



HAL
open science

Minimum-cost distances in spatial analysis

J.M. Huriot, T.E. Smith, J.F. Thisse

► **To cite this version:**

J.M. Huriot, T.E. Smith, J.F. Thisse. Minimum-cost distances in spatial analysis. [Research Report] Institut de mathématiques économiques (IME). 1988, 37 p., figures, bibliographie. hal-01544213

HAL Id: hal-01544213

<https://hal.science/hal-01544213>

Submitted on 21 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

I.M.E.

EQUIPE DE RECHERCHE ASSOCIEE AU C.N.R.S.

DOCUMENT DE TRAVAIL



INSTITUT DE MATHEMATIQUES ECONOMIQUES

UNIVERSITE DE DIJON

FACULTE DE SCIENCE ECONOMIQUE ET DE GESTION

4, BOULEVARD GABRIEL — 21000 DIJON

n° 108

**MINIMUM-COST DISTANCES IN SPATIAL
ANALYSIS**

J.M. HURIOT, T.E. SMITH and J.F. THISSE

June 1988

LIBRARY
-S. 101
1988
JUN 10
C 3. 70

MINIMUM-COST DISTANCES IN SPATIAL ANALYSIS

by J-M.Huriot*, T.E.Smith**, and J-F.Thisse***

June, 1988

ABSTRACT

An axiomatic approach to distance is developed which focuses on those behavioral concepts of distance related to movement in space. In particular, spatial movement by behaving units is postulated to involve a choice from among some set of abstract *trips* in space, and implicitly, to involve the minimization of some relevant notion of *trip costs*. In this context, the relevant behavioral notion of distance in space is taken to be the *minimum-cost distance* generated by this choice process. These trip-cost concepts extend the classical notions of paths, path lengths, and shortest paths in metric spaces. Hence many of the analytical results of the paper involve extensions of classical shortest-path distance properties to minimum-cost distances. In addition to these extensions, a characterization theorem is given which specifies the possible functional relationships between trip costs and their associated path lengths. These relationships include most functional forms which are commonly employed in the literature.

* Professor of Economics, University of Bourgogne, France.

** Professor of Regional Science, University of Pennsylvania, U.S.A..

*** Professor of Regional Science, CORE, Belgium.

1. INTRODUCTION

All models of human spatial behavior are based on some underlying notion of distance. Our most familiar concept of distance is of course euclidean (or straight-line) distance between points, which represents the *shortest-path distance* in euclidean space. But the notion of shortest-path distance is itself much broader, and indeed is meaningful for essentially any kind of spatial movement [as studied for example in Smith (1987a,1987b)]. However, from a behavioral viewpoint, not even shortest-path distances are sufficiently general to cover the full range of distance concepts which are relevant for movement in space. In airline travel, for example, it is clear that while shortest-path concepts (such as great-circle routes on the globe) are useful for determining flight times, they are often poorly correlated with actual ticket prices. Hence from the viewpoint of potential air travelers, the *economic distance* between points is often shaped more by forces of supply and demand than by physical separation alone. With this in mind, the central objective of the present paper is to extend the classical notion of shortest-path distances to a more general class of *minimum-cost distances*, which is sufficiently broad to include most types of economic distances observed in practice. This extension is primarily motivated by the desire to incorporate explicit economic concepts of space within the more general axiomatic approach to geographic space proposed by Beguin and Thisse (1979). However, this framework is potentially meaningful for many types of non-economic distance as well, including both social and psychological concepts of distance.

To model such behavioral notions of distance, it is useful to begin by considering the act of spatial movement itself. For in any given behavioral situation, the relevant notion of distance can depend critically on the type of movement involved (work trip by a commuter, vacation trip by a household, or commodity shipment by a firm). Moreover, each movement decision generally involves a choice from among some set of options (such as route and mode choice by a commuter, or carrier choice by a commodity shipper). Hence our present approach is to model all such options in terms of abstract *trips*, and to treat each movement decision as a choice from some given set of feasible trips. In this context, it is also postulated that there exists some relevant measure of *travel cost* associated with each potential trip. Hence, to the

extent that such trip choices are based on cost-minimizing behavior, one may take the minimum trip costs between locations to define the relevant behavioral notion of *distance* in space.

Within this framework, the specific goals of the present paper are to formalize this notion of *minimum-cost distance*, and to study a number of its more important properties from a behavioral viewpoint. This study is largely inspired by the classical development of shortest-path distances in metric spaces [as summarized in Busemann (1955), Blumenthal (1970), Beguin and Thisse (1979), and Smith (1987a,1987b)], and closely parallels that literature. In particular, we begin by assuming that trips can be meaningfully added together to form *trip chains* (in a manner analogous to concatenations of paths). It is then assumed that such trip chains can always be treated as sequences of individual trips, so that (from a cost-minimizing viewpoint) the relevant cost of a trip chain never exceeds the total cost of its component trips. However, unlike simple path length, there may be significant *economies of trip chaining*. Finally, given these structural assumptions, the *minimum-cost distance* between any two locations is defined (in a manner completely paralleling shortest-path distance) to be the greatest lower bound on all trip costs between these locations.

Our first result (Theorem 3.1 below) is to show that the well-known triangularity property of shortest-path distances extends to all minimum-cost distances, i.e. that such distances must always be *quasimetrics*. This general formulation has the additional analytical advantage of ensuring that minimum-cost distances are well defined whenever there exist finite-cost trips between all locations. But from a behavioral viewpoint this formulation leaves open the question of whether such distances can actually be realized by any trips, i.e. whether there exist *minimum-cost trips*. Hence, as an extension of the well-known existence results for shortest paths (geodesics) between locations, we next consider the existence question for minimum-cost trips between locations. But unlike the classical notion of paths (which are generally assumed to be continuous functions with respect to the neighborhood topology generated by shortest-path distance), there is no prior structure on the set of abstract trips. Hence to establish meaningful results, we first postulate the existence of some appropriate measure of *trip similarity* by which trips can be compared. In this context, our main result (Theorem 3.2

below) is to establish regularity conditions on the similarity structure of trips (both with respect to their cost attributes and origin-destination attributes) which ensure the existence of minimum-cost trips between all locations.

Next, we consider the possible relationships between trip costs and path lengths associated with these trips. In certain cases, these concepts may of course be identical. For example, if the relevant set of trips for a given car owner correspond to the set of alternative route choices leading to his desired destination, and if the relevant cost variable is travel time, then minimum-cost trips in this case may correspond precisely to shortest routes in the given transport network (assuming that traffic congestion and other factors are not significant). More generally, the relevant notion of trip cost may be some monotone increasing function of some underlying measure of path length. In this context, our central result (Theorem 4.1 below) is to identify the precise nature of the functional relationships which are possible between trip costs and path lengths. In addition, the corresponding functional relations between minimum-cost distances and shortest-path distances are also established (Theorems 5.1 and 5.2 below).

Finally, in the concluding section of the paper, we briefly discuss some of the limitations and possible extensions of our approach. In particular, we consider behavioral situations in which the assumption of economies of trip chaining may fail to hold. This is especially relevant when the psychological costs of longer trips are significantly greater than those of shorter trips. We also describe situations in which trip costs may not be monotone increasing in their associated path lengths.

To develop our results, we begin in the next section with a review of the basic concepts of *paths* and *path lengths* in general spaces. This review serves to clarify the formal parallel between these concepts and the more general concepts of trips and trip costs in the present paper. In addition, the results summarized here provide the basis for the comparative analysis of trip costs and path lengths in sections 4 and 5 below.

2. SHORTEST-PATH DISTANCES

The following development of relevant distance concepts is given in much greater detail in Smith (1987a,1987b). Let R , R_+ , and R_{++} denote the *real numbers*, *nonnegative reals*, and *positive reals*, respectively, and let Z_{++} denote the *positive integers*. Throughout the analysis, we take X to be a given nonempty set of abstract *locations*.

2.1. Distances

A nonnegative function, $d: X \rightarrow R_+$, is designated as a *distance* on X iff d satisfies the following condition for all $x, y \in X$:

$$D1. \text{ (Distinguishability)} \quad d(x,y) = 0 \Leftrightarrow x = y$$

The pair (X,d) is designated as the corresponding *distance space*. A distance, d , is designated as a *semimetric* iff d also satisfies

$$D2. \text{ (Symmetry)} \quad d(x,y) = d(y,x)$$

for all $x, y \in X$. Similarly, d is designated as a *quasimetric* iff

$$D3. \text{ (Triangularity)} \quad d(x,z) \leq d(x,y) + d(y,z)$$

holds for all $x, y, z \in X$. Each distance satisfying both D2 and D3 is designated as a *metric* on X . For any $x \in X$ and $\epsilon > 0$, the set, $S_\epsilon(x) = \{y \in X: d(x,y) < \epsilon\}$, is designated as the *d-neighborhood* of size ϵ about x in X . A subset, $B \subseteq X$, is then said to be *d-bounded* iff $B \subseteq S_\epsilon(x)$ for some $x \in B$ and $\epsilon > 0$. With these additional definitions, a distance, d , is said to be *uniform* iff the following condition is satisfied:

D4. *(Uniformity)* For each *d-bounded* set, $B \subseteq X$, and $\epsilon > 0$, there exists a sufficiently small $\delta > 0$ such that for all $x, y, z \in X$,

$$(2.1) \quad \max\{d(z,x), d(z,y)\} < \delta \Rightarrow d(x,y) < \epsilon$$

Intuitively, the uniformity condition requires that pairs of locations be close to one another whenever they are both close to a common location. Each distance satisfying (D2,D4) is designated as a *uniform semimetric* [as exemplified by the measures of "trip dissimilarity" in section 3.2 below]. Similarly, each distance satisfying (D3,D4) is designated as a *uniform quasimetric* [as studied in more detail in section 2.5 below]. Sharper characterizations of both uniform semimetrics and uniform quasimetrics are given in Smith (1987a, Proposition 2.3).

Our primary interest in uniform-distance spaces is that they share many of the convergence properties of familiar euclidean spaces. To state these properties, we require the following additional concepts for general distance spaces, (X,d) . First, a sequence (x_n) in X said to be *d-convergent* iff there exists some $x \in X$ such that $d(x,x_n) \rightarrow 0$. Similarly, (x_n) is said to be a *d-Cauchy sequence* iff for each $\epsilon > 0$ there is some $n_\epsilon \in \mathbb{Z}_{++}$ such that $d(x_n, x_m) < \epsilon$ for all $n,m \geq n_\epsilon$. Next, a set, $B \subseteq X$, and is said to be *totally d-bounded* iff for each $\epsilon > 0$ there exists a finite subset $\{x_1, \dots, x_n\} \subseteq B$ such that $B \subseteq \bigcup_{i=1}^n S_\epsilon(x_i)$, [i.e. such that each element of B is no further than distance, ϵ , from some element of $\{x_1, \dots, x_n\}$]. With these definitions, we now consider the following general regularity conditions on distance spaces:

DEFINITION 2.1. A distance space, (X,d) , is said to be:

- (i) *d-complete* iff every d-Cauchy sequence in X is d-convergent.
- (ii) *totally d-neighborhood bounded* iff every d-neighborhood in X is totally d-bounded.
- (iii) *regular* iff (X,d) is d-complete and totally d-neighborhood bounded.

To interpret the *d-completeness* condition, observe first that by definition each d-Cauchy sequence in X is eventually concentrated in some "arbitrarily small" region of X . Hence d-completeness essentially requires that X have no "missing elements", in the sense that such concentrated sequences eventually converge to elements in X . To interpret the *total d-neighborhood boundedness* condition, observe next that for any totally d-bounded set, $B \subseteq X$, and small $\epsilon > 0$, all elements of B are by definition within distance, ϵ , of some finite subset of B . Hence this condition essentially requires that each d-neighborhood in X be "uniformly approximated" by its finite subsets.

To apply these concepts to uniform-distance spaces, we require the following additional concepts. First, a subset, $B \subseteq X$, is said to be *d-closed* iff B contains each $x \in X$ satisfying the condition that $d(x, x_n) \rightarrow 0$ for some sequence (x_n) in B . Next, B is said to be *sequentially d-compact* iff for each sequence (x_n) in B there exists a subsequence (x'_n) which is *d-convergent* in B , i.e. with $d(x, x'_n) \rightarrow 0$ for some $x \in B$. With these definitions, we now have the following basic property of all regular uniform-distance spaces [Proposition 3.2 in Smith (1987a)]:

PROPOSITION 2.1. *If (X, d) is a regular uniform-distance space, then each d-bounded, d-closed subset, $B \subseteq X$, is sequentially d-compact.*

2.2. Paths

If $[0, 1]$ denotes the closed unit interval in \mathbb{R} , then each function, $p: [0, 1] \rightarrow X$, is designated as a possible *path* in X , where the values of the index set, $[0, 1]$, parameterize some way of moving from the point, $p(0)$, to the point, $p(1)$, in X . Since only the ordering of the index set, $[0, 1]$, is important, we may reparameterize paths in many ways. In particular, if P_X denotes the set of all paths on X , then for any two paths, $p, q \in P_X$, q is said to be a *reparameterization* of p iff there exists a continuous, nondecreasing function, $e: [0, 1] \rightarrow [0, 1]$, with $e(0) = 0$ and $e(1) = 1$ such that $q[e(\alpha)] = p(\alpha)$ for all $\alpha \in [0, 1]$. If in addition, e is one-to-one (and hence a bijection) then paths p and q are said to be *parametrically equivalent*, and written as $p \sim q$. The set of all reparameterizations of p is designated as the *reparameterization class*, $[p]$, for p in P_X . Next we define an appropriate notion of "path addition". In particular, if for any paths, $p_1, p_2 \in P_X$, it is true that p_1 ends exactly where p_2 begins, i.e. if $p_1(1) = p_2(0)$, then it is meaningful to consider the path formed by joining p_2 to p_1 . More precisely, the unique path, $p_1 \circ p_2 \in P_X$, defined for all $\alpha \in [0, 1]$ by

$$(2.2) \quad p_1 \circ p_2(\alpha) = \begin{cases} p_1(2\alpha) & , \quad 0 \leq \alpha \leq 1/2 \\ p_2(2\alpha - 1) & , \quad 1/2 \leq \alpha \leq 1 \end{cases}$$

is designated as the *concatenation* of p_1 and p_2 . This definition is readily extended recursively to all finite concatenations, $p_1 \circ p_2 \circ \dots \circ p_n$, $n \in \mathbb{Z}_{++}$.

Next, if for any elements, $x, y \in X$ and path set, $P \subseteq P_X$, we denote the subset of paths from x to y in P by $P(x, y) = \{p \in P: p(0) = x, p(1) = y\}$, then for any elements, $x, y, z \in X$, and path sets, $P_1, P_2 \in P_X$, the set of path concatenations, $P_1(x, y) \circ P_2(y, z) = \{p_1 \circ p_2: p_1 \in P_1(x, y), p_2 \in P_2(y, z)\}$, is well defined.

As a final path concept, it is important to formalize the notion of a subpath. For any path, $p \in P_X$, and interval, $[\alpha, \beta] \subseteq [0, 1]$, the path, $p\langle\alpha, \beta\rangle \in P_X$, defined by

$$(2.3) \quad p\langle\alpha, \beta\rangle(\sigma) = p[(1-\sigma)\alpha + \sigma\beta] \quad , \quad \sigma \in [0, 1]$$

is designated as the $\langle\alpha, \beta\rangle$ -subpath of p . In particular, it can be shown [Proposition 5.10 in Smith (1987a)] that each path is parametrically equivalent to concatenations of its subpaths, i.e. that for all $p \in P_X$,

$$(2.4) \quad p \sim p\langle 0, \alpha\rangle \circ p\langle \alpha, 1\rangle \quad , \quad \alpha \in [0, 1]$$

Given these path concepts, we can now define those families of paths on a given set, X , which are relevant for the study of shortest-path distances:

DEFINITION 2.2. For any nonempty set, X , a family of paths, $N \subseteq P_X$, is designated as a *path network* on X iff N satisfies the following four conditions for all $x, y, z \in X$, $p \in P_X$, and $[\alpha, \beta] \subseteq [0, 1]$:

- N1. (*Connectedness*) $N(x, y) \neq \emptyset$
- N2. (*Concatenation Closure*) $N(x, y) \circ N(y, z) \subseteq N(x, z)$
- N3. (*Subpath Closure*) $p \in N \Rightarrow p\langle\alpha, \beta\rangle \in N$
- N4. (*Parametric Completeness*) $p \in N \Rightarrow [p] \subseteq N$

Condition N1 asserts that any point, $y \in X$, can be reached from any other point, $x \in X$, by at least one path in N . Condition N2 asserts that concatenations of paths in N always yield paths in N , and similarly, N3 asserts that each subpath of a given path in N is also a meaningful path in N . Finally, condition N4 asserts that all reparameterizations of paths are available in N . Observe also that if for each $x \in X$ we define the *null path*, p_x , to be the unique path in P_X with $p_x(\alpha) = x$ for all $\alpha \in [0, 1]$, then it follows at once

from N1 and N3 that the set of null paths, $P_X^0 = \{p_x : x \in X\}$, on X is contained in every path network on X .

In order to construct path networks, it is convenient to introduce the following concept of a *path base*. In particular, if for any path set, $P \subseteq P_X$, the n^{th} -concatenation power, P^n , is defined recursively by $P^1 = P$ and $P^{n+1} = P \circ P^n$ for all $n \in \mathbb{Z}_{++}$, then we now say that:

DEFINITION 2.3. A set of paths, $P \subseteq P_X$, is designated as a *path base* on X if and only if P satisfies the following two conditions for all $x, y \in X$, $p \in P_X$, and $[\alpha, \beta] \subseteq [0, 1]$:

- B1. (*Weak Connectedness*) $P^n(x, y) \neq \emptyset$ for some $n \in \mathbb{Z}_{++}$
 B2. (*Subpath Closure*) $p \in P \Rightarrow p\langle \alpha, \beta \rangle \in P$

Each path base allows the construction of a path network in a simple way. More specifically, if for each path set, $P \subseteq P_X$, we denote that class of all reparameterizations of paths in P by $[P] = \{[p] : p \in P\}$, then it can be shown [see Proposition 6.1 in Smith (1987a)] that:

PROPOSITION 2.2. For any path base, P , on X , the associated family of paths,

$$(2.5) \quad N(P) = \bigcup_{n \in \mathbb{Z}_{++}} [P^n] \subseteq P_X$$

is a path network on X .

Moreover, $N(P)$ is the smallest path network containing P [i.e. if $P \subseteq N$ for any path network, N , then $N(P) \subseteq N$]. Hence it is natural to designate $N(P)$ as the path network *generated* by P .

2.3. Path Lengths

Given the formal concept of path networks above, our next objective is to construct an appropriate notion of *path length* on networks. We begin by observing that paths may in general have infinite length, and hence may have values ranging over the extended nonnegative real numbers, $R_+(\infty) = R_+ \cup \{\infty\}$.

With this notation, we now say that:

DEFINITION 2.4. (i) For any path network, $N \subseteq P_X$, a function, $l: N \rightarrow R_+(\infty)$, is designated as a *length* on N iff l satisfies the following three conditions for all $p, p_1, p_2 \in N$ with $p_1 \circ p_2 \in P_X$:

L1. (*Nullity*) $p \in P_X^0 \Rightarrow l(p) = 0$

L2. (*Additivity*) $l(p_1 \circ p_2) = l(p_1) + l(p_2)$

L3. (*Invariance*) $p_1 \subseteq [p_2] \Rightarrow l(p_1) = l(p_2)$

(ii) A path $p \in N$ is said to be *l-finite* iff $l(p) < \infty$.

(iii) The class of all lengths on N is denoted by $L(N)$.

Condition L1 simply requires that each null path should have zero length. Condition L2 requires that the length of any path concatenation should be equal to the sum of the lengths of its component paths. Finally, condition L3 requires that the length of any path should be the same as the length of any of its reparameterizations.

2.4. Shortest-Path Distances

Given these general concepts of path networks and path lengths, it is meaningful to consider the possibility of "shortest paths" between points in space. In particular, we now formalize the notion of *shortest-path distances* defined by the lengths of such paths as follows:

DEFINITION 2.5. (i) For any distance function, $d: X^2 \rightarrow R_+$, and path network, $N \subseteq P_X$, if there exists at least one length function, $l \in L(N)$, such that for all $x, y \in X$,

$$(2.6) \quad d(x, y) = \inf\{l(p) : p \in N(x, y)\}$$

then d is designated as a *shortest N-path distance* on X .

(ii) If d is a shortest N -path distance for at least one path network, $N \subseteq P_X$, then d is designated simply as a *shortest-path distance* on X .

In other words, a distance function, $d: X^2 \rightarrow R_+$, is a shortest-path distance on X iff each distance value, $d(x,y)$, corresponds to the greatest lower bound of all path lengths between x and y with respect to some path network, N , on X and length function, ℓ , on N . Notice that this definition does not imply that $d(x,y)$ is actually achieved by the length of any path between x and y . [The existence of shortest paths is considered in section 2.6 below.] For our present purposes, the central property of shortest-path distances is given by the following result [Theorem 5.1 in Smith (1987b)]:

PROPOSITION 2.3. (*Quasimetric Equivalence*) A distance function, $d: X^2 \rightarrow R_+$, is a shortest-path distance on X iff d is a quasimetric.

From a behavioral viewpoint, this result shows that the classical triangularity condition (D3) is synonymous with shortest-path distance. In other words, whenever a given distance, $d: X \rightarrow R_+$, satisfies triangularity, there must always be some implicit path network, N , on X and path length, ℓ , on N , with respect to which d is the shortest-path distance. In general, both the path network, N , and path length, ℓ , may be highly nonunique. However, there is always a unique representative path length for d in the following sense. If $L_d(N)$ denotes the set of path lengths in $L(N)$ for which d is a shortest N -path distance, and if path length, $\ell \in L_d(N)$, is said to be *minimal* iff $\ell(p) \leq \ell'(p)$ for all $\ell' \in L_d(N)$ and $p \in N$, then there always exists a *unique minimal path length* in $L_d(N)$. This minimal path length may be defined as follows. Let the class of finite ordered subsets of $[0,1]$ containing 0 and 1 be denoted by $\Omega = \{\omega = (\alpha_1, \dots, \alpha_n): 0 = \alpha_1 \leq \dots \leq \alpha_n = 1\}$, and for each path, $p \in N$, let

$$(2.7) \quad \ell_d(p|\omega) = \sum_{i=1}^n d[p(\alpha_{i-1}), p(\alpha_i)]$$

denote the ω -length of p with respect to d . Then if d is a shortest N -path distance, it can be shown [Propositions 8.3 and 8.4 in Smith (1987a), and Theorem 5.2 in Smith (1987b)] that the function, $\ell_d: N \rightarrow R_+(\infty)$, defined for all $p \in N$ by

$$(2.8) \quad \ell_d(p) = \sup\{\ell_d(p|\omega): \omega \in \Omega\}$$

is the unique minimal path length in $L_d(N)$. This unique length function (which extends the classical notion of Jordan length in euclidean spaces) is designated as the *d-length* on N . Henceforth we consider only the *d-length* for each shortest N -path distance, d .

2.5. Path Structures

For purposes of analysis it is convenient to summarize the concepts above in terms of the notion of a *path structure*. To do so, observe first that since ℓ_d is uniquely generated by d , it follows at once from (2.6) that for any path network, N , all shortest N -path distances, d , are uniquely characterized by the single condition that

$$(2.9) \quad d(x,y) = \inf\{\ell_d(p) : p \in N(x,y)\}, \quad x,y \in X$$

Given this formal characterization of shortest N -path distances, we now say that:

DEFINITION 2.6. (i) For any nonempty set, X , a triple, $P = (X,N,d)$, is designated as a *path structure* on X iff N is a path network on X , and d is a shortest N -path distance on X .

(ii) If $\ell_d(p) < \infty$ for all $p \in N$, then P is said to be *finite*.

(iii) For any two path structures, $P = (X,N,d)$ and $P' = (X',N',d')$, P' is said to be a *substructure* of P iff $X' \subseteq X$, $N' \subseteq N$, and for all $x,y \in X'$,

$$(2.10) \quad d'(x,y) = \inf\{\ell_d(p) : p \in N'(x,y)\}$$

Numerous examples of path structures and substructures are given in Smith (1987a,1987b). For our present purposes it suffices to consider only those path structures, $P = (X,N,d)$, in which (X,d) is a regular uniform-quasimetric space, and in which all paths in N are continuous with respect to d . More precisely, we now say that a path, $p \in N$, is *d-continuous* if and only if for all convergent sequences (α_n) in $[0,1]$ it is true that $\alpha_n \rightarrow \alpha \Rightarrow d[p(\alpha),p(\alpha_n)] \rightarrow 0$. If the class of all *d-continuous* paths in P_X is denoted by $C_d(X)$, then the desired class of path structures on (X,d) may be defined as follows:

DEFINITION 2.7. (i) A path structure, $P = (X, N, d)$, is designated as a *continuous path structure* on X if and only if $N \subseteq C_d(X)$.

(ii) If in addition, (X, d) is a regular uniform-quasimetric space, then P is designated as a *d-regular path structure*.

One important feature of d-regular path structures, $P = (X, N, d)$, is that the path network, N , is itself a uniform-quasimetric space with respect to an appropriate notion of distance between paths. In particular, if the function, $d^+ : N \rightarrow R_+(\infty)$, defined for all paths, $p, q \in N$, by

$$(2.11) \quad d^+(p, q) = \sup\{d[p(\alpha), q(\alpha)] : \alpha \in [0, 1]\}$$

is designated as the *supremum distance* on N generated by d , then it can be shown that d^+ is in fact a *uniform-quasimetric* on N [Proposition 4.5 in Smith (1987a)], and hence that (N, d^+) constitutes a *uniform-quasimetric space*. Two additional properties of uniform-quasimetric spaces will prove useful in the analysis below. In particular, if a path, $p \in C_d(X)$, is now said to be *uniformly d-continuous* iff for each $\epsilon > 0$ there is some $\delta > 0$ such that for all $\alpha, \beta \in [0, 1]$, $|\alpha - \beta| < \delta \Rightarrow d[p(\alpha), p(\beta)] < \epsilon$, then it can be shown [Proposition 4.4(ii) and Proposition 4.6(ii) in Smith (1987a)] that:

PROPOSITION 2.4. (i) If (X, d) is a uniform-quasimetric space, then for all $x, y \in X$ and all sequences (x_n) in X ,

$$(2.12) \quad \max\{d(x, x_n), d(y, x_n)\} \rightarrow 0 \Rightarrow x = y$$

(ii) In addition, each path, $p \in C_d(X)$, is *uniformly d-continuous*.

Several additional properties of d-regular path structures are important for our purposes. First of all, it can be shown [Proposition 11.3 in Smith (1987a)] that the d-length function, l_d , is itself continuous along paths, i.e. that

PROPOSITION 2.5. For any d-regular path structure, $P = (X, N, d)$, and any l_d -finite path, $p \in N$, the function, $l_d(p\langle 0, \cdot \rangle)$, is continuous.

Moreover, this continuity property in turn implies that it is possible to parameterize all (l_d -finite) paths in a manner directly proportional to their d -length. This notion can be made more precise as follows:

DEFINITION 2.8. (i) For any path structure, $P = (X, N, d)$, and l_d -finite path, $p \in N$, a reparameterization, $p_d \in [p]$, is designated as a *d-standard parameterization* of p iff for all intervals, $[\alpha, \beta] \subseteq [0, 1]$,

$$(2.13) \quad l_d[p_d(\alpha, \beta)] = (\beta - \alpha)l_d(p) .$$

(ii) The subset, $N_d = \{p_d : p \in N\}$, of all d -standard parameterizations of paths in N is designated as the *d-standard representation* of N .

With this definition, we have the following additional property of all d -regular path structures [see Proposition 11.4 in Smith (1987a)]:

PROPOSITION 2.6. For any d -regular path structure, $P = (X, N, d)$, and any l_d -finite path, $p \in N$, there exists a unique *d-standard parameterization* of p in N .

2.6. Shortest Paths

Finally, for purposes of comparison with minimum-cost trips below, we record the following results on the existence of shortest paths in d -regular path structures. To begin with, let us now say that:

DEFINITION 2.9. (i) For any path structure, $P = (X, N, d)$, the shortest N -path distance function, d , is said to be *realizable* iff for each pair of locations, $x, y \in X$, there exists a path, $p \in N(x, y)$, with

$$(2.14) \quad l_d(p) = d(x, y)$$

(ii) Each path, $p \in N(x, y)$, satisfying (2.11) is designated as a *shortest path* from x to y in P .

To ensure the existence of shortest paths, we require the following additional regularity condition. If for any path structure, $P = (X, N, d)$, we

now designate a path sequence (p_n) in N as an l_d -bounded sequence whenever there exists some finite upper bound, $b > 0$, with $l_d(p_n) \leq b$ for all $n \in \mathbb{Z}_{++}$, then we now say that:

DEFINITION 2.10. (i) For any d -regular path structure, $P = (X, N, d)$, a family of paths, $P \subseteq N$, is said to be *finitely closed in P* iff for each $p \in C_d(X)$ and l_d -bounded sequence (p_n) in P ,

$$(2.15) \quad d^+(p, p_n) \rightarrow 0 \Rightarrow p \in P$$

(ii) If the full path network, N , is finitely closed in P , then P is said to be a *finitely closed path structure*.

With this additional condition, it can be shown [Proposition 12.5 in Smith (1987a)] that:

PROPOSITION 2.7. (*Existence of Shortest Paths*) (i) For any substructure, $P' = (X', N', d')$, of a d -regular path structure, $P = (X, N, d)$, if N' is finitely closed in P , then the shortest N' -path distance, d' , is realizable.

(ii) In particular, if P is itself a finitely-closed path structure, then the shortest N -path distance, d , is realizable.

3. MINIMUM-COST DISTANCE

Given this overview of shortest-path distance concepts, we turn now to the major focus of the present paper, namely, to the more general notion of *minimum-cost distance*. To do so, we begin in section 3.1 below by developing a model of abstract trips in terms of the notion of a *trip structure*, which parallels the above concept of a path structure. In addition it is shown that the associated *minimum-cost distance* between possible origins and destinations of trips exhibits the quasimetric property of shortest-path distances.

3.1 Trip Structures

As in section 2 above, we begin with an arbitrary set, X , of *locations* (or *addresses*), x , and consider an abstract collection, T , of possible *trips*, t , between locations in X . Such abstract trips may involve physical travel by individuals, or shipments of economic commodities. More generally, such abstract trips may involve pairwise interactions of any type between behaving units associated with attributes, x , in some relevant space of attributes, X . For our present purposes, the relevant spatial attributes of each trip, $t \in T$, are taken to be its *origin location*, $o(t)$, and its *terminal location*, $r(t)$, in X . All other relevant aspects of trip, t , are taken to be summarized by some given nonnegative measure of *travel cost*, $c(t)$. For physical trips, such costs must implicitly involve some choice of path (or route) in X from $o(t)$ to $r(t)$ [as analyzed in section 4 below], as well as a choice of travel mode (such as car, bus, or plane). For more general types of interactions, such costs may also involve implicit routing choices (through some communication network or organizational decision structure), and choices of interaction modes (such as personal contacts, phone calls, or letters).

Our basic structural hypotheses about trips focus on the possibility of *trip chaining*. In particular, for any two trips, $t_1, t_2 \in T$, if t_1 ends exactly where t_2 begins, i.e. if $r(t_1) = o(t_2)$, then (in a manner completely paralleling path concatenations, $p_1 \circ p_2$) it is postulated that the *trip chain*, $t_1 \otimes t_2$, consisting of trip, t_1 , followed by trip, t_2 , is possible, i.e. that $t_1 \otimes t_2 \in T$. The resulting *trip-chain operator*, \otimes , on T is here taken to be *associative*, i.e. to satisfy the identity, $(t_1 \otimes t_2) \otimes t_3 = t_1 \otimes (t_2 \otimes t_3)$ [so that recursively defined *n-trip chains*, $t_1 \otimes t_2 \otimes \dots \otimes t_n$, are independent of the order in which chains are formed]. Hence, it suffices to consider only simple pairwise trip chains. For each such trip chain, $t_1 \otimes t_2$, it follows by definition that both $o(t_1 \otimes t_2) = o(t_1)$ and $r(t_1 \otimes t_2) = r(t_2)$ hold identically. However, the relevant travel costs associated with trip chains are somewhat more problematic. For unlike the simple physical addition of path lengths, there may in general exist significant *economies of trip chaining*. This is seen most clearly, for example, in the discounting of many types of round-trip fares. More generally, if each possible trip chain is regarded as an economic commodity, then various types of economic discounts may be available in the market place. On the other hand, it may also be argued on economic grounds

that such costs should never exceed the combined costs of individual trips. Indeed if this were the case, then it would be more economical to treat $t_1 \oplus t_2$ as a pair of completely separate trips, t_1 and t_2 . [However, this condition may well be violated in other behavioral contexts, as discussed in section 6 below.] Hence for our present purposes, we now hypothesize that for each pair of trips, $t_1, t_2 \in T$ with $t_1 \oplus t_2 \in T$,

$$(3.1) \quad c(t_1 \oplus t_2) \leq c(t_1) + c(t_2)$$

In addition to this travel-cost assumption, it is of course assumed that the relevant location space, X , is *trip-connected*. More formally, if

$$(3.2) \quad T(x,y) = \{t \in T: o(t) = x \text{ and } r(t) = y\}$$

denotes the set of all trips in T from x to y , then it is assumed that $T(x,y)$ is nonempty for all $x,y \in X$. Finally, in a manner analogous to the notion of null paths in section 2 above, it is convenient to assume the existence of *null trips*, $\theta(x) \in T(x,x)$, at each location, $x \in X$, which may be interpreted as representing the "no trip" option for behaving units located at x . Each null trip is characterized by a zero travel cost, while all other travel costs are positive. In addition, it is assumed that for any pair of distinct locations, $x,y \in X$, there do not exist trips in $T(x,y)$ with travel costs arbitrarily close to zero.

To formalize these assumptions, it is convenient to represent the assignment of origin locations, terminal locations, and travel costs to each trip in T by functions, $o: T \rightarrow X$, $r: T \rightarrow X$, and $c: T \rightarrow R_+$, respectively. Similarly, the assignment of each location in X to a unique null trip in T is representable by a function, $\theta: X \rightarrow T$. Finally, if

$$(3.3) \quad \Delta(T^2) = \{(t_1, t_2) \in T^2: r(t_1) = o(t_2)\}$$

denotes the set of trip pairs in T for which trip chains are defined, then the the trip-chain operator is also representable by a function, $\oplus: \Delta(T^2) \rightarrow T$, with images $\oplus(t_1, t_2) = t_1 \oplus t_2$. Next, to formalize our travel-cost assumption on trip chains, it is convenient to let

$$(3.4) \quad T(x,y,a) = \{t \in T(x,y) : c(t) \leq a\}, \quad a \in \mathbb{R}_+$$

denote the set of trips from x to y with travel costs not exceeding level, a . With these notational conventions, it follows that for any locations $x,y,z \in X$ and trips, $t_1 \in T(x,y)$ and $t_2 \in T(y,z)$, the trip chain, $t_1 \oplus t_2$, is well defined and must be an element of $T[x,z,c(t_1) + c(t_2)]$, i.e. must be a trip from x to z with cost not exceeding $c(t_1) + c(t_2)$. Finally if the set of free trips in T is denoted by, $c^{-1}(0) = \{t \in T : c(t) = 0\}$, then the above assumptions can now be formalized in terms of the following notion of a trip structure:

DEFINITION 3.1. For any nonempty sets, X, T , and collection of functions, $o: T \rightarrow X$, $r: T \rightarrow X$, $c: T \rightarrow \mathbb{R}_+$, $\theta: X \rightarrow T$, and $\oplus: \Delta(T^2) \rightarrow T$, the ordered array, $T = (X, T, o, r, c, \theta, \oplus)$, is designated as a trip structure on X iff the following four conditions are satisfied for all $t \in T$, $x,y,z \in X$, $t_1 \in T(x,y)$ and $t_2 \in T(y,z)$:

- T1. (Trip Connectedness) $T(x,y) \neq \emptyset$
- T2. (Null Trips) $\theta(x) \in T(x,x)$ and $c^{-1}(0) = \theta(X)$
- T3. (Cost Positivity) $\inf\{c(t) : t \in T(x,y)\} = 0 \Rightarrow x = y$
- T4. (Chain Economies) $t_1 \oplus t_2 \in T[x,z,c(t_1) + c(t_2)]$

Given this definition, observe next that by the trip-connectedness property (T1) and the finite nonnegativity of travel costs, the quantities given for all $x,y \in X$ by

$$(3.5) \quad d_c(x,y) = \inf\{c(t) : t \in T(x,y)\}$$

yield a well-defined function, $d_c: X^2 \rightarrow \mathbb{R}_+$. With this definition, our first result is to show that d_c is always a quasimetric on X :

THEOREM 3.1. For each trip structure, $T = (X, T, o, r, \theta, \oplus)$, the function, $d_c: X^2 \rightarrow \mathbb{R}_+$, is a quasimetric on X .

Proof: By the null-trip property (T2) it follows from (3.5) that $d_c(x,x) = c[\theta(x)] = 0$. Similarly, from the cost-positivity property (T3) and the nonnegativity of $c: T \rightarrow \mathbb{R}_+$, it also follows that $x \neq y \Rightarrow d_c(x,y) > 0$. Hence, d_c satisfies distinguishability (D1) and is seen to be a distance on X . To establish triangularity (D5) let us choose any locations, $x,y,z \in X$, and observe from the chain-economy property (T4) that for any $t_1 \in T(x,y)$ and $t_2 \in T(y,z)$ it must be true that $t_1 \oplus t_2 \in T(x,z)$ and that $c(t_1 \oplus t_2) \leq c(t_1) + c(t_2)$. Hence

$$\begin{aligned}
 (3.6) \quad d_c(x,z) &= \inf\{c(t) : t \in T(x,z)\} \\
 &\leq \inf\{c(t_1 \oplus t_2) : t_1 \in T(x,y), t_2 \in T(y,z)\} \\
 &\leq \inf\{c(t_1) + c(t_2) : t_1 \in T(x,y), t_2 \in T(y,z)\} \\
 &= \inf\{c(t_1) : t_1 \in T(x,y)\} + \inf\{c(t_2) : t_2 \in T(y,z)\} \\
 &= d_c(x,y) + d_c(y,z).
 \end{aligned}$$

Hence d_c also satisfies D5, and the result is established. ■

Thus, in a manner completely paralleling the notion of *shortest-path distances* generated by path lengths on spaces (in section 2.4 above), each trip structure is seen to generate a quasimetric space on X in terms of *minimal travel costs*. Hence we now say that

DEFINITION 3.2. (i) For each trip structure, $T = (X, T, o, \tau, c, \theta, \oplus)$, on X , the quasimetric, d_c , in (3.5) is designated as the *minimum-cost distance* on X generated by T .

(ii) The resulting quasimetric space, (X, d_c) , is designated as the *minimum-cost distance space* generated by T .

With this definition, Theorem 3.1 for minimal-cost distances is seen to be completely analogous to Proposition 2.2 above for shortest-path distances. [More explicit relationships between these two concepts will be considered in section 5 below].

3.2. Existence of Minimum-Cost Trips

Given this notion of minimum-cost distance, observe next that as with shortest-path distances, there is no guarantee that minimum-cost distances can actually be realized by any trip. Hence, as a parallel to Definition 2.8 above, we now say that

DEFINITION 3.3. (i) For any trip structure, $T = (X, T, o, r, c, \theta, \Phi)$, the minimum-cost distance, d_c , on X is said to be *realizable* iff for each pair of locations, $x, y \in X$, there exists some trip, $t \in T(x, y)$, with

$$(3.7) \quad c(t) = d_c(x, y)$$

(ii) Each trip, $t \in T(x, y)$, satisfying (3.7) is designated as a *minimum-cost trip* from x to y in T .

In these terms, our objective is to establish conditions on trip structures, T , under which the associated minimum-cost distance function, d_c , on X is always realizable. To do so, we begin by observing that if any trip set, $T(x, y)$, is *finite*, then the existence of minimum-cost trips in $T(x, y)$ is trivial. However, if $T(x, y)$ is *infinite*, then the existence of minimum-cost trips can only be insured by imposing additional structural conditions on $T(x, y)$.

With this in mind, our approach is to assume that it is meaningful to make similarity comparisons between abstract trips in T , and that such comparisons are representable in terms of a measure of "trip dissimilarity", which serves as an appropriate notion of "distance" in T . For example, if the only relevant attributes of trips are their travel costs and their origin locations and terminal locations in the plane, then each trip can be represented by a point in five-dimensional euclidean space. Hence, the euclidean distance in this space might be the appropriate "distance" between trips. More generally, our basic assumption is that the distance on T must satisfy all the conditions of a *uniform-semimetric*:

DEFINITION 3.4. (i) For any nonempty set of abstract trips, T , a function, $\rho: T^2 \rightarrow R_+$, is designated as a *dissimilarity distance* on T iff for each trip, $t \in T$, and positive scalars, $\alpha, \epsilon > 0$, there exists some $\delta > 0$ such that the following three conditions hold for all $t_1, t_2, t_3 \in T$ with $\rho(t, t_i) < \alpha$, $i = 1, 2, 3$:

S1. (*Distinguishability*) $\rho(t_1, t_2) = 0 \Rightarrow t_1 = t_2$

S2. (*Symmetry*) $\rho(t_1, t_2) = \rho(t_2, t_1)$

S3. (*Uniformity*) $\max\{\rho(t_3, t_1), \rho(t_3, t_2)\} < \delta \Rightarrow \rho(t_1, t_2) < \epsilon$

(ii) In addition, if the distance space, (T, ρ) , is both ρ -complete and totally ρ -bounded, then (T, ρ) is designated as a *regular trip space*.

In the present context, the distinguishability condition (S1) implicitly requires that sufficiently many attributes of trips be included in trip comparisons to allow each individual trip to be identified uniquely. The symmetry condition (S2) here implies that similarity comparisons between trips, t_1 and t_2 , do not depend on which trip is labelled "1" or "2". Finally (as discussed in section 2.1), the uniformity condition (S3) implies that similarity comparisons among trips are locally consistent in the sense that if trips t_1 and t_2 are both very similar to t_3 , then they must also be very similar to each other.

Given any dissimilarity distance, ρ , on T , we now introduce an appropriate set of regularity conditions on trip structures, T , with respect to ρ :

DEFINITION 3.5. A trip structure, $T = (X, T, o, r, c, \theta, \oplus)$, is said to be ρ -regular iff (T, ρ) is a regular trip space, and in addition, the following three conditions are satisfied for all distinct locations, $x, y \in X$, all $t \in T$, $a \in R_+$, and all trip sequences (t_n) in $T(x, y, a)$:

T5. (*Cost Continuity*) $\rho(t, t_n) \rightarrow 0 \Rightarrow c(t_n) \rightarrow c(t)$

T6. (*Trip Closure*) $\rho(t, t_n) \rightarrow 0 \Rightarrow t \in T(x, y, a)$

T7. (*Trip Boundedness*) $\rho(t, t_n) \rightarrow \infty \Rightarrow t \notin T(x, y, a)$

The cost continuity condition (T5) simply requires that similar trips between the same pair of locations must exhibit similar travel costs. The trip closure condition (T6) requires that any trip which is arbitrarily close (in terms of similarity) to trips with origin point, x , terminal point, y , and costs not exceeding level, c , must also exhibit these properties. Finally, the trip boundedness condition (T7) implies that trips between the same pair of (distinct) locations can only exhibit arbitrarily large degrees of dissimilarity if the differences in their travel costs are also arbitrarily large. [It is also of interest to note that for ρ -regular trip structures, $T = (X, T, o, \tau, c, \theta, \Theta)$, the cost-positivity condition (T3) can be dispensed with. In particular, since $d_c(x, y)$ is always achieved by some trip $t \in T(x, y)$ for each distinct location pair, $x, y \in X$, and since $t \in T - \theta(X)$ implies from T2 that $c(t) > 0$, it follows from (3.5) that T3 must hold.] Given these regularity conditions on trip structures, we are now ready to establish the existence of minimum-cost trips for such structures:

THEOREM 3.2. *For any given trip structure, $T = (X, T, o, \tau, c, \theta, \Theta)$, and dissimilarity distance, ρ , on T , if T is ρ -regular, then the minimum-cost distance, d_c , for T is realizable.*

Proof: For each location, $x \in X$, it follows at once from T2 that the null trip, $\theta(x)$, is a minimum-cost trip in $T(x, x)$. Hence we need only establish the existence of minimum-cost trips in each trip set, $T(x, y)$, with $x \neq y$. To do so, observe first that if there exists no $t \in T(x, y)$ with $c(t) = d_c(x, y)$, then we may always construct a trip sequence (t_n) in $T(x, y)$ for which the associated cost sequence $(c(t_n))$ is monotone decreasing and satisfies $c(t_n) \rightarrow d_c(x, y)$. In particular, this trip sequence (t_n) is by construction contained in the trip set, $T[x, y, c(t_1)]$. With these observations, our next objective is to show that the set, $T[x, y, c(t_1)]$ is sequentially ρ -compact. To do so, observe first from T7 that $T[x, y, c(t_1)]$ must be ρ -bounded. For if $T[x, y, c(t_1)]$ were not contained in any ρ -neighborhood of t_1 , then one could always construct a trip sequence (t_n°) in $T[x, y, c(t_1)]$ with $\rho(t_1, t_n^\circ) \rightarrow \infty$, which would contradict T7, since $t_1 \in T[x, y, c(t_1)]$ by construction. Hence $T[x, y, c(t_1)]$ is ρ -bounded. But since T6 implies by definition that $T[x, y, c(t_1)]$ is also ρ -closed, and since each regular trip space, (T, ρ) , is by definition a regular uniform-distance space, it then follows from Proposition

2.1 that (T, ρ) must be sequentially ρ -compact. Thus for the sequence (t_n) in $T[x, y, c(t_1)]$ there must exist a ρ -convergent subsequence (t'_n) , which by definition must satisfy $\rho(t, t'_n) \rightarrow 0$ for some $t \in T$. But by T6, it must then be true that $t \in T[x, y, c(t_1)]$, which together with $\rho(t, t'_n) \rightarrow 0$ implies from T5 that $c(t'_n) \rightarrow c(t)$. Finally, since $c(t_n) \rightarrow d_c(x, y)$ implies that $c(t'_n) \rightarrow d_c(x, y)$ must also hold [Bartle, 1964, Lemma 3.11.10], we may conclude that $c(t) = d_c(x, y)$. Hence we have found a minimum-cost trip, $t \in T[x, y, c(t_1)] \subseteq T(x, y)$, for each distinct pair of locations, $x, y \in X$, and the result is established. ■

This existence result can be regarded as an extension of Proposition 2.7 above, in which the *supremum distance*, d^+ , between paths is replaced by the more general notion of *dissimilarity distance*, ρ , between trips. However, a close comparison of these results shows that the rich structure of continuous paths on regular uniform-quasimetric spaces yields much sharper existence conditions than are possible for general trip structures.

4. MINIMUM-COST DISTANCES GENERATED BY PATH LENGTHS

It is clear that physical trips must involve not only a specific origin location, x , and terminal location, y , but also a specific *path* from x to y in the relevant location space, X . Hence for physical trips (or commodity shipments) it is reasonable to assume that travel costs are influenced by some relevant notion of *path length* in X . [Even in more general types of spatial interactions, such as messages through a communication network, alternative paths (or channels) may involve significantly different path lengths in terms of additive time lags or communication costs.] Hence, in such situations, it is of interest to consider conditions under which trip costs may be expressed as functions of their associated path lengths.

With this in mind, our objective in this final section is to establish such conditions for trip structures, $T = (X, T, o, \tau, c, \theta, \oplus)$, in which each trip, $t \in T$, is associated with a path, p , in some underlying continuous path structure, $P = (X, N, d)$. We assume that all such trips are physically realizable, and hence that N contains no infinite paths, i.e. that P is a *finite* path structure. In addition, we assume that (X, d) is a regular

uniform-distance space, and hence that P is a d -regular path structure (Definition 2.6). In this context, it is postulated that there exists a function, $p: T \rightarrow N$, which assigns each trip, $t \in T$, to a unique path, $p(t) \in N$. Moreover, since d -standard parameterizations of every path in N are available in P (by Proposition 2.4), we assume for convenience that every trip is assigned to the d -standard parameterization of a path in N , i.e. that path, $p(t)$, is an element of the d -standard representation, N_d , of N (Definition 2.7). With these observations, if for any trip-dissimilarity measure, ρ , on T we now designate a trip sequence (t_n) in T to be ρ -bounded iff the set $\{t_n: n \in \mathbb{Z}_{++}\}$ is ρ -bounded, and if for each function, p , on T and subset $S \subseteq T$, we denote the image of S by $p(S) = \{p(t): t \in S\}$, then we may formalize such standard path assignments as follows:

DEFINITION 4.1. For any ρ -regular trip structure, $T = (X, T, o, r, c, \theta, \oplus)$, and finite d -regular path structure, $P = (X, N, d)$, on X , a function, $p: T \rightarrow N_d$, is designated as a standard path assignment on T iff the following six conditions are satisfied for all $x, y \in X$, $[\alpha, \beta] \subseteq [0, 1]$, $t \in T - \theta(X)$, $(t_1, t_2) \in \Delta(T^2)$, $p \in N_d$, and trip sequences (t_n) in T :

- A1. (Null Consistency) $p(t) = p_x \in P_x^o \Leftrightarrow t = \theta(x)$
- A2. (Locational Consistency) $t \in T(x, y) \Rightarrow p(t) \in N(x, y)$
- A3. (Trip-Chain Consistency) $p(t_1 \oplus t_2) \sim p(t_1) \circ p(t_2)$
- A4. (Assignment Continuity) $\rho(t, t_n) \rightarrow 0 \Rightarrow d^+[p(t), p(t_n)] \rightarrow 0$
- A5. (Bounded Similarity) $d^+[p(t), p(t_n)] \rightarrow 0 \Rightarrow (t_n)$ is ρ -bounded.
- A6. (Subpath Closure) $p \in p(T) \Rightarrow p(\alpha, \beta) \in p(T)$.

The first three conditions specify the obvious physical consistency relations which must hold between trips and their associated paths, namely that: (A1) each null trip, $\theta(x)$, must correspond to the null path, p_x ; (A2) origin locations and termination locations for each trip, t , must correspond to those of its associated path, $p(t)$; and (A3) paths assigned to trip chains, $t_1 \oplus t_2$, must correspond (under appropriate reparameterizations) to concatenations of the paths assigned to the individual trips, t_1 and t_2 . [Parametric equivalence of the paths, $p(t_1 \oplus t_2)$ and $p(t_1) \circ p(t_2)$ implies that their associated d -standard parameterizations must be identical (Smith, 1987a, Proposition 11.5).] The next two conditions, which both involve perceived dissimilarities

between trips, are more behavioral in nature. The assignment-continuity condition (A4) essentially requires that similar trips be assigned to similar paths. The bounded-similarity condition (A5) may be viewed as a partial converse of condition (A4), which essentially requires that trips assigned to very similar paths cannot themselves be drastically different. Finally, the subpath-closure condition (A6) simply requires that the set of paths, $p(T)$, be sufficiently rich to form the basis for a path network on X [as developed in section 5 below].

Given any standard path assignment, $p: T \rightarrow N_d$, relating trips in T to paths in N_d , it is meaningful to compare the travel cost, $c(t)$, of each trip, $t \in T$, with the d -length, $\ell_d[p(t)]$, of its associated path. With this in mind, we now say that:

DEFINITION 4.2. For any standard path assignment, p , from a ρ -regular trip structure, $T = (X, T, o, r, c, \theta, \otimes)$, to a finite d -regular path structure, $P = (X, N, d)$ on X , the travel-cost function, c , is said to be ℓ_d -generated iff for all trips, $t_1, t_2 \in T$,

$$(4.1) \quad c(t_1) \leq c(t_2) \Leftrightarrow \ell_d[p(t_1)] \leq \ell_d[p(t_2)] .$$

In other words, c is ℓ_d -generated iff travel costs in T are monotonically related to path lengths in P . With these definitions, our main result is to characterize those functional relations between path lengths and travel costs which are consistent with (4.1). To do so, observe first that if the set of non-null trips in T is denoted by, $T_+ = T - \theta(X)$, then the image set of all path lengths corresponding to non-null paths under path assignment, p , is given by $\ell_d[p(T_+)]$. Next, for purposes of the analysis to follow, it is convenient to denote typical path-length values by $\ell, \ell' \in \ell_d[p(T_+)]$. With this notion, recall from the additivity of d -length, ℓ_d , together with the trip-chain closure property (T4) of T_+ and trip-chain consistency property (A3) of p , that the image set, $\ell_d[p(T_+)]$ is *additively closed*, i.e. that $\ell, \ell' \in \ell_d[p(T_+)] \Rightarrow \ell + \ell' \in \ell_d[p(T_+)]$. Hence we now designate a function, $\phi: \ell_d[p(T_+)] \rightarrow \mathbb{R}$, to be *subadditive* iff $\phi(\ell + \ell') \leq \phi(\ell) + \phi(\ell')$ holds for all $\ell, \ell' \in \ell_d[p(T_+)]$. With this notation and terminology, we now have:

THEOREM 4.1. For any ρ -regular trip structure, $T = (X, N, o, r, c, \theta, \oplus)$, finite d -regular structure, $P = (X, N, d)$, and standard path assignment, $p: T \rightarrow N_d$, the travel-cost function, $c: T \rightarrow R_+$, is ℓ_d -generated if and only if there exists a subadditive, continuous, increasing positive function, $\phi: \ell_d[p(T_+)] \rightarrow R_{++}$, such that for all $t \in T_+$,

$$(4.2) \quad c(t) = \phi(\ell_d[p(t)])$$

Proof: First observe that if (4.2) holds for any increasing function, ϕ , then since $\ell_d[p(t)] = 0 \Leftrightarrow p(t) \in P_X^0 \Leftrightarrow t \in \theta(X) \Leftrightarrow c(t) = 0$, it follows that the travel-cost function, c , is automatically ℓ_d -generated in the sense of (4.1). Hence it suffices to show that if the function, c , is ℓ_d -generated then there must exist a subadditive, continuous, increasing function, $\phi: \ell_d[p(T_+)] \rightarrow R_{++}$, satisfying (4.2). To do so, observe first from the finiteness of P that $\ell_d[p(T_+)] \subseteq R_{++}$. Hence if for each positive number, $\ell \in \ell_d[p(T_+)]$, we set

$$(4.3) \quad \phi(\ell) = c(t) \text{ for any } t \in T_+ \text{ with } \ell_d[p(t)] = \ell$$

then it follows at once from (4.1) that for all $t_1, t_2 \in T_+$,

$$(4.4) \quad \begin{aligned} \ell_d[p(t_1)] = \ell_d[p(t_2)] &\Rightarrow \ell_d[p(t_1)] \begin{bmatrix} N \\ M \end{bmatrix} \ell_d[p(t_2)] \\ &\Rightarrow c(t_1) \begin{bmatrix} N \\ M \end{bmatrix} c(t_2) \\ &\Rightarrow c(t_1) = c(t_2) \end{aligned}$$

and hence that (4.3) yields a well defined function, $\phi: \ell_d[p(T_+)] \rightarrow R_{++}$, which by construction satisfies (4.2). Moreover, it follows at once from (4.1) and (4.4) that ϕ must be increasing. Hence it remains to show that ϕ is subadditive and continuous. Turning first to subadditivity, observe that for any $\ell_1, \ell_2 \in \ell_d[p(T_+)]$ with $\ell = \ell_1 + \ell_2 \in \ell_d[p(T_+)]$, there must exist some trip, $t \in T_+$, with $\ell_d[p(t)] = \ell$. Hence, if we now write $p(t) = p_t$, then it follows from Proposition 2.3 above, together with the d -regularity of the path structure, P , that the function, $\ell_d[p_t \langle 0, \cdot \rangle]: [0, 1] \rightarrow [0, \ell]$, is continuous. Hence (by the intermediate-value property of continuous functions) there exists for each value, $\ell_1 > 0$, some $\alpha \in (0, 1)$ with $\ell_d[p_t \langle 0, \alpha \rangle] = \ell_1$. But it

then follows from A2 and A6 that there must exist trips, $t_1 \in T_+[p_t(0), p_t(\alpha)]$ and $t_2 \in T_+[p_t(\alpha), p_t(1)]$, with $p(t_1) = p_t\langle 0, \alpha \rangle$ and $p(t_2) = p_t\langle \alpha, 1 \rangle$, so that in particular, $l_d[p(t_i)] = l_i$, $i = 1, 2$. Moreover, since $(t_1, t_2) \in \Delta(T^2)$, it follows from the additivity of path length, together with A3 and (2.4), that $p(t_1 \oplus t_2) = p(t_1) \circ p(t_2) = p(t) \Rightarrow l_d[p(t_1 \oplus t_2)] = l_d[p(t_1)] + l_d[p(t_2)] = l_d[p(t)]$. Hence by T4, it follows that

$$\begin{aligned}
 (4.5) \quad \phi(l_1 + l_2) &= \phi(l) = \phi(l_d[p(t)]) \\
 &= \phi(l_d[p(t_1 \oplus t_2)]) = c(t_1 \oplus t_2) \\
 &\leq c(t_1) + c(t_2) = \phi(l_1) + \phi(l_2)
 \end{aligned}$$

and we may conclude that ϕ is subadditive. Thus, it remains to establish the continuity of ϕ . To do so, consider any positive value, $l \in l_d[p(T_+)]$, and observe from the monotonicity of ϕ that it suffices to show that $l_n \rightarrow l \Rightarrow \phi(l_n) \rightarrow \phi(l)$ for every sequence (l_n) in $l_d[p(T_+)]$ which is either increasing or decreasing. Suppose first that (l_n) is an increasing sequence. Then again choosing any $t \in T_+$ with $l_d[p(t)] = l > 0$ and letting $p_t = p(t) \in N_d$, it follows from the continuity of $l_d[p_t\langle 0, \cdot \rangle]$ that there exist some $\alpha_n \in (0, 1)$ with $l_n = l_d[p_t\langle 0, \alpha_n \rangle]$. Moreover, since $l_d[p_t] > 0$, it follows from the positive linearity property (2.13) of d-standard parameterizations that each α_n is unique. Hence for this sequence (α_n) we must have

$$(4.6) \quad l_n \rightarrow l \Rightarrow \alpha_n \rightarrow 1.$$

But by conditions A2 and A6, it again follows that there is some trip, $t_n \in T_+[p_t(0), p_t(\alpha_n)]$, with $p(t_n) = p_t\langle 0, \alpha_n \rangle$. For this sequence of trips, we next show that

$$(4.7) \quad d^+[p(t), p(t_n)] \rightarrow 0$$

i.e. that $p(t_n)$ converges to $p(t)$ in the uniform-quasimetric space, (N, d^+) . To do so, recall from expression (2.3) that for each index value, $\sigma \in [0, 1]$, $p(t_n)(\sigma) = p_t\langle 0, \alpha_n \rangle(\sigma) = p_t(\sigma \alpha_n)$. Moreover, since the d-continuity of the function, $p_t: [0, 1] \rightarrow X$, implies from Proposition 2.4(ii) that p_t is uniformly d-continuous, it follows that for each $\epsilon > 0$ there is some $\delta > 0$ such that

$|\alpha - \beta| < \delta \Rightarrow d[p_t(\alpha), p_t(\beta)] < \epsilon$ for all $\alpha, \beta \in [0, 1]$. Hence by (4.6) we may choose $n_\epsilon \in \mathbb{Z}_{++}$ sufficiently large to insure that $1 - \alpha_n < \delta$ for all $n \geq n_\epsilon$, so that for all $\sigma \in [0, 1]$,

$$(4.8) \quad \begin{aligned} n \geq n_\epsilon &\Rightarrow \sigma - \sigma\alpha_n \leq 1 - \alpha_n < \epsilon \\ &\Rightarrow d[p_t(\sigma), p_t(\sigma\alpha_n)] < \epsilon \\ &\Rightarrow d[p(t)(\sigma), p(t_n)(\sigma)] < \epsilon. \end{aligned}$$

Thus, $n \geq n_\epsilon \Rightarrow d^+[p(t), p(t_n)] = \sup\{d[p(t)(\sigma), p(t_n)(\sigma)]: \sigma \in [0, 1]\} < \epsilon$, and it follows that (4.7) must hold. But by A5, this in turn implies that the sequence (t_n) is ρ -bounded, so that (by the same argument as in Theorem 3.1 above) the regularity of (T, ρ) implies the existence of a ρ -convergent subsequence (t'_n) with $\rho(t', t'_n) \rightarrow 0$ for some $t' \in T$. Given this limiting trip, t' , we next show that $p(t') = p(t)$. To do so, let $l'_n = l_d[p(t'_n)]$ for all n , and observe that since (l'_n) is an increasing sequence, and ϕ is an increasing function, it follows that the sequence, $c(t'_n) = \phi(l'_n)$, is also increasing. But since $t'_n \in T_+ \Rightarrow c(t'_n) > 0$, it then follows from T5 that $\rho(t', t'_n) \rightarrow 0 \Rightarrow c(t'_n) \rightarrow c(t') > 0$, and hence from T2 that $t' \in T_+$. Thus by A4, we must have $\rho(t', t'_n) \rightarrow 0 \Rightarrow d^+[p(t'), p(t'_n)] \rightarrow 0$. Moreover, since (4.7) implies that $d^+[p(t), p(t'_n)] \rightarrow 0$, and since (N, d^+) is a uniform-quasimetric space, it then follows from part (i) of Proposition 2.4 that

$$(4.9) \quad \max\{d[p(t'), p(t'_n)], d[p(t), p(t'_n)]\} \rightarrow 0 \Rightarrow p(t') = p(t).$$

This equality of paths, $p(t')$ and $p(t)$, together with $\rho(t', t'_n) \rightarrow 0$, in turn implies from (4.9) together with T5 that

$$(4.10) \quad \begin{aligned} \phi(l'_n) - \phi(l_d[p(t'_n)]) &= c(t'_n), \quad n \in \mathbb{Z}_{++} \\ &\Rightarrow \phi(l'_n) \rightarrow c(t') = \phi(l_d[p(t')]) = \phi(l_d[p(t)]) = \phi(l) \\ &\Rightarrow \phi(l'_n) \rightarrow \phi(l). \end{aligned}$$

Finally, since $(\phi(l'_n))$ is a subsequence of the increasing sequence $(\phi(l_n))$, and since $l_n < l \Rightarrow \phi(l_n) < \phi(l)$, we may conclude from (4.10) that $\phi(l_n) \rightarrow$

$\phi(\ell)$. To establish the same result for each decreasing sequence (ℓ_n) with $\ell_n \rightarrow \ell$, choose $t^\circ \in T_+$ to be any trip with $\ell_d[p(t^\circ)] = \ell_1 > 0$, and let $\alpha \in (0,1)$ and $\alpha_n \in (0,1)$ be defined, respectively, to be the (unique) values with $\ell = \ell_d[p_{t^\circ}\langle 0, \alpha \rangle]$ and $\ell_n = \ell_d[p_{t^\circ}\langle 0, \alpha_n \rangle]$, so that the relation, $\ell_n > \ell \Rightarrow \alpha_n > \alpha$, together with the continuity of $\ell_d[p_{t^\circ}\langle 0, \cdot \rangle]$ now yields the following parallel of (4.6):

$$(4.11) \quad \ell_n \rightarrow \ell \Rightarrow \alpha_n \rightarrow \alpha .$$

Hence, choosing any $t \in T_+$ with $p(t) = p_{t^\circ}\langle 0, \alpha \rangle$ and $t_n \in T_+$ with $p(t_n) = p_{t^\circ}\langle 0, \alpha_n \rangle$, it follows from the uniform d-continuity of path, p_{t° , together with the argument in (4.8), that (4.7) now holds for this choice of trips in T_+ . Thus by A5 and the regularity of (T, ρ) , this again implies the existence of a ρ -convergent subsequence (t'_n) with $\rho(t', t'_n) \rightarrow 0$ for some $t' \in T$. Moreover, since $\ell > 0 \Rightarrow c(t'_n) = \phi(\ell'_n) > \phi(\ell) > 0$, where $\ell'_n = \ell_d[p(t'_n)]$, it again follows that $t' \in T_+$. Hence the same argument leading to (4.9) and (4.10) now shows that $\phi(\ell'_n) \rightarrow \phi(\ell)$ must hold for this decreasing subsequence (ℓ'_n) of (ℓ_n) . But since $\ell_n > \ell \Rightarrow \phi(\ell_n) > \phi(\ell)$, it again follows that $\phi(\ell'_n) \rightarrow \phi(\ell) \Rightarrow \phi(\ell_n) \rightarrow \phi(\ell)$, and we may conclude that ϕ is continuous on $\ell_d[p(T_+)]$. ■

Hence, for any given standard path assignment, p , between T and P , it follows that every ℓ_d -generated travel-cost function, c , for T must be related to ℓ_d (for all non-null trips) by a continuous, subadditive, increasing positive function. In particular, the restriction of any continuous, subadditive, increasing function, $\phi: R_{++} \rightarrow R_{++}$, to the positive domain, $\ell_d[p(T_+)] \subseteq R_{++}$, automatically yields such a function. Hence if we now designate each such function, $\phi: R_{++} \rightarrow R_{++}$, as a *travel-cost generator*, then it is of interest to examine this class of generator functions in more detail. In particular, it is useful to establish sufficient conditions for membership in this class of functions. With this in mind, we now say that:

DEFINITION 4.3. A function, $\phi: R_{++} \rightarrow R_{++}$, is said to exhibit *nonincreasing averages* iff for all $\ell, \ell' \in R_{++}$,

$$(4.12) \quad 0 < \ell < \ell' \Rightarrow \phi(\ell)/\ell \geq \phi(\ell')/\ell' .$$

With this definition, our first result [which closely parallels Theorem 103 in Hardy, Littlewood, and Polya (1934)] is to show that this condition is sufficient to insure subadditivity:

PROPOSITION 4.2. *Each continuous nondecreasing function, $\phi: \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$, exhibiting nonincreasing averages is a trip-cost generator.*

Proof: It suffices from Theorem 4.1 to show that (4.12) implies subadditivity of ϕ . But for all $\ell, \ell' \in \mathbb{R}_{++}$,

$$\begin{aligned}
 (4.13) \quad \frac{\phi(\ell + \ell')}{\ell + \ell'} &\leq \min \left[\frac{\phi(\ell)}{\ell}, \frac{\phi(\ell')}{\ell'} \right] \\
 &\leq \frac{\ell}{\ell + \ell'} \left[\frac{\phi(\ell)}{\ell} \right] + \frac{\ell'}{\ell + \ell'} \left[\frac{\phi(\ell')}{\ell'} \right] \\
 &= \frac{\phi(\ell) + \phi(\ell')}{\ell + \ell'} \\
 &\Rightarrow \phi(\ell + \ell') \leq \phi(\ell) + \phi(\ell') .
 \end{aligned}$$

Hence ϕ is subadditive, and the result is established. ■

As a second result, it is of interest to observe that every increasing concave function, $\phi: \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$, necessarily satisfies these conditions, i.e. that:

PROPOSITION 4.3. *Every increasing concave function, $\phi: \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$, is a trip-cost generator.*

Proof: Since every concave function on \mathbb{R}_{++} is continuous, it suffices from Proposition 4.2 to show that every positive increasing concave function exhibits nonincreasing averages. To do so, observe first that since ϕ is increasing and positive, $\ell \rightarrow 0$ implies that the limit, $\phi(\ell) \rightarrow \alpha$, exists and is nonnegative. Hence setting $\phi(0) = \alpha$, it follows that ϕ has a unique continuous (and concave) extension to \mathbb{R}_+ , and thus that for any $\ell, \ell' \in \mathbb{R}_{++}$

$$(4.14) \quad \phi(\ell) = \phi \left[\frac{\ell'}{\ell + \ell'} (0) + \frac{\ell}{\ell + \ell'} (\ell + \ell') \right]$$

$$\geq \frac{l'}{l+l'} [\phi(0)] + \frac{l}{l+l'} [\phi(l+l')] \geq \frac{l}{l+l'} [\phi(l+l')]$$

$$\Rightarrow \frac{\phi(l)}{l} \geq \frac{\phi(l+l')}{l+l'}$$

and it follows from the arbitrary choice of $l, l' \in R_{++}$ that ϕ must exhibit nonincreasing averages. ■

This class of concave functions (as exemplified by the illustration in Figure 1 below) includes most of the travel cost functions typically employed in the literature. In particular, travel costs typically involve nonzero *start-up costs* which are independent of trip length, and which thus create a discontinuity at $l = 0$ (in view of the zero cost of all null trips). In addition, it has been observed by many writers [see for example, Hoover (1948), Locklin (1954), and Isard (1956)] that direct transport expenditures tend to increase at a *decreasing rate*, reflecting long-haul scale economies, as well as typical quantity-discount pricing policies.

Figure 1

5. MINIMUM-COST DISTANCES AND SHORTEST-PATH DISTANCES

Given the concept of travel costs generated by path lengths, it is of interest in this final section to relate the corresponding concepts of minimum-cost distance and shortest-path distance. To do so, consider any standard path assignment, $p: T \rightarrow N_d$, from a ρ -regular trip structure, $T = (X, T, o, \tau, c, \theta, \oplus)$, to a finite d -regular path structure, $P = (X, N, d)$. If the travel cost function, c , is l_d -generated, then it is natural to conjecture that the associated minimum-cost distance, d_c , is intimately related to the underlying shortest N -path distance, d . However, a moment's reflection shows that this relationship depends critically on the nature of the image set, $p(T)$. In particular, if $p(T)$ generates the path network, N , then (as we shall see in Theorem 5.2 below) there is a direct functional relation between these

two distances. But if N is not generated by $p(T)$, then there is only a weak inequality relation between d_c and d .

To clarify the nature of this relationship, our approach will be to show that each standard path assignment, p , generates a natural substructure, $P_p = (X, N_p, d_p)$, of the path structure, P , for which there always exists a precise functional relationship between d_c and the shortest N_p -path distance, d_p . To do so, we first show that:

LEMMA 5.1. For each standard path assignment, $p: T \rightarrow N_d$, from a p -regular trip structure, T , to a finite d -regular path structure, $P = (X, N, d)$, there exists a unique smallest path network, $N_p \subseteq N$, such that $p(T) = (N_p)_d$, i.e. such that $p(T)$ is the d -standard representation of N_p .

Proof: Observe first that the trip-connectivity condition (T1) of trip structures, T , implies from A2 that $p(T)$ satisfies the connectedness property (B1) of path bases. Moreover, since A6 implies that $p(T)$ is also subpath closed (B2), it follows from Definition 2.3 that $p(T)$ is a path base on X . Hence, by Proposition 2.2, $p(T)$ generates a unique path network, $N_p = N[p(T)]$. In addition, $p(T) \subseteq N_d \subseteq N$ implies from (2.5) [together with the closure properties, N2 and N4, of networks] that $N_p \subseteq N$, and hence by construction that N_p is the smallest path network in N containing $p(T)$. To establish that $p(T)$ is the d -standard representation of N_p , observe first that for any path, $p_1 \circ \dots \circ p_n \in p(T)^n$, $n \in \mathbb{Z}_{++}$, there must exist trips, $t_1, \dots, t_n \in T$ with $p(t_i) = p_i$, $i = 1, \dots, n$. Hence by the chain-closure property (T4) of T and the trip-chain consistency property (A3) of p , it follows that $p_1 \circ \dots \circ p_n = p(t_1) \circ \dots \circ p(t_n) \sim p(t_1 \oplus \dots \oplus t_n) \in p(T)$. But the construction of N_p in (2.5) then implies that each path in N_p must be the reparameterization of some path in $p(T)$. Finally, since $p(T) \subseteq N_d$ implies by definition that each path in $p(T)$ must be its own d -standard parameterization, and since it may be verified [Proposition 11.5 in Smith (1987a)] that the d -standard parameterization of a path is identical with that of each of its reparameterizations (i.e. that $q \in [p] \Rightarrow q_d = p_d$), we may thus conclude that $p(T) = (N_p)_d$. ■

Given this result, observe next if for any given standard path assignment, $p: T \rightarrow N_d$, we define the function, $d_p: X^2 \rightarrow R_+$, for all $x, y \in X$ by

$$(5.1) \quad d_p(x,y) = \inf\{\ell_d(p) : p \in N_p(x,y)\}$$

then it may readily be verified [from the finiteness of the path structure, $P = (X,N,d)$, together with Proposition 9.3 in Smith (1987a)] that the triple, $P_p = (X,N_p,d_p)$, yields a well defined substructure of $P = (X,N,d)$. With these observations, we may now establish the following functional relationship between the minimum-cost distance, d_c , and the shortest N_p -distance, d_p :

THEOREM 5.1. *For any standard path assignment, $p: T \rightarrow N_d$, from a ρ -regular trip structure, $T = (X,T,o,\tau,c,\theta,\oplus)$, to a finite d -regular path structure, $P = (X,N,d)$, if c is ℓ_d -generated and if $\phi: \ell_d[p(T_+)] \rightarrow R_{++}$ is any trip-generator function satisfying (4.2), then for all distinct locations, $x,y \in X$,*

$$(5.2) \quad d_c(x,y) = \phi[d_p(x,y)]$$

Proof: Simply observe that since there exist minimum-cost trips, say $t_{xy} \in T(x,y)$ for each $x,y \in X$ by Theorem 4.1, it follows at once from the increasing monotonicity of ϕ that for all distinct $x,y \in X$ and $t \in T(x,y)$

$$(5.3) \quad \begin{aligned} d_c(x,y) = c(t_{xy}) \leq c(t) &\Rightarrow \ell_d[p(t_{xy})] \leq \ell_d[p(t)] \\ &\Rightarrow \ell_d[p(t_{xy})] = \min\{\ell_d[p(t)] : t \in T(x,y)\} \\ &\Rightarrow \ell_d[p(t_{xy})] = d_p(x,y) \end{aligned}$$

Finally, since $x \neq y \Rightarrow t_{xy} \notin \theta(X)$, it follows from A1 that $p(t_{xy}) \in T_+$, and hence from (4.2) that $d_c(x,y) = c(t_{xy}) = \phi(\ell_d[p(t_{xy})]) = \phi[d_p(x,y)]$. Thus (5.2) holds for all distinct $x,y \in X$, and the result is established. ■

Notice also from the proof of Theorem 5.1 that whenever trip costs, c , are ℓ_d -generated, each minimum-cost trip, t_{xy} , in T must always be assigned to a shortest path, $p(t_{xy})$, from x to y in P_p . As a final consequence of the general result in Theorem 5.1, we obtain the following desired relationships between d_c and the shortest N -path distance, d :

THEOREM 5.2. (i) For any standard path assignment, $\mathbf{p}: T \rightarrow N_d$, from a ρ -regular trip structure, $\mathbf{T} = (X, T, o, \tau, c, \theta, \Theta)$, to a finite d -regular path structure, $\mathbf{P} = (X, N, d)$, if c is ℓ_d -generated and if $\phi: \ell_d[\mathbf{p}(T_+)] \rightarrow R_{++}$ is any trip-generator function of the form (4.2), then for all distinct locations, $x, y \in X$,

$$(5.4) \quad d_c(x, y) \geq \phi[d(x, y)]$$

(ii) If in addition it is true that $N_{\mathbf{p}} = N$, then for all distinct locations $x, y \in X$,

$$(5.5) \quad d_c(x, y) = \phi[d(x, y)].$$

Proof: (i) First, to verify that the values, $\phi[d(x, y)]$, are actually well defined, observe from the subpath-closure property (A6) of $\mathbf{p}(T) \subseteq N$, together with the continuity of the function, $\ell_d[\mathbf{p}\langle 0, \cdot \rangle]$, in Proposition 2.5, that for any value, $\ell_d(p) \in \ell_d[\mathbf{p}(T)]$, it must be true that $(0, \ell_d(p)] \subseteq \ell_d[\mathbf{p}(T_+)]$. But for any distinct $x, y \in X$, it then follows from the connectedness property of T , together with (2.6) and the locational-consistency property (A2) of \mathbf{p} , that $0 < d(x, y) \leq \ell_d(p)$ for each $p \in \mathbf{p}[T(x, y)]$. Hence $d(x, y) \in \ell_d[\mathbf{p}(T_+)]$, and it follows that $\phi[d(x, y)]$ is well defined for each distinct $x, y \in X$. Finally, since $N_{\mathbf{p}}(x, y) \subseteq N(x, y)$ by definition, it follows from (2.6) and (5.1) that $d_{\mathbf{p}}(x, y) \geq d(x, y)$, and we may conclude from (5.2) together with the increasing monotonicity of ϕ that (5.4) must hold.

(ii) To establish the equality in (5.5), observe simply that if $N_{\mathbf{p}} = N$, then $N_{\mathbf{p}}(x, y) = N(x, y) \Rightarrow d_{\mathbf{p}}(x, y) = d(x, y)$ for all $x, y \in X$, so that for this case, (5.2) now implies (5.5). ■

6. CONCLUDING REMARKS

In this final section, we consider briefly a number of possible extensions of the present analysis. To begin with, it is important to emphasize once again that while the above framework was formulated in terms of travel behavior, the basic concepts are meaningful in essentially any behavioral context which involves movement between locations (or transitions between states) in some space. Additional economic examples include production processes, in which efficient transformation sequences from given inputs to desired outputs can be treated as "minimum-cost chains". Similar interpretations can be given to a wide range of social transition processes, such as educational or training programs in which efficient training sequences can also be viewed as minimum-cost chains. Thus, in all those processes where our basic hypothesis of *economies of trip chaining* is appropriate, minimizing behavior induces a well behaved notion of distance on the relevant state space: such as "minimum-transformation-cost" distances on production possibility spaces, or "minimum-transition-cost" distances on social state spaces.

However, it is also important to emphasize the limitations of this framework. In particular, our basic hypothesis of *economies of trip chaining* may fail to hold in many important behavioral contexts where trip chaining is meaningful. This is even true in economic settings such as the airline example mentioned in the introduction. For while airline fares may exhibit economies with respect to trip chaining, such fares are often not the only relevant costs. In particular, if one considers the time costs of travel, then it is clear that flight connections can result in significant time costs which are independent of actual flight times. More generally, there may exist significant *connection costs* in trip chaining which may lead to violations of condition (3.1). Hence as one possible extension of the present framework, it is of interest to consider the minimum-cost distances generated by trip chains with connection costs. Even in simple examples, it is clear that such distances may fail to satisfy the triangularity condition (D3), and hence may even fail to be quasimetrics.

Aside from physical costs such as connection costs, there may also exist many important *psychological costs* of movement which influence behavior. In the simple case of travel, for example, even if all physical costs exhibit

economies of trip chaining, the tiring effect of extended travel on individual trip makers will generally fail to satisfy this condition. In cases where such perceived costs are significant, the relevant notion of "minimum-cost distance" may not only fail to be a quasimetric, but may in fact be very difficult to measure. In particular, when such costs are identifiable only in an ordinal sense, the resulting distance representations will be highly nonunique. Hence, as a second extension of the present analysis, it is of interest to consider the representational properties of ordinal cost structures with respect to minimum-cost distance. Results of this type have been established by Beals and Krantz (1967) for metric representations of perceived distance orderings. Possible extensions of these results to the present framework will be considered in a subsequent paper.

Finally, turning to the relationships between minimum-cost distance and underlying measures of physical distance, it should be clear that the notion of l_d -generated travel costs in Definition 4.2 is highly restrictive. Indeed, even the physical relations between travel time and path length are seldom that simple. In particular, while such relationships may exhibit certain overall monotone tendencies, their exact nature on, say, a given road network is complicated by a host of other factors including road capacity, stop lights, and traffic congestion. More generally, it should be clear that path length is only one among many variables which influence travel costs. Hence the present notion of trip-cost generators should be regarded as only the simplest possible types of structural relations which may be considered.

6. REFERENCES

- Hartle, R.G., (1964) *The Elements of Real Analysis*, Wiley, New York.
- Beals, R. and D.H. Krantz, (1967) "Metrics and Geodesics Induced by Order Relations", *Mathematische Zeitschrift*, Vol.101, pp.285-298.
- Beguin, H. and J.-F. Thisse (1979) "An Axiomatic Approach to Geographical Space", *Geographical Analysis*, Vol. 11, pp.325-341.
- Blumenthal, L.M., (1970) *Distance Geometry*, 2nd Edition, Chelsea, New York.
- Busemann, H., (1955) *Geometry of Geodesics*, Academic Press, New York.
- Hardy, G.H., J.E. Littlewood, and G. Polya, (1934) *Inequalities*, Cambridge University Press, Cambridge, England.
- Hoover, E.M., (1948) *The Location of Economic Activity*, McGraw-Hill, New York
- Isard, W., (1956) *Location and Space Economy*, MIT Press, Cambridge, Massachusetts.
- Locklin, D.P., (1954) *Economics of Transportation*, Irwin, Homewood, Illinois.
- Smith, T.E., (1987a) *Distances in Spatial Analysis: An Axiomatic Approach*
Unpublished Manuscript, Department of Regional Science, University of Pennsylvania, Philadelphia, Pennsylvania.
- _____, (1987b) "Shortest-Path Distances: An Axiomatic Approach", Working Paper, Department of Regional Science, University of Pennsylvania, Philadelphia, Pennsylvania.

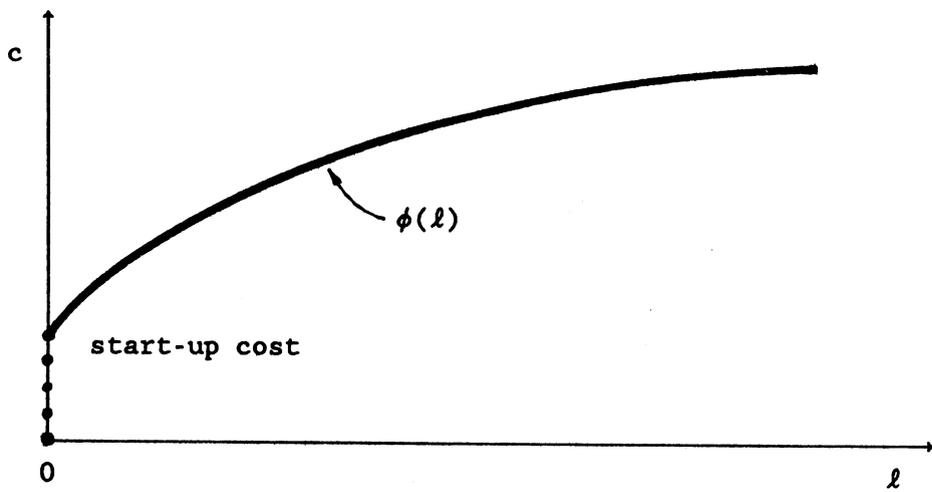


Figure 1. A Concave Trip-Cost Generator