowsky (1980) extend the analysis to ℓ_p distances, and to circles and other general shapes. The solution technique they use is an iterative one based on the Weiszfeld procedure. It is also interesting to note that the new facility, which is currently modelled as an unknown point x, may actually have a significant area (or volume), as in the location of a parking lot to service a set of buildings.

b) Problem (1.1) assumes that the same distance function is associated with every fixed location. However, practical situations may arise where the distance to each customer is more accurately measured by a different function; for example, when travel modes (such as land or air) are not the same for each. This results in a more general problem:

$$\text{minimize } W_m(x) = \sum_{i=1}^{n} w_i d_i (x, a_i) , \tag{1.2a}$$

where d_i is the distance measure associated with destination $a_i, i=1,...,n$. When the d_i are all norms, a well-known class of functions which we shall discuss in the next section, this is called a mixed-norm problem. Hansen, Perreur and Thisse (1980) develop some general properties for such a case. Alternatively, we may have more than one distance function pertaining to each destination; for example, when different products are sent to each customer by different travel modes. This variation is formulated as follows:

$$\text{minimize } W_m(x) = \sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij} d_j (x, a_i) , \tag{1.2b}$$

where L denotes the number of different travel modes, and $w_{ij} \geq 0$ for all i and j, is the appropriate weighting constant for customer i using distance function d_j. Problem (1.2b) is examined by Planchart and Hurter (1975) when $L=2$ and the d_j are the rectangular and Euclidean distances (norms). Note that problem (1.2a) is a special case of (1.2b), in which $w_{ij} = 0$ for all j except one, for $i=1,...,n$.

c) A stochastic extension of the Fermat-Weber problem is considered by Cooper (1974), in which the destinations are no longer predetermined points a_i, but random variables with given probability distributions. The objective in this case becomes the minimization of the sum of weighted expected values of the distances from the source to the a_i. Cooper's model uses the Euclidean norm as the distance function associated with each customer. The same stochastic model is considered by Wesolowsky (1977) with rectangular distances. Problem (1.1) also assumes that the customer demands are deterministic; but in reality, these demands often have a random nature. Aly and White (1978) incorporate this feature in the stochastic location model by considering the weights w_i to be random variables.

d) The Fermat-Weber problem can be viewed as a static model, since the number of customers and their locations and demands are assumed to be constant over a long (infinite) time horizon. Wesolowsky (1973) looks at the dynamic facility location problem, in which these exogenous factors are permitted to change in each period. He employs a dynamic programming algorithm to optimize the sequence of locations of the new facility over a finite planning horizon. Other approaches are discussed by Erlenkotter (1981).

e) Thus far, the new facility can be located anywhere in the geographic space, since the problem is unconstrained. In practical situations, natural and human factors (e.g., land barriers, zoning regulations) tend to restrict, sometimes drastically, the set of feasible locations and routes. Hansen, Peeters and Thisse (1982) allow a very general feasible region in the form of a union of a finite number of convex polygons, and develop an algorithm to solve this constrained problem. Love (1969) uses a gradient method for locating a facility within a convex subset of three-dimensional space. Eckhardt (1975) proposes an amended Weiszfeld procedure when the set of feasible locations is defined by a convex polyhedron in N-dimensional space. Schaefer and Hurter (1974) consider the case where the new facility is constrained to be within a maximum distance of each demand point. Such restrictions, referred to as metric constraints, are of interest when locating services such as police and postal stations. Properties of the optimal solution for several types of constrained problems are given by Hurter, Schaefer and Wendell (1975).

f) Problem (1.1) concerns the location of a single facility. An obvious and important extension includes the multifacility case, in which two or more new depots must be located simultaneously. Each depot can interact with the existing destinations, as well as with the other depots, and all interactions are assumed to be known. Clearly, when no flow exists between pairs of new facilities, the problem reduces to a number of single facility models. A formulation of the unconstrained multifacility model is given below:

$$\text{minimize } WM(X) = \sum_{j=1}^{m} \sum_{i=1}^{n} w_{ij} d(x_j, a_i) + \sum_{r<s}\sum v_{rs} d(x_r, x_s) , \tag{1.3}$$

where m is the number of new depots to be located; n is the number of existing destination points; $X = (x_1,...,x_m)$, where $x_j = (x_{j1},...,x_{jN})^T$ is the unknown location of depot j for $j = 1,...,m$; $a_i = (a_{i1},...,a_{iN})^T$ is the known location of destination i, $i = 1,...,n$; $w_{ij} \geq 0$ is a weighting constant which converts distance between an origin-destination pair into a cost, $i = 1,...,n$, $j = 1,...,m$; $v_{rs} \geq 0$ is a weighting constant which converts distance between an origin-origin pair into a cost, $r = 1,...,m-1$, $s = r+1,...,m$; and d is the distance function.

Ostresh (1977) extends the Weiszfeld procedure to the multifacility case with Euclidean distances, and proves descent properties of the iterations and convergence to the optimal solution. However, his procedure cannot handle vertex iterates (i.e., facilities which coincide). Radó (1988) proves global convergence of a modified version of this algorithm. A new approach is presented by Drezner and Wesolowsky (1978a), in which a trajectory of optimal solutions for a series of perturbed problems is obtained by numerical integration of a set of differential equations. The trajectory begins with a perturbed problem which can be solved easily, and ends with the solution to the original problem.

When d is the rectangular norm, problem (1.3) can be solved by linear programming and related techniques; see Cabot, Francis and Stary (1970), Picard and Ratliff (1978), and Wesolowsky and Love (1971b). However, the number of constraints and variables increases considerably with the problem size. Wesolowsky and Love (1972) use convex programming with a hyperbolic approximation of the rectangular norm, while Juel and Love

(1976) propose an 'edge-descent' algorithm which takes advantage of the convex, piecewise planar shape of the objective function.

A dual problem can be formulated for the single and multifacility models, which provides an alternative method of solution. Kuhn and Kuenne (1962) and Bellman (1965) derive the dual for the single facility Euclidean distance model. A dual for the multifacility Euclidean case is given by Francis and Cabot (1972), while Love and Kraemer (1973) propose a dual decomposition method of solution. A generalization to ℓ_p distances is provided by Love (1974).

g) An important assumption of the multifacility location model is that the interactions between pairs of facilities are all known. In many practical situations, determining these interactions is a main feature of the problem. For example, in locating several warehouses to meet the demands of a set of customers, one generally must determine the best allocation of customers to warehouses as well as the optimal locations of the latter. The optimal number of facilities to service the customers is also usually unknown; however, this can be determined by repeated solution of the location-allocation problem for increasing numbers of new facilities. A formulation of the location-allocation problem without capacity constraints is given below:

$$\text{minimize } \phi = \sum_{j=1}^{m} \sum_{i=1}^{n} w_{ij} d(x_j, a_i) , \tag{1.4}$$

$$\text{subject to } \sum_{j=1}^{m} w_{ij} = w_i, \; i = 1, ..., n .$$

Now the weights w_{ij} representing quantities or flow between facility j and destination i are variables, in addition to the unknown locations x_j of the m facilities. The constraints in problem (1.4) ensure that the demands of each customer (w_i) are satisifed.

Unfortunately, the objective function ϕ has a complex shape which is not amenable to solution by standard methods. With d as the Euclidean norm, Cooper (1967) proves that ϕ is neither convex nor concave. This result generally holds for any distance metric (e.g., see Chapter 7 of Love, Morris and Wesolowsky, 1988); so that ϕ can have several local minima. A rather dramatic illustration is found in Eilon, Watson-Gandy and Christofides (1971). For a 50 customer, 5 depot problem ($n=50$, $m=5$), they obtain 61 local optima by using different initial starting locations for the depots and an adaptive location-allocation heuristic. Several heuristic methods are proposed by Cooper (1963, 1967, 1972). Love (1976) solves the one-dimensional version exactly using dynamic programming. However, except for some special cases, this method cannot be extended efficiently to higher-dimensional spaces. Heuristic methods which attempt to 'jump' over local optima are given by Love and Juel (1982), who also show that the location-allocation problem can be expressed as a concave minimization program. Such programs involve the minimization of a concave function over a compact region, so that the search for an optimal solution can be restricted to the boundary of the feasible region.

An exact solution method is given by Love and Morris (1975a) when distances are rectangular. This procedure uses the property that an optimal solution exists with the new facilities located at discrete intersection points. An algorithm further reduces this candidate set of points. The conditioned problem is then solved exactly with a backtrack programming procedure. Ostresh (1975) considers the two-center problem ($m = 2$) with Euclidean distances. Cavalier and Sherali (1986) examine Euclidean distance location-allocation with uniform demands over convex polygons. A large-scale nonlinear programming approach is used by Murtagh and Niwattisyawong (1982), but this can only guarantee a local optimum dependent on the initial starting locations.

The heuristic methods generally involve the solution of a large number of location problems as the allocations are varied. The use of rectangular distances affords a considerable computational advantage, because of the relative ease of solving the location problems with this metric. For a sample of test problems and computational results for the location-allocation problem with rectangular distances, see Love and Juel (1975).

Physical location problems occur in one, two or three-dimensional space. However, it is interesting to note that location theory is now being applied in areas other than physical distribution systems. Examples include cluster analysis (Cooper 1973), and the spatial analysis of voters' preferences (Riker and Ordeshook, 1973) and customers' preferences in product design problems (Schocker and Srinivasan, 1974). Such cases justify the use of higher-dimensional spaces.

As a final comment, we note that the extensions discussed above to the Fermat-Weber problem, as well as other extensions not included here, give greater flexibility to the original model by allowing different types of cost structures to be estimated more accurately. These extensions can be applied singly or in combinations depending on the practical problem at hand. As an example, one may wish to consider a stochastic location-allocation model with mixed norms to represent a real situation. Of course there is the usual tradeoff; -- the more accurate the model, the more difficult and costly it is generally to solve.

1.2 Distance Functions

The purpose of a distance function d is to give an accurate measure of the separation between any two points in space. In physical location problems, this separation normally signifies the shortest travel distance between pairs of points in the transportation network. Thus, given two points $x,y \in R^N$, the function $d(x,y)$ calculates a distance value. In the most general sense, we see that d represents a mapping of the form,

$$d : R^N \times R^N \rightarrow R \,, \tag{1.2.1}$$

where $R^N \times R^N$ denotes all pairs of points taken from N-dimensional space, with each such pair being assigned by d a value on the real number line.

Herein lies the main difference between continuous location models and network models. The latter type uses actual distances between pairs of points, and by necessity restricts the candidate set of locations of the new facilities to the arcs and nodes forming the transportation network. In realistic problems, the network can contain a large number of nodes and connecting arcs (road segments), so that data storage requirements become excessive. The existence of cycles in the graph requires that a shortest path algorithm be used to find the shortest distance between pairs of nodes. These features plus long computation times for large graphs tend to make network models more cumbersome and expensive. However, because actual distances are used, these models can be made to represent real systems more accurately than the continuous models. (This is achieved by augmenting the number of nodes in the network where necessary). As noted by Francis, McGinnis and White (1983), "you get what you pay for".

Continuous location models, on the other hand, require very little data storage, since distances between pairs of points are now calculated from the coordinates. Of course, these distances only approximate the actual ones. Continuous models are typically easier to analyze. They give useful qualitative insights about the system, which can be used to simplify the network in a second stage of analysis; for example, by removing from consideration a majority of the candidate solution points. Furthermore, there are situations where the continuous location models are more appropriate in their own right; for example, when the set of demand points or candidate solution points includes regions of the space such as line segments, areas or volumes, or when the transportation network has a regular pattern such as a rectangular grid which can be represented with a high precision by some function d.

A distance function should satisfy the following basic properties, where p, q, r are any points in R^N.

$$d(p,q) \geq 0 \ , \tag{1.2.2}$$

$$d(p,q) = 0 \quad \text{if } p=q \ , \tag{1.2.3}$$

and

$$d(p,r) \leq d(p,q) + d(q,r) \ . \tag{1.2.4}$$

The first two conditions are intuitively obvious. The third one, known as the triangle inequality, tells us that the distance between any two points is the shortest one from all possible paths. If d satisfies relations (1.2.2) to (1.2.4), it is termed a weak metric (e.g., see Witzgall, 1964). If in addition d satisfies

$$d(p,q) = 0 \quad \text{implies} \quad p = q \ , \tag{1.2.5}$$

and

$$d(p,q) = d(q,p) \ , \tag{1.2.6}$$

then it is called a metric. Relation (1.2.6) implies a symmetry in the system being modelled, such that the distance from p to q equals the distance from q to p for all pairs of points. However, this condition will not apply when nonsymmetric shipment costs exist, as with travel up and down streams or inclined planes, and travel along one-way streets in an urban area. It should be noted that most location models, including those discussed in the previous section, assume relation (1.2.6) holds; so that d(p,q) is interpreted as the distance between p and q.

Most distance functions employed in continuous location problems belong to the family of norms. A function k is said to be a norm if it satisfies the following properties for any $x, y \in R^N$.

$$k(x) \geq 0 \ , \tag{1.2.7}$$

$$k(x) = 0 \quad \text{if, and only if,} \quad x = 0 \ , \tag{1.2.8}$$

$$k(\alpha x) = |\alpha| k(x) \ , \quad \forall \alpha \in R \ , \tag{1.2.9}$$

and

$$k(x+y) \leq k(x) + k(y) \ . \tag{1.2.10}$$

The value k(x) denotes the distance between the point x and the origin $(0=(0,...,0)^T)$. A norm may be employed to define a metric by setting $d(x,y) = k(x-y)$. However, the converse is not necessarily true, since norms are homogeneous by relation (1.2.9) while metrics do not have to be. A norm provides the following map,

$$k : R^N \rightarrow R \ , \tag{1.2.11}$$

which is clearly restrictive when compared with (1.2.1). Now the distance between points x and y depends in no way on their absolute positions, but only on the vector $(x-y)$ joining them.

A popular distance function in location models is given by the ℓ_p norm, which we define as follows for N-dimensional space:

$$\ell_p(x) = \left[\sum_{i=1}^{N} |x_i|^p \right]^{1/p} , p \geq 1 \ , \tag{1.2.12}$$

where once again $x=(x_1,...,x_N)^T$. The rectangular distance is a special case with $p=1$, while Euclidean distance occurs with $p=2$. Properties and applications of the ℓ_p norm will be of major interest in subsequent chapters. When p is strictly greater than 1, $\ell_p(x)$ belongs to the family of 'round' norms, so-named because their contours contain no flat spots. We shall be examining this important class of norms in detail later on.

Other examples include the weighted one-infinity norm of Ward and Wendell (1980), which is defined as follows:

$$b_\ell(x) = \alpha_1 \ell_1(x) + \alpha_2 \sqrt{2} \ \ell_\infty(x) \ , \tag{1.2.13}$$

where α_1 and α_2 are nonnegative numbers, not both of which are zero, ℓ_1 is the rectangular norm, and ℓ_∞ the Chebychev norm defined as

$$\ell_\infty(x) = \lim_{p \rightarrow +\infty} \ell_p(x) = \max_{1 \leq i \leq N} \{|x_i|\} \ . \tag{1.2.14}$$

The distance function b_ℓ and its components ℓ_1 and ℓ_∞ belong to a general class known as 'block' norms (e.g., see Thisse, Ward and Wendell, 1984, and Ward and Wendell, 1985), so-named because their contours are made up of flat segments; that is, the contours form polytopes in N-dimensional space. The block norm is characterized as follows:

$$b(x) = \min \left\{ \sum_{g=1}^{r} |\beta_g| : x = \sum_{g=1}^{r} \beta_g b_g \right\} . \tag{1.2.15}$$

In the above relation the b_g's with $g = \pm 1, \pm 2, \ldots, \pm r$, are vectors which define the extreme points of the unit contour (or polytope) of b(x), and by symmetry $-b_g = b_{-g}$. The block norm has a geometric interpretation. The b_g's signify the possible directions of travel in the transportation network, and b(x) gives the distance of the shortest path connecting the tail and head of the vector x which follows these permissible directions. A detailed account of the block norm, with proofs of properties and examples, can be found in the references mentioned above. Note also that Thisse, Ward and Wendell (1984) compare some of the properties of block and round norms, and show that a block norm can be made to represent a round one as accurately as desired by increasing the number of extreme points (b_g's) of its polytope. In the limiting sense then, as $r \to \infty$, the block norm becomes a round one.

Other distance functions are found in the literature. Perreur and Thisse (1974) propose the radial and circumferential metrics, and a combination of these two, the circum-radial metric, for approximating star-shaped networks (e.g., the French railway system) and circumferential transportation systems (e.g., ringroads around towns). These functions are called central metrics, as the movement in each case is made partly or completely along rays through an origin. Hodgson, Wong and Honsaker (1987) derive an asymmetric distance (or cost) function for locating facilities on an inclined plane. In this case travel up the slope is more difficult than down, so that the cost of travelling from one point to another is not the same as the cost in the reverse direction. The authors formulate a minisum model to determine the optimal location of prebunching sites on a slope in the logging industry, and develop a Weiszfeld-type iterative solution procedure, proving convergence of their algorithm to the optimal site. When the demand points in a system cover a sufficiently large area, the radius of curvature of the earth's surface can no longer be neglected. Certain distance functions have been proposed for the analysis of location problems on spherical surfaces (e.g., see Aly, Kay and Litwhiler, 1979, Drezner and Wesolowsky, 1978b, and Love and Morris, 1972).

For networks in which backtracking occurs on a regular basis, such as when rectangular floor layouts have a single doorway accessing each department, Juel and Love (1985) propose the use of the hyper-rectilinear distance function. This measure corresponds to the ℓ_p function with $0<p<1$, and occurs in practice when travel distances are generally greater than rectilinear. When $0<p<1$, the ℓ_p distance function is not a norm, and furthermore, it is neither convex nor concave. The authors show that for the hyper-rectilinear case, an optimal solution of the single facility minisum model must occur at an intersection point, which may not be in the convex hull of the existing facilities.

A significant research effort has been directed at the problem of characterizing the optimal solution of the minisum model, the objective being to simplify or narrow the search for this solution by exploiting the properties of the particular distance function. The general approaches taken here can be classified into two categories. The first reduces the set of feasible solutions to a region characterized by the geometry of the existing facilities. The second uses the weight structure or flows between facilities to make deductions on where the optimal location must be.

In the first category, Kuhn (1967) proves that the optimal solution of the Fermat-Weber problem with Euclidean distances on R^2 lies in the convex hull of the existing facilities. This result readily extends to N-dimensional space (Kuhn, 1973). Francis and Cabot (1972) examine the multifacility model with Euclidean distances on R^2, and show that the optimal locations of all the new facilities must be within this same hull. For the more general single and multifacility problems with ℓ_p distances on R^2 and a $p > 1$, Juel and Love (1983) prove once again that the optimal locations must be within the convex hull of the existing facilities.

With any norm on two-dimensional space, at least one optimal solution to the Fermat-Weber problem belongs to the convex hull of the fixed points (Wendell and Hurter, 1973). It should however be stressed that this property no longer holds in higher-dimensional spaces, except when Euclidean distances are used (Plastria, 1984). For the special case of the ℓ_1 norm on R^2, Love and Morris (1975a) prove that at least one optimal solution belongs to a smaller rectangular hull of the existing facilities, and furthermore, attention can be restricted to the intersection points contained in this hull. Hansen, Perreur and Thisse (1980) define an octagonal hull of the existing facilities (which is larger than the convex hull), and show that when a mixed-norm problem on R^2 involves ℓ_p distance functions only, then at least one optimal location belongs to this hull. In a more recent paper, Durier and Michelot (1985) define a metric hull in order to account for nonsymmetric distance measures, which are termed gauges. They show that an optimal location in N-dimensional space can be found in the metric hull of the existing facilities whatever the gauges are.

The second category mentioned above uses the weight structure or flows between facilities to make deductions on where the optimal location should be. Such an approach can result in considerable computational savings in practical situations. An early contribution in this vein is given by Witzgall (1964), who proves the "majority theorem" for single facility location. This states that in the Fermat-Weber model, an existing facility having 50% or more of the total interaction is an optimal base. A refinement of this result is contained in the fixed point optimality criteria proposed by Juel and Love (1981),which apply to the single facility location problem with any set of norms. For the multifacility case, criteria are given by Juel (1983), and Juel and Love (1980), for establishing when facility locations coincide in an optimal solution.

1.3 Empirical Work Related to Distance Functions:

The major research efforts in continuous location theory have been, and are still, in the development of algorithms to solve location problems such as those discussed briefly above. Thus, the researcher normally begins with the assumption that the cost structure and distance function(s) used in his model are a sufficiently accurate representation of some real system. From a practical viewpoint, an accurate measure of distance in the real system is an important requisite to finding an optimal solution. No matter how exact and efficient the solution algorithm may be, the end result would be of questionable value unless the model is an accurate representation of the problem being analyzed.

Very little work in the literature deals with the empirical fitting of distance functions to actual data, although this is clearly a topic of crucial interest in continuous location models. Love and Morris (1972, 1979) present several distance functions which are, for the most part, norms multiplied (weighted) by an inflation factor that helps to account for hills, bends and other forms of 'noise' in the transportation network. They carry out an empirical study in which the best-fitting parameter values are obtained for sets of data from urban and rural regions. An important finding of their study is that an empirical distance function should be tailored to a given region whenever a premium is placed on accuracy. This conclusion resulted from the observed statistical superiority of the weighted ℓ_p norm over the weighted rectangular and Euclidean norms. Thus, the claim by Francis (1967), which, by the way, has been assumed by the majority of researchers in continuous location theory, that the cases of practical interest are the ones where distances are rectangular or Euclidean, is refuted by the findings of Love and Morris.

Other empirical studies are described below. Berens and Körling (1985) examine the road network of the Federal Republic of Germany, and conclude that the weighted Euclidean norm is sufficiently accurate in this case. They further propose that fitting distance functions with two or more parameters is generally unwarranted, since in their opinion the gain in accuracy will be small while the computational work will increase substantially. The conclusions of Berens and Körling are refuted by Love and Morris (1988). Certainly the previous work of Love and Morris (1972, 1979) shows that significant gains in accuracy can be achieved with the weighted ℓ_p norm. The method used by Love and Morris (and Berens and Körling) to find the best-fitting parameter values involves an exhaustive grid search. Thus the number of iterations tends to increase exponentially with the number of unknown parameters, so that the computational work can indeed become rather cumbersome.

Ward and Wendell (1980) fit the weighted one-infinity norm to two sets of data of intercity highway travel used by Love and Morris (1972) for the weighted ℓ_p norm, and observe that their distance function is relatively close in accuracy to the latter. Further empirical work is done by Ward and Wendell (1985) using the general block norm, and the data sets of Love and Morris (1979). Since block norms are linear in their parameters, Ward and Wendell are able to apply standard linear regression techniques to find the best-fitting values.

Another empirical study is carried out by Kolesar, Walker and Hausner (1975), in which travel times are of primary interest. They show that the relation of travel time to distance for fire engines in New York City is nonlinear with economies of scale. Love and Dowling (1985) study the fit of weighted ℓ_p functions in facility layout problems with rectangular flow patterns, and observe that the accuracy of the fit is more sensitive to changes in the inflation factor than to changes in the parameter p.

In order to account for economies or dis-economies of scale inherent in the structure of the transportation network, Love and Morris (1972, 1979) propose a distance function having the ℓ_p norm raised to a power. This results in an extra parameter, and consequently, a better fit to the data. The authors generally observe economies of scale in the road networks examined, although a few cases show dis-economies. These results make intuitive sense. One would usually expect an economy of scale, arising from the larger number of routes available when points are further apart.

It is important to note that the location problem may be more accurately modelled with distance functions raised to a power for purposes other than the particular structure of the road network. That is, economies (dis-economies) of scale may exist for totally different reasons. As an example, one might use different transportation modes depending on the distance between points, which would give rise to nonlinearities in the cost structure. It is generally assumed in such cases that cost is a non-decreasing function of distance. Thus an extension of the Fermat-Weber problem, with distances raised to a power, takes on practical significance. A formulation of this problem is given below:

$$\text{minimize } \; WNL(x) = \sum_{i=1}^{n} w_i \, [d(x, a_i)]^K \, , \tag{1.3.1}$$

where $0<K<1$ for economies of scale, $K = 1$ for continuous returns to scale (the original Weber problem), and $K > 1$ for dis-economies of scale.

Relatively little work can be found in the literature pertaining to problem (1.3.1). An early formulation of this model is given by Cooper (1968), in which d is the Euclidean norm. He develops an iterative solution technique similar to the Weiszfeld procedure. Chen (1984a,b) improves the efficiency and convergence properties of Cooper's algorithm by changing the step-size. He also investigates a more general class of problems in which the cost components can be expressed as non-decreasing functions of the Euclidean distances.

Morris (1981) extends the iterative solution technique to the case where d is the ℓ_p norm, and proves convergence for certain ranges of the parameters. He also shows that the objective function in problem (1.3.1), with d equal to the ℓ_p norm, is neither convex nor concave when $0 < K < 1$, and furthermore, contains several local minima. For this condition the iterative solution procedure may converge to a local minimum which is not the global optimum.

Hansen, Peeters, Richard and Thisse (1985) present a very general algorithm for solving the single facility location problem in which transportation costs are increasing and continuous functions of distance. They call their algorithm Big Square - Small Square. The general idea is to divide the set of feasible solutions in the plane into squares; to calculate bounds for each square by taking the shortest distance between the square and each demand point; to purge those squares whose bounds are no better than the current incumbent solution; and to continue branching to smaller sub-squares and bounding until the length of a side of a square is smaller than a given tolerance. The authors report good computational experience with their algorithm, which is encouraging considering the applicability of their method to general cost structures and general sets of constraints.

1.4 Thesis Objectives

The preceding sections give some insight into the broad nature of continuous location theory. We started with the well-known Fermat-Weber problem, and then discussed several extensions to this classical model. The importance of the distance function in continuous location models was emphasized. We observed that this distance measure should satisfy the properties of the 'weak' metric, a wide class of functions of which the norms are only a subset. Some of the more popular distance functions (e.g., ℓ_p norms, block norms) which appear in the literature were presented. Finally, we summarized the empirical work dealing with the fitting of distance functions to actual transportation networks. The empirical work invariably begins by assuming a given form of the distance measure, and then proceeds to obtain the best-fitting parameter values based on some specified criteria.

In the next chapter, several general mathematical properties of norms are derived, which will be useful in the subsequent analysis. We begin by taking a look at sums of order p, which represent the function $\ell_p(x)$ with the vector x constant and the parameter p treated as a variable. The results obtained here will be of interest when the problem of fitting the ℓ_p norm to a given data set is investigated. A practical classification of norms is presented next, and several properties are derived for this classification scheme. These results are in large part generalizations of known properties for particular norms. Some insights into the fundamental differences between round and block norms are also provided. The properties given here will be useful in our investigation of various minisum models.

The fitting of empirical distance functions was identified in the previous section as a topic of theoretical and practical importance which requires much further research. In Chapter 3 we consider this problem in terms of the mathematical aspects of fitting the weighted ℓ_p norm. At present an exhaustive grid search is employed to find the best-fitting parameter values; e.g., see Love and Morris, 1972, 1979, Love, Truscott and Walker, 1985, and Berens and Körling, 1985. A number of important properties are derived here which will allow more efficient and more accurate searches for these values. In Chapter 4 a positive linear combination of Euclidean and rectangular distances is considered, which we term the weighted one-two norm. It is shown that for practical purposes, this distance measure can be used in place of the weighted ℓ_p norm. Since the weighted one-two norm is linear in its parameters, we are able to develop a simple linear regression model for determining the best-fitting parameter values. Statistical tests are proposed for this model which provide new insights of practical significance.

The remainder of our study is devoted to a broad sample of minisum models. In Chapter 5 we return to the classical Fermat-Weber problem, and generalize certain important properties to the classes of round and block norms. Some generalized results for the multifacility model are also derived. A close look is then taken at the Weiszfeld iterative solution procedure. After much analysis, we extend the global convergence proof of Kuhn (1973) and the local convergence results of Katz (1974) for Euclidean distances to the ℓ_p norm. Chapter 6 investigates the mixed-norm model. An extension of the Weiszfeld procedure is proposed as a solution method, and global and local convergence properties are proven. Optimality criteria which extend the majority theorem of Witzgall (1964) and the results of Juel and Love (1981), are also derived. Finally, in Chapter 7 we consider the minisum problem with cost components which are nonlinear functions of distance (e.g., see model (1.3.1)), and obtain some general properties for this case.

CHAPTER 2

GENERAL MATHEMATICAL RESULTS

In this chapter, we derive several general results which are interesting in their own right, and of which many will be useful for developing properties of location problems in the subsequent chapters. First, we take a close look at sums of order p and the well-known inequality of Jensen. The results here will be of interest when fitting the ℓ_p distance function to a data set is investigated (Chapter 3). We also study the properties of a generalized sum of order p, which would be applicable to a generalization of the ℓ_p distance function. The next section deals with an important class of norms referred to as round norms. Some definitions and properties are developed here, and comparisons are made with the block norm. These results should improve our understanding of distance functions, and will be useful in our analysis of minisum location models. The third section considers functions of norms, and in particular, norms raised to a power. Such distance (or cost) functions have received comparatively little attention in the literature, although their potential benefit in defining more accurate location models has been recognized. Finally, we study directional derivatives and the differentiability of norms and functions of norms. The results obtained here will be useful later on in our analysis of various minisum models.

2.1 A Generalization of Jensen's Inequality

2.1.1 Jensen's Inequality Revisited

A sum of order p is defined as follows (e.g., section 1-16 of Beckenbach and Bellman, 1965):

$$S(y;p) = \left(\sum_{i=1}^{K} y_i^p \right)^{1/p} , \tag{2.1.1}$$

where $y = (y_1, \ldots, y_K)^T$,

$y_i > 0, i=1, \ldots, K$, are any positive values,

and $p \neq 0$ is a real-valued parameter.

Note that the above sum has the form of the distance function $\ell_p(x-a_r)$, where y_i replaces $|x_i - a_{ri}|, i=1, \ldots, K$. The requirement that all the y_i's be strictly positive (i.e., non-zero) is not restrictive, since any zero terms can be deleted from the sum, the remaining terms re-labeled and K decreased accordingly, to give the form in (2.1.1).

The sum of order p satisfies the well-known relation,

$$S(y\,;p_2) < S(y\,;p_1)\,, \quad 0 < p_1 < p_2\,, \quad K \geq 2\,, \tag{2.1.2}$$

which is usually referred to as Jensen's inequality. (For two different proofs of (2.1.2), see Theorem 19 of Hardy, Littlewood and Pólya, 1952, and Beckenbach, 1946.) Beckenbach (1946) also shows that $S(y\,;p)$ is convex in p for $p>0$. His proof utilizes techniques from convex analysis.

We now re-prove Jensen's inequality and the convexity result of Beckenbach by studying the first and second-order partial derivatives of $S(y\,;p)$ with respect to p. Although this approach is less elegant than previous methods, and certainly very tedious, it does enable us to prove in addition that $S(y\,;p)$ is strictly convex in p for $p>0$ under very general conditions, namely $K \geq 2$. We also are able to make some deductions concerning inflection points for $p<0$. An extension of this approach allows us to determine some interesting properties of a 'generalized' sum of order p.

For the simple case where $y = y_1$, a scalar,

$$S(y\,;p) = y_1\,, \ (K=1)\,, \tag{2.1.3}$$

which is constant for varying values of the parameter p. We consider from this point onwards the more interesting case in which $K \geq 2$. The first and second-order partial derivatives of $S(y;p)$ with respect to p are calculated below.

a) **First Derivative**

Let

$$\phi(y;p) := \sum_{i=1}^{K} y_i^p , \tag{2.1.4}$$

so that

$$S(y;p) = [\phi(y;p)]^{1/p} . \tag{2.1.5}$$

Then

$$\begin{aligned}
\frac{\partial}{\partial p} S(y;p) &= \phi^{1/p}\left(-\frac{\ell n\phi}{p^2} + \frac{1}{p\phi}\frac{\partial\phi}{\partial p}\right) \\
&= -\phi^{1/p}\frac{\ell n\phi}{p^2} + \frac{\phi^{\frac{1}{p}-1}}{p}\sum_{i=1}^{K} y_i^p \ell n\, y_i \\
&= \frac{\phi^{(1-p)/p}}{p^2}\left(-\phi\ell n\,\phi + \sum_{i=1}^{K} y_i^p \ell n\; y_i^p\right) \\
&= \frac{\phi^{(1-p)/p}}{p^2}\left(\sum_{i=1}^{K} y_i^p \ell n\left[\frac{y_i^p}{\phi}\right]\right), p \neq 0 .
\end{aligned} \tag{2.1.6}$$

Since

$$0 < \frac{y_i^p}{\phi} = \frac{y_i^p}{\sum_{j=1}^{K} y_j^p} < 1, \quad i=1,\ldots,K,$$

therefore

$$\ell n\left[\frac{y_i^p}{\phi}\right] < 0 , \quad i=1,\ldots,K .$$

Hence

$$\frac{\partial}{\partial p} S(y\,;p) < 0 \ , \tag{2.1.7}$$

wherever this derivative is defined (i.e., $p \neq 0$). At $p=0$, the function $S(y\,;\,p)$ is discontinuous as seen by the following limits:

$$\lim_{p \to 0^-} S(y\,;p) = 0 \ , \tag{2.1.8}$$

and

$$\lim_{p \to 0^+} S(y\,;p) = \infty \ . \tag{2.1.9}$$

Otherwise, $S(y\,;\,p)$ is continuous and differentiable in p. Based on these observations and the inequality (2.1.7), the following result is obtained.

Property 2.1.1

The sum of order p given in (2.1.1) with $K \geq 2$ is a decreasing function of p for $0<p<\infty$; that is

$$S(y\,;p_2) < S(y\,;p_1) \ , \ 0 < p_1 < p_2 \ . \tag{2.1.2}$$

The sum of order p given in (2.1.1) with $K \geq 2$ is also a decreasing function of p in the interval $-\infty < p < 0$: that is,

$$S(y\,;p_2) > S(y\,;p_1) \ , \ p_2 < p_1 < 0 \ . \tag{2.1.10}$$

(Note: Inequality (2.1.10) can be derived directly from (2.1.2), as shown by Beckenbach (1946), by noting that $S(y\,;\,-p) = 1/S(1/y, p)$, where $1/y = (1/y_1, \ldots, 1/y_k)^T$).

b) Second Derivative

Letting

$$\Phi(y\,;p) := \sum_{i=1}^{K} y_i^p \,\ell n \left[\frac{y_i^p}{\phi}\right] < 0 \ , \tag{2.1.11}$$

we can rewrite equation (2.1.6) as

$$\frac{\partial}{\partial p} S(y\,;p) = \frac{\phi^{(1-p)/p}}{p^2} \,\Phi \ . \tag{2.1.12}$$

The second partial derivative with respect to p is then given by,

$$\frac{\partial^2}{\partial p^2} S(y\,;p) = \frac{-2}{p^3}\phi^{\frac{1-p}{p}}\Phi + \frac{\Phi}{p^2}\frac{\partial}{\partial p}\left(\phi^{\frac{1-p}{p}}\right) + \frac{\phi^{\frac{1-p}{p}}}{p^2}\frac{\partial \Phi}{\partial p}\,. \tag{2.1.13}$$

Using elementary calculus, the following results are obtained.

$$\frac{\partial}{\partial p}\left(\phi^{\frac{1-p}{p}}\right) = \frac{\phi^{\frac{1-2p}{p}}}{p^2}(\Phi - p^2\phi')\,, \tag{2.1.14}$$

and

$$\frac{\partial \Phi}{\partial p} = p\phi'' - \phi'\ell n\phi\,, \tag{2.1.15}$$

where

$$\phi' := \frac{\partial}{\partial p}\phi(y\,;p) = \sum_{i=1}^{K} y_i^p\,\ell n\, y_i\,, \tag{2.1.16a}$$

and

$$\phi'' := \frac{\partial^2}{\partial p^2}\phi(y\,;p) = \sum_{i=1}^{K} y_i^p(\ell n\, y_i)^2\,. \tag{2.1.16b}$$

Substituting (2.1.14) and (2.1.15) into equation (2.1.13) gives

$$\frac{\partial^2}{\partial p^2} S(y\,;p) = \frac{1}{p^2}\phi^{\frac{1-2p}{p}}\left[-\frac{2}{p}\phi\Phi + \frac{1}{p^2}\Phi^2 - \Phi\phi' + p\phi\phi'' - \phi\phi'\,\ell n\,\phi\right] \tag{2.1.17}$$

Noting that $\Phi(y\,;p) = p\phi' - \phi\,\ell n\,\phi$, the above equation simplifies to,

$$\frac{\partial^2}{\partial p^2} S(y\,;p) = \frac{1}{p^2}\phi^{\frac{1-2p}{p}}\left[-\frac{2}{p}\phi\Phi + \frac{1}{p^2}\Phi^2 - p(\phi')^2 + p\phi\phi''\right]. \tag{2.1.18}$$

Thus,

$$\frac{\partial^2}{\partial p^2} S(y\,;p) = A(y\,;p)\cdot B(y\,;p)\,,$$

where

$$A(y\,;p) = \frac{1}{p^2}\phi^{\frac{1-2p}{p}}\,,$$

and

$$B(y\,;p) = -\frac{2}{p}\,\phi\Phi + \frac{1}{p^2}\,\Phi^2 - p(\phi')^2 + p\phi\phi'' .$$

Since $\phi(y\,;p) > 0$ for all $p \neq 0$, it is obvious that $A(y\,;p) > 0$ for all $p \neq 0$. We now examine the terms of $B(y\,;p)$. From the analysis of the first derivative, $\partial S/\partial p$, it is clear that $\Phi(y\,;p) < 0$ for all $p \neq 0$. Hence,

$$-\frac{2}{p}\,\phi\Phi \begin{cases} > 0 & \text{for} \quad p > 0\,, \\ < 0 & \text{for} \quad p < 0\,. \end{cases} \tag{2.1.19}$$

The second term of $B(y\,;p)$,

$$\frac{1}{p^2}\,\Phi^2 > 0,\ \forall\, p \neq 0\ . \tag{2.1.20}$$

Finally, we add the third and fourth terms of $B(y\,;p)$ as follows:

$$\begin{aligned}
-p(\phi')^2 + p\phi\phi'' &= p(\phi\phi'' - \phi'^2) \\
&= p\Big[\sum_i y_i^p \sum_j y_j^p(\ell n\, y_j)^2 - \Big(\sum_i y_i^p\, \ell n\, y_i\Big)^2\Big] \\
&= p\Big[\sum_i \sum_j y_i^p y_j^p \{(\ell n\, y_j)^2 - \ell n\, y_i\, \ell n\, y_j\}\Big] \\
&= p\Big[\sum_i \sum_j y_i^p y_j^p\, \ell n\, y_j\,(\ell n\, y_j - \ell n\, y_i)\Big] \\
&= p\Big[\sum\sum_{i<j} y_i^p y_j^p\, \ell n\, y_j\,(\ell n\, y_j - \ell n\, y_i) \\
&\qquad + \sum\sum_{i>j} y_i^p y_j^p\, \ell n\, y_j\,(\ell n\, y_j - \ell n\, y_i)\Big] \qquad (2.1.21) \\
&= p\sum\sum_{i<j} y_i^p y_j^p(\ell n\, y_j - \ell n\, y_i)^2\ .
\end{aligned}$$

Note that $i, j \in \{1, \ldots, K\}$ is understood in the above summations, but omitted to simplify the notation. From equation (2.1.21), we see that

$$-p(\phi')^2 + p\phi\phi'' \begin{cases} \geq 0, \text{ if } p>0 \ , \\ \leq 0, \text{ if } p<0 \ , \end{cases} \tag{2.1.22}$$

with equality occurring if and only if $y_i = y_j$ for all $i, j \in \{1, ..., K\}$. From relations (2.1.19), (2.1.20), and (2.1.22), it is seen that $B(y ; p) > 0$ for $p>0$. Hence, we conclude that

$$\frac{\partial^2}{\partial p^2} S(y ; p) > 0 \ , \ p>0 \ . \tag{2.1.23}$$

Unfortunately, for $p<0$, the sign of $\partial^2 S/\partial p^2$ cannot be determined from the above analysis. This agrees with the known result that $S(y ; p)$ is not convex, or necessarily concave in p for $p<0$, (e.g., see 1-16 of Beckenbach and Bellman, 1965). For example, if $y_1 = y_2 = ... = y_K = a$, then

$$S(y ; p) = K^{1/p} a \ ;$$

$$\therefore \ \frac{\partial S}{\partial p} = -\frac{K^{1/p}a}{p^2} \ell n K \ , \quad \text{and} \quad \frac{\partial^2 S}{\partial p^2} = \frac{K^{1/p} a \ \ell n \ K}{p^3} \left(2 + \frac{\ell n \ K}{p}\right) ;$$

so that $S(y ; p)$ has a unique inflection point in this case at the negative value

$$p = -\frac{1}{2} \ell n \ K \ . \tag{2.1.24}$$

From the inequality (2.1.23), we immediately obtain the following result.

Property 2.1.2

The function $S(y ; p)$ given in equation (2.1.1) with $K \geq 2$ is strictly convex in p, for $0 < p < \infty$.

Property 2.1.2 strengthens the known fact (Beckenbach, 1946) that $S(y ; p)$ is convex in p for $p>0$, and allows us now to state the following strict inequality.

$$S(y ; P) < \sum_{i=1}^{m} \alpha_i S(y ; p_i) \ , \tag{2.1.25}$$

where

$$P = \sum_{i=1}^{m} a_i p_i ,$$

$$a_i > 0 \ , \ i = 1, ..., m, \sum_{i=1}^{m} a_i = 1, \text{ and}$$

$p_i > 0$, $i = 1, ..., m$, are arbitrary (and distinct) values of p.

Returning to equation (2.1.6) for the first derivative, the following interesting observation is made.

$$\lim_{p \to 0^-} \frac{\partial}{\partial p} S(y; p) = \lim_{p \to 0^-} \left\{ \frac{\phi^{\frac{1-p}{p}}}{p^2} \left(\sum_{i=1}^{K} y_i^p \ell n \left[\frac{y_i^p}{\phi} \right] \right) \right\}$$

$$= -\ell n K \lim_{p \to 0^-} \left\{ \frac{\phi^{1/p}}{p^2} \right\} \qquad \left(\because \ y_i^p \to 1, \forall i \right) .$$

But

$$\phi^{1/p} = \left[\sum_{i=1}^{K} y_i^p \right]^{1/p} \begin{cases} \leq y_m K^{1/p}, p < 0, y_m = \min_i (y_i) ; \\ \geq y_M K^{1/p}, p < 0, y_M = \max_i (y_i) . \end{cases}$$

Hence

$$\lim_{p \to 0^-} \frac{\partial}{\partial p} S(y ; p) = -0 . \tag{2.1.26}$$

We can now prove the following fact.

Property 2.1.3

S(y ; p) has at least one inflection point, and hence is neither convex nor concave in p, in the interval $-\infty < p < 0$.

Proof:

We know that S(y ; p) approaches the horizontal asymptote, $y_m = \min_i (y_i)$, from below as $p \to -\infty$ (e.g., Beckenbach, 1946). Hence S(y ; p) is concave in p for sufficiently large negative values of p. It is easily shown that $S(y ; p) \to 0^+$ as $p \to 0^-$. Thus, from equation (2.1.26) we conclude that a $\delta > 0$ exists such that S(y ; p) is convex in p for $p \in (-\delta, 0)$. Therefore the property is proven.

Beckenbach (1946) poses a question concerning the number of possible inflection points of S(y ; p) as a function of p in the interval $(-\infty, 0)$. We see from the above result that there is at least one such point, thus establishing a lower bound of one. It follows that S(y ; p) can never be concave over the entire interval $-\infty < p < 0$, a fact which does not appear to be recognized until now. To illustrate Properties 2.1.2 and 2.1.3, the previous example ($y_1 = y_2 = \ldots = y_K = a$, $S(y ; p) = K^{1/p} a$) is plotted in Figure 2.1.1.

2.1.2 A Generalization

We now introduce a generalization of the sum of order p, defined as follows:

$$T(y ; b, p) = \left[\sum_{i=1}^{K} b_i y_i^p \right]^{1/p} , \qquad (2.1.27)$$

where

$$y = (y_1, \ldots, y_K)^T , \qquad y_i > 0, i=1, \ldots, K ,$$

$$b = (b_1, \ldots, b_K)^T , \qquad b_i > 0, i=1, \ldots, K ,$$

and $\quad p \neq 0.$

(This function is termed a 'weighted' sum in 2.10 of Hardy, Littlewood and Pólya, 1952.) The vector b and the scalar p can be considered as a set of parameter values. If all the weights $b_i = 1$, then T is just the ordinary sum of order p given in (2.1.1). Note that the function T(y; b, p) has the form of a generalized ℓ_p distance given by,

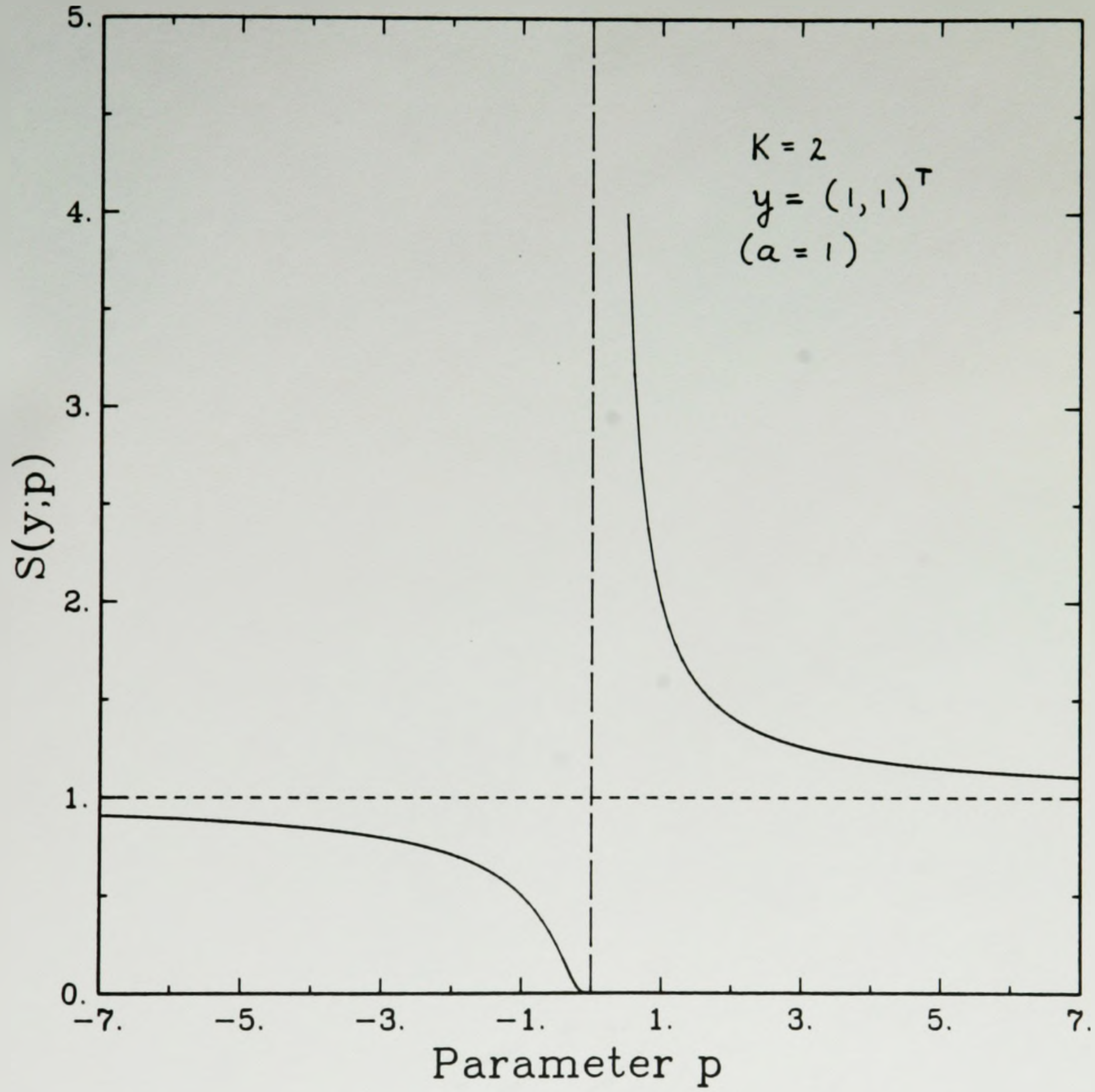


Figure 2.1.1 A Sum of Order p.

$$\ell_{bp}(x - a_r) = \left[\sum_{i=1}^{K} b_i \, | x_i - a_{ri} |^p \right]^{1/p} . \tag{2.1.28}$$

The weights b_i could, for example, represent non-symmetric costs along the axis directions in a location model. Just as for S(y ; p), we are interested in studying the behaviour of the sum T as a function of its parameter p.

Consider first the asymptotic behaviour of T. Letting $y_m = \min_i (y_i)$ and $y_M = \max_i (y_i)$ as before, we obtain,

$$\begin{aligned} \lim_{p \to +\infty} \{T(y ; b, p)\} &= \lim_{p \to +\infty} \left\{ \left[\sum_{i=1}^{K} b_i y_i^p \right]^{1/p} \right\} \\ &= y_M \lim_{p \to +\infty} \left\{ \left[\sum_{i=1}^{K} b_i \left(\frac{y_i}{y_M} \right)^p \right]^{1/p} \right\} \\ &= y_M , \end{aligned} \tag{2.1.29}$$

and similarly,

$$\begin{aligned} \lim_{p \to -\infty} \{T(y ; b, p)\} = y_m \lim_{p \to -\infty} \left\{ \left[\sum_{i=1}^{K} b_i \left(\frac{y_m}{y_i} \right)^{-p} \right]^{1/p} \right\} \\ = y_m . \end{aligned} \tag{2.1.30}$$

Thus, the function T approaches the same horizontal asymptotes irrespective of the positive weights b_i, $i = 1, \ldots, K$.

Without loss in generality, let us assume that all the y_i's have distinct values; that is, $y_i \neq y_j$, $i \neq j$, for all $i, j \in \{1, \ldots, K\}$. (If this is not the case, common terms can be added together and K adjusted accordingly.) Denoting the weights associated with y_m and y_M by b_m and b_M respectively, it is clear from (2.1.29) and (2.1.30) that for $K \geq 2$,

$$\lim_{p \to +\infty} T = \begin{cases} y_M^+ , & \text{if } b_M \geq 1 , \\ y_M^- , & \text{if } 0 < b_M < 1 , \end{cases} \tag{2.1.31}$$

and

$$\lim_{p \to -\infty} T = \begin{cases} y_m^- , & \text{if } b_m \geq 1 , \\ y_m^+ , & \text{if } 0 < b_m < 1 . \end{cases} \tag{2.1.32}$$

Thus, the direction of approach from above or below the horizontal asymptotes y_M, y_m depends on the magnitude of the corresponding weights b_M, b_m.

We now examine the behaviour of T near $p = 0$. Let

$$\beta = \sum_{i=1}^{K} b_i . \tag{2.1.33}$$

There are three possibilities to consider.

(i) $\beta > 1$:

It is readily seen that

$$\lim_{p \to 0^+} T = +\infty, \quad \lim_{p \to 0^-} T = 0 . \tag{2.1.34}$$

(This is the same result as for S(y ; p)).

(ii) $\beta < 1$:

The situation is reversed; i.e.,

$$\lim_{p \to 0^+} T = 0 , \quad \lim_{p \to 0^-} T = +\infty . \tag{2.1.35}$$

(iii) $\beta = 1$:

In this case, T(y; b, p) is called the mean value function. Beckenbach (1946) gives the following result without proof.

$$\lim_{p \to 0} \{T(y ; b, p)\} = \prod_{i=1}^{K} y_i^{b_i} . \tag{2.1.36}$$

It follows that T is continuous at $p=0$ if, and only if, $\beta=1$. The result in (2.1.36) is not immediately obvious; hence we prove it below.

$$\ell n\, T(y\,;b,p) = \frac{1}{p}\,\ell n\left(\sum_{i=1}^{K} b_i\, y_i^p\right).$$

In the limit $p \rightarrow 0$, the denominator and numerator on the right-hand side both go to zero when $\beta=1$, so that by l'Hôpital's rule,

$$\lim_{p\rightarrow 0}\ \ell n T = \lim_{p\rightarrow 0}\left\{\frac{1}{\sum_{i=1}^{K} b_i\, y_i^p}\cdot\sum_{i=1}^{K} b_i\, y_i^p\, \ell n\, y_i\right\}$$

$$= \sum_{i=1}^{K} b_i\, \ell n\, y_i\,.$$

Hence,

$$\lim_{p\rightarrow 0} T \;=\; \exp\left(\sum_{i=1}^{K} b_i\, \ell n\, y_i\right)$$

$$= \prod_{i=1}^{K} y_i^{b_i}\,,$$

confirming equation (2.1.36).

We now calculate the first and second-order partial derivatives of ℓn T with respect to p. Letting

$$a_i = \frac{b_i}{\beta}\quad,\quad i=1,\ldots,K\,, \tag{2.1.37}$$

equation (2.1.27) can be re-written as

$$T(y\,;b,p) = \beta^{1/p}\left[\sum_{i=1}^{K} a_i\, y_i^p\right]^{1/p}, \tag{2.1.38}$$

where $a_i > 0$, $i = 1, \ldots, K$, and

$$\sum_{i=1}^{K} a_i = 1\,.$$

Then

$$\ell n\, T = \frac{1}{p}\,\ell n\,\beta + \frac{1}{p}\,\ell n\left(\sum_{i=1}^{K} a_i\, y_i^p\right). \tag{2.1.39}$$

a) **First Derivative**

$$\frac{\partial}{\partial p}\ell n\, T = -\frac{1}{p^2}\ell n\beta - \frac{1}{p^2}\ell n\left(\sum_{i=1}^{K} a_i\, y_i^p\right) + \frac{1}{p}\cdot \frac{1}{\sum_{i=1}^{K} a_i\, y_i^p}\sum_{i=1}^{K} a_i\, y_i^p\, \ell n y_i$$

$$= \frac{1}{p^2 \sum_{i=1}^{K} a_i\, y_i^p}\left[\sum_{i=1}^{K} a_i\, y_i^p\, \ell n\left(\frac{y_i^p}{\beta\sum_{j=1}^{K} a_j\, y_j^p}\right)\right],\ p \neq 0\ . \tag{2.1.40}$$

Since

$$\frac{\partial}{\partial p}\ell n\, T = \frac{1}{T}\,\frac{\partial T}{\partial p}\ ,$$

we immediately get

$$\frac{\partial T}{\partial p} = \frac{\beta^{1/p}}{p^2}\left[\sum_{i=1}^{K} a_i\, y_i^p\right]^{(1-p)/p}\left[\sum_{i=1}^{K} a_i\, y_i^p\, \ell n\left(\frac{y_i^p}{\beta\sum_{j=1}^{K} a_j\, y_j^p}\right)\right],\quad p \neq 0\ . \tag{2.1.41}$$

It is interesting to note that for $\beta > 1$,

$$\lim_{p\to 0^-}\frac{\partial T}{\partial p} = -\ell n\beta \cdot \lim_{p\to 0^-}\left[\frac{\beta^{1/p}}{p^2}\left(\sum_{i=1}^{K} a_i\, y_i^p\right)^{1/p}\right]$$

$$= -\ell n\,\beta\cdot\prod_{i=1}^{K} y_i^{a_i}\cdot \lim_{p\to 0^-}\left[\frac{\beta^{1/p}}{p^2}\right] \qquad \text{(equation (2.1.36))}$$

$$= -0\ . \tag{2.1.42}$$

This is the same result as obtained for S(y ; p); (see equation (2.1.26)). Thus we can readily show that Property 2.1.3 also holds for the generalized sum T(y; b, p) with $\beta > 1$. Meanwhile, for $\beta < 1$, we obtain in similar fashion the following result.

$$\lim_{p \to 0^+} \frac{\partial T}{\partial p} = +0 .$$

Hence, the function T(y; b, p) with $\beta < 1$ has at least one inflection point in the interval $0 < p < +\infty$.

b) Second Derivative

In the following summations $i, j \in \{1, ..., K\}$ is understood, but omitted to simplify the notation.

$$\frac{\partial^2}{\partial p^2} \ell n\, T = \frac{2}{p^3} \ell n \beta + \frac{2}{p^3} \ell n \Big(\sum_i a_i y_i^p \Big) - \frac{2}{p^2 \sum_i a_i y_i^p} \sum_i a_i y_i^p \ell n y_i$$

$$- \frac{1}{p} \cdot \frac{1}{\Big(\sum_i a_i y_i^p \Big)^2} \Big(\sum_i a_i y_i^p \ell n y_i \Big)^2 + \frac{1}{p \sum_i a_i y_i^p} \cdot \sum_i a_i y_i^p (\ell n y_i)^2 .$$

After some re-arranging this reduces to

$$\frac{\partial^2}{\partial p^2} \ell n\, T = \frac{1}{p^3 \sigma^2} \Bigg[2\sigma \sum_i a_i y_i^p \ell n \Big(\frac{\beta \sigma}{y_i^p} \Big)$$

$$+ p^2 \sum_{j<i} \sum a_i a_j y_i^p y_j^p \Big(\ell n\, y_i - \ell n\, y_j \Big)^2 \Bigg], \quad p \neq 0, \tag{2.1.43}$$

where

$$\sigma = \sum_{i=1}^{K} a_i y_i^p . \tag{2.1.44}$$

By means of the first and second derivatives calculated above, we are able to prove some interesting results, which are extensions of Properties 2.1.1 and 2.1.2 for the ordinary sum of order p. First, let us consider the simple case, $K = 1$. Then,

$$T(y\,; b, p) = b_1^{1/p} y_1 , \tag{2.1.45}$$

$$\frac{\partial T}{\partial p} = -\frac{1}{p^2} b_1^{1/p} y_1 \, \ell n \, b_1, \; p \neq 0 \,, \tag{2.1.46}$$

and

$$\frac{\partial^2 T}{\partial p^2} = b_1^{1/p} y_1 \, \ell n \, b_1 \left[\frac{\ell n \, b_1}{p^4} + \frac{2}{p^3} \right], \; p \neq 0 \,. \tag{2.1.47}$$

The following facts are immediately obvious:

a) If $b_1 > 1$, $\partial T/\partial p < 0$ for all $p \neq 0$, $\partial^2 T/\partial p^2 > 0$ for $p > 0$, and T has a unique inflection point at $p = -1/2 \, \ell n \, b_1 < 0$. Thus, T is decreasing and strictly convex in p, for $0 < p < +\infty$. Meanwhile, for $-\infty < p < 0$, T is again decreasing in p, but concave in the interval $(-\infty, -1/2 \, \ell n \, b_1]$ and convex in the interval $[-1/2 \, \ell n \, b_1, 0)$.

b) If $b_1 = 1$, then $T = y_1$ which is constant for varying p.

c) If $0 < b_1 < 1$, $\partial T/\partial p > 0$ for all $p \neq 0$, $\partial^2 T/\partial p^2 > 0$ for $p < 0$, and T has a unique inflection point at $p = -1/2 \, \ell n \, b_1 > 0$. Thus, T is increasing and strictly convex in p, for $-\infty < p < 0$. Meanwhile for $0 < p < +\infty$, T is again increasing in p, but convex in the interval $(0, -1/2 \, \ell n \, b_1]$ and concave in the interval $[-1/2 \, \ell n \, b_1, +\infty)$. Alternatively, by noting that $b_1 = 1/b_1'$ where $b_1' > 1$, we readily see that the behaviour here is just the mirror image of the first case.

The preceding results give some insights into the more interesting case where $K \geq 2$, to which we now turn.

Theorem 2.1.1

Consider the function T(y ; b, p) defined in (2.1.27), with given (constant) vectors y and b. Assume without loss in generality that $y_M = \max_i (y_i)$ occurs for a unique $M \in \{1, ..., K\}$; i.e., there are no ties. (If this is not the case, add the coefficients (b_i) of the ties to form one term.) Then for $K \geq 2$, T(y ; b, p) is a decreasing function of p for $0 < p < +\infty$, if, and only if,

$b_M \geq 1$, where b_M is the coefficient of y_M. Furthermore, if $b_M \geq 1$, T is also a strictly convex function in p over this same interval.

Proof:

(i) (If) Since $p > 0$, $K \geq 2$ and $b_M \geq 1$, it follows that,

$$\frac{y_i^p}{\beta\sigma} = \frac{y_i^p}{\beta \sum_{j=1}^{K} a_j y_j^p} = \frac{y_i^p}{\sum_{j=1}^{K} b_j y_j^p} < \frac{y_i^p}{b_M y_M^p} \leq 1 \, , \; i = 1, \ldots, K \, . \tag{2.1.48}$$

From equations (2.1.40) and (2.1.43), we see that

$$\frac{\partial \ln T}{\partial p} < 0 \qquad \text{and} \qquad \frac{\partial^2}{\partial p^2} \ln T > 0 \, .$$

Hence, $\ln T$ is decreasing and strictly convex in p for $0 < p < +\infty$. It immediately follows that T is a decreasing function of p in this interval. Furthermore,

$$\frac{\partial^2 T}{\partial p^2} = T \frac{\partial^2}{\partial p^2} \ln T + \frac{1}{T}\left(\frac{\partial T}{\partial p}\right)^2 ; \tag{2.1.49}$$

so that

$$\frac{\partial^2 T}{\partial p^2} \geq T \frac{\partial^2}{\partial p^2} \ln T > 0 \, .$$

Thus, T is also strictly convex in p, for $0 < p < +\infty$. We conclude that $b_M \geq 1$ is a sufficient condition for T to be a decreasing strictly convex function of $p \in (0, \infty)$.

(ii) (Only if) That $b_M \geq 1$ is a necessary condition for T to be decreasing in p immediately follows from the asymptotic behaviour of T as $p \rightarrow +\infty$, shown in (2.1.31). If $b_M < 1$, T approaches y_M asymptotically from below, and hence is increasing and concave for sufficiently large p.

For applications of T as a distance function in location models, one should be interested in the case where all the weights are greater than or equal to one; i.e., $b_i \geq 1$,

$i = 1, \ldots, K$. (Otherwise, distances less than Euclidean are possible.) Theorem 2.1.1 leads to the following useful result for this case.

Corollary 2.1.1

$T(y\,;b,p)$ with $K \geq 2$ is a decreasing function of $p > 0$ for given weights b and all (positive) y, if, and only if, $b_i \geq 1$, $i = 1, \ldots, K$. Furthermore, T is also strictly convex in p under these conditions.

Proof:

Consider any y such that the y_i's are not all equal. Clearly, $b_M \geq 1$ if all the $b_i \geq 1$. By the theorem, we know that T is decreasing and strictly convex in $p > 0$. Now consider any y such that all the y_i's are equal. Then, $T = \beta^{1/p} y_1$, where $\beta = \Sigma^K_{i=1} b_i > 1$ if all the $b_i \geq 1$. From our analysis of the special case ($K = 1$), we know that T is decreasing and strictly convex in $p > 0$. Thus, $b_i \geq 1$, $i = 1, \ldots, K$, is a sufficient condition. That this is also a necessary condition is readily seen by contradiction. Suppose $b_r < 1$, for some $r \in \{1, \ldots, K\}$. Construct a vector y such that $y_i \neq y_j$, $i \neq j$, $\forall\ i,j \in \{1, \ldots, K\}$, and $y_r = \max_i (y_i)$. By the theorem, we know that T is not a decreasing function of $p \in (0, +\infty)$, for this y.

The shape of T as a function of p becomes more complex when the criteria on the weights are changed, as shown in the following result.

Property 2.1.4

Consider a vector of weights b, such that $\beta = \Sigma_i b_i > 1$, and $b_r < 1$ for at least one $r \in \{1, \ldots, K\}$. Then, for any given y there exists a $\delta > 0$ such that T is decreasing and strictly convex in $p \in (0, \delta)$. However, if $y_r = \max_i (y_i)$, and there are no ties, then T is increasing and strictly concave for sufficiently large positive p.

Proof:

Follows immediately from the limit $p \to 0^+$ in (2.1.34) and the limit $p \to +\infty$ in (2.1.31).

Note that the function T described in the preceding result is neither increasing or decreasing in p nor convex or concave in p over the entire interval $0 < p < +\infty$, and that at least one inflection point exists in this interval. This is illustrated in Figure 2.1.2.

The fact that T is a decreasing function of p in the interval $(0, +\infty)$ for all y if, and only if, $b_i \geq 1$, $i = 1, \ldots, K$ (Corollary 2.1.1), has been recognized previously by Hardy, Littlewood and Pólya (1952) in their Theorem 23. However, their proof is completely different than ours, and furthermore, does not show the important result that T is convex in p under these conditions. The third and final case to consider for the weights b is where $\beta = \Sigma_i b_i \leq 1$. In the same Theorem 23, the above authors prove that T is non-decreasing in p over the interval $(0, +\infty)$ for all y if, and only if, this condition holds. Thus, the following property can be given without proof.

Property 2.1.5

A necessary and sufficient condition to have

$$T(y\,;b\,,p_1) \leq T(y\,;b\,,p_2)\,, \qquad 0 < p_1 < p_2\,,$$

for given weights b and all y, is that $\beta \leq 1$. Furthermore, there is strict inequality unless all the y_i are equal and $\beta = 1$.

We make the additional observation that T has at least one inflection point in the interval $0 < p < +\infty$, if $\beta < 1$. (See the discussion following relation (2.1.42).) Thus, T is neither convex nor concave in p under this condition.

Use of negative p when the weighted sum T is a distance function in location models does not appear to have a physical interpretation. However, there may be other

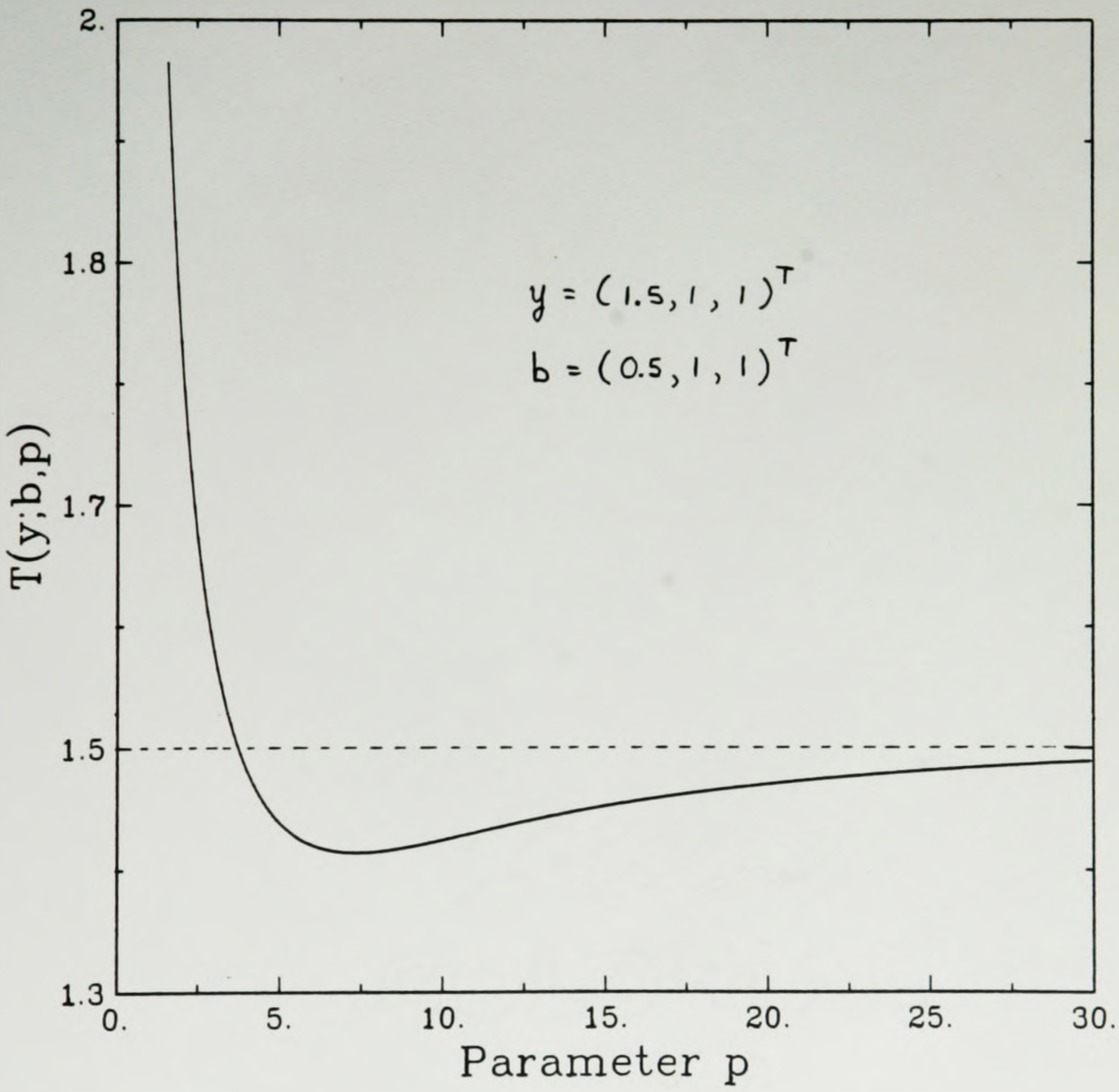


Figure 2.1.2 General Shape of T under Conditions of Property 2.1.4.

situations where $p < 0$ might be considered. In any case, we would like to take full advantage of our lengthy calculations of derivatives. This questionable motivation leads to the following results for $p < 0$.

Theorem 2.1.2

Consider the function $T(y\,;b,p)$ defined in (2.1.27), with given (constant) vectors y and b. Assume without loss in generality that $y_m = \min_i (y_i)$ occurs for a unique $m \in \{1, \ldots, K\}$; i.e., there are no ties. (If this is not the case, add the coefficients (b_i) of the ties to form one term.) Then for $K \geq 2$, $T(y\,;b,p)$ is a decreasing function of p in the interval $(-\infty, 0)$, if, and only if, $b_m \geq 1$, where b_m is the coefficient of y_m. Furthermore, if $b_m \geq 1$, $\ell n\, T$ is a strictly concave function of p over this same interval.

Proof:

(i) (If). Since $p < 0$, $K \geq 2$, and $b_m \geq 1$, it follows that

$$\frac{y_i^p}{\sum_{j=1}^{K} b_j y_j^p} < \frac{y_i^p}{b_m y_m^p} = \frac{1}{b_m}\left(\frac{y_m}{y_i}\right)^{|p|} \leq 1 \;, \qquad i = 1, \ldots, K. \tag{2.1.50}$$

Returning to equations (2.1.40) and (2.1.43), we can readily show that

$$\frac{\partial \ell n\, T}{\partial p} < 0 \qquad \text{and} \qquad \frac{\partial^2 \ell n\, T}{\partial p^2} < 0 \;, \qquad p < 0 .$$

Hence, $\ell n\, T$ is decreasing and strictly concave in p in the interval $(-\infty, 0)$. It immediately follows that T is a decreasing function of p in this interval. We conclude that $b_m \geq 1$ is a sufficient condition for T to be decreasing in p and $\ell n\, T$ strictly concave in p, for $-\infty < p < 0$.

(ii) (Only if). That $b_m \geq 1$ is a necessary condition for T to be decreasing in p immediately follows from the asymptotic behaviour of T as $p \to -\infty$, shown in (2.1.32). If

$b_m < 1$, T approaches y_m asymptotically from above; so that T (or $\ell n\, T$) is increasing and convex for sufficiently large negative values of p.

Corollary 2.1.2

$T(y\,;b,p)$ with $K \geq 2$ is a decreasing function of p in the interval $(-\infty, 0)$ for given weights b and all (positive) y, if, and only if, $b_i \geq 1$, $i = 1, \ldots, K$. Furthermore, $\ell n\, T$ is strictly concave in p under these conditions.

Proof:

Consider any y such that the y_i's are not all equal. Clearly, $b_m \geq 1$ if all the $b_i \geq 1$. By the theorem, we know that T is decreasing and $\ell n T$ is strictly concave in $p \in (-\infty, 0)$. Now consider any y such that all the y_i's are equal. Then, $T = \beta^{1/p} y_1$, where $\beta = \Sigma_i\, b_i > 1$, if all the b_i's ≥ 1. From our analysis of the special case ($K = 1$), we know that T is decreasing in $p \in (-\infty, 0)$. Furthermore,

$$\ell n\, T = \frac{1}{p}\,\ell n\, \beta + \ell n\, y_1, \qquad \frac{\partial \ell n\, T}{\partial p} = -\frac{1}{p^2}\ell n\, \beta, \quad \text{and}$$

$$\frac{\partial^2}{\partial p^2}\,\ell n\, T = \frac{2}{p^3}\ell n\, \beta < 0, \qquad \text{for } p < 0.$$

Thus, $\ell n\, T$ is strictly concave in $p \in (-\infty, 0)$. We conclude that $b_i \geq 1$, $i = 1, \ldots, K$, is a sufficient condition. That this is also a necessary condition is readily seen by contradiction, similar to the procedure in Corollary 2.1.1.

2.2 Norms

2.2.1 Definition of Round Norms

In order to derive additional properties of location models, it is useful to characterize norms in more detail. To this end, we define the unit ball B of a norm k acting on R^N as follows:

$$B = \{x \mid k(x) \leq 1\} \ . \tag{2.2.1}$$

Thus, B is the closed set of points in R^N contained by the unit contour of k. The symmetry property, $k(y) = k(-y)$, implies that if $y \in B$ then so is $-y$. Hence B is a symmetric set of points containing the origin.

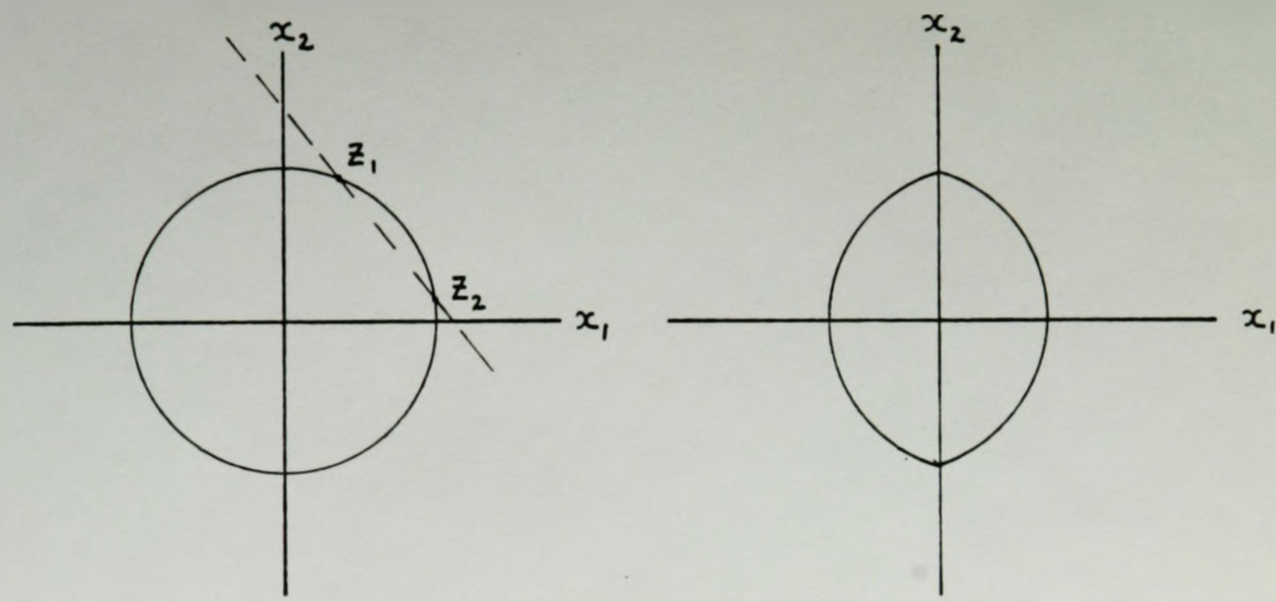
Suppose x_1 and x_2 belong to B, and consider a point $y = \lambda x_1 + (1-\lambda)x_2$ where $\lambda \in [0,1]$. That is, y can be any point along the line segment joining x_1 and x_2. Then, using the triangle inequality and homogeneity properties of norms, we have

$$\begin{aligned} k(y) &= k(\lambda x_1 + (1-\lambda)\, x_2) \\ &\leq \lambda\, k(x_1) + (1-\lambda)\, k(x_2) \leq 1 \ . \end{aligned} \tag{2.2.2}$$

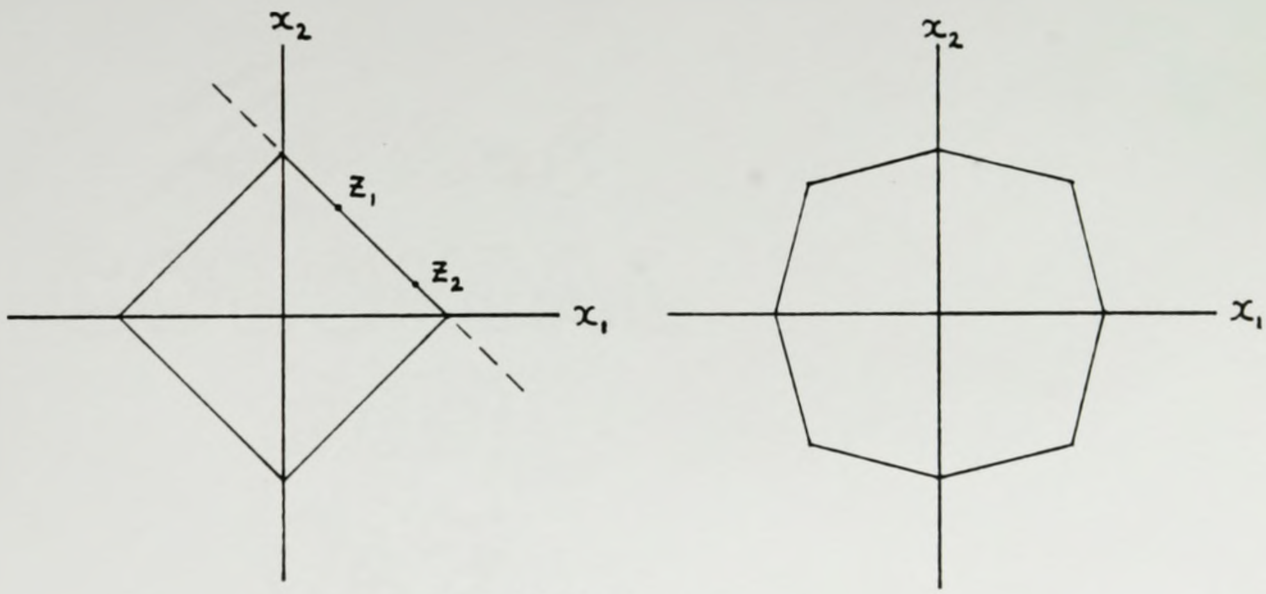
Thus $y \in B$, and we conclude that the unit ball is a convex set. In summary, the unit ball of any norm is a symmetric closed bounded convex set. Furthermore, it can be shown (e.g., Theorem 15.2 of Rockafellar, 1970) that a one-to-one correspondence exists between the norms k and the symmetric closed bounded convex sets B. Thus, a norm is uniquely defined by a unit ball, and vice versa.

Thisse, Ward and Wendell (1984) use the unit ball to distinguish between block and round norms. They classify block norms as those whose contours are polytopes (polygons in R^2), as distinct from round norms whose contours contain no flat spots. This feature is illustrated in Figure 2.2.1. Referring to Figure 2.2.1, let z_1 and z_2 denote two points on some contour C of the norm k. This contour, which is the boundary of a convex set, is given by

$$C = \{x \mid k(x) = b\} \ , \tag{2.2.3}$$



a) Round norms



b) Block norms

Figure 2.2.1 Unit Contours in R^2.

where b is a scalar greater than zero ($b > 0$). Noting that $k(z_1) = k(z_2) = b$, and proceeding as in (2.2.2), we obtain

$$k(\lambda z_1 + (1-\lambda) z_2) \leq b \ , \quad \lambda \in [0,1] \ . \tag{2.2.4}$$

In particular, if C is the unit contour, so that $b = 1$, relation (2.2.4) becomes

$$k(\lambda z_1 + (1-\lambda) z_2) \leq 1 \ , \quad \lambda \in [0,1] \ . \tag{2.2.5}$$

We are now ready to give a formal definition of the round norm, in place of the qualitative description of Thisse, Ward and Wendell (1984).

Definition 2.2.1 A norm k is a round norm if, and only if,

$$k(\lambda z_1 + (1-\lambda) z_2) < 1 \ , \tag{2.2.6}$$

for all z_1, z_2 on the unit contour of k such that $z_1 \neq z_2$, and all λ in the open interval $(0,1)$.

Note that the strict inequality in relation (2.2.6) implies that the unit ball B of a round norm is a strictly convex set. In contrast, the unit ball of a block norm is not strictly convex. If z_1 and z_2 are on the same facet of the unit polytope of some block norm k, then $k(\lambda z_1 + (1-\lambda) z_2) = 1$. Thus, for k(x) a block norm, the $<$ sign in relation (2.2.6) must be replaced by a $\leq$ sign. The strict convexity of B for round norms allows for some useful properties given below. However, before proceeding with these properties, we subdivide round norms into two classes as follows.

Definition 2.2.2 A round norm k(x) which is differentiable at all $x \in R^N$, except $x = 0$ (the origin), is termed a differentiable round norm. Otherwise k(x) is a nondifferentiable round norm.

By definition, all norms must be nondifferentiable at the origin; (e.g., see directional derivatives in Chapter 2 of Juel, 1975). The differentiable round norm has the useful property that its first-order derivatives exist everywhere else. However, this property does not hold for nondifferentiable round norms, or for block norms. For example, if k(x) is a block norm, then clearly $k(x)$ is nondifferentiable along the edges of its polytope contours, while it is differentiable at all other points except $x=0$. In R^2, this means that the block norm is nondifferentiable at the corner points of its polygon contours, while it is differentiable everywhere else except at the origin.

The preceeding discussion leads to a practical classification of norms for use in location models. This classification scheme is shown in Figure 2.2.2. Sample contours of the various norms are illustrated in Figure 2.2.1.

2.2.2 Properties

By means of the triangle inequality and homogeneity properties of any norm k (relations (1.2.9) and (1.2.10)), it follows immediately that $k(\lambda x_1 + (1-\lambda) x_2) \leq \lambda k(x_1) + (1-\lambda) k(x_2)$, for all $x_1, x_2 \in R^N$ and all $\lambda \in [0, 1]$. Hence, k is a convex function of x on R^N. This is a well-known result, (e.g., Fact 1 in Chapter 2 of Juel, 1975, and p. 131 of Rockafellar, 1970), which permits many location problems to be formulated as convex minimization models. In this sub-section, we exploit the strict convexity of the unit ball B to derive stronger convexity results for the case where k is a round norm. This will enable us later on to develop some general properties for minisum models which employ round norms.

We begin by formally proving the equivalence between the round norm and S-norm. The latter is defined by Pelegrin, Michelot and Plastria (1985) as follows:

A norm k on R^N is called an S-norm if, and only if, $k(x_1 + x_2) = k(x_1) + k(x_2)$ implies that $x_2 = \beta x_1$ for some scalar β.

Clearly, β must be a non-negative scalar; otherwise if $x_2 = -|\beta| x_1$, then

$$\begin{aligned} k(x_1 + x_2) &= k(x_1 - |\beta| x_1) \\ &= |1 - |\beta|| k(x_1) \\ &< |1 + |\beta|| k(x_1) = k(x_1) + k(x_2) , \end{aligned}$$

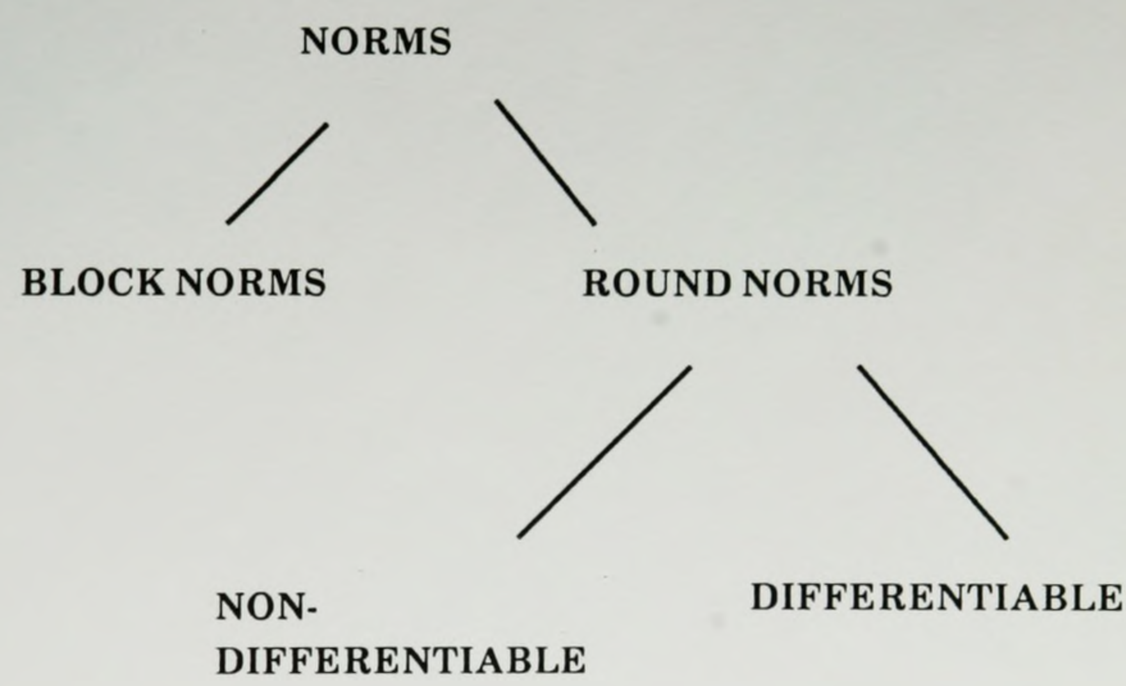


Figure 2.2.2 A practical classification scheme for norms.

for any norm k and $\beta \neq 0$. We proceed with the following basic result.

Property 2.2.1

Let k(x) be a round norm on R^N. Then for $x_1, x_2 \neq 0$, $k(x_1 + x_2) = k(x_1) + k(x_2)$ implies that $x_2 = \beta x_1$ for some scalar $\beta > 0$. Furthermore, if this condition is satisfied by a norm, it is a round norm.

Proof:

Let x_1 and x_2 be any points in R^N other than the origin. Obviously, if $x_2 = \beta x_1$, $\beta > 0$, then for any norm g, we have

$$g(x_1 + x_2) = (1 + \beta)\, g(x_1) = g(x_1) + g(x_2)\ .$$

Now suppose k is a round norm, and choose x_1 and x_2 such that $x_2 \neq \beta x_1$, $\beta > 0$. Then the half-lines from the origin which pass through x_1 and x_2 intersect the unit contour of k at two distinct points, say z_1 and z_2 respectively. We have $x_1 = \mu_1 z_1$, $x_2 = \mu_2 z_2$, where μ_1, μ_2 are positive scalars and $z_1 \neq z_2$. Then, letting $\mu_T = \mu_1 + \mu_2$, $\lambda = \mu_1/\mu_T < 1$, gives

$$\begin{aligned} k(x_1 + x_2) &= \mu_T\, k[\lambda z_1 + (1 - \lambda)\, z_2] \\ &< \mu_T\ . \qquad \text{(relation (2.2.6))} \end{aligned}$$

But

$$\begin{aligned} k(x_1) + k(x_2) &= \mu_1\, k(z_1) + \mu_2\, k(z_2) \\ &= \mu_1 + \mu_2 = \mu_T\ . \end{aligned}$$

Hence,

$$k(x_1 + x_2) < k(x_1) + k(x_2).$$

We conclude that for $k(x_1 + x_2)$ to equal $k(x_1) + k(x_2)$, we must have $x_2 = \beta x_1$, $\beta > 0$.

Now we prove the second part of this property. Suppose g is a norm such that $g(x_1 + x_2) = g(x_1) + g(x_2)$ implies that $x_2 = \beta x_1$, $\beta > 0$. Then, for any x_1, x_2 such that $x_2 \neq \beta x_1$,

$\beta > 0$, we have

$$g(x_1 + x_2) < g(x_1) + g(x_2) .$$

Now choose two distinct points z_1 and z_2 on the unit contour of g. Clearly $z_2 \neq \beta z_1$, $\beta > 0$. Then, letting $\beta = \lambda/(1-\lambda)$, $\lambda \in (0,1)$, we see that $(1-\lambda)z_2 \neq \lambda z_1$. Substituting $x_1 = \lambda z_1$ and $x_2 = (1-\lambda)z_2$ into the above inequality, we get

$$g[\lambda z_1 + (1-\lambda)z_2] < g(\lambda z_1) + g[(1-\lambda)z_2]$$

$$= \lambda\, g(z_1) + (1-\lambda) g\,(z_2) = 1 .$$

Hence, we conclude that g must be a round norm, ending the proof.

We see then, that the round norm introduced by Thisse, Ward and Wendell (1984) and the S-norm defined by Pelegrin, Michelot and Plastria (1985) are one and the same. The next property is taken from Pelegrin, Michelot and Plastria (1985). We introduce some notation first. Let $L(x_1, x_2)$ denote the straight line through any two points $x_1, x_2 \in R^N$, $(x_1 \neq x_2)$, and let (x_1, x_2) denote the open segment connecting x_1 and x_2. Finally, we define

$$L'(x_1, x_2) = L(x_1, x_2) - (x_1, x_2) . \tag{2.2.7}$$

Since the authors do not give a proof of their result, we add one for completeness.

Property 2.2.2

For any point $a \notin L'(x_1, x_2)$, any round norm k, and $x_0 = \lambda x_1 + (1-\lambda)\, x_2$, $\lambda \in (0,1)$, the following inequality holds;

$$k(x_0 - a) < \lambda\, k(x_1 - a) + (1-\lambda)k\,(x_2 - a) . \tag{2.2.8}$$

Proof:

$$k(x_0 - a) = k[\lambda x_1 + (1-\lambda)x_2 - a]$$

$$= k(y_1 + y_2) ,$$

where

$$y_1 = \lambda(x_1 - a) \quad \text{and} \quad y_2 = (1-\lambda)(x_2 - a) .$$

Since $a \notin L'(x_1, x_2)$, clearly $y_1 \neq \beta y_2$, where β is a positive scalar. It follows from Property 2.2.1 that $k(y_1 + y_2) < k(y_1) + k(y_2)$. We conclude that

$$k(x_0 - a) < \lambda k(x_1 - a) + (1 - \lambda)\, k(x_2 - a) \ .$$

The next property is a direct consequence of Property 2.2.2. It is a fundamental result for round norms, which will be seen to have important implications in location models. We give two proofs for this result. The first one is a quick derivation based on Property 2.2.2; while the second one goes back to basics, -the strict convexity of the unit ball of a round norm, and gives some insight into why the result does not hold for a block norm.

Property 2.2.3

Let $k(x)$ be a round norm on R^N, $N > 1$. Then $k(x)$ is strictly convex along any straight line which does not pass through the origin.

Proof:

Let 0 denote the origin, and choose any two points x_1, x_2, $x_1 \neq x_2$, such that $0 \notin L(x_1, x_2)$; i.e., the straight line through x_1 and x_2 does not pass through the origin. Clearly, $0 \notin L'(x_1, x_2)$. Thus, we can substitute $a = 0$ into relation (2.2.8) to obtain

$$k(x_0) < \lambda\, k(x_1) + (1 - \lambda) k\, (x_2) \ , \qquad (2.2.9)$$

where

$$x_0 = \lambda\, x_1 + (1 - \lambda) x_2 \ , \ \ \lambda \in (0,1) \ .$$

Hence we conclude that $k(x)$ is strictly convex along any straight line which does not pass through the origin.

Alternate Proof:

Again choose any two points x_1, x_2, $x_1 \neq x_2$, such that the straight line through x_1 and x_2 does not pass through the origin. Clearly, $x_1 \neq 0$ and $x_2 \neq 0$. We can write $x_1 = \mu_1 z_1$

and $x_2 = \mu_2 z_2$, where μ_1 and μ_2 are positive scalars, and z_1 and z_2 are intersection points of the unit contour of k with the half-lines from the origin through x_1 and x_2 respectively. Since the line through x_1 and x_2 does not pass through the origin, we must have $z_1 \neq z_2$. Without loss in generality, assume that $\mu_1 \leq \mu_2$. Then with $x_0 = \lambda x_1 + (1-\lambda)x_2$, $\lambda \in (0,1)$, $\mu' = \lambda\mu_1 + (1-\lambda)\mu_2$, and $\lambda' = \lambda\mu_1/\mu'$, we obtain

$$\begin{aligned} k(x_0) &= k[\lambda\mu_1 z_1 + (1-\lambda)\mu_2 z_2] \\ &= \mu' k[\lambda' z_1 + (1-\lambda')z_2] \\ &< \mu' \qquad \text{(relation (2.2.6))} \\ &= \lambda\mu_1 + (1-\lambda)\mu_2 \ . \end{aligned} \tag{2.2.10}$$

But $\mu_1 = k(x_1)$ and $\mu_2 = k(x_2)$, so that (2.2.10) implies

$$k(x_0) < \lambda k(x_1) + (1-\lambda)\, k(x_2) \ ,$$

ending the proof.

Returning to relation (2.2.10), let us now suppose that k is a block norm, and futhermore that x_1 and x_2 are chosen so that z_1 and z_2 are on the same facet of the unit polytope of k. In this case, $k[\lambda' z_1 + (1-\lambda')z_2] = 1$, and so, $k(x_0) = \mu' = \lambda k(x_1) + (1-\lambda)\, k(x_2)$. We see that if k is a block norm, and the half lines from the origin through x_1 and x_2 intersect the unit polytope of k on the identical facet, then k varies linearly along the line segment connecting x_1 and x_2. It follows that k is convex, but not strictly convex, along any straight line in R^N.

As noted previously, the ℓ_p distance function is widely used in location models. The following result gives additional information concerning the classification of this important function.

Property 2.2.4

$\ell_p(x)$ is a differentiable round norm for $1 < p < +\infty$.

Proof:

It is well-known that $\ell_p(x)$ is a norm on R^N for $p \geq 1$ and that it is differentiable everywhere except at the origin for $1 < p < +\infty$, (e.g., p. 14 of Juel, 1975). Hence, we need only show that relation (2.2.6) holds for $\ell_p(x)$ to complete the proof. Let x and y be any two points in R^N. Then,

$$\ell_p(x+y) = \left[\sum_{i=1}^{N} |x_i + y_i|^p\right]^{1/p}$$

$$\leq \left[\sum_{i=1}^{N} |x_i|^p\right]^{1/p} + \left[\sum_{i=1}^{N} |y_i|^p\right]^{1/p}$$

$$= \ell_p(x) + \ell_p(y) , \qquad (2.2.11)$$

from the Minkowsky inequality. Furthermore, equality in (2.2.11) occurs if, and only if,

$$x = \mu y , \qquad (2.2.12)$$

where μ is a non-negative scalar. But if z_1 and z_2 are two different points on the unit contour of $\ell_p(x)$, then clearly,

$$z_1 \neq \mu z_2 , \qquad (2.2.13)$$

for any $\mu \geq 0$. This implies that $z_1' \neq \mu z_2'$ for any $\mu \geq 0$, where $z_1' = \lambda z_1$, $z_2' = (1-\lambda)z_2$ and $\lambda \in (0,1)$. Hence,

$$\ell_p[\lambda z_1 + (1-\lambda)z_2] < \ell_p(\lambda z_1) + \ell_p[(1-\lambda)z_2]$$

$$= \lambda\, \ell_p(z_1) + (1-\lambda)\, \ell_p(z_2)$$

$$= 1 . \qquad (2.2.14)$$

Thus, relation (2.2.6) holds, and we conclude that $\ell_p(x)$ is a differentiable round norm.

The next property shows how a nondifferentiable round norm can be constructed.

Property 2.2.5

Let $k_1(x)$ denote a block norm and $k_2(x)$ a differentiable round norm, where $x \in R^N$, $N > 1$. Let $k_3(x)$ be a positive linear combination of $k_1(x)$ and $k_2(x)$; i.e.,

$$k_3(x) = a_1 k_1(x) + a_2 k_2(x), \quad a_1, a_2 > 0 \; . \tag{2.2.15}$$

Then $k_3(x)$ is a nondifferentiable round norm.

Proof:

Since a positive linear combination of norms is itself a norm, we conclude that k_3 is a norm. Since k_1 is a block norm, it is not differentiable at all x other than the origin. Clearly, the same must hold for k_3. Finally, let z_1 and z_2 be any two points on the unit contour of k_3, such that $z_1 \neq z_2$. (Note that z_1 and z_2 are not in general on the unit contour of k_1 or k_2.) Then,

$$k_3(z_1) = a_1 k_1(z_1) + a_2 k_2(z_1) = 1 \; , \quad \text{and}$$

$$k_3(z_2) = a_1 k_1(z_2) + a_2 k_2(z_2) = 1 \; .$$

Two possibilities need to be considered.

a) $z_1 = -z_2$. Then for $\lambda \in (0,1)$,

$$\begin{aligned} k_3[\lambda z_1 + (1-\lambda) z_2] &= k_3[(1-2\lambda) z_2] \\ &= |1-2\lambda| \, k_3(z_2) \\ &= |1-2\lambda| < 1 \; . \end{aligned} \tag{2.2.16a}$$

b) $z_1 \neq -z_2$; so that the line through z_1 and z_2 does not contain the origin. Then for $\lambda \in$ (0,1),

$$\begin{aligned}
k_3[\lambda z_1 + (1-\lambda)z_2] &= a_1 k_1[\lambda z_1 + (1-\lambda)z_2] + a_2 k_2[\lambda z_1 + (1-\lambda)z_2] \\
&\le a_1\{\lambda k_1(z_1) + (1-\lambda)k_1(z_2)\} + a_2 k_2[\lambda z_1 + (1-\lambda)z_2] \\
&\qquad \text{(triangle inequality)} \\
&< a_1\{\lambda k_1(z_1) + (1-\lambda)k_1(z_2)\} + a_2\{\lambda k_2(z_1) + (1-\lambda)k_2(z_2)\} \\
&\qquad \text{(relation (2.2.9) for round norms)} \\
&= \lambda\{a_1 k_1(z_1) + a_2 k_2(z_1)\} + (1-\lambda)\{a_1 k_1(z_2) + a_2 k_2(z_2)\} \\
&= \lambda + 1 - \lambda = 1 \;. \qquad (2.2.16b)
\end{aligned}$$

We see from (2.2.16a) and (2.2.16b) that k_3 satisfies relation (2.2.6), and hence is a round norm. We conclude that $k_3(x)$ is a nondifferentiable round norm.

Property 2.2.5 implies that any block norm or differentiable round norm can be considered as a limiting case of a nondifferentiable round norm. For example,

$$k_1(x) = \frac{1}{a_1} \lim_{a_2 \to 0} k_3(x) \;, \quad \text{and} \quad k_2(x) = \frac{1}{a_2} \lim_{a_1 \to 0} k_3(x) \;, \qquad (2.2.17)$$

where k_1, k_2 and k_3 satisfy (2.2.15). In this sense then, the family of nondifferentiable round norms contains the families of block and differentiable round norms. It will also be evident later that the nondifferentiable round norm exhibits characteristics peculiar to each of the other norms.

2.3 Norms Raised to a Power

As noted in the first chapter, a generalization of the standard Weber problem involves the use of distance functions raised to a power. This allows for a less restrictive cost structure in the model. If the distance function is raised to a power $t \in (0, 1)$, its associated cost component exhibits an economy of scale; if $t = 1$ there is a constant return to scale; finally, if $t > 1$ we have a diseconomy of scale. This section deals with properties of a norm raised to a power. We begin by deriving an inequality, which will be useful subsequently.

Lemma 2.3.1

If $c_i \geq 0, i = 1, ..., M$, then

$$\left[\sum_{i=1}^{M} c_i\right]^t \geq \sum_{i=1}^{M} c_i^t, \qquad \text{for } t > 1; \qquad \text{and}$$

$$\left[\sum_{i=1}^{M} c_i\right]^t \leq \sum_{i=1}^{M} c_i^t, \qquad \text{for } t < 1.$$

(For $t \leq 0$, we assume that $c_i > 0$, $\forall$ i. Otherwise there are undefined terms.) In both relations, equality holds if, and only if, all the c_i's or all but one of the c_i's are zero.

Proof:

If all the c_i's or all but one are equal to zero, it is trivially obvious that the equality sign holds in each case. Hence, we assume from this point that two or more of the c_i's are positive. For $t > 0$, we can delete the c_i's which are equal to zero and re-label the remaining ones; so that without restriction we now assume that

$$c_i > 0, \quad i = 1, ..., M, \qquad \text{and} \qquad M \geq 2.$$

Re-write the left-hand side as follows:

$$\left[\sum_{i=1}^{M} c_i\right]^t = c_1\left[\sum_{i=1}^{M} c_i\right]^{t-1} + c_2\left[\sum_{i=1}^{M} c_i\right]^{t-1} + \ldots + c_M\left[\sum_{i=1}^{M} c_i\right]^{t-1}.$$

If $t > 1$,

$$\left[\sum_{i=1}^{M} c_i\right]^{t-1} > c_j^{t-1}, \quad j = 1, ..., M;$$

so that

$$\left[\sum_{i=1}^{M} c_i\right]^t > c_1^t + c_2^t + \ldots + c_M^t, \quad \text{for } t > 1.$$

If $t < 1$,

$$\left[\sum_{i=1}^{M} c_i\right]^{t-1} < c_j^{t-1}, \quad j = 1, \ldots, M \ ;$$

so that

$$\left[\sum_{i=1}^{M} c_i\right]^{t} < c_1^t + c_2^t + \ldots + c_M^t \ , \quad \text{for } t < 1 \, .$$

Thus the lemma is proven.

Let k(x) denote any norm on R^N, and define the following function,

$$h(x) = [k(x)]^t \, , \tag{2.3.1}$$

where $t \in (-\infty, +\infty)$. We note that

$$\begin{aligned} h(\alpha x) &= [k(\alpha x)]^t \\ &= |\alpha|^t [k(x)]^t = |\alpha|^t h(x) \, , \end{aligned} \tag{2.3.2}$$

so that the homogeneity property of norms is lost unless $t = 1$. Thus, h(x) is not a norm if $t \neq 1$. The following results give additional information concerning the function h(x).

Property 2.3.1

h(x) does not necessarily satisfy the triangle inequality when $t > 1$.

Proof:

Choose three different points x_1, x_2, $x_3 \in R^N$, such that $x_1 - x_2 = \beta(x_2 - x_3)$, $\beta > 0$. In other words, a straight line connects x_1, x_2 and x_3. For this case, we have

$$k(x_1 - x_3) = k(x_1 - x_2) + k(x_2 - x_3) \, .$$

$$\begin{aligned} \therefore \ h(x_1 - x_3) &= [k(x_1 - x_2) + k(x_2 - x_3)]^t \\ &> [k(x_1 - x_2)]^t + [k(x_2 - x_3)]^t \qquad \text{(Lemma 2.3.1)} \\ &= h(x_1 - x_2) + h(x_2 - x_3) \, . \end{aligned} \tag{2.3.3}$$

Thus, the triangle inequality does not hold in this case.

Property 2.3.2

h(x) satisfies the triangle inequality for $0 < t < 1$. Furthermore, the triangle inequality is satisfied in a strict sense.

Proof:

For any three distinct points $x_1, x_2, x_3 \in R^N$, we have

$$
\begin{aligned}
h(x_1 - x_3) &= [k(x_1 - x_3)]^t \\
&\leq [k(x_1 - x_2) + k(x_2 - x_3)]^t \qquad \text{(triangle inequality)} \\
&< [k(x_1 - x_2)]^t + [k(x_2 - x_3)]^t \qquad \text{(Lemma 2.3.1)} \\
&= h(x_1 - x_2) + h(x_2 - x_3) \,.
\end{aligned}
\tag{2.3.4}
$$

Thus, the triangle inequality is satisfied, and furthermore, the inequality is strict for any distinct $x_1, x_2, x_3 \in R^N$.

The preceding two properties generalize results given by Morris (1981), in which k(x) is the ℓ_p-norm, to the case where k(x) can be any norm. In addition, we show in Property 2.3.2 that the triangle inequality is strict for $0 < t < 1$. The following result provides information concerning the classification of a norm raised to a power.

Property 2.3.3

h(x) is a metric for $0 < t < 1$.

Proof:

We verify that the properties of a metric hold. Let p, q, r denote three points in R^N.

(i) Since $k(p-q) \geq 0, \forall\ p, q \in R^N$, it follows that $h(p-q) \geq 0, \ \forall\ p, q \in R^N$.

(ii) $h(p-q) = 0 \Leftrightarrow k(p-q) = 0 \Leftrightarrow p = q$.

(iii) $h(p-q) \leq h(p-r) + h(r-q)$, $\forall\ p, q, r \in R^N$, by Property 2.3.2, (with strict inequality if p, q, r are distinct points).

(iv) $h(p-q) = [k(p-q)]^t = [k(q-p)]^t = h(q-p)$, $\forall\ p, q \in R^N$.

We conclude that h is a metric.

The above result can be generalized as follows: If $\phi(p, q)$ is a metric and $t \in (0, 1)$, then $\Omega(p, q) = [\phi(p, q)]^t$ is also a metric. We show that $\Omega(p, q)$ satisfies the triangle inequality, the remaining properties of a metric being easily verified.

$$\begin{aligned}\Omega(p, q) &= [\phi(p, q)]^t \\ &\leq [\phi(p, r) + \phi(r, q)]^t && \text{(triangle inequality for the metric } \phi) \\ &\leq [\phi(p, r)]^t + [\phi(r, q)]^t && \text{(Lemma 2.3.1)} \\ &= \Omega(p, r) + \Omega(r, q) .\end{aligned}$$

The next property is taken from Pelegrin, Michelot and Plastria (1985). Since they do not give a proof, we provide one for completeness. Also recall that their S-norm is identical to our round norm.

Property 2.3.4

If g is a nondecreasing strictly convex function, and k is a round norm on R^N, then $g(k(x))$ is a strictly convex function of x.

Proof:

First note that $g(u)$, $u \in R^1$, must be an increasing function of u. We see this as follows: Suppose $u_1 < u_2$ and $g(u_1) = g(u_2)$. Then by the strict convexity of g,

$$g(\lambda u_1 + (1-\lambda)u_2) < \lambda g(u_1) + (1-\lambda)g(u_2) = g(u_1), \quad \text{for } 0 < \lambda < 1,$$

which contradicts the nondecreasing property of g. Hence, we must have $g(u_1) < g(u_2)$ for $u_1 < u_2$. Now choose any two points $x_1, x_2 \in R^N$, such that $x_1 \neq x_2$, and $x_1 \neq 0$ (the origin). There are two possibilities to consider.

(i) $x_2 = \beta x_1$, where $\beta \geq 0$ ($\beta \neq 1$).

Letting $x_0 = \lambda x_1 + (1-\lambda)x_2$, $\lambda \in (0, 1)$, we get

$$\begin{aligned} g(k(x_0)) &= g(k[\lambda x_1 + (1-\lambda)x_2]) \\ &= g[(\lambda + (1-\lambda)\beta)k(x_1)] \\ &= g[\lambda k(x_1) + (1-\lambda)k(x_2)] \\ &< \lambda g(k(x_1)) + (1-\lambda)g(k(x_2)), \end{aligned} \tag{2.3.5}$$

since $k(x_1) \neq k(x_2)$ and g is a strictly convex function.

(ii) $x_2 \neq \beta x_1$, where $\beta \geq 0$.

From Property 2.2.2 with $a = 0$, it follows that $k(x_0) < \lambda k(x_1) + (1-\lambda)k(x_2)$. Since g is an increasing strictly convex function, we get

$$\begin{aligned} g(k(x_0)) &< g[\lambda k(x_1) + (1-\lambda)k(x_2)] \\ &\leq \lambda g(k(x_1)) + (1-\lambda)g(k(x_2)), \end{aligned} \tag{2.3.6}$$

with equality in the last line if, and only if, $k(x_1) = k(x_2)$. Combining (2.3.5) and (2.3.6), we conclude that $g(k(x))$ is a strictly convex function of x.

As a consequence of Property 2.3.4, we get the following result.

Property 2.3.5

If $k(x)$ is a round norm and $t > 1$, then $h(x)$ is a strictly convex function of x.

Proof:

Note that $h(x) = g(k(x))$, where the function g is given by

$$g(u) = u^t . \tag{2.3.7}$$

Since g is an increasing strictly convex function for $0 \leq u < +\infty$ and $t > 1$, we immediately conclude from Property 2.3.4 that h is a strictly convex function of x.

Unfortunately, the Properties 2.3.4 and 2.3.5 do not extend to the case where k is a block norm. This is due to the fact that k now has polytope contours (polygon contours in R^2). Since k(x) is constant along the facets of these polytopes, so is g(k(x)), and thus g(k(x)) cannot be a strictly convex function of x. However, Property 2.3.4 can be modified for the case where k is a block norm as follows.

Property 2.3.6

If g is a nondecreasing strictly convex function, and k is a block norm on R^N, then g(k(x)) is a strictly convex function of x along any straight line which is not tangent to a facet of some contour of k. Otherwise g(k(x)) is a convex function of x.

Proof:

Let x_1, x_2 be any two points in R^N, such that the straight line through x_1 and x_2 ($L(x_1, x_2)$) is not tangent to a facet of some contour of k. There are two possibilities to consider.

(i) $k(x_1) \neq k(x_2)$.

Letting $x_0 = \lambda x_1 + (1-\lambda)x_2$, $\lambda \in (0, 1)$, we get

$$g(k(x_0)) = g(k[\lambda x_1 + (1-\lambda)x_2])$$

$$\leq g[\lambda k(x_1) + (1-\lambda)k(x_2)]$$ (using the triangle inequality and the homogeneity property for k, and the nondecreasing property of g)

$$< \lambda g(k(x_1)) + (1-\lambda)\, g(k(x_2)), \tag{2.3.8}$$

since g is a strictly convex function.

(ii) $k(x_1) = k(x_2)$.

Then x_1 and x_2 are on different facets of some polytope contour of k. It immediately follows that

$$k(x_0) < \lambda k(x_1) + (1-\lambda)k(x_2) .$$

In Property 2.3.4, we showed that g must be an increasing function. Thus,

$$g(k(x_0)) < g[\lambda k(x_1) + (1-\lambda)k(x_2)]$$

$$= \lambda g(k(x_1)) + (1-\lambda)g(k(x_2)) . \tag{2.3.9}$$

Combining (2.3.8) and (2.3.9), we see that the first part of the property is proven. Noting that an increasing convex function of the convex function k(x) is itself convex in x proves the second part.

Analogous to Property 2.3.5, we obtain the following special case.

Property 2.3.7

If k(x) is a block norm and $t > 1$, then h(x) is a strictly convex function of x along any straight line which is not tangent to a facet of some contour of k. Otherwise, h(x) is a convex function of x.

2.4 Differentiability and Directional Derivatives

It is usually desirable in optimization models to deal with functions which are differentiable everywhere. Unfortunately, norms are not differentiable at the origin. If k is a norm on R^N, then by the homogeneity property,

$$k(\alpha x) = |\alpha| k(x)$$

where α is any scalar value. Thus, if we plot k as a function of x along any straight line passing through the origin, the slope will have constant magnitude (>0), but opposite sign on each side of the origin. This V-shape is illustrated in Figure 2.4.1(a). Clearly then, the deri-

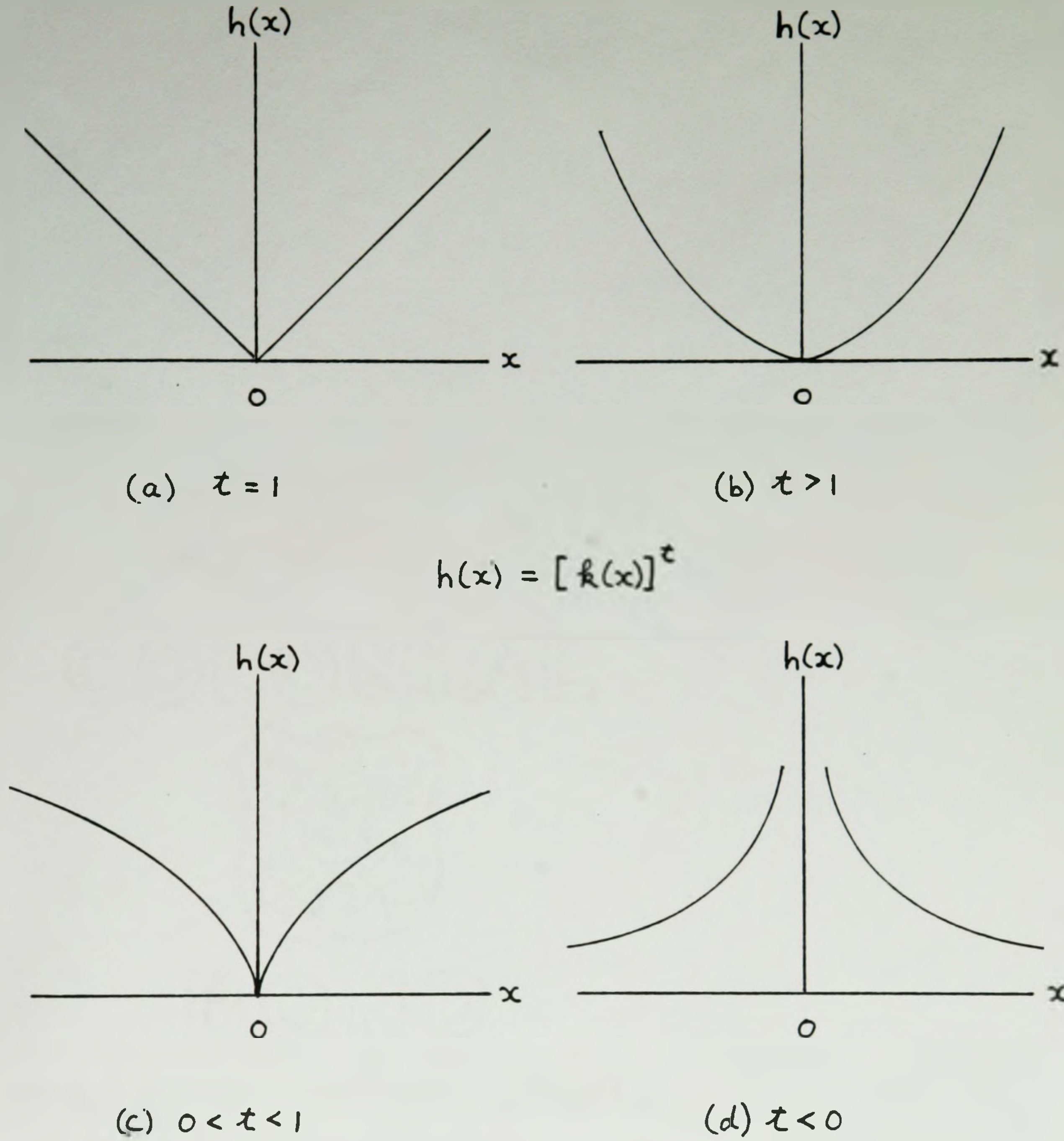


Figure 2.4.1 Profiles of h(x) along a Straight Line in R^N Passing through the Origin.

vatives are not defined at the origin. In addition, if k is a block norm or a nondifferentiable round norm, it has infinitely many other points in R^N where it cannot be differentiated.

To partially circumvent these difficulties, we use the concept of one-sided directional derivatives, defined as follows (e.g., see p. 387 of Kreyszig, 1983): Let f be a real-valued function on R^N, and let x and y be N-vectors, with y having unit length. Then the one-sided directional derivative of the function f at the point x in the direction y is given by

$$f'(x\,;y) = \lim_{\delta \to 0^+} \left[\frac{f(x + \delta y) - f(x)}{\delta} \right] , \tag{2.4.1}$$

if this limit exists.

We see from (2.4.1) that $f'(x\,;y)$ gives the marginal rate of increase of f at point x in the direction y. The term one-sided refers to the fact that the limit is taken as the real variable δ approaches 0 through positive values (i.e., from the right). Since we shall deal only with directional derivatives that are one-sided, this term will be omitted. Furthermore, we shall only consider continuous functions, so that the limit in (2.4.1) will always exist. Thus, the proviso at the end of the definition can also be omitted. Finally we note that the restriction of y to unit length means that

$$\ell_2(y) = \left[\sum_{i=1}^{N} y_i^2 \right]^{1/2} = 1 , \tag{2.4.2}$$

where $y = (y_1, ..., y_N)^T$. If this restriction is deleted, then $\ell_2(y)$ has to be included in the denominator of the expression for which the limit is being taken in (2.4.1).

If f is differentiable at the point x, the directional derivative here is equivalent to the total derivative; i.e.,

$$f'(x\,;y) = \nabla f(x) \cdot y \;, \tag{2.4.3}$$

where $\nabla f(x)$ denotes the gradient vector of f at x, and the "$\cdot$" signifies the inner product (or dot product) of two N-vectors.

We now give a fundamental result for optimization models in which a convex objective function is to be minimized. This result is taken from Shapiro (1979, p. 361), who also provides a proof.

Property 2.4.1

Let f be a real-valued function on R^N, and let x and y be N-vectors, with y having unit length. If f is a convex function, then f has a global minimum at x if, and only if, $f'(x\,;y) \geq 0$ for all directions y.

Alternatively, we can say that the convex function f has a global minimum at x if, and only if,

$$\min_{y} f'(x\,;y) \geq 0\,. \tag{2.4.4}$$

That this is a necessary condition is obvious. If y_* denotes the direction which minimizes $f'(x\,;y)$, and if $f'(x\,;y_*) < 0$, then the function f is decreasing at x in the direction y_* over some finite length. Hence, x cannot be a local (or global) minimum. That relation (2.4.4) is a sufficient condition can be seen as follows. If $\min_y f'(x\,;y) \geq 0$, then by the convexity of f, the point x must be a local minimum. Furthermore, a local minimum of a convex function is also a global one. It is interesting to note that if f is differentiable at x, the relation (2.4.4) reduces to

$$\nabla f(x) = 0\,, \tag{2.4.5}$$

which is the first-order condition defining stationary points of a differentiable function.

Let us now consider any norm k on R^N. It is a well-known result that the directional derivative at the origin is given by

$$k'(0\,;y) = k(y)\,, \tag{2.4.6}$$

(e.g., Juel, 1975, Juel and Love, 1981). This follows readily from the definition of the directional derivative given in (2.4.1).

$$k'(0\,;y) = \lim_{\delta \to 0^+} \left\{ \frac{k(\delta y) - k(0)}{\delta} \right\}$$

$$= \lim_{\delta \to 0^+} \left\{ \frac{\delta\, k(y)}{\delta} \right\} = k(y)\,.$$

If k is a differentiable round norm, it is differentiable everywhere except at the origin; so that for $x \neq 0$,

$$k'(x\,;y) = \nabla k(x) \cdot y\,. \tag{2.4.7}$$

Let us extend the results of the preceding paragraph to functions of a norm. Consider then the case where

$$f(x) = g(k(x))\,. \tag{2.4.8}$$

Here, k is any norm on R^N, and g(u) is a differentiable function for $u \in [0, +\infty)$. By means of the chain rule of calculus, we obtain

$$f'(x\,;y) = g'(k(x))\, k'(x\,;y)\,, \tag{2.4.9}$$

where

$$g'(u) = \frac{dg(u)}{du}\,.$$

Then for $x = 0$, we get

$$f'(0\,;y) = g'(k(0))\, k'(0\,;y) = g'(0)\; k(y)\,, \tag{2.4.10}$$

where use is made of (2.4.6) and the fact that $k(0) = 0$. Note also that $g'(0)$ is the right-sided derivative of g evaluated at 0, since the argument k(x) is non-negative. Furthermore, if k is a differentiable round norm and $x \neq 0$, then

$$f'(x\,;y) = g'(k(x))\, \nabla k(x) \cdot y\,. \tag{2.4.11}$$

As an example, let us return to the function $h(x) = [k(x)]^t$. Here, we have $f = h$ and $g(u) = u^t$. Noting that $g'(0) = 0$ if $t > 1$ and $+\infty$ if $0 < t < 1$, and applying equation (2.4.10) gives

$$h'(0\,;y) = \begin{cases} 0\,, & \text{if } t > 1\,, \\ +\infty\,, & \text{if } 0 < t < 1\,. \end{cases} \tag{2.4.12}$$

The shape of h as a function of x along a straight line through the origin is illustrated in Figures 2.4.1(b) and (c) for the ranges of t above. Equation (2.4.12) generalizes the results obtained by Love and Morris (1978) for the case where k is the Euclidean distance to the case where k can be any norm. Our method is also more concise than theirs.

For the case where $t < 0$, the function h(x) becomes unbounded as x approaches the origin; i.e.,

$$\lim_{x \to 0} h(x) = +\infty, \qquad t < 0 . \tag{2.4.13}$$

This is illustrated in Figure 2.4.1(d). Thus, the directional derivative is undefined at $x = 0$. An infinite cost is associated with h(x) at $x = 0$, and the magnitude of h(x) decreases as the distance k(x) increases. Thus, the use of negative t does not make sense in standard minisum location models. It is not surprising then that t is restricted to positive values in the literature. However, it is interesting to note that for the location of a noxious facility, where the objective is to maximize the minimum distance to a set of fixed points subject to a set of constraints, one might consider instead a minisum criterion with negative t. Such a model would take the form,

$$\begin{aligned} \underset{x}{\text{minimize}}\ W_{NOX}(x) &= \sum_{i=1}^{n} w_i\, h(x - a_i) \\ &= \sum_{i=1}^{n} w_i\, [k(x - a_i)]^t, \quad t < 0, \end{aligned}$$

where the w_i are positive weights, the a_i are the fixed points, and the same set of location constraints apply. The use of such a model in practical situations involving the location of noxious facilities should be of some interest, as an alternative criterion. As an example, consider the location of a polluting facility such as a smoke-stack, where the amount of pollution varies inversely as the distance from the facility. In this case, $t = -1$.

We use (2.4.12) now to obtain some interesting results for norms raised to a power.

Property 2.4.2

Consider the function $h(x) = [k(x)]^t$, where k is a norm on R^N and $t > 1$. Then h(x) is differentiable at the origin, with

$$\nabla h(0) = 0 \ . \tag{2.4.14}$$

Furthermore, if k is a differentiable round norm, then h is differentiable everywhere.

Proof:

From (2.4.12) we see that $h'(0\,;y) = 0$ for all directions y. It immediately follows that h is differentiable at $x = 0$, with $\nabla h(0) = 0$. If in addition, k is a differentiable round norm, then by (2.4.11), $h'(x;y) = t[k(x)]^{t-1}\,\nabla k(x)\cdot y = \nabla h(x)\cdot y$ for all $x \neq 0$ and all y. Thus h is differentiable everywhere in this case.

Property 2.4.3

Consider the function $h(x) = [k(x)]^t$, where k is a norm on R^N and $t < 1$, $(t \neq 0)$. Then h(x) is not differentiable at the origin. Furthermore, h(x) is neither a convex nor concave function of x.

Proof:

For the case $0 < t < 1$, we have $h'(0\,;y) = +\infty$ from (2.4.12), for all directions y. Also h is a finite-valued function of x with $h(0) = 0$. It follows that h is not differentiable at the origin, and that it cannot be convex or concave in x. For the case $t < 0$, it is immediately obvious from (2.4.13) that the same conclusion holds.

From properties (2.3.5) and (2.4.2), we see that h(x) is a strictly convex differentiable function of x, if $t > 1$ and k is a differentiable round norm. This result has practical implications on the optimization of continuous location models. Consider for example, the minisum objective function,

$$W_G(x) = \sum_{i=1}^{n} w_i\, h(x - a_i) ,$$

where the w_i are positive weights and the a_i are fixed points, $i = 1, \ldots, n$. Then W_G is a positive linear sum of strictly convex differentiable terms, and so, is itself a strictly convex differentiable function of x, if $t > 1$ and k is a differentiable round norm. Thus, the optimal location can be found by standard descent techniques. Furthermore, this location will be unique for any set of a_i's (collinear or not). If the minimization is constrained, a convex programming technique will solve it.

Another practical implication is that h(x) can be used as a "smoothing" approximation of a differentiable round norm k(x), by choosing a value of t slightly larger than one; i.e.,

$$t = 1 + \varepsilon,$$

where $0 < \varepsilon << 1$. This provides an alternative to the well-known hyperbolic and hyperboloid approximations used extensively in the literature (e.g., Love and Morris, 1975b, Morris and Verdini, 1979, and Eyster, White and Wierwille, 1973). We illustrate this concept in Figure 2.4.2 for the one-dimensional case, where the hyperbolic approximation $\sqrt{(u^2+\varepsilon)}$ of $|u|$ is compared with our smoothing function $|u|^{1+\varepsilon}$. Note that our approximation is significantly better near the origin ($u=0$), but it becomes inaccurate with $|u|$ sufficiently large.

If k is a block norm or a nondifferentiable round norm, there exist points x other than the origin where k is not differentiable. The next property characterizes this set of points.

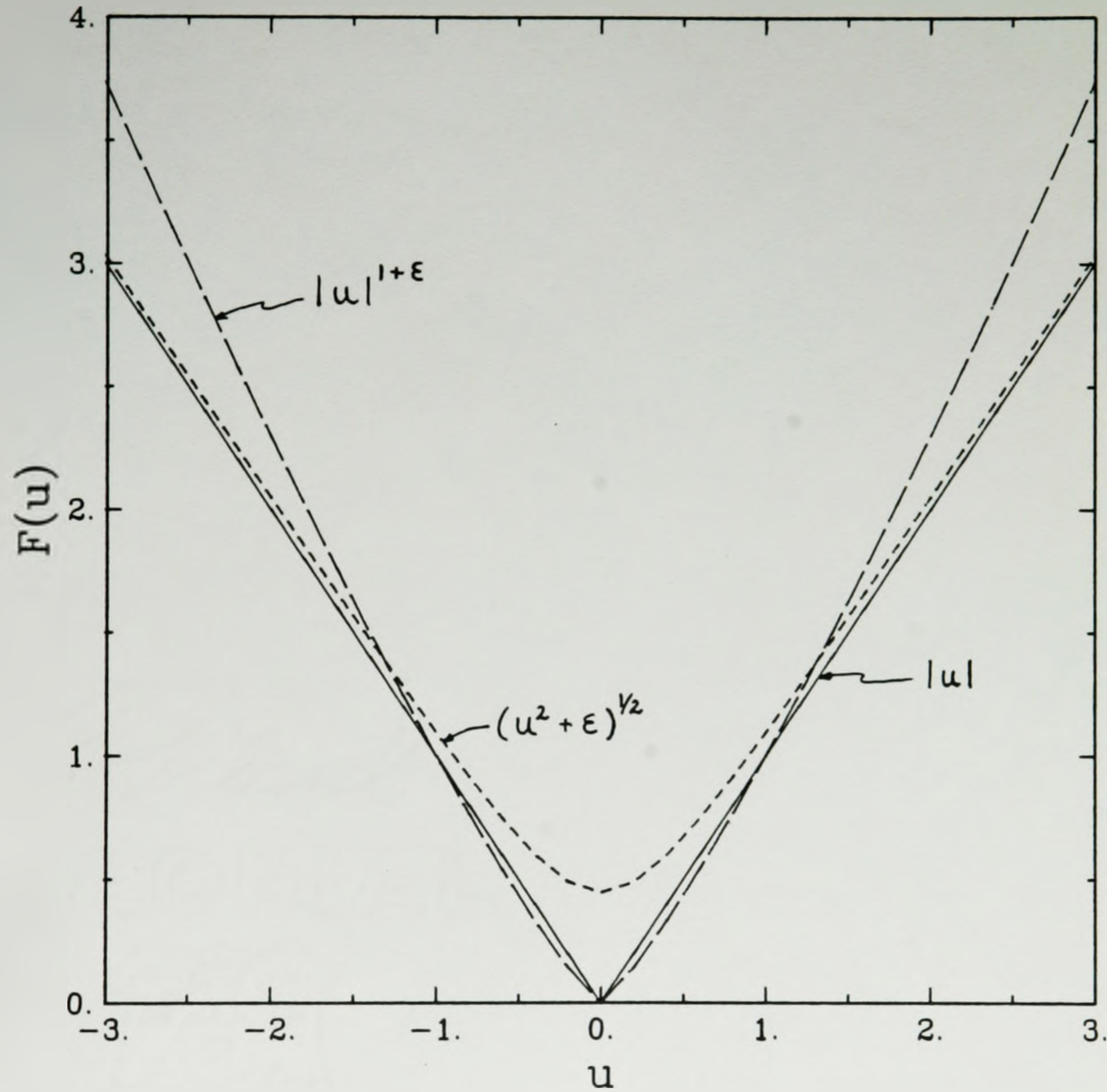


Figure 2.4.2 Smoothing Approximations of the Norm k in One Dimension.

Property 2.4.4

Let k be a norm on R^N which is not differentiable at all points other than the origin, and let S denote the set of points where k is not differentiable. Consider any $x_0 \in S$, $x_0 \neq 0$. Then

$$L_0 \subseteq S, \tag{2.4.15}$$

where L_0 is the straight line passing through the origin and x_0.

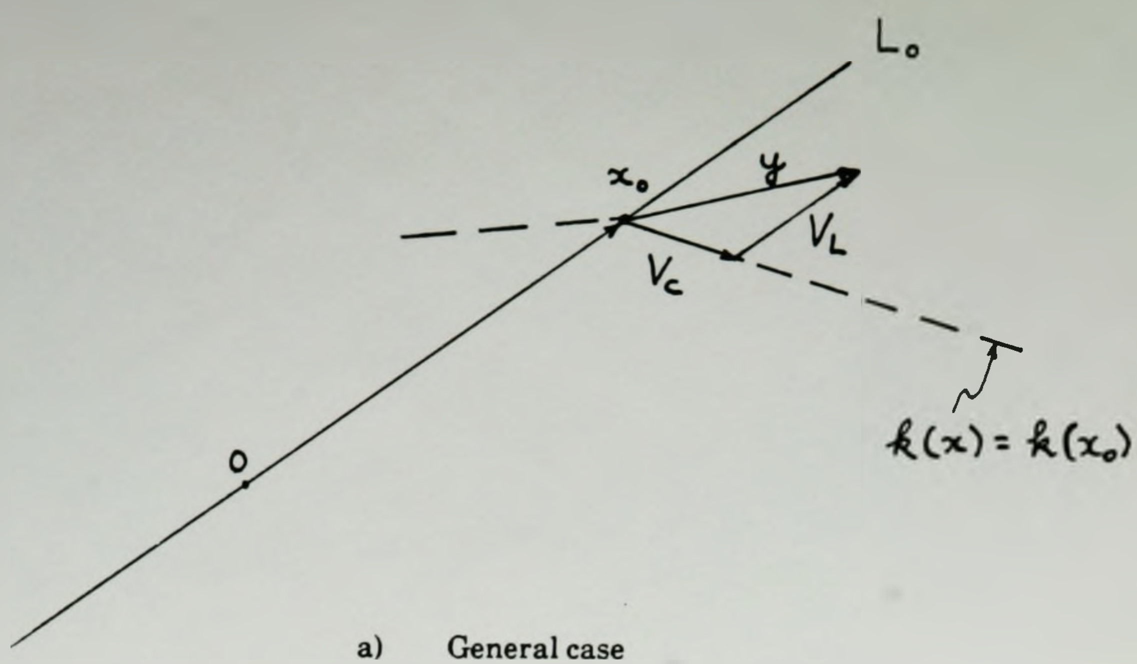
Proof:

We begin by noting that $0 \in S$ for all norms. For the given norm k, we can also choose an $x_0 \in S$ such that $x_0 \neq 0$. Consider the change of variables, $v = x/\mu$, where μ is a non-zero constant. Since k(x) is not differentiable at x_0, it follows that $f(v) = k(\mu v)$ is not differentiable at $v_0 = x_0/\mu$. But $k(\mu v) = |\mu|\, k(v)$, so that k(v) cannot be differentiable at v_0. We conclude that the straight line through the origin and x_0 belongs to the set S.

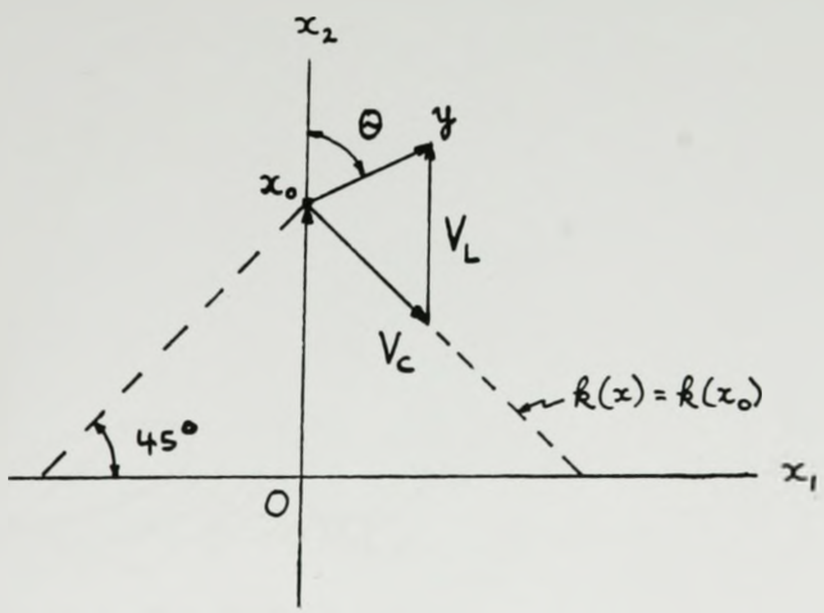
The directional derivative of a norm k(x) at $x = 0$ is given in equation (2.4.6). We would like to extend this result to the other non-differentiable points of k(x). This would be of interest for block norms and nondifferentiable round norms. From Property 2.4.4, we know that these points form straight lines through the origin. Referring to Figure 2.4.3(a), let L_0 denote such a line, and let us calculate the directional derivative at any $x_0 \in L_0$ ($x_0 \neq 0$), in the direction of the unit vector y. The unit vector y can be represented by the unique sum of two vectors, V_L and V_C, where V_L is parallel to L_0 and V_C is tangent to the contour of k at x_0; (see Figure 2.4.3(a)). Thus,

$$y = V_L + V_C; \tag{2.4.16}$$

so that the directional derivative of k at x_0 in the direction y is given by



a) General case



b) Rectangular norm

Figure 2.4.3 Directional Derivative Calculation at a Non-Differentiable Point.

$$k'(x_0;y) = \lim_{\delta \to 0^+} \left\{ \frac{k(x_0+\delta y) - k(x_0)}{\delta} \right\}$$

$$= \lim_{\delta \to 0^+} \left\{ \frac{k(x_0+\delta V_L+\delta V_C) - k(x_0)}{\delta} \right\}$$

$$= \lim_{\delta \to 0^+} \left\{ \frac{k(x_0+\delta V_L) - k(x_0)}{\delta} \right\},$$

since

$$k(x_0+\delta V_L+\delta V_C) = k(x_0+\delta V_L) + O(\delta^2).$$

But

$$k(x_0+\delta V_L) = \begin{cases} k(x_0)-\delta k(V_L), & \text{if } x_0 \text{ and } V_L \text{ have opposite directions}, \\ k(x_0)+\delta k(V_L), & \text{if } x_0 \text{ and } V_L \text{ have the same direction}, \end{cases}$$

where δ is sufficiently small, and use is made of the fact that $k(v_1+v_2) = k(v_1)+k(v_2)$ if $v_2 = \beta v_1, \beta > 0$. We see then that

$$k'(x_0;y) = \begin{cases} -k(V_L), & \text{if } x_0 \text{ and } V_L \text{ have opposite directions}, \\ +k(V_L), & \text{if } x_0 \text{ and } V_L \text{ have the same direction}. \end{cases} \quad (2.4.17)$$

As an example, consider the case where $k(x)$ is the rectangular norm ($\ell_1(x)$) on R^2. The set of points at which $k(x)$ is not differentiable is given by

$$S = \{x \mid x_1 = 0 \quad \text{or} \quad x_2 = 0\},$$

where $x = (x_1, x_2)^T$. That is, the set S consists of the points on the vertical and horizontal axes. Let x_0 be a point on the vertical axis other than the origin, and let θ be the smallest angle at x_0 between the vertical axis pointing away from the origin and the unit direction vector y; (see Figure 2.4.3(b)). Then

$$k(V_L) = \begin{cases} -(\cos\theta+\sin\theta), & \text{if } 135° \le \theta \le 180°, \\ +(\cos\theta+\sin\theta), & \text{if } 0 \le \theta \le 135°. \end{cases}$$

We see from (2.4.17) that

$$k'(x_0;y) = \cos\theta+\sin\theta, \qquad 0 \le \theta \le 180°. \quad (2.4.18)$$

With $y = (y_1, y_2)^T$, we have $\cos\theta = y_2$ and $\sin\theta = |y_1|$, so that (2.4.18) can alternatively be written as

$$k'(x_0; y) = |y_1| + y_2 . \tag{2.4.19}$$

Consider now a minisum location model with fixed points or customers, a_i, $i = 1, \ldots, n$. Each customer a_i has a norm of the form $k_i(x - a_i)$ associated with it. The norm $k_i(x - a_i)$ as a function of x is the same as $k_i(x)$ translated or shifted by the vector a_i. We see that the origin associated with $k_i(x - a_i)$ is at $x = a_i$ $(x - a_i = 0)$. Furthermore, if k_i is a block norm or nondifferentiable round norm, then the straight lines (L_0) containing the non-differentiable points x of $k_i(x - a_i)$ will pass through a_i, (see Property 2.4.4).

Directional derivatives are used in minisum location models to determine optimality criteria at the fixed points (e.g., Juel and Love, 1981, and Juel, 1983). We shall see later that they can be applied to points other than the fixed locations, for block norms and nondifferentiable round norms, to obtain additional optimality criteria at the intersection points. These intersection points occur in R^2 where two (or more) non-differentiable lines (L_0) from different fixed locations cross.

CHAPTER 3

THE ℓ_p NORM

As noted in Chapter 1, the ℓ_p norm is used extensively in the literature on continuous location theory. The most popular distances - rectangular and Euclidean - are specific examples of this norm with $p = 1$ and 2 respectively. In this chapter, we investigate several properties of the ℓ_p norm which are related directly or indirectly to the application of this function in approximating travel distances of road networks.

First we define a directional bias function for norms in general, which is subsequently used to study the directional bias of the ℓ_p norm in detail. An important relation between ℓ_p distances with $1 \leq p \leq 2$ and those with $2 \leq p \leq +\infty$ is established. We then derive some properties of the ℓ_p norm multiplied by an inflation factor, which pertain to certain fitting criteria applied in the literature. These properties should be useful in simplifying the search for the best-fitting parameter values. Finally, we discuss a general procedure for fitting this distance function to actual road data, and illustrate the approach through a case study.

Throughout this chapter, attention is restricted to distance functions acting on the plane (R^2), since this is the most common case occurring in practice. However, the properties given here can be extended in straightforward fashion to higher-dimensional spaces. Also note that the reference axes are always assumed to be mutually orthogonal unless otherwise stated.

3.1 Directional Bias

For any norm k on R^2, we have the following fundamental result.

Property 3.1.1

The dimensionless ratio,

$$r = \frac{k(x)}{\ell_2(x)} \ , \ x \neq 0 \ , \tag{3.1.1}$$

is a function of θ alone, where θ is the angle specifying the vector $x = (x_1, x_2)^T$; that is,

$$\theta = \tan^{-1}(x_2 / x_1) \ . \tag{3.1.2}$$

(See Figure 3.1.1a.)

Proof:

For the points $x \neq 0$ on any half-line ending at the origin, we have

$$k(x) = c\,\ell_2(x),$$

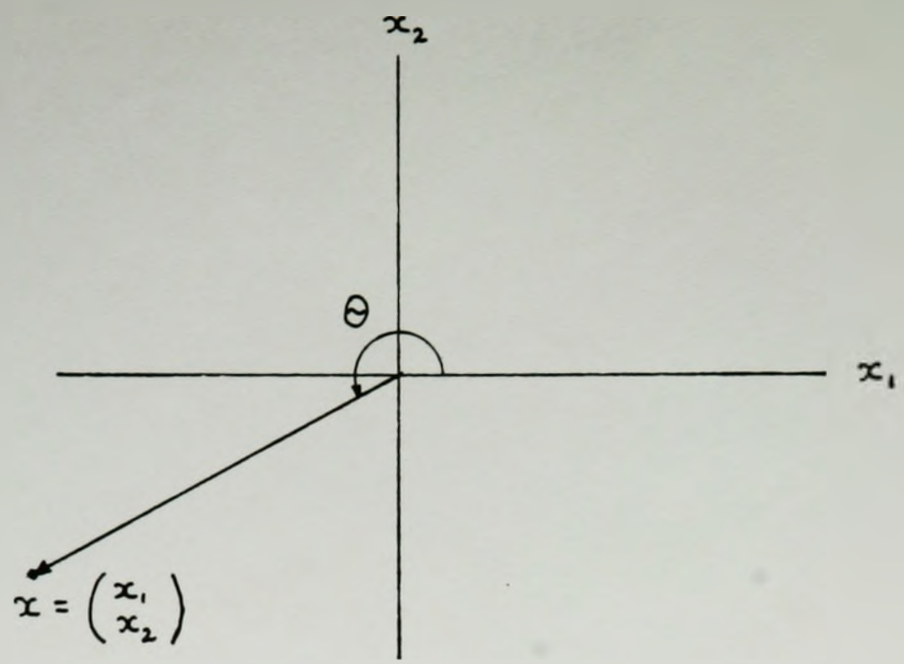
where c is a constant. It immediately follows that $r = r(\theta)$.

We shall call $r(\theta)$ the directional bias function of the norm k. In a qualitative sense, this function can be thought of as a measure of the relative difficulty of travel in any direction. If $r(\theta_1) > r(\theta_0)$, then one must travel a longer distance along a line at angle θ_1 with the x_1-axis than along a line at angle θ_0 with the x_1-axis, to cover the same Euclidean distance between pairs of points. In the physical world, the shortest possible path between two points is the straight-line or Euclidean path. Hence, for norms used to approximate actual travel distances, the directional bias function should satisfy the following relation,

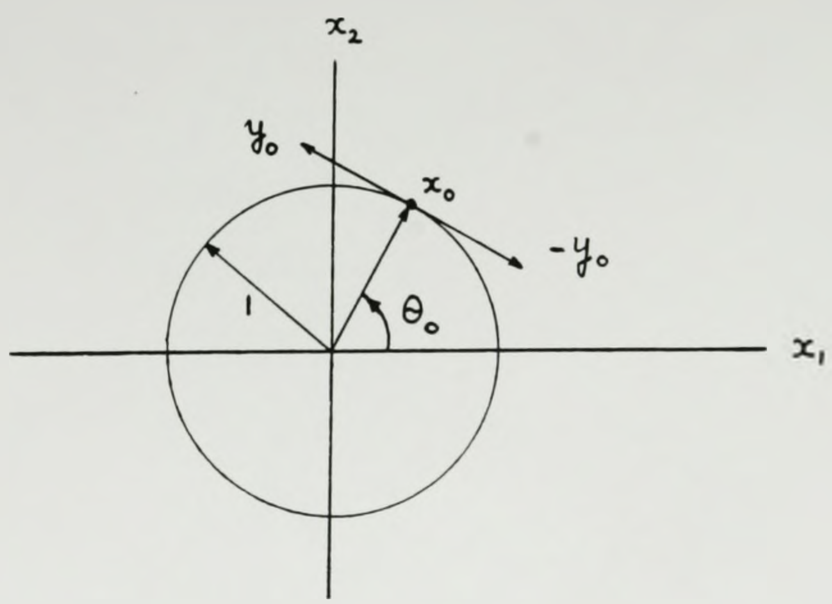
$$r(\theta) \geq 1 \ , \ \forall\, \theta \ . \tag{3.1.3}$$

Otherwise, distances shorter than Euclidean are possible. The differentiability of r at a specified angle θ_0 depends on the differentiability of k at x_0 (Figure 3.1.1b). We shall discuss this relationship in more detail later on.

The traditional method of illustrating and comparing the directional bias of norms is by means of the unit circle (e.g., see Figure 10.1 in Love, Morris and Wesolowsky, 1988). The function $r(\theta)$ provides a formal definition of directional bias, and a new way of



a) Constant along any specified orientation θ



b) Differentiability dependent on type of norm

Figure 3.1.1 Determining the Directional Bias Function of a Norm.

representing it graphically which we believe is more informative and easier to interpret than the unit circle, since the relevant information is now contained in a standard plot of a function of one independent variable (θ).

From the symmetry property of norms, it follows that

$$r(\theta) = \frac{k(x)}{\ell_2(x)}$$

$$= \frac{k(-x)}{\ell_2(-x)} = r(\theta + \pi) . \qquad (3.1.4)$$

Thus, $r(\theta)$ has a periodicity of π/n where n must be some integer greater than or equal to one.

Let us consider now the directional bias of the ℓ_p norm, denoted by $r_p(\theta)$ where $p \geq 1$. Using the definition in (3.1.1), we obtain

$$r_p(\theta) = \frac{\ell_p(x)}{\ell_2(x)} = \frac{1}{\ell_2(x)} \left[|x_1|^p + |x_2|^p \right]^{1/p} . \qquad (3.1.5)$$

Since

$$\cos\theta = \frac{x_1}{\ell_2(x)} \text{ and } \sin\theta = \frac{x_2}{\ell_2(x)} , \qquad (3.1.6)$$

equation (3.1.5) can be rewritten in the form,

$$r_p(\theta) = \left[\left(\frac{|x_1|}{\ell_2(x)} \right)^p + \left(\frac{|x_2|}{\ell_2(x)} \right)^p \right]^{1/p}$$

$$= \left[|\cos\theta|^p + |\sin\theta|^p \right]^{1/p} . \qquad (3.1.7)$$

Alternatively, we see that the ℓ_p norm can be expressed in terms of the Euclidean distance and the angle of travel (θ) as follows:

$$\ell_p(x) = \ell_2(x) \left[|\cos\theta|^p + |\sin\theta|^p \right]^{1/p} . \qquad (3.1.8)$$

Examples of the directional bias function for different values of p include

$$r_1(\theta) = |\cos\theta| + |\sin\theta| , \qquad (3.1.9)$$

and

$$r_2(\theta) = \left[|\cos\theta|^2 + |\sin\theta|^2 \right]^{1/2} = 1 , \tag{3.1.10}$$

for the rectangular and Euclidean norms respectively.

Some useful properties of $r_p(\theta)$ are derived below.

Property 3.1.2

$r_p(\theta)$ is periodic with period $\pi/2$ ($=90°$).

Proof:

$$\begin{aligned} r_p(\theta + \pi/2) &= \left[|\cos(\theta + \pi/2)|^p + |\sin(\theta + \pi/2)|^p \right]^{1/p} \\ &= \left[|-\sin\theta|^p + |\cos\theta|^p \right]^{1/p} \\ &= \left[|\cos\theta|^p + |\sin\theta|^p \right]^{1/p} = r_p(\theta) . \end{aligned} \tag{3.1.11}$$

Property 3.1.3

For any real Ω,

$$r_p(\pi/4 + \Omega) = r_p(\pi/4 - \Omega) . \tag{3.1.12}$$

Proof:

This follows immediately from the observation that $\cos(\pi/4 + \Omega) = \sin(\pi/4 - \Omega)$ and $\sin(\pi/4 + \Omega) = \cos(\pi/4 - \Omega)$.

From the two preceding results, we see that $r_p(\theta)$ is the mirror image of itself about the line $\theta = \pi/4$, and that this function has a period of $\pi/2$. Hence, it is only necessary to consider θ in the interval $[0, \pi/4]$, (i.e., 0 to 45°). Noting that $|\cos\theta| = \cos\theta$ and $|\sin\theta| = \sin\theta$ for $\theta \in [0, \pi/2]$, we readily obtain the following expressions for the first and second-order derivatives of $r_p(\theta)$:

$$\frac{dr_p(\theta)}{d\theta} = \frac{\sin 2\theta}{2[r_p(\theta)]^{p-1}} \cdot \left(-\cos^{p-2}\theta + \sin^{p-2}\theta \right) , \tag{3.1.13}$$

and

$$\frac{d^2 r_p(\theta)}{d\theta^2} = -r_p(\theta) + (p-1)\,[r_p(\theta)]^{1-2p}\,\frac{(\sin 2\theta)^{p-2}}{2^{p-2}} , \tag{3.1.14}$$

where $0 < \theta < \pi/2$.

Property 3.1.4

In the interval $0 \leq \theta \leq \pi/4$, r_p is a strictly increasing function of θ if $0 < p < 2$, while it is strictly decreasing in θ if $p > 2$.

Proof:

For $0 < \theta < \pi/4$, we have $\cos\theta > \sin\theta$; so that $\cos^{p-2}\theta > \sin^{p-2}\theta$ if $p > 2$, while $\cos^{p-2}\theta < \sin^{p-2}\theta$ if $p < 2$. From (3.1.13) it follows that for $0 < \theta < \pi/4$,

$$\frac{dr_p}{d\theta} > 0 , \text{ if } p < 2 \ (p \neq 0) , \tag{3.1.15a}$$

and

$$\frac{dr_p}{d\theta} < 0 , \text{ if } p > 2 . \tag{3.1.15b}$$

Hence, the property is proven.

Property 3.1.5

If $p > 1$ and $p \neq 2$, then r_p has a unique inflection point (θ_*) in the interval $0 \leq \theta \leq \pi/4$.

Proof:

First consider the case $1 < p < 2$. Since r_p is strictly increasing (Property 3.1.4) and so is $\sin 2\theta$, we see from (3.1.14) that $d^2r_p/d\theta^2$ is the sum of two strictly decreasing terms. Hence, $d^2r_p/d\theta^2$ is strictly decreasing in $\theta \in (0, \pi/4)$. Furthermore,

$$\lim_{\theta \to 0^+} \frac{d^2 r_p}{d\theta^2} = +\infty , \quad (1 < p < 2) , \tag{3.1.16}$$

and

$$\frac{d^2 r_p(\pi/4)}{d\theta^2} = (p-2)\, 2^{\frac{1}{p}-\frac{1}{2}} < 0 , (1 < p < 2) . \tag{3.1.17}$$

Therefore, a unique θ_* exists such that $d^2r_p(\theta_*) / d\theta^2 = 0$, and θ_* is an inflection point.

Now consider $p > 2$. Since r_p is strictly decreasing (Property 3.1.4), it follows that $d^2r_p/d\theta^2$ is the sum of two strictly increasing terms. Hence, $d^2r_p/d\theta^2$ is strictly increasing in $\theta \in (0, \pi/4)$. Furthermore,

$$\frac{d^2 r_p(0)}{d\theta^2} = -1 , \quad (p > 2) , \tag{3.1.18}$$

and using (3.1.17),

$$\frac{d^2 r_p(\pi/4)}{d\theta^2} > 0 , \quad (p > 2) . \tag{3.1.19}$$

Once again we conclude that a unique inflection point θ_* exists with $d^2r_p(\theta_*)/d\theta^2 = 0$, thereby ending the proof.

The shape of $r_p(\theta)$ is illustrated in Figure 3.1.2 for various values of p, and for θ in the range $[0,\pi/2]$, i.e., one complete cycle (Property 3.1.2). From Properties 3.1.3 and 3.1.4, it follows that r_p has its maximum value at $\theta = \pi/4$ and minimum value at $\theta = 0, \pi/2$, if $0 < p < 2$, while the converse holds if $p > 2$. Defining the direction of greatest (least) difficulty as the value of θ which maximizes (minimizes) r_p, we see that for $0 < p < 2$ the direction of greatest difficulty is at 45° to the axes ($\theta = \pi/4, 3\pi/4, 5\pi/4, 7\pi/4$), and the direction

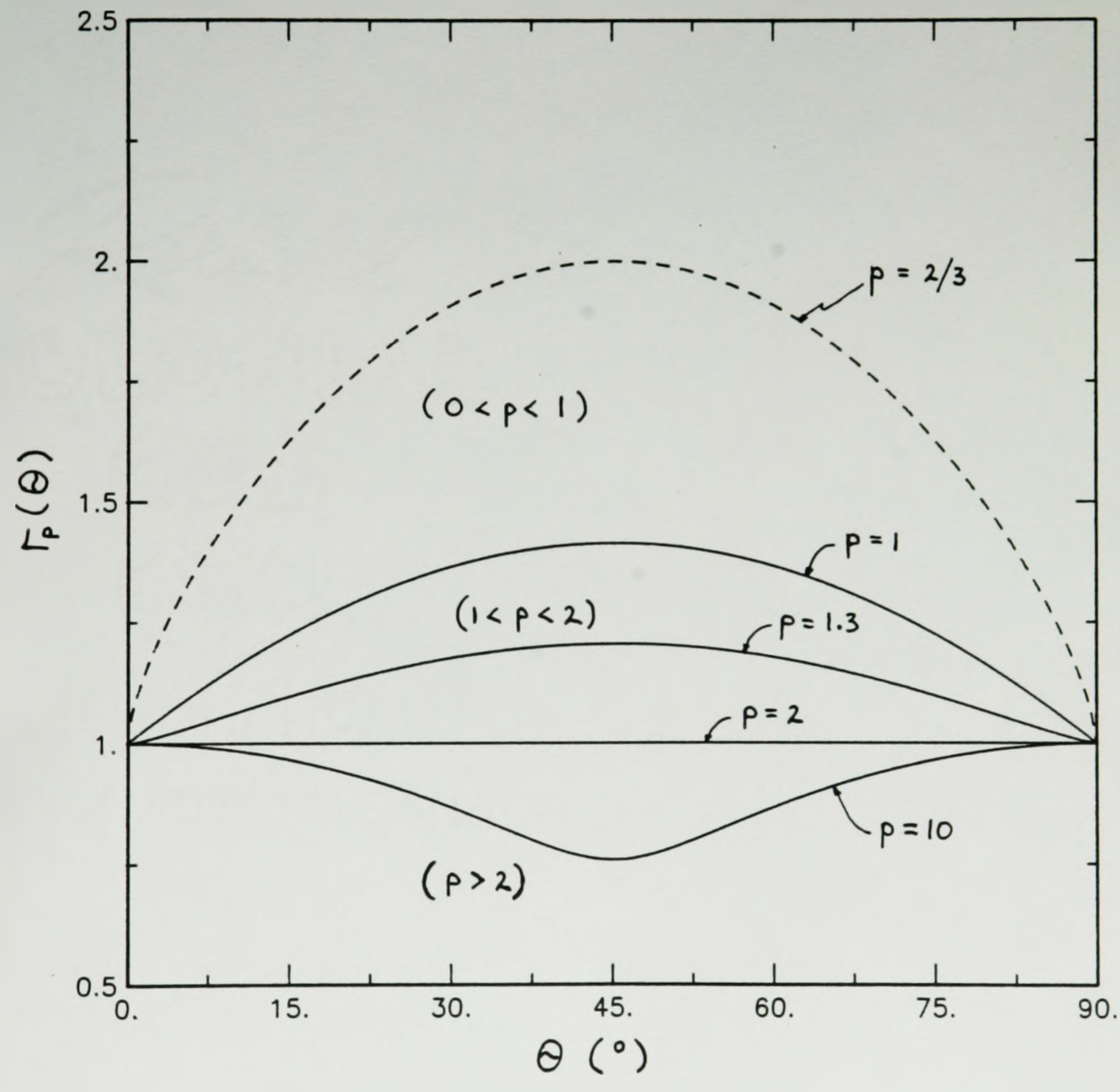


Figure 3.1.2 Directional Bias Function of $\ell_p(x)$.

of least difficulty is parallel to the axes ($\theta=0$, $\pi/2$, π, $3\pi/2$). In other words, the distance $\ell_p(x-y)$ between any two points x and y separated by a straight line segment of fixed length $\ell_2(x-y)$, is maximized if this line segment is at 45° to the axes, and minimized if it is parallel to an axis. The converse holds when $p>2$.

We note the following characteristics of the directional bias function at $\theta=0$, $\pi/4$, $\pi/2$, which will be useful in the subsequent discussion.

$$r_p(0) = r_p(\pi/2) = 1 \ , \quad p > 0 \ ; \tag{3.1.20}$$

$$r_p(\pi/4) = 2^{\frac{1}{p}-\frac{1}{2}} \ , \quad \forall p \neq 0 \ ; \tag{3.1.21}$$

$$\frac{dr_p(0)}{d\theta} = \frac{dr_p(\pi/2)}{d\theta} = 0 \ , \quad 1 < p < +\infty \ ; \tag{3.1.22}$$

$$\frac{dr_p(\pi/4)}{d\theta} = 0 \ , \quad \forall p \neq 0 \ . \tag{3.1.23}$$

Relation (3.1.23) can be calculated directly from the functional form of the first derivative in (3.1.13), while (3.1.22) is readily obtained after rewriting the first derivative as follows:

$$\frac{dr_p(\theta)}{d\theta} = \frac{1}{[r_p(\theta)]^{p-1}} \left(-\cos^{p-1}\theta \sin\theta + \cos\theta \sin^{p-1}\theta \right) , \quad 0<\theta<\pi/2 \ . \tag{3.1.24}$$

Since the functional form of $dr_p/d\theta$ given in (3.1.24) (or (3.1.13)) is valid only in the interval $0 < \theta < \pi/2$, the slopes calculated at 0 and $\pi/2$ are in actuality right and left-sided derivatives respectively. However, since the slopes at 0 and $\pi/2$ are equal in (3.1.22), we conclude from the periodicity of $r_p(\theta)$ (Property 3.1.2) that r_p has a two-sided derivative at $\theta=0$ and $\pi/2$ for $p>1$; i.e., r_p is differentiable here. This is not the case if $0<p\leq 1$, as seen by the following limits. For $p=1$ (rectangular distance), we have

$$\lim_{\theta\to 0^+} \frac{dr_1}{d\theta} = \lim_{\theta\to 0^+} \left| -\sin\theta + \cos\theta \right| = 1 \ , \tag{3.1.25a}$$

while

$$\lim_{\theta \to \frac{\pi}{2}^-} \frac{dr_1}{d\theta} = -1 \,. \tag{3.1.25b}$$

For $0 < p < 1$ (hyper-rectilinear distance),

$$\lim_{\theta \to 0^+} \frac{dr_p}{d\theta} = \lim_{\theta \to 0^+} \left\{ \frac{1}{[r_p(\theta)]^{p-1}} \left(-\cos^{p-1}\theta \sin\theta + \cos\theta \sin^{p-1}\theta \right) \right\}$$

$$= \lim_{\theta \to 0^+} \left\{ \frac{-\cos^{p-1}\theta \sin\theta}{[r_p(\theta)]^{p-1}} \right\} + \lim_{\theta \to 0^+} \left\{ \frac{\cos\theta \sin^{p-1}\theta}{[r_p(\theta)]^{p-1}} \right\}$$

$$= \lim_{\theta \to 0^+} \{-\sin\theta\} + \lim_{\theta \to 0^+} \{\sin^{p-1}\theta\} \qquad \text{(equation (3.1.20))}$$

$$= +\infty \,, \tag{3.1.26a}$$

while similarly,

$$\lim_{\theta \to \frac{\pi}{2}^-} \frac{dr_p}{d\theta} = \lim_{\theta \to \frac{\pi}{2}^-} \left\{ -\cos^{p-1}\theta + \cos\theta \right\} = -\infty \,. \tag{3.1.26b}$$

Hence we conclude, using the preceding limits and the periodicity of $r_p(\theta)$ (Property 3.1.2), that the right and left-sided derivatives are not equal at 0 or $\pi/2$, and thus r_p is not differentiable at $\theta = n\pi/2$, $n = 0, \pm 1, \pm 2, \ldots$, when $0 < p \leq 1$.

We see now that the directional bias function of the ℓ_p norm is differentiable for all θ if, and only if, $p > 1$. This result can be extended to general norms as follows.

Theorem 3.1.1

The directional bias function $r(\theta)$ of the norm k is differentiable for all θ if, and only if, k is a differentiable round norm.

Proof:

Without loss in generality, consider only the points x on the circle of unit radius centered at the origin ($\ell_2(x) = 1$). Let x_0 denote such a point, θ_0 be the angle specifying the vector x_0, and y_0 be a unit vector tangent to the circle at x_0. (See Figure 3.1.1b). Let $d/d\theta^+$ and $d/d\theta^-$ denote the right-sided and left-sided derivatives respectively. Referring to Figure 3.1.1b, we have in general

$$\frac{dr(\theta_0)}{d\theta^+} = k(x_0 ; y_0) , \tag{3.1.27a}$$

and

$$\frac{dr(\theta_0)}{d\theta^-} = -k(x_0 ; -y_0) . \tag{3.1.27b}$$

If k is a differentiable round norm, then by (2.4.7),

$$k(x_0 ; y_0) = \nabla k(x_0) \cdot y_0 = -k(x_0 ; -y_0) ; \tag{3.1.28}$$

so that

$$\frac{dr(\theta_0)}{d\theta^+} = \frac{dr(\theta_0)}{d\theta^-} . \tag{3.1.29}$$

Hence, k a differentiable round norm is a sufficient condition for r to be differentiable at all values of θ.

If k is not a differentiable round norm, and furthermore x_0 is chosen such that $\nabla k(x_0)$ is undefined, then we must have

$$k(x_0 ; y_0) \neq -k(x_0 ; -y_0) . \tag{3.1.30}$$

(Since k is differentiable in the radial direction through x_0, we would otherwise conclude that $\nabla k(x_0)$ exists, which is a contradiction.)

Therefore,

$$\frac{dr(\theta_0)}{d\theta^+} \neq \frac{dr(\theta_0)}{d\theta^-}, \qquad (3.1.31)$$

so that r is not differentiable at θ_0. Hence, k a differentiable round norm is also a necessary condition for r to be differentiable at all values of θ.

Theorem 3.1.1 provides another way of viewing the different classes of norms discussed in Chapter 2 (see Figure 2.2.2). If k is a differentiable round norm its directional bias function r is a smoothly varying function of θ. On the other hand, if k is a non-differentiable round norm or a block norm, then $r(\theta)$ has sharp corners at certain values of θ where the slope changes by a discrete amount. This characteristic is illustrated in Figures 3.1.3a and b, where one cycle of $r_p(\theta)$ is plotted for sample ℓ_p norms with $p > 1$, and compared with $r(\theta)$ of weighted one-infinity norms (see (1.2.13)) whose parameters are adjusted so that $r_p(\theta) = r(\theta)$ at $\theta = 0, \pi/4, \pi/2$.

A practical observation can be made concerning the ℓ_p norm ($p \geq 1$), after a close scrutiny of Figure 3.1.2 and equations (3.1.20) to (3.1.23). First we note that $r_p(\theta)$ is a decreasing function of p except at the boundaries $\theta = 0$ and $\pi/2$ in Figure 3.1.2, where it is constant. (This result is a direct consequence of Property 2.3.1.) When $1 \leq p < 2$, the direction of greatest difficulty is at $\theta = \pi/4$ (45° to the axes), and the direction of least difficulty is at $\theta = 0, \pi/2$ (parallel to the axes). Meanwhile, when $p > 2$ this situation is reversed, signifying a phase change of 45°. One should expect therefore that the norm $\ell_q(x)$, where q is some value greater than 2, can be accurately approximated by the norm $\sigma\ell_p(x')$, where x' gives the coordinates of the point x in a new set of axes obtained by a 45° rotation of the original axes, σ is a scaling factor less than 1, and p takes on a value in the open interval (1,2). This expectation is reinforced by the fact that $dr_p/d\theta = 0$ at $\theta = 0, \pi/4$, and $\pi/2$, for all $p > 1$ (equations (3.1.22), (3.1.23)).

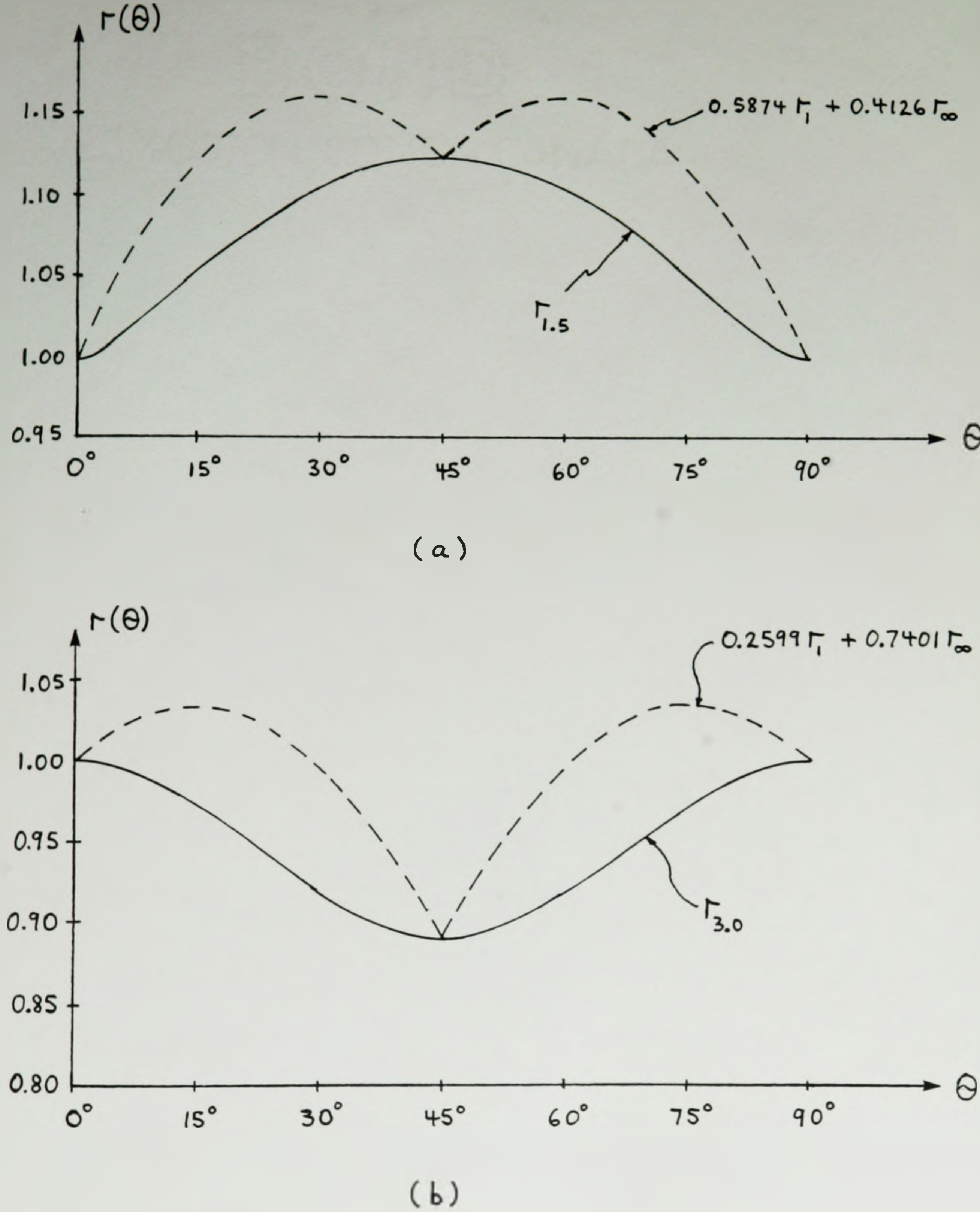


Figure 3.1.3 Comparison of Directional Bias Functions.

We proceed now to investigate the accuracy of such an approximation. Define a 'normalized' difference between $\sigma\ell_p(x')$ and $\ell_q(x)$ as follows;

$$\Delta = \frac{\sigma\,\ell_p(x') - \ell_q(x)}{\ell_2(x)} \ , \ \forall x \neq 0 \ . \tag{3.1.32}$$

Since Euclidean distances are preserved under an orthogonal transformation and a rotation in R^2 is such a transformation (Shields, 1969, p. 285), therefore

$$\ell_2(x') = \ell_2(x) \ . \tag{3.1.33}$$

Thus, we obtain

$$\frac{\ell_q(x)}{\ell_2(x)} = r_q(\theta) \ , \tag{3.1.34}$$

where it is recalled that $x = (x_1, x_2)^T$ and $\theta = \tan^{-1}(x_2/x_1)$; and

$$\frac{\ell_p(x')}{\ell_2(x)} = \frac{\ell_p(x')}{\ell_2(x')} = r_p(\theta - \pi/4) \ , \tag{3.1.35}$$

since the axes are rotated 45°. Hence, for a given value of $q > 2$, and specified values of p and σ, Δ is a function of θ alone; i.e.,

$$\Delta(\theta) = \sigma\, r_p(\theta - \pi/4) - r_q(\theta) \ . \tag{3.1.36}$$

Due to the symmetry of the directional bias function of the ℓ_p norm (Properties 3.1.2 and 3.1.3), it suffices to consider θ in the range,

$$0 \leq \theta \leq \pi/4 \ . \tag{3.1.37}$$

The question now is how to specify for a given $r_q(\theta)$, $(q > 2)$, the values of σ and p in the approximation function $\sigma r_p(\theta - \pi/4)$. We choose a simple method. Impose the following 'boundary' conditions,

$$\Delta(0) = \Delta(\pi/4) = 0 \ , \tag{3.1.38}$$

to obtain two equations to solve for the two unknowns, σ and p. With (3.1.20), (3.1.21) and Property 3.1.2, the boundary conditions become

$$\Delta(0) = \sigma \cdot 2^{\frac{1}{p} - \frac{1}{2}} - 1 = 0 \ , \tag{3.1.39a}$$

and

$$\Delta(\pi/4) = \sigma - 2^{\frac{1}{q}-\frac{1}{2}} = 0 \ . \qquad (3.1.39b)$$

Solving for σ and p gives

$$\sigma = 2^{\frac{1}{q}-\frac{1}{2}} , \qquad (3.1.40)$$

and

$$p = q/(q-1) \ . \qquad (3.1.41)$$

Besides the zero value imposed on Δ at the boundaries, θ = 0 and π/4, it is easily shown using (3.1.22), (3.1.23) and Property 3.1.2 that

$$\frac{d\Delta(0)}{d\theta} = \frac{d\Delta(\pi/4)}{d\theta} = 0 \ . \qquad (3.1.42)$$

As a result of the boundary conditions in (3.1.38) and (3.1.42), and the fact that $r_p(\theta-\pi/4)$ and $r_q(\theta)$ have the same general concave/convex shape arising from a unique inflection point (Property 3.1.5), we expect the difference function Δ to be small. This implies that the approximation gives a good fit. Let us determine now the accuracy of this fit.

For $0 \le \theta \le \pi/4$, we have

$$r_q(\theta) = [\cos^q\theta + \sin^q\theta]^{1/q} , \qquad (3.1.43)$$

and

$$r_p(\theta - \pi/4) = \left[\left| \cos(\theta-\pi/4) \right|^p + \left| \sin(\theta-\pi/4) \right|^p \right]^{1/p}$$

$$= \left[\cos^p(\pi/4-\theta) + \sin^p(\pi/4-\theta) \right]^{1/p}$$

$$= \frac{1}{\sqrt{2}} \left[(\cos\theta + \sin\theta)^p + (\cos\theta - \sin\theta)^p \right]^{1/p} .$$

Also note from (3.1.41) that

$$q = p/(p-1) \ .$$

Using the above equations and (3.1.40), it follows that Δ can be re-written in terms of θ and p alone. In order to signify that the difference function has equations (3.1.40) and (3.1.41) imposed on the parameters σ and p, we denote it as $\delta(\theta;p)$. It is readily seen that

$$\delta(\theta; p) = 2^{-\frac{1}{p}} \left| (\cos\theta + \sin\theta)^p + (\cos\theta - \sin\theta)^p \right|^{1/p}$$
$$- \left| (\cos\theta)^{p/(p-1)} + (\sin\theta)^{p/(p-1)} \right|^{(p-1)/p} , \tag{3.1.43}$$

where

$$0 \le \theta \le \pi/4 \quad \text{and} \quad 1 < p < 2 \; . \tag{3.1.44}$$

We can now carry out a numerical search over the ranges given in (3.1.44) to determine the maximum absolute magnitude of δ, and the values of θ and p where this occurs. First note the following limiting cases:

$$\begin{aligned} \lim_{p \to 1^+} \delta(\theta;p) &= \cos\theta - \lim_{p \to 1^+} \left\{ \left| (\cos\theta)^{p/(p-1)} + (\sin\theta)^{p/(p-1)} \right|^{(p-1)/p} \right\} \\ &= \cos\theta - \max\{\cos\theta, \sin\theta\} \\ &= \cos\theta - \cos\theta \qquad (0 \le \theta \le \pi/4) \\ &= 0 \; ; \end{aligned} \tag{3.1.45}$$

$$\begin{aligned} \lim_{p \to 2^-} \delta(\theta;p) &= \delta(\theta;2) \\ &= \frac{1}{\sqrt{2}} \left| (\cos\theta + \sin\theta)^2 + (\cos\theta - \sin\theta)^2 \right|^{1/2} - 1 \\ &= 1 - 1 = 0 \; . \end{aligned} \tag{3.1.46}$$

Also, by (3.1.38),

$$\delta(0 \; ; p) = \delta(\pi/4 \; ; p) = 0 \; . \tag{3.1.47}$$

Thus, the p and θ which maximize $|\delta|$ must be at an interior point, remote from the boundaries of the rectangle defined by the ranges in (3.1.44). This facilitates the numerical search, since we can now bypass large values of the exponent $p/(p-1)$ $(=q)$ in (3.1.43) when p approaches 1^+.

The difference function $\delta(\theta;p)$ was evaluated on the computer over a finely-divided grid covering the points (θ,p) defined by (3.1.44). The results are summarized in Figures 3.1.4 and 3.1.5, where $\max_\theta |\delta|$ and the θ_* at which this occurs are plotted as functions of p. Observe that

$$\max_{\theta,p} |\delta| = 0.027818 \ , \tag{3.1.48}$$

at

$$p_{**} = 1.2355 \quad \text{and} \quad \theta_{**} = 33.493^\circ \ , \tag{3.1.49}$$

where the maximizing values p_{**} and θ_{**} are obtained by the grid search to an accuracy of 0.0001 and 0.001° respectively. In Figure 3.1.6, δ is plotted as a function of θ for sample fixed p. The profiles here are unimodal in shape, and θ_* shifts to the left for increasing values of p (see also Figure 3.1.5). It is also interesting to note that

$$\delta(\theta\,;p) \geq 0, \quad \forall\ \theta\,,p\ , \tag{3.1.50}$$

a result confirmed by the exhaustive grid search. Finally, in Figure 3.1.7, one cycle of $r_q(\theta)$ and its approximation by $\sigma\, r_p(\theta - \pi/4)$ are shown for the sample case, $p = 1.2$ ($q = 6$).

We see from the above results that the directional bias function $r_q(\theta)$ for any $q > 2$ is accurately approximated by $\sigma r_p(\theta - \pi/4)$, where the scaling factor σ and $p \in (1,2)$ are given in (3.1.40) and (3.1.41). This leads to the following important conclusion.

> The norm $\ell_q(x)$, where $q>2$ and $x \in R^2$, can be replaced for all practical purposes by the norm $\sigma\ell_p(x')$, where x' is the vector of coordinates of x measured in a new set of axes rotated 45° from the original, $\sigma = 2^{1/q - 1/2} < 1$, and $p = q/(q-1) \in (1,2)$.

In quantitative terms, we obtain from (3.1.48) and (3.1.50) the following bounds on the difference;

$$0 \leq \sigma\,\ell_p(x') - \ell_q(x) \leq 0.027818\,\ell_2(x)\ . \tag{3.1.51}$$

Recall that certain constraints were imposed on the normalized difference Δ (see (3.1.38)), in order to specify σ and p for a given q. If these restrictions are removed, and

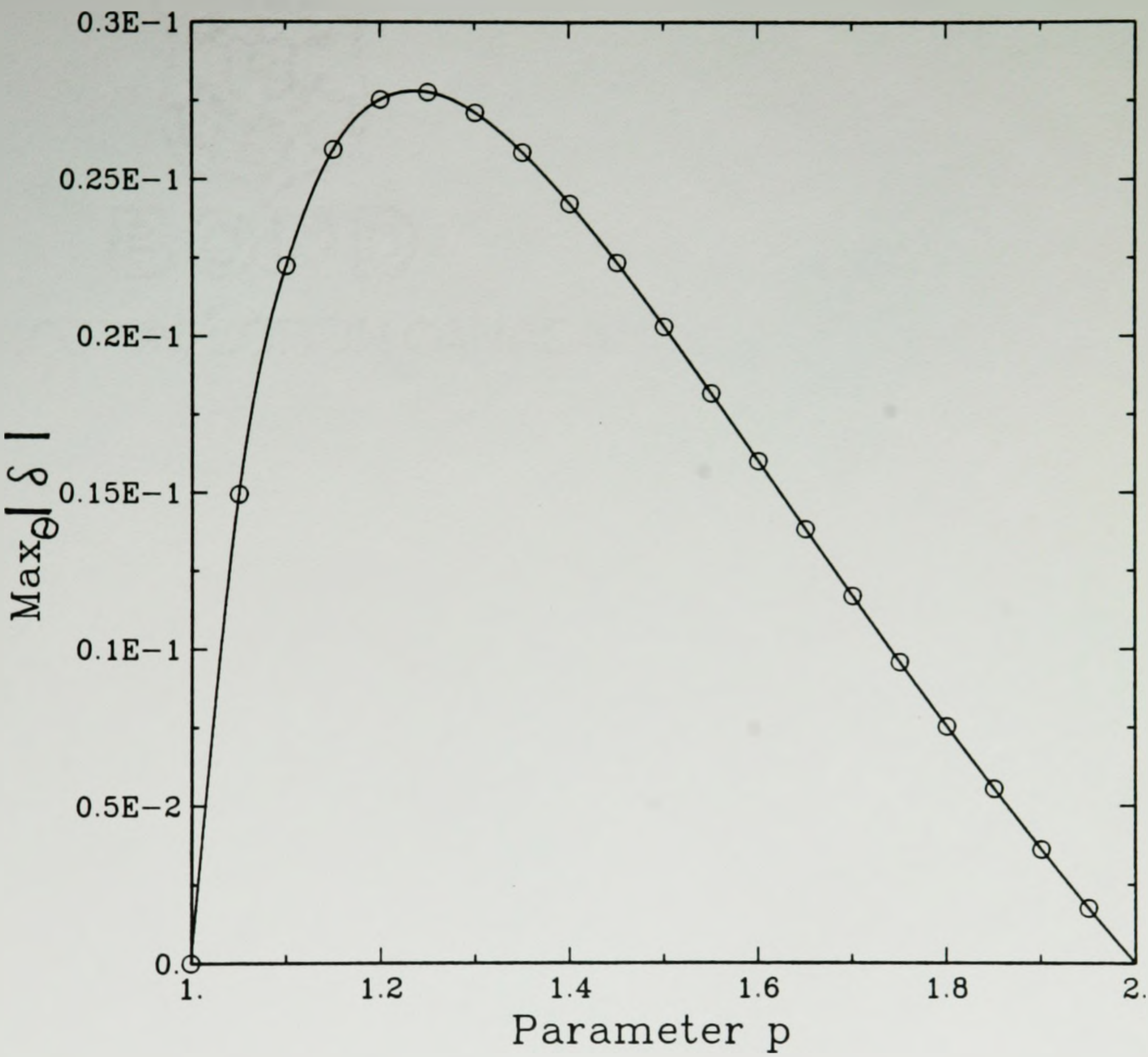


Figure 3.1.4 Maximizing | δ(θ ; p) | over θ.

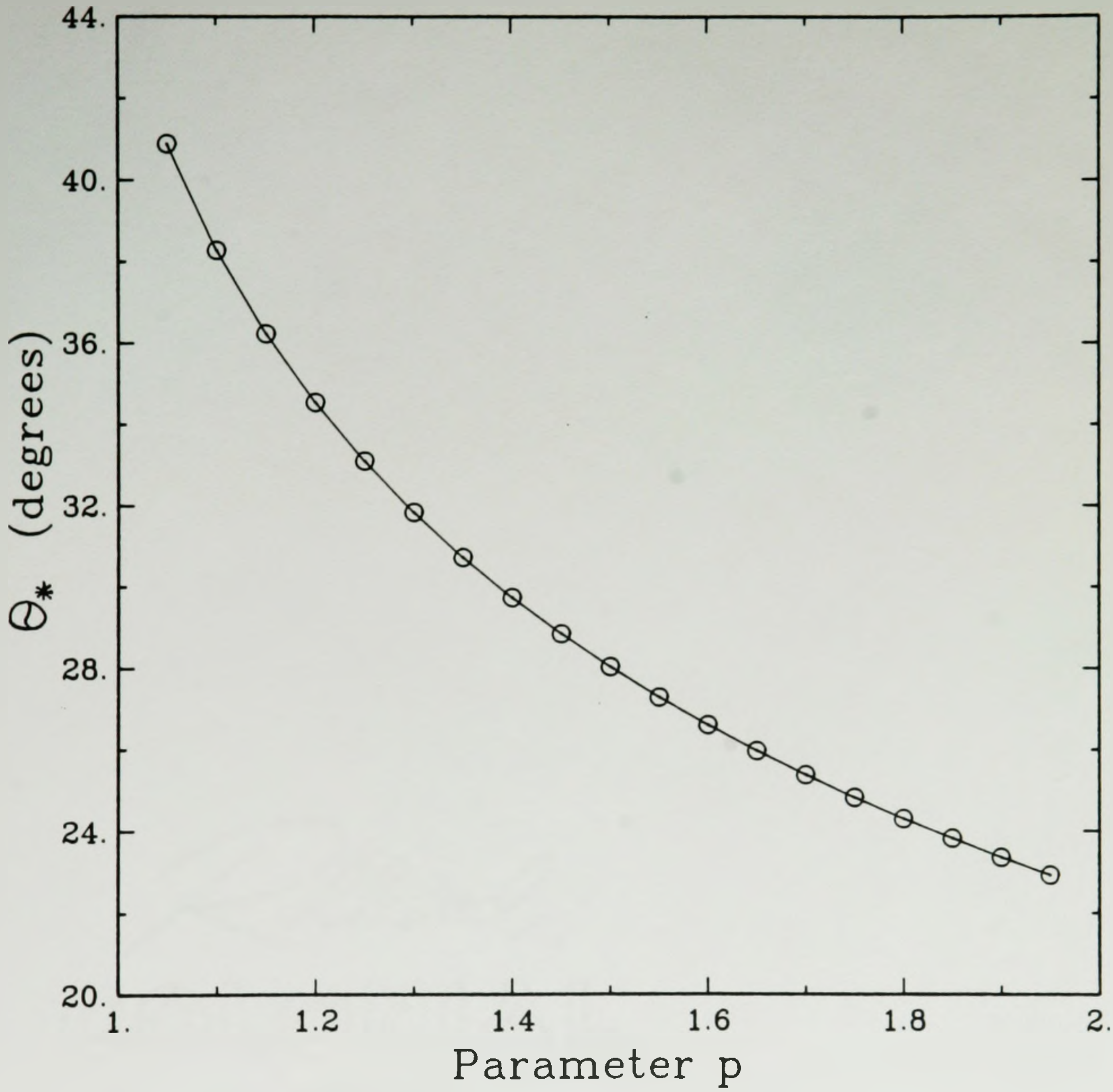


Figure 3.1.5 Maximizing Value of θ.

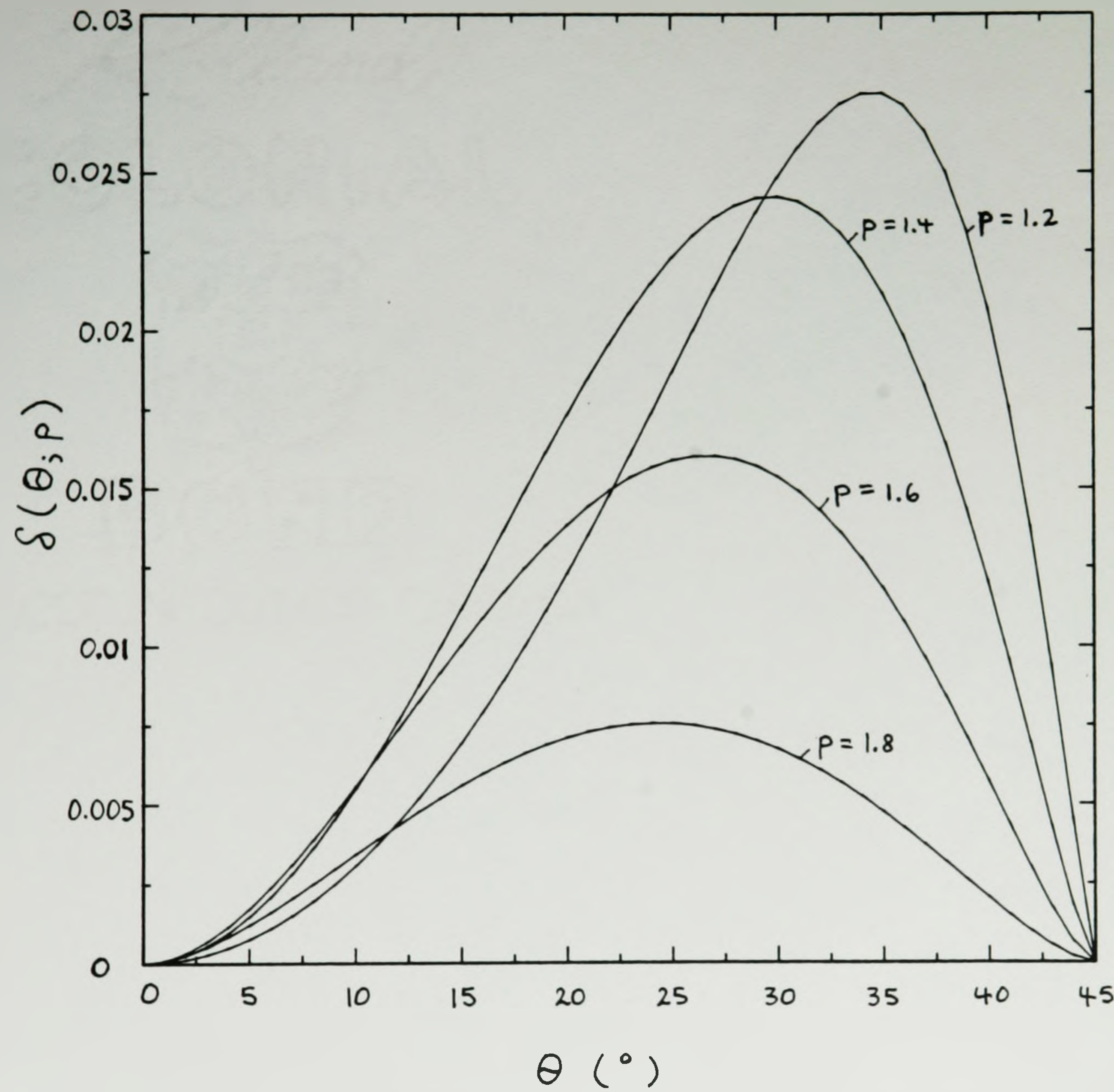


Figure 3.1.6 Profiles of δ(θ ; p) for Fixed p.

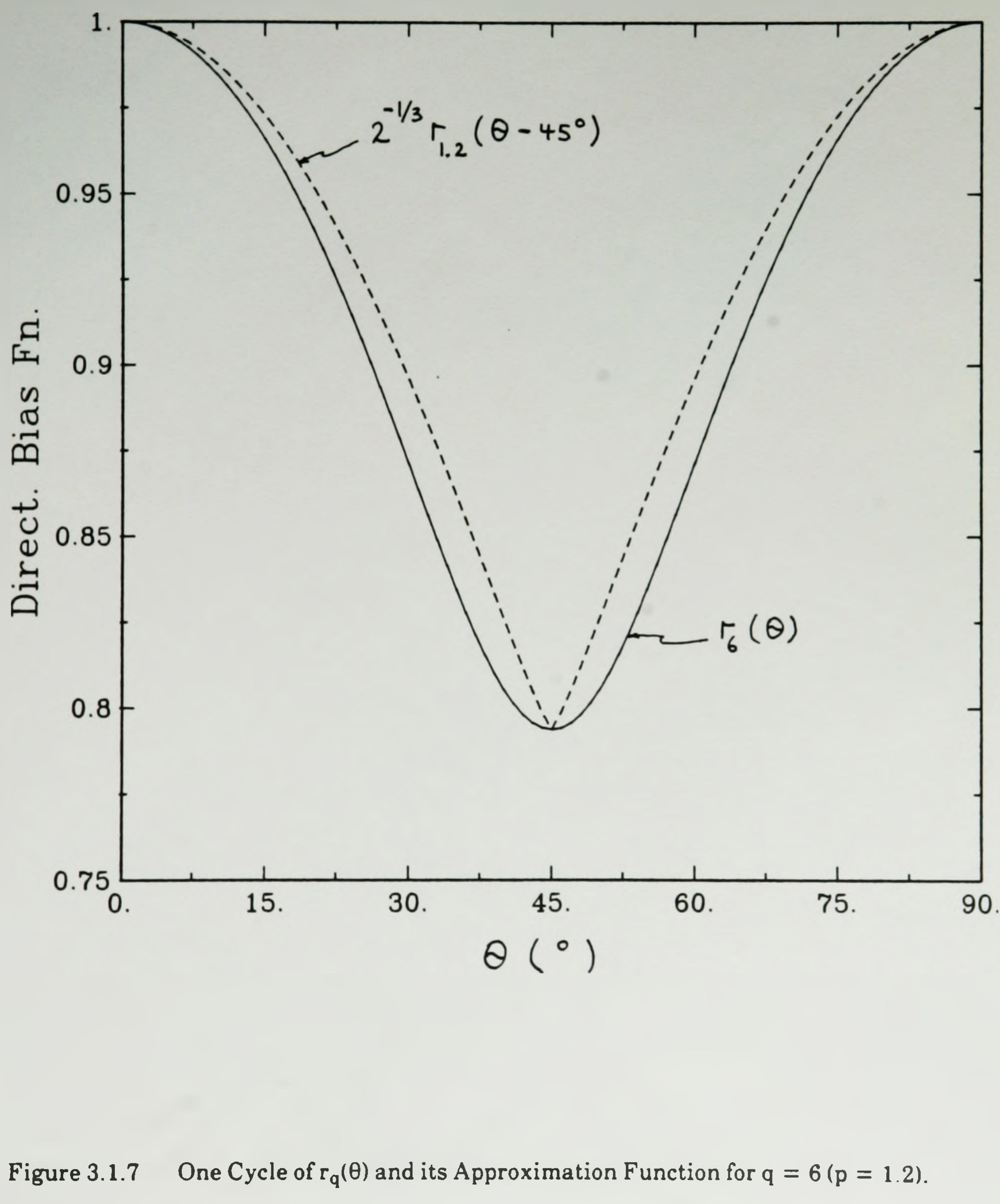


Figure 3.1.7 One Cycle of $r_q(\theta)$ and its Approximation Function for $q = 6$ ($p = 1.2$).

instead σ and p are chosen to minimize the maximum absolute magnitude of Δ, the approximation will be improved. This result strengthens the conclusion above; namely, that for practical problems, 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) after rotating the axes 45°.

As a final comment on this topic, we note that when p is given by (3.1.41), $\ell_p(\cdot)$ is the polar of $\ell_q(\cdot)$. Thus, the bounds in (3.1.51) provide a quantitative relation between $\ell_q(x)$ and its polar acting on transformed coordinates.

3.2 Fitting the Weighted ℓ_p Norm

The weighted ℓ_p norm was introduced in Chapter 1, where we noted the successful use of this function in estimating actual distances from several road networks. We shall see below that the weighted ℓ_p norm has two parameters, an inflation factor τ and the parameter p of the ℓ_p norm, which need to be specified. Two criteria are mentioned in the literature for fitting the unknown parameters of a distance function to a set of data (e.g., see Love and Morris, 1972, p. 64). Applied to our distance function, these goodness- of-fit criteria are as follows:

Criterion 1

$$\underset{\tau,\, p}{\text{minimize}}\ AD_\ell = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} |d_\ell(a_i, a_j) - A_{ij}|; \tag{3.2.1}$$

Criterion 2

$$\underset{\tau,\, p}{\text{minimize}}\ SD_\ell = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} [d_\ell(a_i, a_j) - A_{ij}]^2 / A_{ij}; \tag{3.2.2}$$

where n = the number of fixed points (cities, destinations, customers) in the road network, which are chosen for the data set;

A_{ij} = the actual distance (road miles) between the fixed points a_i and a_j;

$d_\ell(a_i,a_j) = \tau\, \ell_p(a_i-a_j)$ is the weighted ℓ_p norm used to estimate the distance (road miles) between a_i and a_j;

and $\tau, p>0$ are the unknown parameters.

The first criterion involves minimization of the sum of absolute deviations between estimated and actual distances. As noted in Love and Morris (1972), the implication here is that the empirical function tends to estimate greater actual distances relatively more accurately than shorter distances. On the other hand, the second criterion, which involves minimization of a weighted sum of squared deviations, achieves a greater sensitivity for shorter distances through the weights A_{ij}^{-1}. It also possesses attractive statistical properties (Love and Morris, 1972). We see that the two criteria measure goodness-of-fit in significantly different ways.

As noted in Chapter 1, the minimization of AD_ℓ or SD_ℓ is currently carried out by an exhaustive grid search over a 'safe' range of parameter values. Since the calculation of AD_ℓ or SD_ℓ at each grid point involves $O(n^2)$ operations, this procedure becomes very time consuming and inefficient for large samples. It appears that no effort has been made yet to improve on the brute- force approach. Thus, the purpose of this section is primarily to derive properties which will permit us to find the best-fitting values of τ and p for the two criteria in an efficient manner.

The first two results pertain to the behaviour of AD_ℓ and SD_ℓ as functions of the inflation factor τ alone (i.e., p is fixed).

Property 3.2.1

AD_ℓ is a convex function of τ.

Proof:

Denote the terms in the summation defining AD_ℓ by

$$g_{ij}(\tau, p) = |d_\ell(a_i, a_j) - A_{ij}| = |\tau \ell_p(a_i - a_j) - A_{ij}|, \tag{3.2.3}$$

$$i, j = 1, \ldots, n, \; i < j.$$

Then for any i,j,

$$\frac{\partial}{\partial \tau} g_{ij} = \begin{cases} -\ell_p(a_i - a_j), & \text{if } \tau < A_{ij} / \ell_p(a_i - a_j), \\ +\ell_p(a_i - a_j), & \text{if } \tau > A_{ij} / \ell_p(a_i - a_j). \end{cases} \tag{3.2.4}$$

We see that the slope $\partial g_{ij}/\partial \tau$ is non-decreasing in τ, and hence g_{ij} is a convex function of τ. Thus, AD_ℓ is the sum of convex terms in τ, and is itself a convex function of τ.

Property 3.2.2

SD_ℓ is a strictly convex function of τ.

Proof:

Denote the terms in the summation defining SD_ℓ by

$$h_{ij}(\tau, p) = \frac{[d_\ell(a_i, a_j) - A_{ij}]^2}{A_{ij}} = \frac{[\tau \ell_p(a_i - a_j) - A_{ij}]^2}{A_{ij}},$$

$$i, j = 1, \ldots, n, \; i < j. \tag{3.2.5}$$

Then for any i,j,

$$\frac{\partial}{\partial \tau} h_{ij} = \frac{2}{A_{ij}} [\tau \ell_p(a_i - a_j) - A_{ij}] \, \ell_p(a_i - a_j), \tag{3.2.6}$$

and

$$\frac{\partial^2}{\partial \tau^2} h_{ij} = \frac{2}{A_{ij}} [\ell_p(a_i - a_j)]^2 > 0. \tag{3.2.7}$$

Thus, h_{ij} is a strictly convex function of τ. Since SD_ℓ is a sum of terms which are strictly convex in τ, then it is also a strictly convex function of τ.

We shall see later that Properties 3.2.1 and 3.2.2 are very useful in streamlining the search for the best-fitting values of τ and p. In the meantime, let us examine AD_ℓ and SD_ℓ as functions of p (i.e., τ is fixed). This turns out to be a relatively complicated problem.

Property 3.2.3

Consider any term $g_{ij}(\tau,p)$ in the sum AD_ℓ (see (3.2.3)) as a function of p in the open interval $(0,+\infty)$. If the vector a_i-a_j is parallel to an axis, then g_{ij} is constant. Otherwise, there are two possibilities:

(i) if $A_{ij}>\tau\max\{|a_{i1}-a_{j1}|, |a_{i2}-a_{j2}|\} = \tau\ell_\infty(a_i-a_j)$, then g_{ij} is a unimodal function of p, strictly convex over the interval $0<p\leq p_{ij}$ and strictly concave for $p\geq p_{ij}$, where p_{ij} is the unique value of p such that

$$\min_p g_{ij}(\tau, p) = g_{ij}(\tau, p_{ij}) = 0 \; ; \tag{3.2.8}$$

(ii) if $A_{ij}\leq\tau\ell_\infty(a_i-a_j)$, then g_{ij} is a decreasing strictly convex function of p with a minimum approached asymptotically as $p\rightarrow+\infty$.

Proof:

Suppose a_i-a_j is parallel to the x_1-axis, so that $a_{i2}-a_{j2} = 0$. Then clearly, $d_\ell(a_i,a_j) = \tau|a_{i1}-a_{j1}|$ for all $p>0$. A similar result holds if a_i-a_j is parallel to the x_2-axis. Hence, g_{ij} is a constant function of p when $a_i - a_j$ is parallel to an axis.

On the other hand, if this is not the case, then $|a_{it}- a_{jt}|>0$, $t = 1,2$. By Properties 2.1.1 and 2.1.2, it follows that $d_\ell(a_i,a_j)$ is a decreasing strictly convex function of $p\in(0,+\infty)$. Also $\lim_{p\rightarrow0^+}d_\ell(a_i,a_j) = +\infty$, and $\lim_{p\rightarrow\infty}d_\ell(a_i,a_j) = \tau\ell_\infty(a_i- a_j)$, using equations (2.1.9) and (2.1.29). The remainder of the proof is now obvious.

Property 3.2.4

The previous result applies to any term $h_{ij}(\tau,p)$ in the sum SD_ℓ (see (3.2.5)), except when a_i-a_j is not parallel to an axis and $A_{ij}>\tau\ell_\infty(a_i- a_j)$. For this case, h_{ij} is also a unimodal function of p with minimum at p_{ij}. However, h_{ij} is strictly convex over the interval $0<p\leq p_{ij}$

and strictly concave for $p \geq \rho_{ij}$, where ρ_{ij} is the unique inflection point such that

$$\frac{\partial^2}{\partial p^2} h_{ij}(\tau, \rho_{ij}) = 0 \ . \tag{3.2.9}$$

Furthermore,

$$\rho_{ij} > p_{ij} \, . \tag{3.2.10}$$

Proof:

If $a_i - a_j$ is parallel to an axis, then g_{ij} is a constant function of p (Property 3.2.3). Therefore, $h_{ij} = g_{ij}{}^2/A_{ij}$ is also constant in p.

Now consider $a_i - a_j$ not parallel to an axis. To simplify the notation let $\beta(p) := \ell_p(a_i - a_j)$, $f_1 := |a_{i1} - a_{j1}| > 0$, $f_2 := |a_{i2} - a_{j2}| > 0$, and $\beta'(p)$ and $\beta''(p)$ give the first and second-order derivatives of β. We have

$$h_{ij}(\tau, p) = [\tau \beta(p) - A_{ij}]^2 / A_{ij} \, , \tag{3.2.11a}$$

$$\frac{\partial}{\partial p} h_{ij}(\tau, p) = \frac{2\tau}{A_{ij}} (\tau \beta(p) - A_{ij}) \beta'(p) \, , \tag{3.2.11b}$$

and

$$\frac{\partial^2}{\partial p^2} h_{ij}(\tau, p) = \frac{2\tau}{A_{ij}} [\tau (\beta'(p))^2 + (\tau \beta(p) - A_{ij}) \beta''(p)] \, . \tag{3.2.11c}$$

Furthermore, by Properties 2.1.1 and 2.1.2, it follows that

$$\beta'(p) < 0 \ \text{ and } \ \beta''(p) > 0 \, , \ \forall \, p \in (0, +\infty) \, . \tag{3.2.12}$$

If $A_{ij} \leq \tau \ell_\infty(a_i - a_j)$, then

$$\begin{aligned} \tau \beta(p) - A_{ij} &> \tau \ell_\infty(a_i - a_j) - A_{ij} \, , \quad 0 < p < +\infty \, , \quad \text{(Property 2.1.1)} \\ &\geq 0 \, . \end{aligned} \tag{3.2.13}$$

Hence,

$$\frac{\partial h_{ij}}{\partial p} < 0 \ \text{ and } \ \frac{\partial^2 h_{ij}}{\partial p^2} > 0, \ 0 < p < +\infty \, . \tag{3.2.14}$$

Clearly then, h_{ij} is a decreasing strictly convex function of p with the minimum $([\tau \ell_\infty(a_i - a_j) - A_{ij}]^2 / A_{ij})$ approached asymptotically as $p \to +\infty$.

Now consider the case where $A_{ij} > \tau\ell_\infty(a_i - a_j)$. It is readily seen 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 at least one inflection point ρ_{ij}, such that $\partial^2 h_{ij}(\tau, \rho_{ij})/\partial p^2 = 0$. We now show that ρ_{ij} is unique, and furthermore, $\rho_{ij} > p_{ij}$. Using equation (2.1.6), rewrite $\beta'(p)$ as follows:

$$\beta'(p) = \beta(p) \cdot H(p), \tag{3.2.15}$$

where

$$H(p) = \frac{1}{p^2 (f_1^p + f_2^p)} \left[f_1^p \ell n \left(\frac{f_1^p}{f_1^p + f_2^p} \right) + f_2^p \ell n \left(\frac{f_2^p}{f_1^p + f_2^p} \right) \right]. \tag{3.2.16}$$

Therefore,

$$\begin{aligned} \beta''(p) &= \beta'(p) H(p) + \beta(p) H'(p) \\ &= \beta(p) [H^2(p) + H'(p)]. \end{aligned} \tag{3.2.17}$$

Substituting (3.2.15) and (3.2.17) into (3.2.11c), and equating to zero, it follows that ρ_{ij} must solve the equation

$$\frac{A_{ij}}{\tau \beta(p)} = \frac{2H^2(p) + H'(p)}{H^2(p) + H'(p)}. \tag{3.2.18}$$

But

$$\begin{aligned} 2H^2(p) + H'(p) &> H^2(p) + H'(p) \\ &> 0 \quad (\because \beta''(p) > 0), \end{aligned}$$

so that

$$\frac{2H^2(p) + H'(p)}{H^2(p) + H'(p)} > 1, \; \forall p > 0. \tag{3.2.19}$$

Hence,

$$A_{ij} > \tau \beta (\rho_{ij}). \tag{3.2.20}$$

Since

$$A_{ij} = \tau \beta (p_{ij}), \tag{3.2.21}$$

by the definition of p_{ij} (see Property 3.2.3), and β is a decreasing function of p, we must have

$$\rho_{ij} > p_{ij}.$$

A lengthy computation, the details of which are left to Appendix A, reveals that the right-hand side of equation (3.2.18) is a decreasing function of p (>0). The left-hand side, $A_{ij}/\tau\beta(p)$, is clearly an increasing function of p. It follows that ρ_{ij} is unique, and that h_{ij} is strictly convex in p for $0<p\leq\rho_{ij}$ and strictly concave in p for $p\geq\rho_{ij}$. Also note that

$$\frac{\partial h_{ij}}{\partial p} \begin{cases} <0\,, & \text{if } 0<p<p_{ij}\,, \\ =0\,, & \text{if } p=p_{ij}\,, \\ >0\,, & \text{if } p>p_{ij}\,, \end{cases} \tag{3.2.22}$$

and hence, h_{ij} is a unimodal function of p with minimum at p_{ij}. This ends the proof.

The shapes of g_{ij} and h_{ij} for varying p are illustrated in Figure 3.2.1a) and b), for the case where the vector a_i-a_j is not parallel to an axis. Since AD_ℓ and SD_ℓ are the sums of terms g_{ij} and h_{ij} respectively, each term being in general neither convex nor concave in p, we obtain the following important result.

Property 3.2.5

Consider the sums AD_ℓ and SD_ℓ as functions of p in the open interval $(0,+\infty)$; i.e., the inflation factor τ is fixed. In general AD_ℓ and SD_ℓ are neither convex nor concave in p, and may have more than one local minimum or maximum.

As a consequence of the above property, there is no easy way to find a value of p which minimizes AD_ℓ or SD_ℓ globally for a given τ. One is forced essentially to do a thorough numerical search over a safe range. It would be advantageous to restrict this search by specifying the smallest interval of p in which the global optimum is known to occur. This is the purpose of the following results.

Property 3.2.6

If $\tau\leq 1$ and at least one of the vectors a_i-a_j is not parallel to an axis, then

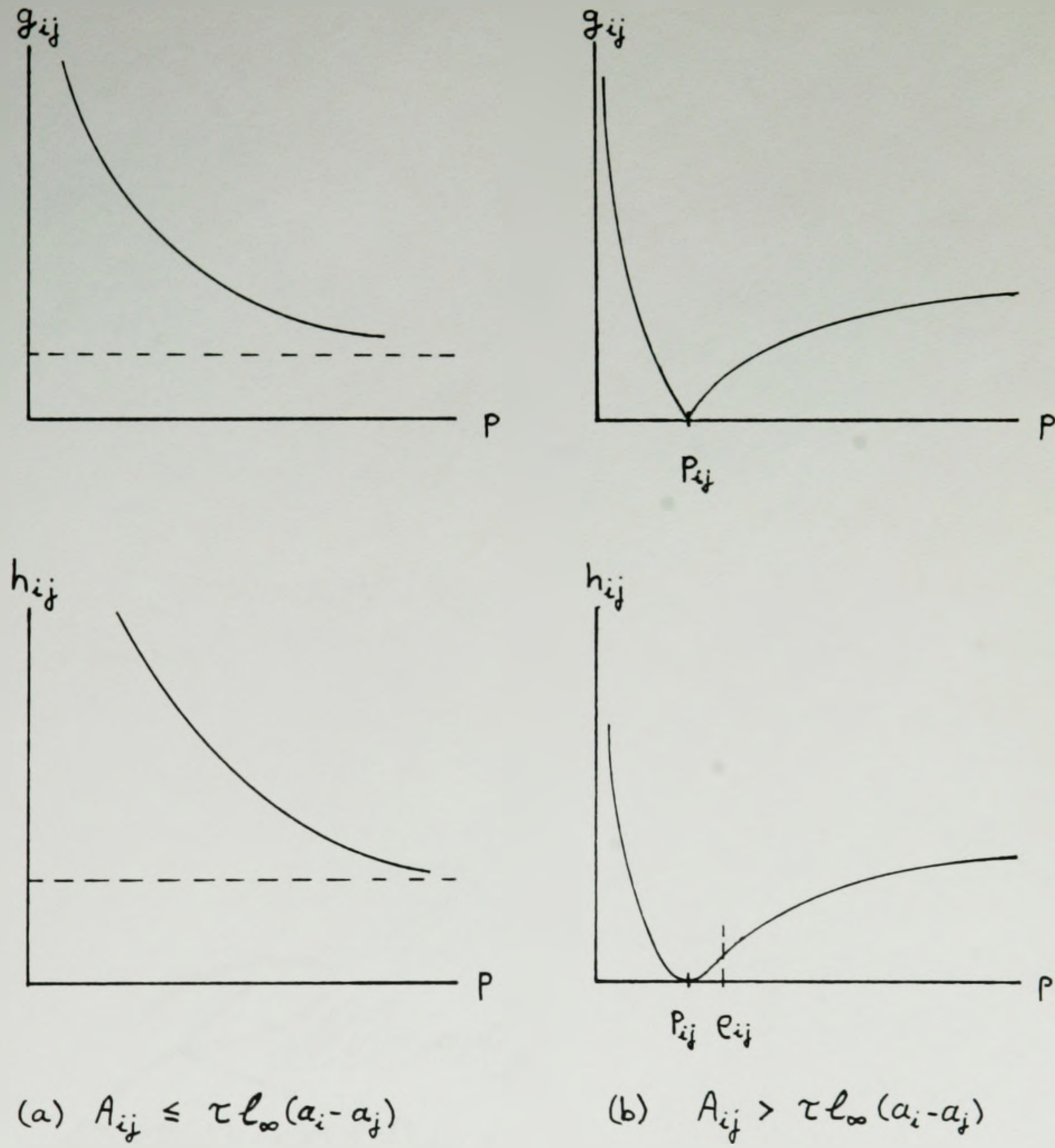


Figure 3.2.1 General Shape of g_{ij} and h_{ij}.

$$0 < p_1^*, p_2^* \leq 2,$$

where p_1* and p_2* are any values of p which minimize AD_ℓ and SD_ℓ respectively for the specified τ.

Proof:

The proof relies on the fact that the shortest distance between two points is the Euclidean (straight-line) distance between them. Hence

$$A_{ij} \geq \ell_2(a_i - a_j), \quad \forall\, i,j. \tag{3.2.23}$$

Since $\ell_p(a_i-a_j)$ is a non-increasing function of p (Property 2.1.1), and $\tau \leq 1$, therefore

$$\begin{aligned} A_{ij} - \tau\ell_p(a_i - a_j) &\geq A_{ij} - \tau\ell_2(a_i - a_j) \\ &\geq 0, \quad \forall\ i,j, \text{and } p > 2. \end{aligned} \tag{3.2.24}$$

Furthermore, the first inequality in (3.2.24) is satisfied strictly for each pair (i,j) with a_i-a_j not parallel to an axis. Thus,

$$\begin{aligned} AD_\ell(p > 2) &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} |\tau\ell_p(a_i - a_j) - A_{ij}| \\ &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (A_{ij} - \tau\ell_p(a_i - a_j)) \\ &> \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (A_{ij} - \tau\ell_2(a_i - a_j)) \\ &= AD_\ell(p = 2). \end{aligned} \tag{3.2.25}$$

Similarly,

$$SD_\ell(p > 2) > SD_\ell(p = 2). \tag{3.2.26}$$

Therefore, we conclude that any value of p which minimizes AD_ℓ or SD_ℓ lies in the interval (0,2].

Property 3.2.7

Let $K = \{(i,j)|a_i - a_j$ is not parallel to an axis, $i<j\}$, and assume that K is non-empty. Also assume that $A_{ij} > \tau\ell_\infty(a_i - a_j)$, $\forall(i,j) \in K$, where τ has some specified value. Then

$$p_m \le p_1^*, p_2^* \le p_M$$

where

$$p_m = \min_{(i,j) \in K} p_{ij} \text{ and } p_M = \max_{(i,j) \in K} p_{ij}.$$

Proof:

For pairs $(i,j) \notin K$, the corresponding terms g_{ij} and h_{ij} are constant functions of p by Properties 3.2.3 and 3.2.4. For pairs $(i,j) \in K$, g_{ij} and h_{ij} are decreasing for $0 < p \le p_{ij}$, while these terms are increasing for $p \ge p_{ij}$, again using Properties 3.2.3 and 3.2.4. It follows that AD_ℓ and SD_ℓ are decreasing functions of p for $0 < p \le p_m$, and increasing functions of p for $p \ge p_M$. Therefore, we conclude that any value of p which minimizes AD_ℓ or SD_ℓ for the specified τ, lies in the interval $[p_m, p_M]$.

The preceding result has some practical implications. First note that the p_{ij} can be obtained with relative ease, using standard techniques such as interval bisection or Newton-Raphson's method (e.g., see Dahlquist and Björck, 1974, Chapter 6), since g_{ij} (or h_{ij}) is a unimodal function of p. With the interval $[p_m, p_M]$ specified, we now have lower and upper bounds on the values of p_1^* and p_2^*, so that the search for p_1^* and p_2^* should be confined to this interval. This provides a substantial improvement over the current practice, in which a safe range for the search is left to the arbitrary discretion of the analyst.

The range $[p_m, p_M]$ also characterizes or describes the road network in a new way. If the width of the interval, measured by $p_M - p_m$, is small, we can regard the road network as being consistent with the distance function d_ℓ, in that d_ℓ approximates individual travel distances consistently with a high degree of accuracy. Thus, the width of the interval $[p_m, p_M]$

can be considered as a measure of the consistency of the road network with the distance function d_ℓ. We also observe that the specification of $[p_m, p_M]$ is especially useful for consistent road networks, since it reduces the search for p_1^* and p_2^* to a narrow interval in this case.

Let us now examine another aspect of the overall problem, namely that of finding τ_1^* and τ_2^*, the values of τ which minimize AD_ℓ and SD_ℓ respectively for a given p. The convexity results of Properties 3.2.1 and 3.2.2 can be put to good use here. First we consider the sum SD_ℓ which is strictly convex in τ (Property 3.2.2). Thus, a necessary and sufficient condition for minimizing SD_ℓ with p fixed is given by

$$\frac{\partial SD_\ell}{\partial \tau} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{A_{ij}} [\tau \ell_p(a_i - a_j) - A_{ij}] \ell_p(a_i - a_j)$$
$$= 0\,, \tag{3.2.27}$$

which provides the closed form solution,

$$\tau_2^* = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \ell_p(a_i - a_j)}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} [\ell_p(a_i - a_j)]^2 / A_{ij}}\,. \tag{3.2.28}$$

Next consider the sum AD_ℓ which is convex in τ (Property 3.2.1). We have

$$\frac{\partial AD_\ell}{\partial \tau} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sign}[\tau \ell_p(a_i - a_j) - A_{ij}] \ell_p(a_i - a_j)\,, \tag{3.2.29}$$

wherever this derivative exists, so that AD_ℓ is also piecewise linear in τ with discontinuities in the slope at

$$\tau_{ij} = A_{ij} / \ell_p(a_i - a_j)\,, \quad i,j = 1, \ldots, n,\ i < j\,. \tag{3.2.30}$$

It is interesting to note that the shape of AD_ℓ for varying τ is similar to that of the objective function in one coordinate for the unconstrained single facility minisum problem with rectangular distances (e.g., see Love, Morris and Wesolowsky, 1988, p. 18-22). Hence an analogous solution method can be used to obtain $\tau_1{}^*$. An outline of this method is given below.

Algorithm 3.2.1 {Finding $\tau_1{}^*$ for given p}

Step 1: Calculate each τ_{ij} using (3.2.30).

Step 2: Sequence and re-label the pairs $(\tau_{ij}, \ell_p(a_i-a_j))$ as (τ_r, ℓ_r), $r = 1, \ldots, n(n-1)/2$, such that $\tau_1 \leq \tau_2 \leq \ldots \leq \tau_{n(n-1)/2}$; i.e., the τ_{ij} are arranged in non-decreasing order.

Step 3: (Finding $\tau_1{}^*$)

Set $\quad i = 0,$

$$S_i = -\sum_{r=i+1}^{n(n-1)/2} \ell_r .$$

Repeat

$$i \leftarrow i + 1,$$

$$S_i \leftarrow S_{i-1} + 2\,\ell_i$$

until $S_i \geq 0$.

If $S_i = 0$ then $\tau_1{}^* \in [\tau_i, \tau_{i+1}]$,

else $(S_i > 0)$ $\tau_1{}^* = \tau_i$.

It should be clear that $\tau_1{}^*$ and $\tau_2{}^*$ are functions of p; furthermore, the curves $\tau_1{}^*(p)$ and $\tau_2{}^*(p)$ are readily obtained using Algorithm 3.2.1 and equation (3.2.28) respectively. With this information we can immediately calculate $AD_\ell{}^*$ and $SD_\ell{}^*$, the minimum values of AD_ℓ and SD_ℓ as functions of p; i.e.,

$$AD_\ell{}^*(p) = AD_\ell(\tau_1{}^*(p), p)\,,\; SD_\ell{}^*(p) = SD_\ell(\tau_2{}^*(p), p)\,. \qquad (3.2.31)$$

Thus, the criteria specified in (3.2.1) and (3.2.2) are now reduced to minimization problems in one variable.

Using the preceding result, we outline a strategy for finding the global minimizers of AD_ℓ and SD_ℓ, denoted by $(\tau_1{}^{**}, p_1{}^{**})$ and $(\tau_2{}^{**}, p_2{}^{**})$ respectively.

Algorithm 3.2.2 {Solving Criteria 1 and 2}

Step 1: Determine the curve of $AD_\ell{}^*(p)$ (or $SD_\ell{}^*(p)$), with a small enough increment Δp to identify all sub-intervals containing a local minimum. Delete those sub-intervals which obviously do not possess a global solution, and label the remaining ones 1, . . . , M.

Step 2: Do the following for $i = 1, \ldots, M$.

Divide sub-interval i using a smaller increment of p. With the additional points, reduce the width of the sub-interval containing the local minimum (denoted by $p^{(i)}$). Repeat this process until $p^{(i)}$ is calculated to the desired accuracy.

Step 3: Set $p_1{}^{**} = p^{(k)}$ and $\tau_1{}^{**} = \tau_1{}^*(p^{(k)})$, where $AD_\ell{}^*(p^{(k)}) = \min_i\{AD_\ell{}^*(p^{(i)})\}$. (For SD_ℓ, set $p_2{}^{**} = p^{(k)}$ and $\tau_2{}^{**} = \tau_2{}^*(p^{(k)})$, where $SD_\ell{}^*(p^{(k)}) = \min_i\{SD_\ell{}^*(p^{(i)})\}$.)

The preceding algorithm does not specify a range of p in step 1 for $AD_\ell{}^*$ (or $SD_\ell{}^*$), which guarantees that a global solution will eventually be found. This question is addressed in the next section.

3.3 General Considerations on the Use of the Weighted ℓ_p Norm

Section 3.1 discusses the directional bias of norms in general, and the ℓ_p norm in particular. Implicit in this discussion is the fact that the directional bias function pertains to a particular set of orthogonal reference axes. Thus, characteristics such as the directions of greatest and least difficulty of the distance function are measured relative to the given axes. Except for the weighted Euclidean norm (see (3.1.10)), rotating the axes results in a different directional bias, or alternatively, the distance between any two points varies under this rotation.

In Section 3.2, where fitting of the parameters τ and p is discussed, we assume that the reference axes are pre-specified. The resulting estimates $(\tau_1{}^{**}, p_1{}^{**})$ or $(\tau_2{}^{**}, p_2{}^{**})$ obviously depend on this choice of axes. Consider, as an example, the hypothetical case where the roads in a transportation network form a perfect rectangular grid and the destinations a_i are all situated at intersection points of the roads. If the axes are chosen parallel to the grid, then $\tau_1^{**} = \tau_2^{**} = 1$, $p_1^{**} = p_2^{**} = 1$, and actual distances are predicted exactly. On the other hand, if the axes are specified at 45° to the grid, then $\tau_1^{**} = \tau_2^{**} = \sqrt{2}$ and $p_1^{**} = p_2^{**} = +\infty$, with actual distances being predicted exactly once again. However, for any other choice of axes, the parameters will take on intermediate values, and the predicted distances will not coincide with the actual.

Clearly, the specification of the reference axes is an important part of any empirical study. One must recognize the dual relation between the distance function and the reference axes. Both are required in order to obtain a specific form of the directional bias. The choice of axes and distance function should be made after a careful study of the road network. Based on the predominant pattern of the roads, one should ascertain the directions which are easiest and most difficult to travel in. The axes and the distance function should be chosen accordingly to coincide with this directional bias.

Specification of the reference axes for the distance function based on an identification and examination of the patterns in the road network, does not appear to be a consideration in the empirical studies described in the literature (e.g., Love and Morris, 1972, 1979, and 1988, Love, Truscott and Walker, 1985, Ward and Wendell, 1980 and 1985). In other words, the axes are chosen arbitrarily without examining the physical nature of the system. Before showing the advantages and usefulness of our approach for the weighted ℓ_p norm, some definitions are in order.

Definition 3.3.1

The normalized travel distance between destinations a_i and a_j in the data set is given by

$$a_{ij} = A_{ij}/\ell_2(a_i - a_j), \quad i,j = 1, \ldots, n, \quad i < j \,. \tag{3.3.1}$$

Definition 3.3.2

A set of axes having orientation γ means that these axes are rotated counter-clockwise by an angle γ from true east and north.

Definition 3.3.3.

Let $R(\theta;\gamma_0)$ denote a function of θ, where the angle θ is measured relative to a set of axes having orientation γ_0 (see Figure 3.3.1). Then $R(\theta;\gamma_0)$ is said to have a rectangular bias if, and only if, the following conditions are satisfied:

(i) $R(\theta + \pi/2; \gamma_0) = R(\theta; \gamma_0)$, $\forall\theta$ {periodicity of $\pi/2$};

(ii) $R(\pi/4 - \Omega;\gamma_0) = R(\pi/4 + \Omega;\gamma_0)$, $0 \le \Omega \le \pi/4$ {symmetry property};

(iii) R is non-decreasing for $\theta \in [0, \pi/4]$ and non-increasing for $\theta \in [\pi/4, \pi/2]$ {unimodal cycle with maximum at $\theta = \pi/4$}.

Definition 3.3.4

A transportation network has a predominant rectangular pattern relative to a set of axes with orientation γ_0 if, and only if, the following relation is satisfied:

$$\alpha(\theta; \gamma_0) = \beta_0 + \beta_1 R(\theta; \gamma_0) + \epsilon(\theta; \gamma_0); \tag{3.3.2}$$

where $\beta_0, \beta_1 \ge 0$ are parameters with at least one β_i strictly positive, α is the normalized travel distance from any point q to any point s, where vector $(s - q)$ has direction θ (Figure 3.3.1), R is a function with rectangular bias, ϵ is an independent error term with mean zero, and $(\theta;\gamma_0)$ denotes that the angle θ is measured relative to the axes with orientation γ_0.

An obvious example of rectangular bias occurs when

$$R(\theta; \gamma_0) = r_p(\theta; \gamma_0), \quad 0 < p \le 2 \,. \tag{3.3.3}$$

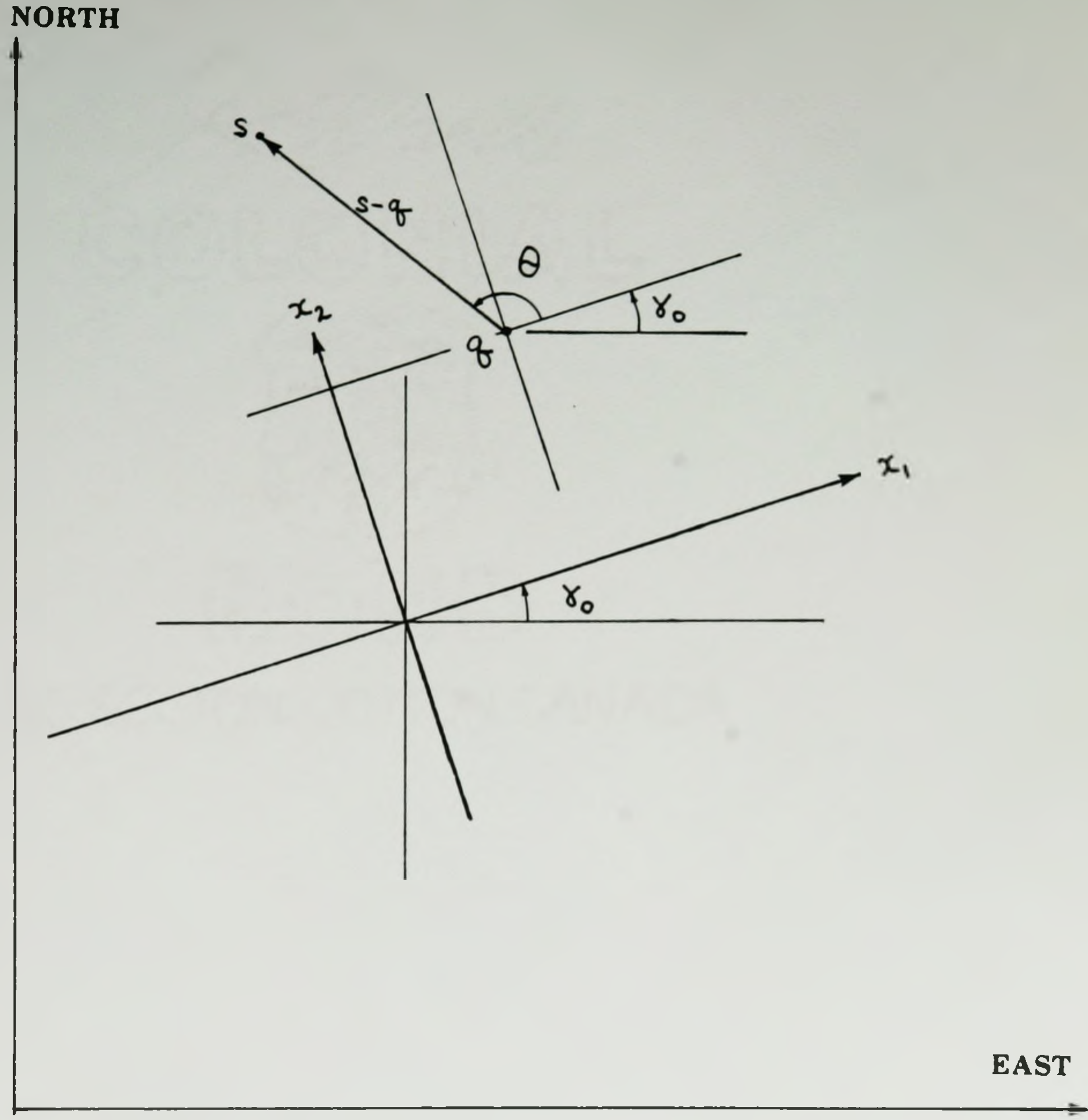


Figure 3.3.1 Reference Axes (x_1, x_2) with Orientation γ_0.

Typically we expect the road network to have an underlying rectangular grid (identified as the predominant pattern), offset by some angle γ_0 to the true east and north directions. We shall observe this condition in our case study of the road system in southern Ontario, at the end of the chapter. The normalized travel distance would be modelled in this case as

$$\alpha(\theta;\gamma_0) = \beta_0 + \beta_1 r_1(\theta;\gamma_0) + \epsilon(\theta;\gamma_0)\ . \tag{3.3.4}$$

In a transportation network with a predominant rectangular pattern, the direction of greatest difficulty is at $\gamma_0 + m\pi/4$, m = 1,3,5,7, while the direction of least difficulty is at $\gamma_0 + m\pi/2$, m = 0,1,2,3. This signifies that for pairs of points separated by the same straight-line distance, the actual travel distances are generally greatest at 45° to the set of axes with orientation γ_0 and least parallel to these axes. For the special case where p = 2 in (3.3.3), we have

$$\alpha(\theta;\gamma_0) = (\beta_0 + \beta_1) + \epsilon(\theta;\gamma_0)\ ; \tag{3.3.5}$$

i.e., the normalized travel distance is a constant plus an error term. This signifies a highly-developed network, with travel in any direction having the same degree of difficulty on average. Also note that $R(\theta;\gamma_0)$ does not necessarily represent the directional bias of a norm. For example, if the road system has one-way streets or obstacles resulting in a lot of back-tracking, then R might belong to a hyper-rectilinear distance function (i.e., $p<1$ in (3.3.3)).

Consider again the weighted ℓ_p norm, with directional bias function $\tau r_p(\theta;\gamma)$, where γ now specifies the orientation of the reference axes pertaining to the distance function. We can specify a third criterion for fitting the parameters τ and p, based on a minimization of the sum of squared normalized deviations, as follows.

Criterion 3

$$\underset{\tau,\,p}{\text{minimize}}\quad SND_\ell = \sum_{i=1}^{n-1}\ \sum_{j=i+1}^{n} [\tau r_p(\theta_{ij};\gamma) - \alpha_{ij}]^2, \tag{3.3.6}$$

where

$$\theta_{ij} = \tan^{-1}\left(\frac{a_{j2}(\gamma) - a_{i2}(\gamma)}{a_{j1}(\gamma) - a_{i1}(\gamma)}\right), \ \forall\ i,j, \tag{3.3.7}$$

the γ in (3.3.7) indicating that the coordinates of a_i and a_j are measured in the reference axes of the distance function.

The advantage of choosing γ to comply with the physical nature of the transportation network being modelled is shown in the following important result.

Theorem 3.3.1

Suppose we have a transportation network with a predominant rectangular pattern as in (3.3.2), and the weighted ℓ_p distance function is used to model travel distances in this network. If $\gamma = \gamma_0$ (i.e., the reference axes of the distance function coincide with those of the network), and the sample size of destinations in the data set is sufficiently large (i.e., the asymptotic limiting case $n \to \infty$), then

$$0 < p_3^{**} \leq 2\,, \tag{3.3.8}$$

where p_3^{**} is the value of p at a global minimum of SND_ℓ.

Proof:

The proof is by contradiction. Consider a value of $p > 2$, and let

$$\begin{aligned} \bar{a}_{ij} &= \beta_0 + \beta_1 R(\theta_{ij}\,;\,\gamma_0) \\ &= a_{ij} - \epsilon(\theta_{ij}\,;\,\gamma_0) \quad , \quad \forall\, i,j\,. \end{aligned} \tag{3.3.9}$$

Noting that $\gamma = \gamma_0$, and deleting both for notational convenience, we obtain

$$\begin{aligned} SND_\ell &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} [\tau r_p(\theta_{ij}) - \bar{a}_{ij} + \bar{a}_{ij} - a_{ij}]^2 \\ &= \sum\sum_{i<j} (\tau r_p(\theta_{ij}) - \bar{a}_{ij})^2 + \sum\sum_{i<j} \epsilon^2(\theta_{ij}) - 2\sum\sum_{i<j} (\tau r_p(\theta_{ij}) - \bar{a}_{ij})\,\epsilon(\theta_{ij})\,. \end{aligned} \tag{3.3.10}$$

In the limiting case, $n \to \infty$,

$$\sum_{i<j}\sum (\tau r_p(\theta_{ij}) - \bar{a}_{ij})\,\epsilon(\theta_{ij}) \to 0 \,, \tag{3.3.11}$$

since $\epsilon(\theta)$ is an independent random variable with mean of zero. Thus, for sufficiently large samples,

$$SND_\ell \simeq \sum_{i<j}\sum (\tau r_p(\theta_{ij}) - \bar{a}_{ij})^2 + \sum_{i<j}\sum \epsilon^2(\theta_{ij}) \,. \tag{3.3.12}$$

The second summation in (3.3.12) is a constant, so that we only need to consider the first summation for the minimization of SND_ℓ. Clearly, a τ' can be found (for $p>2$), such that

$$\begin{aligned} |\tau' r_2(\theta) - \bar{a}(\theta)| &= |\tau' - \bar{a}(\theta)| \\ &\le |\tau r_p(\theta) - \bar{a}(\theta)|, \quad \forall\theta \,. \end{aligned} \tag{3.3.13}$$

(This is illustrated in Figure 3.3.2.) It follows that (3.3.8) must be true.

Corollary 3.3.1

If in addition $R(\theta;\gamma_0) \equiv r_1(\theta;\gamma_0)$, then

$$1 \le p_3^{**} \le 2 \,. \tag{3.3.14}$$

Proof:

We know from the preceding theorem that $0 < p_3^{**} \le 2$. Therefore it only remains to be shown that $p_3^{**} \ge 1$. Deleting γ_0 again for notational convenience, we have

$$\begin{aligned} \bar{a}(\theta) &= \beta_0 + \beta_1 r_1(\theta) \\ &= \beta_T[\lambda + (1-\lambda)\, r_1(\theta)], \end{aligned} \tag{3.3.15}$$

where

$$\beta_T = \beta_0 + \beta_1, \tag{3.3.16}$$

and

$$0 \le \lambda = \beta_0/\beta_T \le 1 \,. \tag{3.3.17}$$

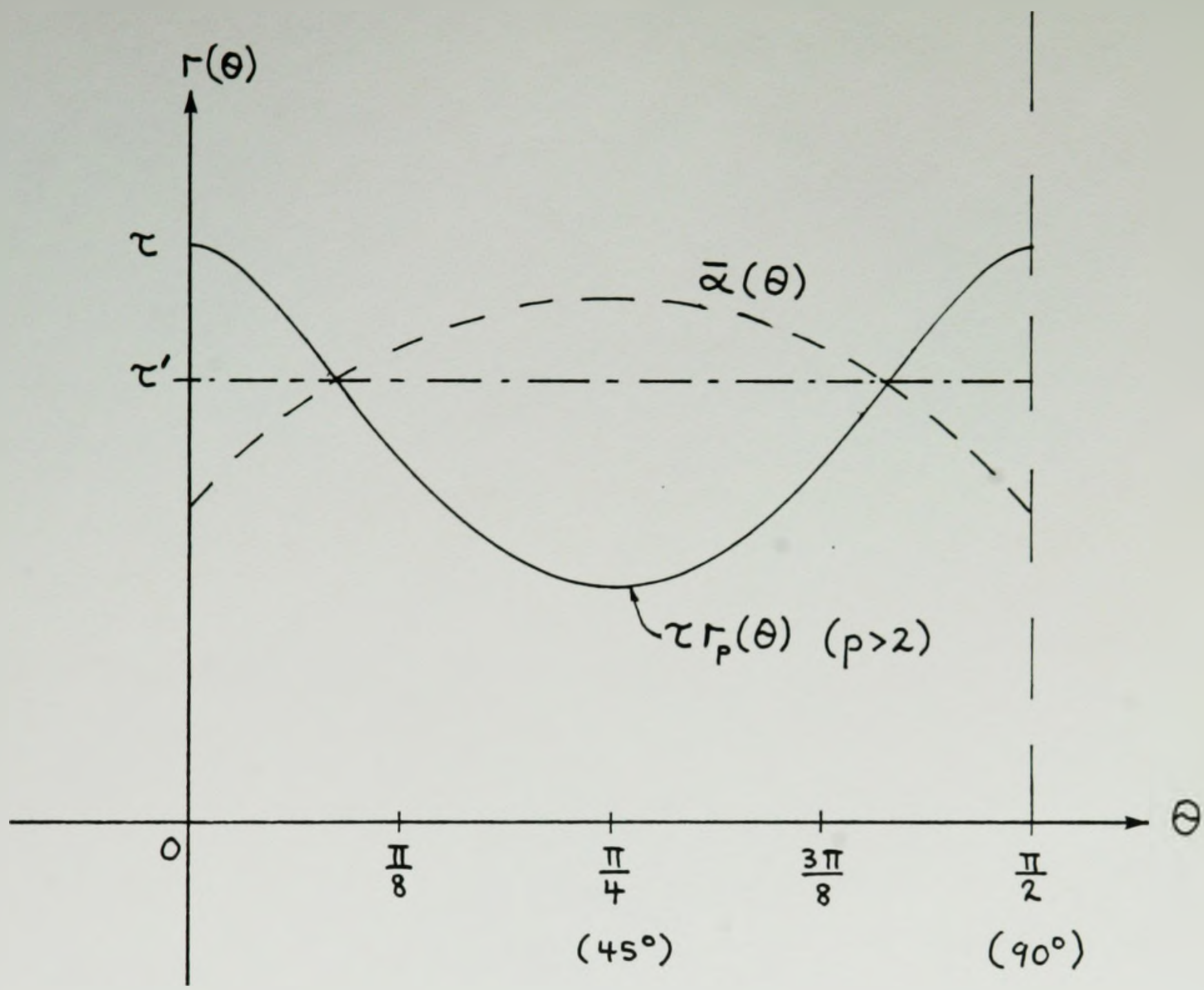


Figure 3.3.2 Directional Bias Functions which are Out-of-Phase by 45°.

Since travel distance is always at least as large as the straight-line distance, and $\bar{\alpha}(0) = \beta_0 + \beta_1 = \beta_T$, it follows that

$$\beta_T \geq 1 \tag{3.3.18}$$

The general shape of $\bar{\alpha}$ is illustrated in Figure 3.3.3, where by symmetry we only need to consider $0 \leq \theta \leq \pi/4$. Referring to Figure 3.3.3, it should be clear that for $\tau > 0$ and $p \in (0,1)$, a τ' can always be found such that

$$|\tau' r_1(\theta) - \bar{\alpha}(\theta)| \leq |\tau r_p(\theta) - \bar{\alpha}(\theta)| \, . \tag{3.3.19}$$

Hence, a global optimum of SND_ℓ exists with $p_3^{**} \geq 1$, ending the proof.

Using a similar method as in Property 3.2.2, it is readily shown that SND_ℓ is a strictly convex function of the inflation factor τ. We can proceed as in the derivation of (3.2.28), to obtain a closed form solution for τ_3^*, the minimizer of SND_ℓ for a fixed p. Thus, we obtain

$$\tau_3^*(p\,;\gamma) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \alpha_{ij}\, r_p(\theta_{ij}\,;\gamma)}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} r_p^2(\theta_{ij}\,;\gamma)} \, . \tag{3.3.20}$$

The preceding theorem and its corollary are readily extended to the criteria 1 and 2 given in (3.2.1) and (3.2.2) respectively. Thus, p_3^{**} can be replaced by p_t^{**}, $t = 1,2,3$, in (3.3.8) and (3.3.14). These results lead to the following general procedure for modelling travel distances in a transportation network with the weighted ℓ_p distance function.

Step 1: Verify that the transportation network has a predominant rectangular pattern with respect to a set of axes having some orientation (γ_0). If this is not the case, a different distance function should be considered.

Step 2: Orient the reference axes of the weighted ℓ_p distance function to coincide with those of the network (i.e, $\gamma = \gamma_0$).

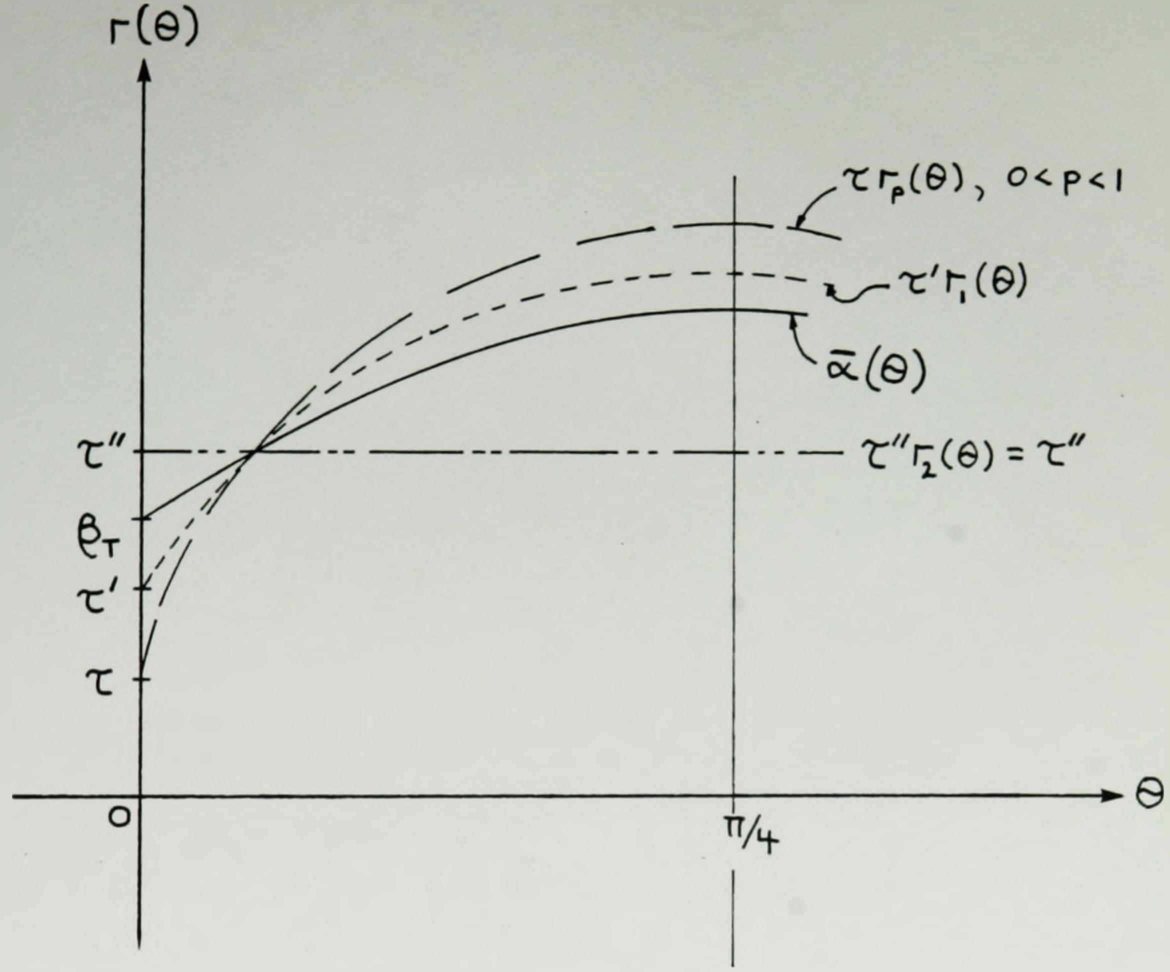


Figure 3.3.3 Fitting $r_p(\theta)$ to a Road Network with Underlying Rectangular Grid.

Step 3: Determine the best-fitting values of the parameters τ and p for one or more of the criteria given above, where p is restricted to the interval (0,2] or [1,2] in accordance with Theorem 3.3.1 and Corollary 3.3.1.

In brief, the procedure first involves a verification that the weighted ℓ_p function is appropriate for the network being modelled. Next the axes are rotated so that the distance function is in phase with the network. Only at this point, are we ready to solve for the parameters τ and p. The advantages of such an approach can be summarized as follows:

(i) the directional bias inherent in the road network is reproduced by the distance function;

(ii) because the reference axes coincide with those of the network, we can expect the best overall fit of the weighted ℓ_p function to actual distances; and

(iii) the search for the minimizing values of τ and p (e.g., using Algorithm 3.2.2) can be done efficiently, since p is effectively restricted by Theorem 3.3.1 and its corollary to a small interval.

As a simple example, consider an underlying pattern of roads intersecting at an oblique angle, so that the network does not have a rectangular bias. Use of the weighted ℓ_p function in this case would lead to relatively poor estimates of actual distances. Instead, we would be better off fitting the parameters τ and p of a modified function of the form $\tau\ell_p(Ax;\gamma_0)$. Here A is a non-singular, non-orthogonal 2×2 matrix which transforms the coordinates in a set of reference axes with orientation γ_0, to reproduce the oblique pattern of the road network.

To illustrate the general considerations given above on the use of the weighted ℓ_p function, we present a case study of the road system covering the central and eastern parts of southern Ontario. Eighteen representative cities are chosen from this region to form the data set. These cities are listed in Table 3.3.1, with their coordinates measured in the base axes pointing true east and north. From an inspection of the official road map, a section of which is

Table 3.3.1 – Cities Forming the Data Set

City No.	City Name	Coordinates (1/4" Unit)[(a),(b)] x_1	x_2
1	Windsor	2.0	−5.3
2	Sarnia	12.7	9.3
3	Chatham	15.35	−3.2
4	London	31.0	8.8
5	Kitchener/Waterloo	42.9	18.7
6	Brantford	46.6	11.6
7	Hamilton	53.2	13.9
8	Toronto	61.0	23.2
9	Fort Erie	68.0	6.7
10	St. Catharines	63.0	12.0
11	Stratford	35.3	17.0
12	Goderich	24.2	25.3
13	Barry	56.0	38.7
14	Owen Sound	36.7	43.0
15	Peterborough	77.3	36.7
16	Belleville	92.1	34.3
17	Ottawa	117.7	61.6
18	Cornwall	132.5	53.8

(a) SCALE: 9.5 units = 50 km.

(b) Coordinates measured in base axes (east, north).

shown in Figure 3.3.4, we can discern an approximate rectangular pattern underlying the network. The base of this rectangular grid is formed in large part by the number 401 Highway, which is a major route in the network following the north shore of Lake Ontario and the St. Lawrence River. Therefore, using our general procedure, we should rotate the reference axes of the distance function to parallel the 401 Highway.

The effect of rotating the reference axes on the minimum value of SD_ℓ (Criterion 2), and the corresponding parameter values τ_2^{**} and p_2^{**}, is shown in Table 3.3.2. The main point of interest here is the sensitivity of p_2^{**} to the axis orientation γ, and the fact that p_2^{**} takes on a rather low value (1.4583) at $\gamma = 22.5°$. This suggests a substantial rectangular bias in the road system, at an orientation in line with the mean direction of the 401 Highway, thus confirming our earlier conclusion based on an inspection of the map. Also note that the fit, measured by the minimum value of SD_ℓ, is optimized at $\gamma = 22.5°$.

Referring to Table 3.3.2, we see that if the axes are arbitrarily set parallel to the east and north directions (i.e., $\gamma = 0°$), a value of p_2^{**} close to 2.0 is obtained. This can be explained by the fact that the distance function is now severely out of phase with the underlying pattern or structure of the network. The arbitrary specification of the reference axes in this fashion is common practice in the literature. For example, Love, Truscott and Walker (1985) use the east and north directions as the reference axes in their empirical study, which covers basically the same geographic region as ours. Not surprisingly, they also obtain a value of p_2^{**} close to 2.0. This leads to the erroneous conclusion that travel distances are essentially Euclidean multiplied by an inflation factor; i.e., there is virtually no rectangular bias in the road system. Thus, the case study confirms quite dramatically the usefulness of our general procedure for fitting the weighted ℓ_p distance function.

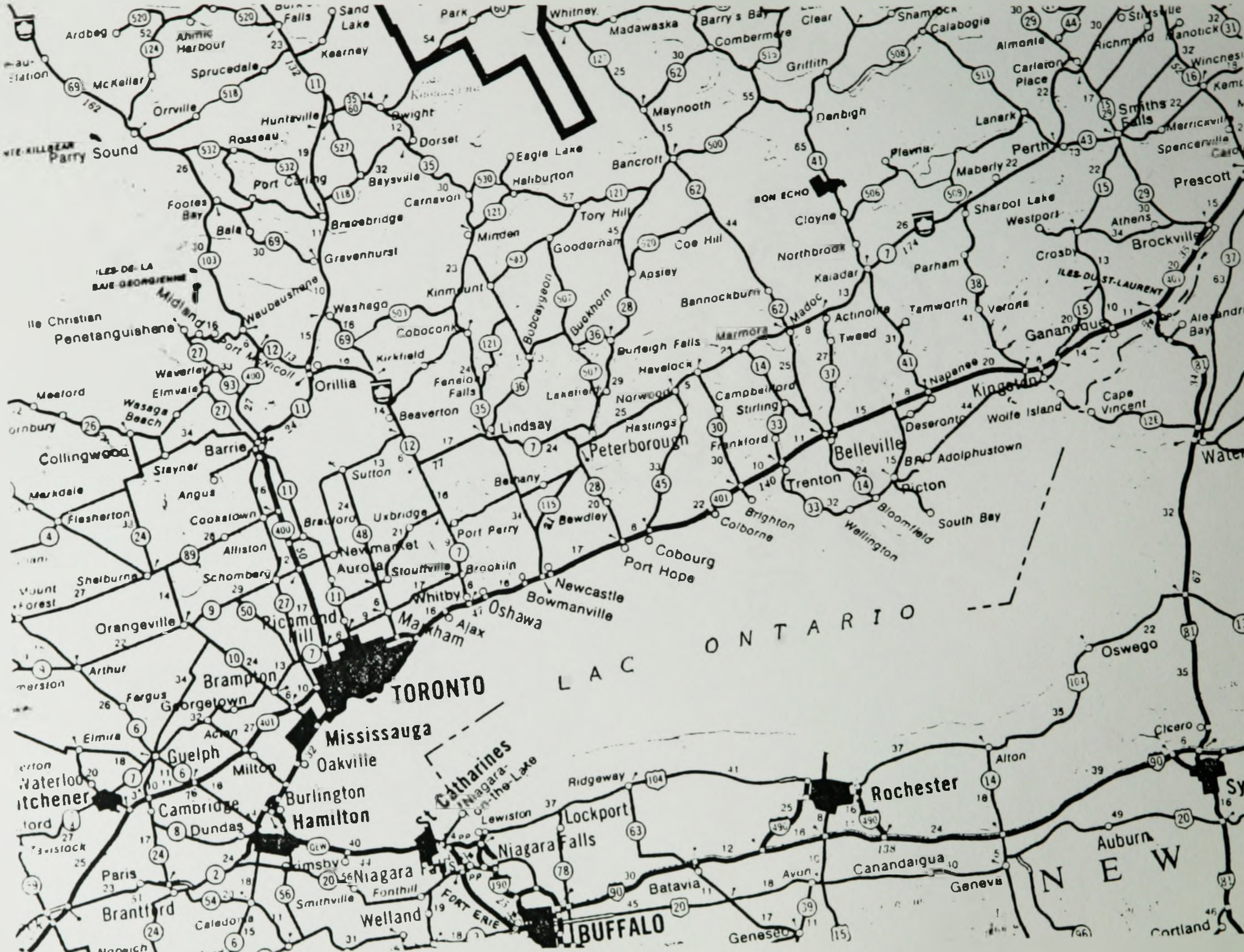


Figure 3.3.4 Section of Road Map for Southern Ontario.

Table 3.3.2 – Effect of Orientation (γ) on Best Fit

γ(°)	τ_2**	p_2**	min SD_ℓ
0	1.1118	1.9174	417.71
5	1.0874	1.7210	407.22
10	1.0765	1.6059	389.79
15	1.0715	1.5173	367.85
20	1.0715	1.4635	349.82
22.5	1.0733	1.4583	347.40
25	1.0756	1.4675	350.86
30	1.0826	1.5458	375.21
40	1.1158	1.9286	417.83
50	1.1693	2.5243	396.71
60	1.2211	3.0832	357.20
65	1.2353	3.2055	351.38
70	1.2324	3.1122	360.69
80	1.1778	2.4794	401.74

CHAPTER 4

THE WEIGHTED ONE-TWO NORM

The Euclidean and rectangular norms are the most commonly-used distance functions in continuous location models. Since the actual routes in the physical problem are not likely to follow purely straight-line or rectilinear paths, a logical extension would be to consider a positive linear combination of these two distances. Since this hybrid function is a positive sum of norms, it is also a norm. The purpose of this chapter is to examine our new distance measure, which we term the weighted one-two norm and denote as follows:

$$k_H(x; b_0, b_1) = b_0\, \ell_2(x) + b_1 \ell_1(x), \tag{4.1}$$

where $x \in R^N$; $b_0, b_1 \geq 0$, with at least one of these parameters being strictly positive; and ℓ_1, ℓ_2 are the rectangular and Euclidean norms.

Letting

$$b_T = b_0 + b_1 > 0, \tag{4.2}$$

we can rewrite (4.1) as

$$k_H(x; b_0, b_1) = b_T\, k_h(x; a_0), \tag{4.3}$$

with

$$a_0 = b_0/b_T,\ a_1 = b_1/b_T,\ a_0 + a_1 = 1, \tag{4.4}$$

and

$$k_h(x; a_0) = a_0\, \ell_2(x) + (1 - a_0)\, \ell_1(x)\,. \tag{4.5}$$

This allows an interesting physical interpretation of the weighted one-two norm. The shortest route between any two points q and r on a transportation network can be viewed as a composition of straight-line segments parallel to the vector $(q - r)$ and rectilinear segments. That is, part of the route follows a direction parallel to $q - r$, while the remainder is along a rectangular grid. Loosely-speaking, the parameter a_0 gives the proportion of the route which is Euclidean, while the remaining proportion, $a_1 = 1 - a_0$, is rectilinear. The parameter b_T takes on the role of an inflation factor, similar to τ in the weighted ℓ_p norm. For a transportation network with an underlying rectangular grid (i.e., $R \equiv r_1$ in (3.3.2)), this interpretation is intuitively appealing, since a typical trip moves along the grid part of the way and diagonal roads the rest.

We begin this chapter with a look at the directional bias function of the weighted one-two norm on $R^2(N = 2)$. This leads to an important observation concerning the relation between the families of norms, $k_h(x; a_0)$, $0 \le a_0 \le 1$, and $\ell_p(x)$, $1 \le p \le 2$. Next, we use our hybrid norm to develop a simple linear regression model for describing travel distances in a transportation network. Some novel applications are derived for this model, based on standard statistical tests. These concepts are illustrated with a continuation of the case study at the end of Chapter 3.

4.1 Directional Bias

From the definition in (3.1.1), the directional bias function of the weighted one-two norm with $b_T = 1$ is given by

$$\begin{aligned} r_h(\theta; a_0) &= k_h(x; a_0)/\ell_2(x) \\ &= a_0 + a_1\, \ell_1(x)/\ell_2(x) \\ &= a_0 + a_1\, r_1(\theta), \end{aligned} \tag{4.1.1}$$

where θ is defined in (3.1.2). Recalling that

$$0 \le a_0 \le 1, \text{ and } a_0 + a_1 = 1, \tag{4.1.2}$$

we see that r_h is a convex combination of the directional bias functions of the rectangular norm ($r_1(\theta)$) and Euclidean norm ($r_2(\theta) = 1$).

The following properties are immediately obvious from the results in Section 3.1, and are therefore given without proof.

Property 4.1.1

$r_h(\theta; \alpha_0)$ is periodic with period $\pi/2$.

Property 4.1.2

For any real Ω,

$$r_h(\pi/4 + \Omega\,; \alpha_0) = r_h(\pi/4 - \Omega; \alpha_0).$$

Property 4.1.3

If $\alpha_0 < 1$, then r_h is an increasing function of θ in the interval $[0, \pi/4]$.

From the preceding results and Definition 3.3.3, it is clear that $r_h(\theta; \alpha_0)$ exhibits a rectangular bias. We also observe that

$$r_h(\theta; 0) = r_1(\theta), \tag{4.1.3}$$

and

$$r_h(\theta; 1) = 1; \tag{4.1.4}$$

so that these two functions are precisely the same as $r_p(\theta)$ with $p = 1$ and 2 respectively. It is important to keep in mind that the reference axes for r_h and r_p are assumed to have the same orientation γ, and that γ is omitted to simplify the notation.

Recall from Section 3.1 that $r_p(\theta)$ is a decreasing function of p, except at $\theta = m\pi/2$, $m = 0, \pm 1, \pm 2, \ldots$, where it is constant. An analogous result holds for r_h, as shown below.

Property 4.1.4

r_h is a non-increasing function of its parameter α_0.

Proof:

Since $\alpha_1 = 1 - \alpha_0$, we can rewrite (4.1.1) as

$$r_h(\theta; \alpha_0) = r_1(\theta) + \alpha_0 (1 - r_1(\theta)). \tag{4.1.5}$$

Therefore,

$$\frac{\partial}{\partial \alpha_0} [r_h (\theta; \alpha_0)] = 1 - r_1(\theta) . \tag{4.1.6}$$

But $r_1(\theta) \geq 1$, with equality only at $\theta = m\pi/2$, where m is an integer. It follows that

$$\frac{\partial}{\partial \alpha_0} [r_h (\theta; \alpha_0)] \leq 0 , \tag{4.1.7}$$

and hence r_h is a non-increasing function of α_0. Furthermore, r_h is decreasing in α_0, except at $\theta = m\pi/2$, $m = 0, \pm 1, \pm 2, \ldots$, where it is constant.

We see then that r_h and r_p have the same type of bias (rectangular), for $\alpha_0 \in [0,1]$ and $p \in [1,2]$. Furthermore, from Property 4.1.4, equations (4.1.3) and (4.1.4), and the results in Section 3.1, it follows that a one-to-one correspondence can be established between r_h and r_p with α_0 and p increasing in their respective intervals [0,1] and [1,2] and where the amplitudes of r_h and r_p are decreasing simultaneously. Thus, we might expect that our hybrid norm closely approximates the ℓ_p norm, for an appropriate choice of α_0 as an increasing function of p. Before investigating this relation, we note a basic difference in the shapes of $r_h(\theta; \alpha_0)$ and $r_p(\theta)$. Whereas the latter has an inflection point in the interval $0 \leq \theta \leq \pi/4$ by Property 3.1.5, no such point exists in the former as shown below.

Property 4.1.5

r_h is a strictly concave function of θ in the interval $0 \leq \theta \leq \pi/2$, for any $\alpha_0 < 1$.

Proof:

For $0 \leq \theta \leq \pi/2$, we have

$$r_1(\theta) = \cos\theta + \sin\theta,$$

and thus,

$$r_h(\theta; \alpha_0) = \alpha_0 + \alpha_1 (\cos\theta + \sin\theta).$$

Since $\cos\theta$ and $\sin\theta$ are both strictly concave in this interval, and also, $\alpha_1 = 1 - \alpha_0 > 0$, the result follows immediately.

Let us now investigate the accuracy of the hybrid norm $k_h(x; \alpha_0)$ as an approximation of $\ell_p(x)$ for $1 \leq p \leq 2$, where the parameter α_0 is a function of p yet to be determined. As in Section 3.1, we use a normalized difference, defined here by

$$\Delta = \frac{k_h(x; \alpha_0) - \ell_p(x)}{\ell_2(x)} \quad , \quad \forall x \neq 0 . \tag{4.1.8}$$

Since α_0 is a function of p, this can be rewritten in the form,

$$\Delta(\theta; p) = r_h(\theta; \alpha_0(p)) - r_p(\theta). \tag{4.1.9}$$

Due to the periodicity and symmetry of the directional bias functions (Properties 3.1.2, 3.1.3, 4.1.1 and 4.1.2), it suffices to consider θ in the range,

$$0 \leq \theta \leq \pi/4 . \tag{4.1.10}$$

At $\theta = 0$, we have $r_p(0) = 1$ by (3.1.20), and furthermore,

$$r_h(0; \alpha_0) = \alpha_0 + \alpha_1 r_1(0)$$

$$= \alpha_0 + \alpha_1 = 1. \tag{4.1.11}$$

Thus,

$$\Delta(0; p) = 0, \tag{4.1.12}$$

independently of the choice of α_0 for a given p. In order to specify the form of $\alpha_0(p)$, we adopt a similar procedure as in Section 3.1 (see (3.1.38)), by imposing the following boundary condition:

$$\Delta(\pi/4; p) = 0 \ . \tag{4.1.13}$$

Noting that

$$\Delta(\pi/4\, ; p) = a_0 + (1 - a_0)\sqrt{2} - 2^{1/p - 1/2} \ ,$$

we can readily solve equation (4.1.13) to obtain

$$a_0 = \frac{2 - 2^{1/p}}{2 - \sqrt{2}} \ . \tag{4.1.14}$$

To signify that the difference function satisfies the specific constraint in (4.1.13), we denote it by $\delta(\theta; p)$. Noting that $|\cos\theta| = \cos\theta$ and $|\sin\theta| = \sin\theta$ for $0 \le \theta \le \pi/4$, and using (4.1.14), it is easily shown that

$$\delta(\theta\, ; p) = \frac{2 - 2^{1/p}}{2 - \sqrt{2}} + \left(\frac{2^{1/p} - \sqrt{2}}{2 - \sqrt{2}}\right)(\cos\theta + \sin\theta) - [\cos^p\theta + \sin^p\theta]^{1/p} \ , \tag{4.1.15}$$

where the ranges of interest are given by

$$0 \le \theta \le \pi/4 \quad \text{and} \quad 1 \le p \le 2 \tag{4.1.16}$$

The difference function $\delta(\theta;p)$ was evaluated on the computer over a finely-divided grid covering the points (θ,p) defined by (4.1.16). The results are summarized in Figures 4.1.1 and 4.1.2, where $\max_\theta|\delta|$ and the θ_* at which this occurs are plotted as functions of p. Observe that

$$\max_{\theta,p} |\delta| = 0.015431 \ , \tag{4.1.17}$$

at

$$p_{**} = 1.3498 \text{ and } \theta_{**} = 10.1637° \ , \tag{4.1.18}$$

where the maximizing values p_{**} and θ_{**} are obtained by the grid search to an accuracy of 0.0001 and 0.0001° respectively. In Figure 4.1.3, δ is plotted as a function of θ for sample fixed p. Observe that the profiles here are unimodal in shape. From Figure 4.1.2 we see that the value θ_* which maximizes $|\delta|$ for a given p, shifts to the right for increasing p. It is also interesting to note that

$$\delta(\theta; p) \ge 0, \ \forall\, \theta, p \ , \tag{4.1.19}$$

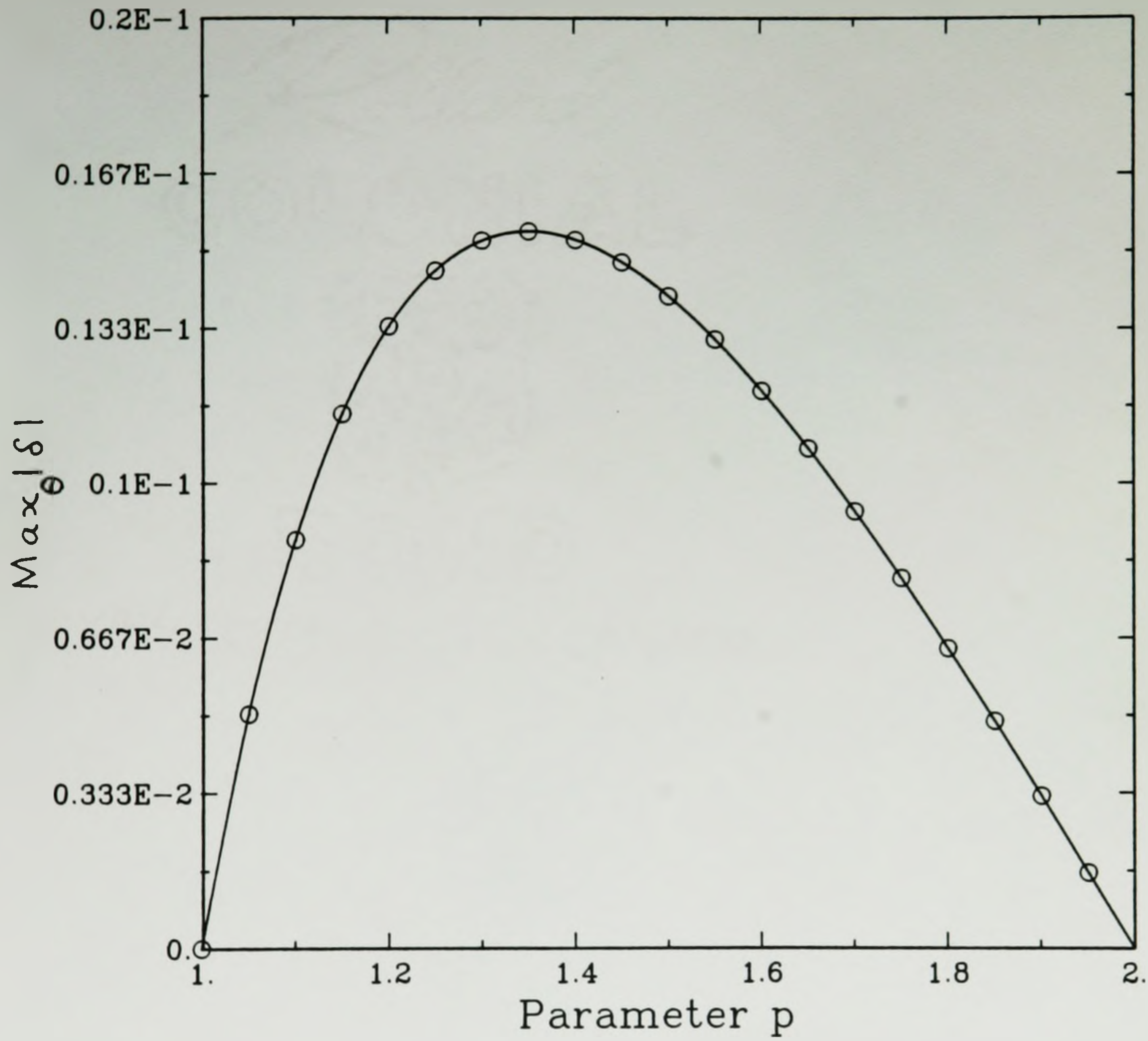


Figure 4.1.1 Maximizing $|\delta(\theta\,;\,p)|$ over θ.

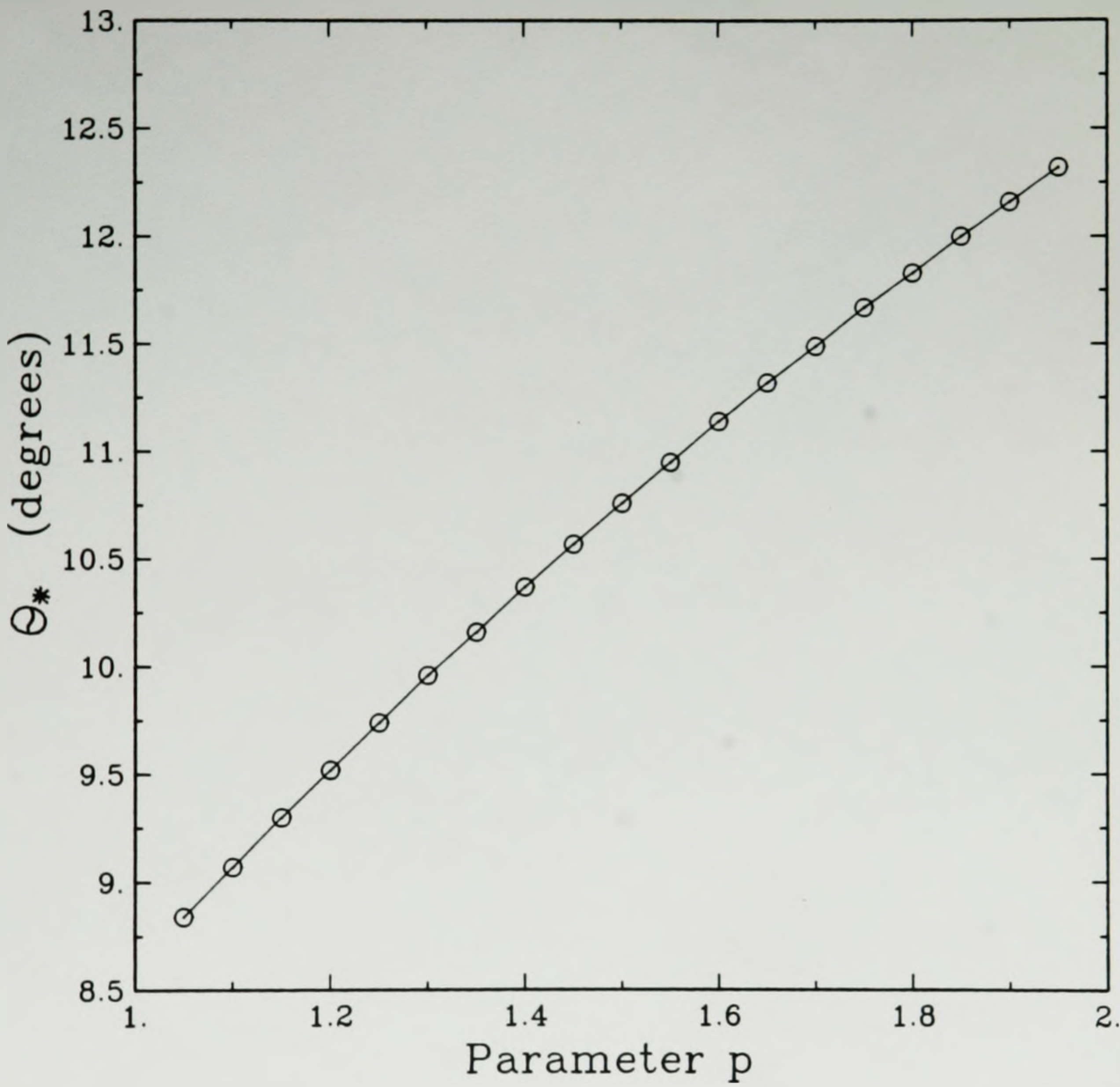


Figure 4.1.2 Maximizing Value of θ.

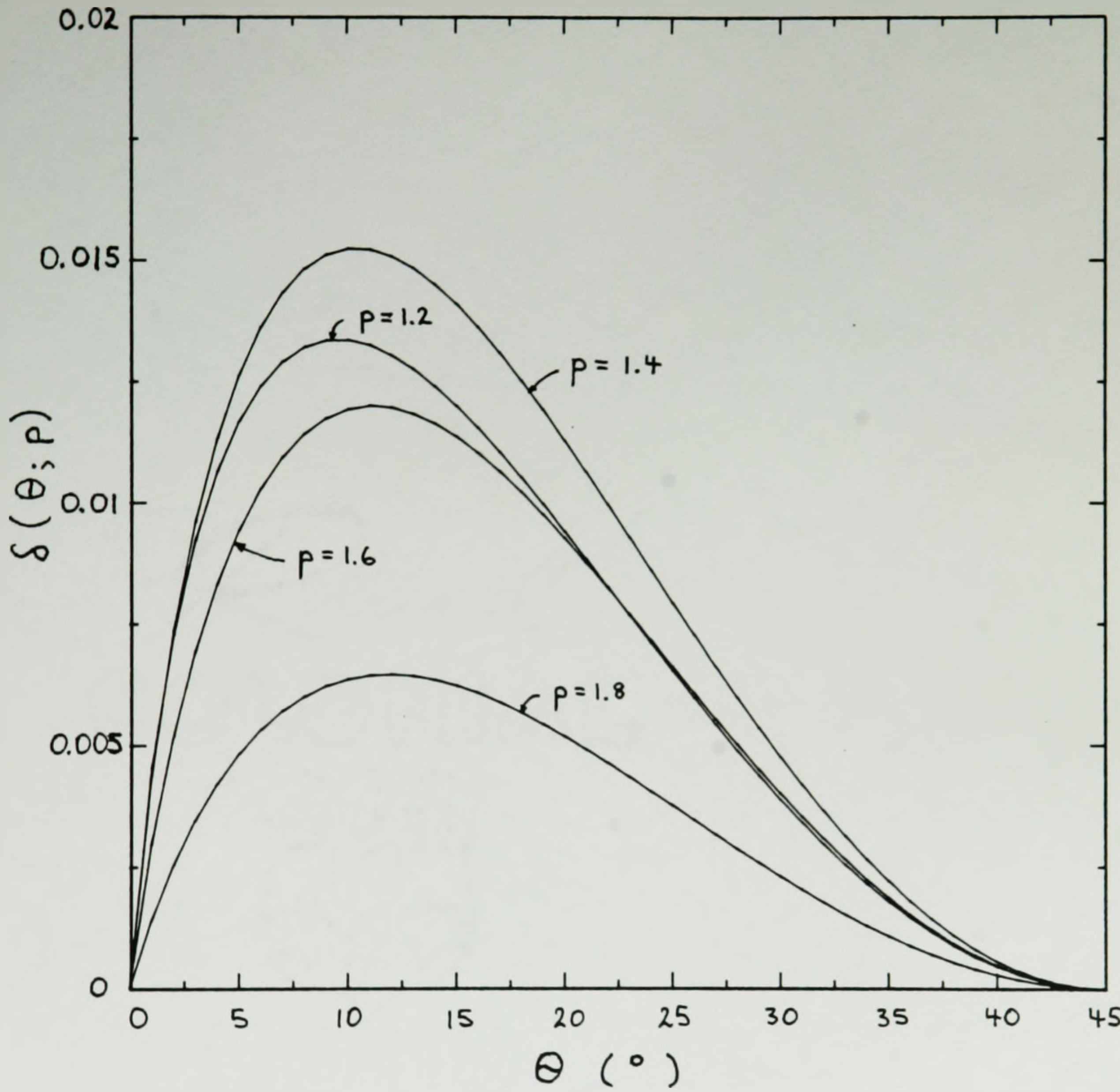


Figure 4.1.3 Profiles of $\delta(\theta;\, p)$ for Fixed p.

a result confirmed by the exhaustive grid search. Finally, in Figure 4.1.4, we show one cycle of $r_p(\theta)$ and its approximation by $r_h(\theta; a_0)$ for the sample case, $p = 1.4$ ($a_0 = 0.6134135$).

Returning to Figure 4.1.2, it is seen that θ_* varies in a nearly linear manner from approximately 8.8° to 12.3° as p increases from 1.05 to 1.95. These low values for θ_* can be explained qualitatively as follows. At 0°, the curve of r_p has a zero slope (equation (3.1.22)), while the right-sided derivative of r_h is given by

$$\frac{\partial}{\partial\theta^+} r_h(0; a_0) = a_1 \frac{d}{d\theta^+} r_1(0) = a_1 = 1 - a_0 . \tag{4.1.20}$$

Thus, a gap between the two curves is formed, and increases in size as θ moves to the right of 0° (e.g., see Figure 4.1.4). However, the second-order derivative of r_p with respect to θ is very large in the vicinity of 0° (equation (3.1.16)); so that the slope of r_p quickly catches up to and surpasses the slope of r_h, resulting in a transition from an increasing to decreasing gap size at a low value of θ_*.

Consider a transportation network with a predominant rectangular pattern, such that $R \equiv r_1$ for the model in (3.3.2). In this case, the weighted one-two norm (k_H) will give a better fit than the weighted ℓ_p norm (d_ℓ). This is readily seen if we substitute

$$r_H(\theta; b_0, b_1) = k_H(x; b_0, b_1)/\ell_2(x) = b_T\, r_h(\theta; a_0), \tag{4.1.21}$$

in place of $\tau r_p(\theta)$ in equation (3.3.12). Then the first term on the right-hand side of (3.3.12) goes to zero asymptotically as the sample size of destinations $n \to \infty$, since $b_0 \to \beta_0$ and $b_1 \to \beta_1$. Thus, for sufficiently large sample sizes, the sum of squared normalized deviations will be smaller with k_H than with d_ℓ. Alternatively, we can explain this property in physical terms as follows. For a transportation network with an underlying rectangular grid, the expected travel distance should increase at a positive rate as we move away from a direction of least

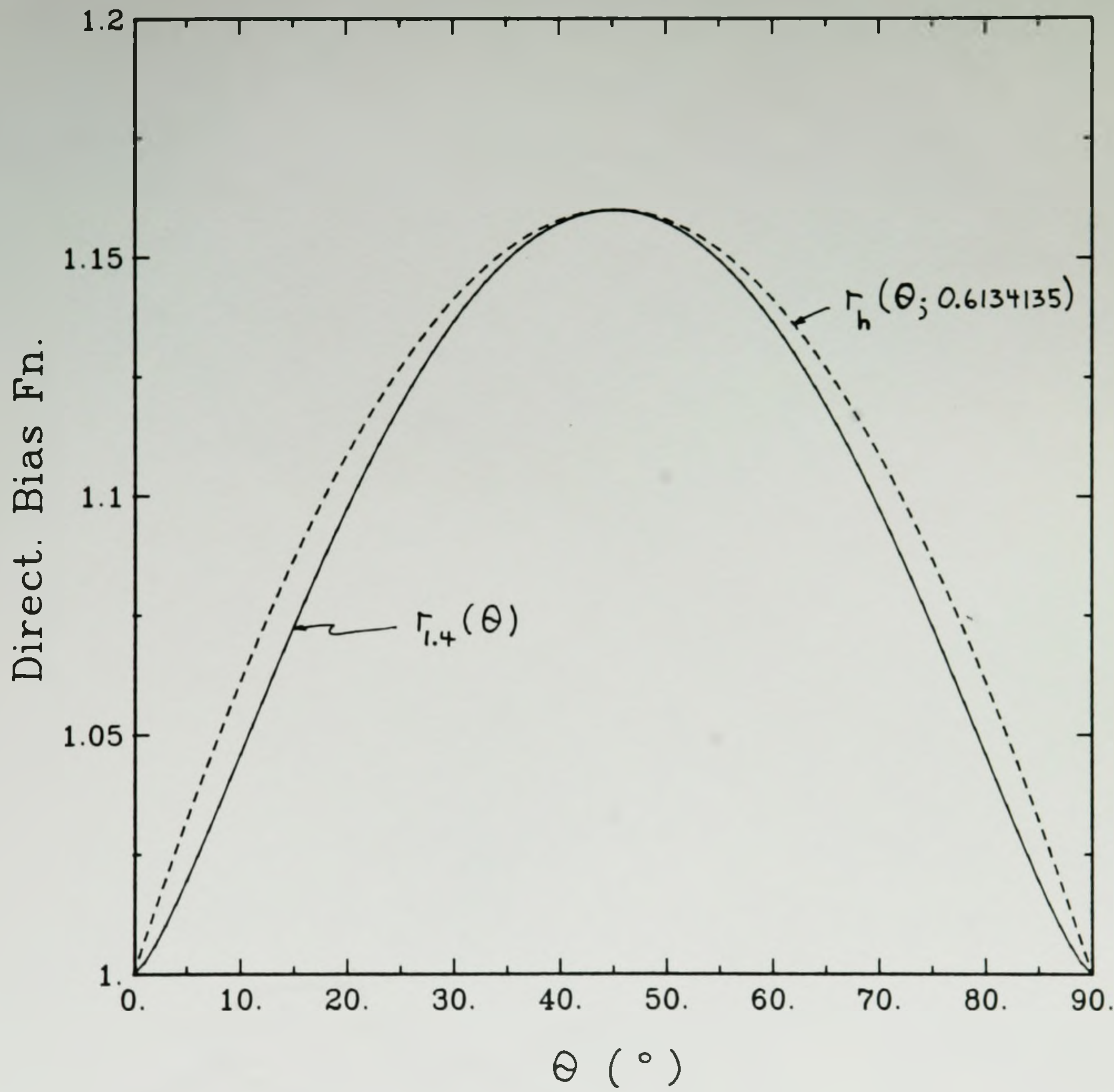


Figure 4.1.4 One Cycle of $r_p(\theta)$ and its Approximation Function for $p = 1.4$.

difficulty (i.e., parallel to a reference axis). Thus, the positive slope in equation (4.1.20) is more appropriate than the zero slope of r_p at $\theta = 0$. In this respect, the shape of r_h is better-suited for such networks than that of r_p.

Combining (4.1.17) and (4.1.19), we obtain the following bounds:

$$0 \le k_h(x; \alpha_0) - \ell_p(x) \le 0.015431\, \ell_2(x), \tag{4.1.22}$$

where x is any point in R^2, $1 \le p \le 2$, and α_0 is given by equation (4.1.14). We see then that k_h is an accurate approximation function of ℓ_p. Furthermore, recall that a constraint was imposed on the normalized difference Δ (see (4.1.13)), in order to obtain the form of α_0 in (4.1.14). If this restriction is removed, and instead α_0 is chosen to minimize the maximum absolute magnitude of Δ, the approximation will be improved. Thus, we conclude in summary the following important result:

> The weighted ℓ_p norm with $1 \le p \le 2$ and $x \in R^2$ can be replaced for all practical purposes by a weighted one-two norm.

4.2 A Linear Regression Model

We begin by hypothesizing the following model for actual travel distances:

$$A(q,r) = \beta_0 \ell_2(q-r) + \beta_1 \ell_1(q-r) + e(q-r), \tag{4.2.1}$$

where

q,r are any two points in the plane,

A(q,r) equals the travel distance between q and r,

$\beta_0, \beta_1 \ge 0$, with at least one of these parameters being positive,

ℓ_1, ℓ_2 are, as before, the rectangular and Euclidean norms, and

$e(q-r)$ is an independent error term which is assumed to be normally distributed with mean zero.

It is assumed here that the reference axes for measuring the coordinates of q and r have a known fixed orientation (γ_0). Also recall the physical requirement, $A(q,r) \geq \ell_2(q-r)$, $\forall$ q,r, which implies that $\beta_0 + \beta_1 \geq 1$, similarly as in (3.3.18).

Let us consider the error term more closely. The vector $(q-r)$ has a direction θ and a magnitude $\|q-r\| = \ell_2(q-r)$. One would logically expect that for a given θ, the variance of the error term should increase as $\|q-r\|$ becomes larger. In other words, the variation in travel distance is greater for pairs of points which are further apart. Thus, we assume that a normalized error defined as

$$\epsilon(\theta) = e(q-r)/\|q-r\|, \tag{4.2.2}$$

has a standard deviation given by $\sigma(\theta)$. Making the further simplifying assumption that

$$\sigma(\theta) = \sigma, \forall \theta, \tag{4.2.3}$$

where σ is a constant, it follows that ϵ is a normal random variable with zero mean and a constant variance, σ^2; i.e.,

$$\epsilon \sim N(0, \sigma^2). \tag{4.2.4}$$

Now dividing both sides of (4.2.1) by $\ell_2(q-r)$, we obtain

$$\alpha(\theta) = \beta_0 + \beta_1 r_1(\theta) + \epsilon, \tag{4.2.5}$$

where $\alpha(\theta) = A(q,r)/\ell_2(q-r)$ is the normalized travel distance (see Definition 3.3.1), and ϵ is an independent random variable with distribution given by (4.2.4). This is precisely the same model as in (3.3.4), for a transportation network with a predominant rectangular pattern (Definition 3.3.4) and $R \equiv r_1$; except now we specify in addition that ϵ is normally distributed with constant variance.

The formulation in (4.2.5) provides a simple linear regression model with one independent variable, $r_1(\theta)$. For any particular set of data, we can readily calculate the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ of β_0 and β_1, respectively. It is a well-known fact (e.g., Neter, Wasserman and Kutner, 1985, p. 39) that $\hat{\beta}_0$ and $\hat{\beta}_1$ are the best linear unbiased estimators of the model parameters. (Note, however, that the assumption of constant variance must hold. Otherwise, the weighted least squares estimators should be used after determining the functional form of $\sigma(\theta)$.) Furthermore, since the expected normalized travel distance is given by

$$\bar{\alpha}(\theta) = E[\alpha(\theta)] = \beta_0 + \beta_1 r_1(\theta), \tag{4.2.6}$$

it follows that $\hat{\beta}_0$ and $\hat{\beta}_1$ can be used as the coefficients of the weighted one-two norm, k_H (equation (4.1)), to approximate actual distances in the network. Then $k_H(x; \hat{\beta}_0, \hat{\beta}_1)$ estimates the mean or expected travel distance between any two points q and r, such that $q - r = x$.

We now propose a few applications of the model in (4.2.5), which provide new ways of analyzing the physical nature of the transportation network under consideration. These applications rely on standard methods of linear regression analysis. In the following statistical tests, H_0 and H_1 denote the null and alternate hypotheses respectively. The details pertaining to these tests are omitted here, since they can be found in any standard text on linear regression (e.g., see Chapter 3 of Neter, Wasserman and Kutner, 1985).

Test 1 {Directional Bias}

Here we consider the decision problem,

$$H_0 : \beta_1 = 0 \quad \text{versus} \quad H_1 : \beta_1 > 0. \tag{4.2.7}$$

When H_0 is rejected in favour of H_1, we conclude that the network has a statistically-significant rectangular bias. This implies that the underlying rectangular pattern of roads in the network contributes on average to the total travel distance between pairs of points.

Test 2 {Diagonal Roads}

Next consider the decision problem,

$$H_0 : \beta_0 = 0 \quad \text{versus} \; H_1 : \beta_0 > 0. \tag{4.2.8}$$

If H_0 is rejected in favour of H_1, we conclude that the Euclidean component of the travel distance is statistically significant. This implies in turn that the diagonal roads traversing the network contribute on average to the total travel distance between pairs of points.

Test 3 {Outliers}

Under the assumption of constant variance for the error term, a substantial deviation of an actual normalized distance from the estimated mean $(\hat{\beta}_0 + \hat{\beta}_1 r_1(\theta))$ may signify the presence of an outlier. Such outliers are detected from an analysis of the residuals. Since our model has only one independent variable, it is possible to plot the residuals and identify any outliers visually. A more rigorous method involves the use of standardized residuals (e.g., see Belsley, Kuh and Welsch, 1980). It is crucial that the outliers be identified, since they can have an excessive influence on the estimates $(\hat{\beta}_0, \hat{\beta}_1)$ of the model parameters, and hence on the distance function itself.

In addition, the outliers provide important information concerning the physical nature of the transportation network being modelled. If several of them are associated with the same destination, say a_r, then this implies that the distance function obtained for the population in general does not accurately estimate travel distances to a_r (and its environs) from the other points in the network. To remedy such a situation, we should custom-fit a separate distance function for a_r alone, using the subset of data associated with a_r. This results in a mixed-norm model, which will be discussed in further detail in Chapter 6. The importance of the mixed-norm model is that it allows a more accurate representation of the real system. In this respect, we are closing the gap between continuous and discrete location models.

To illustrate the use of our linear regression model, let us continue the case study of the road network in southern Ontario discussed at the end of Chapter 3. The 18 cities listed in

Table 3.3.1 provided 153 ($= {}^{18}C_2$) travel distances for the analysis, which was carried out with the Minitab computer package. Based on our previous observations on the nature of the road system, an orientation of 22.5° was chosen for the reference axes of the model.

The least squares estimators of the coefficients β_0, β_1, were found to be

$$\hat{\beta}_0 = 0.7577 \quad \text{and} \quad \hat{\beta}_1 = 0.3284 .$$

From the output of standardized residuals, Fort Erie and St. Catharines were identified as the source of several outliers. A scan of the map shows that these two cities are located south of Lake Ontario, and that the lake provides a large obstacle between them and the other destinations in the data set. Hence, we conclude that a separate distance function should be used for these two cities. The data points corresponding to the Sarnia-Windsor and Peterborough-Barrie links were also identified as significant outliers. This is due to the fact that Lake St. Clair and Lake Simcoe result in local barriers to travel between these two pairs.

Deleting Fort Erie and St. Catharines plus the two links mentioned above, and repeating the regression analysis with the remaining data, we obtained least squares estimators,

$$\hat{\beta}_0 = 0.7786 \quad \text{and} \quad \hat{\beta}_1 = 0.2881 .$$

Subsequent removal of outliers was observed to have little effect on these values. Finally, we note that the t-test values associated with $\hat{\beta}_0$ and $\hat{\beta}_1$ for the reduced data set, respectively 12.72 and 5.64, provide strong statistical evidence in support of the alternate hypothesis in Tests 1 and 2 above.

CHAPTER 5

SINGLE FACILITY LOCATION WITH THE ℓ_p NORM

The single facility unconstrained minisum location problem, commonly referred to as the Weber problem, was introduced briefly in Chapter 1; (see model (1.1)). We re-state it here for the more restrictive case where the distance function is given by a norm k on R^N. Thus, we obtain the following model:

$$\text{Minimize } W(x) = \sum_{i=1}^{n} w_i \, k(x - a_i), \tag{5.1}$$

where $a_i = (a_{i1}, ..., a_{iN})^T$ is the known position of the i^{th} destination or fixed point, $i = 1, ..., n$; n is the number of fixed points; $x = (x_1, ..., x_N)^T$ is the unknown position of the new facility; w_i is a positive weighting constant which converts distance travelled between the new facility and the i^{th} customer into a cost, for $i = 1, ..., n$; and $k(\cdot)$ is a norm used to measure the distance between any two points in R^N.

In this chapter, we begin with the minisum model in (5.1), and develop some general properties for the case where k is a round norm. These results provide some insights into the differences between models which use round norms and those which use block norms. We then consider the special case where k is the ℓ_p norm. The model in (5.1) becomes

$$\text{minimize } W(x) = \sum_{i=1}^{n} w_i \, \ell_p(x - a_i), \tag{5.2}$$

where $p \geq 1$, and w_i, a_i, $i = 1, ..., n$, and x are defined above. When the parameter p has a value in the open interval $(1, +\infty)$, $\ell_p(\cdot)$ is a differentiable round norm (Property 2.2.4). For $p = 1$ and $+\infty$, we have the rectangular and Chebychev norms, both of which are block norms. A Weiszfeld-type solution algorithm for the model in (5.2) is analyzed in great detail for global convergence properties and local convergence rates, and some interesting results are obtained.

5.1 Properties of the Minisum Problem

In this section, we derive properties pertaining to the optimal solution of model (5.1). Conditions which guarantee that this solution is unique have been obtained by Pelegrin, Michelot and Plastria (1985), for a more general minisum problem. We provide a different proof of these results which is geared to the specific model in (5.1). This also permits us to analyze the behaviour of the objective function in more detail and the properties of the optimal solution when it is not unique. We first consider the case where k is a round norm, then a block norm. Finally, we make use of the uniqueness results to deduce properties concerning the location of the optimal solution in relation to the fixed points. Extensions to the multifacility problem are also discussed.

Property 5.1.1

Consider the minisum problem given in (5.1), where the a_i, $i = 1, \ldots, n$, $(n > 1)$, are collinear points, and k is a round norm. Then the objective function W(x) is convex piecewise linear along the straight line joining the a_i, and strictly convex everywhere else.

Proof:

Since W(x) is the sum of convex functions, it is also convex. Let L_1 denote the straight line passing through the fixed points, and choose a point $x \in L_1$, such that $x \neq a_i$, $i = 1, \ldots, n$. Then

$$k(x - a_i) = c\,\ell_2(x - a_i), \qquad i = 1, \ldots, n \ , \tag{5.1.1}$$

where c is a positive constant. Let y denote a unit vector parallel to L_1. The directional derivative of W at x in the direction y is given by

$$W'(x;y) = c\left(\sum_{i\in J_1} w_i - \sum_{i\in J_2} w_i\right), \tag{5.1.2}$$

where $\quad J_1 = \{j \mid (x-a_j)\cdot y > 0\}$,

and $\quad J_2 = \{j \mid (x-a_j)\cdot y < 0\}$.

We see that $W'(x;y)$ is constant at all points in the open segment (a_{i_1}, a_{i_2}) of L_1, where a_{i_1} and a_{i_2} are adjacent fixed points on L_1. Furthermore, $W'(x;y)$ changes by a discrete amount between adjacent segments, since one of the indices gets transferred between J_1 and J_2. It follows then that W(x) is convex piecewise linear on L_1.

Now consider a point $x \notin L_1$, and draw any straight line L_2 through x. At least $(n-1)$ of the a_i's are not contained in L_2. Choose one of these; say, $a_r \notin L_2$. Since k is a round norm, it follows from Property 2.2.3 that $k(x-a_r)$ is strictly convex along L_2. Thus, W(x) is the sum of n convex terms, of which at least $(n-1)$ are strictly convex along L_2. It follows that W(x) is strictly convex everywhere except along the straight line through the fixed points, thereby ending the proof.

Corollary 5.1.1

For the collinear case considered above, suppose an optimal solution x^* exists such that $x^* \notin L_1$. Then x^* is the only optimal solution. If on the other hand, $x^* \in L_1$, there are two possibilities:

i) x^* is uniquely located at a fixed point; or

ii) all the points on the closed line segment $[a_{i_1}, a_{i_2}]$ are optimal, where a_{i_1} and a_{i_2} are adjacent fixed points on L_1.

Proof:

Follows immediately from Property 5.1.1.

Property 5.1.2

Consider the minisum problem given in (5.1), where k is a round norm, and the a_i, $i = 1, \ldots, n$, are non-collinear points this time. Then the objective function W(x) is a strictly convex function of x.

Proof:

Let L denote any straight line in R^N. Since the a_i are non-collinear, there must be at least one fixed point, say a_r, such that $a_r \notin L$. Since k is a round norm, it follows from Property 2.2.3 that $k(x - a_r)$ is strictly convex along L. Thus W(x) is the sum of n convex terms, at least one of which is strictly convex along L. We conclude then that W(x) is a strictly convex function of x.

Corollary 5.1.2

The optimal solution x* of the minisum model given in (5.1) is unique when k is a round norm and the fixed points a_i are non-collinear.

Proof:

Follows immediately from Property 5.1.2.

Corollary 5.1.2 gives sufficient conditions for the optimal solution of the classical minisum problem to be unique. It is also interesting to note that Properties 5.1.1 and 5.1.2 generalize the Result 3 of Francis and Cabot (1972), where k is the Euclidean norm on R^2, to the case where k is any round norm on R^N. Consider now the case where k is a block norm. The uniqueness of the optimal solution can no longer be guaranteed, as made evident by the following result, which applies irrespective of whether the a_i are collinear or not.

Property 5.1.3

Let k be a block norm in the minisum model (5.1). Then the objective function W(x) is convex piecewise linear in x along any straight line in R^N.

Proof:

Along any straight line in R^N, W(x) is the sum of n convex piecewise linear terms. (See the discussion pertaining to block norms in Section 2.2.) Hence W(x) is itself a convex piecewise linear function of x along the line.

Let us consider now the minisum model (5.1) where k is a block norm on R^2 ($N = 2$). Recalling Property 2.4.4, we draw through each of the fixed points a_i the straight lines along which $k(x - a_i)$ is non-differentiable. This is illustrated in Figure 5.1.1 when k is the weighted one-infinity norm. We see that these lines form in general polygons of various shapes and sizes. Let us define a 'small box' as any polygon bounded by these lines, which does not contain within it other such polygons. In other words, a small box is any cell formed by the straight lines we have drawn through the a_i.

In Theorem 6 of Thisse, Ward and Wendell (1984), it is shown that an optimal solution must occur at one of the intersection or fixed points located at the corners of the small boxes. We can prove this result quite readily using Property 5.1.3. The objective function W(x) is convex piecewise linear along any straight line L. Furthermore, the directional derivative $W'(x\,;y)$, where y is a unit vector parallel to L and $x \in L$, only changes, and by a discrete amount, when L crosses from one small box into an adjacent one. Suppose then that

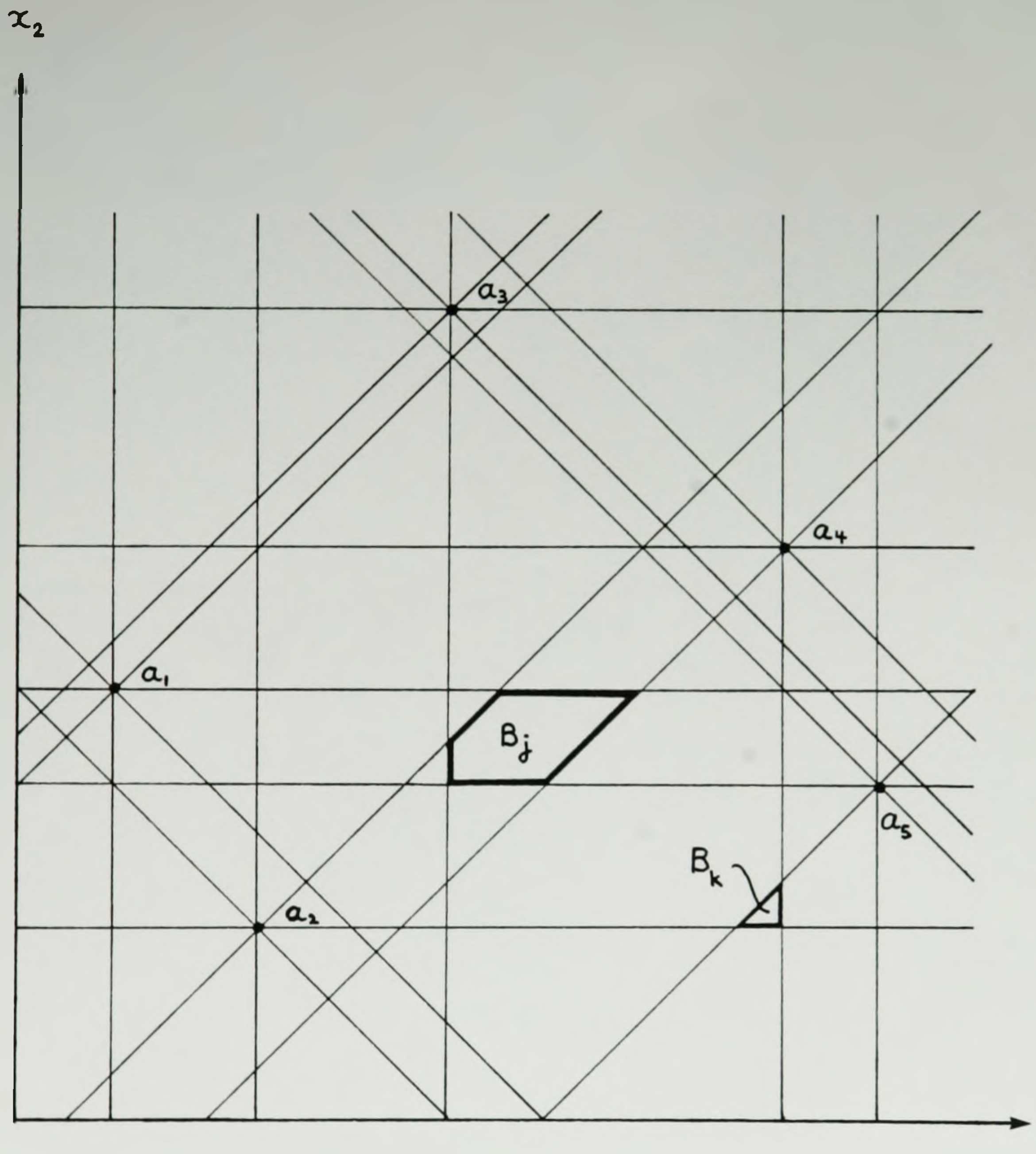


Figure 5.1.1 Non-Differentiable Points and Small Boxes for the Weighted One-Infinity Norm.

an optimal solution occurs at a point x_0 belonging to a small box labelled B_j. If x_0 is an interior point of B_j, draw the straight line L to pass through x_0 and a corner point x_c of B_j. Clearly $W'(x\,;y) = 0$ on the open line segment of L contained in B_j. (Otherwise x_0 cannot be optimal.) Thus the corner point x_c is also optimal. A similar line of reasoning holds if x_0 is on an edge of B_j, and L is chosen to coincide with this edge. Thus we conclude as in Thisse, Ward and Wendell (1984) that an optimal solution occurs at a corner point.

Based on the preceding discussion, we can also characterize the optimal solution of model (5.1), when k is a block norm on R^2. There are three possibilities:

i) the optimal solution occurs uniquely at an intersection or fixed point;

ii) the points along one edge of a small box are all optimal; or

iii) all the points belonging to a small box are optimal.

Since Property 5.1.3 holds in R^N, it is also interesting to note that the above results can be generalized to higher-dimensional spaces.

The following properties deal with the location of the optimal solution in relation to the convex hull of the fixed points, denoted by c.h. $\{a_1, \ldots, a_n\}$.

Property 5.1.4

Consider the minisum model (5.1), where k is a round norm on R^2. Any optimal solution must lie in the convex hull of the fixed points.

Proof:

Consider first the case where the a_i, $i = 1, \ldots, n$, are non-collinear points. From Corollary 5.1.2, we see that the optimal solution x^* is unique. For any norm, an optimal solution must exist within the convex hull of the fixed points (Corollary 4 of Wendell and Hurter, 1973). Hence, we conclude that $x^* \in$ c.h. $\{a_1, \ldots, a_n\}$. Now consider the case where the

a_i, i = 1, ..., n, are collinear points. Suppose an optimal solution x^* exists outside the convex hull of the fixed points. By Corollary 5.1.1, it follows that x^* is the unique optimal. But this contradicts Corollary 4 of Wendell and Hurter (1973), which states that an optimal solution can be found in c.h. $\{a_1, ..., a_n\}$. Hence, we conclude for the collinear case that any optimal solution $x^* \in$ c.h. $\{a_1, ..., a_n\}$.

Property 5.1.5

Consider the minisum model (5.1), where k is the Euclidean norm on R^N ($k \equiv \ell_2$). Any optimal solution must lie in the convex hull of the fixed points.

Proof:

First we note that k is a round norm (Property 2.2.4). The remainder of the proof is identical to that of Property 5.1.4, except that Corollary 3 of Wendell and Hurter (1973) replaces their Corollary 4.

Juel and Love (1983) show that all optimal solutions of the minisum model (5.2) in two-dimensional space (i.e., k is the ℓ_p norm on R^2), must occur in the convex hull of the fixed points when $p > 1$. Their proof relies on properties of the directional derivative of the ℓ_p norm. We see that Property 5.1.4 generalizes this result to the case where k is any round norm on R^2.

The hull properties discussed above for single facility location can be readily extended to the multifacility case. The multifacility problem was introduced briefly in Chapter 1; (see model (1.3)). We reformulate it here with the distance function being given by a norm k:

$$\text{minimize } WM(X) = \sum_{j=1}^{m} \sum_{i=1}^{n} w_{ij} k(x_j - a_i) + \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} v_{jk} k(x_j - x_k), \tag{5.3}$$

where m is the number of new facilities to be located; n is the number of existing destinations (or fixed points); $X = (x_1, ..., x_m)$ is an $Nm \times 1$ column vector with $x_j = (x_{j1}, ..., x_{jN})^T$ being the unknown position of new facility j, for $j = 1, ..., m$; $a_i = (a_{i1}, ..., a_{iN})^T$ is the known position of the i^{th} destination, $i = 1, ..., n$; $w_{ij} \geq 0$ is a weighting constant which converts distance between new facility j and destination i into a cost, for $i = 1, ..., n$, $j = 1, ..., m$; $v_{jk} \geq 0$ is a weighting constant which converts distance between new facilities j and k into a cost, for $j = 1, ..., m-1$, $k = j+1, ..., m$; and finally $k(\cdot)$ is an appropriate norm used to measure the distance between any two points in R^N.

As discussed in Francis and Cabot (1972, p. 338), the new facilities must be chained in order that the model (5.3) be well-formulated. This means that each new facility j must be linked to some existing facility i, either directly ($w_{ij} > 0$), or indirectly through a chain of new facilities $j, j_1, ..., j_p$, such that $v_{jj_1} > 0$, $v_{j_1j_2} > 0, ..., v_{j_{p-1}j_p} > 0$, $w_{ij_p} > 0$, (and $v_{kl} = v_{lk}$ when $k > l$). Otherwise at least two new facilities can be located coincident to each other anywhere in R^N, without affecting the solution. Henceforth, we assume that the multifacility problem is well-formulated; i.e., the new facilities are chained.

The next result extends Property 4 of Francis and Cabot (1972), where k is the ℓ_2-norm on R^2, to the general case where k is any round norm on R^N. To simplify the notation, let

$$f_j(x_j) = \sum_{i=1}^{n} w_{ij} k(x_j - a_i), \qquad j = 1, ..., m, \tag{5.1.3}$$

and

$$f_o(X) = \sum_{j=1}^{m-1} \sum_{r=j+1}^{m} v_{jr} k(x_j - x_r), \tag{5.1.4}$$

so that

$$WM(X) = \sum_{j=1}^{m} f_j(x_j) + f_o(X) . \tag{5.1.5}$$

Property 5.1.6

Consider the multifacility model (5.3), where k is a round norm on R^N. The objective function WM(X) is strictly convex if, and only if, for j = 1, ..., m, the set $S_j = \{a_i \mid w_{ij} > 0\}$ is nonempty, and the points in each set S_j are not collinear.

Proof:

The proof is a straightforward extension of the one given by Francis and Cabot (1972). The details are given here for completeness.

(i) (If). If S_j is nonempty and the points in S_j are not collinear, for j = 1, ..., m, it follows using Property 5.1.2 that $f_j(x_j)$ is strictly convex on R^N, j = 1, ..., m. Thus, $\Sigma^m_{j=1} f_j(x_j)$ is strictly convex on R^{Nm}. Let X_1 and X_2 be distinct points in R^{Nm}, with $X_1 = (x_1, ..., x_m)$ and $X_2 = (y_1, ..., y_m)$. Using the triangle inequality, we obtain for $\lambda \in (0, 1)$,

$$\begin{aligned} f_o(\lambda X_1 + (1-\lambda)X_2) &= \sum_{j<r}\sum v_{jr}\, k[\lambda x_j + (1-\lambda)y_j - \lambda x_r - (1-\lambda)y_r] \\ &= \sum_{j<r}\sum v_{jr}\, k[\lambda(x_j - x_r) + (1-\lambda)(y_j - y_r)] \\ &\leq \lambda \sum_{j<r}\sum v_{jr}\, k(x_j - x_r) + (1-\lambda)\sum_{j<r}\sum v_{jr}\, k(y_j - y_r) \\ &= \lambda f_o(X_1) + (1-\lambda) f_o(X_2) \; . \end{aligned} \tag{5.1.6}$$

Thus f_o is convex on R^{Nm}. We see that WM is the sum of a convex and a strictly convex function, both having domain R^{Nm}, and thus, WM is strictly convex on R^{Nm}.

(ii) (Only if). Consider first the case where at least one set S_j is empty, and without loss of generality, suppose S_m is empty. Let $v_m = v_{1m} + ... + v_{m-1,m}$. Then,

$$WM(0, ..., 0, x_m) = v_m\, k(x_m) + \sum_{j=1}^{n-1} f_j(0) \; . \tag{5.1.7}$$

Since $k(x_m)$ varies linearly on any half-line H in R^N beginning at the origin, it follows that WM will be linear on the line segment joining $(0, ..., 0, x_m^1)$ and $(0, ..., 0, x_m^2)$ in R^{Nm}, where x_m^1 and x_m^2 are distinct points on H. This contradicts the fact that WM is strictly convex, and so, each set S_j must be nonempty.

Now suppose that the points in at least one set S_j are collinear; without loss of generality, assume the points in S_m are collinear, and that $w_{1m} > 0$. Define

$$w'_{1m} = \sum_{j=1}^{m-1} v_{jm} + w_{1m} = v_m + w_{1m},$$

$$w'_{im} = w_{im}, \qquad i = 2, ..., n,$$

and

$$f(x_m) = \sum_{i=1}^{n} w'_{im} k(x_m - a_i) .$$

Then

$$WM(a_1, ..., a_1, x_m) = \sum_{j=1}^{m-1} f_j(a_1) + f(x_m) . \tag{5.1.8}$$

Using Property 5.1.1, it follows that f is convex piecewise linear on the line L in R^N containing S_m. Thus, distinct points x_m^1 and x_m^2 in L may be chosen such that f is linear on the line segment in R^N joining x_m^1 and x_m^2. It follows from (5.1.8) that f will be linear on the line segment joining $(a_1, ..., a_1, x_m^1)$ and $(a_1, ..., a_1, x_m^2)$ in R^{Nm}. This contradicts the fact that WM is strictly convex, and so, the points in each set S_j are not collinear.

Property 5.1.6 provides a sufficient condition for guaranteeing that WM(X) has a unique minimum. That is, if k is a round norm on R^N, and S_j is nonempty and the points contained in S_j are not collinear for $j = 1, ..., m$, then WM is minimized at a unique X^* in R^{Nm}. Of course this is not a necessary condition, since WM does not have to be strictly convex to have a unique optimal. It is also interesting to note that the sufficient condition for strict convexity given in the above property does not depend in any manner upon f_0.

The next result considers the case where k is a block norm. We see that Property 5.1.3 for the single facility objective function W(x) extends quite readily to WM(X).

Property 5.1.7

Consider the multifacility model (5.3), where k is a block norm on R^N. Then WM(X) is convex piecewise linear along any straight line in R^{Nm}.

Proof:

From Property 5.1.3, it follows that $f_j(x_j)$ is convex piecewise linear along any line in R^N, for $j = 1, ..., m$. Thus, $\Sigma^m_{j=1} f_j(x_j)$ is convex piecewise linear on any line in R^{Nm}. Returning to relation (5.1.6), we see that

$$f_o(\lambda X_1 + (1-\lambda)X_2) = \sum_{j<r}\sum v_{jr}\, k[\lambda(x_j - x_r) + (1-\lambda)(y_j - y_r)]$$

where $X_1 = (x_1, ..., x_m)$ and $X_2 = (y_1, ..., y_m)$ are two distinct points in R^{Nm}, and $0 \leq \lambda \leq 1$. As λ varies from 0 to 1, the argument of f_o describes the line segment joining X_1 and X_2, while the argument of k describes the line segment in R^N joining the points $(x_j - x_r)$ and $(y_j - y_r)$ for each term in the double summation. For a fixed increment on the line segment joining X_1 and X_2 in R^{Nm}, we obtain the same proportional increment on the line segment joining $(x_j - x_r)$ and $(y_j - y_r)$ in R^N, for $1 \leq j < r \leq m$. Using this fact and Property 5.1.3, it follows that f_o is the sum of convex piecewise linear functions on the line segment joining X_1 and X_2, and hence is also convex piecewise linear on this line segment. Thus, we conclude that WM is convex, piecewise linear along any straight line in R^{Nm}.

Some localization results have been obtained for the multifacility problem. Francis and Cabot (1972, Property 2) prove by induction that any optimal solution to the model (5.3), with k as the Euclidean norm on R^2, must have all the new facilities located in the convex hull of the fixed points. Juel and Love (1983) extend this result to the case where k is an ℓ_p norm on R^2, and $1 < p < +\infty$. Meanwhile, Hansen, Perreur and Thisse (1980) show that an optimal solution exists with all the new facilities in the convex hull of the fixed points when k is any norm on R^2. (However, optimal locations may also exist outside the hull.) The next result generalizes the one by Juel and Love (1983) for the ℓ_p norm on R^2 ($1<p<+\infty$) to any round norm on R^2.

Property 5.1.8

Consider the multifacility model (5.3), where k is a round norm on R^2, and let $X^* = (x_1^*, ..., x_m^*)$ denote an optimal solution. Then

$$\{x_1^*, ..., x_m^*\} \subset \text{c.h.}\{a_1, ..., a_n\} .$$

Proof:

From the convexity property of WM, it follows that at least one optimal solution exists. Assume that such a solution $(x_1^*, ..., x_m^*)$ has at least one new facility located outside c.h. $\{a_1, ..., a_n\}$. Then c.h. $\{a_1, ..., a_n, x_1^*, ..., x_m^*\} \supset$ c.h. $\{a_1, ..., a_n\}$. Clearly some of the new facility locations correspond to extreme points of the larger hull. Take one of these points, say $x \notin$ c.h. $\{a_1, ..., a_n\}$. If one new facility is located at x, then by Property 5.1.4, a better solution can be obtained by moving this facility to some location in the convex hull of the fixed points and the remaining new facilities. This contradicts the optimality assumption for $(x_1^*, ..., x_m^*)$. If two or more new facilities are located at x, they can be combined and treated as a single new facility to arrive at the same contradiction. Hence, we conclude that all the new facilities must be located in c.h. $\{a_1, ..., a_n\}$ for an optimal solution.

The next result generalizes the one by Francis and Cabot (1972) for the Euclidean norm on R^2 to higher-dimensional spaces.

Property 5.1.9

Consider the multifacility model (5.3), where k is the Euclidean norm on R^N ($k = \ell_2$), and let $X^* = (x_1{}^*, \ldots, x_m{}^*)$ denote an optimal solution. Then

$$\{x_1^*, \ldots, x_m^*\} \subset \text{c.h.}\{a_1, \ldots, a_n\} .$$

Proof:

The proof is identical to the previous one, except that Property 5.1.5 replaces Property 5.1.4.

5.2 The Weiszfeld Procedure Revisited

5.2.1 One-Point Iterative Methods

The iterative solution technique developed by Weiszfeld (1937) to solve the single facility minisum model with Euclidean distances has received considerable attention in the literature. As noted in Chapter 1, this algorithm was re-discovered several years later independently by Miehle (1958), Kuhn and Kuenne (1962), and Cooper (1963). The main advantages of the Weiszfeld procedure are its simplicity and ease of programming, and the fact that the iterations give progressively better solutions. The disadvantages of the Weiszfeld procedure include the fact that it will fail if one of the iterates happens to be a fixed point (Kuhn, 1973), and the local convergence rate is generally linear (Katz, 1974). In exceptional cases which occur only when the optimal location is at a fixed point, this rate may be quadratic or sublinear. By local convergence, we are referring to the behaviour of the iterates (x^q) when they are sufficiently close to the optimal solution (x^*).

It is worthwhile noting that the Weiszfeld procedure belongs to a broader class of solution techniques known as one-point iteration methods. (For a general discussion of one-point iteration methods, the reader is referred to Dahlquist and Bjorck, 1974, Chapter 6.)

This class of solution techniques can be described as follows. Given a general system of N nonlinear equations in N unknowns,

$$f_i(x_1, ..., x_N) = 0 , \qquad i = 1, ..., N, \tag{5.2.1}$$

we rewrite the system in the form,

$$x_i = \phi_i(x_1, ..., x_N) , \qquad i = 1, ..., N; \tag{5.2.2}$$

and then proceed to solve for the unknowns by the following iterative sequence,

$$x_i^{q+1} = \phi_i(x_1^q, ..., x_N^q) , \qquad i = 1, ..., N. \tag{5.2.3}$$

Here, $q = 0, 1, 2, ...$, is used to specify the iteration number, and $(x_1^o, ..., x_N^o)$ gives our initial estimate of the solution for the starting point of the iterations. The system of equations in (5.2.1) can normally be put in the form (5.2.2) in many different ways, not all of which will necessarily yield sequences that converge to the solution. The trick then is to find the functions ϕ_i with good convergence properties. One can rewrite (5.2.3) in vector notation as follows:

$$x^{q+1} = \phi(x^q) , \qquad q = 0, 1, 2, ..., \tag{5.2.4}$$

where $x = (x_1, ..., x_N)^T$, and $\phi(x) = (\phi_1(x), ..., \phi_N(x))^T$.

We see that the term 'one-point' derives from the fact that the iteration function ϕ uses only the current iterate (x^q) to determine the succeeding one (x^{q+1}). This is the simplest possible form which ϕ can take. More generally, ϕ is a function of m points, which are not necessarily successive iterates, so that

$$x^{q+1} = \phi(y^q, y^{q-1}, ..., y^{q-m+1}). \tag{5.2.5}$$

This is referred to as an m-point iteration method. The principal requirement of an iterative method such as the one given in (5.2.4) or (5.2.5) is that the sequence generated should converge to a solution (or root) of the system of equations given in (5.2.1), for any arbitrary starting point. Letting $x^* = (x_1^*, ..., x_N^*)^T$ denote such a solution, this means that

$$\lim_{q\to\infty} x^q = x^* , \tag{5.2.6}$$

Global conv.

for arbitrary $x^o \in R^N$. This is referred to as the global convergence property of the iterative method. (For a general discussion of global convergence, see Luenberger, 1973, Chapter 6.)

In the Weiszfeld procedure, the system of equations (5.2.1) is given by the first-order conditions for a stationary point of a differentiable function; namely,

$$\nabla W(x) = 0 \ , \tag{5.2.7}$$

where $\nabla = (\partial/\partial x_1, \ldots, \partial/\partial x_N)^T$ denotes the gradient vector, $W(x) = \Sigma^n_{i=1} w_i \ell_2(x - a_i)$ is the objective function of the minisum problem with Euclidean distances, $\ell_2(\cdot)$ is the Euclidean norm on R^N, and the remaining symbols are as defined in model (5.1). The iteration function ϕ used to solve (5.2.7) will be discussed later.

Kuhn (1973) shows that the Weiszfeld procedure converges globally to the solution x^* of (5.2.7) provided that an iterate x^q does not fall on one of the fixed points a_i. If this occurs, the iteration function ϕ becomes undefined due to division by zero. We shall investigate this problem in greater detail later. In the meantime, it suffices to note that an x^q will coincide with an a_i only for a denumerable number of starting points (Kuhn, 1973). Theoretically then, the probability of a vertex iterate occurring would be zero for a randomly chosen starting point and a computer with infinite accuracy. Hence in practice we should expect this problem to occur very rarely, and so, we should not be too concerned about it. If by chance a vertex iterate does occur, one can always re-start the iterations at an x^o slightly removed from the fixed point in question. Ostresh (1978) proceeds in this manner to resolve the problem of vertex iterates, by defining a new step when an iterate falls on a fixed point. With this modification, global convergence is guaranteed for any starting point.

In an earlier paper, Ostresh (1977) extends the Weiszfeld procedure to the multi-facility minisum problem with Euclidean distances, and shows that the descent property of the algorithm also holds in this case. By descent property, we mean that the iterative method gives lower values of the associated objective function from one iteration to the next, provided the new iterate does not equal its predecessor. For the multifacility case, the problem of vertex iterates becomes more complex, since the new facilities can coincide with each other anywhere in the location space (R^N). If two or more new facilities coincide or a new facility coincides with an existing one, the Weiszfeld procedure will fail in a similar manner as the single facility case when an iterate lands on a fixed point. The popular way of circumventing this problem is to use a smoothing function in place of the distance measure. Eyster, White and Wierwille (1973) introduce a hyperboloid approximation of the Euclidean norm. Alternatively, the hyperbolic approximation can be used (e.g., see Love, 1969, 1974, Wesolowsky and Love, 1972, and Love and Morris, 1975b). These smoothing functions have the computationally-appealing property that they are infinitely differentiable everywhere. Although this approach eliminates the problem of vertex iterates, it has the disadvantage that the solution obtained can be a considerable distance away from the solution of the original model.

The local convergence properties of an iterative method such as the one in (5.2.4) are a measure of the ultimate speed of convergence, when the iterates are within a sufficiently small neighbourhood of the solution x^*. However, they have to get there first. Therefore, one should establish beforehand that the algorithm is globally convergent. Generally, the local convergence properties are of interest when we wish to determine the relative advantage of one algorithm to another. Katz (1974) shows that the Weiszfeld procedure has a local convergence rate which is always linear if x^* is not a fixed point. His results apply to the single facility minisum problem with Euclidean distances in R^N. In mathematical terms, this means

$$\lim_{q \to \infty} \frac{\|x^{q+1} - x^*\|}{\|x^q - x^*\|} = c < 1 \ , \tag{5.2.8}$$

where c is a positive constant giving the asymptotic convergence rate, and $\|\cdot\|$ denotes the Euclidean norm in R^N. For example, if $c = \frac{1}{2}$, the distance between x^q and x^* is approximately halved after every iteration for sufficiently large values of q.

We can use more sophisticated one-point iterative methods to improve the convergence rate when x^q is close to x^*. However, these methods require several more computations at each step. It can be shown that a one-point iterative method of order r always requires the computation of all quantities related to the functions $f_i(x_1, \ldots, x_N)$, $i = 1, \ldots, N$, up to and including the $(r-1)^{th}$ order partial derivatives, at each iteration (Dahlquist and Björck, 1974). This becomes very cumbersome for values of $r > 2$, unless the higher-order derivatives are easily computed. An r^{th}-order method implies that

$$\lim_{q \to \infty} \frac{\|x^{q+1} - x^*\|}{\|x^q - x^*\|^r} = c \quad (\neq 0) . \tag{5.2.9}$$

An example of a one-point iterative method of higher order would be the Newton-Raphson method generalized to N dimensions, for which $r = 2$. Katz (1974) uses Steffensen's method, which is also of second order, on several sample problems to obtain much better convergence rates near the optimal solution than by the Weiszfeld procedure. However, it should be noted that Steffensen's iterations are not known to be globally convergent. This leads to the idea of using hybrid algorithms to solve the minisum location problem, in which we begin with the Weiszfeld procedure, and then switch to a more sophisticated technique to accelerate the convergence when the iterates are close to the optimal location.

Morris and Verdini (1979) extend the Weiszfeld procedure to the minisum model with ℓ_p distances. We shall proceed next to prove global convergence of their generalized algorithm. Currently such a proof exists only for the special case of Euclidean distances (Kuhn, 1973), or when a hyperbolic approximation of the ℓ_p norm is used (Morris, 1978, 1981). The smoothing function has the advantage that the singularities in the iteration function are eliminated. Subsequently in section (5.3), we extend the local convergence results of Katz (1974) for the Euclidean norm to the ℓ_p norm. Aside from academic interest, our aims serve a practical purpose as well. As noted previously, the use of the hyperbolic approximation of the ℓ_p norm, or other smoothing functions for that matter, may result in solutions which are considerably removed from the optimal solution to the original problem. Hence, a global convergence proof pertaining to the original (un-approximated) model would be useful. Finally, a knowledge of the local convergence properties will enable us to design hybrid algorithms and choose acceleration methods more effectively.

5.2.2 Global Convergence Proof

Let us consider now the single facility minisum model (5.2), where distances are given by the ℓ_p norm on R^N. Substituting the functional form of the ℓ_p norm, we can rewrite the objective function as follows:

$$W(x) = \sum_{i=1}^{n} w_i \left(\sum_{j=1}^{N} |x_j - a_{ij}|^p \right)^{1/p} , \qquad p \geq 1 . \tag{5.2.10}$$

When $p = 1$, so that rectangular distances are being used, the minimization problem can be separated in the N dimensions and solved in an efficient manner (e.g., see Love, Morris and Wesolowsky, 1988). If the Chebychev norm is used ($p = +\infty$), the problem can be reformulated to rectangular distances by a rotation of the axes, and solved in a similar fashion. Hence, in applying the Weiszfeld iterative procedure to minimize W(x), we restrict attention in practice to problems where the parameter p has a value in the following range,

$$1 < p < +\infty . \tag{5.2.11}$$

Recall that the ℓ_p function is a differentiable round norm for values of p in this range (Property 2.2.4).

If a solution occurs at an x* which is not a fixed point, then W(x) is differentiable at x*, and the first-order necessary conditions for a stationary point require that

$$\frac{\partial W(x^*)}{\partial x_t} = 0 , \qquad t = 1, \ldots, N . \tag{5.2.12}$$

Note that the system of equations in (5.2.12) is analogous to the system given in (5.2.1), where one-point iterative methods were being discussed. Since W is a convex function of x, the equations (5.2.12) are also sufficient conditions for x* to be an optimal location. Evaluating the partial derivatives of W at x*, we rewrite (5.2.12) as follows:

$$\sum_{i=1}^{n} w_i \operatorname{sign}(x_t^* - a_{it}) \frac{|x_t^* - a_{it}|^{p-1}}{[\ell_p(x^* - a_i)]^{p-1}} = 0 , \qquad t = 1, \ldots, N , \tag{5.2.13}$$

where again $x^* = (x_1^*, \ldots, x_N^*)^T$. The procedure now is to re-arrange (5.2.13) in an analogous form as (5.2.2). One of the many ways this can be accomplished is to note first that

$$(x_t - a_{it}) = \operatorname{sign}(x_t - a_{it}) | x_t - a_{it} | ; \tag{5.2.14}$$

so that (5.2.13) becomes

$$\sum_{i=1}^{n} w_i (x_t^* - a_{it}) \frac{|x_t^* - a_{it}|^{p-2}}{[\ell_p(x^* - a_i)]^{p-1}} = 0 , \qquad t = 1, \ldots, N , \tag{5.2.15}$$

and we readily obtain

$$x_t^* = \frac{\sum_{i=1}^{n} w_i | x_t^* - a_{it}|^{p-2} a_{it} \,/\, [\ell_p(x^* - a_i)]^{p-1}}{\sum_{i=1}^{n} w_i | x_t^* - a_{it}|^{p-2} \,/\, [\ell_p(x^* - a_i)]^{p-1}} , \qquad t = 1, \ldots, N . \tag{5.2.16}$$

As in (5.2.3), the above set of equations suggests the following iterative scheme:

$$x_t^{q+1} = \frac{\sum_{i=1}^{n} w_i | x_t^q - a_{it}|^{p-2} a_{it} \,/\, [\ell_p(x^q - a_i)]^{p-1}}{\sum_{i=1}^{n} w_i | x_t^q - a_{it}|^{p-2} \,/\, [\ell_p(x^q - a_i)]^{p-1}} , \qquad t = 1, \ldots, N , \tag{5.2.17}$$

where the superscript q = 0, 1, 2, ..., denotes the iteration number.

Letting

$$\gamma_{it}(x) = \frac{w_i |x_t - a_{it}|^{p-2}}{[\ell_p(x - a_i)]^{p-1}}, \qquad i = 1, \ldots, n, \quad t = 1, \ldots, N, \tag{5.2.18}$$

we can rewrite (5.2.17) in the compact form,

$$x_t^{q+1} = \frac{\sum_{i=1}^{n} \gamma_{it}(x^q)\, a_{it}}{\sum_{i=1}^{n} \gamma_{it}(x^q)}, \qquad t = 1, \ldots, N. \tag{5.2.19}$$

Note that the iteration function vector $\phi(x)$ in (5.2.4) is now defined as

$$\phi(x) = (\phi_1(x), \ldots, \phi_N(x))^T, \tag{5.2.20a}$$

where

$$\phi_t(x) = \frac{\sum_{i=1}^{n} \gamma_{it}(x)\, a_{it}}{\sum_{i=1}^{n} \gamma_{it}(x)}, \qquad t = 1, \ldots, N. \tag{5.2.20b}$$

For the special case where $p = 2$, the iterative scheme in (5.2.17) simplifies to

$$x_t^{q+1} = \frac{\sum_{i=1}^{n} w_i\, a_{it} \,/\, \ell_2(x^q - a_i)}{\sum_{i=1}^{n} w_i \,/\, \ell_2(x^q - a_i)}, \qquad t = 1, \ldots, N. \tag{5.2.21}$$

This is the well-known Weiszfeld procedure for the minisum model with Euclidean distances, for which several references are noted in subsection 5.2.1. The iterative method in (5.2.17) extends the Weiszfeld procedure to the general case where distances are given by an ℓ_p norm. A similar formulation for the hyperbolic approximation of the ℓ_p norm is given by Morris (1978), Morris and Verdini (1979), and Love, Morris and Wesolowsky (1988).

Returning to the equations (5.2.19) and letting

$$\lambda_{it}(x) = \frac{\gamma_{it}(x)}{\sum_{j=1}^{n} \gamma_{jt}(x)}, \qquad i = 1, \ldots, n, \quad t = 1, \ldots, N, \tag{5.2.22}$$

we see that

$$x_t^{q+1} = \sum_{i=1}^{n} \lambda_{it}(x^q)\, a_{it}\,, \qquad t = 1, \ldots, N. \tag{5.2.23}$$

Since

$$\lambda_{it}(x) \geq 0\,, \qquad \forall\ i, t\,, \tag{5.2.24}$$

and

$$\sum_{i=1}^{n} \lambda_{it}(x) = 1\,, \qquad \forall\ t\,, \tag{5.2.25}$$

we conclude the important result that x_t^{q+1} is a convex combination of a_{it}, $i = 1, \ldots, n$, in each iteration step, for $t = 1, \ldots, N$. Furthermore, for the special case, $p = 2$,

$$\gamma_{i1}(x) = \gamma_{i2}(x) = \ldots = \gamma_{iN}(x) = \gamma_i(x)\,, \qquad i = 1, \ldots, n\,, \tag{5.2.26}$$

so that

$$\lambda_{it}(x) = \frac{\gamma_i(x)}{\sum_{j=1}^{n} \gamma_j(x)} = \lambda_i(x)\,, \qquad \forall\ t, i\,, \tag{5.2.27}$$

and

$$x_t^{q+1} = \sum_{i=1}^{n} \lambda_i(x^q)\, a_{it}\,, \qquad t = 1, \ldots, N \quad (p = 2). \tag{5.2.28}$$

Hence, x_1^{q+1} is the same convex combination of the a_{i1}'s as x_2^{q+1} is of the a_{i2}'s, ..., and x_N^{q+1} is of the a_{iN}'s. All the iterates x^{q+1}, $q = 0, 1, 2, \ldots$, must therefore fall within the convex hull of the fixed points a_i. Kuhn (1973) uses this result to prove that the optimal solution for single facility minisum problems with Euclidean distances lies in the convex hull of the a_i.

However, if $p \neq 2$, then equations (5.2.26), and thus (5.2.27) and (5.2.28), do not hold in general. In other words, x_1^{q+1} will not be the same convex combination of the a_{i1}'s as x_2^{q+1} is of the a_{i2}'s, and so on. Hence, we can only conclude for $p \neq 2$ that all the iterates x^{q+1}, $q = 0, 1, 2, \ldots$, will fall in a bounded hypercube containing the fixed points; that is,

$$\min_i \{a_{it}\} \leq x_t^{q+1} \leq \max_i \{a_{it}\}\,, \tag{5.2.29}$$

for $t=1, \ldots, N$ and $q=0,1,2,\ldots$.

Consider now the case where $a_{1t} = a_{2t} = \ldots = a_{nt}$, for some $t \in \{1, \ldots, N\}$. Then, from (5.2.23) and (5.2.25), it follows that

$$x_t^{q+1} = a_{1t}, \qquad q = 0, 1, 2, \ldots. \tag{5.2.30}$$

Hence, each iterate x^{q+1} lies in the hyperplane, $x_t - a_{1t} = 0$, and the problem reduces to one in $(N-1)$ dimensions. We assume without restriction that all problems are reduced in this manner to the minimum number of required dimensions.

As noted by Morris (1981), no complete proof of global convergence for the ℓ_p norm has previously been published except for the case where $p = 2$; i.e., the iterative method in (5.2.21). Our purpose then is to extend the global convergence proof to the iterative method in (5.2.17) for ℓ_p norms in general. Much of the analysis to follow is based on and motivated by the work of Kuhn (1973) for Euclidean distances. However, our convergence proof requires some new approaches due to complications we shall see later arising from the following fundamental result.

Property 5.2.1

If $p < 2$, the iteration function $\phi_t(x)$ is undefined along the hyperplanes,

$$x_t - a_{it} = 0, \qquad i = 1, \ldots, n, \tag{5.2.31}$$

for $t = 1, \ldots, N$. Whereas if $p \geq 2$, the iteration function $\phi_t(x)$ is undefined only at the fixed points a_i, $i = 1, \ldots, n$, for $t = 1, \ldots, N$.

Proof:

If $p < 2$, then $|x_t - a_{it}|^{p-2} \rightarrow +\infty$, as x approaches any point on the hyperplane, $x_t - a_{it} = 0$, in R^N. Note that this hyperplane also includes the fixed point a_i. If $p \geq 2$, division by zero within the functional form of $\phi_t(x)$ will only occur if $\ell_p(x-a_i) = 0$; i.e., $x = a_i$, for some i. Aside from the above singular points where $\phi_t(x)$ cannot be computed directly, we see that the functional form of $\phi_t(x)$ is well-defined and continuous. Hence, the property follows.

Property 5.2.1 reveals a basic difference between a Weiszfeld procedure with Euclidean distances and one with ℓ_p distances, where $p < 2$. In the former case, the iteration functions $\phi_t(x)$, $t = 1, \ldots, N$, are singular only at the fixed locations a_i, $i = 1, \ldots, n$. However, in the latter case, we have to contend with singularity on the hyperplanes, $x_t - a_{it} = 0$, $i = 1, \ldots, n$, $t = 1, \ldots, N$. It would be advantageous in proving global convergence for each iteration function $\phi_t(x)$ to be continuous. Hence, we study the behaviour of $\phi_t(x)$ in the vicinity of its singular points to see if it can be made continuous at these points. The following three results deal with this question.

Property 5.2.2

Let p have a value in the range, $1 \le p < 2$. Then the iteration function $\phi_t(x)$, $t \in \{1, \ldots, N\}$, is continuous if, and only if, we set $\phi_t(x) = a_{it}$ at all points x on the hyperplane $x_t - a_{it} = 0$, for $i = 1, \ldots, n$.

Proof:

This follows immediately from the fact that

$$\lim_{x_t \to a_{rt}} \phi_t(x) = a_{rt}, \qquad \forall\ r, t, \tag{5.2.32}$$

and $\phi_t(x)$ is continuous everywhere else.

Property 5.2.3

Let $p = 2$. Then the iteration function $\phi_t(x)$, $t \in \{1, \ldots, N\}$, is continuous if, and only if, we set $\phi_t(a_i) = a_{it}$, for $i = 1, \ldots, n$.

Proof:

This follows immediately from the fact that

$$\lim_{x \to a_r} \phi_t(x) = a_{rt} , \qquad \forall\ r, t , \tag{5.2.33}$$

and $\phi_t(x)$ is continuous everywhere else.

Property 5.2.3 was previously recognized by Kuhn (1973). Using vector notation, we see that $\phi(a_i) = a_i$, $i = 1, \ldots, n$, in order that the mapping $\phi: x \to \phi(x)$ be continuous for $p = 2$. As seen from Property 5.2.2, this result also holds for $1 \leq p < 2$, but there are additional requirements on nN hyperplanes in R^N as well. For the case where $p > 2$, the following rather surprising result is obtained.

Property 5.2.4

Let p have a value in the range, $2 < p < +\infty$. Then the iteration function $\phi_t(x)$, $t \in \{1, \ldots, N\}$, cannot in general be made continuous at its singular points a_i, $i = 1, \ldots, n$.

Proof:

Let H denote the hyperplane, $x_t - a_{rt} = 0$, $r \in \{1, \ldots, n\}$, and consider the following limits:

$$\lim_{\substack{x \to a_r \\ x \in H}} \phi_t(x) = \frac{\sum_{i \neq r} w_i \,|a_{rt} - a_{it}|^{p-2} a_{it} \,/\, [\ell_p(a_r - a_i)]^{p-1}}{\sum_{i \neq r} w_i \,|a_{rt} - a_{it}|^{p-2} \,/\, [\ell_p(a_r - a_i)]^{p-1}} \;; \tag{5.2.34}$$

$$\lim_{\substack{x \to a_r \\ x \notin H}} \phi_t(x) = a_{rt} . \tag{5.2.35}$$

The first limit is easily obtained from (5.2.17). The second one follows from the observation that

$$\lim_{\substack{x \to a_r \\ x \notin H}} \left\{ \frac{|x_t - a_{rt}|^{p-2}}{[\ell_p(x-a_r)]^{p-1}} \right\} = \lim_{\substack{x \to a_r \\ x \notin H}} \left\{ \frac{|x_t - a_{rt}|^{p-2}}{[\tau |x_t - a_{rt}|]^{p-1}} \right\}$$

$$= \lim_{\substack{x \to a_r \\ x \notin H}} \left\{ \frac{1}{\tau^{p-1} |x_t - a_{rt}|} \right\}$$

$$= +\infty , \tag{5.2.36}$$

where $\ell_p(x-a_r) = \tau\,|x_t - a_{rt}|$, and $\tau > 0$ depends on the direction of approach to a_r. Since the two limits are not in general equal, the property is proven.

In summary, we define the following iteration procedure:

i) For $1 \le p < 2$,

$$x_t^{q+1} = \begin{cases} \phi_t(x^q) & \text{if } x_t^q - a_{it} \neq 0, \quad i = 1, \ldots, n, \\ a_{it} & \text{if } x_t^q - a_{it} = 0, \quad i \in \{1, \ldots, n\}, \end{cases}$$

$$t = 1, \ldots, N. \tag{5.2.37}$$

ii) For $p \ge 2$,

$$x^{q+1} = \begin{cases} \phi(x^q) & \text{if } x^q \neq a_i, \quad i = 1, \ldots, n, \\ a_i & \text{if } x^q = a_i, \quad i \in \{1, \ldots, n\}. \end{cases} \tag{5.2.38}$$

From Property 5.2.2, we see that the mapping given in (5.2.37) is continuous. From Properties 5.2.3 and 5.2.4, it follows that the mapping in (5.2.38) is continuous for $p = 2$, and discontinuous in general at the fixed points a_i for $p > 2$. Denote the procedure given in (5.2.37) or (5.2.38) symbolically as

$$T: x \rightarrow T(x) \quad (x \in R^N). \tag{5.2.39}$$

Clearly, T is just our iteration function vector ϕ specified in (5.2.20a) and (5.2.20b), with the singularities of its components, the ϕ_t, accounted for. The remaining properties deal with the iteration procedure T as defined in (5.2.37) or (5.2.38).

Property 5.2.5

The map, $T: x \rightarrow T(x)$, lies in a compact set.

Proof:

From equations (5.2.23), (5.2.24) and (5.2.25), it is seen that x_t^{q+1} is a convex combination of the a_{it}, $i = 1, \ldots, n$, if x^q is not a singular point of ϕ_t. Furthermore, if x^q is a singular point of ϕ_t, then $x_t^{q+1} = a_{rt}$, $r \in \{1, \ldots, n\}$, from (5.2.37) and (5.2.38). We conclude that all the iterates except possibly the starting point will fall in a bounded hypercube, such that (5.2.29) is satisfied.

Before proceeding to the next property, we introduce the following notation:

$$H_{it} = \{x \mid x_t - a_{it} = 0\}, \qquad i = 1, \ldots, n, \quad t = 1, \ldots, N; \tag{5.2.40}$$

$$H_t = \bigcup_{i=1}^{n} H_{it}, \qquad t = 1, \ldots, N; \tag{5.2.41}$$

and

$$S = \begin{cases} \bigcup_{t=1}^{N} H_t, & \text{if } 1 \le p < 2, \\ \{a_1, \ldots, a_n\}, & \text{if } p \ge 2. \end{cases} \tag{5.2.42}$$

In other words, S is the set of points, $x \in R^N$, where ϕ is singular. Let D denote the set of x where the objective function W is not differentiable. If $p > 1$, then $D = \{a_1, \ldots, a_n\}$. If $p = 1$, then $D = \cup^N_{t=1} H_t$. It is interesting to note that $D = S$ if $p = 1$ or $p \ge 2$, but $D \subset S$ if $1 < p < 2$.

Property 5.2.6

Let x^* denote an optimal solution of the model (5.2.10). If $x^q = x^*$ then $x^{q+1} = x^*$ as well as all subsequent iterations. If $x^q \notin S$ and $x^{q+1} = x^q$, then $x^q = x^*$.

Proof:

If $x^* \notin S$, then $\nabla W(x^*) = 0$ implies that $x^* = \phi(x^*) = (\phi_1(x^*), \ldots, \phi_N(x^*))^T$. Therefore, if $x^q = x^*$, then $x^{q+1} = T(x^q) = \phi(x^*) = x^*$, as well as all subsequent iterations. Consider now the case where $x^* \in S$. If $p \geq 2$, then $x^* = a_r$, for some $r \in \{1, \ldots, n\}$. Hence, if $x^q = x^*$, then by (5.2.38), $x^{q+1} = a_r = x^*$. If $1 \leq p < 2$, and $V_* = \{t \mid \phi_t(x^*)$ is non-singular$\}$, we obtain the following: the complement of V_*, V'_*, is non-empty; $x_r^* = a_{i_r r}\ \forall\, r \in V'_*$; and $x_r^* = \phi_r(x^*)\ \forall\, r \in V_*$. Hence, if $x^q = x^*$, then by (5.2.37), $x_r^{q+1} = a_{i_r r} = x_r^*\ \forall\, r \in V'_*$, and $x_r^{q+1} = \phi_r(x^*) = x_r^*\ \forall\, r \in V_*$; so that $x^{q+1} = x^*$. For the last statement of the property, we have $x^{q+1} = T(x^q) = \phi(x^q)$, since $x^q \notin S$. Hence, $x^q = \phi(x^q)$, which implies that $\nabla W(x^q) = 0$. We conclude that $x^q = x^*$.

The next result shows that each iteration moves in a descent direction of the objective function, W.

Property 5.2.7

If $p = 2$, then x^{q+1} lies on a vector from x^q pointing in the direction of steepest descent of W at x^q. Otherwise, if $p \neq 2$, this vector points in a descent direction which is generally not the steepest descent. In both cases, we assume $x^{q+1} \neq x^q$.

Proof:

Equations (5.2.19) and (5.2.20b) can be rewritten in the form,

$$x_t^{q+1} = \phi_t(x^q) = x_t^q - \frac{\sum_{i=1}^{n} \gamma_{it}(x^q) \cdot (x_t^q - a_{it})}{\sum_{i=1}^{n} \gamma_{it}(x^q)}, \qquad t = 1, \ldots, N.$$

But

$$\gamma_{it}(x)\cdot(x_t - a_{it}) = \frac{w_i \operatorname{sign}(x_t - a_{it})|x_t - a_{it}|^{p-1}}{[\ell_p(x - a_i)]^{p-1}}$$

$$= w_i \nabla_t \ell_p(x - a_i), \qquad \forall\ i, t,$$

where

$$\nabla_t := \frac{\partial}{\partial x_t}.$$

Letting

$$s_t(x) = \sum_{i=1}^{n} \gamma_{it}(x), \qquad t = 1, \ldots, N, \tag{5.2.43}$$

we see that

$$x_t^{q+1} = x_t^q - \frac{1}{s_t(x^q)} \nabla_t W(x^q), \qquad t = 1, \ldots, N, \tag{5.2.44}$$

provided, of course, x^q is not a singular point of ϕ_t. Since $x^{q+1} \neq x^q$, it follows that x^q is not at a fixed location; i.e., $x^q \neq a_i$, $i = 1, \ldots, n$. Thus, x^q can be a singular point of ϕ_t only if $1 \leq p < 2$ and $x_t^q = a_{it}$ for some $i \in \{1, \ldots, n\}$. In this case, $s_t(x^q) = +\infty$, and $x_t^{q+1} = a_{it}$, as well as all subsequent iterations. However, a non-empty set $J \subseteq \{1, \ldots, N\}$ must exist such that $s_r(x^q)$ has a finite value for all $r \in J$ (i.e., x^q is not a singular point of ϕ_r), since $x^{q+1} \neq x^q$. We see that

$$s_t(x^q) > 0, \qquad t = 1, \ldots, N; \tag{5.2.45}$$

also, $s_t(x^q)$ is finite valued for all $t \in \{1, \ldots, N\}$ if $p \geq 2$, and finite valued for at least some t if $1 \leq p < 2$. Furthermore, by (5.2.26), it follows that

$$s_1(x) = s_2(x) = \ldots = s_N(x) = s(x), \qquad p = 2; \tag{5.2.46}$$

but this does not generally hold if $p \neq 2$. We conclude from (5.2.44) and (5.2.45) that x^{q+1} lies on a vector from x^q pointing in a descent direction of W at x^q. For $p = 2$, this is the steepest descent direction, since by (5.2.46),

$$x^{q+1} - x^q = -\frac{1}{s(x^q)} \nabla W(x^q) \qquad (p = 2). \tag{5.2.47}$$

For $p \neq 2$, the above equation does not hold in general, so that the descent is not in the steepest direction.

The preceding result shows that the iterations move in downward directions along the surface of the objective function W. We can rewrite (5.2.44) in the form,

$$x^{q+1} = x^q - [M(x^q)]^{-1} \nabla W(x^q) , \tag{5.2.48}$$

in a similar manner as Morris and Verdini (1979) for the hyperbolic approximation of the ℓ_p norm, to show that the iterative scheme is actually a modified gradient descent method with pre-determined step size. The modification matrix $[M(x^q)]^{-1}$ is a diagonal matrix with (non-negative) diagonal elements given by $1/s_t(x^q)$, $t = 1, \ldots, N$. As noted by Kuhn (1973), a problem affecting global convergence may arise if the iterates 'overshoot'; that is, the step-size may be too large, causing W to increase between iterations. The following important result shows that overshooting cannot occur for a certain range of p.

Property 5.2.8 *{Descent Property}*

If $1 \leq p \leq 2$ and $x^{q+1} \neq x^q$, then $W(x^{q+1}) < W(x^q)$.

Proof:

For a given x^q, let $\gamma_{it}{}^q := \gamma_{it}(x^q)$, $i = 1, \ldots, n$, $t = 1, \ldots, N$. Then, from (5.2.19) we have

$$x_t^{q+1} = \sum_{i=1}^{n} \gamma_{it}^q a_{it} / \sum_{i=1}^{n} \gamma_{it}^q ,$$

provided $x^q \notin H_t$ if $1 \leq p < 2$ and $x^q \notin S$ if $p \geq 2$ (see relations (5.2.40), (5.2.41), and (5.2.42)). Let $V_q = \{t \mid \phi_t(x^q) \text{ is non-singular}\}$. If $p \geq 2$, then since $x^{q+1} \neq x^q$, we must have $x^q \neq a_i$, $i = 1, \ldots, n$; i.e., $x^q \notin S$. Clearly, $V_q = \{1, \ldots, N\}$ if $p \geq 2$. On the other hand, if $1 \leq p < 2$, then since $x^{q+1} \neq x^q$, we must have $V_q \subseteq \{1, \ldots, N\}$ and $V_q \neq 0$.

For $t \in V_q$, define

$$g_t(x_t) = \sum_{i=1}^{n} \gamma_{it}^{q}(x_t - a_{it})^2 . \tag{5.2.49}$$

Then $g_t(x_t)$ is a strictly convex function of x_t, and has a unique minimum at x_t^{q+1}. Thus, for all $t \in V_q$,

$$g_t(x_t^{q+1}) \le g_t(x_t^q)$$

$$= \sum_{i=1}^{n} w_i |x_t^q - a_{it}|^p [\ell_p(x^q - a_i)]^{1-p} . \tag{5.2.50}$$

At least one of the inequalities given by (5.2.50) must be satisfied in a strict sense, since $x^{q+1} \neq x^q$. Now for $s \in V'_q$, where V'_q is the complement of V_q, we have $x_s^{q+1} = x_s^q$. (Note that $V'_q = 0$ if $p \geq 2$.) Thus,

$$h_s(x_s^{q+1}) = h_s(x_s^q), \quad \forall\ s \in V_q' , \tag{5.2.51}$$

where

$$h_s(x_s) = \sum_{i=1}^{n} w_i |x_s - a_{is}|^p [\ell_p(x^q - a_i)]^{1-p} . \tag{5.2.52}$$

Combining (5.2.50) and (5.2.51), gives

$$\sum_{t \in V_q} g_t(x_t^{q+1}) + \sum_{s \in V_q'} h_s(x_s^{q+1}) < \sum_{t \in V_q} g_t(x_t^q) + \sum_{s \in V_q'} h_s(x_s^q)$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{N} w_i |x_k^q - a_{ik}|^p [\ell_p(x^q - a_i)]^{1-p}$$

$$= \sum_{i=1}^{n} w_i\ \ell_p(x^q - a_i) = W(x^q) . \tag{5.2.53}$$

Now consider the left-hand side of the above relation, and $p \neq 2$:

$$\sum_{t \in V_q} g_t(x_t^{q+1}) + \sum_{s \in V_q'} h_s(x_s^{q+1})$$

$$= \sum_{i=1}^{n} \sum_{t \in V_q} w_i | x_t^q - a_{it}|^{p-2} [\ell_p(x^q - a_i)]^{1-p} \cdot \quad (x_t^{q+1} - a_{it})^2$$

$$+ \sum_{i=1}^{n} \sum_{s \in V_q'} w_i | x_s^{q+1} - a_{is}|^{p} [\ell_p(x^q - a_i)]^{1-p} \qquad (V_q' = 0 \text{ if } p \geq 2)$$

$$= \sum_{\substack{i=1 \\ |x_k^q - a_{ik}| \neq 0}}^{n} \sum_{k=1}^{N} w_i | x_k^q - a_{ik}|^{p-2} [\ell_p(x^q - a_i)]^{1-p} \cdot \quad (x_k^{q+1} - a_{ik})^2 \qquad (x_s^{q+1} = x_s^q, \forall s \in V_q')$$

$$= \sum_{\substack{i=1 \\ |x_k^q - a_{ik}| \neq 0}}^{n} \sum_{k=1}^{N} w_i [\ell_p(x^q - a_i)]^{1-p} \cdot \quad [| x_k^q - a_{ik} |^p]^{\frac{p-2}{p}} \; [|x_k^{q+1} - a_{ik} |^p]^{\frac{2}{p}}$$

$$\geq \sum_{i=1}^{n} \sum_{k=1}^{N} w_i [\ell_p(x^q - a_i)]^{1-p} \left\{ \left(\frac{p-2}{p} \right) |x_k^q - a_{ik}|^p + \frac{2}{p} | x_k^{q+1} - a_{ik}|^p \right\}$$

for $p < 2$, (Beckenbach and Bellman, 1965, Chapter 1, 14.(7))

$$= \sum_{i=1}^{n} w_i \left(1 - \frac{2}{p} \right) \ell_p(x^q - a_i) + \frac{2}{p} \sum_{i=1}^{n} w_i [\ell_p(x^q - a_i)]^{1-p} [\ell_p(x^{q+1} - a_i)]^p$$

$$\geq \sum_{i=1}^{n} w_i \left(1 - \frac{2}{p} \right) \ell_p(x^q - a_i) + \frac{2}{p} \sum_{i=1}^{n} w_i \{(1-p)\, \ell_p(x^q - a_i) + p\ell_p(x^{q+1} - a_i)\}$$

for $p \geq 1$, (Beckenbach and Bellman, 1965, Chapter 1, 14.(7))

$$= - \sum_{i=1}^{n} w_i \, \ell_p(x^q - a_i) + 2 \sum_{i=1}^{n} w_i \, \ell_p(x^{q+1} - a_i)$$

$$= -W(x^q) + 2W(x^{q+1}) \, . \qquad (5.2.54a)$$

For $p = 2$, we have

$$\sum_{t \in V_q} g_t(x_t^{q+1}) + \sum_{s \in V_q'} h_s(x_s^{q+1})$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{N} w_i \, [\ell_2(x^q - a_i)]^{-1} \, (x_k^{q+1} - a_{ik})^2 \qquad (V_q' = 0)$$

$$= \sum_{i=1}^{n} w_i \, [\ell_2(x^q - a_i)]^{-1} [\ell_2(x^{q+1} - a_i)]^2$$

$$\geq \sum_{i=1}^{n} w_i \, \{-\ell_2(x^q - a_i) + 2\ell_2(x^{q+1} - a_i)\}$$

(Beckenbach and Bellman, 1965, Chapter 1, 14. (7))

$$= -W(x^q) + 2W(x^{q+1}) \, . \qquad (5.2.54b)$$

Comparing (5.2.54a) and (5.2.54b) with (5.2.53) gives

$$-W(x^q) + 2W(x^{q+1}) < W(x^q), \qquad 1 \leq p \leq 2 \, .$$

Hence

$$W(x^{q+1}) < W(x^q), \qquad 1 \leq p \leq 2 \, , \qquad (5.2.55)$$

thus proving the descent property of the algorithm for values of p in this range.

For values of p greater than 2, the descent property does not hold at all times. This is due to the fact that the first inequality in (5.2.54a) is reversed for $p > 2$. Consider the following simple example with four fixed points in two dimensions: $a_1 = (0, 0)$, $a_2 = (0, 10)$, $a_3 = (10, 10)$, $a_4 = (10, 0)$, $w_1 = w_2 = 2$, and $w_3 = w_4 = 1$. Let the starting point of the iterations be $x^o = (0, 5)$. It is readily seen from (5.2.17) that the iterates will oscillate between (10, 5) and (0, 5) in all subsequent iterations, for any value of $p > 2$. This unstable behaviour is illustrated in Figure 5.2.1, as well as the first few iterations when $x^o = (0, 9)^T$ and $p = 3$. If x^o is a point on the vertical line through a_1 and a_2 or a_3 and a_4, then the consecutive iterates will oscillate between these two lines to infinity or until one of them lands on an a_i. It is clear

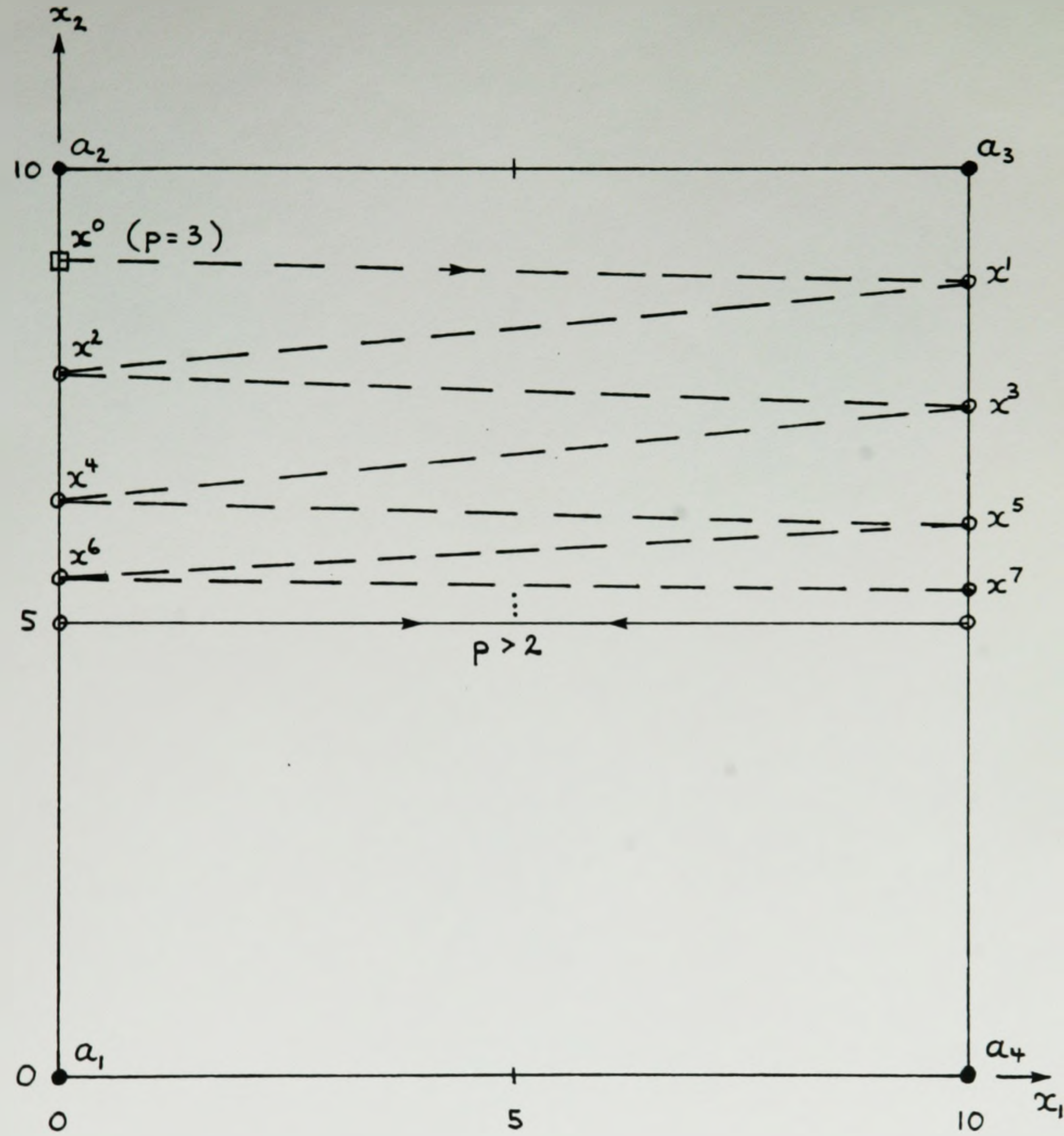


Figure 5.2.1 Unstable Trajectories for $p > 2$.

then from this simple example and Property 5.2.7, that although the iterates move in descent directions, they can overshoot if $p > 2$ (i.e., $W(x^{q+1}) > W(x^q)$).

A similar situation applies if $p < 1$ ($p \neq 0$). Recall in this case that the ℓ_p distance function is no longer a norm. It can readily be shown that Property 5.2.7 still applies, so that the iterates move in descent directions of the objective function W. However, we cannot guarantee that overshooting will not occur, since the second inequality in (5.2.54a) is reversed when $p < 1$.

If $x^q = a_i$ for some q, then $x^{q+1} = T(a_i) = a_i$, so that $W(x^{q+1}) = W(x^q)$. As a result, the fixed points belong to the solution set Γ referred to in the general global convergence proof for descent algorithms of Luenberger (1973, Chapter 6), which implies in turn that a subsequence of the iterates may converge to any a_i. This other potential difficulty of the Weiszfeld procedure was recognized by Kuhn (1973) for Euclidean distances ($p = 2$). For the case $1 \leq p < 2$, the problem is complicated further by the possibility that $T(Q) = Q$ at non-optimal points $Q \in S$, where S is the union of hyperplanes, $x_t - a_{it} = 0$, as defined in (5.2.42). For example, consider the intersection point B in N dimensions defined as follows:

$$B = (a_{i_1 1}, a_{i_2 2}, \ldots, a_{i_N N}) , \qquad (5.2.56)$$

where $i_j \in \{1, \ldots, n\}$, $j = 1, \ldots, N$,

and $B \neq a_i$, $i = 1, \ldots, n$.

If $1 \leq p < 2$, then $T(B) = B$, and thus all the intersection points belong to the solution set Γ. As another example, consider a hypothetical point Q on the hyperplane $x_1 - a_{11} = 0$, but on no other hyperplanes of S, satisfying

$$\frac{\partial W(Q)}{\partial x_2} = \frac{\partial W(Q)}{\partial x_3} = \ldots = \frac{\partial W(Q)}{\partial x_N} = 0 .$$

If $1 \leq p < 2$, then $T(Q) = Q$, and $Q \in \Gamma$. Using similar hypothetical examples, it follows that we must investigate the possibility of convergence to any non-optimal point belonging to S.

Before addressing the potential problem of convergence to a non-optimal point in S, we prove two useful lemmas.

Lemma 5.2.1

Let $p \in [1, 2]$, and consider any sequence x^q, $q = 0, 1, 2, \ldots$, generated by the map $T: x \rightarrow T(x)$. Then $\{x^q\}$ and all the subsequences thereof converge to one and the same point.

Proof:

First consider the case where $x^{q+1} = T(x^q) = x^q$ for some q. Clearly the sequence repeats from that point, thus verifying the lemma for this case. Hence, we only need to consider sequences where $x^{q+1} \neq x^q$ for all q.

With the possible exception of x^o, the sequence x^q lies in a compact set defined by a bounded hypercube (Property 5.2.5). Hence, by the Bolzano-Weierstrasz Theorem, there exists at least one point P and a subsequence x^{r_ℓ} such that $\lim_{\ell \to \infty} x^{r_\ell} = P$. To prove the lemma, we must show that there is at most one such P. This is done by contradiction.

Suppose there are M subsequences ($M \geq 2$) of x^q which converge to distinct points $P_1, \ldots, P_M$. Consider the first subsequence x^{r_ℓ}. Then

$$\lim_{\ell \to \infty} x^{r_\ell} = P_1 \ . \tag{5.2.57}$$

By the monotonicity of W on the entire sequence x^q (Property 5.2.8), we also have

$$W(x^o) > W(x^1) > \ldots > W(P_1) \ . \tag{5.2.58}$$

Now construct a δ-neighbourhood around P_1, isolating it from the other P_i. Then it is clear that we can choose our subsequence $x^{r_\ell} \rightarrow P_1$ such that $T(x^{r_\ell}) \notin \Delta_1$ for all ℓ, where Δ_1 denotes the δ-neighbourhood. Invoking the Bolzano-Weierstrasz Theorem once again, we conclude that the subsequence $T(x^{r_\ell})$ must converge to at least one point, say $P'_1 \notin \Delta_1$; i.e.,

$$\lim_{\ell \to \infty} T(x^{r_\ell}) = P_1' \neq P_1 . \tag{5.2.59}$$

But

$$\lim_{\ell \to \infty} T(x^{r_\ell}) = T(P_1) , \tag{5.2.60}$$

by (5.2.57) and the fact that T is a continuous mapping. Since $P_1 \neq T(P_1)$, then by Property 5.2.8,

$$W(T(P_1)) < W(P_1) . \tag{5.2.61}$$

But this clearly results in a contradiction of the monotonicity of W on the entire sequence x^q, (relation (5.2.58)). Hence we conclude that there cannot be M subsequences ($M \geq 2$) which converge to distinct points, $P_1, \ldots, P_M$. Thus, the sequence x^q converges to a unique point.

Lemma 5.2.1 reveals an important property concerning the nature of any sequence x^q generated by the Weiszfeld procedure, and is a stronger result than previously recognized. The uniqueness of the convergence point is based principally on the strict monotonicity of W on x^q. The lemma applies even when more than one optimal location x^* exists, which can occur when $p=1$ or when the a_i are collinear. We cannot extend the lemma a priori to values of $p > 2$, since the descent property can be violated and T is not continuous at the fixed points in this case.

The second lemma provides a sufficient condition for non-convergence of the sequence x^q to a specified point Q, which will be useful in our investigation of the singular points on the hyperplanes $x_t - a_{it} = 0$ when $1 \leq p < 2$.

Lemma 5.2.2

Let $Q = (Q_1, \ldots, Q_N)^T$ be any point in R^N. Consider a sequence x^q, $q = 0, 1, 2, \ldots$, generated by the map T, such that $x_t^q \neq Q_t$ for all q and some $t \in \{1, \ldots, N\}$. If

$$\lim_{\substack{x \to Q \\ x_t - Q_t \neq 0}} \left\{ \frac{|\phi_t(x) - Q_t|}{|x_t - Q_t|} \right\} > 1 , \tag{5.2.62}$$

then the given sequence does not converge to Q.

Proof:

Suppose that the sequence x^q converges to Q. Then for any δ-neighbourhood of Q, where $\delta > 0$ can be made arbitrarily small, an M can be found such that

$$x^q \in \Delta_Q , \qquad \forall \, q > M , \tag{5.2.63}$$

where Δ_Q denotes the δ-neighbourhood. But for sufficiently small δ, we have by (5.2.62),

$$\frac{|x_t^{q+1} - Q_t|}{|x_t^q - Q_t|} > 1 , \qquad \forall \, x^q \in \Delta_Q . \tag{5.2.64}$$

Hence, an s exists such that if $x^q \in \Delta_Q$ then $x^{q+s} \notin \Delta_Q$, which contradicts (5.2.63). We conclude therefore that the sequence x^q does not converge to Q.

Informally, the above result says that if an iterate happens to land inside Δ_Q, then the subsequent iterates will eventually be kicked out by T. In order to apply the preceding lemma at the singular points of the iteration functions ϕ_t, we use (5.2.17) to obtain

$$x_t^{q+1} - a_{rt} = \frac{\sum_{i=1}^{n} w_i |x_t^q - a_{it}|^{p-2} a_{it} / [\ell_p(x^q - a_i)]^{p-1}}{\sum_{i=1}^{n} w_i |x_t^q - a_{it}|^{p-2} / [\ell_p(x^q - a_i)]^{p-1}} - a_{rt}$$

$$= \frac{\sum_{i \neq r} w_i |x_t^q - a_{it}|^{p-2} (a_{it} - a_{rt}) / [\ell_p(x^q - a_i)]^{p-1}}{\sum_{i=1}^{n} w_i |x_t^q - a_{it}|^{p-2} / [\ell_p(x^q - a_i)]^{p-1}} ;$$

$$\therefore \frac{|x_t^{q+1} - a_{rt}|}{|x_t^q - a_{rt}|} = \frac{\left| \sum_{i \neq r} w_i |x_t^q - a_{it}|^{p-2} (a_{rt} - a_{it}) / [\ell_p(x^q - a_i)]^{p-1} \right|}{\left| w_r \dfrac{|x_t^q - a_{rt}|^{p-1}}{[\ell_p(x^q - a_r)]^{p-1}} + |x_t^q - a_{rt}| \sum_{i \neq r} w_i \dfrac{|x_t^q - a_{it}|^{p-2}}{[\ell_p(x^q - a_i)]^{p-1}} \right|},$$

$$r \in \{1, ..., n\}, \quad t = 1, ..., N . \qquad (5.2.65)$$

We are primarily interested in sequences where none of the iterates coincide with singular points of the ϕ_t. As will be seen later, the probability of an iterate landing exactly on a singular point is very low, for a randomly-chosen x^o, and theoretically it is zero if the sequence is calculated with infinite accuracy. The following definition distinguishes between the two fundamental types of sequences.

Definition 5.2.1

A sequence x^q, $q = 0, 1, 2, ...$, is termed regular if $x^q \notin S$ for all q, where S is the set of singular points defined in (5.2.42). Otherwise, x^q, $q = 0, 1, 2, ...$, is a non-regular sequence.

It is important to note that if $x^q \in S$ for some q, then $x^{q+k} \in S$, $k = 1, 2, ...$, as seen by (5.2.37) or (5.2.38). The next result shows that a regular sequence will never converge to a non-optimal point in S.

Property 5.2.9

Consider any regular sequence x^q of the map T, with $1 \leq p \leq 2$. Let $Q \in S$ be a non-optimal location; i.e., $Q \neq x^*$. Then $\{x^q\}$ does not converge to Q.

Proof:

We consider three cases as follows:

i) $1 < p < 2, Q \neq a_i, i = 1, ..., n.$

Since Q is not a fixed point, the first derivatives $\partial W(Q)/\partial x_k$ are defined for all k. Furthermore, since $Q \neq x^*$, at least one of these derivatives is non-zero, say $\partial W(Q)/\partial x_t$. Thus,

$$\frac{\partial W(Q)}{\partial x_t} = \sum_{i=1}^{n} w_i \operatorname{sign}(Q_t - a_{it}) \left\{ \frac{|Q_t - a_{it}|^{p-1}}{[\ell_p(Q - a_i)]^{p-1}} \right\} \neq 0 , \tag{5.2.66}$$

where $Q = (Q_1, ..., Q_N)^T$ as before.

Suppose $\{x^q\}$ converges to Q. Then

$$\lim_{q \to \infty} x^q = \lim_{q \to \infty} T(x^q) = Q , \tag{5.2.67}$$

which implies by the continuity of T that

$$T(Q) = Q . \tag{5.2.68}$$

If $Q_t \neq a_{it} \; \forall \, i$, then clearly Q is not a singular point of ϕ_t, and we obtain from (5.2.66),

$$\phi_t(Q) \neq Q_t . \tag{5.2.69}$$

Hence $T(Q) \neq Q$, contradicting the supposition that $\{x^q\}$ converges to Q. We conclude that $Q_t = a_{rt}$ for some $r \in \{1, ..., n\}$, if convergence of $\{x^q\}$ to Q is to take place. Then (5.2.66) becomes

$$\frac{\partial W(Q)}{\partial x_t} = \sum_{i \neq r} w_i \operatorname{sign}(a_{rt} - a_{it}) \left\{ \frac{|a_{rt} - a_{it}|^{p-1}}{[\ell_p(Q - a_i)]^{p-1}} \right\} \neq 0 . \tag{5.2.70}$$

Using (5.2.65), it is readily seen that

$$\lim_{\substack{x \to Q \\ x_t - a_{rt} \neq 0}} \frac{|\phi_t(x) - a_{rt}|}{|x_t - a_{rt}|} = +\infty , \tag{5.2.71}$$

so that convergence of $\{x^q\}$ to Q cannot occur due to Lemma 5.2.2.

ii) $p = 1, Q \neq a_i, i = 1, \ldots, n.$

Again, suppose that $\{x^q\}$ converges to Q. Using a similar reasoning as in case (i), we conclude that for $t \in \{1, \ldots, N\}$, if $Q_t \neq a_{it}\ \forall\, i$, then $\partial W(Q)/\partial x_t = 0$. Let $J_Q = \{s \mid Q_s - a_{r_s s} = 0\}$, where $r_s \in \{1, \ldots, n\}$. Without loss in generality, assume that for each $s \in J_Q$, there is a unique r_s such that $Q_s - a_{r_s s} = 0$. This is always the case if the fixed points do not share common coordinates. (The proof can be easily modified otherwise.) Since $Q \neq x^*$, J_Q is a non-empty set.

The directional derivative of W evaluated at Q in the direction of the unit vector $y = (y_1, \ldots, y_N)^T$ is calculated as follows:

$$W'(Q; y) = \lim_{\delta \to 0^+} \left\{ \frac{W(Q+\delta y) - W(Q)}{\delta} \right\}$$

$$= \lim_{\delta \to 0^+} \left\{ \frac{\sum_{i=1}^{n} w_i \sum_{j=1}^{N} (|Q_j + \delta y_j - a_{ij}| - |Q_j - a_{ij}|)}{\delta} \right\}$$

$$= \lim_{\delta \to 0^+} \left\{ \frac{\sum_{s \in J_Q} \sum_{i=1}^{n} w_i (|a_{r_s s} + \delta y_s - a_{is}| - |a_{r_s s} - a_{is}|)}{\delta} \right.$$

$$\left. + \frac{\sum_{t \in J_Q'} \sum_{i=1}^{n} w_i (|Q_t + \delta y_t - a_{it}| - |Q_t - a_{it}|)}{\delta} \right\}$$

$$= \sum_{s \in J_Q} \left\{ w_{r_s} |y_s| + \sum_{i \neq r_s} w_i \operatorname{sign}(a_{r_s s} - a_{is}) y_s \right\}$$

$$+ \sum_{t \in J_Q'} \sum_{i=1}^{n} w_i \operatorname{sign}(Q_t - a_{it}) y_t \,. \qquad (5.2.72)$$

where

$$J_Q' = \{1, \ldots, N\} - J_Q$$

is the complement of J_Q. But

$$\frac{\partial W(Q)}{\partial x_t} = \sum_{i=1}^{n} w_i \operatorname{sign}(Q_t - a_{it}) = 0 , \qquad \forall\ t \in J_Q' . \tag{5.2.73}$$

Hence, equation (5.2.72) becomes

$$W'(Q\,;y) = \sum_{s \in J_Q} \left\{ w_{r_s} |y_s| + \sum_{i \neq r_s} w_i \operatorname{sign}(a_{r_s s} - a_{is}) y_s \right\} . \tag{5.2.74}$$

Since Q is a non-optimal location, we must have by Property 2.4.1,

$$\min_y W'(Q\,;y) < 0 . \tag{5.2.75}$$

Hence

$$\min_y W'(Q\,;y) = \min_y \left[\sum_{s \in J_Q} \left\{ w_{r_s} |y_s| - \left| \sum_{i \neq r_s} w_i \operatorname{sign}(a_{r_s s} - a_{is}) \right| \cdot |y_s| \right\} \right] < 0 , \tag{5.2.76}$$

which implies that

$$\frac{\left| \sum_{i \neq r_s} w_i \operatorname{sign}(a_{r_s s} - a_{is}) \right|}{w_{r_s}} > 1 \tag{5.2.77}$$

for at least one $s \in J_Q$, say $s = \sigma$. Using (5.2.65) with $p = 1$, $t = \sigma$ and $r = r_\sigma$, it is readily seen that

$$\lim_{\substack{x \to Q \\ x_\sigma - a_{r_\sigma \sigma} \neq 0}} \frac{|\phi_\sigma(x) - a_{r_\sigma \sigma}|}{|x_\sigma - a_{r_\sigma \sigma}|} = \frac{\left| \sum_{i \neq r_\sigma} w_i \operatorname{sign}(a_{r_\sigma \sigma} - a_{i\sigma}) \right|}{w_{r_\sigma}}$$

$$> 1 ; \tag{5.2.78}$$

so that convergence of $\{x^q\}$ to Q cannot occur due to Lemma 5.2.2.

iii) $1 \leq p \leq 2$, $Q = a_r$ for some $r \in \{1, \ldots, n\}$.

Let

$$W_r(x) = W(x) - w_r \ell_p(x-a_r) = \sum_{i \neq r} w_i \ell_p(x-a_i) . \tag{5.2.79}$$

Then the directional derivative of W at $Q = a_r$ in the direction y is given by

$$W'(a_r ; y) = \nabla W_r(a_r) \cdot y + w_r \ell_p(y) . \tag{5.2.80}$$

(Note that if $p = 1$, we are assuming without loss in generality as in case (ii) that the fixed points do not share common coordinates. Otherwise, $\nabla W_r(a_r)$ may not be defined, and the above equation would have to be modified slightly.) Let

$$f(y) = \frac{-\nabla W_r(a_r) \cdot y}{w_r \ell_p(y)} \tag{5.2.81}$$

Then the convergence of $\{x^q\}$ to a_r must be along a unique asymptotic direction V such that

$$f(V) = \max_y f(y) \tag{5.2.82}$$

(see Property 5.3.8 in the next subsection).

As shown in Juel and Love (1981), a necessary and sufficient condition for a_r to be an optimal solution is that $f(V) \leq 1$. Since a_r is not optimal in our case, it follows that

$$f(V) = \frac{-\nabla W_r(a_r) \cdot V}{w_r \ell_p(V)} > 1 .$$

$$\therefore \quad \nabla W_r(a_r) \cdot V + w_r \ell_p(V) < 0 . \tag{5.2.83}$$

Comparing (5.2.80) and (5.2.83), we see that $W'(a_r ; V) < 0$. Hence convergence of $\{x^q\}$ to a_r cannot occur, since this would contradict Property 5.2.8 (descent property).

Since cases (i), (ii) and (iii) exhaust all possibilities, the proof is complete.

At last we are ready to prove global convergence of the Weiszfeld procedure. However, this shall be under the proviso that p has a value in the range [1, 2], and $x^q \notin S$ for all q; i.e., the sequence is regular.

Theorem 5.2.1 *(Global Convergence)*

Let x^q, $q = 0, 1, 2, \ldots$, be a regular sequence generated by the map T for a value of p in the closed interval [1, 2]. Then $\{x^q\}$ converges to an optimal solution of the single facility location problem; i.e.,

$$\lim_{q \to \infty} x^q = x^* \; . \tag{5.2.84}$$

Proof:

By Lemma 5.2.1, it follows that $\{x^q\}$ converges to a unique point, $P \in R^N$, so that

$$\lim_{q \to \infty} x^q = P \; . \tag{5.2.85}$$

To prove the theorem, we must show that $P = x^*$.

If $x^{q+1} = x^q$ for some q, then the sequence repeats from that point and $P = x^q$. Since $x^q \notin S$, $P = x^*$ by Property 5.2.6. Otherwise by Property 5.2.8,

$$W(x^0) > W(x^1) > \ldots > W(x^q) > \ldots > W(x^*) \; .$$

Hence

$$\lim_{q \to \infty} [W(x^q) - W(T(x^q))] = 0 \; . \tag{5.2.86}$$

Since the continuity of T and (5.2.85) imply

$$\lim_{q \to \infty} T(x^q) = T(P) \; , \tag{5.2.87}$$

it follows that

$$W(P) - W(T(P)) = 0 \; . \tag{5.2.88}$$

Therefore, by Property 5.2.8, we must have $P = T(P)$. If $P \notin S$, then $P = x^*$ by Property 5.2.6. If $P \in S$, then P cannot be a non-optimal location by Property 5.2.9. Again $P = x^*$, and the theorem is proved.

Consider now a non-regular sequence x^q, $q = 0, 1, 2, \ldots$. If $p = 2$ (or $p > 2$), this implies that x^q coincides with a fixed point for some iteration number q, as well as all subsequent iterations, -- not an interesting situation. On the other hand, if $1 \leq p < 2$, then $\{x^q\}$ is restricted from some iteration number onwards to motion in a subspace of R^N, defined by the intersection of one or more hyperplanes of the form $x_t - a_{it} = 0$. The following corollary gives an analogous result as Theorem 5.2.1, for a non-regular sequence when $1 \leq p < 2$. First we introduce the following notation. Let

$$J = \{r \mid x_r^s - a_{i_r r} = 0\},$$

and

$$H = \bigcap_{r \in J} H_{i_r r},$$

where s is some iteration number such that $J \neq 0$, $i_r \in \{1, ..., n\}$ depends on r, and $H_{i_r r}$ is defined in (5.2.40).

Corollary 5.2.1

Consider a non-regular sequence x^q, $q = 0, 1, 2, ...,$ of the map T with $p \in [1, 2)$. Then if all subsequent iterations after x^s do not fall on any hyperplanes H_{it} not already included in H, the sequence converges to an optimal solution for the subspace H.

Proof:

By Lemma 5.2.1, it follows that $\{x^q\}$ converges to a unique point, $P \in R^N$. Furthermore, by the definition of T it follows that x^{s+k}, $k = 0, 1, 2, ...,$ lies in H, hence $P \in H$. The remaining steps are essentially the same as in the theorem.

For $p = 2$, there is only a denumerable set of starting points x^o such that $\{x^q\}$ will terminate at an a_i after a finite number of iterations (Kuhn, 1973). This result can be extended to $p > 2$. However, for $1 \leq p < 2$, the singular points of the iteration function vector ϕ comprise the hyperplanes H_{it}; and we obtain the following.

Property 5.2.10

For $p \in [1, 2)$, the sequence x^q, $q = 0, 1, 2, ...,$ converges to x^*, except for a set of starting points x^o which is dense as the set R^{N-1}.

Proof:

Consider a point $Q \in S$ such that if $x^r = Q$ for some r, the sequence results in a non-optimal solution. Then

$$\{x^o \mid x^1 = Q\}$$

is finite, since x^o solves a system of algebraic equations. It follows that

$$\{x^o \mid x^r = Q \quad \text{for some } r\}$$

is denumerable. Let

$$S' = \{Q \in S \mid \text{ if } x^r = Q \quad \text{then } \lim_{q \to \infty} x^q \neq x^*\}.$$

Since S′ is dense as the set R^{N-1}, we conclude that

$$\{x^o \mid x^r \in S' \quad \text{for some } r\}$$

is also dense as the set R^{N-1}.

Since x^o can be any point in R^N, the likelihood that the algorithm will not converge to x^* for an arbitrarily chosen starting point should be very low (zero theoretically if the sequence is calculated with unlimited accuracy). As a consequence of Property 5.2.10, we are well-advised to use double precision arithmetic when $1 \leq p < 2$. A topic for future consideration would be the use of a variable step-size when an iterate lands on a singular point, extending the results of Ostresh (1978) for $p = 2$. As a final comment, we note that although global convergence of a regular sequence is not guaranteed for $p > 2$ (e.g., see counter-examples in Figure 5.2.1), this is not a practical limitation, since we only need to consider values of p in the interval [1, 2] for properly oriented axes (Chapter 3).

5.3 Local Convergence Rates of the Weiszfeld Procedure

Now that global convergence has been proven for $1 \leq p \leq 2$, we turn our attention to the behaviour of the sequence x^q when the iterates are close to an optimal solution x^*. Katz (1974) studies the local convergence rates of the Weiszfeld procedure for the single facility minisum problem in N-dimensions with Euclidean distances ($p = 2$). For this case he shows that the local convergence is always linear if x^* is not a fixed point. Furthermore, for $N=2$, the upper asymptotic convergence bound (λ_M) takes on a value in the range $1/2 \leq \lambda_M < 1$. If x^* is a destination, the local convergence rate is usually linear, but it can be quadratic or sublinear in certain cases.

The only published results concerning local convergence of the Weiszfeld procedure appear to be those given by Katz (1974). Our objective then is to extend these results to the single facility minisum problem with ℓ_p distances, where $1 \leq p \leq 2$. We shall soon see that the analysis is considerably more complex, and that a basically different methodology than the one by Katz (1974) is required, because of the more cumbersome form of the iteration functions ϕ_t for general p. (Compare the functional forms in (5.2.17) and (5.2.21)). Interestingly though, the general results obtained by Katz (1974) for Euclidean distances also apply when p takes on a value in the open interval (1, 2), but a different situation holds for $p = 1$. In the first case the ℓ_p function is a round norm, whereas it becomes a block norm when $p = 1$.

We study the local convergence rates in great detail for the two-dimensional problem ($N = 2$), since location in the plane occurs most commonly in practice. The analysis also leads to some interesting observations for values of p outside the interval [1,2]. The results are then extended to the single facility minisum problem in R^N.

5.3.1 Convergence to a Non-Singular Point

We shall assume here as in Katz (1974) that the fixed points or destinations a_i are non-collinear. For $p > 1$, this guarantees a unique solution x^* (Corollary 5.1.2). The collinear case is trivial to solve in two dimensions (see Corollary 5.1.1 and Property 5.1.4), so that a Weiszfeld iterative procedure would not be required here. In the following analysis, the same notation is used as in section 5.2, unless stated otherwise.

Recall that the iteration functions are given by

$$\phi_t(x) = \sum_{i=1}^{n} \gamma_{it}(x)\, a_{it} \Big/ \sum_{i=1}^{n} \gamma_{it}(x) , \quad t = 1, \ldots, N , \tag{5.3.1}$$

where

$$\gamma_{it}(x) = \frac{w_i \,| x_t - a_{it} |^{p-2}}{[\ell_p (x - a_i)]^{p-1}} , \quad i = 1, \ldots, n , \quad t = 1, \ldots, N . \tag{5.3.2}$$

Let us consider first the case where x^* is not a singular point of the ϕ_t; i.e., $x^* \notin S$. Since the first partial derivatives of the objective function $W(x)$ are equal to zero at x^*, it follows immediately that

$$x_t^* = \phi_t(x^*) , \quad t = 1, \ldots, N , \tag{5.3.3}$$

or in vector notation,

$$x^* = \phi(x^*) . \tag{5.3.4}$$

For $p \in [1, 2]$, the iteration functions are infinitely differentiable at any $x \notin S$. Thus, we can rewrite $\phi(x)$ in a δ-neighbourhood of x^* in terms of its Taylor series expansion at x^*. For sufficiently small δ, the higher-order terms in the series become insignificant. Letting $\| \cdot \|$ denote the Euclidean distance, we obtain

$$\begin{aligned} \phi(x) &= \phi(x^*) + \phi'(x^*) \cdot (x - x^*) + O(\| x - x^* \|^2) \\ &= x^* + \phi'(x^*) \cdot (x - x^*) + O(\| x - x^* \|^2) , \qquad x \in \Delta , \end{aligned} \tag{5.3.5}$$

where Δ denotes the δ-neighbourhood of x^*, and $\phi'(x^*)$ is the $N \times N$ matrix of first partials of ϕ evaluated at x^*; i.e.,

$$\phi'(x^*) = \begin{bmatrix} \dfrac{\partial \phi_1 (x^*)}{\partial x_1} & \cdots & \dfrac{\partial \phi_1 (x^*)}{\partial x_N} \\ \vdots & & \vdots \\ \dfrac{\partial \phi_N (x^*)}{\partial x_1} & \cdots & \dfrac{\partial \phi_N (x^*)}{\partial x_N} \end{bmatrix} . \tag{5.3.6}$$

We proceed now to calculate the elements of the above matrix. For general x,

$$\frac{\partial \Phi_t(x)}{\partial x_j} = \frac{\partial}{\partial x_j}\left[\frac{\sum_i \gamma_{it}(x)\cdot a_{it}}{\sum_i \gamma_{it}(x)}\right]$$

$$= \frac{\sum_i \frac{\partial}{\partial x_j}[\gamma_{it}(x)]\cdot a_{it}}{\sum_i \gamma_{it}(x)} - \frac{\sum_i \gamma_{it}(x)\cdot a_{it}}{\left[\sum_i \gamma_{it}(x)\right]^2}\cdot \sum_i \frac{\partial}{\partial x_j}[\gamma_{it}(x)]$$

$$= \frac{1}{s_t(x)}\left\{\sum_i \frac{\partial}{\partial x_j}[\gamma_{it}(x)]\cdot a_{it} - \frac{\sum_i \gamma_{it}(x)\cdot a_{it}}{\sum_i \gamma_{it}(x)}\cdot \sum_i \frac{\partial}{\partial x_j}[\gamma_{it}(x)]\right\} , \qquad (5.3.7)$$

where

$$s_t(x) = \sum_i \gamma_{it}(x) ,$$

and the summations are understood to be over the index set $\{1, \ldots, n\}$. At $x = x^*$, the above expression reduces by means of (5.3.3) to

$$\frac{\partial \Phi_t(x^*)}{\partial x_j} = \frac{1}{s_t(x^*)}\sum_{i=1}^{n}\left[\frac{\partial}{\partial x_j}\gamma_{it}(x^*)\right]\cdot(a_{it} - x_t^*) . \qquad (5.3.8)$$

Using standard calculus, we obtain

$$\frac{\partial}{\partial x_j}[\gamma_{it}(x)] = \frac{(1-p)\, w_i \,\mathrm{sign}(x_j - a_{ij})|x_j - a_{ij}|^{p-1}|x_t - a_{it}|^{p-2}}{[\ell_p(x - a_i)]^{2p-1}} , \quad j \neq t , \qquad (5.3.9)$$

and

$$\frac{\partial}{\partial x_t}[\gamma_{it}(x)] = \frac{-w_i\,\mathrm{sign}(x_t - a_{it})\,|x_t - a_{it}|^{p-3}}{[\ell_p(x - a_i)]^{p-1}}\cdot\left[(2-p) + \frac{(p-1)|x_t - a_{it}|^{p}}{[\ell_p(x - a_i)]^{p}}\right] . \qquad (5.3.10)$$

Thus,

$$\frac{\partial \phi_t(x^*)}{\partial x_j} = \frac{(p-1)}{s_t(x^*)} \sum_{i=1}^{n} \frac{w_i \operatorname{sign}(x_j^* - a_{ij}) \operatorname{sign}(x_t^* - a_{it}) \, |x_j^* - a_{ij}|^{p-1} |x_t^* - a_{it}|^{p-1}}{[\ell_p(x^* - a_i)]^{2p-1}},$$

$$j \neq t, \tag{5.3.11}$$

and

$$\frac{\partial \phi_t(x^*)}{\partial x_t} = \frac{1}{s_t(x^*)} \sum_{i=1}^{n} w_i \frac{|x_t^* - a_{it}|^{p-2}}{[\ell_p(x^* - a_i)]^{p-1}} \left\{ (2-p) + (p-1) \frac{|x_t^* - a_{it}|^{p}}{[\ell_p(x^* - a_i)]^{p}} \right\}$$

$$= (2-p) + \frac{(p-1)}{s_t(x^*)} \cdot \sum_{i=1}^{n} \gamma_{it}(x^*) \frac{|x_t^* - a_{it}|^{p}}{[\ell_p(x^* - a_i)]^{p}}. \tag{5.3.12}$$

Using equations (5.3.11) and (5.3.12), we can readily construct the $N \times N$ matrix $\phi'(x^*)$. For the planar case ($N = 2$),

$$\phi'(x^*) = \begin{vmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{vmatrix}, \tag{5.3.13}$$

where

$$\phi_{11} = (2-p) + \frac{(p-1)}{s_1(x^*)} \sum_{i=1}^{n} \gamma_{i1}(x^*) \frac{|x_1^* - a_{i1}|^{p}}{[\ell_p(x^* - a_i)]^{p}},$$

$$\phi_{12} = \frac{(p-1)}{s_1(x^*)} \sum_{i=1}^{n} w_i \operatorname{sign}(x_1^* - a_{i1}) \operatorname{sign}(x_2^* - a_{i2}) \frac{|x_1^* - a_{i1}|^{p-1} |x_2^* - a_{i2}|^{p-1}}{[\ell_p(x^* - a_i)]^{2p-1}},$$

$$\phi_{21} = \frac{(p-1)}{s_2(x^*)} \sum_{i=1}^{n} w_i \operatorname{sign}(x_1^* - a_{i1}) \operatorname{sign}(x_2^* - a_{i2}) \frac{|x_1^* - a_{i1}|^{p-1} |x_2^* - a_{i2}|^{p-1}}{[\ell_p(x^* - a_i)]^{2p-1}},$$

and

$$\phi_{22} = (2-p) + \frac{(p-1)}{s_2(x^*)} \sum_{i=1}^{n} \gamma_{i2}(x^*) \frac{|x_2^* - a_{i2}|^{p}}{[\ell_p(x^* - a_i)]^{p}}.$$

When distances are Euclidean, this matrix reduces to the special form,

$$\phi'(x^*) = \frac{1}{s(x^*)} \begin{bmatrix} \sum_{i=1}^{n} \frac{w_i (x_1^* - a_{i1})^2}{[\ell_2 (x^* - a_i)]^3} & \sum_{i=1}^{n} \frac{w_i (x_1^* - a_{i1})(x_2^* - a_{i2})}{[\ell_2 (x^* - a_i)]^3} \\ \sum_{i=1}^{n} \frac{w_i (x_1^* - a_{i1})(x_2^* - a_{i2})}{[\ell_2 (x^* - a_i)]^3} & \sum_{i=1}^{n} \frac{w_i (x_2^* - a_{i2})^2}{[\ell_2 (x^* - a_i)]^3} \end{bmatrix}, \quad (5.3.14)$$

$$p = 2,$$

where

$$s(x) = \sum_{i=1}^{n} w_i / \ell_2 (x - a_i) .$$

The matrix in (5.3.14) agrees with the result obtained by Katz (1974).

The eigenvalues of $\phi'(x^*)$ are important quantities, since they determine the ultimate rate of convergence of the iteration procedure. Denoting the eigenvalues by λ_j, $j = 1, \ldots, N$, a necessary condition for convergence to x^* is that

$$\mu = \max_j \{|\lambda_j|\} \leq 1 \quad (5.3.15)$$

(e.g., see Dahlquist and Björck, 1974, or Ortega and Rheinboldt, 1970). Furthermore, if this relation holds strictly, the local convergence rate will be linear or better, since μ equals the upper asymptotic convergence bound (Katz, 1974).

We now derive some properties of the 2×2 matrix $\phi'(x^*)$ in (5.3.13), which will be useful in characterizing its eigenvalues, λ_1 and λ_2.

Property 5.3.1

The eigenvalues of the matrix in (5.3.13) are real.

Proof:

For $p = 2$, $\phi'(x^*)$ is a real symmetric matrix, as seen by (5.3.14). Hence, the eigenvalues are real in this case (Finkbeiner, 1972, Theorem 5.19).

For $p \neq 2$, $\phi'(x^*)$ is no longer symmetric, since $s_1(x^*) \neq s_2(x^*)$ in general. Hence, the desired result is not immediately obvious. Consider the characteristic equation,

$$\det[\phi'(x^*) - \lambda I] = 0 , \tag{5.3.16}$$

where I is the identity matrix, λ denotes an eigenvalue, and 'det' symbolizes the determinant. For the 2×2 matrix, we obtain the quadratic equation,

$$\lambda^2 - \lambda(\phi_{11} + \phi_{22}) + \phi_{11}\phi_{22} - \phi_{21}\phi_{12} = 0 , \tag{5.3.17}$$

whose roots are

$$\lambda_1 = \frac{1}{2}(\phi_{11} + \phi_{22}) - \frac{1}{2}\sqrt{(\phi_{11} - \phi_{22})^2 + 4\phi_{21}\phi_{12}} , \tag{5.3.18a}$$

and

$$\lambda_2 = \frac{1}{2}(\phi_{11} + \phi_{22}) + \frac{1}{2}\sqrt{(\phi_{11} - \phi_{22})^2 + 4\phi_{21}\phi_{12}} . \tag{5.3.18b}$$

Since

$$(\phi_{11} - \phi_{22})^2 \geq 0 ,$$

and

$$\phi_{21}\phi_{12} = \frac{(p-1)^2}{s_1(x^*)s_2(x^*)} \left| \sum_{i=1}^{n} w_i \,\mathrm{sign}(x_1^* - a_{i1})\mathrm{sign}(x_2^* - a_{i2}) \frac{|x_1^* - a_{i1}|^{p-1}|x_2^* - a_{i2}|^{p-1}}{[\ell_p(x^* - a_i)]^{2p-1}} \right|^2$$

$$\geq 0 ,$$

it follows that

$$(\phi_{11} - \phi_{22})^2 + 4\phi_{21}\phi_{12} \geq 0 .$$

Hence, λ_1 and λ_2 are real valued.

Property 5.3.2

For $1 \leq p \leq 2$,

$$\det[\phi'(x^*)] > 0 ,$$

where $\phi'(x^*)$ is the matrix in (5.3.13) and we assume that the fixed points a_i are non-collinear.

Proof:

Let

$$F_1(x) := \sum_{i=1}^{n} \gamma_{i1}(x) \frac{|x_1 - a_{i1}|^p}{[\ell_p(x - a_i)]^p} = \sum_{i=1}^{n} \frac{w_i |x_1 - a_{i1}|^{2p-2}}{[\ell_p(x - a_i)]^{2p-1}}, \tag{5.3.19}$$

$$F_2(x) := \sum_{i=1}^{n} \gamma_{i2}(x) \frac{|x_2 - a_{i2}|^p}{[\ell_p(x - a_i)]^p} = \sum_{i=1}^{n} \frac{w_i |x_2 - a_{i2}|^{2p-2}}{[\ell_p(x - a_i)]^{2p-1}}, \tag{5.3.20}$$

and

$$H(x) := \sum_{i=1}^{n} w_i \,\mathrm{sign}(x_1 - a_{i1})\mathrm{sign}(x_2 - a_{i2}) \frac{|x_1 - a_{i1}|^{p-1}|x_2 - a_{i2}|^{p-1}}{[\ell_p(x - a_i)]^{2p-1}}. \tag{5.3.21}$$

Using Schwarz's inequality, we obtain

$$|H(x)| \le \left(\sum_{i=1}^{n} w_i \frac{|x_1 - a_{i1}|^{2p-2}}{[\ell_p(x - a_i)]^{2p-1}}\right)^{1/2} \left(\sum_{i=1}^{n} w_i \frac{|x_2 - a_{i2}|^{2p-2}}{[\ell_p(x - a_i)]^{2p-1}}\right)^{1/2}$$

$$= [F_1(x) \cdot F_2(x)]^{1/2}. \tag{5.3.22}$$

Equality holds above if, and only if,

$$\mathrm{sign}(x_1 - a_{i1})|x_1 - a_{i1}|^{p-1} = c\,\mathrm{sign}(x_2 - a_{i2})|x_2 - a_{i2}|^{p-1}, \tag{5.3.23}$$

for some scalar c and $i = 1, \ldots, n$. For $p > 1$, this implies that

$$|x_1 - a_{i1}| = |c|^{\frac{1}{p-1}} |x_2 - a_{i2}|, \quad \forall i,$$

so that (5.3.23) can be rewritten in the form,

$$(x_1 - a_{i1}) = c'(x_2 - a_{i2}), \quad \forall i, \tag{5.3.24}$$

where

$$c' = \mathrm{sign}(c) \cdot |c|^{\frac{1}{p-1}}.$$

Thus, equality holds in (5.3.22) for $p > 1$ if, and only if, the a_i are collinear.

Now consider the determinant of $\phi'(x^*)$.

$$\det[\phi'(x^*)] = \phi_{11}\phi_{22} - \phi_{21}\phi_{12}$$

$$= (2-p)^2 + (2-p)(p-1)\left[\frac{F_1(x^*)}{s_1(x^*)} + \frac{F_2(x^*)}{s_2(x^*)}\right]$$

$$+ \frac{(p-1)^2}{s_1(x^*)\,s_2(x^*)}[F_1(x^*)\,F_2(x^*) - \{H(x^*)\}^2]$$

$$\geq (2-p)^2 + (2-p)(p-1)\left[\frac{F_1(x^*)}{s_1(x^*)} + \frac{F_2(x^*)}{s_2(x^*)}\right]$$

(relation (5.3.22))

$$\geq 0 \,, \qquad 1 \leq p \leq 2 \,. \tag{5.3.25}$$

Since the a_i are non-collinear, the first inequality in (5.3.25) is satisfied strictly for $p > 1$. Furthermore, the second inequality is satisfied strictly for $1 \leq p < 2$. Hence, we conclude that $\det[\phi'(x^*)] > 0$ for $1 \leq p \leq 2$.

Now consider the trace of the matrix $\phi'(x^*)$, denoted by $\mathrm{tr}\,[\phi'(x^*)]$. For $p = 2$, we see from (5.3.14) that $\mathrm{tr}\,[\phi'(x^*)] = 1$. This result also holds for higher dimensions ($N > 2$). However, from (5.3.13) we note the following interesting fact:

$$\lim_{p \to 1} \mathrm{tr}[\phi'(x^*)] = 2 \;; \tag{5.3.26a}$$

while for $N > 2$, it is readily shown that

$$\lim_{p \to 1} \mathrm{tr}\,[\phi'(x^*)] = N \,. \tag{5.3.26b}$$

We are assuming here that x^* remains a non-singular point, since $\phi'(x^*)$ is undefined otherwise. Thus, the trace of $\phi'(x^*)$ varies as a function of p. The following result places bounds on this function.

Property 5.3.3

If $p \in [1, 2]$, then

$$0 < \mathrm{tr}\,[\phi'(x^*)] \leq 2 \,, \tag{5.3.27}$$

where $\phi'(x^*)$ is the 2×2 matrix in (5.3.13). Furthermore, if $p \neq 1$, the upper bound is satisfied strictly.

Proof:

$$\mathrm{tr}[\phi'(x^*)] = (4-2p) + (p-1)\left[\frac{F_1(x^*)}{s_1(x^*)} + \frac{F_2(x^*)}{s_2(x^*)}\right]. \tag{5.3.28}$$

Since

$$\frac{F_j(x^*)}{s_j(x^*)} > 0, \qquad \forall\ j, \tag{5.3.29}$$

it is immediately obvious that

$$\mathrm{tr}[\phi'(x^*)] > 0.$$

Rewriting

$$\frac{F_j(x^*)}{s_j(x^*)} = \sum_{i=1}^{n} \beta_{ij}(x^*)|x_j^* - a_{ij}|^p / [\ell_p(x^* - a_i)]^p, \tag{5.3.30}$$

where

$$\beta_{ij}(x) = \gamma_{ij}(x) / \sum_{k=1}^{n} \gamma_{kj}(x), \qquad \forall\ i,j, \tag{5.3.31}$$

we obtain

$$\frac{F_j(x^*)}{s_j(x^*)} \le \max_{1 \le i \le n} \left\{ \frac{|x_j^* - a_{ij}|^p}{[\ell_p(x^* - a_i)]^p} \right\} \le 1, \tag{5.3.32}$$

with at least one of the inequalities being satisfied strictly for each j. Thus,

$$\mathrm{tr}[\phi'(x^*)] \le (4-2p) + (p-1) \times 2 = 2, \tag{5.3.33}$$

with equality only when $p = 1$.

For $N > 2$, the preceding property is generalized in a straightforward manner to obtain

$$0 < \mathrm{tr}[\phi'(x^*)] \le N, \quad p \in [1, 2], \tag{5.3.34}$$

with the upper bound being satisfied strictly except when $p = 1$.

We have insufficient information at this point to make any conclusions concerning the local convergence rate of the iteration procedure. However, the following observations lead to an interesting resolution of this problem. From equation (5.3.11) and the second-order partial derivatives given in Chapter 7 (set s = p in (7.2.5)), it is readily seen that

$$\frac{\partial \phi_t(x^*)}{\partial x_j} = -\frac{1}{s_t(x^*)}\frac{\partial^2 W(x^*)}{\partial x_j \partial x_t}, \qquad j \neq t. \tag{5.3.35}$$

From (5.3.12) and (7.2.4), we obtain

$$\frac{\partial \phi_t(x^*)}{\partial x_t} = 1-(1-p)\left|\frac{1}{s_t(x^*)}\sum_{i=1}^{n}\gamma_{it}(x^*)\frac{|x_t^*-a_{it}|^p}{[\ell_p(x^*-a_i)]^p}-1\right|$$

$$= 1-\frac{(1-p)}{s_t(x^*)}\sum_{i=1}^{n}\gamma_{it}(x^*)\left|\frac{|x_t^*-a_{it}|^p}{[\ell_p(x^*-a_i)]^p}-1\right|$$

$$= 1-\frac{1}{s_t(x^*)}\frac{\partial^2 W(x^*)}{\partial x_t^2}, \qquad t = 1,\dots,N. \tag{5.3.36}$$

Thus, the $N \times N$ matrix $\phi'(x^*)$ (see (5.3.6)) takes the form,

$$\phi'(x^*) = \begin{bmatrix} 1-\frac{1}{s_1(x^*)}\frac{\partial^2 W(x^*)}{\partial x_1^2} & -\frac{1}{s_1(x^*)}\frac{\partial^2 W(x^*)}{\partial x_2 \partial x_1} & \cdots & -\frac{1}{s_1(x^*)}\frac{\partial^2 W(x^*)}{\partial x_N \partial x_1} \\ -\frac{1}{s_2(x^*)}\frac{\partial^2 W(x^*)}{\partial x_1 \partial x_2} & 1-\frac{1}{s_2(x^*)}\frac{\partial^2 W(x^*)}{\partial x_2^2} & \cdots & -\frac{1}{s_2(x^*)}\frac{\partial^2 W(x^*)}{\partial x_N \partial x_2} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{s_N(x^*)}\frac{\partial^2 W(x^*)}{\partial x_1 \partial x_N} & -\frac{1}{s_N(x^*)}\frac{\partial^2 W(x^*)}{\partial x_2 \partial x_N} & \cdots & 1-\frac{1}{s_N(x^*)}\frac{\partial^2 W(x^*)}{\partial x_N^2} \end{bmatrix}. \tag{5.3.37}$$

Let us consider again location in the plane (N = 2), where $\phi'(x^*)$ is the 2×2 matrix given in (5.3.13). Using (5.3.37), we see that

$$\phi'(x^*) = I - B(x^*) , \tag{5.3.38}$$

where

$$B(x^*) = \begin{bmatrix} \frac{1}{s_1(x^*)} \frac{\partial^2 W(x^*)}{\partial x_1^2} & \frac{1}{s_1(x^*)} \frac{\partial^2 W(x^*)}{\partial x_2 \partial x_1} \\ \frac{1}{s_2(x^*)} \frac{\partial^2 W(x^*)}{\partial x_1 \partial x_2} & \frac{1}{s_2(x^*)} \frac{\partial^2 W(x^*)}{\partial x_2^2} \end{bmatrix} . \tag{5.3.39}$$

Letting μ_1 and μ_2 denote the eigenvalues of $B(x^*)$, we obtain the following preliminary result.

Lemma 5.3.1

If $p > 1$, and the fixed points a_i are non-collinear, then μ_1 and μ_2 are positive; i.e.,

$$\mu_1 , \mu_2 > 0 .$$

Proof:

By Property (5.1.2), we know that W is a strictly convex function of x. Thus,

$$\frac{\partial^2 W(x^*)}{\partial x_1^2} > 0 , \qquad \frac{\partial^2 W(x^*)}{\partial x_2^2} > 0 ,$$

and

$$\frac{\partial^2 W(x^*)}{\partial x_1^2} \frac{\partial^2 W(x^*)}{\partial x_2^2} - \left(\frac{\partial^2 W(x^*)}{\partial x_1 \partial x_2} \right)^2 > 0 .$$

Also note that $s_1(x^*), s_2(x^*) > 0$. It follows that

$$\begin{aligned} \mu_1 \mu_2 &= \det [B(x^*)] \\ &= \frac{1}{s_1(x^*) s_2(x^*)} \left[\frac{\partial^2 W(x^*)}{\partial x_1^2} \frac{\partial^2 W(x^*)}{\partial x_2^2} - \left(\frac{\partial^2 W(x^*)}{\partial x_1 \partial x_2} \right)^2 \right] \\ &> 0 . \end{aligned} \tag{5.3.40}$$

Furthermore,

$$\mu_1 + \mu_2 = \mathrm{tr}[B(x^*)]$$

$$= \frac{1}{s_1(x^*)} \frac{\partial^2 W(x^*)}{\partial x_1^2} + \frac{1}{s_2(x^*)} \frac{\partial^2 W(x^*)}{\partial x_2^2}$$

$$> 0 . \tag{5.3.41}$$

Using a similar method as in Property 5.3.1, it is readily shown that μ_1 and μ_2 are real. Hence, we conclude that $\mu_1, \mu_2 > 0$.

The main result follows at last.

Theorem 5.3.1

Let p take on a value in the range $1 < p \leq 2$ and the fixed points a_i be non-collinear. Then the eigenvalues of the 2×2 matrix $\phi'(x^*)$ in (5.3.13) satisfy the following relation:

$$0 < \lambda_1, \lambda_2 < 1 . \tag{5.3.42}$$

Hence, the asymptotic convergence rate to the non-singular point x* is linear.

Proof:

The eigenvalues, λ_1 and λ_2, are real by Property 5.3.1. From Properties 5.3.2 and 5.3.3, it follows that

$$\lambda_1 \lambda_2 > 0 , \tag{5.3.43}$$

and

$$0 < \lambda_1 + \lambda_2 < 2 . \tag{5.3.44}$$

Hence

$$\lambda_1, \lambda_2 > 0 . \tag{5.3.45}$$

Since $\phi'(x^*) = I - B(x^*)$, we also have

$$\lambda_1 = 1 - \mu_1, \quad \lambda_2 = 1 - \mu_2 . \tag{5.3.46}$$

But $\mu_1, \mu_2 > 0$ by Lemma 5.3.1, and hence

$$\lambda_1, \lambda_2 < 1 . \tag{5.3.47}$$

Combining (5.3.45) and (5.3.47), we obtain the desired relation (5.3.42).

To show that the local convergence rate to x* is linear, we employ a direct extension of the method in Katz (1974). Let

$$x = U\tilde{x} \tag{5.3.48}$$

denote the transformation from the original x-coordinates to new coordinates ($\tilde{x}$) with respect to the eigenvectors of $\phi'(x^*)$ (columns of U). Then, pre-multiplying (5.3.5) by U^{-1} and using standard linear algebra (e.g., see Stephenson, 1966, Chapter 6), it follows that

$$\begin{aligned}\tilde{x}^{q+1} &= \tilde{x}^* + U^{-1}\phi'(x^*)\, U \cdot (\tilde{x}^q - \tilde{x}^*) + O(\|x^q - x^*\|^2)\\ &= \tilde{x}^* + \mathrm{diag}(\lambda_1, \lambda_2) \cdot (\tilde{x}^q - \tilde{x}^*) + O(\|x^q - x^*\|^2) ,\end{aligned} \tag{5.3.49}$$

where

$$\mathrm{diag}(\lambda_1, \lambda_2) := \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} .$$

Thus,

$$(\tilde{x}_1^{q+1} - \tilde{x}_1^*)^2 + (\tilde{x}_2^{q+1} - \tilde{x}_2^*)^2 = \lambda_1^2(\tilde{x}_1^q - \tilde{x}_1^*)^2 + \lambda_2^2(\tilde{x}_2^q - \tilde{x}_2^*)^2 + O(\|x^q - x^*\|^3) . \tag{5.3.50}$$

Recall from equations (5.3.18a) and (5.3.18b) that $\lambda_1 \leq \lambda_2$ by definition. Therefore, from the preceding relation, we obtain

$$|\lambda_1| \cdot \|\tilde{x}^q - \tilde{x}^*\| \leq \|\tilde{x}^{q+1} - \tilde{x}^*\| \leq |\lambda_2| \cdot \|\tilde{x}^q - \tilde{x}^*\| , \tag{5.3.51}$$

where higher-order terms have been neglected, and the Euclidean norm ($\|\cdot\|$) now acts on the transformed coordinates. Hence, $|\lambda_1|$ and $|\lambda_2|$ give the lower and upper asymptotic convergence bounds respectively. Since relation (5.3.42) holds, we conclude that the local convergence rate to the non-singular optimum x* is linear.

A few comments are required concerning the transformation in (5.3.48). When $p = 2$, we see from (5.3.14) that $\phi'(x^*)$ is a symmetric matrix. Hence, the eigenvectors in U are orthogonal (e.g., Stephenson, 1966, Chapter 5). Furthermore, if the eigenvectors are normalised to unit length, then $U^T = U^{-1}$ for this case. However, when $p \neq 2$, $\phi'(x^*)$ loses its symmetry, so that the transformation becomes non-orthogonal in general. For the special case where $\lambda_1 = \lambda_2$, the off-diagonal elements ϕ_{12} and ϕ_{21} must equal zero; hence, $\phi'(x^*)$ is already in diagonalized form and $U = I$. In all cases, the inverse U^{-1} exists and equation (5.3.49) is obtainable. We now proceed to illustrate relation (5.3.42) with a few simple numerical examples.

Example 1:

Four fixed locations are given which form the corners of a square: $a_1 = (-1, -1)$, $a_2 = (1, -1)$, $a_3 = (1, 1)$, $a_4 = (-1, 1)$. The weight at each fixed location is the same; i.e., $w_1 = w_2 = w_3 = w_4 = w$. For $p > 1$, there is a unique optimal solution at the centre of the square; $x^* = (0,0)$. For $p = 1$, all points contained in the square are optimal. The elements of $\phi'(x^*)$ are easily calculated, yielding the diagonal matrix,

$$\phi'(x^*) = \begin{bmatrix} \frac{3}{2} - \frac{1}{2}p & 0 \\ 0 & \frac{3}{2} - \frac{1}{2}p \end{bmatrix} .$$

Thus,

$$\lambda_1 = \lambda_2 = \frac{3}{2} - \frac{1}{2}p, \quad p \geq 1 .$$

Example 2:

The fixed points remain the same, but the weights are now $w_1 = w_3 = w$, and $w_2 = w_4 = 1/2\,w$. The optimal location is unchanged from the previous example; $x^* = (0,0)$. We obtain

$$\gamma_{i1}(x^*) = \gamma_{i2}(x^*) = \tau w_i, \quad i = 1, \ldots, 4 ,$$

where $\tau = 2^{(1-p)/p}$. Thus,

$$s_1(x^*) = s_2(x^*) = \sum_{i=1}^{4} \tau w_i = 3\,\tau w \ .$$

After some simple calculations, we find

$$\phi'(x^*) = \begin{vmatrix} \frac{3}{2} - \frac{1}{2}p & \frac{p-1}{6} \\ \frac{p-1}{6} & \frac{3}{2} - \frac{1}{2}p \end{vmatrix} .$$

Solving for the eigenvalues gives

$$\lambda_1 = \frac{5}{3} - \frac{2}{3}p, \quad \lambda_2 = \frac{4}{3} - \frac{1}{3}p, \quad p \geq 1 .$$

Example 3:

In the previous examples, the eigenvalues vary as linear functions of p, because of the inherent symmetry of the fixed points and the weights. We move the four fixed points now to the corners of a rectangle as follows: $a_1 = (-3, -1)$, $a_2 = (3, -1)$, $a_3 = (3,1)$, $a_4 = (-3, 1)$. Let the weights be the same at each a_i; $w_1 = w_2 = w_3 = w_4 = w$. Thus, the optimal location remains at $x^* = (0, 0)$. The elements of $\phi'(x^*)$ are readily calculated, to obtain

$$\phi'(x^*) = \begin{vmatrix} 2 - \left(\frac{3^p + p}{1 + 3^p}\right) & 0 \\ 0 & 2 - \left(\frac{1 + 3^p \cdot p}{1 + 3^p}\right) \end{vmatrix} .$$

Hence,

$$\lambda_1 = 2 - \left(\frac{1 + 3^p \cdot p}{1 + 3^p}\right), \ \lambda_2 = 2 - \left(\frac{3^p + p}{1 + 3^p}\right), \quad p \geq 1 .$$

Example 4:

The a_i remain unchanged from the previous example, but now $w_1 = w_3 = w$ and $w_2 = w_4 = 1/2\ w$. The calculations proceed as follows.

$$\gamma_{i1}(x^*) = \frac{w_i |x_1^* - a_{i1}|^{p-2}}{[\ell_p(x^* - a_i)]^{p-1}} = \frac{w_i \cdot 3^{p-2}}{[1 + 3^p]^{(p-1)/p}} \quad , i = 1, \ldots, 4 ;$$

$$s_1(x^*) = \sum_{i=1}^{4} \gamma_{i1}(x^*) = \tau_1 w, \quad \text{where } \tau_1 = \frac{3^{p-1}}{[1 + 3^p]^{(p-1)/p}} ;$$

$$\gamma_{i2}(x^*) = \frac{w_i |x_2^* - a_{i2}|^{p-2}}{[\ell_p(x^* - a_i)]^{p-1}} = \frac{w_i}{[1 + 3^p]^{(p-1)/p}} \quad , i = 1, \ldots, 4 ;$$

$$s_2(x^*) = \sum_{i=1}^{4} \gamma_{i2}(x^*) = \tau_2 w, \quad \text{where } \tau_2 = \frac{3}{[1 + 3^p]^{(p-1)/p}} .$$

The diagonal elements ϕ_{11} and ϕ_{22} are easily verified to be the same as in the previous example. The off-diagonal elements are given by

$$\phi_{12} = \frac{(p-1)}{\tau_i w} \cdot \frac{3^{p-1}}{(1 + 3^p)^{(2p-1)/p}} \cdot \{w_1 - w_2 + w_3 - w_4\}$$

$$= (p-1)/(1 + 3^p), \text{ and}$$

$$\phi_{21} = \frac{s_1(x^*)}{s_2(x^*)} \phi_{12} = 3^{p-2} \frac{(p-1)}{(1 + 3^p)} .$$

Thus,

$$\phi'(x^*) = \begin{vmatrix} 2 - \left(\dfrac{3^p + p}{1 + 3^p} \right) & \dfrac{p-1}{1 + 3^p} \\ 3^{p-2} \dfrac{(p-1)}{(1 + 3^p)} & 2 - \left(\dfrac{1 + 3^p \cdot p}{1 + 3^p} \right) \end{vmatrix} .$$

Solving for the eigenvalues, we obtain

$$\lambda_1 = \frac{1}{2}(3 - p) - \frac{1}{2}\left(\frac{p-1}{1 + 3^p} \right) \sqrt{(3^p - 1)^2 + (4)3^{p-2}} ,$$

and

$$\lambda_2 = \frac{1}{2}(3 - p) + \frac{1}{2}\left(\frac{p-1}{1 + 3^p} \right) \sqrt{(3^p - 1)^2 + (4)3^{p-2}} , \quad p \geq 1 .$$

As an illustration, the eigenvalues in Example 3 are plotted as functions of p in Figure 5.3.1. It is interesting to note that the preceding examples illustrate a tendency for the eigenvalues to increase, and hence for the local convergence rate to decrease, as p decreases from a value of 2 to 1. Furthermore, as $p \to 1^+$, λ_1 and λ_2 both converge to unity ($\lambda_1, \lambda_2 \to 1^-$). This occurs in all cases where x^* remains a non-singular point, as seen by the following result.

Property 5.3.4

$$\lim_{p \to 1^+} \phi'(x^*) = I, \tag{5.3.52}$$

provided a $\delta > 0$ exists such that $x^* \notin S$ for $1 < p < 1 + \delta$, where S is the set of singular points of ϕ defined in (5.2.42).

Proof:

Since $x^* \notin S$ for $1 < p < 1 + \delta$, the limit is defined. Referring to (5.3.13), we see that the coefficient of $(p-1)$ in each element of $\phi'(x^*)$ is bounded, and hence all terms containing $(p-1)$ go to zero as $p \to 1^+$. Thus, $\phi_{12}, \phi_{21} \to 0$ and $\phi_{11}, \phi_{22} \to 1^-$, giving the desired result for $N = 2$. The same principle readily extends to higher dimensions, so that

$$\lim_{p \to 1^+} \phi'(x^*) = I \text{ for } N \geq 2 .$$

The preceding property has some practical significance. For p slightly greater than 1, the asymptotic convergence to x^* will be at a slow linear rate, since λ_1 and λ_2 have values close to 1. Thus, an acceleration technique such as Steffensen's method would be most advantageous in this case, to finish the iteration sequence. In general, we expect the usefulness of such acceleration methods to increase as p takes on lower values in the interval (1, 2).

Also note that if $p = 1$ and there is an optimal location $x^* \notin S$, then this solution is not unique, and furthermore, a δ-neighbourhood ($\delta > 0$) of x^* exists such that all x inside this neighbourhood are optimal; (see the discussion following Property 5.1.3). Hence, if an iterate

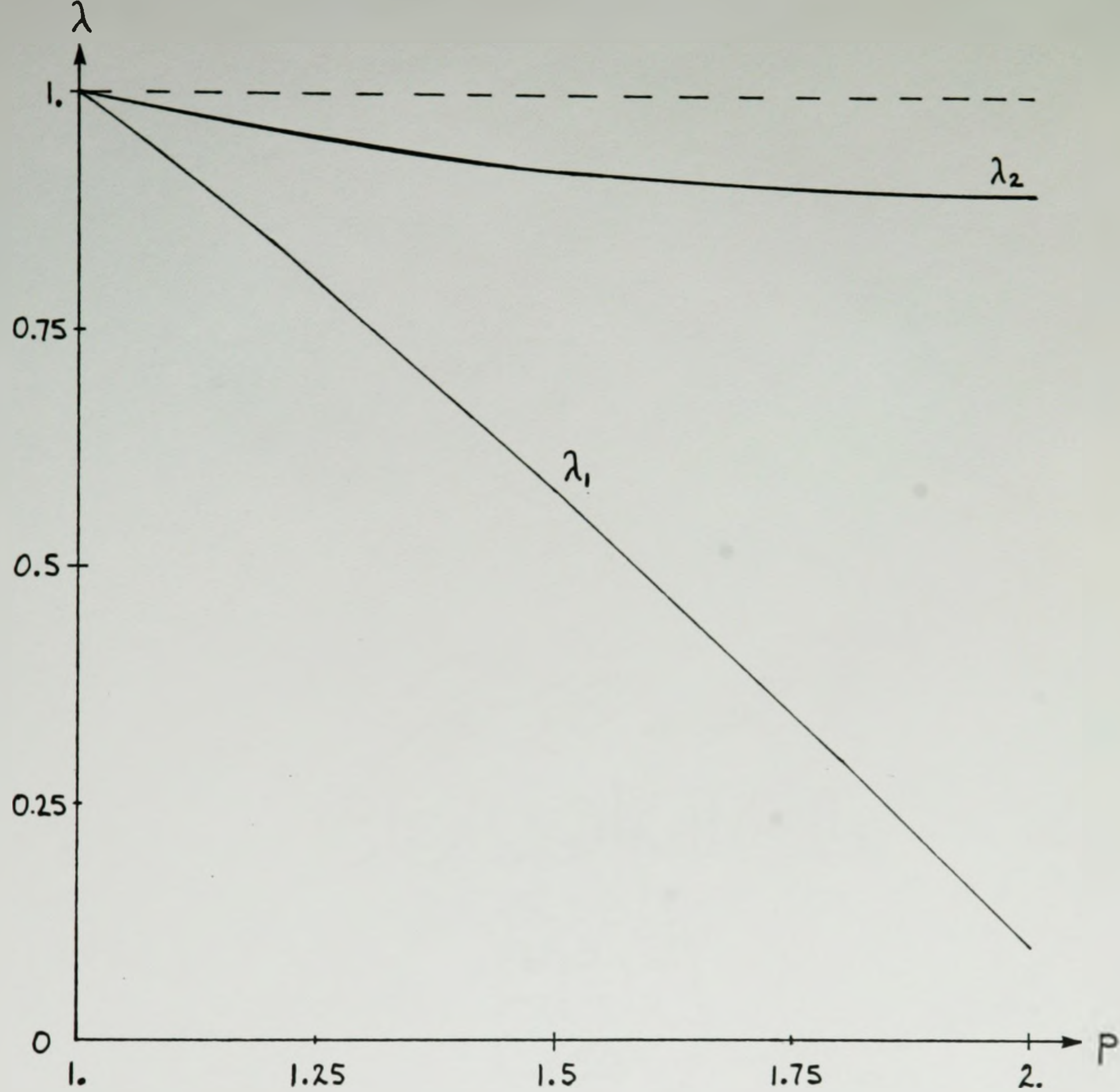


Figure 5.3.1 Eigenvalues of $\phi'(x^*)$ for Example 3.

lands on a point $x_0 \neq x^*$ inside the δ-neighbourhood, the sequence will remain there and never converge to x^*. Such an outcome relates to the fact that eigenvalues λ_1 and λ_2 are precisely equal to one in this case.

An interesting behaviour occurs in the preceding examples when $p > 2$. We see that

$$|\lambda_1| > 1 , \quad p > M , \tag{5.3.53}$$

where $M > 2$ is a sufficiently large value. Thus, the iteration procedure will not converge to x^* in these examples, when p is sufficiently large. This result is formalized below.

Theorem 5.3.2

Convergence of the iteration procedure to x^* will not occur in general for sufficiently large values of p exceeding 2.

Proof: (for $N = 2$)

Let the weights and the geometry of the fixed points be arranged such that

$$\{i \mid |x_1^* - a_{i1}| > |x_2^* - a_{i2}|\}$$

is empty for all $p \in (2, +\infty)$. (Note that examples 1 to 4 satisfy this condition.) It is easily verified that

$$\lim_{p \to +\infty} \operatorname{tr}[\phi'(x^*)] = -\infty .$$

Since $\lambda_1 = \min\{\lambda_1, \lambda_2\}$, and $\lambda_1 + \lambda_2 = \operatorname{tr}[\phi'(x^*)]$, it follows that an $M > 2$ exists such that (5.3.53) holds. Hence, the iteration procedure will not converge to x^* for $p > M$, and the theorem is proven.

The proof above is readily extended to higher dimensions, so that the theorem applies for $N \geq 2$. Some comments are appropriate at this time:

(i) Since $0 < \lambda_1, \lambda_2 < 1$ at $p = 2$ (Theorem 5.3.1), it follows from the continuity of the eigenvalues as functions of p that an $r > 0$ exists such that local convergence also occurs in the interval $2 < p < 2 + r$. However, global convergence of the iteration procedure is guaranteed only for $1 \le p \le 2$, as evidenced by the counter-example in Section 5.2.

(ii) Based on their computational experience, Morris and Verdini (1979) conclude that convergence can be expected for $p > 2$. However, we see now by Theorem 5.3.2 that the iteration procedure will not generally converge to x^*, even locally, if values of p are used which are too high.

5.3.2 Convergence to a Singular Point

We consider now the case where the optimal solution occurs at a singular point of the iteration functions ϕ_t; i.e., $x^* \in S$. For Euclidean distances ($p = 2$), Katz (1974) studies the local convergence of the iterative procedure when x^* is at a fixed point. He shows that the convergence rate is normally linear, but under certain conditions, it can be superlinear or sublinear. We wish to extend these results to ℓ_p distances in general, where $1 \le p \le 2$. Recall that for $p = 2$, the singular points occur only at the fixed locations; whereas for $1 \le p < 2$, the problem becomes more complicated because all points on the hyperplanes, $x_t - a_{it} = 0$, are now included in S.

When $x^* \in S$, one or more of the iteration functions ϕ_t is undefined, as well as its derivatives, so that the Taylor series expansion given in (5.3.5) is no longer feasible. Thus, another method is required to study the behaviour of the iterates close to x^*. In order to simplify the notation in the subsequent analysis, we restrict attention to location in the plane ($N = 2$), although the method and results are readily extended to higher-dimensional spaces.

5.3.2.1 Optimal Location at an Intersection Point

Consider an intersection point given by

$$B = (a_{r1}, a_{s2})^T, \tag{5.3.54}$$

where r, s $\in$ {1, . . ., n} and r $\neq$ s. By definition (see (5.2.56)), B is not a destination. Without loss in generality, assume that $a_{i1} \neq a_{r1}$, $\forall\, i \neq r$, and $a_{i2} \neq a_{s2}$, $\forall\, i \neq s$. (Otherwise minor changes are required in the calculations below.) We suppose now that

$$x^* = B\ . \tag{5.3.55}$$

The iteration function $\phi_1(x)$ for the x_1-coordinate can be rewritten in the form,

$$\phi_1(x) = \frac{\dfrac{w_r|x_1 - a_{r1}|^{p-2}\cdot a_{r1}}{[\ell_p(x-a_r)]^{p-1}} + \sum\limits_{i\neq r}\dfrac{w_i|x_1-a_{i1}|^{p-2}\cdot a_{i1}}{[\ell_p(x-a_i)]^{p-1}}}{\dfrac{w_r|x_1-a_{r1}|^{p-2}}{[\ell_p(x-a_r)]^{p-1}} + \sum\limits_{i\neq r}\dfrac{w_i|x_1-a_{i1}|^{p-2}}{[\ell_p(x-a_i)]^{p-1}}}$$

$$= \frac{a_{r1} + \sigma_{r1}(x)H_{r1}(x)}{1 + \sigma_{r1}(x)h_{r1}(x)} \qquad (x_1 \neq a_{r1})\ , \tag{5.3.56}$$

where

$$\sigma_{jt}(x) := \frac{[\ell_p(x-a_j)]^{p-1}}{w_j|x_t - a_{jt}|^{p-2}} = \frac{1}{\gamma_{jt}(x)}\ , \tag{5.3.57}$$

$$H_{jt}(x) := \sum_{i\neq j}\frac{w_i|x_t-a_{it}|^{p-2}\cdot a_{it}}{[\ell_p(x-a_i)]^{p-1}}\ , \tag{5.3.58}$$

and

$$h_{jt}(x) := \sum_{i\neq j}\frac{w_i|x_t-a_{it}|^{p-2}}{[\ell_p(x-a_i)]^{p-1}}\ , \tag{5.3.59}$$

for all j and t.

We are interested in the behaviour of the iteration functions ϕ_t when x is 'close' to $x^* = B$. Noting that

$$\sigma_{r1}(x) = O(|x_1 - a_{r1}|^{2-p}) , \qquad x \in \Delta_B , \tag{5.3.60}$$

where Δ_B is a small δ-neighbourhood of B, we can express the right-hand side in (5.3.56) in terms of a power series for $p<2$, to obtain

$$\begin{aligned}\phi_1(x) &= [a_{r1} + \sigma_{r1}(x) H_{r1}(x)] \cdot [1 - \sigma_{r1}(x) h_{r1}(x) + \sigma^2_{r1}(x) h^2_{r1}(x) + O(\sigma^3_{r1}(x))] \\ &= a_{r1} + [H_{r1}(x) - h_{r1}(x) a_{r1}]\, \sigma_{r1}(x) - h_{r1}(x)[H_{r1}(x) - h_{r1}(x) a_{r1}]\, \sigma^2_{r1}(x) + O(\sigma^3_{r1}(x)).\end{aligned} \tag{5.3.61}$$

Let

$$g_r(x) := \frac{[\ell_p(x - a_r)]^{p-1}}{w_r} , \tag{5.3.62}$$

and

$$E_{rt}(x) := H_{rt}(x) - h_{rt}(x)\, a_{rt} , \ \forall\, t. \tag{5.3.63}$$

Then

$$\sigma_{r1}(x) = g_r(x) \cdot |x_1 - a_{r1}|^{2-p} , \tag{5.3.64}$$

and equation (5.3.61) becomes

$$\begin{aligned}\phi_1(x) = a_{r1} &+ E_{r1}(x)\, g_r(x) |x_1 - a_{r1}|^{2-p} - h_{r1}(x)\, E_{r1}(x)\, g^2_r(x) |x_1 - a_{r1}|^{4-2p} \\ &+ O(E_{r1}(x) |x_1 - a_{r1}|^{6-3p}).\end{aligned} \tag{5.3.65}$$

It is important to note that for $p > 1$,

$$\begin{aligned} E_{r1}(B) &= H_{r1}(B) - h_{r1}(B)\, a_{r1} \\ &= \sum_{i \neq r} \frac{w_i |a_{r1} - a_{i1}|^{p-2} \cdot a_{i1}}{[\ell_p(B - a_i)]^{p-1}} - \sum_{i \neq r} \frac{w_i |a_{r1} - a_{i1}|^{p-2} \cdot a_{r1}}{[\ell_p(B - a_i)]^{p-1}} \\ &= - \sum_{i \neq r} w_i \operatorname{sign}(a_{r1} - a_{i1}) \frac{|a_{r1} - a_{i1}|^{p-1}}{[\ell_p(B - a_i)]^{p-1}} \\ &= - \frac{\partial W(B)}{\partial x_1} \\ &= 0 , \end{aligned} \tag{5.3.66}$$

since $x^* = B$. For $x \in \Delta_B$, we apply the Mean Value Theorem and (5.3.66) to obtain

$$E_{r1}(x) = E_{r1}(B) + \nabla_1 E_{r1}(Q)\cdot(x_1 - a_{r1}) + \nabla_2 E_{r1}(Q)\cdot(x_2 - a_{s2})$$

$$= \nabla_1 E_{r1}(Q)\cdot(x_1 - a_{r1}) + \nabla_2 E_{r1}(Q)\cdot(x_2 - a_{s2}), \qquad (5.3.67)$$

where $Q \in \Delta_B$, and ∇_t denotes $\partial/\partial x_t$. Returning to (5.3.65), we see that

$$\phi_1(x) - a_{r1} = g_r(x)[\nabla_1 E_{r1}(Q)\cdot(x_1 - a_{r1}) + \nabla_2 E_{r1}(Q)\cdot(x_2 - a_{s2})]\cdot|x_1 - a_{r1}|^{2-p}$$

$$+ O(\|x - B\|\cdot|x_1 - a_{r1}|^{4-2p}). \qquad (5.3.68)$$

Hence,

$$\frac{\phi_1(x) - a_{r1}}{\|x - B\|} = O(|x_1 - a_{r1}|^{2-p}). \qquad (5.3.69)$$

Similarly, it can be shown that

$$\frac{\phi_2(x) - a_{s2}}{\|x - B\|} = O(|x_2 - a_{s2}|^{2-p}). \qquad (5.3.70)$$

Furthermore, if B is an intersection point in R^N (see (5.2.56)), where $N > 2$, the same procedure can be used as above to obtain a similar result as in (5.3.69) or (5.3.70) for each ϕ_t, $t = 1, \ldots, N$. This leads to the following intriguing fact.

Property 5.3.5

Suppose the optimal location x^* occurs at an intersection point for some value of p in the open interval (1, 2). Then the local convergence rate to x^* is superlinear.

Rectangular distances ($p = 1$) must be treated separately. This is due to the fact that when $p = 1$, $\partial W/\partial x_t$ is undefined on the hyperplanes $x_t - a_{it} = 0$, and not just at the fixed points; so that (5.3.66) no longer applies. It is well-known that an optimal solution in this case always occurs at an intersection point or a fixed point. Let us consider again $x^* = B$, an intersection point, and also let us assume that the optimal solution is unique. (Recall that this is not guaranteed for the non-collinear case if $p = 1$.) Using (5.3.65) and invoking the optimality criteria at B derived in Property 6.2.3 (set $L = 1$), we see that

$$\phi_1(x) - a_{r1} = E_{r1}(B)\, g_r(B)\, |x_1 - a_{r1}| + O(\|x - B\|^2), \tag{5.3.71}$$

where

$$0 \le |E_{ri}(B)\, g_r(B)| = \frac{1}{w_r} \left| \sum_{i \ne r} w_i \operatorname{sign}(a_{r1} - a_{i1}) \right| < 1. \tag{5.3.72}$$

The upper bound must be satisfied strictly since x^* is unique. A similar result applies for the other coordinates x_t. Furthermore, the first inequality in (5.3.72) is generally satisfied in a strict sense, so that we obtain the following result.

Property 5.3.6

Let $p = 1$, and suppose that x^* occurs uniquely at an intersection point,

$$B = (a_{r_1 1}, a_{r_2 2}, \ldots, a_{r_N N})^T .$$

(See the definition in (5.2.56).) For the special case where

$$\sum_{i \ne r_t} w_i \operatorname{sign}(a_{r_t t} - a_{it}) = 0\,, \quad t = 1, \ldots, N,$$

the local convergence rate to x^* is quadratic. Otherwise, the rate is linear.

Consider the possibility now that x^* occurs at a singular point which is not an intersection or a fixed point, for $1 < p < 2$. Say for example $x^*_k = a_{rk}$ for some $r \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, N\}$, but $x^*_t \ne a_{it}$ for all i and all $t \ne k$. From the preceding analysis it follows that the local convergence rate to x^* will be superlinear in the x_k direction. However, based on sub-section 5.3.1, it will only be linear in the subspace comprising the other directions. Hence, the overall rate of convergence is linear.

5.3.2.2 Optimal Location at a Destination

Consider the case where the optimal solution coincides with a fixed point; that is,

$$x^* = a_r \tag{5.3.73}$$

for some $r \in \{1, \ldots, n\}$. Our objective here is to analyze the local behaviour of the iteration functions ϕ_t when x is in a (small) δ-neighbourhood of a_r.

Letting Δ_r denote this neighbourhood, we first observe that for all t,

$$\sigma_{rt}(x) = \frac{1}{w_r}[\ell_p(x-a_r)]^{p-1} \cdot |x_t - a_{rt}|^{2-p}$$

$$\leq \frac{1}{w_r}[\ell_p(x-a_r)]^{p-1} \cdot [\ell_p(x-a_r)]^{2-p} \qquad (1 \leq p \leq 2)$$

$$= \frac{1}{w_r}\ell_p(x-a_r)$$

$$= O(\|x-a_r\|) . \tag{5.3.74}$$

Thus, $\phi_t(x)$ can be expressed as a power series using the same procedure as in sub-section 5.3.2.1, to obtain

$$\phi_t(x) = a_{rt} + E_{rt}(x)\,\sigma_{rt}(x) - h_{rt}(x)E_{rt}(x)\sigma_{rt}^2(x)$$
$$+ O(\sigma_{rt}^3(x)) \quad , x \in \Delta_r \quad . \tag{5.3.75}$$

Expressing $E_{rt}(x)$ in terms of its Taylor series at a_r, and using (5.3.74), we see that

$$\phi_t(x) = a_{rt} + E_{rt}(a_r)\,\sigma_{rt}(x) + O(\|x - a_r\|^2),\ x \in \Delta_r. \tag{5.3.76}$$

Furthermore,

$$E_{rt}(a_r) = \sum_{i \neq r} w_i \frac{|a_{rt}-a_{it}|^{p-2}}{[\ell_p(a_r-a_i)]^{p-1}} \cdot (a_{it}-a_{rt})$$

$$= -\sum_{i \neq r} w_i\, \text{sign}(a_{rt}-a_{it}) \frac{|a_{rt}-a_{it}|^{p-1}}{[\ell_p(a_r-a_i)]^{p-1}}$$

$$= -\frac{\partial W_r(a_r)}{\partial x_t} , \tag{5.3.77}$$

where $W_r(x)$ is defined in (5.2.79), and in general, $E_{rt}(a_r) \neq 0$.

Let x^s, $s = 0, 1, 2, \ldots$, be any regular sequence generated by the iteration procedure, which converges to a_r. Then an iteration number q exists such that $x^s \in \Delta_r$ for $s = q, q+1, \ldots$. Using (5.3.76), we obtain for $N = 2$,

$$x_1^{q+1} - a_{r1} = E_{r1}(a_r) \cdot \sigma_{r1}(x^q) + O(\|x^q - a_r\|^2) , \tag{5.3.78a}$$

and

$$x_2^{q+1} - a_{r2} = E_{r2}(a_r) \cdot \sigma_{r2}(x^q) + O(\|x^q - a_r\|^2) . \tag{5.3.78b}$$

Consider the case where

$$- \nabla W_r(a_r) = (E_{r1}(a_r), E_{r2}(a_r))^T \neq 0 , \tag{5.3.79}$$

and without loss in generality, assume that $E_{r1}(a_r) \neq 0$. (Note that if $\nabla W_r(a_r) = 0$, the local convergence rate will be quadratic.) Then dividing (5.3.78b) by (5.3.78a) gives

$$\begin{aligned} \frac{x_2^{q+1} - a_{r2}}{x_1^{q+1} - a_{r1}} &= \frac{E_{r2}(a_r) \cdot \sigma_{r2}(x^q)}{E_{r1}(a_r) \cdot \sigma_{r1}(x^q)} + O(\|x^q - a_r\|) \\ &= \frac{E_{r2}(a_r)}{E_{r1}(a_r)} \cdot \frac{|x_2^q - a_{r2}|^{2-p}}{|x_1^q - a_{r1}|^{2-p}} + O(\|x^q - a_r\|) . \end{aligned} \tag{5.3.80}$$

Let

$$b := E_{r2}(a_r)/E_{r1}(a_r) , \tag{5.3.81}$$

and

$$\tan \theta_s := \frac{x_2^s - a_{r2}}{x_1^s - a_{r1}} , \qquad s = q, q+1, \ldots . \tag{5.3.82}$$

Then equation (5.3.80) can be rewritten as

$$\tan \theta_{q+1} = b |\tan \theta_q|^{2-p} + O(\|x^q - a_r\|) . \tag{5.3.83}$$

If $b = 0$ ($E_{r2}(a_r) = 0$), then $\tan \theta_{q+1} = O(\|x^q - a_r\|)$, which can be made arbitrarily small by increasing q. We conclude that $\tan \theta_{q+1}$ approaches zero asymptotically in this case. On the other hand, if $b \neq 0$, a recursive argument can be used as follows:

$$\tan\theta_{q+2} = b\,|\tan\theta_{q+1}|^{2-p} + O(\|x^{q+1} - a_r\|)$$

$$= b\,|b\,|\tan\theta_q|^{2-p} + O(\|x^q - a_r\|)|^{2-p} + O(\|x^{q+1} - a_r\|)$$

$$= b \cdot |b|^{2-p} \cdot |\tan\theta_q|^{(2-p)^2} + O(\delta)\;; \tag{5.3.84a}$$

and proceeding in this manner,

$$\tan\theta_{q+m} = b \cdot |b|^{2-p} \cdot |b|^{(2-p)^2} \cdot \ldots \cdot |b|^{(2-p)^{m-1}} \cdot |\tan\theta_q|^{(2-p)^m} + O(\delta)\,. \tag{5.3.84b}$$

Noting that the geometric series,

$$1 + (2-p) + (2-p)^2 + \ldots + (2-p)^{m-1} = \frac{1-(2-p)^m}{p-1}\;,\; p \neq 1\;, \tag{5.3.85}$$

we rewrite (5.3.84b) as

$$\tan\theta_{q+m} = \mathrm{sign}\,(b) \cdot |b|^{\lambda} \cdot |\tan\theta_q|^{(2-p)^m} + O(\delta)\;, \tag{5.3.86}$$

where

$$\lambda = \frac{1-(2-p)^m}{p-1}\;,\; \text{and}\;\; 1 < p \leq 2\;.$$

But δ can be made arbitrarily small by increasing q, and $(2-p)^m$ can be made arbitrarily small by increasing m. Hence, we conclude that

$$\lim_{s\to\infty} \tan\theta_s = \mathrm{sign}\,(b) \cdot |b|^{\frac{1}{p-1}}\;,\; 1 < p \leq 2\;. \tag{5.3.87}$$

If $p = 1$, equation (5.3.84b) becomes

$$\tan\theta_{q+m} = \mathrm{sign}\,(b) \cdot |b|^{m} \cdot |\tan\theta_q| + O(\delta)\;,$$

so that

$$\lim_{s\to\infty} \tan\theta_s = \begin{cases} 0\;, & \text{if } |b| < 1\;, \\ +\infty\;, & \text{if } b > 1\;, \qquad (p = 1)\;. \\ -\infty\;, & \text{if } b < -1\;, \end{cases} \tag{5.3.88}$$

For the special case where $b = +1$ or -1, it follows that $\tan\theta_s$ does not have a unique asymptotic value when $p = 1$.

We see from the preceding analysis that the series $\{x^s\}$ converges to a_r along an asymptotic "direction of approach" which depends only on b. Also note from equations (5.3.78a) and (5.3.78b) that the quadrant of the approach angle is uniquely determined by the signs of $E_{r1}(a_r)$ and $E_{r2}(a_r)$. For Euclidean distances, equation (5.3.87) becomes

$$\lim_{s\to\infty} \tan\theta_s = \text{sign}(b)\cdot|b|$$

$$= b = E_{r2}(a_r)/E_{r1}(a_r)\ , \qquad p=2\ ,$$

which is the same result obtained by Katz (1974).

We summarize the results obtained above by the following.

Property 5.3.7

Let p take a value in the range $1 < p \le 2$, and let $\{x^s\}$ be a regular sequence which converges to the fixed point a_r. Then $\{x^s\}$ converges to a_r along an asymptotic direction of approach uniquely defined by θ_*, where

$$\tan\theta_* = \text{sign}\left(\frac{E_{r2}(a_r)}{E_{r1}(a_r)}\right)\cdot\left|\frac{E_{r2}(a_r)}{E_{r1}(a_r)}\right|^{\frac{1}{p-1}}\ , \tag{5.3.89}$$

and θ_* is located in the quadrant defined by the signs of $E_{r1}(a_r)$ and $E_{r2}(a_r)$.

For $p = 1$, the direction of approach is along the x_1-axis if $|E_{r1}(a_r)| > |E_{r2}(a_r)|$, along the x_2-axis if $|E_{r1}(a_r)| < |E_{r2}(a_r)|$, and along an indeterminate θ if $|E_{r1}(a_r)| = |E_{r2}(a_r)|$. Again, the quadrant is uniquely defined by the signs of $E_{r1}(a_r)$ and $E_{r2}(a_r)$.

Alternatively, the direction of approach can be specified by the unit vector V, defined as follows:

$$V = \left(\text{sign}(E_{r1}(a_r))\cdot\frac{|E_{r1}(a_r)|^{\frac{1}{p-1}}}{D}\ ,\ \text{sign}(E_{r2}(a_r))\cdot\frac{|E_{r2}(a_r)|^{\frac{1}{p-1}}}{D}\right)^T\ , \tag{5.3.90}$$

where

$$D = [|E_{r1}(a_r)|^{\frac{2}{p-1}} + |E_{r2}(a_r)|^{\frac{2}{p-1}}]^{1/2} . \tag{5.3.91}$$

Equations (5.3.90) and (5.3.91) are generalized for higher-dimensional spaces in a straight-forward manner.

Consider the following problem,

$$\max_y \left\{ f(y) = \frac{\varepsilon_r \cdot y}{w_r \ell_p(y)} \right\} , \tag{5.3.92}$$

where

$$\varepsilon_r = (E_{r1}(a_r), E_{r2}(a_r))^T = - \nabla W_r(a_r) , \tag{5.3.93}$$

and y is a unit vector. It is a well-known fact that

$$\max_y f(y) = \frac{1}{w_r} \ell_q(\varepsilon_r) , \tag{5.3.94}$$

where $q = p/(p-1)$ and $\ell_q(\cdot)$ is the polar of the ℓ_p norm (e.g., see Juel and Love, 1981). But

$$f(V) = \frac{\left[|E_{r1}(a_r)| \cdot \dfrac{|E_{r1}(a_r)|^{\frac{1}{p-1}}}{D} + |E_{r2}(a_r)| \cdot \dfrac{|E_{r2}(a_r)|^{\frac{1}{p-1}}}{D} \right]}{\dfrac{w_r}{D} \cdot \left[|E_{r1}(a_r)|^{p/(p-1)} + |E_{r2}(a_r)|^{p/(p-1)} \right]^{1/p}}$$

$$= \frac{1}{w_r} \left[|E_{r1}(a_r)|^{p/(p-1)} + |E_{r2}(a_r)|^{p/(p-1)} \right]^{1-\frac{1}{p}}$$

$$= \frac{1}{w_r} [|E_{r1}(a_r)|^{q} + |E_{r2}(a_r)|^{q}]^{1/q}$$

$$= \frac{1}{w_r} \ell_q(\varepsilon_r) . \tag{5.3.95}$$

Comparing (5.3.94) and (5.3.95), we obtain the following interesting result.

Property 5.3.8

The asymptotic direction of approach V at the fixed point a_r maximizes the function $f(y)$; that is,

$$f(V) = \max_y \left\{ f(y) = \frac{\varepsilon_r \cdot y}{w_r \, \ell_p(y)} \right\} . \tag{5.3.96}$$

Note that $\varepsilon_r \cdot y$ gives the descent rate in the direction y due to the component W_r of the objective function W, while $w_r \, \ell_p(y)$ gives the ascent rate of the component $w_r \, \ell_p(x - a_r)$. Hence, we see that V is the direction which maximizes the descent rate of W_r relative to the ascent rate of $w_r \, \ell_p(x - a_r)$ at a_r; or loosely-speaking, V can be regarded as a direction of minimum ascent (or maximum descent) of the objective function W at a_r.

Juel and Love (1981) prove that a necessary and sufficient condition for a_r to be an optimal solution is that

$$\frac{1}{w_r} \ell_q(\varepsilon_r) \le 1 , \qquad x^* = a_r . \tag{5.3.97}$$

On the other hand, if a_r is not optimal, it follows that

$$\frac{1}{w_r} \ell_q(\varepsilon_r) > 1 , \qquad x^* \neq a_r . \tag{5.3.98}$$

Consider a regular sequence $\{x^s\}$ which converges to a_r. Since V is the asymptotic direction of approach for this sequence, then

$$\lim_{s \to \infty} \left\{ \frac{x^s - a_r}{\|x^s - a_r\|} \right\} = V . \tag{5.3.99}$$

Let $V = (v_1, v_2)^T$, where the components v_1 and v_2 are defined in (5.3.90). Neglecting higher-order terms in the limit as $s \to \infty$, it follows from (5.3.99) that

$$\begin{aligned} x_t^{s+1} - a_{rt} &= v_t \cdot \|x^{s+1} - a_r\| \\ &= (x_t^s - a_{rt}) \cdot \frac{\|x^{s+1} - a_r\|}{\|x^s - a_r\|} , \ \forall t, \end{aligned} \tag{5.3.100}$$

and so, (5.3.78a) can be rewritten in the form,

$$(x_i^s - a_{ri}) \cdot \frac{\|x^{s+1} - a_r\|}{\|x^s - a_r\|} = E_{r1}(a_r) \cdot \frac{1}{w_r} [\ell_p(x^s - a_r)]^{p-1} \cdot |x_1^s - a_{r1}|^{2-p}$$

$$= E_{r1}(a_r) \cdot \frac{[\ell_p(x^s - a_r)]^{p-1} \cdot (x_1^s - a_{r1})}{w_r \operatorname{sign}(x_1^s - a_{r1}) |x_1^s - a_{r1}|^{p-1}}$$

$$= E_{r1}(a_r) \cdot v_1 \cdot \left(\frac{[\ell_p(x^s - a_r)]^{p-1} \|x^s - a_r\|}{w_r \operatorname{sign}(x_1^s - a_{r1}) |x_1^s - a_{r1}|^{p-1}} \right).$$

Hence,

$$\lim_{s \to \infty} \left\{ \frac{\|x^{s+1} - a_r\|}{\|x^s - a_r\|^2} \cdot \frac{w_r |x_1^s - a_{r1}|^p}{[\ell_p(x^s - a_r)]^{p-1}} \right\} = E_{r1}(a_r) \cdot v_1 . \tag{5.3.101a}$$

Similarly, we obtain

$$\lim_{s \to \infty} \left\{ \frac{\|x^{s+1} - a_r\|}{\|x^s - a_r\|^2} \cdot \frac{w_r |x_2^s - a_{r2}|^p}{[\ell_p(x^s - a_r)]^{p-1}} \right\} = E_{r2}(a_r) \cdot v_2 . \tag{5.3.101b}$$

Adding equations (5.3.101a) and (5.3.101b) gives

$$\lim_{s \to \infty} \left\{ \frac{\|x^{s+1} - a_r\|}{\|x^s - a_r\|^2} \cdot w_r \, \ell_p(x^s - a_r) \right\} = \varepsilon_r \cdot V .$$

But

$$\ell_p(x^s - a_r) \to \|x^s - a_r\| \, \ell_p(V) , \tag{5.3.102}$$

from (5.3.99), and we finally conclude that

$$\lim_{s \to \infty} \frac{\|x^{s+1} - a_r\|}{\|x^s - a_r\|} = \frac{\varepsilon_r \cdot V}{w_r \, \ell_p(V)} = \frac{1}{w_r} \, \ell_q(\varepsilon_r) , \tag{5.3.103}$$

where the second equality is obtained from (5.3.95). This leads to the following important result.

Theorem 5.3.3

Let p take a value in the range $1 \le p \le 2$, and let $x^* = a_r$, $r \in \{1, \ldots, n\}$, be the unique optimal solution. Then the local convergence rate to a_r is linear with asymptotic convergence factor $\mu = \ell_q(\varepsilon_r)/w_r$, except for two special cases. If $\ell_q(\varepsilon_r) = 0$ ($\nabla W_r(a_r) = 0$), the rate is quadratic; if $\ell_q(\varepsilon_r) = w_r$, the rate is sublinear.

Proof:

If $0 < \ell_q(\varepsilon_r)/w_r < 1$, we see from equation (5.3.103) that convergence is linear with asymptotic convergence factor $\mu = \ell_q(\varepsilon_r)/w_r$. If $\ell_q(\varepsilon_r) = 0$, then $\varepsilon_r = 0$, and it follows from (5.3.76) that the local convergence rate is quadratic. If $\ell_q(\varepsilon_r) = w_r$, then from equation (5.3.103) and the fact that global convergence is assured by Theorem 5.2.1, we conclude that the convergence must be sublinear. From (5.3.97), we see that all possibilities have been considered.

Theorem 5.3.3 generalizes a result obtained by Katz (1974, Theorem 2) for Euclidean distances, to the case where p can have any value in the closed interval [1,2]. Although a similar result is obtained for the generalized problem, the analysis turns out to be considerably more complex when values of p other than 2 are used. Also note that the analysis can be extended to higher-dimensional location spaces ($N > 2$) in a straightforward manner, so that Theorem 5.3.3 holds for $N \geq 2$.

If a_r is not an optimal solution, then the asymptotic convergence factor, $\mu > 1$, by (5.3.98). Thus, a regular sequence will never converge to a non-optimal fixed point. This provides an alternate proof to the one given in Property 5.2.9 (for case iii). It is also interesting to note that if an iterate x^q lands in a sufficiently small δ-neighbourhood of a_r, a non-optimal fixed point, then the sequence will linger near a_r for several iterations, but will ultimately move away along a direction of departure tending to θ_*.

5.3.3 Non-Singular Optimum in N-Dimensional Space

Thus far, we have concentrated on the local convergence properties of our iterative solution procedure in R^2. When x^* was an intersection point or a fixed point, the convergence rates derived in subsection 5.3.2 (Properties 5.3.5 and 5.3.6, and Theorem 5.3.3) were seen to apply readily to higher-dimensional spaces as well. However, the local convergence rate when x^* is a non-singular point in R^N still remains an open question. Our objective then is to generalize Theorem 5.3.1 to the minisum problem in R^N. The concept of an asymptotic direction of approach, introduced when x^* was a fixed point, will be useful here.

We begin by defining a diagonal matrix,

$$W = \begin{bmatrix} (s_1(x^*))^{1/2} & 0 & 0 & \cdots & 0 \\ 0 & (s_2(x^*))^{1/2} & 0 & \cdots & 0 \\ 0 & 0 & (s_3(x^*))^{1/2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & (s_N(x^*))^{1/2} \end{bmatrix} . \tag{5.3.104}$$

Then, recalling the form of $\phi'(x^*)$ in (5.3.37), it is readily seen that

$$A = W\,[I - \phi'(x^*)]\,W^{-1} \tag{5.3.105}$$

is symmetric and positive definite, provided that $p > 1$ and the a_i are non-collinear. This leads to the following important result.

Property 5.3.9

Consider the $N \times N$ matrix $\phi'(x^*)$, and let $p > 1$ and the a_i be non-collinear. Then (a) the eigenvalues of $\phi'(x^*)$ are real, (b) the algebraically largest eigenvalue of $\phi'(x^*)$ is less than one, and (c) the set of eigenvectors for $\phi'(x^*)$ forms a basis in R^N.

Proof:

This property is a direct application of Theorem 2-2.1 in Hageman and Young (1981). However, since they do not provide a proof of their theorem, we give one here for completeness.

Denote the eigenvalues of A by μ_t, $t = 1, \ldots, N$, and those of $\phi'(x^*)$ by λ_t, $t = 1, \ldots, N$. Since A is a symmetric and positive definite matrix, therefore

$$\mu_t > 0, \qquad t = 1, \ldots, N. \tag{5.3.106}$$

But from (5.3.105), A and $[I - \phi'(x^*)]$ are similar matrices, and thus have identical eigenvalues. It immediately follows that the λ_t are real, and

$$1 - \lambda_t = \mu_t, \qquad t = 1, \ldots, N. \tag{5.3.107}$$

Combining (5.3.106) and (5.3.107), we see that

$$\lambda_t < 1, \qquad t = 1, \ldots, N. \tag{5.3.108}$$

Now let the columns of the $N \times N$ matrix D denote the eigenvectors of A. Then the eigenvectors of $\phi'(x^*)$ are given by the columns of $W^{-1}D$. Since the eigenvectors in D include a basis for R^N, it is immediately clear that those in $W^{-1}D$ also form a basis for R^N, thereby ending the proof.

Theorem 5.3.4 *{Local Convergence in R^N}*

Let p take on a value in the range $1 < p \leq 2$, the a_i be non-collinear, and the optimal solution occur at a non-singular point x^*. Then the asymptotic convergence rate to x^* is linear.

Proof:

We prove this result in a heuristic fashion. Consider a perturbed problem in which an additional fixed point a_{n+1} is placed within a (small) δ-neighbourhood of x^*. Let us adjust the weight w_{n+1} to ensure that a_{n+1} ($\neq x^*$) now gives the optimal location, while perturbing the original problem as little as possible; i.e.,

$$\frac{1}{w_{n+1}} \ell_q(\varepsilon_{n+1}) = 1 , \tag{5.3.109}$$

where

$$\varepsilon_{n+1} = -\nabla W(a_{n+1}),$$

$$q = p/(p-1),$$

and W is the objective function of the original problem (see (5.3.97)). Then, from the results in subsection 5.3.2 (Property 5.3.7) and global convergence (Theorem 5.2.1), it follows that any regular sequence $\{x^s\}$ will converge to a_{n+1} along a unique asymptotic direction of approach.

Now consider a series of perturbed problems, such that $a_{n+1} \rightarrow x^*$ and w_{n+1} is adjusted according to (5.3.109). Each problem in this series similarly has a unique asymptotic direction of approach to a_{n+1}. Also note that $\varepsilon_{n+1} \rightarrow 0$ since $\nabla W(x^*) = 0$. Thus, $w_{n+1} \rightarrow 0$, and the series converges to the original problem. We conclude therefore that the sequence $\{x^s\}$ converges to x^* in the original problem along a single asymptotic direction. But this is possible if, and only if, the dominant eigenvalue of $\phi'(x^*)$ is positive; i.e., the spectral radius,

$$\rho = \max_{1 \le t \le N} |\lambda_t| \tag{5.3.110}$$

is associated with a positive eigenvalue. (Otherwise there would be two directions of approach.) Using (5.3.108), it follows that

$$\rho < 1, \tag{5.3.111}$$

and hence, we conclude that the asymptotic convergence rate is linear.

CHAPTER 6

THE MIXED-NORM MODEL

The mixed-norm problem was introduced in paragraph (b) of Chapter 1, where we noted that such a model should be considered when flows using different transportation modes are associated with individual customers. In this chapter, a specific form of the mixed-norm problem is studied, in which distances travelled by the various transportation modes are adequately approximated by different ℓ_p norms. The resulting model can be formulated as follows:

$$\text{minimize } W_m(x) = \sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij}\, \ell_{p_j}(x - a_i), \tag{6.1}$$

where

$p_j \geq 1, j = 1, \ldots, L$, are distinct values of the parameter p in the ℓ_p distance function;

$w_{ij} > 0$, $j = 1, \ldots, L$, $i = 1, \ldots, n$, are weighting constants which convert distance travelled between the new facility (x) and destination a_i using transportation mode j into a cost; and

$L \geq 2$.

The restriction above that none of the w_{ij} can be zero is not limiting in a practical sense (since the w_{ij} can be made arbitrarily small), and has the advantage of simplifying the notation in the subsequent analysis.

We proceed next to solve model (6.1), by developing a simple extension of the Weiszfeld procedure for a single ℓ_p norm (Chapter 5). Global convergence of our iterative solution algorithm is proven when the p_j all fall within a certain range of values. Local convergence properties of the algorithm are also investigated.

In Section 6.2, criteria are derived which verify the optimality of the fixed points in model (6.1). We also propose the use of intersection point optimality criteria when one of the p_j's equals unity, and compute these for our problem. A typical application is examined having $L=2$, $p_1=1$ and $p_2=2$. This particular model could be used to represent material handling costs, for example on a shop floor, when flow from a source moves partly along rectangular aisles, while the remainder travels along conveyance equipment linking each destination to the source by a straight path.

6.1 Solution by an Extended Weiszfeld Procedure

The Weiszfeld procedure in Chapter 5 for the standard minisum problem with a single ℓ_p norm is readily extended to our mixed-norm model. As in (5.2.12), we begin by supposing that an optimal solution occurs at a differentiable point, $x^* = (x^*_1, \ldots, x^*_N)^T$, so that the following set of first-order necessary conditions must be satisfied:

$$\frac{\partial}{\partial x_t} W_m(x^*) = 0 \,, \qquad t = 1, \ldots, N. \tag{6.1.1}$$

Since W_m is a positive linear combination of norms, it is a convex function of x. Hence, the above system of equations also gives a sufficient condition for x^* to be an optimal location. Evaluating the partial derivatives of W_m at x^*, we get

$$\sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij} \operatorname{sign}(x^*_t - a_{it}) \frac{|x^*_t - a_{it}|^{p_j - 1}}{[\ell_{p_j}(x^* - a_i)]^{p_j - 1}} = 0 \,, \qquad t = 1, \ldots, N. \tag{6.1.2}$$

Following the same substitution as in (5.2.14) and (5.2.15), the above equations are then re-arranged in analogous form as (5.2.16) to obtain:

$$x_t^* = \frac{\sum_{i=1}^{n} a_{it} \sum_{j=1}^{L} w_{ij} \frac{|x_t^* - a_{it}|^{p_j - 2}}{[\ell_{p_j}(x^* - a_i)]^{p_j - 1}}}{\sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij} \frac{|x_t^* - a_{it}|^{p_j - 2}}{[\ell_{p_j}(x^* - a_i)]^{p_j - 1}}}, \qquad t = 1, \ldots, N. \tag{6.1.3}$$

Thus, replacing x^* on the right-hand side by x^q and on the left-hand side by x^{q+1}, where $q = 0, 1, 2, \ldots$, denotes the iteration number, we have the one-point iterative method given by

$$x^{q+1} = \phi(x^q), \tag{6.1.4}$$

where

$$\phi(x) = (\phi_1(x), \ldots, \phi_N(x))^T, \tag{6.1.5}$$

$$\phi_t(x) = \frac{\sum_{i=1}^{n} \beta_{it}(x) a_{it}}{\sum_{i=1}^{n} \beta_{it}(x)}, \qquad t = 1, \ldots, N, \tag{6.1.6}$$

and

$$\beta_{it}(x) = \sum_{j=1}^{L} w_{ij} \frac{|x_t - a_{it}|^{p_j - 2}}{[\ell_{p_j}(x - a_i)]^{p_j - 1}}, \qquad i = 1, \ldots, n, \quad t = 1, \ldots, N. \tag{6.1.7}$$

Note that if $L = 1$ (i.e., a single ℓ_p norm), the iteration method reduces to the Weiszfeld procedure in (5.2.17) with $w_{i1} = w_i$, $p_1 = p$, and $\beta_{it}(x) \equiv \gamma_{it}(x)$, for all i, t.

Letting

$$\lambda_{it}(x) = \beta_{it}(x) / \sum_{k=1}^{n} \beta_{kt}(x), \qquad \forall\ i, t, \tag{6.1.8}$$

the iterative scheme can be rewritten in the compact form,

$$x_t^{q+1} = \sum_{i=1}^{n} \lambda_{it}(x^q) a_{it}, \qquad t = 1, \ldots, N, \tag{6.1.9}$$

where

$$\lambda_{it}(x) \geq 0, \qquad \forall\ i, t, \tag{6.1.10}$$

and

$$\sum_{i=1}^{n} \lambda_{it}(x) = 1 , \qquad \forall \ t . \tag{6.1.11}$$

We see that this is precisely the same form as in (5.2.23), and thus, x_t^{q+1} is once again a convex combination of a_{it}, $i = 1, \ldots, n$, for each $t \in \{1, \ldots, N\}$. Hence, the same conclusion is reached; namely, that all the iterates x^{q+1} fall in a bounded hypercube containing the fixed points. More precisely, we have

$$\min_i \{a_{it}\} \leq x_t^{q+1} \leq \max_i \{a_{it}\} , \tag{6.1.12}$$

for $t = 1, \ldots, N$ and $q = 0, 1, 2, \ldots$.

Let us proceed now with an analysis of our extended Weiszfeld procedure for the mixed-norm problem. The main objective here will be to determine the conditions which guarantee global convergence of the algorithm to an optimal solution. Several of the results which follow are a straightforward extension of those given in Section 5.2, for $L = 1$. Complete proofs for these results are therefore omitted.

Property 6.1.1

If $p_r < 2$ for some $r \in \{1, \ldots, L\}$, the iteration function $\phi_t(x)$ is undefined along the hyperplanes,

$$x_t - a_{it} = 0 , \qquad i = 1, \ldots, n,$$

for $t = 1, \ldots, N$. Whereas if $p_j \geq 2$ for all j, then ϕ_t is undefined only at the fixed points a_i, $i = 1, \ldots, n$, for $t = 1, \ldots, N$.

Proof:

Follows immediately from Property 5.2.1, and the fact that $w_{ij} > 0$ for all i and j.

The iteration functions ϕ_t are well-behaved and continuous everywhere, except at the singular points noted above, where they are undefined. The next three results deal with the local behaviour of the ϕ_t at their singular points.

Property 6.1.2

Consider the case where $p_r < 2$ for some $r \in \{1, \ldots, L\}$. Then for any $t \in \{1, \ldots, N\}$, ϕ_t is a continuous function of x if, and only if, we set $\phi_t(x) = a_{it}$ at all points x on the hyperplane $x_t - a_{it} = 0$, for $i = 1, \ldots, n$.

Proof:

Since $L \geq 2$, at least one of the p_j's < 2. The remainder of the proof is a straightforward extension of Property 5.2.2.

Property 6.1.3

Consider the case where $p_j \geq 2$, $j = 1, \ldots, L$, with $p_r = 2$ for some $r \in \{1, \ldots, L\}$. Then for any $t \in \{1, \ldots, N\}$, ϕ_t is a continuous function of x if, and only if, we set $\phi_t(a_i) = a_{it}$, for $i = 1, \ldots, n$.

Proof:

A straightforward extension of Property 5.2.3.

Property 6.1.4

Consider the final possibility where $p_j > 2$, for all j. Then ϕ_t, $t = 1, \ldots, N$, cannot in general be made continuous at the singular points a_i, $i = 1, \ldots, n$.

Proof:

A straightforward extension of Property 5.2.4.

Based on the preceding properties, we define the following iteration scheme:

i) If $p_r < 2$ for one or more $r \in \{1, \ldots, L\}$,

$$x_t^{q+1} = \begin{cases} \phi_t(x^q) & \text{if} \quad x_t^q - a_{it} \neq 0, \quad i = 1, \ldots, n, \\ a_{st} & \text{if} \quad x_t^q - a_{st} = 0, \quad s \in \{1, \ldots, n\}, \end{cases} \tag{6.1.13}$$

$$t = 1, \ldots, N, \qquad q = 0, 1, 2, \ldots \ .$$

ii) If $p_j \geq 2, \ \forall \ j$,

$$x^{q+1} = \begin{cases} \phi(x^q) & \text{if} \quad x^q \neq a_i, \quad i = 1, \ldots, n, \\ a_s & \text{if} \quad x^q = a_s, \quad s \in \{1, \ldots, n\}, \end{cases} \tag{6.1.14}$$

$$q = 0, 1, 2, \ldots \ .$$

Let us denote the procedure given in (6.1.13) or (6.1.14) symbolically as

$$T : x \rightarrow T(x) \qquad (x \in R^N). \tag{6.1.15}$$

Then T is a continuous mapping if $p_r \leq 2$ for some $r \in \{1, \ldots, L\}$ by Properties 6.1.2 and 6.1.3, and discontinuous in general at the fixed points a_i otherwise, by Property 6.1.4.

Property 6.1.5

The map, $T : x \rightarrow T(x)$, lies in a compact set.

Proof:

Using an identical reasoning as in Property 5.2.5, we conclude that all the iterates except possibly the starting point fall in a bounded hypercube, such that (6.1.12) is satisfied.

Adopting the same notation as in (5.2.40) and (5.2.41), the set of singular points of the vector function ϕ is given by

$$S = \begin{cases} \bigcup_{t=1}^{N} H_t & \text{if } p_r < 2 \text{ for one or more } r \in \{1, \ldots, L\}, \\ \{a_1, \ldots, a_n\} & \text{otherwise} . \end{cases} \tag{6.1.16}$$

Then Definition 5.2.1 can be used to distinguish between the two basic types of sequences, regular and non-regular.

Property 6.1.6

Let x^* denote an optimal solution of model (6.1). If $x^q = x^*$ then $x^{q+1} = x^*$, as well as all subsequent iterations. If $x^q \notin S$ and $x^{q+1} = x^q$, then $x^q = x^*$.

Proof:

A direct extension of Property 5.2.6.

Property 6.1.7

Each iterate moves in a descent direction of W_m, provided $x^{q+1} \neq x^q$, $q = 0, 1, 2, \ldots$.

Proof:

Equation (6.1.4) can be rewritten in the form,

$$x_t^{q+1} = \phi_t(x^q) = x_t^q - \frac{\sum_{i=1}^{n} \beta_{it}(x^q) \cdot (x_t^q - a_{it})}{\sum_{i=1}^{n} \beta_{it}(x^q)} , \qquad t = 1, \ldots, N .$$

But

$$\beta_{it}(x)\cdot(x_t - a_{it}) = \sum_{j=1}^{L} w_{ij}\,\mathrm{sign}(x_t - a_{it})\,\frac{|x_t - a_{it}|^{p_j - 1}}{[\ell_{p_j}(x - a_i)]^{p_j - 1}}$$

$$= \sum_{j=1}^{L} w_{ij}\,\nabla_t\,\ell_{p_j}(x - a_i), \qquad \forall\ i, t.$$

Thus, letting

$$s_t(x) = \sum_{i=1}^{n} \beta_{it}(x)\ , \qquad t = 1,\ \ldots,\ N, \tag{6.1.17}$$

we obtain

$$x_t^{q+1} = x_t^q - \frac{1}{s_t(x^q)} \sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij}\,\nabla_t\,\ell_{p_j}(x^q - a_i)$$

$$= x_t^q - \frac{1}{s_t(x^q)}\,\nabla_t\,W_m(x^q)\ , \qquad t = 1,\ \ldots,\ N. \tag{6.1.18}$$

This assumes, of course, that x^q is not a singular point of any of the ϕ_t.

Since $x^{q+1} \neq x^q$, it follows that $x^q \neq a_i$, $i = 1, \ldots, n$. Thus, $x^q \in S$ only if $p_r < 2$ for some $r \in \{1,\ldots,L\}$ and $x_t^q = a_{i_t t}$ for all $t \in J_q$, where J_q is a non-empty subset of $\{1,\ldots,N\}$. Clearly, the complement $J'_q = \{1,\ldots,N\} - J_q$ is a non-empty set, for otherwise, $x^{q+1} = x^q$. If $t \in J_q$, then $s_t(x^q) \to +\infty$; and if $t \in J'_q$, then $s_t(x^q)$ is positive and finite-valued. We see then that

$$s_t(x^q) > 0\ , \qquad t = 1,\ \ldots,\ N, \qquad q = 0, 1, 2, \ldots\ . \tag{6.1.19}$$

Furthermore, $s_t(x^q)$ is finite-valued for all t if $p_j \geq 2$, $j = 1, \ldots, L$, and finite-valued for at least one t if $p_r < 2$ for some $r \in \{1,\ldots,L\}$. Hence, we conclude that the iterates follow descent directions of W_m.

The iteration procedure can be rewritten in the same form as (5.2.48), where once again, the modification matrix $[M(x^q)]^{-1}$ is diagonal with non-negative diagonal elements $(1/s_t(x^q))$. We see now that the iterates move in descent directions with pre-determined step-size, just as in the single-norm case (L = 1). If the step-size is too large, over-shooting will occur. The next result gives a sufficient condition (analogous to Property 5.2.8), to guarantee that this never occurs.

Property 6.1.8 *{Descent Property}*

If $1 \le p_j \le 2$, $j = 1, \ldots, L$, and $x^{q+1} \neq x^q$, then $W(x^{q+1}) < W(x^q)$.

Proof:

Since $L \ge 2$ and the p_j's take on distinct values, it is clear that at least one $p_j < 2$. Hence, by (6.1.16),

$$S = \bigcup_{t=1}^{N} H_t . \qquad (6.1.20)$$

Let $V_q = \{t \mid \phi_t(x^q) \text{ is non-singular}\}$. Since $x^{q+1} \neq x^q$, V_q is a non-empty set.

For $t \in V_q$, define

$$g_t(x_t) = \sum_{i=1}^{n} \beta_{it}^{q}(x_t - a_{it})^2 , \qquad (6.1.21)$$

where $\beta_{it}^{q} := \beta_{it}(x^q)$, $\forall$ i, t, and a given x^q. For $s \in V'_q$ (the complement of V_q), define

$$h_s(x_s) = \sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij} |x_s - a_{is}|^{p_j} [\ell_{p_j}(x^q - a_i)]^{1-p_j} . \qquad (6.1.22)$$

Then using similar steps as in Property 5.2.8, we obtain

$$\sum_{t \in V_q} g_t(x_t^{q+1}) + \sum_{s \in V'_q} h_s(x_s^{q+1}) < W_m(x^q) . \qquad (6.1.23)$$

Consider the left-hand side of the above relation:

$$\sum_{t \in V_q} g_t(x_t^{q+1}) + \sum_{s \in V'_q} h_s(x_s^{q+1})$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{L} \sum_{t \in V_q} w_{ij} |x_t^q - a_{it}|^{p_j - 2} [\ell_{p_j}(x^q - a_i)]^{1-p_j} (x_t^{q+1} - a_{it})^2$$

$$+ \sum_{i=1}^{n} \sum_{j=1}^{L} \sum_{s \in V'_q} w_{ij} |x_s^{q+1} - a_{is}|^{p_j} [\ell_{p_j}(x^q - a_i)]^{1-p_j}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{L} \sum_{\substack{k=1 \\ |x_k^q - a_{ik}| \neq 0}}^{N} w_{ij} [\ell_{p_j}(x^q - a_i)]^{1-p_j} [|x_k^q - a_{ik}|^{p_j}]^{\frac{p_j - 2}{p_j}} [|x_k^{q+1} - a_{ik}|^{p_j}]^{\frac{2}{p_j}}$$

$$(x_s^{q+1} = x_s^q, \ \forall \ s \in V'_q)$$

$$\geq \sum_{i=1}^{n} \sum_{j=1}^{L} \sum_{\substack{k=1 \\ |x_k^q - a_{ik}| \neq 0}}^{N} w_{ij} [\ell_{p_j}(x^q - a_i)]^{1-p_j} \left\{ \left(\frac{p_j - 2}{p_j} \right) |x_k^q - a_{ik}|^{p_j} + \frac{2}{p_j} |x_k^{q+1} - a_{ik}|^{p_j} \right\}$$

if $p_j \leq 2$, $\forall j$ (Beckenbach and Bellman, 1965, Ch. 1, 14.(7)),

$$= \sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij} \left(1 - \frac{2}{p_j} \right) \ell_{p_j}(x^q - a_i) + 2 \sum_{i=1}^{n} \sum_{j=1}^{L} \frac{w_{ij}}{p_j} [\ell_{p_j}(x^q - a_i)]^{1-p_j} [\ell_{p_j}(x^{q+1} - a_i)]^{p_j}$$

$$\geq \sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij} \left(1 - \frac{2}{p_j} \right) \ell_{p_j}(x^q - a_i) + 2 \sum_{i=1}^{n} \sum_{j=1}^{L} \frac{w_{ij}}{p_j} \{ (1 - p_j) \ell_{p_j}(x^q - a_i) + p_j \ell_{p_j}(x^{q+1} - a_i) \}$$

if $p_j \geq 1$, $\forall j$ (Beckenbach and Bellman, 1965, Ch. 1, 14.(7)),

$$= \sum_{i=1}^{n} \sum_{j=1}^{L} [-w_{ij} \ \ell_{p_j}(x^q - a_i) + 2 w_{ij} \ \ell_{p_j}(x^{q+1} - a_i)]$$

$$= -W_m(x^q) + 2W_m(x^{q+1}) \quad . \tag{6.1.24}$$

Comparing (6.1.23) and (6.1.24) gives

$$-W_m(x^q) + 2W_m(x^{q+1}) < W_m(x^q) , \quad \text{if } 1 \le p_j \le 2, \ \forall j \ .$$

Hence,

$$W_m(x^{q+1}) < W_m(x^q) ,$$

ending the proof.

We now see that the descent property is guaranteed if all the p_j lie in the closed interval [1,2]. As in Section 5.2 (for $L=1$), we also need to address the problem of possible convergence of a given sequence $\{x^q\}$ to a non-optimal singular point. Two preliminary results are given next, which follow immediately from Lemmas 5.2.1 and 5.2.2.

Lemma 6.1.1

Let $p_j \in [1,2], j = 1, ..., L$, and consider any sequence x^q, $q = 0, 1, 2, ...,$ generated by the map T. Then $\{x^q\}$ and all the subsequences thereof converge to one and the same point.

Proof:

Identical to Lemma 5.2.1, with Properties 5.2.5 and 5.2.8 replaced by Properties 6.1.5 and 6.1.8 respectively.

Lemma 6.1.2

Consider any point $Q = (Q_1, ..., Q_N)^T$, and a sequence $\{x^q\}$ such that $x_t^q \neq Q_t$ for all q and some index $t \in \{1, ..., N\}$. Then relation (5.2.62) gives a sufficient condition for non-convergence of $\{x^q\}$ to Q.

Proof:

Same as for Lemma 5.2.2.

We now show that a regular sequence will never converge to a non-optimal singular point.

Property 6.1.9

Suppose $p_j \in [1,2]$, $j = 1, ..., L$, and let $Q \in S$ be a non-optimal location; i.e., $Q \neq x^*$. Then any regular sequence $\{x^q\}$ of T does not converge to Q.

Proof:

Using (6.1.6) and (6.1.7), we can rewrite the iterative transformation as follows:

$$x_t^{q+1} - a_{rt} = \phi_t(x^q) - a_{rt}$$

$$= \sum_{i \neq r} \beta_{it}(x^q)(a_{it} - a_{rt}) \Big/ \sum_{i=1}^{n} \beta_{it}(x^q)$$

$$= \frac{\displaystyle\sum_{i \neq r} (a_{it} - a_{rt}) \sum_{j=1}^{L} w_{ij} \frac{|x_t^q - a_{it}|^{p_j - 2}}{[\ell_{p_j}(x^q - a_i)]^{p_j - 1}}}{\displaystyle\sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij} \frac{|x_t^q - a_{it}|^{p_j - 2}}{[\ell_{p_j}(x^q - a_i)]^{p_j - 1}}}, \qquad (6.1.25)$$

for $r \in \{1, ..., n\}$ and $t \in \{1, ..., N\}$. Hence,

$$\frac{|x_t^{q+1} - a_{rt}|}{|x_t^q - a_{rt}|} = \frac{\left| \displaystyle\sum_{i \neq r} (a_{it} - a_{rt}) \sum_{j=1}^{L} w_{ij} \frac{|x_t^q - a_{it}|^{p_j - 2}}{[\ell_{p_j}(x^q - a_i)]^{p_j - 1}} \right|}{\displaystyle\sum_{j=1}^{L} w_{rj} \frac{|x_t^q - a_{rt}|^{p_j - 1}}{[\ell_{p_j}(x^q - a_r)]^{p_j - 1}} + |x_t^q - a_{rt}| \sum_{i \neq r} \sum_{j=1}^{L} w_{ij} \frac{|x_t^q - a_{it}|^{p_j - 2}}{[\ell_{p_j}(x^q - a_i)]^{p_j - 1}}}. \qquad (6.1.26)$$

Without loss in generality, assume that the p_j are arranged in increasing order. We also assume that the fixed points do not share common coordinates. (The proof can be easily modified otherwise.) Then, three cases are considered.

i) $p_1 > 1$, $Q \neq a_i$, $i = 1, \ldots, n$.

Using the same procedure as in case (i) of Property 5.2.9, we conclude that, if convergence of $\{x^q\}$ to Q takes place, then an r and t can be found such that

$$\frac{\partial W_m(Q)}{\partial x_t} \neq 0, \quad \text{and} \quad Q_t = a_{rt}.$$

Hence,

$$\frac{\partial W_m(Q)}{\partial x_t} = \sum_{i \neq r} \sum_{j=1}^{L} w_{ij} \operatorname{sign}(a_{rt} - a_{it}) \frac{|a_{rt} - a_{it}|^{p_j - 1}}{[\ell_{p_j}(Q - a_i)]^{p_j - 1}} \neq 0. \tag{6.1.27}$$

Using (6.1.26) and (6.1.27), it readily follows that

$$\lim_{\substack{x \to Q \\ x_t - a_{rt} \neq 0}} \frac{|\phi_t(x) - a_{rt}|}{|x_t - a_{rt}|} = +\infty; \tag{6.1.28}$$

so that convergence of $\{x^q\}$ to Q cannot occur by Lemma 6.1.2.

ii) $p_1 = 1$, $Q \neq a_i$, $i = 1, \ldots, n$.

Again, suppose that $\{x^q\}$ converges to Q. Using a similar reasoning as in case (i), we conclude that for $t \in \{1, \ldots, N\}$, if $Q_t \neq a_{it}$, $\forall i$, then $\partial W_m(Q)/\partial x_t = 0$. Let $J_Q = \{s \mid Q_s - a_{r_s s} = 0\}$, where $r_s \in \{1, \ldots, n\}$. Since $Q \in S$, J_Q is a non-empty set.

The directional derivative of W_m at Q in the direction $y = (y_1, \ldots, y_N)^T$ is calculated as follows:

$$W'_m(Q;y) = \lim_{\delta \to 0^+} \left\{ \frac{W_m(Q+\delta y) - W_m(Q)}{\delta} \right\}$$

$$= \lim_{\delta \to 0^+} \left\{ \frac{\sum_i \sum_j w_{ij} \ell_{p_j}(Q+\delta y - a_i) - \sum_i \sum_j w_{ij} \ell_{p_j}(Q - a_i)}{\delta} \right\}$$

$$= \lim_{\delta \to 0^+} \left\{ \frac{\sum_{i=1}^{n} w_{i1} [\ell_1(Q+\delta y - a_i) - \ell_1(Q-a_i)]}{\delta} \right.$$

$$\left. + \frac{\sum_{i=1}^{n} \sum_{j=2}^{L} w_{ij} [\ell_{p_j}(Q+\delta y - a_i) - \ell_{p_j}(Q-a_i)]}{\delta} \right\}$$

$$= \sum_{s \in J_Q} \{ w_{r_s 1} |y_s| + \sum_{i \neq r_s} w_{i1} \operatorname{sign}(a_{r_s s} - a_{is}) y_s \}$$

$$+ \sum_{t \in J'_Q} \sum_{i=1}^{n} w_{i1} \operatorname{sign}(Q_t - a_{it}) y_t + \sum_{i=1}^{n} \sum_{j=2}^{L} w_{ij} \nabla \ell_{p_j}(Q - a_i) \cdot y$$

$$= \sum_{s \in J_Q} \left\{ w_{r_s 1} |y_s| + \sum_{i \neq r_s} \left[w_{i1} \operatorname{sign}(a_{r_s s} - a_{is}) \right. \right.$$

$$\left. \left. + \sum_{j=2}^{L} w_{ij} \operatorname{sign}(a_{r_s s} - a_{is}) \frac{|a_{r_s s} - a_{is}|^{p_j - 1}}{[\ell_{p_j}(Q - a_i)]^{p_j - 1}} \right] |y_s| \right\}$$

$$+ \sum_{t \in J'_Q} \frac{\partial W_m(Q)}{\partial x_t} \cdot y_t \,.$$

But from above, we see that

$$\frac{\partial W_m(Q)}{\partial x_t} = 0, \qquad \forall \;\; t \in J_Q' .$$

Hence,

$$W_m'(Q;y) = \sum_{s\in J_Q} \left\{ w_{r_s 1}|y_s| + \sum_{i\neq r_s} \left[w_{i1}\,\mathrm{sign}(a_{r_s s}-a_{is}) + \sum_{j=2}^{L} w_{ij}\,\mathrm{sign}(a_{r_s s}-a_{is}) \frac{|a_{r_s s}-a_{is}|^{p_j-1}}{[\ell_{p_j}(Q-a_i)]^{p_j-1}} \right] y_s \right\} . \tag{6.1.29}$$

Since Q is a non-optimal location, therefore by Property 2.4.1,

$$\min_y W_m'(Q;y) < 0 . \tag{6.1.30}$$

This implies

$$\frac{1}{w_{r_s 1}} \left| \sum_{i\neq r_s} \left(w_{i1}\,\mathrm{sign}(a_{r_s s}-a_{is}) + \sum_{j=2}^{L} w_{ij}\,\mathrm{sign}(a_{r_s s}-a_{is}) \frac{|a_{r_s s}-a_{is}|^{p_j-1}}{[\ell_{p_j}(Q-a_i)]^{p_j-1}} \right) \right| > 1 , \tag{6.1.31}$$

for at least one $s \in J_Q$, say $s = h$. Using (6.1.26) with $p_1 = 1$, $t = h$, and $r = r_h$, and (6.1.31), it is readily seen that

$$\lim_{\substack{x\to Q \\ x_h - a_{r_h h} \neq 0}} \frac{|\phi_h(x) - a_{r_h h}|}{|x_h - a_{r_h h}|} > 1 . \tag{6.1.32}$$

Thus, convergence of $\{x^q\}$ to Q cannot occur due to Lemma 6.1.2.

iii) $Q = a_r$ for some $r \in \{1, \ldots, n\}$.

Consider the function

$$U_r = \sum_{j=1}^{L} w_{rj}\,\ell_{p_j}(x - a_r) , \tag{6.1.33}$$

which gives the contribution to W_m from the cost components due to a_r. For x along some ray from a_r with direction θ, it is readily verified that

$$U_r = w_r \ell_{p(\theta)}(x - a_r) , \tag{6.1.34}$$

where $p(\theta) \in (1,2)$ and

$$w_r = \sum_{j=1}^{L} w_{rj} .$$

Furthermore, from the unit circles of the ℓ_{p_j}, $j = 1, \ldots, L$, and considering the results obtained for the weighted one-two norm (Chapter 4), it follows that $w_r \ell_{p_*}(x - a_r)$ is a close approximation of U_r, where $p_* \in (1,2)$ is a mean value of $p(\theta)$. Thus, the iterates in a δ-neighbourhood of a_r behave to a first approximation as if the single cost $w_r \ell_{p_*}(x - a_r)$ is associated with a_r. Based on Property 5.2.9 (case (iii)), we see that $\{x^q\}$ will not converge to a_r, since this would have to be along a descent direction of W_m, thereby violating Property 6.1.8.

Since cases (i), (ii) and (iii) exhaust all possibilities, the proof is complete.

We are finally ready to prove global convergence of our algorithm. This extends the result given in Theorem 5.2.1 for the single norm to the more general mixed-norm model.

Theorem 6.1.1

Let $p_j \in [1,2]$, $j = 1, \ldots, L$. Then any regular sequence x^q, $q = 0, 1, 2, \ldots$, converges to an optimal solution; i.e.,

$$\lim_{q \to \infty} x^q = x^* . \tag{6.1.35}$$

Proof:

The same as Theorem 5.2.1, with W_m instead of W, Lemma 6.1.1 replacing Lemma 5.2.1, and Properties 6.1.6, 6.1.8 and 6.1.9 replacing Properties 5.2.6, 5.2.8 and 5.2.9 respectively.

It is interesting to consider the case where $\{x^q\}$ is a non-regular sequence. Using the same notation as in Corollary 5.2.1, we obtain the following immediate result.

Corollary 6.1.1

Let $p_j \in [1,2]$, $j = 1, \ldots, L$, and consider any non-regular sequence x^q, $q = 0, 1, 2, \ldots$. If all subsequent iterates after x^s do not fall on any hyperplanes H_{it} not already included in H, then $\{x^q\}$ converges to a solution which is optimal in the subspace H.

As a final comment on global convergence, we note that Property 5.2.10 readily extends to our mixed-norm problem as follows.

Property 6.1.10

If $p_r \in [1,2)$ for at least one $r \in \{1, \ldots, L\}$, then $\{x^q\}$ is a regular sequence, except for a set of starting points x^0 which is dense as R^{N-1}.

The steps of the proof are the same as in Property 5.2.10. We conclude that if $p_j \in [1,2]$ for all j, the likelihood that the algorithm will not converge to x^* for arbitrary x^0 is very low (zero theoretically if the sequence is calculated with unlimited accuracy). Once again, the use of double precision arithmetic is recommended.

We take a quick look now at local convergence rates, restricting attention to the case where W_m is strictly convex (i.e., the a_i are non-collinear), and the optimal solution x^* does not occur at a singular point. Let $\phi'(x^*)$ denote the $N \times N$ matrix of first partials of the vector ϕ evaluated at x^*. Then, using an analogous procedure as in subsection 5.3.1, it is readily seen that

$$\phi'(x^*) = [\phi_{kt}] , \tag{6.1.36}$$

where

$$\phi_{kt} = \frac{\partial \phi_k(x^*)}{\partial x_t} = \delta_{kt} - \frac{1}{s_k(x^*)} \frac{\partial^2 W_m(x^*)}{\partial x_t \partial x_k}, \tag{6.1.37}$$

$$\delta_{kt} = \begin{cases} 1, & \text{if } k = t, \\ 0, & \text{otherwise,} \end{cases} \quad \text{is the Kronecker delta,} \tag{6.1.38}$$

and

$$k, t = 1, \ldots, N.$$

Note that $\phi'(x^*)$ has the same form as in equation (5.3.37), except the new objective function W_m replaces W and $s_i(x)$ is now given by (6.1.17).

For the two-dimensional case ($N = 2$), the elements of $\phi'(x^*)$ are given by

$$\phi_{11} = \frac{1}{s_1(x^*)} \sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij} \frac{|x_1^* - a_{i1}|^{p_j - 2}}{[\ell_{p_j}(x^* - a_i)]^{p_j - 1}} \left\{ (2 - p_j) + (p_j - 1) \frac{|x_1^* - a_{i1}|^{p_j}}{[\ell_{p_j}(x^* - a_i)]^{p_j}} \right\}, \tag{6.1.39a}$$

$$\phi_{22} = \frac{1}{s_2(x^*)} \sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij} \frac{|x_2^* - a_{i2}|^{p_j - 2}}{[\ell_{p_j}(x^* - a_i)]^{p_j - 1}} \left\{ (2 - p_j) + (p_j - 1) \frac{|x_2^* - a_{i2}|^{p_j}}{[\ell_{p_j}(x^* - a_i)]^{p_j}} \right\}, \tag{6.1.39b}$$

$$\phi_{12} = \frac{1}{s_1(x^*)} \sum_{i=1}^{n} \sum_{j=1}^{L} (p_j - 1) w_{ij} \,\mathrm{sign}(x_1^* - a_{i1}) \,\mathrm{sign}(x_2^* - a_{i2}) \frac{|x_1^* - a_{i1}|^{p_j - 1} |x_2^* - a_{i2}|^{p_j - 1}}{[\ell_{p_j}(x^* - a_i)]^{2p_j - 1}}, \tag{6.1.39c}$$

and

$$\phi_{21} = \frac{s_1(x^*)}{s_2(x^*)} \phi_{12}. \tag{6.1.39d}$$

Consider the case where $p_j \in [1,2]$ for all j, so that global convergence to x^* of any regular sequence is guaranteed by Theorem 6.1.1. Clearly, ϕ_{11} and ϕ_{22} are positive-valued, so that

$$\lambda_1 + \lambda_2 = \mathrm{tr}[\phi'(x^*)] > 0, \tag{6.1.40}$$

where λ_1 and λ_2 are the eigenvalues of $\phi'(x^*)$. Next, we show that the determinant of $\phi'(x^*)$ is positive. Letting

$$F_t(x) = \sum_{i=1}^{n} \sum_{j=1}^{L} (p_j - 1)\, w_{ij} \frac{|x_t - a_{it}|^{2p_j - 2}}{[\ell_{p_j}(x - a_i)]^{2p_j - 1}} ,$$

$$G_t(x) = \sum_{i=1}^{n} \sum_{j=1}^{L} (2 - p_j)\, w_{ij} \frac{|x_t - a_{it}|^{p_j - 2}}{[\ell_{p_j}(x - a_i)]^{p_j - 1}} ,$$

$$t = 1, 2,$$

and

$$H(x) = \sum_{i=1}^{n} \sum_{j=1}^{L} (p_j - 1)\, w_{ij}\, \mathrm{sign}(x_1 - a_{i1}) \mathrm{sign}(x_2 - a_{i2}) \frac{|x_1 - a_{i1}|^{p_j - 1} |x_2 - a_{i2}|^{p_j - 1}}{[\ell_{p_j}(x - a_i)]^{2p_j - 1}} ,$$

the following expression is obtained:

$$\det[\phi'(x^*)] = \frac{1}{s_1(x^*)\, s_2(x^*)} \{G_1(x^*)\, G_2(x^*) + G_1(x^*) F_2(x^*)$$

$$+ G_2(x^*) F_1(x^*) + F_1(x^*)\, F_2(x^*) - H^2(x^*)\}. \qquad (6.1.41)$$

Applying Schwarz's inequality, we readily show that

$$H^2(x^*) < F_1(x^*)\, F_2(x^*) , \qquad (6.1.42)$$

with strict inequality holding even if the a_i are collinear. Furthermore, since F_1, F_2, G_1 and G_2 are all clearly positive, it now follows that

$$\lambda_1\, \lambda_2 = \det[\phi'(x^*)] > 0 . \qquad (6.1.43)$$

Comparing relations (6.1.40) and (6.1.43) leads to the result,

$$\lambda_t > 0 , \qquad t = 1, 2 . \qquad (6.1.44)$$

From (6.1.37) we see that

$$\phi'(x^*) = I - B_m(x^*) , \qquad (6.1.45)$$

where

$$B_m(x^*) = \begin{bmatrix} \dfrac{1}{s_1(x^*)} \dfrac{\partial^2 W_m(x^*)}{\partial x_1^2} & \dfrac{1}{s_1(x^*)} \dfrac{\partial^2 W_m(x^*)}{\partial x_2 \partial x_1} \\ \dfrac{1}{s_2(x^*)} \dfrac{\partial^2 W_m(x^*)}{\partial x_1 \partial x_2} & \dfrac{1}{s_2(x^*)} \dfrac{\partial^2 W_m(x^*)}{\partial x_2^2} \end{bmatrix} . \tag{6.1.46}$$

Furthermore, it can be shown in exactly the same way as for Lemma 5.3.1 that the eigenvalues of $B_m(x^*)$ are positive. Denoting these eigenvalues by μ_1 and μ_2, and recalling that

$$\lambda_t = 1 - \mu_t , \qquad t = 1, 2 , \tag{6.1.47}$$

leads to the result

$$\lambda_t < 1 , \qquad t = 1, 2 . \tag{6.1.48}$$

Comparing (6.1.44) and (6.1.48), we finally conclude that

$$0 < \lambda_t < 1 , \qquad t = 1, 2 . \tag{6.1.49}$$

Hence, the following important result is obtained.

Theorem 6.1.2

Consider the mixed-norm problem in two dimensions, such that $p_j \in [1,2]$, $j = 1, ..., L$, the a_i are non-collinear, and $x^* \notin S$. Then the local convergence rate to x^* is linear.

As a final comment on local convergence rates, we note that the results in subsection 5.3.3 are readily extended to the mixed-norm model. This follows from the previous observation that $\phi'(x^*)$ has an analogous form in both cases. In particular note that

$$\lambda_t < 1 , \qquad t = 1, \ldots, N , \tag{6.1.50}$$

for the more general condition where $p_j \geq 1$ (but not necessarily ≤ 2), for all j.

6.2 Optimality Criteria at Non-Differentiable Points

In this section, we are interested in deriving optimality criteria at points $x \in D$, where D denotes the set of non-differentiable points of W_m. Consider first the case where $p_j > 1$ for all j. Then the cost component U_r (see equation (6.1.33)) associated with any fixed point a_r is a positive linear combination of differentiable round norms centered at a_r, and hence is itself a differentiable round norm centered at a_r. It follows that $D = \{a_1, ..., a_n\}$ for this case. On the other hand, if one of the p_j's equals 1, then D becomes the union of hyperplanes $x_t - a_{it} = 0$, $i = 1, ..., n$, $t = 1, ..., N$. However, when x is neither an intersection nor a fixed point, there exists at least one direction x_h where W_m is differentiable; and hence $\partial W_m(x)/\partial x_h = 0$ is a necessary condition for x to be optimal. It follows in the latter case that optimality criteria are practical only at the intersection and fixed points.

Once again, we make the standard assumption that the a_i do not share any common coordinates with one another. If this is not so, the criteria given below can be easily modified, where applicable, to suit the individual case. Let us first consider the fixed point a_r, $r \in \{1, ..., n\}$. Using relation (2.4.6), it follows that the directional derivative here is given by

$$W'_m(a_r\,; y) = \sum_{j=1}^{L} w_{rj}\, \ell_{p_j}(y) - \sum_{t=1}^{N} B_{rt}\, y_t \,, \tag{6.2.1}$$

where

$$B_{rt} = -\sum_{i \neq r} \sum_{j=1}^{L} w_{ij}\, \mathrm{sign}(a_{rt} - a_{it}) \frac{|a_{rt} - a_{it}|^{p_j - 1}}{[\ell_{p_j}(a_r - a_i)]^{p_j - 1}} \,, \quad t = 1, \ldots, N \,. \tag{6.2.2}$$

Letting

$$B_r := (B_{r1}, \ldots, B_{rN}),$$

$$w_r := \sum_{j=1}^{L} w_{rj},$$

$$a_{rj} := w_{rj} / w_r \,, \qquad j = 1, \ldots, L,$$

we rewrite (6.2.1) as,

$$W'_m(a_r\,; y) = w_r \sum_{j=1}^{L} a_{rj}\, \ell_{p_j}(y) - B_r \cdot y \;. \tag{6.2.3}$$

A necessary and sufficient condition for a_r to be an optimal solution is that $W'_m(a_r\,; y) \geq 0$ for all directions y (Property 2.4.1). This immediately implies

$$\max_{y} \left| \frac{1}{w_r} \frac{B_r \cdot y}{L_r(y)} \right| \le 1 , \tag{6.2.4}$$

where the norm,

$$L_r(x) := \sum_{j=1}^{L} a_{rj} \ell_{p_j}(x) . \tag{6.2.5}$$

Note that L_r is a convex combination of the ℓ_{p_j}'s. Using the definition of the polar and letting $L_r^{\,o}$ denote the polar of L_r, the optimality criterion at a_r given in (6.2.4) becomes

$$\frac{1}{w_r} L_r^o(B_r) \le 1 . \tag{6.2.6}$$

Relation (6.2.6) is identical in form to the optimality criterion of Juel and Love (1981), where a single arbitrary norm is associated with each fixed point. The total weight (w_r) at a_r has the same interpretation as the simple cost coefficient in the single-norm case. However, the polar of an arbitrary convex combination of ℓ_p norms is not readily available in closed form, and obtaining $L_r^{\,o}$ for a specific problem proves to be a cumbersome task (e.g., see Juel, 1975). Thus, fixed point optimality criteria have limited practical use in the mixed norm problem. At the end of this section, we demonstrate the application of (6.2.6) when L_r is a convex combination of ℓ_1 and ℓ_2.

Suppose now that $p_j \in [1,2]$ for all j. Then a lower bound on the directional derivative of W_m at a_r can be obtained in the following manner. Note that

$$\ell_{p_j}(y) \ge \ell_2(y) = 1 , \quad j = 1, \ldots, L , \tag{6.2.7}$$

and $B_r \cdot y$ is maximized when the unit vector y has the same direction as B_r; i.e.,

$$y = B_r / \|B_r\| . \tag{6.2.8}$$

Using (6.2.1), (6.2.7) and (6.2.8), it immediately follows that for any y,

$$W_m'(a_r ; y) \ge \sum_{j=1}^{L} w_{rj} - B_r \cdot B_r / \|B_r\|$$

$$= w_r - \|B_r\| . \tag{6.2.9}$$

Thus, a sufficient condition for a_r to be an optimal solution is that

$$w_r \geq \|B_r\| = \left| \sum_{t=1}^{N} B_{rt}^2 \right|^{1/2} . \tag{6.2.10}$$

The above criterion provides a practical (but obviously less accurate) alternative to relation (6.2.6).

Witzgall (1964) proves a "Majority Theorem" which applies to the Weber problem with distance function given by a metric. This theorem states that a fixed point is always optimal if it accounts for 50% or more of the total interaction. One can easily extend this result to a generalized version of the mixed-norm model, as shown below.

Property 6.2.1

Consider the problem,

$$\text{minimize } W_m(x) = \sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij}\, d_j(x, a_i) ,$$

where d_j is a metric, $j = 1, \ldots, L$, and the other symbols are as before. Then a sufficient condition for a_r to be an optimal solution for some $r \in \{1, \ldots, n\}$ is that

$$w_{rj} \geq \sum_{i \neq r} w_{ij} , \quad j = 1, \ldots, L . \tag{6.2.11}$$

Proof:

From the triangle inequality and symmetry properties of the metric, we obtain

$$|d_j(x, a_i) - d_j(a_r, a_i)| \leq d_j(x, a_r) , \quad \forall\, j, i, x . \tag{6.2.12}$$

Hence,

$$W_m(x) - W_m(a_r) = \sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij} d_j(x, a_i) - \sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij} d_j(a_r, a_i)$$

$$= \sum_{i \neq r} \sum_{j=1}^{L} w_{ij}[d_j(x, a_i) - d_j(a_r, a_i)] + \sum_{j=1}^{L} w_{rj}\, d_j(x, a_r)$$

$$\geq -\sum_{i \neq r} \sum_{j=1}^{L} w_{ij} |d_j(x, a_i) - d_j(a_r, a_i)| + \sum_{j=1}^{L} w_{rj}\, d_j(x, a_r)$$

$$\geq \sum_{j=1}^{L} \sum_{i \neq r} w_{ij}[d_j(x, a_r) - | d_j(x, a_i) - d_j(a_r, a_i)|]$$

(relation (6.2.11))

$$\geq \quad 0 \qquad \text{(relation (6.2.12))} \; .$$

Therefore, a_r must be an optimal solution if the w_{rj} satisfy (6.2.11).

Returning to the mixed-norm model (6.1), we now provide a sufficient condition for a_r to be optimal which is based on the combined effect of the w_{rj}, $j = 1, \ldots, L$, as measured by the total weight w_r at a_r.

Property 6.2.2

Consider the mixed-norm model (6.1) with $1 \leq p_j \leq 2$ for all j. Suppose that an $r \in \{1, \ldots, n\}$ can be found such that

$$w_r \geq \left(\frac{\sqrt{N}}{1 + \sqrt{N}} \right) w \, , \tag{6.2.13}$$

where

$$w = \sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij} \; .$$

Then a_r is an optimal solution.

Proof:

For any $x \in R^N$,

$$W_m(x) - W_m(a_r) = S_1 + S_2,$$

where

$$S_1 = \sum_{j=1}^{L} w_{rj}\, \ell_{p_j}(x - a_r),$$

and

$$S_2 = \sum_{i \neq r} \sum_{j=1}^{L} w_{ij} [\ell_{p_j}(x - a_i) - \ell_{p_j}(a_r - a_i)] .$$

Since $1 \leq p_j \leq 2$, $\forall j$, it follows that for any $z \in R^N$,

$$\ell_2(z) \leq \ell_{p_j}(z) \leq \ell_1(z), \qquad j = 1, \ldots, L. \tag{6.2.14}$$

Furthermore, in N-dimensional space,

$$\ell_1(z) \leq \sqrt{N}\; \ell_2(z) . \tag{6.2.15}$$

Using (6.2.14) and (6.2.15), we see that

$$S_1 \geq \sum_{j=1}^{L} w_{rj}\, \ell_2(x - a_r) = w_r\, \ell_2(x - a_r), \tag{6.2.16}$$

and

$$S_2 \geq - \sum_{i \neq r} \sum_{j=1}^{L} w_{ij} |\ell_{p_j}(x - a_i) - \ell_{p_j}(a_r - a_i)|$$

$$\geq - \sum_{i \neq r} \sum_{j=1}^{L} w_{ij}\, \ell_{p_j}(x - a_r) \qquad \text{(triangle inequality)}$$

$$\geq - \sqrt{N}\; \ell_2(x - a_r) \sum_{i \neq r} \sum_{j=1}^{L} w_{ij} . \tag{6.2.17}$$

Substituting

$$\sum_{i \neq r} \sum_{j=1}^{L} w_{ij} = w - w_r, \tag{6.2.18}$$

into (6.2.17) and combining with (6.2.16) gives

$$S_1 + S_2 \geq [(1 + \sqrt{N})w_r - \sqrt{N}\; w]\, \ell_2(x - a_r) . \tag{6.2.19}$$

Hence, if w_r satisfies (6.2.13), we immediately obtain

$$W_m(x) - W_m(a_r) \geq 0 ;$$

so that a_r must be an optimal solution.

It is informative to note the following alternative proof of Property 6.2.2, based on the directional derivative of W_m at a_r. From (6.2.2), it is clear that

$$|B_{rt}| \leq \sum_{i \neq r} \sum_{j=1}^{L} w_{ij} = w - w_r , \quad \forall\, t. \tag{6.2.20}$$

Using the lower bound on $W'_m(a_r ; y)$ in (6.2.9), and imposing (6.2.13), we thus obtain

$$\begin{aligned} W'_m(a_r ; y) &\geq w_r - \left(\sum_{t=1}^{N} B_{rt}^2 \right)^{1/2} \\ &\geq w_r - \sqrt{N}(w - w_r) \\ &= (1 + \sqrt{N})w_r - \sqrt{N} w \geq 0 , \quad \forall\, y , \end{aligned}$$

therefore concluding that a_r is an optimal solution.

For location in the plane ($N = 2$), it now follows that a sufficient condition for a_r, $r \in \{1, ..., n\}$, to be optimal is that

$$w_r \geq \left(\frac{\sqrt{2}}{1 + \sqrt{2}} \right) w \approx 0.5858\, w . \tag{6.2.21}$$

However, we note from (6.2.13) that as N increases, w_r must become a larger fraction of the cumulative weight w in order to guarantee the optimality of a_r. Hence, the usefulness of Property 6.2.2 is limited to lower dimensional spaces.

Let us assume without loss in generality that the p_j's are arranged in increasing order, and consider the case where $p_1 = 1$. Then, by Property 2.2.5, L_r defined in (6.2.5) is a nondifferentiable round norm. For minisum models employing block or nondifferentiable round norms, one can derive optimality criteria at any nondifferentiable point of the objective function by considering the directional derivative in a similar manner as at the fixed points. However, it appears that such criteria are not provided in the published literature, other than for the fixed points. In the next result, optimality criteria are specified at the intersection points.

Property 6.2.3

Consider the mixed-norm model (6.1) with $p_1 = 1$, and let Q denote an intersection point; i.e.,

$$Q = (a_{r_1 1}, a_{r_2 2}, \ldots, a_{r_N N})^T$$

where $r_s \in \{1, \ldots, n\}$ depends on $s = 1, \ldots, N$. (Recall that $Q \neq a_i$, $i = 1, \ldots, n$, by definition of an intersection point.) Then Q is an optimal solution if, and only if,

$$\min_{1 \leq s \leq N} \{C_s\} \geq 0 , \tag{6.2.22}$$

where

$$C_s = w_{r_s 1} - \left| \sum_{i \neq r_s} \left| w_{i1} \operatorname{sign}(a_{r_s s} - a_{is}) + \sum_{j=2}^{L} w_{ij} \operatorname{sign}(a_{r_s s} - a_{is}) \frac{|a_{r_s s} - a_{is}|^{p_j - 1}}{[\ell_{p_j}(Q - a_i)]^{p_j - 1}} \right| \right| , \tag{6.2.23}$$

$s = 1, \ldots, N$.

Proof:

Referring to the directional derivative calculated in (6.1.29), and noting that $J_Q = \{1, \ldots, N\}$ for the intersection point Q, it follows that

$$\min_y W_m'(Q ; y) = \min_y \left\{ \sum_{s=1}^{N} C_s |y_s| \right\} . \tag{6.2.24}$$

Thus, Q is optimal if, and only if, $C_s \geq 0$ for all s. We conclude that (6.2.22) is both a necessary and sufficient condition for Q to be an optimal solution.

An Application:

Let us consider a typical mixed-norm problem in R^2, which utilizes the rectangular and Euclidean norms. The model is stated as follows:

$$\text{minimize } W_m(x) = \sum_{i=1}^{n} \sum_{j=1}^{L} w_{ij}\, \ell_{p_j}(x - a_i) ,$$

where N = 2, L = 2, $p_1 = 1$, and $p_2 = 2$. Alternatively, the objective function can be written as

$$W_m(x) = \sum_{i=1}^{n} w_i\, L_i(x - a_i) , \tag{6.2.25}$$

with

$$w_i = w_{i1} + w_{i2} ,$$

$$L_i(z) = \alpha_i\, \ell_1(z) + \beta_i\, \ell_2(z) , \qquad \forall\ z \in R^2 ,$$

$$0 < \alpha_i = w_{i1}/w_i < 1 , \qquad \beta_i = w_{i2}/w_i = 1 - \alpha_i ,$$

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

We see that L_i is a convex combination of the ℓ_1 and ℓ_2 norms. Furthermore, if α_i has the same value for all i ($\alpha_i = \alpha$, i = 1, ..., n), the problem reduces to a standard minisum model with distances given by a weighted one-two norm ($L(z) = \alpha\, \ell_1(z) + (1 - \alpha)\, \ell_2(z)$).

The fixed point optimality criteria can be derived directly from relation (6.2.6), using the polar for positive linear combinations of ℓ_1 and ℓ_2 given by Juel (1975, p. 15). However, it is simpler to back-track a few steps, and utilize the special structure of the directional derivative for this case. The criteria thus obtained have a different form, which is more compact and easier to implement.

From (6.2.3),

$$\begin{aligned} W'_m(a_r ; y) &= w_r[\alpha_r\, \ell_i(y) + \beta_r\, \ell_2(y)] - B_r \cdot y \\ &= w_r[\alpha_r\, \ell_1(y) + \beta_r] - B_r \cdot y , \end{aligned} \tag{6.2.26}$$

since y is a unit vector ($\ell_2(y) = 1$). Letting θ denote the direction of y, we have

$$y = (\cos\theta,\ \sin\theta)^T ; \tag{6.2.27}$$

so that (6.2.26) becomes

$$W'_m(a_r\,;y) = w_r[\,\alpha_r(|\cos\theta| + |\sin\theta|) + \beta_r\,] - B_{r1}\cos\theta - B_{r2}\sin\theta\,. \tag{6.2.28}$$

Clearly, the directional derivative at a_r will be minimized if, and only if,

$$\text{sign}(\cos\theta) = \text{sign}(B_{r1}), \quad \text{sign}(\sin\theta) = \text{sign}(B_{r2}); \tag{6.2.29}$$

in which case,

$$W'_m(a_r\,;y) = (\alpha_r w_r - |B_{r1}|)\,|\cos\theta| + (\alpha_r w_r - |B_{r2}|)\,|\sin\theta| + (1-\alpha_r)w_r\,. \tag{6.2.30}$$

Also note that the direction which minimizes $W'_m(a_r\,;y)$ is restricted to a specific quadrant by (6.2.29).

Suppose that $\alpha_r w_r \geq \min\{|B_{r1}|, |B_{r2}|\}$, and without loss in generality, assume that $|B_{r1}| = \min\{|B_{r1}|, |B_{r2}|\}$. Then the first term on the right-hand side of equation (6.2.30) is non-negative, and clearly, a θ which minimizes the directional derivative at a_r satisfies $(|\cos\theta_*|, |\sin\theta_*|) = (0, 1)$. Hence, we obtain

$$\min_y \{W'_m(a_r\,;y)\} = -\max\{|B_{r1}|, |B_{r2}|\} + w_r\,. \tag{6.2.31}$$

The other possibility to consider is that $\alpha_r w_r < \min\{|B_{r1}|, |B_{r2}|\}$. For this case, it is readily shown using elementary calculus that

$$\min_y \{W'_m(a_r\,;y)\} = -[(\alpha_r w_r - |B_{r1}|)^2 + (\alpha_r w_r - |B_{r2}|)^2\,]^{1/2} + (1-\alpha_r)w_r\,. \tag{6.2.32}$$

From (6.2.31) and (6.2.32), we immediately obtain the following optimality criterion at a fixed point.

Property 6.2.4

Consider the mixed-norm problem with objective function defined in (6.2.25). Then $x = a_r$, $r \in \{1, \ldots, n\}$, is an optimal solution if, and only if,

$$w_r \geq \begin{cases} \max\{|B_{r1}|, |B_{r2}|\}, & \text{if } a_r w_r \geq \min\{|B_{r1}|, |B_{r2}|\}, \\ \left(\frac{1}{1-a_r}\right)[(a_r w_r - |B_{r1}|)^2 + (a_r w_r - |B_{r2}|)^2]^{1/2}, & \text{otherwise.} \end{cases} \tag{6.2.33}$$

It is interesting to examine relation (6.2.33) for the limiting cases, $a_r \to 1^-$ ($L_r \to \ell_1$), and $a_r \to 0^+$ ($L_r \to \ell_2$). Assuming that $w_r \neq |B_{r1}|$ or $w_r \neq |B_{r2}|$, we see that

$$\lim_{a_r \to 1^-} \left|\left(\frac{1}{1-a_r}\right)[(a_r w_r - |B_{r1}|)^2 + (a_r w_r - |B_{r2}|)^2]^{1/2}\right| = +\infty .$$

Hence, the optimality criterion at a_r reduces in the first case to testing whether or not

$$w_r \geq \max\{|B_{r1}|, |B_{r2}|\} . \tag{6.2.34}$$

For the second case we have $a_r w_r \to 0$, and hence, the limiting form of (6.2.33) becomes

$$w_r \geq (B_{r1}^2 + B_{r2}^2)^{1/2} . \tag{6.2.35}$$

Note that (6.2.34) and (6.2.35) are the optimality criteria at a_r with L_r as the rectangular and Euclidean norms respectively.

The following results give some useful information concerning the sensitivity of an optimal solution at a fixed point.

Property 6.2.5

For the mixed-norm problem with objective function defined by (6.2.25), suppose that an $r \in \{1, \ldots, n\}$ exists such that $w_r > \max\{|B_{r1}|, |B_{r2}|\}$. Then an $a^* \in [0, 1)$ can be found such that $x = a_r$ is an optimal solution for all a_r in the interval $[a^*, 1]$.

Proof:

Since $w_r > \max\{|B_{r1}|, |B_{r2}|\}$, it follows that an $a^* \in [0, 1)$ exists such that

$$a^* w_r = \min\{|B_{r1}|, |B_{r2}|\} .$$

Then, for any $a_r \in [a^*, 1]$, we see that the criterion in (6.2.33) is satisfied. Hence, by Property 6.2.4, a_r is an optimal solution for all $a_r \in [a^*, 1]$.

Property 6.2.6

For the mixed-norm problem with objective function defined by (6.2.25), suppose that an $r \in \{1, \ldots, n\}$ exists such that $w_r \geq (B_{r1}^{\ 2} + B_{r2}^{\ 2})^{1/2}$. Then $x = a_r$ is an optimal solution for all a_r in the interval [0, 1].

Proof:

Without loss in generality, assume that

$$|B_{r1}| = \min\{|B_{r1}|, |B_{r2}|\} .$$

There are two possibilities to consider.

(i) $|B_{r1}| = 0$.

Then $w_r \geq |B_{r2}| = \max\{|B_{r1}|, |B_{r2}|\}$, and $a_r w_r \geq \min\{|B_{r1}|, |B_{r2}|\}$ for any $a_r \in [0, 1]$. If follows from Property 6.2.4 that a_r is an optimal solution for all $a_r \in [0, 1]$.

(ii) $|B_{r1}| > 0$.

Let

$$h_1(a_r) := a_r w_r - |B_{r1}|, \tag{6.2.36a}$$

$$h_2(a_r) := a_r w_r - |B_{r2}|, \tag{6.2.36b}$$

and

$$g(a_r) := \left(\frac{1}{1 - a_r}\right)[h_1^2(a_r) + h_2^2(a_r)]^{1/2}. \tag{6.2.37}$$

The first-order derivative of g is given by,

$$\frac{dg(a_r)}{da_r} = \frac{1}{(1 - a_r)^2}\left|(h_1^2 + h_2^2)^{1/2} + \frac{(1 - a_r)w_r}{(h_1^2 + h_2^2)^{1/2}} \cdot (h_1 + h_2)\right| . \tag{6.2.38}$$

Now, since $w_r \geq (B_{r1}^{\ 2} + B_{r2}^{\ 2})^{1/2}$, it follows that

$$w_r > |B_{r2}| = \max\{|B_{r1}|, |B_{r2}|\} . \tag{6.2.39}$$

Then, an $a^* \in (0, 1)$ exists such that

$$a^* w_r = |B_{r1}| = \min\{|B_{r1}|, |B_{r2}|\} \ . \tag{6.2.40}$$

From (6.2.39), (6.2.40) and Property 6.2.4, we conclude that a_r is an optimal solution for all $a_r \in [a^*, 1]$.

Next consider a_r in the interval $[0, a^*)$. Clearly, $h_1 < 0$ and $h_2 < 0$; so that

$$h_1 + h_2 = -(|h_1| + |h_2|) \ . \tag{6.2.41}$$

Substituting $a_r = 0$ in (6.2.38), we obtain

$$\begin{aligned} \frac{dg(0)}{da_r} &= \left| (B_{r1}^2 + B_{r2}^2)^{1/2} - \frac{w_r}{(B_{r1}^2 + B_{r2}^2)^{1/2}} (|B_{r1}| + |B_{r2}|) \right| \\ &\le [(B_{r1}^2 + B_{r2}^2)^{1/2} - (|B_{r1}| + |B_{r2}|)] \\ &< 0 . \qquad \text{(Property 2.1.1)} \end{aligned} \tag{6.2.42}$$

Also, $g(0) = (B_{r1}{}^2 + B_{r2}{}^2)^{1/2}$, and hence

$$w_r \ge g(0) \ . \tag{6.2.43}$$

From (6.2.42) and (6.2.43), it follows that an $a^{**} > 0$ exists such that $w_r \ge g(a_r)$ for all $a_r \in [0, a^{**}]$. Furthermore, let a^{**} denote the largest value for which this is true. We show by contradiction that $a^{**} \ge a^*$. Suppose $a^{**} < a^*$. Then, using (6.2.41), we have

$$\begin{aligned} \frac{dg(a^{**})}{da_r} &= \frac{1}{(1-a^{**})^2} \left| (h_1^2(a^{**}) + h_2^2(a^{**}))^{1/2} - \frac{w_r}{g(a^{**})} (|h_1(a^{**})| + |h_2(a^{**})|) \right| \\ &\le \frac{1}{(1-a^{**})^2} [(h_1^2(a^{**}) + h_2^2(a^{**}))^{1/2} - (|h_1(a^{**})| + |h_2(a^{**})|)] \\ &< 0 . \qquad \text{(Property 2.1.1)} \end{aligned} \tag{6.2.44}$$

But this contradicts the fact that a^{**} is the largest value such that $w_r \ge g(a_r)$ for all $a_r \in [0, a^{**}]$. Hence, $a^{**} \ge a^*$, and we conclude from Property 6.2.4 that a_r is an optimal solution for all $a_r \in [0, a^*)$.

Thus, combining results, we see that $x = a_r$ is optimal for $0 \le a_r \le 1$.

Property 6.2.7

For the mixed-norm problem with objective function defined by (6.2.25), suppose that an $r \in \{1, \ldots, n\}$ exists such that a_r is optimal for $\alpha_r = \alpha_{r1}$ and $\alpha_r = \alpha_{r2}$, where α_{r1}, α_{r2} are values in the interval $[0, 1]$ and $\alpha_{r1} < \alpha_{r2}$. Then a_r is optimal for all $\alpha_r \in [\alpha_{r1}, \alpha_{r2}]$.

Proof:

A straightforward modification of the one given for Property 6.2.6.

The next result provides the criterion for testing the optimality of an intersection point in our example.

Property 6.2.8

Consider the mixed-norm model with objective function given in (6.2.25), and let Q denote an intersection point; i.e.

$$Q = (a_{r1}, a_{s2})^T,$$

where $r, s \in \{1, \ldots, n\}$, $r \neq s$. (Recall that $Q \neq a_i$, $i = 1, \ldots, n$, by definition of an intersection point.) Let

$$C_1 = w_{r1} - \left| \sum_{i \neq r} \left| w_{i1}\,\mathrm{sign}(a_{r1} - a_{i1}) + w_{i2} \frac{(a_{r1} - a_{i1})}{\ell_2(Q - a_i)} \right| \right| , \tag{6.2.45a}$$

$$C_2 = w_{s1} - \left| \sum_{i \neq s} \left| w_{i1}\,\mathrm{sign}(a_{s2} - a_{i2}) + w_{i2} \frac{(a_{s2} - a_{i2})}{\ell_2(Q - a_i)} \right| \right| . \tag{6.2.45b}$$

Then, Q is an optimal solution if, and only if,

$$\min\{C_1, C_2\} \geq 0 . \tag{6.2.46}$$

Proof:

This is a direct application of Property 6.2.3.

CHAPTER 7

A GENERALIZED MINISUM PROBLEM

An extension of the single facility minisum location problem (or Weber problem) has the distance function raised to some power K. This generalization was discussed briefly in Chapter 1 (see model (1.3.1)), where we noted that economies of scale are introduced in the model if $0 < K < 1$, while dis-economies occur when $K > 1$. Finally, if $K = 1$, we are back to the original Weber problem with its constant returns to scale. The purpose of the new formulation is to provide a more accurate representation of the cost structure in the real problem. As explained in Chapter 1, the actual costs will often exhibit a nonlinear relation with the distance function.

In this chapter, a new model is investigated which generalizes the minisum problem with distances measured by a norm even further. This model, which we appropriately name the "mixed-power" problem, is formulated as follows:

$$\text{minimize } W_G(x) = \sum_{i=1}^{n} w_i \, [k(x - a_i)]^{K_i} , \tag{7.1}$$

where w_i, a_i, $i = 1, \ldots, n$, $k(\cdot)$ and x are the same as in model (5.1), and $K_i > 0$, $i = 1, \ldots, n$, is the power associated with the i^{th} fixed point or destination. If all the K_i are equal, we have the original extension of the minisum problem. However, by allowing different values of the K_i for each customer, we are providing a greater flexibility in the cost structure. This recognizes the fact that economies or dis-economies may differ among customers, due to the use of different transportation modes and other possible factors.

We begin this chapter by deriving some general properties of the mixed-power problem. Next, we look at the specific case where k is the ℓ_p norm. The model then becomes

$$\text{minimize } W_G(x) = \sum_{i=1}^{n} w_i [\ell_p(x-a_i)]^{K_i}, \quad p \geq 1. \tag{7.2}$$

To our best knowledge, the mixed-power problem has not been formulated previously in the published literature on location theory. Since this model allows the cost component of each customer to have a different functional form, it can be considered analogous to the mixed-norm problem. We could complicate matters further, but hopefully increase the accuracy of the model as well, by mixing norms and powers simultaneously. For example, we might consider the following model:

$$\text{minimize } W_{GM}(x) = \sum_{i=1}^{n} w_i [k_i(x-a_i)]^{K_i}, \tag{7.3}$$

where k_i is the norm associated with customer i, and all other symbols are the same as above.

7.1 General Properties

We first derive some properties related to the shape of the objective function. These results generalize ones obtained by Morris (1981) for the case where k is an ℓ_p norm and all the K_i are equal. Unless stated otherwise, the location problem takes place in N-dimensional space (R^N).

Property 7.1.1

Let k be a round norm in model (7.1). If $K_i \geq 1$, i = 1, ..., n, with at least one of these inequalities satisfied strictly, then W_G is a strictly convex function of x.

Proof:

$W_G(x)$ is a positive linear combination of convex terms, with one or more of these terms having $K_r > 1$, $r \in \{1,...,n\}$, and being strictly convex by Property 2.3.5. Hence, $W_G(x)$ is strictly convex.

Property 7.1.2

If $0 < K_r < 1$, where $r \in \{1,\dots,n\}$, then the fixed point a_r is a local minimum of $W_G(x)$.

Proof:

Letting

$$h_i(x) := [k(x)]^{K_i}, \qquad i = 1, \dots, n, \tag{7.1.1}$$

we can rewrite the objective function as

$$W_G(x) = \sum_{i=1}^{n} w_i h_i(x - a_i) . \tag{7.1.2}$$

Then the directional derivative of W_G evaluated at $x = a_r$ in the direction y is given by

$$W'_G(a_r ; y) = \sum_{i \neq r} w_i h'_i(a_r - a_i ; y) + w_r h'_r(0 ; y) . \tag{7.1.3}$$

But for all $i \neq r$ (see equation (2.4.9)),

$$h'_i(a_r - a_i ; y) = K_i[k(a_r - a_i)]^{K_i - 1} k'(a_r - a_i ; y) , \tag{7.1.4}$$

which is finite-valued; while for $i = r$,

$$h'_r(0 ; y) = +\infty , \tag{7.1.5}$$

by equation (2.4.12). It follows that

$$W'_G(a_r ; y) = +\infty , \tag{7.1.6}$$

for all directions y, and hence, the fixed point a_r is a local minimum of W_G.

Property 7.1.3

If $0 < K_r < 1$ for one or more $r \in \{1,\dots,n\}$, then $W_G(x)$ is neither convex nor concave.

Proof:

By Property 7.1.2, a local minimum occurs at $x = a_r$, an interior point, so that W_G cannot be a concave function of x. Since W_G is bounded in any compact set containing a_r,and $W_G'(a_r\,;y) = +\infty$ for all directions y (equation (7.1.6)), it follows that W_G cannot be a convex function of x either.

Returning to Property 7.1.1, we see that the optimal solution (x^*) must be unique, if k is a round norm and $K_i \geq 1, \forall\, i$, with at least one of these inequalities satisfied in a strict sense. The uniqueness of x^* is guaranteed here for any arrangement of the fixed points-- even the collinear case. Suppose now that k is a block norm. Then W_G will no longer be a strictly convex function of x, so that x^* may not be unique. However, from Property 2.3.7 it follows that the optimal solutions must all lie on a facet of some polytope contour of $k(x - a_r)$, for each $r \in \{1,...,n\}$ such that $K_r > 1$. This result provides an easy way of checking the uniqueness of the optimal solution for problems in R^2 (N=2). Say that we have found an optimal solution at x_0. Now draw the edge of the polygon contour of $k(x - a_r)$ passing through x_0 for each $r \in \{1,...,n\}$ having $K_r > 1$. If any two of these edges are not parallel, then x_0 must be the unique solution. In higher-dimensional spaces, this verification step becomes more difficult, since we are now dealing with the intersection of hyperplanes instead of edges.

The uniqueness properties discussed above are illustrated with two simple examples in R^2. First consider a problem having two fixed points, $a_1 = (0,1)^T$ and $a_2 = (2,0)^T$, $w_1 = w_2 = w$, and $K_1 = K_2 = K$. Suppose that k is a round norm. Then if $K = 1$, the set of optimal solutions consists of all the points on the line segment joining a_1 and a_2. However, if $K > 1$, then $x^* = (1, 1/2)^T$ is the unique solution. Now suppose that $k \equiv \ell_1$ (a block norm). If $K = 1$, all the points contained in the rectangle with vertices $(0,0)^T$, $(0,1)^T$, $(2,1)^T$ and $(2,0)^T$ are optimal. If $K > 1$, x^* occurs at a reduced set consisting of the points on the 45° line segment through $(1, 1/2)^T$ bounded by the sides of the rectangle (see Figure 7.1.1(a)). We see that x^* is not unique here even when $K > 1$.

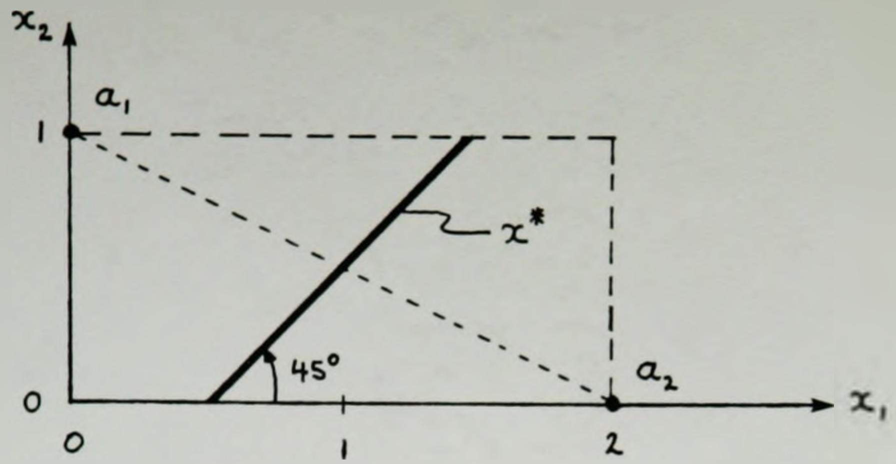
For the second example, add two more destinations at the unoccupied vertices of the rectangle; that is, $a_3 = (0,0)^T$ and $a_4 = (2,1)^T$. Let $w_i = w$ and $K_i = K$, $i = 1,\ldots,4$. Then the optimal solution is uniquely given by $x^* = (1, 1/2)^T$ for k a round norm and $K \geq 1$. If $k = \ell_1$ and $K = 1$, all the points within the rectangle are optimal as in the first example. But if $K > 1$, $x^* = (1, 1/2)^T$ becomes the unique solution, since the edges passing through x^* of the polygon contours for $k\,(x - a_1)$ and $k\,(x - a_2)$ are not parallel to those of $k\,(x - a_3)$ and $k\,(x - a_4)$, as shown in Figure 7.1.1(b).

Let us consider now the case where $0 < K_r < 1$ for at least one of the indices r. In a small δ-neighbourhood of the fixed point a_r, the directional derivative of W_G is dominated by the contribution from the cost component associated with a_r. Thus, along any line segment through a_r and contained in the δ-neighbourhood, W_G has the basic shape shown in Figure 2.4.1(c). This confirms graphically Property 7.1.3. Since W_G is neither a convex nor concave function of x, the optimal solution will not be unique in general. By Property 7.1.2, a local optimum occurs at $x = a_r$, which cannot be ruled out a priori as the global optimum. A global solution may also exist at a point other than the a_i.

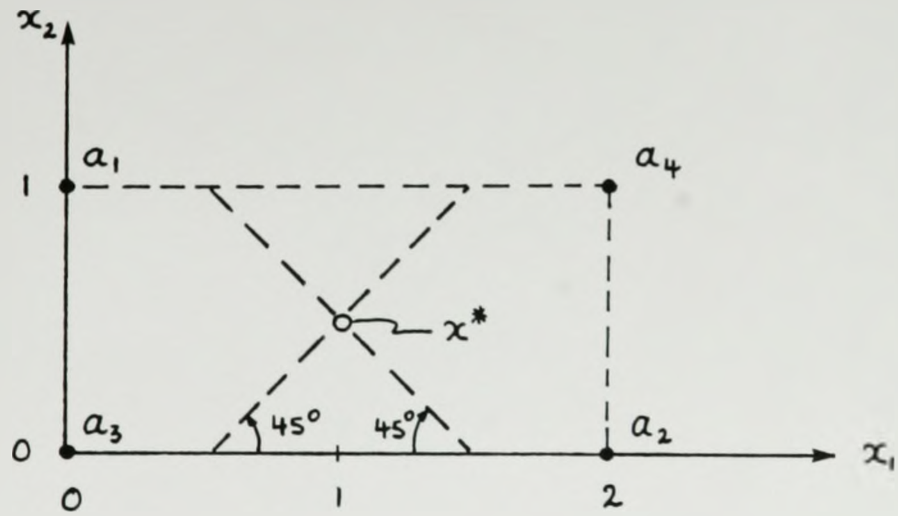
The following result is rather interesting, in that it shows a tendency for the optimal solution to move to a fixed point for sufficiently small values of the K_i.

Property 7.1.4

In the limiting case $K_i \rightarrow 0^+$, $\forall\, i$, the optimal solution of model (7.1) occurs at $x^* = a_s$, where $w_s = \max_i \{w_i\}$.



a) Nonunique



b) Unique

Figure 7.1.1 Uniqueness of the Optimal Solution (x^*) Illustrated for Rectangular Norm and $K > 1$.

Proof:

Denote this limiting case by 'lim'. Then for $x \neq a_i$, $i = 1,...,n$, we have

$$\lim W_G(x) = \sum_{i=1}^{n} w_i \; ; \tag{7.1.7}$$

while for $x = a_r$, $r \in \{1,...,n\}$,

$$\begin{aligned} \lim W_G(x) &= \lim W_G(a_r) \\ &= \lim \left\{ \sum_{i \neq r} w_i \, [k(a_r - a_i)]^{K_i} \right\} \\ &= \sum_{i \neq r} w_i \, . \end{aligned} \tag{7.1.8}$$

It follows that

$$\min_x \left\{ \lim W_G(x) \right\} = \sum_{i \neq s} w_i \; , \tag{7.1.9}$$

which occurs at $x^* = a_s$.

As $K_i \to 0^+$, $i = 1,...,n$, we see from (7.1.7) that there is a flattening effect on the objective function. The cost component associated with each destination a_i becomes insensitive to the distance travelled, because of the extreme economies of scale resulting from the low values of the K_i. We also note from (7.1.7) and (7.1.8) that the cusps at the fixed points (see Figure 2.4.1(c)) become more pronounced as the K_i are decreased. This fact is recognized by Morris (1981) for the case where k is the ℓ_p norm, and illustrated in his Figure 2 with a one-dimensional example. We see now that the optimal solution tends to be for sufficiently small values of the K_i at the destination with the largest weight.

If all the $K_i = 1$, we return to the standard minisum problem,

$$\text{minimize } W(x) = \sum_{i=1}^{n} w_i \, k(x - a_i) \, .$$

For this model, a fixed point a_r will be optimal if, and only if, the following criterion proven by Juel and Love (1981) is satisfied:

$$w_r \geq k^\circ \left[\sum_{i \neq r} w_i \nabla k(a_r - a_i) \right] , \qquad (7.1.10)$$

where k° denotes the polar of the norm k, $k(x - a_i)$ is differentiable at $x = a_r$ for all $i \neq r$, and $\nabla k(a_r - a_i)$ denotes the gradient of $k(x)$ evaluated at $x = a_r - a_i$ (or alternatively, the gradient of $f_i(x) = k(x - a_i)$ evaluated at $x = a_r$) for all $i \neq r$. The above result derives from the fact that the directional derivative $W'(a_r ; y)$ must be greater than or equal to zero for all unit vector directions y, when $x = a_r$ is an optimal solution. (For further discussion, see the derivation of fixed point optimality criteria for the mixed-norm problem in Chapter 6.) Since W is a convex function of x, the requirement, $W'(a_r ; y) \geq 0, \forall y$, is both a necessary and sufficient condition for optimality at a_r (Property 2.4.1). As a comparison, note that $W_G'(a_r ; y) = +\infty, \forall y$, if $0 < K_r < 1$ (equation (7.1.6)). But since $W_G(x)$ is not convex, we can only conclude that a_r is a local minimum.

The fixed point optimality criterion in (7.1.10) for the standard minisum problem shows that if the weight w_r is sufficiently large relative to the weights and geometry of the other fixed points, then $x^* = a_r$. In fact, by the majority theorem of Witzgall (1964), a_r is guaranteed to be an optimal solution if

$$w_r \geq \frac{1}{2} \sum_{i=1}^{n} w_i .$$

However, the optimality criterion in (7.1.10) is often satisfied in practice at a much lower value of w_r, as seen in the examples given by Juel and Love (1981).

Consider once again the mixed-power model (7.1) in which $K_i \geq 1$ for all i, with at least one of these inequalities satisfied strictly. Furthermore, assume that $J = \{i \mid K_i = 1\}$ is a non-empty set. Then, optimality criteria at the fixed points a_r, where $r \in J$, can be derived in a similar manner as for the standard minisum problem. First we calculate the directional derivative of W_G at a_r in the direction y.

$$W'_G(a_r\,;y) = \sum_{i \neq r} w_i h'_i(a_r - a_i\,; y) + w_r k(y)$$

$$= y^T.\left| \sum_{i \neq r} w_i K_i [k(a_r - a_i)]^{K_i - 1} \nabla k(a_r - a_i) \right| + w_r k(y) \,, \qquad (7.1.11)$$

where equation (2.4.6) is used in the first step, and we assume that $k\,(x - a_i)$ is differentiable at $x = a_r, \forall\, i \neq r$, and use equation (2.4.11) in the second step. A necessary condition for an optimal solution to occur at a_r is that $W_G'\,(a_r\,; y) \geq 0$ for all directions y. Letting

$$v_i = w_i K_i [k(a_r - a_i)]^{K_i - 1} \,, \forall\, i \neq r \,, \qquad (7.1.12)$$

we see that this implies

$$y^T.\left| \sum_{i \neq r} v_i \nabla k(a_r - a_i) \right| + w_r k(y) \geq 0 \,, \forall\, y \,,$$

or

$$w_r \geq \frac{-y^T.\left| \sum_{i \neq r} v_i \nabla k(a_r - a_i) \right|}{k(y)} \,, \forall\, y \,. \qquad (7.1.13)$$

Substituting $z = -y$, and noting that $k(-z) = k(z)$, we can rewrite (7.1.13) as

$$w_r \geq \max_z \left\{ \frac{z^T.\left| \sum_{i \neq r} v_i \nabla k(a_r - a_i) \right|}{k(z)} \right\}$$

$$= k^\circ \left| \sum_{i \neq r} v_i \nabla k(a_r - a_i) \right| \,, \qquad (7.1.14)$$

by definition of the polar. Since W_G is a convex function of x when all the $K_i \geq 1$, it follows that (7.1.14) is both a necessary and sufficient condition for an optimal solution to occur at a_r, for any $r \in J$. Comparing (7.1.14) with (7.1.10), it is interesting to note that the optimality criterion at any fixed point a_r, $r \in J$, corresponds to the one for the standard minisum problem with adjusted weights v_i used in place of the w_i for $i \neq r$.

Thus far we have observed that a_r is a local minimum of W_G if $0 < K_r < 1$, and that the criterion (7.1.14) can be used to test for the optimality of a_r when $K_r = 1$. If

$K_i \geq 1, \forall i$, then this criterion is both a necessary and sufficient condition for a global minimum at a_r. However, if any of the K_i are in the interval (0,1), then (7.1.14) is a necessary condition but not a sufficient one, since W_G is no longer a convex function of x. To complete this topic, consider now a fixed point a_r with $K_r > 1$. We prove below that the directional derivative $W_G'(a_r; y)$ is independent of the weight w_r in this case. Thus, the optimality criterion at a_r ($W_G'(a_r; y) \geq 0$ for all y) is unaffected by any increase in w_r. We conclude that fixed point optimality criteria are not relevant at any destinations a_r with power $K_r > 1$.

Property 7.1.5

Consider model (7.1) where $K_r > 1$ for some $r \in \{1, ..., n\}$. Then the directional derivative of W_G evaluated at the fixed point a_r in any direction y is independent of the weight w_r.

Proof:

$$W'_G(a_r; y) = \sum_{i \neq r} w_i h'_i(a_r - a_i; y) + w_r h'_r(0; y).$$

But $h'_r(0; y) = 0$, by equation (2.4.12) with $t = K_r > 1$ and $h(x) \equiv h_r(x)$. Thus

$$W'_G(a_r; y) = \sum_{i \neq r} w_i h'_i(a_r - a_i; y), \tag{7.1.15}$$

which is independent of w_r.

In general, a direction y can be found such that $W'_G(a_r; y) < 0$ in (7.1.15). As an example, assume that $k(x - a_i)$ is differentiable at $x = a_r$ for all $i \neq r$. Then (7.1.15) becomes

$$W'_G(a_r; y) = y^T \cdot \left| \sum_{i \neq r} v_i \nabla k(a_r - a_i) \right|,$$

where v_i is defined in (7.1.12). But

$$\sum_{i \neq r} v_i \nabla k(a_r - a_i) = \nabla W_{Gr}(a_r), \tag{7.1.16}$$

where

$$W_{Gr}(x) = \sum_{i \neq r} w_i [k(x - a_i)]^{K_i} , \tag{7.1.17}$$

and thus,

$$W'_G(a_r ; y) = y^T \cdot \nabla W_{Gr}(a_r), \quad K_r > 1 . \tag{7.1.18}$$

Except for the special case where $\nabla W_{Gr}(a_r) = 0$, we can choose y to be the steepest descent direction; i.e.,

$$y = - \nabla W_{Gr}(a_r) / \|\nabla W_{Gr}(a_r)\| , \tag{7.1.19}$$

so that

$$W'_G(a_r ; y) = - \|\nabla W_{Gr}(a_r)\| < 0 . \tag{7.1.20}$$

It follows that the fixed point a_r, with $K_r > 1$, is never an optimal solution, except for very special choices of the weights and geometry of the other destinations.

In Chapter 5 we showed that any optimal solution of the Weber problem in R^2 must lie in the convex hull of the fixed points, when the distance function is a round norm (Property 5.1.4). We also showed that this result holds in R^N when the Euclidean norm is used (Property 5.1.5). Surprisingly enough these localization results extend to a general class of minisum problems, of which model (7.1) is a specific case. The key observation here is that an optimal solution of the general problem also solves a related Weber problem.

Theorem 7.1.1

Consider the following minisum model,

$$\text{minimize } W_g(x) = \sum_{i=1}^{n} w_i \, g_i \, (k(x - a_i)) , \tag{7.1.21}$$

where $g_i(u)$ is an increasing differentiable function of u in the interval $[0, +\infty)$, $i = 1, \ldots, n$, and k is a round norm on R^2. Then any optimal solution must lie in the convex hull of the fixed points.

Proof:

Let x^* denote an optimal solution. If x^* is a fixed point, it is automatically within the convex hull. Therefore, we only need to consider the case, $x^* \neq a_i$, $i = 1, ..., n$.

If k is a nondifferentiable round norm, the objective function W_g may not be differentiable at x^*. Hence, in order not to lose generality, we need to consider the directional derivative, $W'_g(x^* ; y)$. The necessary condition for a local minimum must be satisfied at x^*; i.e.,

$$\min_y \; W'_g(x^* ; y) \geq 0 . \tag{7.1.22}$$

Using equation (2.4.9), we see that

$$W'_g(x^* ; y) = \sum_{i=1}^{n} w_i \, g'_i \, (k(x^* - a_i)) \, k'(x^* - a_i ; y) , \tag{7.1.23}$$

where $g'_i(u) = dg_i(u)/du$, $i = 1, ..., n$. Now let

$$v_i = w_i \, g'_i \, (k(x^* - a_i)) \;, \quad i = 1, ..., n . \tag{7.1.24}$$

Since the g_i are increasing functions, then $g'_i(k(x^* - a_i)) > 0$ for all i, so that

$$v_i > 0 \;, \quad i = 1, ..., n . \tag{7.1.25}$$

Combining (7.1.22), (7.1.23) and (7.1.24), it follows that

$$\min_y \left| \sum_{i=1}^{n} v_i \; k'(x^* - a_i ; y) \right| \geq 0 . \tag{7.1.26}$$

Consider the following Weber problem,

$$\text{minimize } W(x) = \sum_{i=1}^{n} v_i \, k(x - a_i) , \tag{7.1.27}$$

where the v_i are adjusted positive weights defined in (7.1.24). By the inequality (7.1.26), we conclude that x^* is also an optimal solution of this related problem. But all optimal solutions of (7.1.27) must lie in the convex hull of the fixed points, by Property 5.1.4. Hence, $x^* \in \text{c.h.}\{a_1, ..., a_n\}$, ending the proof.

In model (7.1), $g_i(u) = u^{K_i}$ with $K_i > 0$, $i = 1, ..., n$. Since these are increasing, differentiable functions of u in the interval $[0, +\infty)$, we immediately obtain the following result.

Corollary 7.1.1

Let k be a round norm on R^2 in model (7.1). Then all optimal solutions must lie in the convex hull of the fixed points.

The preceding theorem and its corollary apply in N-dimensional space when k is the Euclidean norm, as shown next.

Theorem 7.1.2

Consider the following minisum model,

$$\text{minimize } W_g(x) = \sum_{i=1}^{n} w_i\, g_i(\ell_2(x - a_i)), \tag{7.1.28}$$

where $g_i(u)$ is an increasing differentiable function of u in the interval $[0, +\infty)$, $i = 1, ..., n$, and ℓ_2 is the Euclidean norm on R^N. Then any optimal solution must lie in the convex hull of the fixed points.

Proof:

As in the preceding theorem, we only need to consider the case $x^* \neq a_i$, $i = 1, ..., n$. The directional derivative of W_g at x^* can be calculated, with the proof proceeding in a similar manner as before. However, since the Euclidean norm is a differentiable round norm (Property 2.2.4), it follows that W_g is differentiable at x^*, and we can use instead the first-order necessary condition,

$$\nabla W_g(x^*) = 0 . \tag{7.1.29}$$

But

$$\nabla W_g(x^*) = \sum_{i=1}^{n} v_i \, \nabla \ell_2(x^* - a_i), \tag{7.1.30}$$

where the v_i are adjusted positive weights defined in (7.1.24) with $k \equiv \ell_2$. Thus

$$\sum_{i=1}^{n} v_i \, \nabla \ell_2(x^* - a_i) = 0 \; . \tag{7.1.31}$$

It follows that x^* is also an optimal solution of the related Weber problem,

$$\text{minimize } W(x) = \sum_{i=1}^{n} v_i \; \ell_2(x - a_i) \; .$$

Therefore, by Property 5.1.5, $x^* \in \text{c.h.}\{a_1, \ldots, a_n\}$.

The next result is immediately obvious.

Corollary 7.1.2

Let k be the Euclidean norm on R^N in model (7.1) (or alternatively, $p = 2$ in model (7.2)). Then all optimal solutions must lie in the convex hull of the fixed points.

The preceding localization theorems require that the g_i be increasing differentiable functions. This is not restrictive in a practical sense, since we normally expect costs to increase with distance travelled, and any function can always be approximated by a differentiable one to the degree of accuracy desired. It is also interesting to note that the objective function (W_g) does not have to be a convex function of x since the g_i are not restricted in this manner, and yet all optimal solutions will be in the convex hull. We can go one step further, using the same reasoning as in the theorems, to observe that all local minima of W_g lie in the convex hull.

7.2 Applications with the ℓ_p Norm

In this section, we investigate the mixed-power problem (7.2), where k is now the ℓ_p norm. Letting $K_i = p/s_i$, $s_i > 0$, $i = 1, \ldots, n$, model (7.2) can be rewritten in the form,

$$\text{minimize } W_{\bar{G}}(x) = \sum_{i=1}^{n} w_i \, [\ell_p(x-a_i)]^{p/s_i}$$

$$= \sum_{i=1}^{n} w_i \, \ell_{p,s_i}(x-a_i) \; ; \qquad (7.2.1)$$

where

$$\ell_{p,s}(x-a_i) = \left| \sum_{j=1}^{N} |x_j - a_{ij}|^p \right|^{\frac{1}{s}}, \qquad s > 0, \qquad (7.2.2)$$

is a distance function on R^N first introduced by Love and Morris (1972, 1979) for $N = 2$.

We initially calculate the first and second-order partial derivatives of the $\ell_{p,s}$ function, since these will be required in the subsequent analysis. Using standard calculus, we obtain after some re-arranging the following results at points x where the derivatives are defined:

$$\frac{\partial}{\partial x_j} \ell_{p,s}(x) = \frac{p}{s} [\ell_p(x)]^{\frac{p(1-s)}{s}} \operatorname{sign}(x_j) |x_j|^{p-1}, \qquad j = 1, \ldots, N; \qquad (7.2.3)$$

$$\frac{\partial^2}{\partial x_j^2} \ell_{p,s}(x) = \frac{p}{s} [\ell_p(x)]^{\frac{p(1-2s)}{s}} |x_j|^{p-2} \left\{ (p-1)[\ell_p(x)]^p + \frac{p}{s}(1-s)|x_j|^p \right\}, \qquad (7.2.4)$$

$$j = 1, \ldots, N;$$

and

$$\frac{\partial^2}{\partial x_j \partial x_k} \ell_{p,s}(x) = \frac{p^2}{s^2}(1-s)[\ell_p(x)]^{\frac{p(1-2s)}{s}} \operatorname{sign}(x_j) \operatorname{sign}(x_k) |x_j|^{p-1} |x_k|^{p-1}, \qquad (7.2.5)$$

$$j, k = 1, \ldots, N, \quad j \neq k.$$

From Property 2.3.5, it follows that $\ell_{p,s}$ is a strictly convex function of x when p > 1 and p/s > 1 (p − s > 0). However, if p = 1 and 1/s > 1, then $\ell_{1,s}(x)$ is convex but not strictly so, by Property 2.3.6. It is instructive to verify these results using the Hessian matrix of second-order derivatives given in (7.2.4) and (7.2.5). This is done below for the two-dimensional case (N = 2), thereby extending the convexity proof of El-Shaieb (1978, Theorem 1) for p = 2 to general values of p ≥ 1.

We consider first the second-order derivatives $\partial^2\ell_{p,s}(x)/\partial x_j^2$. From (7.2.4), it follows that

$$\frac{\partial^2}{\partial x_j^2}\ell_{p,s}(x) = B_j(x)\cdot A_j(x),$$

where

$$\begin{aligned} B_j(x) &= \frac{p}{s}\,[\ell_p(x)]^{\frac{p(1-2s)}{s}}\,|x_j|^{p-2} \\ &> 0, \quad \text{if } x_j \neq 0, \end{aligned} \tag{7.2.6}$$

and

$$\begin{aligned} A_j(x) &= (p-1)[\ell_p(x)]^p + \frac{p}{s}(1-s)|x_j|^p \\ &= (p-1)\sum_{t=1}^{N}|x_t|^p + \frac{p}{s}(1-s)|x_j|^p \\ &= (p-1)\sum_{t\neq j}|x_t|^p + \frac{(p-s)}{s}|x_j|^p \\ &> 0, \quad \text{if } x_j \neq 0, \end{aligned} \tag{7.2.7}$$

since p − s > 0. Hence, if p ≥ 1 and p/s > 1, then

$$\frac{\partial^2}{\partial x_j^2}\ell_{p,s}(x) > 0, \qquad x_j \neq 0, \qquad j = 1, \ldots, N. \tag{7.2.8}$$

The Hessian matrix for N = 2 is given by

$$H_2(x\,;p,s) = \begin{bmatrix} \dfrac{\partial^2}{\partial x_1^2}\,\ell_{p,s}(x) & \dfrac{\partial^2}{\partial x_1 \partial x_2}\,\ell_{p,s}(x) \\ \\ \dfrac{\partial^2}{\partial x_2 \partial x_1}\,\ell_{p,s}(x) & \dfrac{\partial^2}{\partial x_2^2}\,\ell_{p,s}(x) \end{bmatrix}, \tag{7.2.9}$$

where

$$\ell_{p,s}(x) = [\,|x_1|^p + |x_2|^p]^{1/s} .$$

The determinant of H_2 is readily obtained using (7.2.4) and (7.2.5). After a straightforward computation, we get

$$\det(H_2) = \frac{p^2}{s^3}(p-1)(p-s)[\,\ell_p(x)]^{\frac{2p(1-s)}{s}}\,|x_1|^{p-2}|x_2|^{p-2} . \tag{7.2.10}$$

For any x with $x_1 \neq 0$ and $x_2 \neq 0$, it follows that

$$\det(H_2) > 0\,, \qquad \text{if } p > 1\,, \tag{7.2.11}$$

since $p-s > 0$, and

$$\det(H_2) = 0\,, \qquad \text{if } p = 1\,. \tag{7.2.12}$$

From (7.2.8), (7.2.11) and the first-order differentiability of $\ell_{p,s}(x)$ for all x (Property 2.4.2), we conclude that $\ell_{p,s}$ is a strictly convex function of x if $p > 1$ and $p/s > 1$. However, since $\det(H_2) = 0$ if $p = 1$, it follows that $\ell_{1,s}$ is convex but not in a strict sense, if $1/s > 1$. Thus, Properties 2.3.5 and 2.3.6 are verified by means of the Hessian matrix for the case where k is the ℓ_p norm on R^2.

If $p > 1$ (and finite), and the location problem is in R^2, we can apply Corollary 7.1.1 to obtain the result that any optimal solution of model (7.2) must lie in the convex hull of the fixed points. Consider now the case where $p = 1$, so that the cost components contain rectangular distances raised to a power. Using (7.2.1) and (7.2.2), the model becomes

$$\text{minimize } W_G(x) = \sum_{i=1}^{n} w_i\, \ell_{1,s_i}(x - a_i)\,, \tag{7.2.13}$$

where

$$\ell_{1,s}(x) = \left[\sum_{j=1}^{N} |x_j| \right]^{1/s}, \qquad s > 0. \tag{7.2.14}$$

We shall soon see that if $s_i \geq 1$ (or $K_i \leq 1$) for all i with at least one of these inequalities satisfied strictly, then any optimal solution must occur at one of a finite number of intersection or fixed point locations.

The first and second-order partial derivatives of $\ell_{1,s}$ are obtained directly from (7.2.3), (7.2.4) and (7.2.5) with p = 1. Thus, at points x where the derivatives are defined we have

$$\frac{\partial}{\partial x_j} \ell_{1,s}(x) = \frac{1}{s} [\ell_i(x)]^{\frac{(1-s)}{s}} \operatorname{sign}(x_j), \qquad j = 1, \ldots, N; \tag{7.2.15}$$

$$\frac{\partial^2}{\partial x_j^2} \ell_{1,s}(x) = \frac{(1-s)}{s^2} [\ell_1(x)]^{\frac{(1-2s)}{s}}, \qquad j = 1, \ldots, N; \tag{7.2.16}$$

and

$$\frac{\partial^2}{\partial x_j \partial x_k} \ell_{1,s}(x) = \frac{(1-s)}{s^2} [\ell_i(x)]^{\frac{(1-2s)}{s}} \operatorname{sign}(x_j) \operatorname{sign}(x_k), \tag{7.2.17}$$

$$j, k = 1, \ldots, N, \qquad j \neq k.$$

It is important to note that the first partial derivative in (7.2.15) is undefined on the hyperplane $x_j = 0$, since $\operatorname{sign}(x_j)$ is undefined here. (An exception occurs at $x = 0$ when $s < 1$, in which case $\partial \ell_{1,s}(0)/\partial x_j = 0$.) Hence, higher-order derivatives in which $\partial/\partial x_j$ appears at least once are also undefined on the hyperplane $x_j = 0$.

We are now ready to prove the following useful result.

Property 7.2.1

Consider the minisum problem given in (7.2.13), where the $\ell_{1_{s_i}}$ are distance functions on R^N. If $s_i \geq 1$ for all i with at least one of these inequalities satisfied strictly, then any optimal solution must occur at an intersection point or a fixed point.

Proof:

Suppose that a local minimum occurs at an $x^o = (x_1{}^o, \ldots, x_N{}^o)^T$ which is not an intersection point or a fixed point. Then one or more indices $r \in \{1, \ldots, N\}$ exist such that

$$x_r^0 \neq a_{ir}, \qquad i = 1, \ldots, n. \tag{7.2.18}$$

Hence, the objective function W_G is infinitely differentiable in the x_r direction at x^o, and furthermore using (7.2.16),

$$\frac{\partial^2}{\partial x_r^2} W_G(x^0) = \sum_{i=1}^{n} w_i \frac{(1-s_i)}{s_i^2} [\ell_1(x^0 - a_i)]^{\frac{(1-2s_i)}{s_i}}$$

$$< 0, \tag{7.2.19}$$

since $s_i \geq 1, \forall i$, with at least one $s_i > 1$. This implies that W_G is strictly concave in the x_r direction at x^o, which contradicts the supposition that x^o is a local minimum. We conclude that all local minima of W_G must occur at an intersection point or fixed point, and hence any global solution occurs here as well.

The key observation in the preceding proof is that for any x which is not an intersection or fixed point, the objective function is infinitely differentiable in at least one direction. This concept is extendable to the minisum problem involving functions of a general block norm. In Theorem 6 of Thisse, Ward and Wendell (1984), a different method is used to prove that an optimal solution can always be found at an intersection or fixed point, when the cost components being summed are non-decreasing concave functions of a block norm, and

$N = 2$. If in addition these functions are strictly concave, the authors show that the optimal solutions can only occur at the fixed points and the intersection points contained in the convex hull of the fixed points.

Suppose now that $0 < p < 1$, so that the objective function in (7.2.1) is a weighted sum of hyper-rectilinear distances raised to different powers. Juel and Love (1985) prove that for the standard Weber problem with hyper-rectilinear distances, an optimal solution always occurs at an intersection or fixed point. They also show that the optimal location may not lie within the convex hull of the existing facilities, even when $N = 2$. We extend this intersection point property to the mixed-power problem, for a certain range of values of the parameters s_i. The proof is analogous to the one for Property 7.2.1.

Property 7.2.2

Consider the minisum problem in (7.2.1), where the ℓ_{p,s_i} are distance functions on R^N. If $0 < p < 1$, and $s_i \geq p$ for all i, with at least one of the $s_i > p$, then any optimal solution must coincide with an intersection point or a fixed point.

Proof:

Suppose that a local minimum occurs at an $x^o = (x_1{}^o, \ldots, x_N{}^o)^T$ which is not an intersection point or a fixed point. Then $x_r{}^o \neq a_{ir}$, $i = 1, \ldots, n$, for at least one $r \in \{1, \ldots, N\}$. It follows that W_G is infinitely differentiable at x^o in the x_r direction, and using (7.2.4) and (7.2.7),

$$\frac{\partial^2}{\partial x_r^2} W_G(x^0) = \sum_{i=1}^{n} \frac{w_i p}{s_i} [\ell_p(x^0 - a_i)]^{\frac{p(1-2s_i)}{s_i}} |x_r^0 - a_{ir}|^{p-2} A_{ir}(x^0 - a_i), \tag{7.2.20}$$

where

$$A_{ir}(x^0 - a_i) = (p-1)\sum_{t\neq r}|x_t^0 - a_{it}|^p + \frac{(p-s_i)}{s_i}|x_r^0 - a_{ir}|^p$$

$$\leq \frac{(p-s_i)}{s_i}|x_r^0 - a_{ir}|^p \qquad (p<1)$$

$$\leq \; 0\,, \qquad i = 1,\ldots,N \qquad (s_i \geq p)\,. \tag{7.2.21}$$

Since one or more of the $s_i > p$, there exists at least one $q \in \{1, \ldots, n\}$ such that $A_{qr}(x^o - a_q) < 0$. It follows that

$$\frac{\partial^2}{\partial x_r^2} W_G(x^o) < 0\,, \tag{7.2.22}$$

and using the same reasoning as in the preceding property, we conclude that any optimal solution occurs at an intersection point or fixed point.

If $s_i = p$, $i = 1, \ldots, n$, we return to the standard Weber problem. Note that in this case,

$$A_{ir}(x^0 - a_i) = (p-1)\sum_{t\neq r}|x_t^0 - a_{it}|^p\,, \qquad i = 1,\ldots,n\,. \tag{7.2.23}$$

For $p < 1$, we have $A_{ir}(x^o - a_i) \leq 0$ for all i. Furthermore, except for the trivial problem where the fixed points all lie on a straight line parallel to the x_r-axis, there must be again at least one $q \in \{1, \ldots, n\}$ such that $A_{qr}(x^o - a_q) < 0$. Hence, (7.2.22) applies here as well, and we conclude the following important result.

Property 7.2.3

If $s_i = p$, $\forall\, i$, $0 < p < 1$, and the a_i do not all lie on a straight line parallel to one of the axes, then any optimal solution of the minisum problem in (7.2.1) coincides with an intersection point or fixed point.

Juel and Love (1985, Property 3) show that an optimal location for the preceding problem can always be found at an intersection or fixed point. Property 7.2.3 gives a stronger result; we see now that an optimal solution cannot exist elsewhere.

APPENDIX A

(To be read in conjunction with Property 3.2.4)

Property A.1

Let

$$G(p) = \frac{2\,H^2(p) + H'(p)}{H^2(p) + H'(p)} \,, \qquad (A.1)$$

where H(p) is defined in equation (3.2.16), with f_1 and f_2 denoting positive constants. Then G(p) is a decreasing function of p in the interval $(0, +\infty)$.

Proof:

To prove that G(p) is a decreasing function of $p \in (0, +\infty)$, it suffices to show that $G'(p) < 0$. Rewriting (A.1) in the form,

$$G(p) = 1 + \frac{H^2}{H^2 + H'} \,,$$

we obtain using standard calculus,

$$G'(p) = \frac{2HH'}{H^2 + H'} - \frac{H^2}{(H^2 + H')^2} \cdot (2HH' + H'')$$

$$= \frac{2H(H')^2 - H^2H''}{(H^2 + H')^2} \,. \qquad (A.2)$$

Since f_1 and f_2 are both positive constants, it immediately follows from (3.2.16) that

$$H(p) < 0 \,, \quad \forall \ p \in (0, +\infty) \,. \qquad (A.3)$$

Thus, $G'(p) < 0$ if, and only if,

$$2(H')^2 - HH'' > 0 \,. \qquad (A.4)$$

Without loss in generality, assume that $f_1 \geq f_2$. Returning to equation (3.2.16), we have

$$H(p) = \frac{1}{p^2}\left| \frac{1}{1+u^p}\, \ell n\left(\frac{1}{1+u^p}\right) + \frac{u^p}{1+u^p}\, \ell n\left(\frac{u^p}{1+u^p}\right)\right| , \tag{A.5}$$

where

$$0 < u = f_2/f_1 \leq 1 . \tag{A.6}$$

Letting

$$\phi(p) := 1 + u^p , \tag{A.7}$$

and noting that

$$\phi'(p) = u^p\, \ell n\, u , \tag{A.8}$$

equation (A.5) can be written as

$$H = -\frac{\ell n\, \phi}{p^2} + \frac{1}{p\phi}\phi' . \tag{A.9}$$

Again using standard calculus, we obtain

$$H' = \frac{2\ell n\, \phi}{p^3} - \frac{2\phi'}{p^2\phi} - \frac{1}{p\phi^2}(\phi')^2 + \frac{1}{p\phi}\phi'' . \tag{A.10}$$

and

$$H'' = -\frac{6\ell n\, \phi}{p^4} + \frac{6}{p^3\phi}\phi' + \frac{3}{p^2\phi^2}(\phi')^2 - \frac{3}{p^2\phi}\phi''$$

$$- \frac{3}{p\phi^2}\phi'\phi'' + \frac{2}{p\phi^3}(\phi')^3 + \frac{1}{p\phi}\phi''' . \tag{A.11}$$

With equations (A.10) and (A.11), the left-hand side of (A.4) becomes after some simplification and re-arranging:

$$2(H')^2 - HH'' = \sum_{i=1}^{5} S_i , \tag{A.12}$$

where

$$S_1 = -\frac{5(\phi')^2 \ell n\,\phi}{p^4\phi^2} + \frac{5\phi'' \ell n\,\phi}{p^4\phi} + \frac{5(\phi')^3}{p^3\phi^3} - \frac{5\phi'\phi''}{p^3\phi^2}\,, \tag{A.13a}$$

$$S_2 = -\frac{(\phi')^2\phi''}{p^2\phi^3} + \frac{2(\phi'')^2}{p^2\phi^2}\,, \tag{A.13b}$$

$$S_3 = -\frac{3\phi'\phi'' \ell n\,\phi}{p^3\phi^2} + \frac{2(\phi')^3 \ell n\,\phi}{p^3\phi^3}\,, \tag{A.13c}$$

$$S_4 = -\frac{\phi'\phi'''}{p^2\phi^2} + \frac{\phi''' \ell n\,\phi}{p^3\phi}\,, \tag{A.13d}$$

and

$$S_5 = \frac{2(\ell n\,\phi)^2}{p^6} - \frac{4\phi' \ell n\,\phi}{p^5\phi} + \frac{2(\phi')^2}{p^4\phi^2}\,. \tag{A.13e}$$

Recalling equations (A.7) and (A.8), and also noting that

$$\phi'' = u^p(\ell n\ u)^2\,, \qquad \phi''' = u^p(\ell n\ u)^3\,, \tag{A.14}$$

we can rewrite the S_i in terms of p and the constant u as follows:

$$S_1 = \frac{5u^p}{p^3(1+u^p)^3}\left|\frac{(1+u^p)\,\ell n(1+u^p)\,(\ell n\,u)^2}{p} - u^p(\ell n\,u)^3\right|, \tag{A.15a}$$

$$S_2 = \frac{u^{2p}(\ell n\,u)^4(2+u^p)}{p^2(1+u^p)^3}\,. \tag{A.15b}$$

$$S_3 = -\frac{u^{2p}\ell n(1+u^p)(\ell n\,u)^3(3+u^p)}{p^3(1+u^p)^3}\,, \tag{A.15c}$$

$$S_4 = \frac{u^p(\ell n\,u)^3}{p^2(1+u^p)^2}\left|-u^p\,\ell n\,u + \frac{(1+u^p)\,\ell n(1+u^p)}{p}\right|, \tag{A.15d}$$

and

$$S_5 = \frac{2}{p^4}\left|\frac{\ell n(1+u^p)}{p} - \frac{u^p\,\ell n\,u}{1+u^p}\right|^2. \tag{A.15e}$$

Since $0 < u \leq 1$ from relation (A.6), therefore $\ell n\,u \leq 0$ and $\ell n(1+u^p) > 0$. Hence, it is readily seen that,

$$S_j \geq 0\,, \qquad j = 1,2,3\,, \tag{A.16a}$$

$$S_4 \leq 0\,, \tag{A.16b}$$

and

$$S_5 > 0\,. \tag{A.16c}$$

In order to prove that

$$\Sigma_{i=1}^{5}\, S_i > 0\,,$$

it is thus sufficient to show that

$$\Sigma_{i=1}^{4}\, S_i \geq 0\,.$$

After a series of straightforward algebraic steps, we obtain

$$S_2 + S_3 + S_4 = \frac{1}{p^3(1+u^p)^3}\,[p\,u^{2p}(\ell n\,u)^4 - (u^{2p} - u^p)\,(\ell n\,u)^3\,\ell n(1+u^p)]\,. \tag{A.17}$$

Then, using the inequality,

$$x > \ell n(1+x)\,, \qquad \forall\, x > 0\,,$$

and deleting non-negative terms, it follows that

$$\begin{aligned} S_2 + S_3 + S_4 &\geq \frac{u^p\,(\ell n\,u)^3\,\ell n(1+u^p)}{p^3(1+u^p)^3} \\ &\geq \frac{u^{2p}\,(\ell n\,u)^3}{p^3(1+u^p)^3}\,. \end{aligned} \tag{A.18}$$

Combining (A.15a) and (A.18), we now see that

$$\begin{aligned} \Sigma_{i=1}^{4}\, S_i &\geq \frac{5u^p\,\ell n(1+u^p)(\ell n\,u)^2}{p^4(1+u^p)^2} - \frac{4u^{2p}\,(\ell n\,u)^3}{p^3(1+u^p)^3} \\ &\geq 0\,. \end{aligned} \tag{A.19}$$

Combining (A.16c) and (A.19) gives

$$\Sigma_{i=1}^{5}\, S_i > 0\,.$$

Thus, the inequality (A.4) is satisfied, ending the proof.

BIBLIOGRAPHY

Aly, A.A., Kay, D.C. and Litwhiler, D.W., (1979), "Location Dominance on Spherical Surfaces," Operations Research, Vol.27, No. 5, pp.972-981.

Aly, A.A. and White, J.A., (1978), "Probabilistic Formulations of the Multifacility Weber Problem," Naval Research Logistics Quarterly, Vol.25, No.3, pp.531-547.

Baxter, J., (1981), "Local Optima Avoidance in Depot Location," Journal of the Operational Research Society, Vol.32, No.9, pp.815-819.

Baxter, J., (1984), "Depot Location: A Technique for the Avoidance of Local Optima," European Journal of Operational Research, Vol.18, No.2, pp.208-214.

Beckenbach, E.F., (1946), "An Inequality of Jensen," American Mathematical Monthly, Vol.53, pp. 501-505.

Beckenbach, E.F. and Bellman, R., (1961), An Introduction to Inequalities, Random House, New York.

Beckenbach, E.F. and Bellman, R., (1965), Inequalities, second revised printing, Springer-Verlag, Berlin.

Bellman, R., (1965), "An Application of Dynamic Programming to Location-Allocation Problems," SIAM Review, Vol.7, No.1, pp.126-128.

Belsley, D.A., Kuh, E. and Welsch, R.E., (1980), Regression Diagnostics, John Wiley and Sons, New York.

Berens, W. and Körling, F.-J., (1985), "Estimating Road Distances by Mathematical Functions," European Journal of Operational Research, Vol.21, No.1, pp.54-56.

Buchanan, D.J., (1988), "Interactive Computer Graphical Approaches to Some Maximin and Minimax Location Problems," Ph.D. Thesis, McMaster University, Hamilton.

Cabot, A.V., Francis, R.L. and Stary, M.A., (1970), "A Network Flow Solution to a Rectilinear Distance Facility Location Problem," AIIE Transactions, Vol.2, No.2, pp.132-141.

Calamai, P.H. and Conn, A.R., (1980), "A Stable Algorithm for Solving the Multifacility Location Problem Involving Euclidean Distances," SIAM Journal on Science, Statistics and Computers, Vol.1, No.4, pp.512-526.

Cavalier, T.M. and Sherali, H.D., (1986), "Euclidean Distance Location-Allocation Problems with Uniform Demands over Convex Polygons," Transportation Science, Vol.20, No.2, pp.107-116.

Chen, R., (1984a), "Location Problems with Costs Being Sums of Powers of Euclidean Distances," Computers & Operations Research, Vol.11, No.3, pp.285-294.

Chen, R., (1984b), "Solution of Location Problems with Radial Cost Functions," Computers & Mathematics with Applications, Vol.10, No.1, pp.87-94.

Christofides, N. and Eilon, S., (1969), "Expected Distances in Distribution Problems," Operational Research Quarterly, Vol.20, No.4, pp.437-443.

Cooper, L., (1963), "Location-Allocation Problems," Operations Research, Vol.11, No.3, pp.37-52.

Cooper, L., (1964), "Heuristic Methods for Location-Allocation Problems," SIAM Review, Vol.6, No.1, pp.37-53.

Cooper, L., (1967), "Solutions of Generalized Locational Equilibrium Models," Journal of Regional Science, Vol.7, No.1, pp.1-18.

Cooper, L., (1968), "An Extension of the Generalized Weber Problem," Journal of Regional Science, Vol.8, No.2, pp.181-197.

Cooper, L., (1972), "The Transportation-Location Problem," Operations Research, Vol.20, No.1, pp.94-108.

Cooper, L., (1973), "N-Dimensional Location Models: An Application to Cluster Analysis," Journal of Regional Science, Vol.13, No.1, pp.41-54.

Cooper, L., (1974), "A Random Locational Equilibrium Problem," Journal of Regional Science, Vol.14, No.1, pp.47-54.

Dahlquist, G. and Björck, Å., (1974), Numerical Methods, translated by Anderson, N., Prentice-Hall, Englewood Cliffs, New Jersey.

Domschke, W. and Drexl, A., (1985), Location and Layout Planning: An International Bibliography, Springer-Verlag, New York.

Drezner, Z., (1985), "Sensitivity Analysis of the Optimal Location of a Facility," Naval Research Logistics Quarterly, Vol.32, No.2, pp.209-224.

Drezner, Z. and Wesolowsky, G.O., (1978a), "A Trajectory Method for the Optimization of the Multi-Facility Location Problem with ℓ_p Distances," Management Science, Vol.24, No.14, pp.1507-1514.

Drezner, Z. and Wesolowsky, G.O., (1978b), "Facility Location on a Sphere," Journal of the Operational Research Society, Vol.29, No.10, pp.997-1004.

Drezner, Z. and Wesolowsky, G.O., (1980), "Optimal Location of a Facility Relative to Area Demands," Naval Research Logistics Quarterly, Vol.27, No.2, pp.199-206.

Durier, R. and Michelot, C., (1985), "Geometrical Properties of the Fermat-Weber Problem," European Journal of Operational Research, Vol.20, No.3, pp.332-343.

Eckhardt, U., (1975), "On an Optimization Problem Related to Minimal Surfaces with Obstacles," in Optimization and Optimal Control, Burlisch, R., Oettli, W. and Stoer, J. (eds.), Lecture Notes in Math., Vol.477, Springer-Verlag, Berlin, pp.95-101.

Eilon, S., Watson-Gandy, C.D.T. and Christofides, N., (1971), Distribution Management: Mathematical Modelling and Practical Analysis, Griffin, London.

El-Shaieb, A.M., (1978), "The Single Source Weber Problem – Survey and Extensions," Journal of the Operational Research Society, Vol.29, No. 5, pp. 469-476.

Erlenkotter, D., (1981), "A Comparative Study of Approaches to Dynamic Location Problems," European Journal of Operational Research, Vol. 6, No. 2, pp. 133-143.

Eyster, J., White, J. and Wierwille, W., (1973), "On Solving Multifacility Location Problems Using a Hyperboloid Approximation Procedure," AIIE Transactions, Vol.5, No.1, pp.1-6.

Finkbeiner, D.T., (1972), Elements of Linear Algebra, W.H. Freeman and Company, San Francisco.

Francis, R.L., (1967), "Some Aspects of a Minimax Location Problem," Operations Research, Vol.15, No.6, pp.1163-1169.

Francis, R.L. and Cabot, A., (1972), "Properties of a Multifacility Location Problem Involving Euclidean Distances," Naval Research Logistics Quarterly, Vol.19, No.2, pp.335-353.

Francis, R.L., McGinnis, L.F. and White, J.A., (1983), "Locational Analysis," European Journal of Operational Research, Vol.12, No.3, pp.220-252.

Francis, R.L. and White, J.A., (1974), Facility Layout and Location – An Analytical Approach, Prentice Hall, Englewood Cliffs, New Jersey.

Ginsburgh, V. and Hansen, P., (1974), "Procedures for the Reduction of Errors in Road Network Data," Operational Research Quarterly, Vol.25, No.2, pp.321-322.

Hageman, L.A. and Young, D.M., (1981), Applied Iterative Methods, Academic Press, New York.

Hansen, P., Peeters, D., Richard, D. and Thisse, J.-F., (1985), "The Minisum and Minimax Location Problems Revisited," Operations Research, Vol.33, No.6, pp.1251-1265.

Hansen, P., Peeters, D. and Thisse, J.-F., (1981), "Some Localization Theorems for a Constrained Weber Problem." Journal of Regional Science, Vol.21, No.1, pp.103-115.

Hansen, P., Peeters, D. and Thisse, J.-F., (1982), "An Algorithm for a Constrained Weber Problem," Management Science, Vol.28, No.11, pp.1285-1295.

Hansen, P., Peeters, D. and Thisse, J.-F., (1983), "Public Facility Location Models: A Selective Survey," in Thisse, J.-F. and Zoller, H.G. (eds.), Location Analysis of Public Facilities, North-Holland, Amsterdam.

Hansen, P., Perreur, J. and Thisse, J.-F., (1980), "Location Theory, Dominance and Convexity: Some Further Results," Operations Research, Vol.28, No.5, pp.1241-1250.

Hardy, G.H., Littlewood, J.E. and Pólya, G., (1952), Inequalities, 2nd ed., Cambridge University Press, Cambridge.

Harris, B., (1976), "Speeding Up Iterative Algorithms - The Generalized Weber Problem," Journal of Regional Science, Vol.16, No.3, pp.411-413.

Hodgson, M.J., Wong, R.T. and Honsaker, J., (1987), "The P-Centroid Problem on an Inclined Plane," Operations Research, Vol.35, No.2, pp.221-233.

Huriot, J.M. and Perreur, J., (1973), "On the Weber Problem with Rectangular Distance: A Comment," Management Science, Vol.20, No.3, pp.418-419.

Hurter, A.P., Schaefer, M.K. and Wendell, R.E., (1975), "Solutions of Constrained Location Problems," Management Science, Vol.22, No.1, pp.51-56.

Jacoby, S.L.S., Kowalik, J.S. and Pizzo, J.T., (1972), Iterative Methods for Nonlinear Optimization Problems, Prentice-Hall, Englewood Cliffs, New Jersey.

Juel, H., (1975), "Properties of Location Models," Operations Research Technical Report No.3, Graduate School of Business, University of Wisconsin, Madison, WI.

Juel, H., (1982), "A Note on Solving Multifacility Location Problems Involving Euclidean Distances," Naval Research Logistics Quarterly, Vol.29, No.1, pp.179-180.

Juel, H., (1983), "Coincident Optima for Two-Facility Weber Problems," Transportation Science, Vol.17, No.1, pp.110-113.

Juel, H. and Love, R.F., (1976), "An Efficient Computational Procedure for Solving the Multifacility Rectilinear Facilities Location Problem," Operational Research Quarterly, Vol.27, No.3, pp.697-703.

Juel, H. and Love, R.F., (1980), "Sufficient Conditions for Optimal Facility Locations to Coincide," Transportation Science, Vol.14, No.2, pp.125-129.

Juel, H. and Love, R.F., (1981), "Fixed Point Optimality Criteria for the Location Problem with Arbitrary Norms," Journal of the Operational Research Society, Vol.32, No.10, pp.891-897.

Juel, H. and Love, R.F., (1983), "Hull Properties in Location Problems," European Journal of Operational Research, Vol.12, No.3, pp.262-265.

Juel, H. and Love, R.F., (1985), "The Facility Location Problem for Hyper-Rectilinear Distances," AIIE Transactions, Vol.17, No.1, pp.94-98.

Katz, I.N., (1969), "On the Convergence of a Numerical Scheme for Solving Some Locational Equilibrium Problems," SIAM Journal of Applied Mathematics, Vol.17, No.6, pp.1224-1231.

Katz, I.N., (1974), "Local Convergence in Fermat's Problem," Mathematical Programming, Vol.6, No.1, pp.89-104.

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, Vol.23, No.4, pp.614-627.

Kreyszig, E., (1983), Advanced Engineering Mathematics, 5th ed., John Wiley and Sons, New York.

Kuhn, H.W., (1967), "On a Pair of Dual Nonlinear Problems," Chapter 3 in J. Abadie (ed), Nonlinear Programming, John Wiley and Sons, New York.

Kuhn, H.W., (1973), "A Note on Fermat's Problem," Mathematical Programming, Vol.4, No.1, pp.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, Vol.4, No.2, pp.21-34.

Love, R.F., (1968), "A Note on the Convexity of Siting Depots," The International Journal of Production Research, Vol.6, No.2, pp.153-154.

Love, R.F., (1969), "Locating Facilities in Three-Dimensional Space by Convex Programming," Naval Research Logistics Quarterly, Vol.16, No.4, pp.503-516.

Love, R.F., (1972), "A Computational Procedure for Optimally Locating a Facility with Respect to Several Rectangular Regions," Journal of Regional Science, Vol.12, No.2, pp.233-242.

Love, R.F., (1974), "The Dual of a Hyperbolic Approximation to the Generalized Constrained Multifacility Location Problem with ℓ_p Distances," Management Science, Vol.21, No.1, pp.22-33.

Love, R.F., (1976), "One-Dimensional Facility Location-Allocation Using Dynamic Programming," Management Science, Vol.22, No.5, pp.614-617.

Love, R.F. and Dowling, P., (1985), "Optimal Weighted ℓ_p Norm Parameters for Facilities Layout Distance Characterizations," Management Science, Vol.31, No.2, pp.200-206.

Love, R.F. and Juel, H., (1975), "Test Problems and Computational Results for Five Rectangular Location-Allocation Algorithms," Working Paper 4-75-13, Graduate School of Business, University of Wisconsin, Madison, WI.

Love, R.F. and Juel, H., (1982), "Properties and Solution Methods for Large Location-Allocation Problems," Journal of the Operational Research Society, Vol.33, No.5, pp.443-452.

Love, R.F. and Kraemer, S.A., (1973), "A Dual Decomposition Method for Minimizing Transportation Costs in Multi-Facility Location Problems," Transportation Science, Vol.7, No.4, pp.297-316.

Love, R.F. and Morris, J.G., (1972), "Modelling Inter-City Road Distances by Mathematical Functions," Operational Research Quarterly, Vol.23, No.1, pp.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, Vol.22, No.3, pp.441-453.

Love, R.F. and Morris, J.G., (1975b), "Solving Constrained Multifacility Location Problems Involving ℓ_p Distances Using Convex Programming," Operations Research, Vol.23, No.3, pp.581-587.

Love, R.F. and Morris, J.G., (1978), "On Convexity Proofs in Location Theory," Naval Research Logistics Quarterly, Vol.25, No.1, pp.179-181.

Love, R.F. and Morris, J.G., (1979), "Mathematical Models of Road Travel Distances," Management Science, Vol.25, No.2, pp.130-139.

Love, R.F. and Morris, J.G., (1988), "On Estimating Road Distances by Mathematical Functions," European Journal of Operational Research, Vol.36, No.2, pp.251-253.

Love, R.F., Morris, J.G. and Wesolowsky, G.O., (1988), Facilities Location: Models and Methods, North-Holland, New York.

Love, R.F., Truscott, W.G. and Walker, J.H., (1985), "Terminal Location Problem: A Case Study Supporting the Status Quo," Journal of the Operational Research Society, Vol.36, No.2, pp.131-136.

Love, R.F. and Yerex, L., (1976), "An Application of a Facilities Location Model in the Prestressed Concrete Industry," Interfaces, Vol.6, No.1, pp.45-49.

Luenberger, D.G., (1973), Introduction to Linear and Nonlinear Programming, Addison-Wesley, Reading.

Miehle, W., (1958), "Link-length Minimization in Networks," Operations Research, Vol.6, No.2, pp.232-243.

Morris, J.G., (1978), "Analysis of Generalized Empirical 'Distance' Function for Use in Location Problems," International Symposium on Locational Decisions, Banff, Alberta.

Morris, J.G., (1981), "Convergence of the Weiszfeld Algorithm for Weber Problems Using a Generalized 'Distance' Function," Operations Research, Vol.29, No.1, pp.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, Vol.27, No.6, pp.1180-1188.

Murtagh, B.A. and Niwattisyawong, S.R., (1982), "An Efficient Method for the Multi-Depot Location-Allocation Problem," Journal of the Operational Research Society, Vol.33, No.7, pp.629-634.

Neter, J., Wasserman, W. and Kutner, M.H., (1985), Applied Linear Statistical Models, 2nd ed., Irwin, Homewood, Illinois.

Ortega, J.M. and Rheinboldt, W.C., (1970), Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York.

Ostresh, L.M., (1975), "An Efficient Algorithm for Solving the Two Center Location-Allocation Problem," Journal of Regional Science, Vol. 15, No. 2, pp. 209-216.

Ostresh, L.M., (1977), "The Multifacility Location Problem: Applications and Descent Theorems," Journal of Regional Science, Vol. 17, No. 3, pp. 409-419.

Ostresh, L.M., (1978), "On the Convergence of a Class of Iterative Methods for Solving the Weber Location Problem," Operations Research, Vol. 26, No. 4, pp. 597-609.

Pelegrin, B., Michelot, C. and Plastria, F., (1985), "On the Uniqueness of Optimal Solutions in Continuous Location Theory," European Journal of Operational Research, Vol. 20, No. 3, pp. 327-331.

Perreur, J. and Thisse, J., (1974), "Central Metrics and Optimal Location," Journal of Regional Science, Vol. 14, No. 3, pp. 411-421.

Picard, J.-C. and Ratliff, H.D., (1978), "A Cut Approach to the Rectilinear Distance Facility Location Problem," Operations Research, Vol. 26, No. 3, pp. 422-433.

Planchart, A. and Hurter, A.P., (1975), "An Efficient Algorithm for the Solution of the Weber Problem with Mixed Norms," SIAM Journal on Control, Vol. 13, No. 3, pp. 650-665.

Plastria, F., (1984), "Localization in Single Facility Location," European Journal of Operational Research, Vol. 18, No. 2, pp. 215-219.

Plastria, F., (1987), "Solving General Continuous Single Facility Location Problems by Cutting Planes," European Journal of Operational Research, Vol. 29, No. 1, pp. 98-110.

Radó, F., (1988), "The Euclidean Multifacility Location Problem", Operations Research, Vol. 36, No. 3, pp. 485-492.

Riker, W.H. and Ordeshook, P.C., (1973), An Introduction to Positive Political Theory, Prentice-Hall, Englewood Cliffs, N.J.

Rockafellar, R.T., (1970), Convex Analysis, Princeton University Press, Princeton.

Schaefer, M.K. and Hurter, A.P., (1974), "An Algorithm for the Solution of a Location Problem with Metric Constraints," Naval Research Logistics Quarterly, Vol. 21, No. 4, pp. 625-636.

Schocker, A.D. and Srinivasan, V., (1974), "A Consumer-Based Methodology for the Identification of New Product Ideas," Management Science, Vol. 20, No. 6, pp. 921-937.

Shapiro, J.F., (1979), Mathematical Programming: Structures and Algorithms, John Wiley & Sons, New York.

Shields, P.C., (1969), Elementary Linear Algebra, Worth Publishers, New York.

Stephenson, G., (1966), Matrices, Sets and Groups: An Introduction for Students of Science and Engineering, American Elsevier Publishing Company, 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, Vol.32, No.6, pp.1309-1327.

Ward, J.E. and Wendell, R.E., (1980), "A New Norm for Measuring Distance Which Yields Linear Location Problems," Operations Research, Vol.28, No.3, Part II, pp.836-844.

Ward, J.E. and Wendell, R.E., (1985), "Using Block Norms for Location Modeling," Operations Research, Vol.33, No.5, pp.1074-1090.

Weiszfeld, E., (1937), "Sur le point lequel la somme des distances de n points donnés est minimum," Tohoku Mathematical Journal, Vol.43, pp.355-386.

Wendell, R.E. and Hurter, A.P., (1973), "Location Theory, Dominance, and Convexity," Operations Research, Vol.21, No.1, pp.314-320.

Wesolowsky, G.O., (1973), "Dynamic Facility Location," Management Science, Vol.19, No.11, pp.1241-1248.

Wesolowsky, G.O., (1977), "The Weber Problem with Rectangular Distances and Randomly Distributed Destinations," Journal of Regional Science, Vol.17, No.1, pp.53-60.

Wesolowsky, G.O. and Love, R.F., (1971a), "Location of Facilities with Rectangular Distances Among Point and Area Destinations," Naval Research Logistics Quarterly, Vol.18, No.1, pp.83-90.

Wesolowsky, G.O. and Love, R.F., (1971b), "The Optimal Location of New Facilities Using Rectangular Distances," Operations Research, Vol.19, No.1, pp.124-130.

Wesolowsky, G.O. and Love, R.F., (1972), "A Nonlinear Approximation Method for Solving a Generalized Rectangular Distance Weber Problem," Management Science, Vol.18, No.11, pp.656-663.

Witzgall, C., (1964), "Optimal Location of a Central Facility: Mathematical Models and Concepts," National Bureau of Standards Report 8388, Gaithersburg, MD.