TRANSPORTATION NETWORKS EQUILIBRATION

WITH DISCRETE CHOICE MODELS

by

Yosef Sheffi

B.Sc., The Technion, Israel Institute of Technology
(1975)

S.M., Massachusetts Institute of Technology
(1977)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1978

© Yosef Sheffi, 1978

Signature of Author. . . . . .
Department of Civil Engineering, May 5, 1978

Certified by . . . . . .
Thesis Supervisor

Certified by . . . . . .
Thesis Supervisor

Accepted by . . . . . .
Chairman
Departmental Committee on Graduate Students
of the Department of Civil Engineering

TRANSPORTATION NETWORKS EQUILIBRATION

WITH DISCRETE CHOICE MODELS

by

YOSEF SHEFFI

Submitted to the Department of Civil Engineering
on May 5, 1978 in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

ABSTRACT

The transportation planning process has been traditionally performed on a sequential, heuristic basis, with each step having a methodology of its own. This thesis suggests a unified approach, and an algorithm, within which many transportation equilibrium analyses can be carried out, using a disaggregate demand model (the multinomial probit) as an integral part of the equilibration procedure. The conditions of equilibrium in the passenger transportation market are identified and defined, the problem is cast as a mathematical program and an efficient algorithm for its solution is introduced.

The approach consists in reducing the equilibration problem to a network assignment problem over a modified network (termed hypernetwork). All choices faced by tripmakers (e.g., taking a trip, mode, destination, route, etc.) are viewed as choice of an alternative path on this (abstract) hypernetwork, to which the network formulation and equilibration algorithms are applied.

THESIS SUPERVISOR:	Fred Moavenzadeh
TITLE:	Professor of Civil Engineering
THESIS SUPERVISOR:	Carlos F. Daganzo
TITLE:	Assistant Professor of Civil Engineering

ACKNOWLEDGEMENTS

I should like to thank my thesis committee chairman, Professor Fred Moavenzadeh, whose constant patience, advice and support throughout my graduate school years at M.I.T. were invaluable, and are greatly appreciated.

Professor Carlos Daganzo supervised this thesis. His teaching, guidance and most of all, his friendship, have been most rewarding and stimulating. I would like to thank him and his wife, Val, for their wonderful hospitality during the summer of 1977, when many of the ideas of this dissertation were formulated.

Further thanks go to my thesis committee members, Professor Joseph Sussman, Professor Ann Friedlaender and Professor Steven Lerman as well as Professor Marvin Manheim and Professor Moshe Ben-Akiva for their constructive comments.

Finally, I would like to thank my wife, Anat, for her constant encouragement and understanding.

TABLE OF CONTENTS

Page

TITLE PAGE..... 1

ABSTRACT..... 2

ACKNOWLEDGEMENT..... 3

CONTENTS..... 4

LIST OF FIGURES..... 6

LIST OF TABLES..... 6

CHAPTER 1: INTRODUCTION..... 7

1.1 OBJECTIVE AND SCOPE..... 7

1.2 REVIEW OF EQUILIBRIUM TRANSPORTATION PLANNING METHODOLOGY..... 8

1.3 THE HYPERNETWORK CONCEPT.....13

1.4 THESIS OUTLINE.....20

CHAPTER 2: CHOICE THEORY AND EQUILIBRIUM BACKGROUND.....22

2.1 CHOICE THEORY BACKGROUND.....23

2.1.1 Disaggregate Demand Models and the Probit Integral.....23

2.1.2 The Approximation of the Probit Integral.....26

2.2 AGGREGATION WITH MULTINOMIAL PROBIT.....29

2.2.1 Aggregation Over Individuals and Market Segmentation...29

2.2.2 Spatial Aggregation through Normal Approximation.....32

2.2.3 Aggregation of Alternatives.....36

2.3 EQUILIBRIUM OVER NETWORKS.....39

2.3.1 On Stochastic Models of Traffic Assignment.....39

2.3.2 Approximating S-U-E by U-E for Congested Networks.....44

2.3.3 An Algorithm for the Spatial Aggregation Problem of Traffic Assignment.....48

CHAPTER 3: THE TRANSPORTATION MARKET AS A HYPERNETWORK.....58

3.1 A MODIFIED ALGORITHM FOR THE SPATIAL AGGREGATION PROBLEM.....59

3.1.1 The Modified Algorithm.....59

3.1.2 Computational Considerations.....63

Page

3.2 FORMULATING THE TRANSPORTATION MARKET AS A NETWORK.....69
3.2.1 Equilibrium Over Hypernetworks.....69
3.2.2 The Hypercentroid and Hyperzone.....74
3.2.3 The Hypernetwork Structure.....84
3.3 ASSIGNMENT OVER HYPERNETWORK.....91
3.3.1 The Hypernetwork Assignment Algorithm.....91
3.3.2 Properties of the Equilibrium Solution.....96
3.3.3 Example.....100

CHAPTER 4: SUMMARY AND APPLICATIONS.....113
4.1 SUMMARY OF THE APPROACH AND MODELLING ASSUMPTIONS.....113
4.2 APPLICATIONS AND EXTENSIONS OF HYPERNETWORK METHODOLOGY.....116

REFERENCES.....119

APPENDIX.....125

LIST OF FIGURES

Figure		Page
1-1	Mode and Route Choice Hypernetwork	17
1-2	Mode and Route Choice Hypernetwork (Independent Hyperpath Representation)	17
1-3	Mode, Destination and Route Choice Hypernetwork	19
2-1	Determination of the Moments of Intrazonal Travel Distance	34
2-2	Comparison of Stochastic Network Loading Methods	45
2-3	Network Representation for the Multicentroid Assignment Problem	51
3-1	A Modified Algorithm for the Continuous Representation	61
3-2	Approximate Relative Processing Time Savings vs. Average Number of Centroids per Zone	67
3-3	Spatial Aggregation Network Representation	75
3-4	A Hyperzone Example	77
3-5	Mode and Route Choice Hypernetwork Example	80
3-6	Equivalent Hypernetwork Representations	87
3-7	An Algorithm for Assignment Over Hypernetworks	95
3-8	Hypernetwork for the Numerical Example	102
3-9	Graphical Solution of the Example	105
3-10	Divergence Pattern of Traditional Equilibration Method	107
3-11	First Iteration of the Algorithm for the Example Problem	110

LIST OF TABLES

Table		Page
3-1	Divergence Pattern of Traditional Equilibration Scheme	106
3-2	The Convergence of the Hypernetwork Equilibrium Algorithm	111

CHAPTER 1

INTRODUCTION

1.1 OBJECTIVE AND SCOPE

This thesis presents a framework and an algorithm, within which many transportation equilibrium analyses can be carried out. It departs from the main line of thought in previous transportation planning research and applications in that it suggests that passenger transportation market forecasting problems can be dealt with in a unified way. The solution methodology consists in interpreting the sequence of choices faced by an individual about to take (or not to take) a trip, as a choice of path on an abstract network (hypernetwork).

The hypernetwork's methodology eliminates some of the biases and inconsistencies that limit the theoretical basis of existing analysis techniques and that are a major source of the high cost associated with such procedures. This alternative approach simplifies the analysis conceptually and promises significant cost savings. In this thesis, the equilibrium conditions are defined, the equilibration problem formulated, and an efficient algorithm for its solution is developed. The mathematically consistent formulation and algorithmic solution of the transportation market equilibrium equations greatly enhances the potential of disaggregate demand models since it is now possible to avoid their gross mispredictions, when applied to congested and capacitated transportation systems.

Although the urban passenger transportation planning process is used throughout this thesis as an example, it should be noted that the algorithmic

framework is not restricted to either macroscopic or microscopic applications. It can be used to study the equilibrium flow pattern over an entire urban area, or to design the capacity and location of an isolated parking lot.

The next section introduces the notion of equilibrium and discusses the transportation planning process.

1.2 REVIEW OF EQUILIBRIUM TRANSPORTATION PLANNING METHODOLOGY

The term "equilibrium" as used in this thesis (and as it is commonly used among transportation engineers and planners) refers to a consistent pattern of flows and level of service (LOS) over the transportation market. Predicting such equilibrium situation (the equilibration process) is not trivial since the performance of the transportation system (the LOS) depends, in general, on the volume (flows) of users in the system and vice versa. The function relating the LOS to the flow is commonly referred to as the "supply" side of the transportation "market".

It is useful to distinguish between the economic concept of supply and the meaning of the same term here since this distinction would clarify the scope of this thesis. The economic term relates the reaction of the consumers and firms to the market stimulus, while the latter refers to technological relationships. In the context of transportation, the economic term of supply describes the reaction of government and operators to the LOS in the system (in terms of changing capacity and performance). Such reactions are long-run phenomena and are not included among the market forces under consideration in this thesis. In contrast, the term supply in this thesis means performance function, as explained above.

A similar distinction holds with regard to the demand side of the

transportation market. Demand funtions relate the reactions of the users (passengers in our case) to the LOS offered by the system. The users reactions considered in this thesis consists in the short-run travel decisions (e.g., not to take a trip, change mode, change route, etc.). Longer-run phenomena (e.g., residential relocation, car ownership level changes, etc.) are not included in this analysis.[1] Thus, the analysis of the transportation market involves the equilibration of the short run travel decisions with the system performance (supply) functions.

Modelling the abovementioned equilibrium is the classic problem of passenger transportation planning. Sheffi and Daganzo (1978a) review a sample of the huge body of literature dealing with this so-called "transportation planning process". The process is typically modeled and conducted as a four step analysis, including the prediction of trip generation, trip distribution, modal split, and traffic assignment, where each step is associated with a methodology of its own.

The most widely used model of urban passenger transportation is the "UTPS"[2], which is a battery of computer programs designed to perform the abovementioned process. There are other computer packages that attempt to perform the transportation market equilibration process, such as "DODOTRANS"[3] (which, unlike the early versions of the UTPS, is an explicit

---

[1]Manheim (1978) terms these long-run forces on the demand side "Type II", and the abovementioned long-run supply relationships "Type III". Note, however, that technically the methodology can be extended to include this.

[2]"UTPS" stands for "UMTA Transportation Planning System"; see USDOT/UMTA (1974) for a description of the model system, and, for example, Brand (1972) for a discussion of the methodology.

[3]"DODOTRANS" stands for "Decision Oriented Data Organizer for Transportation Analysis, see Manheim and Ruiter (1970).

equilibration package). A review of many of these packages can be found in Peat, Marwick, Mitchell and Co. (1973).

Although generally accepted and widely used, the abovementioned four step process was severely criticized in the literature in the last several years. Some of the criticism is general and points out the deficiencies of all large scale models [e.g., Alonso (1968), Bolan (1970) and Lee (1973)]. Some of it is directed at specific models used in process[4]. Yet other researchers have based their criticism on a more fundamental issue -- the statistical and behavioral assumptions that underlie the treatment of the demand side. The latter line of criticism led to the so-called disaggregate travel demand models which by using individuals as the study units attempted to capture travellers' behavior[5]. Subsection 2.2.1 of this thesis contains a brief overview of the foundation and functional forms of these models.

Because of their apparent advantages, disaggregate models have gained popularity among planners and are increasingly used in practice[6]. However, disaggregate demand models have raised a set of new unresolved issues. The first of these is the aggregation problem, i.e., how to use disaggregate models to get aggregate predictions. Koppelman (1976) reviews several methodologies, none of which produces satisfactory solution [see for example Bouthelier and Daganzo (1978)]. The second difficulty (which is related to the aggregation problem) is in incorporating these

---

[4]For review of references concerning criticism of specific models see Sheffi and Daganzo (1978a).

[5]See for example Lave (1969), Reichman and Stopher (1971), McGillivray (1972), Charles River Associates (1972), Ben-Akiva (1974), and Richards (1974).

[6]Sheffi and Daganzo (1978a) review applications of disaggregate models to issues such as trip generation, trip distribution, modal split, traffic assignment, residential location and freight shipping.

models within an equilibration framework. The third one is that most disaggregate modelling effort has been with models, such as logit, that involve sometimes unrealistic assumptions and often fail to capture reasonable user behavior (see Subsection 2.2.1).

In addition, some of the issues that arise from the heuristic nature of the four step transportation planning process still remain, as happens for instance, with its failure to represent the transportation system LOS consistently, throughout all the steps. In order to circumvent this problem, it is suggested that the model system should be iterated several times, in order to achieve a state of equilibrium (which is not formally defined). However, due to the high computation costs involved, seldom is this done in practice. These iterations are the source of a body of literature concerning "feedback-loops" and "accessibility measures".[7]

Lastly, there are some specific problems in addition to the aggregation and equilibration issues. For instance, even though it was realized that market segmentation might enhance predictions [e.g., Lovelock (1975), and Nicholaidis, Wachs and Golob (1977)] and is commonly used in practice, no firm guidelines are given as to the necessary extent of the segmentation. Similarly, in the traffic assignment step, no definite criteria exist on how to represent the network, i.e., how to locate the zone centroids and select the number and characteristics of dummy centroid connectors.

In summary, the following issues are identified:

1. Concerning disaggregate demand models --

   1a. The aggregation problem (including market segmentation)

[7] See Sheffi and Daganzo (1978a) for references with regard to this point.

1b. Incorporation in equilibrium analysis

1c. Alternatives to logit.

2. Concerning the traditional process in general --

2a. Equilibrium formulation and equilibration procedure

2b. Consistency throughout the steps

2c. Network representation.

The objective of this thesis is to provide a framework within which these issues can be resolved. Some of these problems have recently been tackled. McFadden (1977) has introduced a generalized logit (The General Extreme Value Model) -- a model that eliminates some of the theoretical weaknesses of logit, while Daganzo, Bouthelier, and Sheffi (1977a and 1977b) have developed a numerical solution to probit -- both reasonable alternatives to logit (issue 1c). The network representation problem has been solved by Daganzo (1977c) through a continuous approximation of the interzonal trip-end impedance (issue 2c). Some of the abovementioned problems are partially solved, such as the aggregation problem with the aggregation method introduced by Bouthelier and Daganzo (1978) as a multivariate extention of the work of McFadden and Reid (1975) and Westin (1974), and yet some of them remain unsolved (incorporating disaggregate demand model within a formal and efficient equilibration scheme). This thesis uses some of these results and some new ideas, to formulate a solution (at least partial) to the abovementioned problems.

Several equilibrium models have been recently developed. The first ones dealt (rigorously) with the route choice and netowrk equilibrium only [e.g., Nguyen (1974) and LeBlanc (1975)] by casting the problem as a mathematical program. Ruiter and Ben-Akiva (1977) developed a complete

equilibrium forecasting system incorporating an integrated set of production oriented disaggregate models, and a conceptually similar model system was used by Jacobson (1977); both methods, however, are not guaranteed to produce the desired results (in terms of convergence to a defined equilibrium). A formal solution of the equilibration problem over a transportation corridor, using disaggregate demand models was obtained by Talvitie and Hasan (1977). Their approach consists in formulating the equilibration as a fixed point problem and solving it utilizing the algorithm proposed by Scarf (1973).

The approach taken in this thesis is to view and formulate all the choice processes as route choice processes over an abstract network (hypernetwork) and use an efficient mathematical-programming procedure to derive the equilibrium solution. The general concept of hypernetworks is introduced in the next subsection below.

1.3 THE HYPERNETWORK CONCEPT

In this thesis, the various alternatives opened to travelers in the transportation market (e.g., mode, route, destination, etc.) are viewed as paths in a hypothetical network (a hypernetwork) made up of link characterized by disutilities. It is assumed that, as in route choice problems, users select the shortest route (i.e., the alternative with the lowest disutility) from their origin to their destination. (This is merely a restatement of the utility maximization principle of choice theory).

The ideate of hypernetworks has been latent in the literature for some time. As early as 1972, at the Williamsburg conference, A. G. Wilson (1973) noted:

> "...It is tempting as computer capacity expands to think of assigning on multimodal networks, in effect, possibly directly on routes on an abstract modal basis....This is another class of mathematical aggregation problems."

Manheim (1973) tried to formulate the transportation planning process as a network assignment problem, using logit path-choice model (in the form of Dial's (1971) STOCH algorithm) imbedded in an incremental assignment equilibration.[8,9]

Dafermos (1976) suggested an integrated equilibrium flow model for transportation planning, based, again, on visualizing the whole transportation planning process as a solution to a network assignment problem. In her words,

> "...We adopt the natural behavioral assumption that each user chooses his origin, his destination, as well as his path as to minimize his "travel cost." Of course, "travel cost" should be interpreted in a very liberal fashion. In reality additional factors such as "attractiveness" of the origins (residential areas) and destinations (places of work) have to be taken into account but this can be incorporated into the model as "travel cost" by a straight-forward modification of the network....Interestingly, we establish a mathematical equivalency which reduces integrated transportation problems for a network into assignment problems for a modified network".

Dafermos' model, although very similar to the hypernetwork concept, is not quite as general for she was working exclusively with deterministic

---

[8] A similar technique, but for the traffic assignment step only, was formulated and tested recently by Fisk (1978).

[9] This approach does not solve any of the issues discussed in the preceeding section, due to the use of logit (see Subsection 2.2.1 for a review of logit's flaws and also Schneider (1973), Burrell (1976), Florian and Fox (1976) and Daganzo and Sheffi (1977) for a critique of Dial's assignment method) and the heuristic incremental equilibration procedure (see Yagar (1976) and Ferland, Florian and Achim (1975) for a discussion of the inconsistencies of incremental methods).

travel costs over the modified network. This explicitly excludes many demand models from the realm of applications of her model since, as it is assumed with deterministic equilibrium traffic assignment methods, users are identical (this excludes disaggregate demand models), fully informed (which excludes logit, probit, and other stochastic models) individuals, making consistently perfect decisions.

The well known elastic demand traffic assignment problem formulated by Beckman et al. (1956) can be solved with existing fixed demand (fixed trip rate) traffic assignment algorithms on an expanded network, as shown by Danzig et al. (1976). Such an expanded network can be viewed as a hypernetwork since it includes (in addition to the street network) dummy links going from each origin to each destination in order to represent the no-travel alternative.

Sheffi and Daganzo (1978a) cite additional references of formulations combining several steps of the planning process which can be viewed as hypernetwork formulations.

The hypernetwork concept, as developed in this thesis, is intimately related to Multinomial Probit (MNP) models, and thus present the same advantages and disadvantages of MNP models. Namely, MNP models and hypernetworks solve, or at least alleviate, the market segmentation problem, as explained in Sec. 2 of this thesis. The old issue of the proper step sequence (i.e., should mode choice be predicted before destination choice, after it or simultaneously with it), posed by Ben-Akiva (1974) (who demonstrated the feasibility of a simultaneous approach) has already been (indirectly) addressed in Sec. 1.2. The hypernetwork approach is equivalent to a simultaneous MNP choice model whose covariance matrix can be

studied visually. The hypernetwork idea is the key to performing supply-demand equilibration with disaggregate demand models over the whole transportation market, on a mathematically consistent basis (heuristic equilibration technique based on feedback loops do not necessarily converge).

To illustrate the hypernetwork methodology developed in this thesis, assume for instance that one is concerned with a modal split and route choice problem for a single origin-destination pair, and to further facilitate the concept, assume that there is only one transit mode and two automobile routes. Figure 1-1 presents a possible configuration of the hypernetwork corresponding to such problems.

In this figure there are three hyperpaths corresponding to the three alternatives. The "costs" over links OA and OB represent the flow independent components of the disutility of the two modes (e.g., socio-economic-related disutility components, comfort, privacy, etc.) and links AD and BD are associated with the actual travel impedance (e.g., travel time) of the two alternatives. Choice of, say, the top route in the figure, implies that the shortest route through the hypernetwork consists in driving a car through Route 1 of the street network.

In the most general case, link disutilities may be random (i.e., perceived and measured utilities are distinguished), flow-dependent (e.g., travel time under congested conditions or transit dwell times), fixed (e.g., transit fare, parking fee or transfer disutility) and/or multi-attributed. As will become apparent in the subsequent chapters, computational efficiency considerations require the modelling of links exhibiting flow-dependent and flow independent disutilities in a different way (see Subsection 2.3.2). Disutilities are also assumed to be additive so

Figure 1-1 Mode and Route Choice Hypernetwork

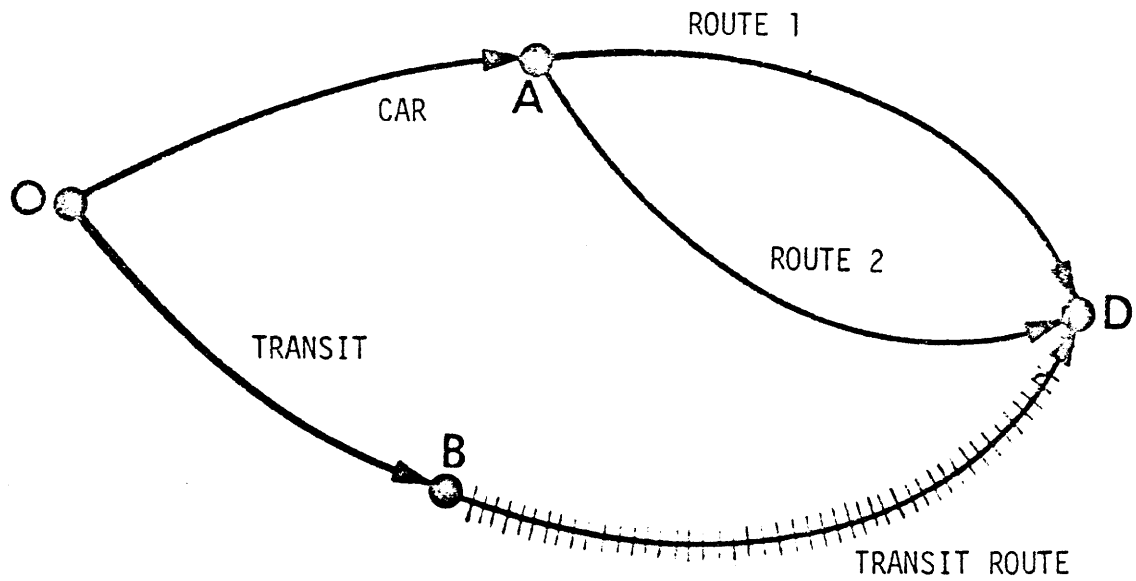
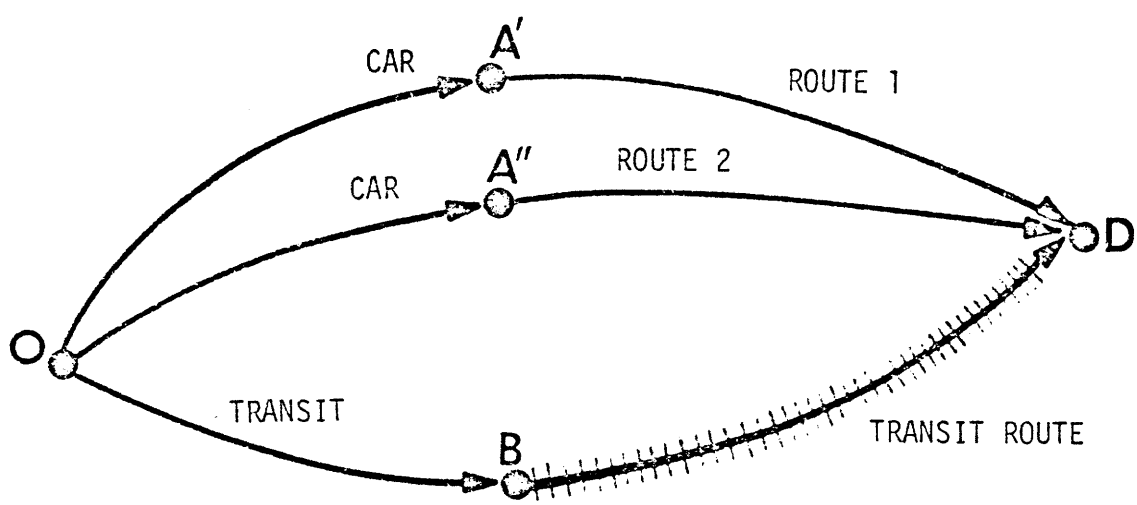


Figure 1-2 Mode and Route Choice Hypernetwork
(Independent Hyperpaths Representation)



that the disutility of an alternative (hyperpath) equals the sum of the disutilities of the links that make it up.[11]

By changing the structure of a hypernetwork, one can affect the probabilistic structure of the corresponding choice model (this will be seen in Subsection 2.3.1 which explains the effect of network topology on the choice probabilities). For instance, Figure 1-2 displays an alternative representation of the choice situation depicted in Figure 1-1, corresponding to a view of the three hyperpaths as alternatives exhibiting statistically independent utilities.[12]

Figure 1-3 illustrates a more complicated choice problem represented as a hypernetwork. It displays a (single origin) problem of combined modal split, route and destination choice, where a fraction of the population does not have access to the car mode. The links representing the street network and transit lines are associated with travel impedance. All other links represent other dimensions of travel choice and are associated with the corresponding disutility [e.g., links D_iD (i=1,2,3) are associated with destinations attraction variables]. Note that O_2 does not have access to the street network, in order to represent market segments that do not own automobiles. The number of hyperpaths in this hypernetwork is larger than in the preceeding example (in fact, in real problems this number is large enough as to preclude total enumeration of all possible hyperpaths).

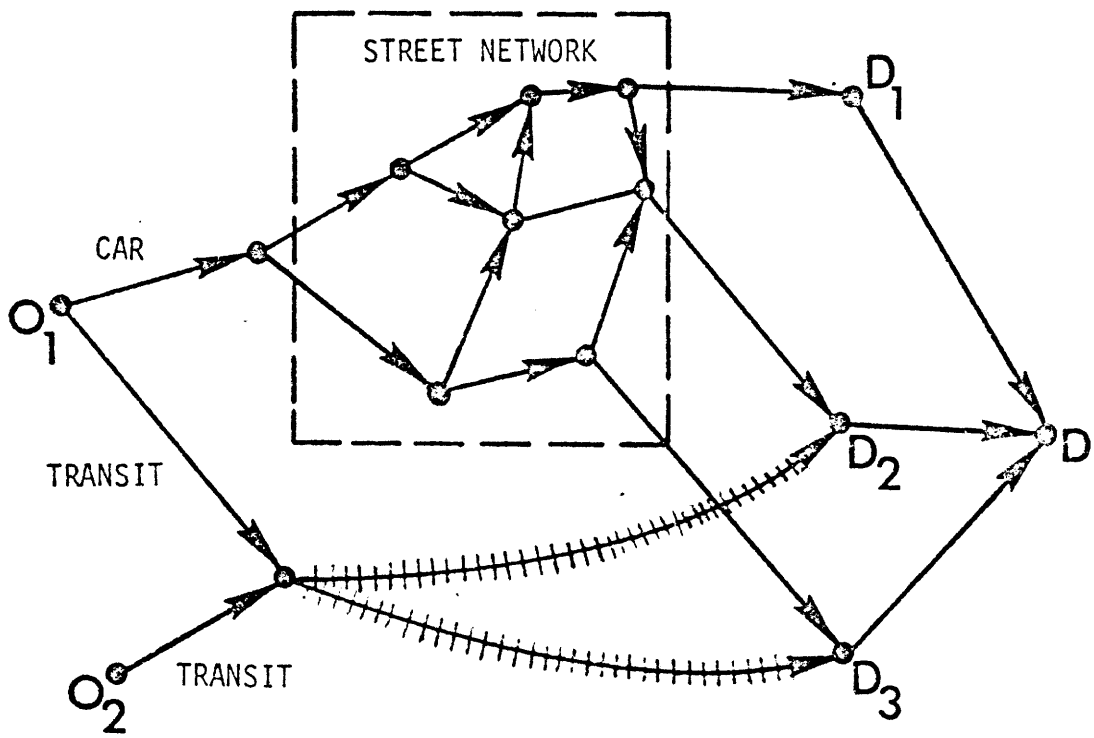
---

[11] A discussion of the additivity assumption is included in Section 3.3.

[12] This discussion corresponds to a restriction on the randomness in the process where link utilities are viewed as statistically independent variates. This point is discussed in detail in subsequent chapters of the thesis. As shown later in the thesis, such a representation corresponds approximately to the so called "independence from irrelevant alternatives" property [due to Luce (1959)].

Figure 1-3

Mode, Destination and Route Choice Hypernetwork



The hypernetwork includes one origin, two market segments (one without access to the car mode) and three destinations (one is not served by transit).

These examples were intended to demonstrate that it is possible to construct a hypernetwork for many choice problems and that different market segments can be adequately handled by appropriate representation (as will be shown later, one doesn not have to enumerate all hyperpaths in order to solve for the equilibrium flows).

1.4 THESIS OUTLINE

This research provides a framework and an algorithm for equilibrium analysis of hypernetworks. Chapter 1 has presented the problem and its context, while Chapter 2 reviews and further develops some specific analytical tools needed in the following one. Chapter 3 presents the body of the research and Chapter 4 summarizes the results and suggests further research needs.

Chapter 2 is divided into three main parts: choice theory background (Sec. 2.1), aggregation issues (Sec. 2.2), and network equilibration methodology background (Sec. 2.3). Following an introduction outlining Chapter 2, Section 2.1 presents the MNP disaggregate demand model and the method used for evaluating the MNP choice probabilities. Sec. 2.2 explores some of the MNP model's advantageous properties with regard to various aggregation problems. Section 2.3 covers the necessary network assignment and equilibration background. It includes a review of the MNP-based theory of stochastic equilibration, the rationale of the deterministic modelling of congested links and an algorithm for the spatial aggregation traffic assignment which is the basis for the equilibration algorithm presented in Chapter 3.

Chapter 3 starts with a modification of the abovementioned algorithm (Sec. 3.1). The quilibrium conditions and the hypernetwork concept are explored in Sec. 3.2, and the modified algorithm (of Sec. 3.1) is applied to the hypernetwork in Sec. 3.3. The latter section also discusses the modelling assumptions and illustrates the approach through an example.

The first section of Chapter 4 includes a brief summary of the thesis and the assumption underlying the approach. The last section (Sec. 4.2) discusses some applications and extensions of the methodology, and points out directions for further research.

CHAPTER 2

CHOICE THEORY AND EQUILIBRIUM BACKGROUND

In Chapter 2, the tools needed as background for developing the comprehensive equilibration method presented in Chapter 3 are briefly reviewed. It is divided into three basic parts: choice theory (Sec. 2.1), aggregation issues (Sec. 2.2) and equilibrium theory (Sec. 2.3).

The first subsection of Sec. 2.1 reviews choice theory and the multinomial probit (MNP) equation, while the subsequent subsection (2.1.2) presents the method for evaluating the MNP choice probabilities. Sec. 2.2 deals with the aspects of the MNP model upon which the demand side of the equilibration procedure is based. Subsection 2.2.1 illustrates how aggregation can be performed analytically with MNP models when some of these models' explatory variables are approximately normally distributed across the population. Such a normal approximation is used in Subsection 2.2.2 to obtain the intrazonal travel time distribution of trip ends. The last part of Sec. 2.3 explains the concept of expected maximum utility and its use in obtaining the total utility of the population.

Section 2.3 deals with equilibrium network assignment background. The first subsection reviews a MNP-based traffic assignment (i.e., route choice) model for networks exhibiting stochastic links costs. This subsection also points out the inefficiencies of existing algorithms for traffic assignment over stochastic and congested networks, leading to Subsection 2.3.2, where the rationale for approximating certain components of equilibrium models by deterministic ones, is presented. Such approxi-

mations are exploited in later sections of the thesis to develop the equilibration method. The last subsection of Sec. 2.3 reviews an algorithm, orginally developed to solve the spatial aggregation problem of traffic assignment. This technique serves as a basis for the hypernetwork equilibration method developed in Chapter 3.

2.1 CHOICE THEORY BACKGROUND

2.1.1 Disaggregate Demand Models and the Probit Integral

This subsection introduces the multinomial probit (MNP) disaggregate model of travel choice.[13]

Disaggregate demand models have been the central thrust of travel demand research in the last decade or so due to their following features:

(a) The use of disaggregate data for model estimation is more efficient, implying a reduction in data collection costs.

(b) The estimation is independent of the distribution of the explanatory variables -- making disaggregate models potentially more transferrable and eliminating possible biases due to prior aggregation.

(c) Some of these models are interpreted as utility maximization, giving them a flavor of causality and behavioral realism.

The hypothesis underlying these models is that when confronted with a choice situation, an individual associates a (perceived) level of attractiveness

[13] Detailed review of disaggregate models and travel choice theory can be found in a variety of references including Domencich and McFadden (1975), Manski (1973) and Richards and Ben-Akiva (1975).

(utility[14]) with each available alternative. This utility is a function of the choice maker's characreristics and the alternatives' attributes, and the choice maker is assumed to select the alternative with the greatest utility. Since utilities are not observable, they are modeled as random variables distributed across the population of choice makers.

Most operational models assume a functional form of the utility, which is linear in the parameters and with additive disturbance, i.e., the utility of alternative i to an individual chosen in random from the population, U_i, is given by:

$$U_i = \beta Z_i + \xi_i \qquad [2.1]$$

where β is a vector of parameters, Z_i is a vector of functions of characteristics of the individual under consideration and the attributes of alternative i, and ξ_i is a random variable representing an unobserved disturbance or error term.[15] The term βZ_i is usually denoted V_i and termed the observed utility (or mean utility since without loss of generality, it can be assumed that $E[U_t] = V_t$).

The disaggregate choice model is concerned with estimating the probability of each alternative being selected, given the vector of measured utilities $V = (\ldots, V_i, \ldots)$ and the joint distribution of $\xi = (\ldots, \xi_i, \ldots)$. The probability, P_i, that alternative i is selected by a (randomly chosen) individual, from his choice set S is:

---

[14] The term "utility" is used throughout this thesis to denote this level of attractiveness associated with each alternative. However, the term utility does not correspond exactly to the general meaning of this term in the economic literature -- see discussion in Subsection 3.3.2.

[15] If the parameters, β, vary from individual to individual (taste variation), the distribution of ξ depends on the characteristics and attributes, Z; such models are discussed by Hausman and Wise (1978), and Albright, Lerman and Manski (1977).

$$P_i = \Pr(\text{choose } i|S) = \Pr(U_i \geq U_j \ ; \ \forall j \varepsilon S) \ ; \ \forall i \varepsilon S \qquad [2.2]$$

Substituting Eq. [2.1] into [2.2], one gets the choice probabilities:

$$P_i = \Pr(\xi_j \leq V_i - V_j + \xi_i \ ; \ \forall j \varepsilon S)$$

$$= \int_t \Pr\{[\xi_i \varepsilon (t, t+dt)] \bigcap_{\forall j \varepsilon S} [\xi_j \leq (t+V_i-V_j)]\} dt \qquad [2.3a]$$

Letting $F(\ldots,t_i,\ldots)$ denote the joint cummulative distribution function of the disturbance vector ξ, and $F_i(\ldots,t_j,\ldots)$ its partial derivative with respect to t_i (assumed to exist), one can rewrite Eq. [2.3a]:

$$P_i = \int_{-\infty}^{\infty} F_i(\ldots,t+V_i-V_j,\ldots)dt. \qquad [2.3b]$$

In order to solve Eq. [2.3b] one has to assume a probability law for the disturbance vector, $\xi = (\ldots,\xi_i,\ldots)$. If the ξ_i's are assumed to be independent and identically distributed (iid) Gumbel variates, Eq. [2.3b] reduces to the well known multinomial logit (MNL) formula [see for example Beilner and Jacobs (1972) and McFadden (1973)].

The major drawback of the MNL model is that it exhibits the irrelevance from Independent Alternatives (IIA) property (Luce, 1959) that have been shown to produce unacceptably counter-intuitive results when applied to certain choice situations.[16]

---

[16] Mayberry (1970) carried the consequences of the IIA property to an extreme in the context of mode choice with the well known "Blue Bus-Red Bus" (contrived) example, Daganzo and Sheffi (1977) showed its undesirable consequences for route choice situations (see also Footnote 9), and Sheffi (1978a) discussed the failure of the MNL for the case of integer ordered alternatives.

Multinomial Probit (MNP) models provide a more attractive alternative to the MNL model, as they are based on the hypothesis that the random vector ξ is multivariate normal (MVN) distributed . Since the MVN distribution admits a full parametrization in terms of covariance matrix, correlation among alternatives can be captured (thus obviating the IIA property). Furthermore, some very powerful results with regard to the aggregation problem can be developed, because the MVN family is closed under linear transformation. Substituting the MVN probability law in Eq. [2.3] the choice probability becomes:

$$P_i = \int_{t_1=-\infty}^{t_1+V_i-V_1} \cdots \int_{t_{i-1}=-\infty}^{t_{i-1}+V_i-V_{i-1}} \int_{t_i=-\infty}^{\infty} \int_{t_{i+1}=-\infty}^{t_{i+1}-V_i+V_{i+1}} \cdots \int_{t_I=-\infty}^{t_I+V_i-V_I} MVN(t)\,dt_1 \ldots dt_I \qquad [2.4]$$

where:

$$MVN(t) = (\pi/2)^{I/2} \frac{1}{\sqrt{|\Sigma|}} \exp(-1/2\ t^T \Sigma^{-1} t) ,$$

and I is the number of alternatives in the choice set S, Σ is the covariance matrix of ξ and, as mentioned above, $E[\xi] = 0$.

Although Eq. [2.4] is not easy to evaluate, an approximate solution method was recently proposed and tested by Daganzo, Bouthelier and Sheffi (1977a and 1977b). This solution method is reviewed in the next subsection below.

2.1.2 The Approximation of the Probit Integral

The solution of the integral [2.4] is approximated based on some formulae suggested by Clark (1961) to evaluate the MVN distribution function. This section reviews these formulae and their use to calculate choice

probabilities.

Clark's formulae approximate the first four moments of the distribution of $\max(U_1,\ldots,U_I)$, where the I random variables have an unrestricted joint normal distribution.

Let U_1, U_2 and U_3 be MVN distributed with means V_1, V_2 and V_3, variances σ_1^2, σ_2^2 and σ_3^2, and correlation coefficients ρ_{12}, ρ_{13} and ρ_{23}. Then if ν_i is the ith moment about zero of the random variable, $\max(U_1,U_2)$, and $\rho[U_3, \max(U_1,U_2)]$ is the coefficient of linear correlation between the new variable and U_3, Clark showed that:

$$\nu_1 = V_1\Phi(\gamma) + V_2\Phi(-\gamma) + a\phi(\gamma) \qquad [2.5a]$$

$$\nu_2 = (V_1^2 + \sigma_1^2)\Phi(\gamma) + (V_2^2 + \sigma_2^2)\Phi(-\gamma) + (V_1 + V_2)a\phi(\gamma) \qquad [2.5b]$$

and

$$\rho[U_3,\max(U_1,U_2)] = [\sigma_1\rho_{13}\Phi(\gamma) + \sigma_2\rho_{23}\Phi(-\gamma)]/(\nu_2 - \nu_1^2)^{1/2} \qquad [2.5c]$$

where:

$\phi(x) = (2\pi)^{-1/2}\exp(-X^2/2)$ (the standard normal distribution)

$\Phi(x) = \int_{-\infty}^{x} \phi(t)dt$ (the standard cumulative normal curve)

$a^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}$ (the variance of the difference V_1-V_2)

and

$$\gamma = (V_1 - V_2)/a.$$

If now one approximates the distribution of the maximum of two normal random variables by a normal distribution, one has,

$$\max(U_1,U_2) \sim N(\nu_1, \nu_2 - \nu_1^2), \qquad [2.6]$$

and Equations [2.5 through 2.6] can be used recursively to obtain the

approximate distribution of the maximum of I variates by calculating the mean vector and covariance matrix of $[U_1,\ldots,U_{I-2},\max(U_{I-1},U_I)]$ and repeating the process to calculate them for $[U_1,\ldots,\max(U_{I-2}, \max(U_{I-1},U_I))]$, $[U_1,\ldots,\max(U_{I-3},\max(U_{I-2},\max(U_{I-2},\max(U_{I-1},U_I)))]$, etc. Thus, after I-1 iterations, one can obtain the approximate mean and variance, V_{max} and σ^2_{max}, of the maximum.

If U_i is the last variable to be considered, we have at the last iteration:

$$V_{-i} = E[\max(U_1,\ldots,U_{i-1},U_{i+1},\ldots,U_I)] \qquad [2.7a]$$

$$\sigma^2_{-i} = \mathrm{var}[\max(U_1,\ldots,U_{i-1},U_{i+1},\ldots,U_I)] \qquad [2.7b]$$

$$\rho_{-i,i} = \mathrm{corr}[U_i,\max(U_i,\ldots,U_{i-1},U_{i+1},\ldots,U_1)] \qquad [2.7c]$$

It is now possible to calculate the probability that the ith variate is actually the largest (i.e., the probability that the ith alternative is chosen), P_i:

$$P_i = \Pr\{U_i \geq \max_{\forall j\neq i} [U_j]\} = \Pr\{[\max(U_1,\ldots,U_{i-1},U_{i+1},\ldots,U_I)] - U_i \leq 0\}$$

$$= \Phi\left(\frac{V_i - V_{-i}}{\sigma_1^2+\sigma_{-i}^2-2\sigma_i\sigma_{-i}\rho_{-i,i}}\right) \quad . \qquad [2.8]$$

Due to the error introduced by Eq. [2.6], Eq. [2.8] is only approximate. However, as shown by Daganzo, Bouthelier and Sheffi (1976b), the latter is a good approximation for forecasting purposes for a wide range of conditions.

Further results with regard to the approximation method, its accuracy,

its use in MNP models' estimation and related statistical issues are explored in detail by Bouthelier (1978).

The next section reviews the use of MNP models for various aggregation purposes.

2.2 AGGREGATION WITH MULTINOMIAL PROBIT

2.2.1 Aggregation Over Individuals and Market Segmentation

The usefulness of travel demand models for planning and policy analysis lies in the model's ability to generate an aggregate demand function for each alternative, i.e., the share of the population choosing each alternative as a function of the alternative's attributes. The MNP model enables the user to perform such aggregation analytically, as shown below.

In accordance with the notation introduced in Subsection 2.1 denote the I-vector of observed utilities, V, ($V=\beta Z$, where β is a I x J matrix of calibrated coefficient and Z is the J-vector of explanatory variables)[17], and the calibrated (disaggregate) choice model $P_i(v)$.

The predicted population share of alternative i, R_i, is the expectation of the choice probability over the joint probability density function (p.d.f.) in the population, of the vector of measured utilities or:

$$R_i = E_v[P_i(v)] = \int_V P_i(v) \cdot f_V(v) dv \qquad [2.9]$$

where $f_V(v) = f_{V_1,\ldots,V_I}(v_1,\ldots,v_I)$ is the joint p.d.f. of V across the

[17] These notations deviate from the more common ones, in the travel demand literature, where β is denoted as a vector and Z as a matrix. Though the product βZ is identical in both cases, this form lends itself to a clearer explanation of the aggregation method reviewed below.

population.

The above multiple integral is, in general, difficult to compute and its numerical evaluation is prohibitively expansive in most cases. Koppelman (1976) reviews most of these numerical methods.

However, as discussed by Bouthelier (1978), R_i equals the probability that an individual sampled at random from the population, selects alternative i, and the utility vector for such individual, U*, can be constructed as a multivariate convolution of the probability density functions of V and ξ. For some p.d.f.'s such as the MVN, the distribution of the convolution is known and a simple choice model can be applied to groups of individuals.

Let $G_{U*}(u)$ denote the cummulative distribution of U*, and G^i_{U*} its partial derivative with respect to u^*_i. In a fashion parallel to the derivation of [2.1] through [2.4], one gets:

$$R_i = \Pr\ (U^*_i \geq U^*_j \quad ; \quad \forall j)$$

$$= \int_{t=-\infty}^{\infty} G^i_{U*}\ (\ldots, t+\bar{V}_i - \bar{V}_j, \ldots)dt \qquad [2.10]$$

where $\bar{V}_j = E(U^*_j)$, and Eq. [2.10], the aggregate share, exactly parallels Eq. [2.3b], the disaggregate choice probability.

When the MNP model is used [$\xi \sim MVN(0, \Sigma_\xi)$] and if V is approximately normally distributed [$V \sim MVN(\bar{V}, \xi_V)$], the convolution p.d.f. becomes:

$$g_{U*}(t) = \Pr\ \{t \leq (V + \xi) \leq t + dt\} = MVN(\bar{V},\ \Sigma_V + \Sigma_\xi) \qquad [2.11]$$

where $g_{U*}(t)$ is the p.d.f. of U* at t $[g_{U*}(t)=\partial G_{U*}(t)/\partial t]$ and assuming that V and ξ are independent random vectors.[18] The p.d.f. of the aggregate utilities can be linearly transformed, assuming that the explanatory variables are MVN distributed $[Z \sim MVN(\bar{Z},\Sigma_z)]$, using the definitional relationship $V = \beta Z$ to express the joint distribution of the aggregate utilities as:

$$g_{U*}(u^*) = MVN\ (\beta\bar{Z},\ \Sigma_\xi + \beta\Sigma_z\ \beta^T) \qquad [2.12]$$

and the Clark method can be applied to Eq. [2.10] to get the aggregate share.

Bouthelier and Daganzo (1978) (following McFadden and Reid's (1975) and Westin's (1974) works with regard to binary choices) extend this solution, freeing it from the assumption that all attributes Z have to be normally distributed (a hardly acceptable assumption especially for binary and other discrete members of Z) by simply conditioning on the values of the non-normal variables. Partitioning Z into its normally distributed members Z' and non-normally distributed members Z", $Z^T = (Z'^T,\ Z''^T)$, the conditional MVN distributions of Z, given that Z"=z" becomes:

$$(Z|Z'' = z'') \sim MVN[\bar{Z}(z''),\ \Sigma_z(z'')],$$

or

$$(Z|Z'' = z'') \sim MVN \left\{ \begin{bmatrix} \bar{Z}'(z'') \\ z'' \end{bmatrix} ,\ \begin{bmatrix} \Sigma_{z'}(z'') & \dots & 0 \\ 0 & \dots & 0 \end{bmatrix} \right\}$$

where the conditional notations are self-explanatory. Then, Eq. [2.12] becomes:

---

[18] The assumption of the independence between V and ξ implies that the aggregation procedure described here does not apply to a model including taste variations, where ξ is specifically assumed to be a function of V.

$$g_{U^*|Z''}(u^*|z'') = MVN[\beta\bar{Z}(z''), \ \Sigma_\xi + \beta\Sigma_z(z'')\beta^T] \qquad]2.14]$$

and

$$R_i = \sum_{k=1}^{K} \frac{q_k}{q} R_i(z''_k) \qquad [2.15]$$

where $R_i(z''_k)$ is the probability of choice derived from [2.14], z''_k is the value of the vector Z" in k^{th} combination, q is the population size, q_k is the group size to which the kth combination applies, and there are K such combinations.

Typically, only a small number of the attributes' p.d.f.'s would not be approximately normal and K, the number of market segments, would be relatively small. Furthermore, the segmentation criterion is clear since it consists in classification with regard to discrete variables.

Another application of normal approximation and MNP models is discussed in the next section which is concerned with spatial aggregation.

2.2.2 Spatial Aggregation through Normal Approximation

The spatial aggregation errors in traffic assignment studies arise from the representation of the population of each zone in the study area as a point - the "centroid". Generally, it is reasonable to assume that the finer the division of the study area into zones, the more accurate the representation is. However, the computational burden of the network analysis increases dramatically with additional nodes and centroids.

In this Subsection, this problem is tackled through a representation of the intra-zonal travel impedances by a continuous approximation, following the work of Daganzo (1977c). The approach consists in assuming that the population is uniformly distributed over a zone and that the intrazonal impedance distribution over the population approximately follows a MVN

probability law. This approximation is used in Subsection 2.3.3 below and in the equilibration algorithm of Chapter 3.

The issue is presented through an example of one zone. Consider the a x b rectangular zone shown in Figure 2-1, including three access nodes consecutively number 1 to 3. Assume further that the intrazonal street network consists of a uniform grid parallel to the zone sides.

The distance traveled from a random point in the zone, 0, to access node i, D_i, is the sum of two uniformly distributed and independent random variables, A_i, [between o and a] and B_i, [between o and b] (see Figure 2-1b). Thus, the expected distance from a random point to any of the access nodes, $E[D_i]$, is:

$$E[D_i] = E[A_i + B_i] = \frac{a + b}{2} \; ; \quad i = 1,2,3. \qquad [2.16]$$

The variance of D_i is given by:

$$\sigma_1^2 = \mathrm{var}[A_i + B_i] = \frac{a^2 + b^2}{2} \; ; \quad i = 1,2,3. \qquad [2.17]$$

The covariance between the distances to any two access nodes i and j is:

$$\sigma_{ij} = \mathrm{cov}[(A_i + B_i),(A_j + B_j)] = \mathrm{cov}[B_i,B_j] + \mathrm{cov}[A_i,A_j] \qquad [2.18]$$

since A_i and B_j are independent.

Calculating, for example, the first term of the above equation, one gets:

$$\mathrm{cov}[B_i,B_j] = E(B_iB_j] - E[B_i]\,E[B_j]$$

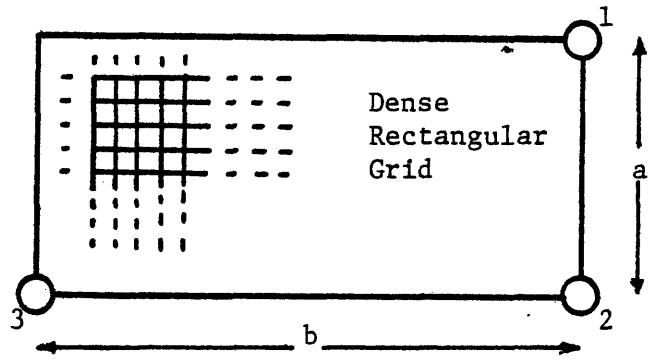
$$= \left(\int_{y_i=0}^{b} y_i(b - y_i) \cdot \frac{1}{b}\, dy_i\right) - \frac{b}{2}\,\frac{b}{2} = -\frac{b^2}{12} \; . \qquad [2.19]$$

Calculating the rest of the moments of D in a similar fashion, and applying a normal approximation to the joint p.d.f. of D, one gets:

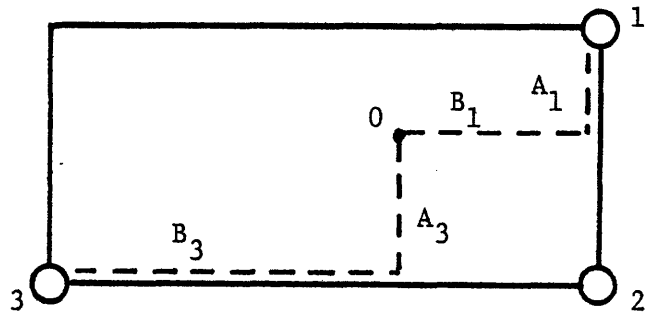
Figure 2-1

Determination of the Moments of Intrazonal Travel Distance

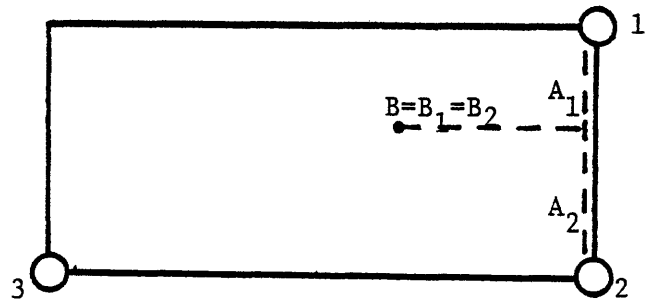
2-1a Zone Dimensions and Representation



2-1b Determination of Moments for Access Links 1 and 3



2-1c Determination of Moments for Access Links 1 and 2



$$D = \begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} \sim \text{MVN} \left\{ \begin{bmatrix} \frac{a+b}{2} \\ \frac{a+b}{2} \\ \frac{a+b}{2} \end{bmatrix} \begin{bmatrix} \frac{a^2+b^2}{12} & \frac{a^2-b^2}{12} & \frac{-a^2-b^2}{12} \\ \frac{a^2-b^2}{12} & \frac{a^2+b^2}{12} & \frac{b^2-a^2}{12} \\ \frac{-a^2-b^2}{12} & \frac{b^2-a^2}{12} & \frac{a^2+b^2}{12} \end{bmatrix} \right\} . \qquad [2.20]$$

Daganzo (1977c) develops a graphical technique for computing the moments of the distance from a random point to the access nodes for zones of any shape. He assumes that intrazonal travel can take place in any direction and uses the uniform distribution assumption to derive the moments with a graphical integration technique.

The moments of the intrazonal distance p.d.f. can be also obtained through an off-line Monte-Carlo simulation. Using the simulation technique one does not have to assume that travel can take place in all directions at the same speed, or over a grid. Also the uniform population distribution assumption can be relaxed and any p.d.f. might be used to represent the distribution of origins or destinations within a zone.

To obtain the moments with a simulation approach, the intrazonal network is represented and the population is sampled. Once a realization is drawn, the minimum path tree from the random point (the realization) to the access nodes is computed and the impedance to all access nodes recorded. After a large enough sample has been drawn, the moments of the impedance vector are estimated using the sample moments.

Since in most instances the geometry of the local street network only has a second order effect, a simple pythagorean expression can replace the minimum-path-tree calculation in each drawing.

Bouthelier and Daganzo (1978) discussed the accuracy of the normal approximation for the case of uniform population distribution. In the context of this thesis, it is important to note that the intrazonal trip-end impedance distribution can be reasonably approximated by a MVN normal distribution for other population distribution patterns too [see Daganzo (1977c)].

2.2.3 Aggregation of Alternatives

This subsection deals with the expected utility of an individual chosen in random from the population. This expected utility is used in Chapter 3 as an integral part of the equilibration procedure (where it is multiplied by the population size to generate a measure of "total utility").

To find the mean utility, note that (as mentioned in Subsection 2.2.1) the aggregate share of an alternative i equals the probability that an individual sampled at random from the population selects alternative i, based on the distribution of the aggregate utilities vector $U^* = (\ldots, U_j^*, \ldots)$. In accordance with the definition of random utility models, it is clear that the utility of a group of alternatives (facing a randomly chosen individual) is the utility of the chosen alternative, i.e., the maximum-utility alternative. Denoting the utility of the chosen alternative by $\tilde{U}^*$ we have:

$$\tilde{U}^* = \max_{j \varepsilon S} \{U_j^*\} \qquad [2.21]$$

where S is the aggregate choice set (including I alternatives).

Since U^* is a random vector, $\tilde{U}^*$ is a random variable and to find its mean, one has to use the expectation operator. Thus denoting the popula-

tion size by q:

$$\text{Total Utility} = q \cdot E[\tilde{U}^*] = q \cdot E[\max_{j \varepsilon S} \{U_j^*\}]. \qquad [2.22]$$

The computation of $E[\tilde{U}^*]$ does not require any additional effort when the Clark method (see Subsection 2.1.2) is applied to evaluate the alternatives' aggregate shares. As evident from the description of the method, the expected maximum utility, $E[\tilde{U}^*]$, can be computed as an integral part of the technique.

The expected maximum utility has been used in similar contexts of several researchers. Sheffi (1977a) and Sheffi and Daganzo (1978b) utilized it in the context of stochastic network assignment models, which is somewhat parallel to the use of $E[\tilde{U}^*]$ in this thesis -- over the hypernetwork. Other researchers, including Harris and Tanner (1974), Williams (1977) and Ben-Akiva and Lerman (1977) have used it at the disaggregate demand model level as a measure of accessibility.

As an aside, and to facilitate some intuitive notion of the expected maximum utility, note that it has two properties:

a) It is monotonic with respect to the size of the choice set:

$$E[\max\{\tilde{U}^*, U_{I+1}^*\}] \geq E[\tilde{U}^*] \qquad [2.23]$$

b) Its marginal with respect to the mean utility of an alternative equals the choice probability (aggregate share) of this alternative:

$$\frac{\partial}{\partial \bar{V}_i^*} E[\tilde{U}^*] = R_i \qquad [2.24]$$

where (see Subsection 2.2.1) $\bar{V}_i = E[U_i^*]$ and R_i is the aggregate share of alternative i (see Eq. [2.10]).

Both of the abovementioned properties of the expected maximum utility mean that the total utility (as defined by Eq. [2.22]) is a reasonable measure of the aggregate attractiveness of a given system to all users. The first property holds for non-interacting alternatives (i.e., when the introduction of a new alternative does not decrease the measured utility, V, of any existing alternative)[19]. It means that as the choice set is expanded, the expected maximum utility increases. The second property holds provided that the distribution of U* is translationary invariant [Williams (1977)][20]. For the MNP model, this restriction means that the variance of

---

[19] To see that [2.24] holds, one can write the random variables explicitly, i.e., for every individual in the population:

$$\max\{U_1^*,\ldots,U_I^*,\ U_{I+1}^*\} \geq \max\{U_1^*,\ldots,U_I^*\},$$

and the expectation operator will obviously yield Eq. [2.23].

[20] To prove [2.24], write explicitly the expression for $E[\tilde{U}^*] = E[\max_{j\varepsilon S}\{U_j^*\}]$:

$$E[\tilde{U}^*] = \int\cdots\int \max_{j\varepsilon S}(U_j^*)\, g_{U^*}(u_1,\ldots,u_I)\,du_1\cdots du_I$$

where $g_{U^*}(\cdot)$ is the p.d.f. of U* (see Eq. [2.11]). Interchanging integration and differentiation (utilizing the independence assumption) the marginal expected maximum utility becomes:

$$\frac{\partial E[\tilde{U}^*]}{\partial \bar{V}_i^*} = \int\cdots\int \{\frac{\partial}{\partial \bar{V}_i^*}\ [\max_{j\varepsilon S}(U_j^*)]\, g_{U^*}(u_1,\ldots,u_I)\,du_1\cdots du_I.$$

Now, define the set of U*'s, S_i such that $U_i^* \geq (U_j^*;\ \forall j\varepsilon S)$, and denote the indicator function of S_i by $ID_i(u_1,\ldots,u_I)$. That is $ID_i(u_1,\ldots,u_I) = 1$ if $(u_1,\ldots,u_I)\varepsilon S_i$ and zero otherwise. Since:

$$\frac{\partial}{\partial \bar{V}_i}\ \max_{j\varepsilon S}(U_j^*) = ID_i(u_1,\ldots,u_I),$$

the marginal expected utility can be expressed as:

$$\frac{\partial E[U^*]}{\partial \bar{V}_i} = \int\cdots\int ID_i(u_1,\ldots,u_I)\, g_{U^*}(u_1,\ldots,u_I)\,du_1\cdots du_I$$

$$= P_r[(u_1,\ldots,u_I)\varepsilon S_i] = P_r[U_i^* \geq (U_j^*;\ \forall j\varepsilon S)] = R_i$$

where the last equality follows from the definition of R_i (see Eq. [2.10]).

U^* is not a function of its mean vector. (A similar assumption was utilized already in Subsection 2.2.1 -- see Footnote 18.) Property (b) means that as the measured utility of an alternative is improving, the expected utility in the system increases (since R_i is the share, which is non-negative by definition).

Though quite straightforward, the above mentioned properties do not hold for other measures that were intended to approximate the expected utility [e.g., the weighted sum used by McLynn (1976) and others) -- see Ben-Akiva (1977)].

This concludes the review of the choice theory background and the various aggregation issues that are part of the equilibrium analysis. The next section is devoted to some background on network equilibrium analysis.

2.3 EQUILIBRIUM OVER NETWORKS

2.3.1 On Stochastic Models of Traffic Assignment

A theory of traffic assignment over network exhibiting stochastic links costs has been developed by Daganzo and Sheffi (1977). This Subsection includes a brief overview of this theory, as it provides a link between MNP choice models and network (and hypernetwork) representation and topology.

Equilibrium traffic assignment models attempt to achieve the following user-equilibrium (U-E), the (differently phrased) definition of which, was given by Wardrop (1951):

"At equilibrium no user can improve his travel time[21] by unilaterally changing routes."

Recognizing the somewhat unrealistic behavioral assumptions underlying this definition Daganzo and Sheffi (1977) expanded upon this principle, distinguishing between the measured and perceived travel time and suggested the following definition of stochastic user equilibrium (S-U-E):

"At equilibrium (S-U-E) no user <u>believes</u> he can improve his travel time by unilaterally changing routes."

In their work, the travel time on route k between origin r and destination s, as perceived by a randomly selected motorist, is modeled as a random variable -- t_k^{rs}. They also postulated that $E[t_k^{rs}] = T_k^{rs}$, where T_k^{rs} denotes the measured travel time. Using these notations, the above S-U-E definitions can be formalized, based on the weak law of large numbers, as follows:

$$\Pr\{t_k^{rs} \leq t_h^{rs} \;,\; \forall h \neq k \mid T\} = \frac{x_k^{rs}}{\sum_h x_h^{rs}} \;;\quad \forall r,s,k \qquad [2.25]$$

where x_k^{rs} is the flow on route k between origin r and destination s, and $T = (\ldots, T_k^{rs}, \ldots)$.

For a user sampled at random from the population, the LHS of Eq. [2.25] is the probability the route k is chosen (it parallels Eq. [2.2] which applies to any choice) and since it is a function of the measured travel times, T^{rs}, which in turn depends on the flow pattern $x = (\ldots, x_k^{rs}, \ldots)$,

[21] For the discussion of network assignment (throughout Sections 2.3 and 3.1), the terms travel time and travel cost are used interchangeably, meaning negative travel utility (disutility), without loss of generality.

Eq. [2.25] is an equilibrium equation which merely states the S-U-E principle. This equation can be shown to be a generalization of the abovementioned U-E principle. This result is extended in Chapter 3 to define the equilibrium over the whole transportation market.

To evaluate the LHS of Eq. [2.25] one has to assume a probability law for $t = (\ldots, t_k^{rs}, \ldots)$[22]. Postulating that non-overlapping sections of road are perceived independently by tripmakers and that sections of equal length are perceived in identical fashion, Daganzo and Sheffi have shown that the perceived <u>links</u> travel times, t'_{ij}, are approximately normally distributed:

$$t'_{ij} \sim N(T'_{ij}, \theta T'_{ij}) \qquad [2.26]$$

where θ is the variance of t_{ij} on a road section of unit length. Defining the network incidence matrix, as a matrix with entries, $\delta_{ij,k}$, given by:

$$\delta_{ij,k} = \begin{cases} 1 \text{ if link ij belongs to route k} \\ 0 \text{ otherwise,} \end{cases}$$

one can write:

$$t_k = \sum_{\forall ij} t'_{ij} \cdot \delta_{ij,k} \quad ;$$

or using vector notation:

$$t = t' \cdot \Delta. \qquad [2.27]$$

---

[22] For the rest of the discussion in this Subsection, we deal with a single arbitrary O-D pair as we concentrate on the probability of choice <u>given</u> the measured travel times, and the superscripts denoting O-D pair are dropped. This problem was termed S-N-L (stochastic network loading) by Daganzo and Sheffi (1977).

Since t' is a vector of mutually independent normal random variables, and [2.27] is a linear transformation, the vector of route travel times, t, is a MVN random variate.

By definition, E[t] = T and the covariance matrix of t is easily obtained. Letting t_{kp} and T_{kp} denote respectively the perceived and measured travel times onthe road sections shared by the routes k and p, it is easy to see that:

$$\mathrm{var}(t_k) = \theta T_k \qquad [2.28a]$$

and

$$\mathrm{cov}(t_k, t_p) = \mathrm{var}(t_{kp}) = \theta T_{kp} \ . \qquad [2.28b]$$

The important conclusion from Eqs. [2.28] is that the distribution of t (and therefore the flow pattern) is not only determined by the measured travel times, T (like any traffic assignment procedure) and the accuracy of people's perception of time, θ, (like any stochastic assignment method), but by the <u>topology of the network</u> as well.

Since Eqs. [2.25] and [2.28] define a probit (MNP) model with utilities equal to <u>minus</u> travel time, the route choice probabilities can be obtained (for small networks) with the results of Subsection 2.1.2.

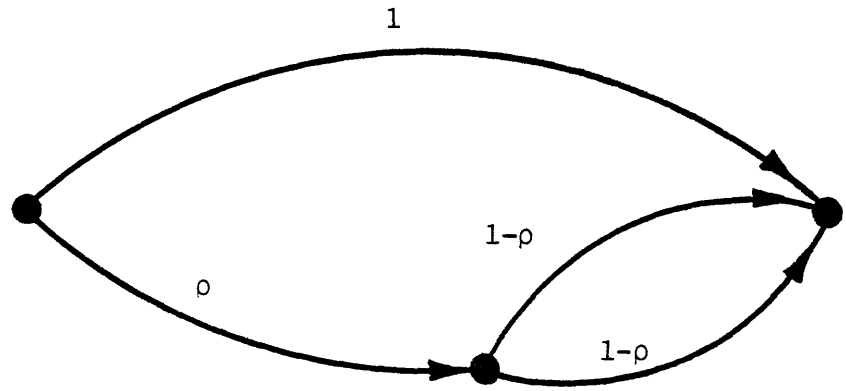
In the context of this thesis, it is important to note that logit based models of route choice produce unreasonable results because the logit formula cannot capture well the topology of the network. This point is explained below.

Consider the simple network shown in Fig. 2-2. The argument used by

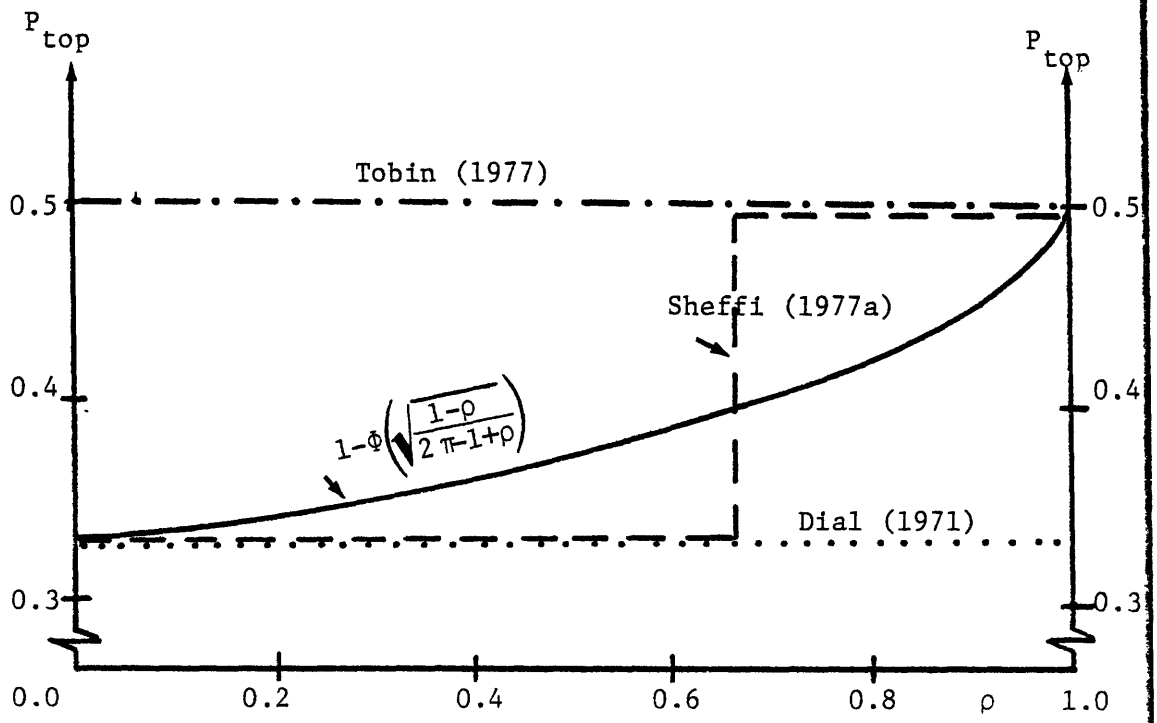
Figure 2-2

Comparison of Stochastic Network Loading Methods

2-2a Network Example (Numbers on links represent impedance)



2.2b Fraction of Flow on Top Route of Fig. 2.2a, P_{top}, vs. the Overlap Between the Bottom Routes, ρ



most authors[23] when discussing Dial's logit-based method is that the fraction of flow using the top route, P_{top}, should behave as (see Figure 2-2a):

$$P_{top} \to 1/2 \text{ when } \rho \to 1$$

and $$P_{top} \to 1/3 \text{ when } \rho \to 0.$$

None of the logit-based models produces a systematic dependency of P_{top} on ρ in this example. Dial's (1971) model predicts $P_{top} = 1/3$, regardless of ρ; Tobin's (1977) "Arrival Likelihood" model predicts $P_{top} = 1/2$, regardless of ρ [this last model is similar to an earlier one suggested by Gunnarson (1972)]; and the model used by Sheffi (1977b)[24] would produce $P_{top} = 1/2$ or $P_{top} = 1/3$, depending on an ad-hoc modeler's definition of the network.

The three abovementioned approaches are demonstrated in Figure 2-2b where the MNP approach is also plotted; and as can be seen from the figure, this latter curve coincides with what most authors dealing with stochastic assignment (see Footnote 23) described as a desired result.

Thus, MNP is theoretically more attractive than logit. However, while logit models can be put into analytical assignment algorithms, the MNP approach cannot, as yet, since its straightfoward application requires path-enumeration which is prohibitively expensive for large networks. A practical approach to MNP assignment has been suggested by Daganzo and

---

[23] This deficiency of logit-based models has been pointed out by many authors including Schneider (1973), Burrell (1976) and Floridan and Fox (1976), with regard to Dial's model, and by Sheffi (1978b), with regard to Tobin's (1977) model.

[24] See also Moavenzadeh, Sheffi and Brademayer (1977).

Sheffi, based on a simulation of perceived link travel times[25]. Although such a technique can be applied to networks exhibiting flow dependent travel times and stochastic effects, its rate of convergence is slow in congested networks. Thus, rejecting all the abovementioned methods, the approach taken in this thesis is to approximate the S-U-E flow pattern over congested networks with a U-E (deterministic) flow pattern and yet recognize the important stochastic aspects of the problem.

An intuitive justiciation of this simplification is given in the next subsection.

2.3.2 Approximating S-U-E by U-E for Congested Networks

This subsection explains the rationale for using the Wardropian user equilibrium (U-E) flow pattern to approximate the stochastic user equilibrium (S-U-E) flow pattern presented in the last subsection. The argument is based on the system's behavior near capacity.

Using the notation of the latter subsection (since we are concerned here with the street network), the S-U-E equilibrium equations are:

$$x_k^{rs} = q^{rs} P_k^{rs} (T^{rs}, \Sigma^{rs}) \quad ; \quad \forall r,s,k \qquad [2.29]$$

and

$$T_k^{rs} = \sum_{ij} \delta_{ij,k}^{rs} \cdot T'_{ij} \left(\sum_{r,s} \sum_{m} x_m^{rs} \delta_{ij,m}^{rs} \right) \quad ; \quad \forall r,s,k \qquad [2.30a]$$

$$(\Sigma^{rs})_{pq} = \theta T_{pq}^{rs} \quad ; \quad \forall r,s,p,q \qquad [2.30b]$$

[25] This algorithm is a modification of the simulation algorithm suggested by Burrell (1968), Von Falkenhausen (1966) and Wildermuth (1972).

where q^{rs} is the trip interchange rate (i.e., $q^{rs} = \sum_h x_h^{rs}$), P_k^{rs} is the MNP route choice probability (i.e., $P_k^{rs}(T^{rs}, \Sigma^{rs}) = Pr\{t_k^{rs} < t_h^{rs}; \forall h \neq k\}$ and $t^{rs} \sim MVN[T^{rs}, \Sigma^{rs}]$), $\delta_{ij,m}^{rs} = 1$ if link ij belongs to the mth route from r to s, $\delta_{ij,m}^{rs} = 0$ otherwise and $T'_{ij}(\cdot)$ is the volume-delay curve for link ij.

If $T'_k(\cdot) = T'_k$, Eqs. [2.29]-[2.30] reduce to stochastic assignment over uncongested network, and Subsection 2.3.1 reviewed the solution of this problem. It was also mentioned in the abovementioned subsection that if $\theta \to 0$ the problem reduces to the well-researched U-E equilibrium and available algorithms [e.g., Nguyen (1974), LeBlanc (1975)] can be used to solve for such an equilibrium. Following Daganzo (1977d) this subsection demonstrates that the U-E condition is a reasonable approximation for S-U-E as the network links approach capacity, even if $\theta \neq 0$.

It is well known from both queueing theory and experience that realistic volume-delay curves should rapidly increase as the link approaches capacity [e.g., see Daganzo (1977a)]. In other words, acceptable curves must satisfy:

$$\lim_{x'_{ij} \to C_{ij}} \frac{d}{dx'_{ij}} T'_{ij}(x'_{ij}) = \infty \qquad [2.31]$$

where C_{ij} denotes the link capacity, and $T'_{ij}(\cdot)$ is non-negative, increasing and strictly convex.

Assume that a network is uniformly congested and in user equilibrium. Then, at the limit (heavy congestion), a small change in the flow pattern, results in such large travel time changes that the relative merits of alternative routes become obvious and perceived "correctly" (in accordance with the measured travel times) by the users. Under these conditions,

the U-E and S-U-E flow patterns must coincide.

In other words, if $x' = (\ldots, x'_{ij}, \ldots)$ is a U-E flow pattern such that $x'_{ij} \to C_{ij}$ and $\frac{d}{dx'_{ij}} T'_{ij}(x)\Big|_{x=x'_{ij}} \to \infty$, x' is also a S-U-E flow pattern.

This holds because if any given user (one unit of flow) between r and s were to change from route k to route k*, the travel time on route k* would increase by:

$$\sum_{ij} \delta^{rs}_{ij,k^*}(1 - \delta^{rs}_{ij,k}) \frac{dT'_{ij}(x)}{dx}\Bigg|_{x=x'_{ij}} \to \infty \qquad [2.32a]$$

and the travel time on route k would change by:

$$\sum_{ij} \delta^{rs}_{ij,k} (\delta^{rs}_{ij,k^*} - 1) \frac{dT'_{ij}(x)}{dx}\Bigg|_{x=x'_{ij}} \to -\infty \qquad [2.32b]$$

Obviously, for partially congested networks, the approximation would be better with increased congestion, and as the perception variance decreases (i.e., as $\theta \to 0$ -- see Eq. [2.30b]).

This conclusion is in agreement with the validation experiment carried out by Florian and Nguyen (1976). They tested (deterministic) U-E methods on an urban area (the city of Winnipeg) and reported:

> "... The results... show clearly that, the higher the predicted volume, the better the fit between predicted and observed volume."

Since users imperfect perception of link travel times was not modeled in this study, their deterministic model predicted better, over congested parts of the network, where the S-U-E is well approximated by deterministic U-E, as argued in this subsection.

The other argument for using deterministic methods to assign trips

to the street network is that no better method is available. As mentioned in the preceeding subsection, path enemeration of the street network is computationally infeasible and simulation methods are inefficient. This is the reason that deterministic methods are always applied in practice, for congested networks.

The next subsection discusses the last piece of background material needed for the development of the equilibrium method. It deals with an application of the convex programming formulation and solution of the (deterministic) traffic assignment problem to the spatial aggregation problem.

2.3.3 An Algorithm for the Spatial Aggregation Problem of Traffic Assignment

This subsection reviews a traffic assignment algorithm which was proposed by Leblanc et al. (1975) and modified by Daganzo (1977a) to count for finite link capacities as well. It further presents two algorithms [Daganzo (1977b and 1977c)] that can account for continuous distributions of population, thus solving the spatial aggregation problem of traffic assignment [26].

Consider a network (directed graph) consisting of a set of nodes J containing a set of centroids C and a set of nodes N (i.e., $\{C\} \cap \{N\} = \phi$ and $\{C\} \cup \{N\} = \{J\}$), and a set of links, L, joining nodes (i.e., $(ij) \varepsilon L$ if there is a link from i to j). Denoting the link flows (which are the objects of this analysis) by x_{ij}, the link volume-delay curves by $T_{ij}(\cdot)$,

[26] The Appendix to this thesis contains a more detailed discussion of the formulation of the traffic assignment problem as a minimization program and the convex combinations algorithm for its solution.

the link capacities by C_{ij}, and the given (fixed) centroid to centroid trip interchange by q_{rs}, the user equilibrium can be shown to be [see for example Jorgenson (1953), or Beckman et al., (1956)] the solution of the following mathematical program:

$$\min \sum_{ij} \int_0^{x_{ij}} T_{ij}(\omega)\,d\omega \qquad [2.33]$$

s.t.

$$\sum_{\forall j} x_{ij} - \sum_{\forall \ell} x_{\ell i} = D_i \qquad \forall i \varepsilon J$$

$$0 \leq x_{ij} \leq C_{ij} \qquad \forall (ij) \varepsilon L$$

where $D_i = \sum_{\forall r} q_{ri} - \sum_{\forall s} q_{is}$ if $i \varepsilon C$ and $D_i = 0$ otherwise.

LeBlanc et al. (1975) applied the Frank Wolfe (1956) convex combinations algorithm to solve the Program [2.33]. The algorithm steps (including Daganzo's (1977a) modification -- see Step 2) are the following:

Step 0. Initialization

Determine an initial link flow pattern $\{x_{ij}\}$ and the associated link travel times $\{T_{ij}\}$.

Step 1. Direction Finding

Perform an "All-or-Nothing" assignment using the current link costs, T_{ij}; label the resulting flow pattern $\{y_{ij}\}$.

Step 2. Step Size Determination

Find the value of α^* that minimizes:

$$\sum_{(i,j)\varepsilon L} \int_0^{x_{ij}+\alpha(y_{ij}-x_{ij})} T_{ij}(\omega)\,d\omega$$

s.t.

$$\alpha \leq \alpha_{max} = \min_{x_{ij}<y_{ij}} \{(C_{ij} - x_{ij})/(y_{ij} - x_{ij})\} \quad \text{and } 0 \leq \alpha \leq 1.$$

Step 3. Updating

$$x_{ij}^{next} \leftarrow x_{ij} + \alpha^*(y_{ij} - x_{ij}) \qquad \forall (ij) \varepsilon L$$

$$T_{ij}^{next} \leftarrow T_{ij}\ (x_{ij}^{next}) \qquad \forall (ij) \varepsilon L$$

Step 4. Stopping Test

If convergence has not been achieved, go to Step 1; otherwise, the current $\{x_{ij}\}$ is the equilibrium flow pattern.

The details of the stopping test are not important for the following discussion (some measure of similarity between x_{ij}^{next} and x_{ij} can be used) in this subsection.

It should be noted that Step 1 of the F-W algorithm is what limits the size of the problem to be solved, and Step 3, though complicated looking, uses up relatively small amounts of computer time.

The approach taken by Daganzo (1977b) is to introduce several centroids per zone instead of a single one, thereby reducing the spatial aggregation bias. A straightforward application of the above algorithm to such a representation would have been very expensive and thus the F-W algorithm had to be changed.

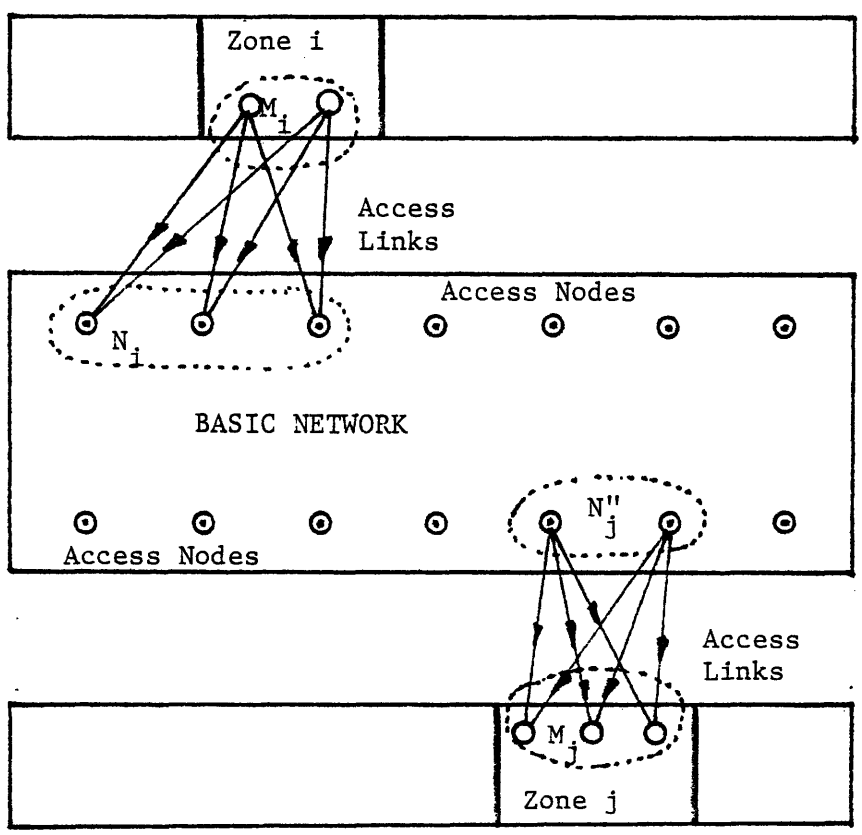
In order to discuss the streamlined algorithm, the network representation used should be explained in more detail. Denote the set of nodes that can be reached in one step (traveling on one link only) from centroid r, as N_r' (these are termed "outbound" access nodes) and the set of nodes from which centroid s can be reached in one step N_s'' ("inbound" access nodes). Obviously, $\{N_r' \cap N_s'\}$ needs not be empty, $\{N_r'\} \neq \{N_r''\}$ in general, and $\{N_r'\}$, $\{N_s''\} \subseteq \{N\}$. The links connecting between centroids and access nodes

are termed access links (in the traditional network representation these are the dummy centroid connectors) and are assumed to exhibit no congestion effects since they represent the somewhat ubiquitous intrazonal street network. The set of access links is denoted L' ($\{L'\} \subseteq \{L\}$), and the complementary subset, $\bar{L}'$, denotes the links of the street network (the network between access nodes) which is termed the basic network.

For the discussion of the multicentroid representation some more notations are needed. Let each zone r be divided into m_r subzones each associated with a single subcentroid. The set of subcentroids of zone r is denoted M_r. The trip interchanges, q'_{rs}, between subcentroids are arranged in a sub O-D matrix, Q'. Figure 2-3 illustrates a schematic

Figure 2-3

Network Representation for the Multicentroid Assignment Problem



representation of the various network components.

In order to perform the "All-or-Nothing" assignment required in Step 1 (and Step 0) of the F-W algorithm on a network such as the one depicted in Figure 2-3, Daganzo (1977b) utilized the optimality principle of dynamic programming to decompose the problem into two stages. In the first stage, the travel times over the shortest paths between all possible access nodes combinations are computed. The matrix of trip costs for every access node pair is referred to as the skim tree.

Stage Two consists in finding the best access nodes to connect any given pair of subcentroids. Since a shortest path is also shortest for any pair of intermediate nodes, all one needs to find the access nodes used by the shortest path between any given subcentroid pair is the skim tree (and the access links travel time). Once the access node pair used by a subcentroid pair has been found, one adds the corresponding trip interchange to the number of trips already assigned (from the other subcentroid pairs) to the access node pair under consideration. In the end one will have a trip interchange table, Q", between access nodes, which is termed the access table. Stage Two is completed when one assigns the entries of the access table, q''_{rs}, to the appropriate paths identified in Stage One (this is done either by using the paths found in Stage One or by recomputing the shortest paths between all access nodes) to get the flow pattern $\{y_{ij}\}$.

Thus, Step 1 of the algorithm would consist of the following:[27]

[27] Note that Stage Two is broken here into two steps (1b and 1c).

Step 1. <u>Direction Finding</u>

Step 1a. Obtain the minimum travel cost between each access node pair based on the current link costs $\{T_{ij}\}$.

Step 1b. For each subcentroid pair, find the access nodes that result in the least travel cost from subcentroid to subcentroid; allocate the trip interchanges q'_{rs} to such access nodes and obtain the total travel cost. Repeat the process for all subcentroid pairs to obtain the access table, q''_{rs}.

Step 1c. Load the access table, Q", onto the network by performing a new "All-or-Nothing" assignment between all access nodes. This yields a set of link flows $\{y_{ij}\}$.

The computational advantages of this decomposition are discussed in Subsection 3.1.2 of the following chapter.

The rest of the algorithmic steps are, of course, not affected by the decomposition of the Direction Finding Step. However, for reasons that will be apparent in the sequel, the line search (Step 2) is decomposed as follows:

$$\min_{0<\alpha<\bar{\alpha}} \Big[\sum_{(ij)\varepsilon L'} T_{ij}[x_{ij} + \alpha(y_{ij}-x_{ij})] + \sum_{(ij)\varepsilon \bar{L}'} \int_0^{x_{ij}+\alpha(y_{ij}-x_{ij})} T_{ij}(\omega)d\omega \qquad [2.34a]$$

where $\bar{\alpha} = \min(1,\alpha_{max})$, and $\alpha_{max} = \min_{x_{ij}\leq y_{ij}} \{(C_{ij} - x_{ij})/(y_{ij} - x_{ij})\}$, [2.34b]

since for $(ij)\varepsilon L'$(the set of access links), $\int_0^{x_{ij}} T_{ij}(\omega)d\omega = x_{ij}\cdot T_{ij}$.

Eliminating constants from the above minimization (see Appendix A), Eq. [2.34] can be written as:

$$\min_{0\leq \alpha \leq \bar{\alpha}} [\alpha(L_y-L_x) + \sum_{(ij)\varepsilon \bar{L}'} \int_0^{x_{ij}+\alpha(y_{ij}-x_{ij})} T_{ij}(\omega)d\omega] \qquad [2.35a]$$

where:

$$L_y = \sum_{(ij)\varepsilon L'} T_{ij}\cdot y_{ij} \quad \text{and} \quad L_x = \sum_{(ij)\varepsilon L'} T_{ij}\cdot x_{ij} \qquad [2.35b]$$

L_y, the <u>total access cost</u>, can be readily obtained in Stage Two of the streamlined "All-or-Nothing" method, by adding up the product of the sub O-D table, q'_{rs}, and the sum of the costs of the two access links used between each subcentroid pair rs as pair rs is considered in Step 1b. These values are accumulated as Stage Two proceeds.

This decomposition brings about one value to be updated in Step 3, that replaces the updating of all access links costs:

$$L_x^{next} = L_x + \alpha^*(L_y - L_x). \qquad [2.36]$$

Figure A-2 of the Appendix describes the multicentroid streamlined algoritm in detail.

Even though this algorithm ameliocrates the large increase in computational cost that results from an increase in the number of centroids, the latter cannot be drastically increased. Daganzo (1977c) overcomes this limitation by using a continuum approximation for the distribution of intrazonal trip ends, which is practically, equivalent to using an infinite number of subcentroids. This approach is reviewed below.

The approach is based on the observation that Stage Two consists in calculating the access table, $\{q''_{rs}\}$, and the total access costs, L_y, (see Step 1b above) and that by representing origin and destination densities by continuous functions, q''_{rs} and L_y can be obtained mathematically.

Let t_{0,r_i} denote the intrazonal travel time, for a person in zone r chosen at random, to access node $r_i, (r_i \varepsilon N'_r)$, and let $t_{s_j,D}$ denote the intrazonal "inbound" travel time from access node $s_j, (s_j \varepsilon N''_s)$ to a random destination in zone s. Using these notations note that the distribution

of the random vectors representing the intrazonal travel time to/from the access nodes can be approximated, using the methods of Subsection 2.2.2, by (see Eq. [2.20] for example):

$$t_0 \sim MVN(E[t_0], \Sigma_0) \qquad [2.37a]$$

and

$$t_D \sim MVN(E[t_D], \Sigma_D) \qquad [2.37b]$$

where $t_0 = (\ldots, t_{0,r_i}, \ldots)$ and $t_D = (\ldots, t_{s_j,D}, \ldots)$ †. The techniques discussed in Subsection 2.2.2 can be used to obtain the moments of the abovementioned distribution (i.e., $E(t_{0,r_i})$, $var(t_{0,r_i}, t_{D,r_i})$ and $cov(t_{0,r_i}, t_{0,r_j})$ for the origin zone, and $E(t_{s_i,D})$, $var(t_{s_i,D})$ and $cov(t_{s_i,D}, t_{s_j,D})$ for the destination zone).

Given the skim tree entries, T_{ij}, the probability, $P(r_i,s_j)$ and thus the volume using each access node pair r_i,s_j can be determined, using a MNP model, i.e.:

$$P(r_i,s_j) = P_r \{t^*_{r_i,s_j} \leq t^*_{r_h,s_\ell}; \ \forall r_h \varepsilon N'_r \text{ and } s_\ell \varepsilon N''_s\} \qquad [2.38]$$

where:

$$t^*_{r_h,s_\ell} = t_{0,r_h} + T_{t_h,s_\ell} + t_{s_\ell,D} \ . \qquad [2.39]$$

Thus, for a given zone pair, r-s, the p.d.f. of the vector $t^*_{rs} = (\ldots, t^*_{r_h,s_\ell}, \ldots)$ is given by:

---

†Note that t_0 and t_D are independent random vectors, but their entries are not.

$$t^*_{rs} \sim MVN(E[t^*_{rs}],\ \Sigma^*_{rs}) \qquad [2.40]$$

where the entries of the vector of means and covariance matrix in Eq. [2.40] are given by:

$$E[t^*_{r_h,s_\ell}] = E[t_{0,r_h}] + T_{r_h,s_\ell} + E[t_{s_\ell,D}] \qquad [2.41a]$$

$$var[t^*_{r_h,s_\ell}] = var[t_{0,r_h}] + var[t_{s_\ell,D}] \qquad [2.41b]$$

$$cov[t^*_{r_h,s_\ell},\ t^*_{r_i,s_j}] = cov[t_{0,r_h},\ t_{0,r_i}] + cov[t_{D,s_\ell},\ t_{D,s_j}]. \qquad [2.41c]$$

The volume of tripmakers using access node pair (r_i,s_j) (for zone pair r-s) is thus given by:

$$q''_{r_i,s_j} = P(r_i,s_j)\cdot q_{rs}\ . \qquad [2.42]$$

To complete the calculation of Step 1b, the total access costs, L_y, have to be obtained. The total access cost[28] for a given zone pair, L_y^{rs}, can be expressed as:

$$L_y^{rs} = \text{(total travel cost)} - \text{(total cost on the basic links)}$$

$$= q_{rs}\ \text{(mean cost of the } \underline{\text{chosen}} \text{ route)} - \sum_{\substack{r_i \varepsilon N'_r \\ s_j \varepsilon N''_s}} q''_{r_i s_j}\ T_{r_i s_j}\ .$$

[28] The terms travel time and cost are used interchangeably here, without loss of generality - see Footnote 21.

As shown in Subsection 2.2.3, the average cost over the chosen route is given by $-E[\max\{-t^*_{rs}\}]$ (since travel time (cost) is negative utility), or $E[\min\{t^*_{rs}\}]$ which is readily obtained from the MNP approximation method reviewed in Subsection 2.1.2.

Thus,

$$L_y^{rs} = q_{rs}E[\min\{t^*_{rs}\}] - \sum_{\substack{r_i \varepsilon N'_r \\ s_j \varepsilon N''_s}} q''_{r_i s_j} T_{r_i s_j} \qquad [2.43]$$

and the decomposed algorithm can be applied as described above.

This concludes the review of the tools needed for developing the equilibration approach, which is the subject of the next chapter. Section 2.1 presented the MNP model and the computation of the associated choice probabilities. Section 2.2 discussed three aggregation issues that are an integral part of the equilibration scheme. The last section of Chapter 2 reviewed a decomposition of the Frank-Wolfe algorithm which is the basic algorithmic approach for equilibrating the whole transportation market.

CHAPTER 3

THE TRANSPORTATION MARKET AS A HYPERNETWORK

This chapter includes the main research results of this thesis. Section 3.1 presents a modified algorithm for the spatial aggregation traffic assignment problem presented in the last subsection. It includes some computational cost considerations that lead to a modified algorithm. This algorithm is the basis for the hypernetwork equilibration procedure developed in Section 3.3.

Section 3.2 includes the formulation of the transportation market as a hypernetwork. The first subsection introduces the equations governing the equilibration in this market, as a generalization of the S-U-E conditions presented in Section 2.3, and also the general equilibrium conditions for any probabilistic disaggregate demand model. The concepts of hypercentroid and hyperzone are introduced in Subsection 3.2.2. The hypercentroid (a subcentroid of the hypernetwork) is defined as a point in a space (the hypercube or hyperzone) spanned by the explanatory variables in the utility functions of the disaggregate demand model used. Subsection 3.2.3 explores the structure of the hypernetwork, the interpretation of hyperpaths, the use of the aggregated MNP model to assign trips to hyperpaths, and related modelling issues. This section draws heavily from the background material introduced in Sections 2.2 and 2.3.1.

The equilibration algorithm is presented in Section 3.3. The first subsection includes the necessary generalizations of the algorithm

presented in Section 3.1 to be applicable to the hypernetwork of Section 3.2. Subsection 3.3.2 includes a discussion of some properties of the equilibrium solution. These properties are an extension of the results associated with the formulation of the deterministic network equilibration problem as a mathematical program. This subsection also discusses some of the modelling assumptions that enable the network formulation of the problem. The last subsection of Section 3.3 presents a numerical example which illustrates many of the concepts discussed in this thesis.

3.1 A MODIFIED ALGORITHM FOR THE SPATIAL AGGREGATION PROBLEM

3.1.1 The Modified Algorithm

In this section, the streamlined F-W algorithm described in Subsection 2.3.3 is modified two ways. Step 1 is carried out by origin (or destination) and the convergence criterion is based on the associated objective function.

The first modification is a matter of computational efficiency. The original version of the algorithm proposed by Daganzo (1976b) requires the computation of the shortest path between all access nodes twice in every iteration (Steps 1a and 1c), unless all shortest paths are simultaneously stored during the execution of Step 1b (which in most cases would be more expensive than recomputing them). The modification described below eliminates this computational burden for most network representations.

The second modification of the algorithm is the adaptation of the

original convex combinations (Frank-Wolfe) algorithm property, that at each iteration, the solution of the Direction Finding Phase (Step 1) provides a lower bound to the optimal value of the (U-E) problem (See Eq. [2.33] and the Appendix). Hence, the difference between the value of the objective function and this lower bound can be used to construct a convenient stopping rule.

To explain the modification of the Direction Finding Step, the notion of <u>zonal access table</u> is introduced. The zonal access table for zone r is the matrix Q''_r, the entries of which $[(q''_{r_i s_j} \equiv (Q''_r)]$, are the trip interchange rates between all "outbound" access nodes of zone r (the set N'_r) and all other "inbound" access nodes (the set N"). Similarly, let L^r_y denote the total access cost between zones r and all other zones and let x^r_{ij} denote the flow on basic link ij ($ij \varepsilon \bar{L}'$) with origin r. Each time the Direction Finding Step is executed (i.e., each iteration), the shortest path trees from each origin zone's access nodes to all other access nodes are computed and the corresponding (zonal) skim tree obtained. Given this skim tree, the zonal access table entries, $q''_{r_i s_j}$ ($r_i \varepsilon N'_r$; $s_j \varepsilon N''$), are obtained using Eq. [2.42]. These entries are assigned to the basic network before a new origin zone is handled, yielding an increment of flow y^r_{ij}. As the step proceeds, these increments are summed up to obtain the new links flows $y_{ij} = \sum_r y^r_{ij}$. Similarly, the access cost L^r_y is obtained (Eq. [2.43]) [29] for the origin zone under consideration and the (zonal) access costs are summed over all zones as the algorithm proceeds to handle all origins, i.e., $L_y = \sum_r L^r_y$.

The algorithm, including the stopping rule (discussed below), is

[29] Naturally, $L^r_y = \sum_s L^{rs}_y$.

Figure 3-1

A MODIFIED ALGORITHM FOR THE CONTINUOUS APPROXIMATION OF TRIP END IMPEDANCES

STEP 0. INITIALIZATION

Determine initial link costs $\{T_{ij}\}$, initial total access costs, L_x, and the associated link flows $\{x_{ij}\}$.

STEP 1. DIRECTION FINDING

For every origin zone r:

a. Obtain the minimum travel time from each access node r_i to all other access nodes, based on the current T_{ij}'s (store the associated shortest paths).

b. 1. Find the volume using each access node pair, $q''_{r_i s_j}$.

2. Assign $q''_{r_i s_j}$ to the shortest path trees between all r_i and all inbound access nodes s_j. This yields y^r_{ij}.

3. Obtain the access cost L^r_y.

As Step 1 proceeds, obtain $y_{ij} = \sum_r y^r_{ij}$; $L_y = \sum_r L^r_y$.

STEP 2. STOPPING TEST

If $\left| L_y - L_x + \sum_{(ij)\varepsilon\bar{L}'} (y_{ij} - x_{ij})T_{ij} \right| \leq \varepsilon$ STOP; otherwise, go to Step 3.

STEP 3. STEP SIZE DETERMINATION

Find α^* that is the solution of:

$$\min_{0 \leq \alpha \leq \bar{\alpha}} \left[\alpha[L_y - L_x] + \sum_{(ij)\bar{L}'} \int_0^{x_{ij}+\alpha(y_{ij}-x_{ij})} T_{ij}(\omega)d\omega \right]$$

where $\bar{\alpha} = \min(1,\alpha_{max})$, and $\alpha_{max} = \min_{x_{ij}<y_{ij}} \{(C_{ij}-x_{ij})/(y_{ij}-x_{ij})\}$.

STEP 4. UPDATING

Obtain the sets of flows, $\{x_{ij}\}$, costs, $\{T_{ij}\}$, and total access costs, L_x, for the next iteration:

$$x^{next}_{ij} = x_{ij} + \alpha^*(y_{ij} - x_{ij})$$

$$T^{next}_{ij} = T_{ij}(x^{next}_{ij})$$

$$L^{next}_x = L_x + \alpha^*(L_y - L_x).$$

Go to Step 1.

described in Figure 3-1. (This description corresponds to the Frank-Wolfe (F-W) algorithm as given in the Appendix.) Note that only the shortest paths trees rooted at a single origin have to be stored at any given time. Subsection 3.1.2 discusses this and various other computational issues.

Step 0, the initialization step, is completely identical to Step 1, but one starts with any feasible solution, typically a flow pattern that is based on $T_{ij} = T_{ij}(0)$, $\forall ij$. Note also that during the execution of Step 1, the flow information that is stored with regard to each link in the basic network includes the current flow level, x_{ij}, and the intermediate flow level, $\sum_r y_{ij}^r$ (where the sum includes only those origins already considered at any point during the execution of Step 1).

The stopping test in Figure 3-1 is based on a predetermined tolerance level ε. The algorithm terminates when the difference between the above-mentioned lower bound and the current value of the objective function at a given iteration is less than ε. For the original Frank-Wolfe (F-W) algorithm, this can be shown (see Appendix) to be equivalent to:

$$\left| \sum_{ij} (y_{ij} - x_{ij}) \cdot T_{ij} \right| \leq \varepsilon \qquad [3.1a]$$

where the link costs and flows correspond to their values in a given iteration. Since the LHS of Eq. [3-1a] is merely the difference between the total cost over the network between the current solution $\{x_{ij}\}$ and the linearized-problem solution $\{y_{ij}\}$ and since the corresponding access costs, in our case, are given by L_x and L_y, the stopping criterion becomes:

$$\left| L_y - L_x + \sum_{(ij)\varepsilon\bar{L}'} (y_{ij} - x_{ij})\, T_{ij} \right| \leq \varepsilon \ . \qquad [3.1b]$$

The line search that determines the step size (Step 3 in Figure 3-1) can be carried out using any standard technique such as Golden Section, Fibbonaci or Bolzano search.[30]

The algorithm as described in Figure 3-1 is the basis for the equilibrium over the transportation market - the hypernetwork, that is the subject of this chapter. Before expanding the network representation to include dimensions of travel choice other than route choice, some computational considerations are given below. This discussion centers on the spatial aggregation algorithms for traffic assignment, but bears directly upon the hypernetwork equilibration algorithm presented in Section 3.3 since basically the same algorithm is used there.

3.1.2 Computational Considerations

The preceeding subsection discussed a modification of the streamlined version of the F-W algorithm presented in Sec. 2.3. In this subsection, some of the computational aspects of these algorithms are compared and some computer implementation issues, bearing upon the computational costs, are discussed.

To avoid confusion in the following discussion, the algorithm described in the beginning of Subsection 2.3.3 is referred to, in this section, as the original algorithm . The decomposition algorithms suggested by Daganzo (1977b and 1977c) are referred to as the streamlined algorithm and the algorithm depicted in Fig. 3-1 is termed the modified algorithm. The

[30] See for example Zangwill (1964) or Avriel (1976).

comparisons below are made with respect to two network representations: the multicentroid problem (whose network representation is depicted in Fig. 2-3) and the continuum approximation problem (discussed in Subsection 2.2.2).

The objective of the first comparison is to demonstrate the reduction in the computational cost of Step 1 of the algorithm resulting from the streamlining of the F-W algorithm.[31] Thus, the first comparison is between the original algorithm and the modified algorithm both applied to the multicentroid network representation, where each zone is represented by m centroids (m may represent the average number of subcentroids per zone).[32]

Let a be the total number of access nodes in the network and let c denote the number of centroids. The approximate processing time of access pathfinding and obtaining L_y for the algorithm given in Figure 3-1 can be shown [if one uses a simple dynamic programming algorithm, for instance -- see Daganzo (1977b)] to be:

$$P_1 = k(\frac{a}{c})[a + mc]mc \qquad [3.2]$$

where k is the time it takes to perform some elementary calculations (specifically one sum of three quantities and two multiplications needed

[31] Step 1, the direction finding, is the most crucial step of the algorithms described in this thesis, in the sense that it takes the most of the computation time and is what limits the capabilities of the algorithm with regard to larger networks.

[32] The algorithms are compared with regard to the multicentroid problem rather than the continuous representation since the original version of the F-W algorithm cannot handle the latter. Note, however, that Figure 3-1 includes the modified algorithm as applied to the continuum approximation problem, but it can be trivially modified to apply to the multicentroid problem.

for access pathfinding and obtaining L_y). To the above time, one should add the computation time needed to obtain the skim tree:

$$P_2 = \tau \cdot k' \cdot a \cdot n \qquad [3.3]$$

where n is the number of nodes in the basic network and k' is the computation time incurred in adding one branch to the tree of the shortest paths (which would be comparable to k and likely smaller). The coefficient, τ, stands for the average number of zones to which an access node belongs. In general, the computation of the shortest paths from an access node that is shared by more than one zone would have to be carried out separately for each of the zones sharing this access node (and the coefficient τ captures this effect); this point is further explained and commented upon when the second comparison is made.

The total approximate computation time is the sum of the abovementioned times, i.e.,

$$P = k(\frac{a}{c})(a + mc)mc + \tau \cdot k' \cdot a \cdot n \qquad [3.4a]$$

Assuming $k \simeq k'$ and $a = \Psi \cdot c$, the computation time of Step 1 of the modified algorithm becomes:

$$P = k'\Psi[mc^2(\Psi + m) + \tau \cdot cn]. \qquad [3.4b]$$

For the same problem (the multicentroid network representation), the approximate computation time using the original algorithm, P', is given by :

$$P' = k' (mc + n)mc. \qquad [3.5]$$

The difference in computation time between the original and modified algorithms, relative to the computation time of the original algorithm (when both are applied to the multicentroid problem) is given by $(P' - P)/P$. For $\Psi=1$ (i.e., the number of zones is equivalent to the number of access nodes), the relative costs are:

$$\frac{P' - P}{P} = \frac{(\frac{n}{c})(m - \tau) - m}{m(m + 1) + \tau\ (\frac{n}{c})} \qquad [3.6]$$

This function is depicted in Figure 3-2, where the relative cost increase with the original algorithm is drawn versus m, for different values of $(\frac{n}{c})$ and for $\tau = 1.5$ (e.g., on the average, 50% of the access nodes belong to two zones or 30% belong to two zones and 10% to three zones, etc.). For values of m (the average number of subcentroids per zone) between 3 and 5 and values of $(\frac{n}{c})$ between 10 and 20, the relative cost difference are about 50% to 100% of the computation time for the modified (streamlined) algorithm. The algorithm is advantageous for values of m greater than τ, for example, if m=1 the network representation is the traditional one for which the original F-W algorithm is more efficient.

Thus, it can be concluded that the decomposition of Step 1 of the F-W algorithm reduces the increase in computational cost associated with addition to centroids to a network.

The second comparison discussed in this subsection deals with the streamlined algorithm [Daganzo (1977c)] and the modified algorithm suggested in the preceeding subsection. Here the comparison can be made with regard to the problem of traffic assignment with continuous approximation of the intrazonal travel times.

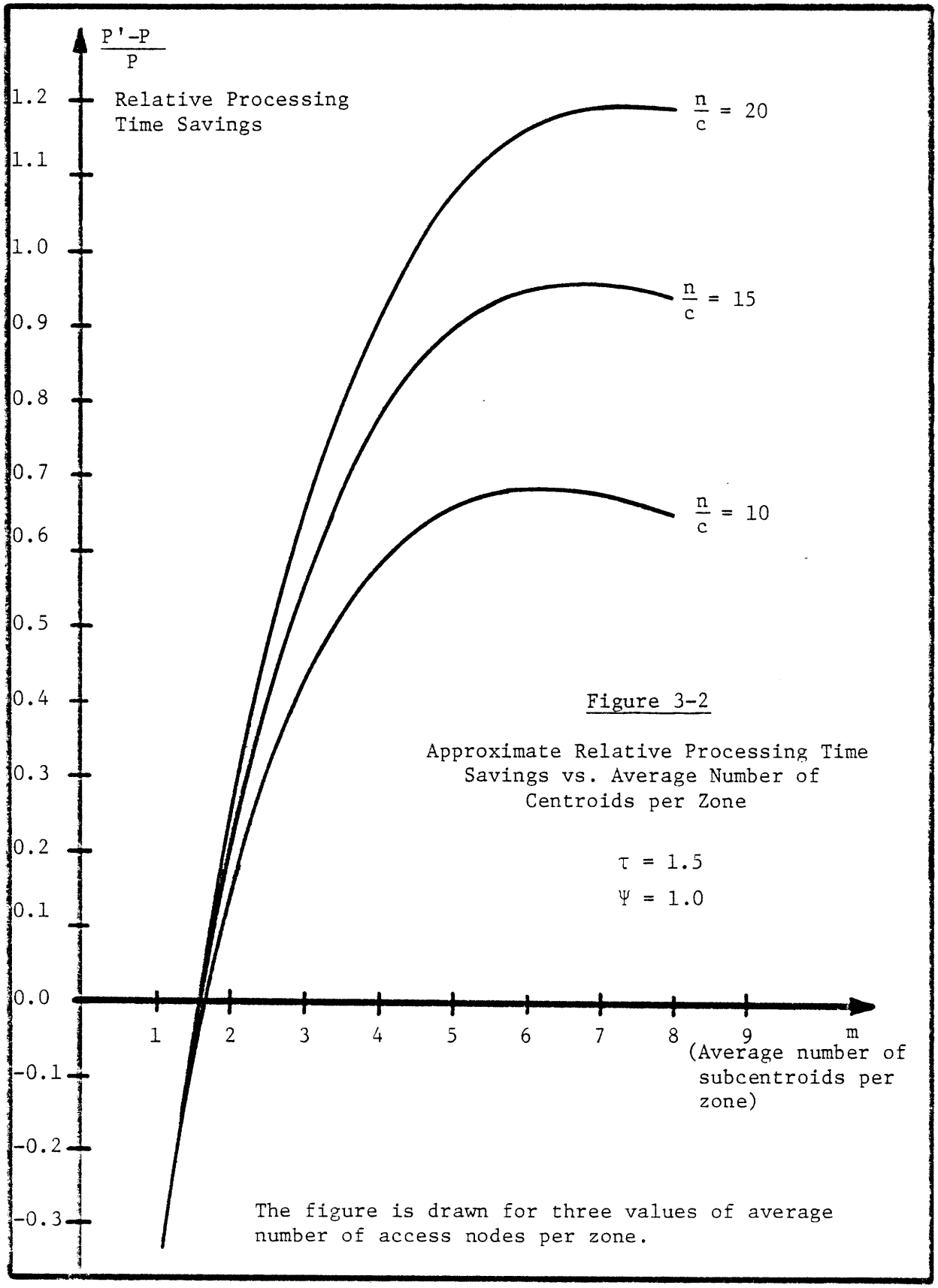


Figure 3-2

Approximate Relative Processing Time Savings vs. Average Number of Centroids per Zone

The modified algorithm (depicted in Figure 3-1) reduces the core requirements of the streamlined algorithm when the latter is applied in storage mode (i.e., Step 1c uses the path computed in Step 1a) by approximately a factor of c. (More specifically c/Ψ since the modified version stores on the average Ψ minimum path trees at every iteration.)

Transforming the MNP CPU-time formula developed by Daganzo, Bouthelier and Sheffi (1977b), the modified algorithm can be compared to the streamlined one when the latter is applied with double minimum path computation at every step. In both algorithms, the access table is given by $q''_{r_i s_j} = q_{rs} \cdot P(r_i, s_j)$ which is calculated through the Clark approximation (see Subsection 2.1.2). Using the notation developed in this subsection the total processing time for Step 1 (of the modified algorithm) is given by:

$$P'' = k''\left(\frac{a}{c}\right)^4 c^2 + k' \cdot \tau\, n\, a.^{\dagger} \qquad [3.7]$$

For the streamlined algorithm, the total processing for Step 1 is given by the same formula, but with $\tau = 2$, since the shortest paths are computed exactly twice at each iteration. Thus, if τ, the average number of zones to which an access node is connected, exceeds 2, the streamlined version should be preferred.

A slight change in the modified algorithm, in conjunction with careful implementation, can reduce the computational costs of the modified algorithm for $\tau > 2$. This can be accomplished by taking advantage of the

---

†Based on the results of Daganzo, Bouthelier and Sheffi (1977b) it can be estimated that $k'' \simeq (1/6)10^{-5}$CPU minutes, on the M.I.T. IBM 370/168 computer.

geography of the study area and ordering adjacent zones subsequently in the zones list. The algorithm can be modified to keep in core a list of the shortest path trees rooted at the last zone (or several zones) dealt with. This list should be scanned prior to computing the skim tree for a new origin zone, as some of the entries of the latter can be found in this list. When the calculation with a given zone are finished, the above-mentioned list of shortest paths is updated and the next origin zone considered in the same fashion.

This concludes the discussion of algorithms related to the spatial aggregation problem of traffic assignment. The most promising approach (especially for $\tau \leq 2$) seems to be the algorithm depicted in Figure 3-1, with the shortest path storage strategy just described.

The next section, 3.2, includes the formulation of the problem. The equilibrium equations are presented in the first section, while the second one introduces the concept of hypercentroid and hypercube that are key to the formulation of the transportation market equilibration as a hypernetwork assignment problem. Subsection 3.2.3 discusses some issues with regard to the topology of the hypernetwork and related modelling topics.

3.2 FORMULATING THE TRANSPORTATION MARKET AS A NETWORK

3.2.1 Equilibrium Conditions

In this subsection, the equilibrium conditions are defined and related to the hypernetwork representation of the transportation market, introduced in Section 1.3.

Although the equilibrium problem has been addressed in the literature

(see Section 1.2), the equilibrium conditions over the whole transportation market (when the demand side is modeled as a probabilistic proposition, based on random utility theory) do not appear to have been formalized. Therefore, the following definition is proposed :

Equilibrium Criterion

> "At equilibrium no user perceives a possible increase of his utility by unilaterally changing alternatives".

Later on in this subsection, this definition is shown to be a generalization of the Stochastic User Equilibrium principle (see Subsection 2.3.1) of traffic assignment which, in turn, is a generalization of Wardrop's (1952) User Equilibrium rule.

The equilibrium solution is obtained by solving the two systems of equations representing the demand and supply relationships. Let $P_i(z)$ denote the disaggregate demand model, i.e., the probability that a user characterized by a given combination of attributes, z, will choose travel alternative i (which may be a combination of frequency, mode, access node, route, destination, or any other alternative of interest). Also let $f_Z(z)$ denote the joint p.d.f. of Z. Using these notations (see also Subsection 2.2.1 and Eq. [2.9] in particular), the equilibrium equations are :

Demand

$$\int_Z P_i(z) \cdot f_Z(z)dz = \frac{x_i}{\sum_{\forall j} x_j} \qquad [3.8a]$$

where x_i is the number of users selecting alternative i, and $\sum_{\forall j} x_j$ is a known, fixed quantity (the population size). This equation is, of course, merely a statement of the weak law of large numbers. It states that the (predicted) market share, equals the expectation of the choice probability

with respect to the distribution of the attributes (the aggregation integral).

Supply

$$Z = Z(x) \qquad [3.8b]$$

This equation states that the values of the vector of attributes, Z, are a function of the usage of each of the alternatives.

Equations 3.8 are general and apply to a general disaggregate demand model and a general distribution of the attributes across the population. The equilibrium equations are given below for the case of the multinomial probit model utilized in this thesis.

Without yet specifying a probability law for the utility functions, and using the notation for utility functions introduced in Sec. 2, Eq. [3.8a] can be written as :

Demand

$$\Pr(U_i^* \geq U_j^* \; ; \; \forall j) = \frac{x_i}{\sum_{\forall j} x_j} \qquad [3.9a]$$

where U_j^* is the utility of travel alternative j as perceived by an individual chosen at random from the population. Equations [3.8a] and [3.9a] are, of course, identical. The distribution function of $U^* = (\ldots, U_j^*, \ldots)$ is determined by the (assumed) distribution function of the error term, ξ, (i.e., the disaggregate model) and the distribution function of the measured utility, V, across the population.

If the disaggregate model is MNP and the measured utility is approximately multivariate normally distributed (see Subsection 2.2.1), the distribution function of U* is totally characterized by a vector of means,

$\bar{V}$, and a covariance matrix[33], Σ, and in the general case, one has :

Supply

$$\bar{V} = \bar{V}(x) \quad ; \quad \Sigma = \Sigma(x) \qquad [3.9b]$$

These equations state the vector of mean utilities and the corresponding covariance matrix are functions of the usage of each one of the alternatives. In instances where Σ can be considered independent of x, standard supply modelling techniques can be used to determine Eq. [3.9b].

For Eq. [3.9a] to follow a MNP model, one has to assume that the covariance matrix of the disaggregate model (Σ_ξ -- see Footnote 33) is independent of the vector of means, V, (i.e., no "taste variation" allowed-- see Footnotes 15 and 18). If this condition is not met, the distribution of U* is not MVN and one can not use a MNP model to determine the aggregate shares. Thus, in Eq. [3.9b], not all the components of Σ can be flow-dependent (since $\bar{V}$ is flow dependent), i.e., $\Sigma_\xi \neq \Sigma_\xi(x)$, but in general, $\Sigma_V = \Sigma_V(x)$, and therefore $\Sigma = \Sigma(x)$.

The abovementioned equilibrium definition holds whether the transportation market is represented as a hypernetwork or not. However, it can be viewed as a generalization of the S-U-E principle defined in the preceeding chapter, where the network under consideration represents all dimensions of travel, i.e., a hypernetwork, as shown below.

Assume that a general hypernetwork is composed of links representing various independent dimensions of travel choice. Every link ij of the hypernetwork is associated with a utility level U'_{ij} that is the utility

[33] Note that in accordance with the notation of Subsection 2.2.1, $\Sigma=\Sigma_V+\Sigma_\xi$ or $\Sigma=\Sigma_\xi+\beta\Sigma_Z\beta^T$.

of this link as perceived by a randomly sampled individual from the population[34]. The link utilities are assumed to be normally distributed, i.e., $U' \sim MVN\ (\bar{V}', \Sigma')$.

Following the same linear transformation of Subsection 2.3.1 (see Eq. [2.27]) and defining Δ_{rs} as the link-route incidence matrix for the hypernetwork O-D pair[35] r-s, the vector of hyperpaths utilities, U^{rs}, is given by:

$$U^{rs} = U' \cdot \Delta_{rs} \qquad [3.10a]$$

and

$$U^{rs} \sim MVN\ (\bar{V}^{rs}, \Sigma^{rs}) \qquad [3.10b]$$

where:

$$\bar{V}^{rs} = \bar{V}' \Delta_{rs} \quad ; \quad \Sigma^{rs} = \Delta_{rs}^{T} \cdot \Sigma' \cdot \Delta_{rs} \qquad [3.10c]$$

Once the distribution of the alternative hyperpaths' utilities is given for every O-D pair, the equilibrium equations for the hypernetwork are:

<u>Demand</u>

$$\Pr\{U_i^{rs} \geq U_j^{rs}\ ;\ \forall j \,|\, \bar{V}^{rs}, \Sigma^{rs}\} = \frac{x_i^{rs}}{\sum_{\forall j} x_j^{rs}} \qquad [3.11a]$$

---

[34] Note that this utility was denoted U* in the preceeding discussion; the asterisk is omitted here for clarity of notation.

[35] The specific meaning of origins and destinations in the hypernetwork context is explained in the following subsection. At this point, note that since the following hypernetwork is formulated by O-D pairs, the hyperpath utilities correspond to an individual sampled at random from the population of zone r that is destined to zone s.

where $\sum_{\forall j} x_j^{rs}$ is the trip rate interchange between r and s, and x_j^{rs} is the flow on the jth route between r and s.

Supply

$$\bar{V}^{rs} = \bar{V}^{rs}(x) \quad ; \quad \Sigma^{rs} = \Sigma^{rs}(x). \qquad [3.11b]$$

In formulating the hypernetwork through the Transformation [3.10], it was assumed that links exhibit independent utilities, and that the utility associated with a hyperpath is the sum of the utilities of all links comprising it. If the hyperpaths' utilities are known (as is the case with travel demand models) this assumption is unnecessary. However, as will become apparent in the following sections, the additivity assumption is required with respect to the part of the hypernetwork exhibiting flow-dependent utility. The reason for this is that this part of the hypernetwork would typically include the street network, where path enumeration is prohibitively expensive and thus the flow dependent utility over this part is modeled differently (i.e., by link, where the Transformation [3.10] holds -- see Subsection 2.3.2). This assumption is further discussed and explained in Subsection 3.2.3 through 3.3.2.

Once the equilibrium solution (the solution of Eq. [3.10a] and [3.10b] has been obtained, one can calculate various measures with regard to the systems's evaluation, policy analysis and decision making.

The next subsection introduces the concepts of hypercentroid and hyperzone that replace the traditional centroid-zone network representation.

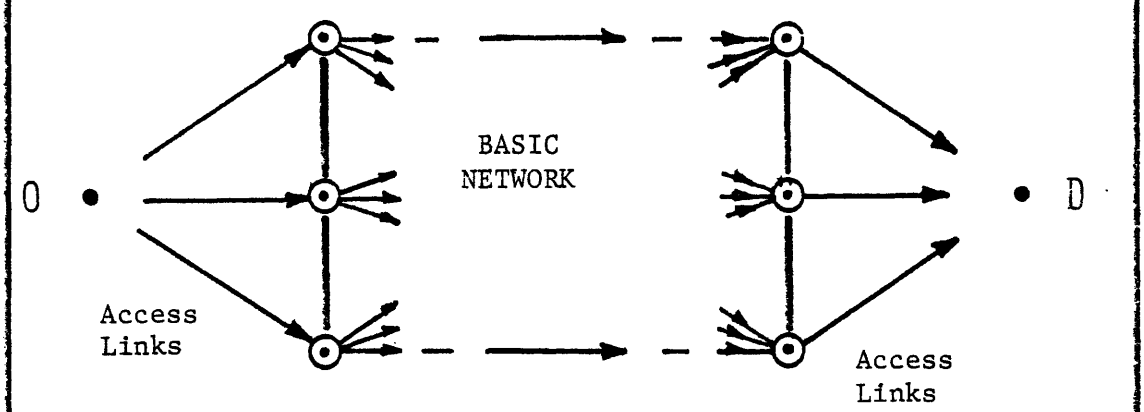
3.2.2 The Hypercentroid and Hyperzone

Consider the network representation used in Section 3.1, depicted in Figure 3-3 for one origin-destination pair. In the traditional net-

work studies (and in the abovementioned multicentroid problem), the points 0 and D stand for centroids, i.e., the network nodes where trips are originated and ended, and where all zonal residential locations and intrazonal destinations are assumed to be concentrated. In the continuum approximation case, they represent a random point in a zone, or a random point in a space, the dimensions of which are the travel times (or costs) to the access nodes. In this section, points such as 0 in Figure 3-3 are referred to as hypercentroids.

Figure 3-3

Spatial Aggregation Network Representation



Each hypercentroid is a point representing an individual or a group of individuals with the same observed and unobserved utility. If one now associates with such a point, a set of coordinates (the different observed attributes and the unobserved utility components), all hypercentroids, can be arranged in a hypercube termed a hyperzone. A hyperzone is associated with a certain zone (either origin or destination).

The number (or density) of people at a given hypercentroid is given by the joint p.d.f. of the observed and unobserved utility components. Thus, any given point in the hyperzone defines a hypercentroid referring to any number of individuals, all identified with the same composition of socio-economic characteristics, alternatives' attributes, and error terms. Each hypercentroid is also associated with a deterministic choice (e.g., of access node, mode, etc.) that is identical to all individuals represented by it (unless there are alternatives with the same perceived utilities). The choice is deterministic since the unobserved part of the utility function is one of the dimensions (coordinates) of the space of which the hypercentroid is a part.

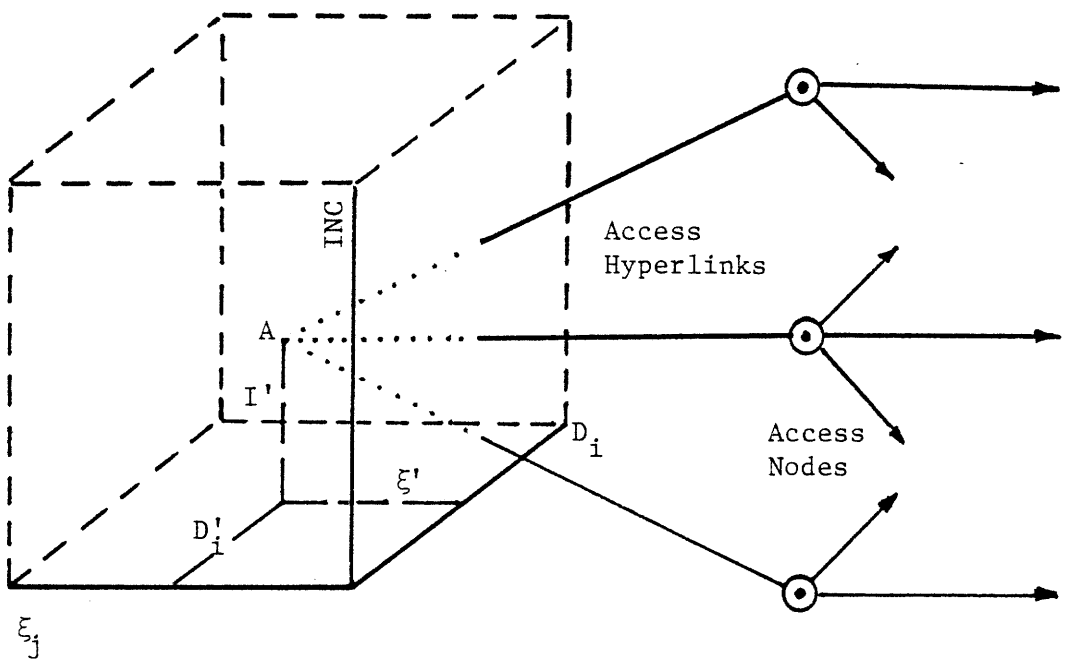
A simple hyperzone is illustrated in Figure 3-4. Point A in this figure represents a group of individuals with income I', distance to a given access node D_i' and error term ξ_j'.

The hyperzone can be visualized as partitioned into several sub-hypercubes, where in each one of these, the combination of coordinates is such that all people in a sub-hypercube choose the same alternative. As explained later in this chapter (see Section 3.3), the equilibration

Figure 3-4

A Hyperzone Example

(Only three dimensions shown: Income (INC), distance to access node i (D_i) and the error term associated with the jth alternative ξ_j)



Point A is a hypercentroid, the population of which has income I', distance D_i' to access node i and error term ξ_j for alternative j.

process can be thought of as finding the boundaries of these sub-hypercubes, and the number of decision makers in each (i.e., the integral of the decision makers' density over the sub-hypercube).

Each hypercentroid pair (one in the origin hyperzone and one in the destination hyperzone) is connected by hyperpaths. In general, each hyperpath is composed of three sequential parts:

a) A sequence of hyperlinks connecting the origin hypercentroid under consideration to an "outbound" access node. Each of these hyperlinks is associated with some utility level, e.g., intrazonal (origin zone) travel time, mode utility (from a mode-choice demand function) etc.

b) A sequence of links through the basic network. Such links are associated with a single measure of impedance (typically travel time) which is a function of the flow over the link, i.e., this links exhibit flow-dependent utility which is measured by a deterministic quantity (the rational behind this treatment of the basic network was given in Subsection 2.3.3).

c) A sequence of hyperlinks connecting "inbound" access links to a destination hypercentroid. Each of the destination hyperlinks might be associated, for example, with the destination intrazonal travel time, or the utility estimated for a destination choice model. (As shown below, each destination zone is actually associated with a single hyperlink of the latter kind.)

In general, to solve for the equilibrium in the transportation market the flow over all these hyperpaths between all hyperzones has to

be determined; thus the travel market equilibration process reduces to an assignment problem over a hypernetwork. This point is further discussed in the following sections.

In the remainder of this subsection and in order to clarify the concept, a simple hypernetwork example, where there is only one hyperzone is provided.

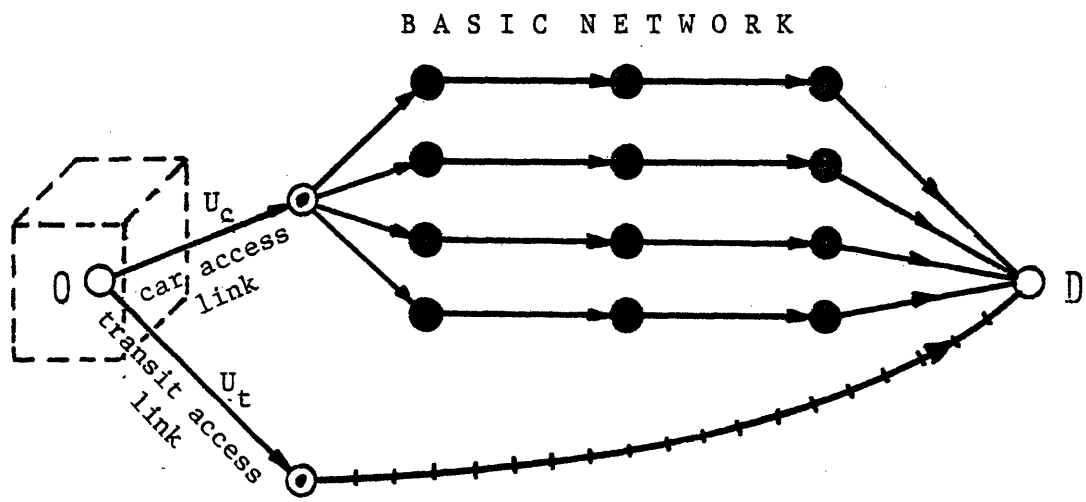
Consider an origin hyperzone, associated with a zone. Assume that the population of this zone is concentrated in one geographical point (say, the zone includes only one multi-story apartment complex). Assume further that there are two available modes (say, car and transit) between this zone and a given point destination (say, an industrial park) and the trip interchange rate to the destination is fixed and known. Thus, there is only one choice that is modelled in addition to route choice through the basic network (for the car mode) -- the choice of mode.

The dimensions of the origin hyperzone (hypercube), in this case, are the explanatory variables appearing in the mode choice mode, including the two error terms. Each point in this hyperzone is a hypercentroid. From each hypercentroid, there are two access hyperlinks connecting to each one of the access nodes (say, the parking lot and the transit station). Such a hypercentroid is shown as point 0 in Figure 3-5 below where the hypernetwork associated with the choice situation under consideration is depicted. The "cost" over the access hyperlinks is given in terms of the (negative) utility function (dis-

utility)[36] associated with each one of the modes (access hyperlinks). The cost over the basic links is given in terms of volume-delay curves.

Figure 3-5

Mode and Route Choice Hypernetwork Example



U_i = Utility of mode i

The infinitesimal trip rate interchange, dq, between the hypercentroid and the destination is known -- it equals the zonal trip inter-

[36] From this point on in the thesis, the attractiveness associated with the travel alternatives is referred to as disutility rather than utility, since travel demand is generally thought of as derived, and since most of the network literature deals with "costs" or "impedance" over links rather than with positive utility or attractiveness.

change rate times the density of people at a given hypercentroid. Note, however, that this representation poses two restrictions that are fundamental to the general approach, since in traffic assignment the "cost" of a hyperpath is the sum of the "cost" over the links comprising it: First, the utility functions have to be additive; and second, the access hyperlinks "cost" (disutility) should be expressable in travel time units to be compatible with the cost over the basic network. Both these points are further discussed in detail in Subsection 3.3.2.

The cost over the access hyperlinks can be given as a mathematical function of the coordinates of the corresponding hypercentroid (this function is given by the underlying demand model). In other words, the cost (in terms of disutility) and the choice, from a given hypercentroid, depend on the location of the hypercentroid in the hyperzone (i.e., on what segment of the population the hypercentroid represents). This point is explained below in more detail.

Consider a hyperplane (in the abovementioned hyperzone) that is parallel to the error terms axes. Such a hyperplane represents a given combination of observed attributes and characteristics (the explanatory variables associated with the measured utility functions in the mode choice model). The number of tripmakers that can be characterized by this set of attributes is given by the joint density function of the explanatory variables. The loading of this hyperplane onto access node, i, is given by the disaggregate mode choice model:

$$P_i(V) = Pr(U_i \leq U_j | V_j \quad ; \quad \forall j) \quad ; \quad \forall i.^{\dagger} \qquad [3.12]$$

---

†This equation parallels Eq. [2.2] but it is written in terms of disutilities rather than utilities (see the last footnote).

where U_i and V_i are the disutility and measured disutility of alternative i, respectively, and (using the notation introduced earlier), $U = V + \xi = \beta Z + \xi$. The entries of the matrix β are known parameters and the vector Z is fixed (by definition) over the hyperplane under consideration; V is thus fixed over the hyperplane and the choice varies from point to point as ξ varies, of course.

In the case of the example under consideration, the observed disutility of the car-hyperlink includes the travel time over the basic network (the skim tree), multiplied by its respective coefficient in the car utility function of the mode choice model. The algorithm described in the preceeding section can be used to solve the equilibration problem resulting from assigning from the abovementioned hyperplane to point D over the hypernetwork illustrated in Figure 3-5.

To execute the algorithm, as described, one needs the total cost over the access links, at every iteration (L_y^r and L_y in Figure 3-1), which is given, for the hyperplane under consideration, by (see Eq. [2.43] and the related argument):

$$L_y = dq_{OD}\, E[\min_{\forall j} \{U_j\}] \qquad [3.13]$$

where U_j is the disutility of alternative j and dq_{OD} is the population in the hyperplane under consideration.

This can be done regardless of the disaggregate demand model form used $[P_i(V)]$.

In order to find the total flow carried by all hyperlinks to a certain access node, the flow from all possible hyperplanes have to be summed up. This is merely the aggregation problem discussed in Subsection 2.2.1. Thus (under the normality assumptions discussed in the latter subsections) the total volume choosing access node i is given by (see Eq. [2.10]):

$$q_i'' = \int_{t=-\infty}^{\infty} G^i_{U*}(\ldots, t+\beta Z_i-\beta Z_j, \ldots)dt \qquad [3.14]$$

where q_i'', the aggregate share of alternative i, is, in this case, the corresponding entry to the access table (see Section 3.1) and $G^i_{U*}(\cdot)$ is defined by Eqs. [2.11] and [2.12].

The aggregate total access cost over all possible hyperplanes is merely the total utility (or total disutility, in this case) presented in Subsection 2.2.3. (It is the product of the flow from all hypercentroids, and the disutility associated with the hyperpaths used by these flows.) One has only to subtract the cost over the basic network from the total disutility to get the total access cost, AU:

$$AU = q[E[\min_{\forall i} \{U_i^*\}] - \sum_i R_i \cdot T_{iD}] \qquad [3.15]$$

where q is the population size, R_i is the share of the population choosing access node i, and T_{iD} is the travel time over the basic network, from access node i to the destination D.

The problem of assignment from a hyperzone parallels the spatial aggregation problem discussed in Subsection 2.3.3 and 3.1.1. To see

the similarity, consider another hyperplane, parallel to the explantory variables axes (and perpendicular to all the error terms axes). In every point on this hyperplane, the mode split due to unobserved characteristics is given, and the problem is to identify those sub-hyperplanes associated with a combination of explanatory variables that yield the same choice. All decision makers contained in this subspace would be assigned to the same entry of the access table. The only difference between the continuum-approximation spatial aggregation problem and the above mentioned one is that the intrazonal distribution in the hyperzone case is given by a joint MVN density function rather than a uniform density over the zone in the case of the spatial aggregation assignment.[37]

This concludes the discussion of the simple example of Figure 3-5. The extension to multi-O-D pairs hypernetworks is trivial because the access hyperlinks costs are flow-independent. Thus, in the next sub-section the representation of several choice models as a hypernetwork (e.g., the decision to take a trip, mode, access node, etc.) is explained and illustrated.

3.2.3 The Hypernetwork Structure

This section explores the hypernetwork representation of the transportation market, in greater detail. The following issues, all related to the hypernetwork topology, are covered in this subsection:

[37] In the continuum approximation case, the population density is assumed to be uniformly distributed, the moments to the access nodes found and then the normal approximation is applied. In the hyperzone case, all dimensions are assumed to be normally distributed over the population to begin with.

a. The role of hypernetworks in estimation.

b. Effect of hypernetwork structure in prediction.

c. Market segmentation.

d. Modelling of through traffic.

The first issue is related to the specification of the MNP model covariance matrix during the model estimation phase.

The linear transformation used in Subsection 3.2.1, $U^{rs} = U' \cdot \Delta_{rs}$, is an example of the choice model specification where it is assumed that a hypernetwork representation of the choice situation has been obtained prior to the estimation. In this regard, and as mentioned in Section 1.2, hypernetworks as a visualization of the choice process may serve as an aid in finding good parametrizations of the covariance matrices associated with the corresponding MNP choice model. This point is further explained below.

The main effort, in econometric studies (estimation), is devoted to the specification of the measured variables, i.e., which variables are to be included in the model and in what functional form. Another part of the specification problem is the specification with respect to the choice set, i.e., which parameters are modelled as generic and which are modelled as alternative-specific.[38]

---

[38] Assuming the usual specification $U=\beta Z+\xi$ where β is an I by J matrix of parameters, Z is a J-vector of the alternatives attributes and decision makers' characreristics and ξ is the disturbance I-vector, an attribute Z_j is called generic if the values of the entries of the associated vector of parameters are constrained to be the same (in the estimation phase) across alternatives, i.e., $\beta_{ij} \equiv \beta_j; \forall i$ (except for the base alternative). If the β_{ij}'s are unrestricted across the alternatives, Z_j is said to be alternative-specific. Some variables can, of course, be partially generic (i.e., constrained across some of the alternative).

The advantage of a generic specification (aside from reducing the computational requirements) is that it permits the prediction of choice probabilities of alternatives that are not observed in the data (say, a new mode) if the new alternative's attributes are given (one can hypothesize that a parameter β_k applies to the variable Z_k if it is in generic form).

In estimating MNP models, one has to deal with the estimation of the entries to the covariance matrix associated with the MVN distribution of ξ. In order to be able to use the estimated model for prediction of the usage of new alternatives, the covariance matrix has to be parametrized, i.e., one has to hypothesize the pattern of correlation among the alternatives. This is no different from any other specification problem concerning generic variables. The abovementioned correlation pattern is assumed, in this thesis, to be independent of the specified measured utilities (see Footnotes 16 and 18 -- this point is also discussed in more detail in Subsection 3.2.2).

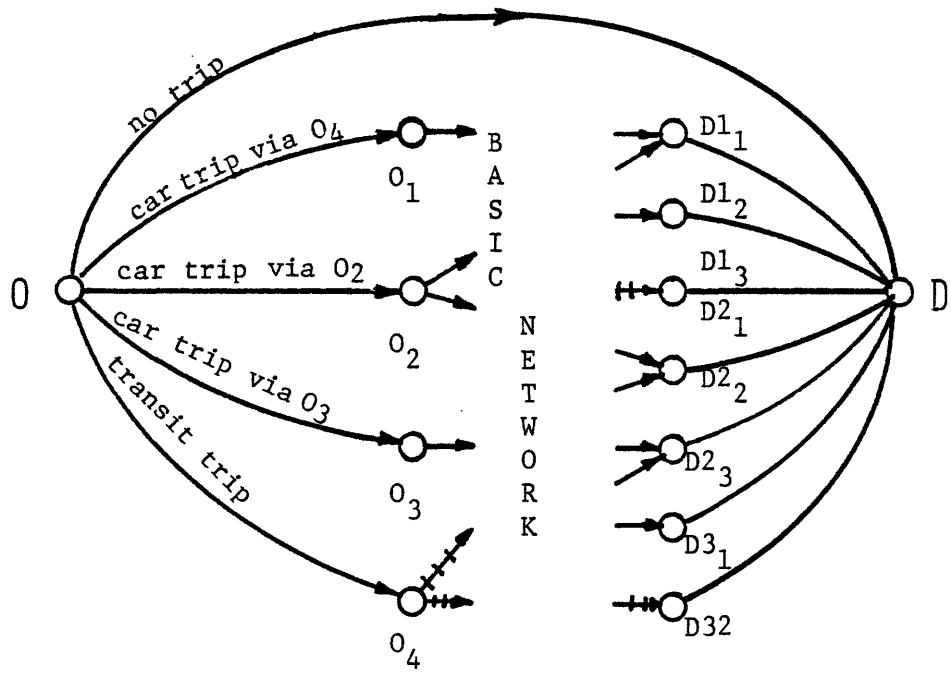
To discuss the effects of hypernetwork structure in prediction, an example of modelling the urban passenger transportation market is given below since the four step process, mentioned in Section 2.1, is widely used.

Figure 3-6 illustrates two possible hypernetwork configurations relating to the decision (at the disaggregate level) to take (or not to take) a trip, mode choice, destination choice, and route choice (through the basic network). Mathematically, these two representations are equivalent, i.e., the demand model can be represented by one origin

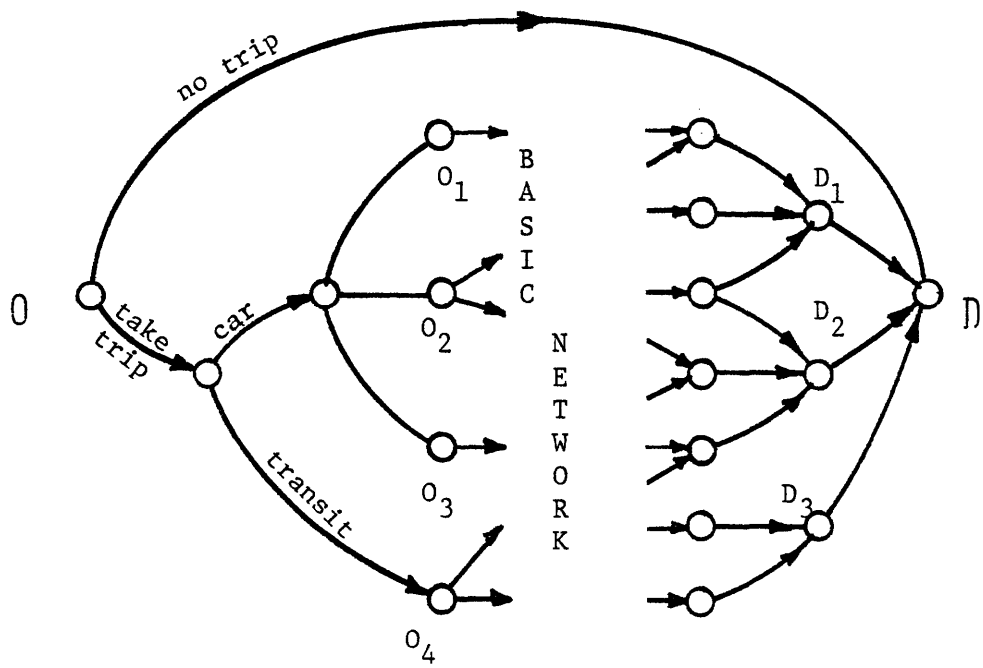
Figure 3-6

Equivalent Hypernetwork Representations of
Trip Generation, Mode, Destination and Route Choice

3.6a



3.6b



hypercentroid and one trip -end hypercentroid as in Figure 3-6a, or by a hypercentroid and a series of hypernodes on the origin side and trip-end side, as in Figure 3-6b. These are merely different visualizations of the same transportation market, as explained below.

Both representations refer to an origin zone with three access nodes to the basic network and one transit station, and three possible destinations. The decision not to take a trip is represented by the top direct hyperlink, from O to D. In Figure 3-6a, there are five hyperlinks coming out of the origin hypercentroid, associated with the disutility of the following choices:

hyperlink O-D: Not taking a trip

hyperlink $O-O_1$: Take a car trip and depart the origin zone through O_1

hyperlink $O-O_2$: Take a car trip and depart the origin zone through O_2

hyperlink $O-O_3$: Take a car trip and depart the origin zone through O_3

hyperlink $O-O_4$: Take a transit trip

The destination hyperlinks (leading to D) are associated with the disutility of choices such as:

hyperlink $D1_1D$: Arrive at destination D1 through access node $D1_1$, or;

hyperlink $D3_2D$: Arrive at destination D3 through access node $D3_2$, etc.

and the "no trip" hyperlink O-D.

The origin hyperzone is spanned by the components of the utility functions associated with the trip generation model (to take or not to take a trip), the modal split model (car vs. transit), and the choice of access node (the associated disutility here, typically includes the intrazonal travel time only). The trip-end hyperzone is spanned by the

intrazonal travel times from the access nodes and the variables appearing in the destination choice model.

Figure 3-6b depicts a different representation of the same problem. If one visualizes users moving from left to right and making a myopic decision every time a node is reached, this representation corresponds to a sequential decision, and the representation in Figure 3-6a to a simultaneous one.

Of course, if one assumes that link disutilities are independent, representations 3-6a and 3-6b will yield different results. This is because, as discussed in Subsection 2.3.1, network topology affects the probabilities of choice. However, both representations can correspond to the same probit model if the proper covariance matrix of link disutilities is selected.

The third issue mentioned at the beginning of this subsection is the market segmentation problem. Even though this is an aggregation issue, it bears upon the hypernetwork representation, as shown below.

In describing the hyperzone concept, it was assumed (see the preceeding subsection) that all explanatory variables appearing in the utility functions are MVN distributed. However, even though the density of most variables can be expected to be well approximated by a normal one, this does not hold with regard to binary variables (such as car ownership). The solution of this difficulty was discussed in Subsection 2.2.1; it consists in conditioning on the non-normal variables.

In the context of hypernetworks, such a conditioning is equivalent to the introduction of additional hyperzones to take care of market segments characterized by discrete attributes. In fact, the hypernetwork in Figure 1-3 illustrated how an additional origin could be used to model one market segment that had no access to the car mode.

In general, different population segments should be represented by different hyperzones, and for every zone there should be as many hyperzones as market segments. This ensures that the distribution of observed and unobserved attributes in a hyperzone remains MVN and that one can use analytical aggregation for calculating the shares and total access disutility in each hyperzone.

The last subject of this subsection is the modelling of traffic passing through hyperzones. Destination zones are used not only as "sinks" but they also serve through traffic to other destination zones.

To model the abovementioned phenomenon, the basic network should be expanded to include links connecting access nodes directly (across zones). The travel time over such links can be the airline distance divided by the intrazonal average speed (if warranted, a local or global street network factor can be used to modify these travel times) and they should be treated as basic links in every aspect during the assignment process.

The hypernetwork example depicted in Figure 3-6 serves as a basis for the description of the hypernetwork assignment algorithm given in the next section. The algorithm is developed in the first subsection as an extension of the spatial aggregation assignment algorithm describ-

ed in Section 3.1. Subsection 3.3.2 discusses some properties of the equilibrium solution and the assumptions leading to it. The last subsection of Section 3.4 gives an example of the equilibration procedure.

3.3 ASSIGNMENT OVER HYPERNETWORKS

3.3.1 The Hypernetwork Assignment Algorithm

The main idea of this thesis is that the problem of equilibration of the transportation market is shown to be equivalent to a traffic assignment on a modified network -- a hypernetwork.

The assignment algorithm is an extension of the one described in Section 3.1 for the spatial aggregation problem of traffic assignment. Instead of using an arbitrary point in each zone as a centroid, the hypercentroid concept, described in the preceeding section, is used[39]. Thus, Step 1 of the algorithm is applied in two stages, for each origin hyperzone. At the first stage the minimum paths from all origin access nodes to all other access nodes are found and the travel times over these paths is the origin zonal skim tree, the entries of which are given by $T_{r_i s_j}$ $(\forall i, s_j)$. The second stage consists in finding the volume, $q''_{r_i s_j}$, using each access node pair r_i-s_j (the access table):

$$q''_{r_i s_j} = q_r \cdot P(O_r, r_i, s_j, D) \qquad [3.16]$$

[39] The spatial aggregation problem is thus a special case of hypernetwork equilibration, when, trip generation, modal split, destination choice, etc., are given and only the car trips are equilibrated.

where q_r is the population in origin zone r, and $P(O_r, r_i, s_j, D)$ is the probability of choosing a hyperpath from hyperzone r to the trip-end hypercentroid D, using access nodes r_i and s_j. The disutility of hyperpath ℓ, $\bar{U}_{h_\ell}$, is given by:[40]

$$\bar{U}_{h_\ell} = U^*_{f_k} + U^*_{m_p} + U^*_{d_q} + (t_{r_i} + t_{s_j} + T_{r_i s_j})\eta \qquad [3.17]$$

where:

$U^*_{f_k}$ = Aggregate (over individuals) disutility[41] of taking k trips.[42] The vector $U^*_f = (U^*_{f_o}, U^*_{f_1})^T \sim MVN(\bar{V}_f, \Sigma_f)$.

$U^*_{m_p}$ = Aggregate disutility associated with mode p; $U^*_m \sim MVN(\bar{V}_m, \Sigma_m)$.

$U^*_{d_q}$ = Aggregate disutility of traveling to destination q; $U^*_d \sim MVN(\bar{V}_d, \Sigma_d)$

t_{r_i}, t_{s_j} = Travel time[43] from/to a random point in zone r/s to/from an access node r_i/s_j, respectively; $t_r \sim MVN(\bar{t}_r, \Sigma_{t_r})$, $t_s \sim MVN(\bar{t}_s, \Sigma_{t_s})$.

$T_{r_i s_j}$ = Minimum travel time over the basic network between access node r_i and access node s_j. (The $r_i s_j$ entry of the skim tree.)

η = Units conversion factor, "utils"/minutes.

For MNP demand models, the density function for each of the above utility vector is MVN -- see Eq. [3.14] and the related discussion. The

---

[40] The description here refers to a frequency-mode-destination-route-access node transportation planning problem, such as the one depicted in Figure 3-6.

[41] An aggregate disutility means, in this context, the disutility as perceived by a decision maker sampled at random from the population.

[42] The hypernetwork in Figure 3-6 refers to k=0 (not taking a trip) and k=1 only; thus it is applicable to say, peak hour travel.

[43] Note that in Subsection 2.3.3, t_{r_i} and t_{s_j} were denoted as t_{0,r_i} and $t_{s_j D}$.

intrazonal travel time distribution is given by an expression similar to Eq. [2.20] (trivially transformed from distance to travel time). The min-path travel time over the basic network refers to a given level of flow over the basic network and is the only deterministic disutility in Eq. [3.17]. The units conversion factor ensures the compatibility of the estimated utility functions with the travel times derived from the network topology. Rather than expressing $\bar{U}_h$ in terms of utility units, it can be expressed in terms of travel time by the transformation: $\bar{U}_h = \frac{1}{\eta}(U_f^* + U_m^* + U_d^*) + t_{r_i} + t_{s_j} + T_{r_i s_j}$. Thus, from every origin hyperzone, there is a list of available hyperpaths to the trip-end hyperzone (denoted D in Figure 3-6). To this list one should add the "No-Trip" hyperpath, thus the hyperpaths' disutilities are given by:

$$\bar{U}_{h_\ell} = \begin{cases} \frac{1}{\eta} U_{f_0}^* & \text{for "No-Trip" hyperpath} \\ \frac{1}{\eta}(U_{f_1}^* + U_{m_p}^* + U_{d_q}^*) + t_{r_i} + t_{s_j} + T_{r_i s_j} & \text{for all hyperpaths associated with taking a trip.} \end{cases} \quad [3.18]$$

Since $\bar{U}_{h_\ell}$ is a linear combination of normal variates (jointly MVN), it is itself MVN and the probability of choosing any hyperpath is given by a probit model:

$$P(O,r_i,s_j,D) = Pr(\bar{U}_{h_p} \leq \bar{U}_{h_n} \; ; \forall n) \quad [3.19]$$

where hyperpath $\bar{U}_{h_p}$ in the above equation uses access nodes r_i and s_j. Eqs. [3.19] can be readily evaluated with the use of the formulae intro-

duced in Subsection 2.1.2.

The entries of the access table are assigned to the basic network, yielding a flow pattern y^r (the set of link flows, y^r_{ij}, over the basic links ij ($ij \varepsilon \bar{L}'$) with origin at r). As Step 1 proceeds, the total link flow on the basic network is accumulated to yield the link flow pattern $y_{ij} = \sum_r y^r_{ij}$.

One quantity has to be computed each time an origin hyperzone is considered - the total access disutility. This quantity was discussed in Subsection 3.2.2 and stands for the product of access hyperpaths flows and disutilities, summed over all hyperpaths. Using the hyperpaths notation introduced in this subsection, the total access disutility, for origin r, AU^r_y (or AU^r_x, depending on its reference to the flow pattern y or x, respectively) is given by:

$$AU^r_y = q_r \cdot E[\min_{\forall \ell} \{\bar{U}_{h_\ell}\}] - \sum_{\substack{r_i \varepsilon N'_r \\ s_j \varepsilon N^h}} q''_{r_i s_j} \cdot T_{r_i s_j} \quad [3\text{-}20]$$

Since $E[\min \{\bar{U}_h\}]$ is available once the access table entries are calculated (see Eqs. [3.16],]3.19] and Subsection 2.1.2), the access disutility, AU^r_y, can be readily obtained as each origin zone is considered, and the total access disutility $AU_y = \sum_r AU^r_y$ accumulated as Step 1 proceeds.

The rest of the algorithmic steps are equivalent to those described in Section 3.1, with the new meaning and interpretation given in Section 3.2 and here. A complete description of the algorithm is given in Figure 3-7.

Figure 3-7
AN ALGORITHM FOR ASSIGNMENT OVER HYPERNETWORKS

Step 0. INITIALIZATION

Determine initial basic links costs $\{T_{ij}\}$, initial access costs AU_x, and the associated links flows $\{x_{ij}\}$.

STEP 1. DIRECTION FINDING

For each origin hyperzone r:

a. Obtain the minimum travel time from each access node r_i to all other access nodes, based on the current T_{ij}'s (store the associated shortest path trees).

b. 1) Find the volume using each access node pair,

$$q''_{r_i s_j} = q_r \cdot P(O_r, r_i, s_j, D).$$

2) Assign $q''_{r_i s_j}$ to the shortest path $r_i \to s_j$ obtained in (a); this yields a flow pattern $\{y^r_{ij}\}$ over the basic network.

3) Obtain the access disutility,

$$AU^r_y = q_r E[\min\{\bar{U}_h\}] - \Sigma\, q''_{r_i s_j} T_{r_i s_j}.$$

As the step proceeds to deal with all origin hyperzones, obtain: $y_{ij} = \sum_r y^r_{ij}$; $AU_y = \sum_r AU^r_y$.

STEP 2. STOPPING TEST

If $\left|AU_y - AU_x + \sum_{(ij)\varepsilon\bar{L}'} (y_{ij} - x_{ij})T_{ij}\right| < \varepsilon$. STOP, o.w. CONTINUE

STEP 3. STEP SIZE DETERMINATION

Find α^* that is the solution of:

$$\min_{0\le\alpha\le\bar{\alpha}} \left[\alpha(AU_y - AU_x) + \sum_{(ij)\varepsilon\bar{L}'} \int_0^{x_{ij}+\alpha(y_{ij}-x_{ij})} T_{ij}(\omega)\,d\omega\right]$$

where $\bar{\alpha} = \min\{1, \alpha_{max}\}$, and $\alpha_{max} = \min_{x_{ij}<y_{ij}} \{(C_{ij} - x_{ij})/(y_{ij} - x_{ij})\}$.

STEP 4. UPDATING

Obtain the sets of flows, $\{x_{ij}\}$, basic links costs, $\{T_{ij}\}$, and the total access disutility, AU_x, for the next iteration.

$$x^{next}_{ij} = x_{ij} + \alpha^*(y_{ij} - x_{ij})$$

$$T^{next}_{ij} = T_{ij}(x^{next}_{ij})$$

$$AU^{next}_x = AU_x + \alpha^*(AU_y - AU_x).$$

GO TO Step 1.

The next subsection describes some analytical aspects of the hypernetwork formulation.

3.3.2 Properties of the Equilibrium Solution

As mentioned in the preceeding subsection, the problem of the equilibration of the transportation market is solved by reducing the problem to an assignment over a hypernetwork. The equilibrium in the transportation market is defined by the equations (using the hyperpath notation introduced in the preceeding subsection):

Demand:

$$\Pr[\bar{U}^r_{h_i} \leq \bar{U}^r_{h_j} \;;\; \forall j \,|\, \bar{V}^r_h, \Sigma^r_h] = \frac{x^r_i}{q_r} \;;\; \forall i,r \qquad [3.21a]$$

Supply:

$$\bar{V}^r_{h_j} = \bar{V}^r_{h_j}(x^r_j) \;;\; \forall j,r$$

$$\Sigma^r_h = \Sigma^r_h(x) \;;\; \forall\, r \qquad [3.21b]$$

where $\bar{U}^r_{h_j}$ is the disutility (expressed in car travel time units) of (alternative) hyperpath i from hypercentroid r to the trip-ends hypercentroid, D, for a person sampled at random from zone r. The mean aggregate disutility, $\bar{V}^r_{h_j}$, is a function of the flow on hyperpath j through the volume-delay curves associated with the basic network, $T_{ij}(x_{ij})$ and in general, the covariance matrix of the hyperpaths' disutilities may be a function of the flow as well (see Subsection 3.2.1).

The solution of the equilibrium equations is the flow pattern over the hypernetwork. As can be seen from Figure 3-7, this flow pattern is given by the solution of the mathematical program:

$$\min\left[\sum_r q_r\left(E[\min_i\{\bar{U}^r_{h_i}\}] - \sum_{h_i} P_{h_i} T_{h_i}\right) + \sum_{(ij)\varepsilon\bar{L}'} \int_0^{x_{ij}} T_{ij}(\omega)d\omega\right] \qquad [3.22]$$

where $P_{h_i} = P(0,r_i,s_j,D)$, $T_{h_i} = T_{r_i s_j}$, and the sum $\sum_{h_i}$ goes over all $r_i \varepsilon N'_r$ and $s_j \varepsilon N''$ (see Eqs. [3.19] and [3.20]). This minimization is, of course, subject to the hypernetwork connectivity and flow conservation constraints.

The Formulation of the equilibration as a network assignment problem enables the use of many results from the theory of (deterministic) network equilibrium. Thus, an equilibrium flow pattern exists if the network has enough capacity to handle the volumes (e.g., if the no-travel alternative hyperlink between each origin and the destination hyperzone has infinite capacity, a solution would always exist). The equilibrium solution is then unique, in the sense that if there are two equilibrium solutions, both will have the same value of the objective function [3.22]. Furthermore, the link flow pattern on the basic network is unique in a strict sense as long as the routes' travel times between any two hyperzones are increasing functions of these routes' flows. this, of course, happens if the flow delay curves on all basic links are strictly increasing functions of the link flow.

The output of the equilibration procedure, in an urban transportation scenario, would include the total flow on the basic network (which, depending upon the hypernetwork representation, may include the flow on the various transit lines), the total number of users (by mode) at each destination, the total number of users of each mode (by origin) and the total number of people at each origin zone who are not taking a trip. The solution is given in terms of hyperlink flows.

Note that the market equilibration can be actually reduced to network assignment with fixed demand, or fixed trip table. The fixed trip table is given in terms of the total population of each origin zone. The elasticity of demand is handled through the hypernetwork structure (hyperpath choice) and not explicitly in the objective function as is the case with the classical formulation of traffic assignment with elastic demands (see Bekcman et al., 1956).

The condition of non-negative cost functions for the hyperlinks is met, by definition, for the basic links. The sign of the access hyperlinks' disutilities does not matter, since the disutilities can be made positive by adding a large enough constant to the disutilities of all access links (note that the solution does not change if one does that).

In the formulation presented in this thesis, it is assumed that the hyperpath disutilities are MVN distributed. This requires that all disaggregate demand models involved be based on a MVN distribution of the associated error term vector (MNP models) and that the p.d.f. of all explanatory variables in the disutility functions (excluding the ones that are conditioned upon in the aggregation process, i.e., modeled as additional hyperzones) is MVN as well (see Subsection 2.2.3).

For the hypernetwork interpretation to hold, the disutility functions have to be additive, as mentioned in the preceeding section. This requirement poses an additional (with respect to traditional studies) constraint on the model; the disutility functions have to be not only linear in the parameter and with additive disturbance term (as is the case with most disaggregate travel demand models) but also additive. In other words, the travel time over the basic network has to be modified

by a generic parameter, η, and has to enter the disutility functions as a linear-additive term. This assumption is essential in order to establish the equivalency between the transportation market and the hypernetwork equilibration problems. Moreover, the evidence with regard to linearity in variables is very mixed.[44]

The hyperpaths' disutility have to be expressed in terms of travel time (or any other measure of travel impedance on the basic network) for the access hyperlinks to be compatible with the basic links. Doing so explicitly might raise some questions in the minds of economic purists since this means the utility is used as a cardinal rather than an ordinal measure. However, the utilities are used in the same fashion in any other random utility model and the approach suggested in this thesis is no different, in this respect.

Another assumption that was used in setting up the hypernetwork is that the covariance matrix in all MNP demand models associated with the hyperlinks is independent of the vector of measured disutilities, V. If this is not the case, and $\Sigma = \Sigma(V)$, (as happens if the demand models include taste variations) the results developed with respect to the

[44]The requirement for utility functions that are linear in the parameters and with additive disturbance term that is used in most disaggregate demand models such as the logit, is rooted (just like the utility additivity requirement) in analytical feasibility. It is only for such specification of the utility functions that the logit log-likelihood function is proven to be unimodal, thereby enabling unique parameters estimates. The behavioral rationale for the traditional assumption and the one added here are identical. In addition, it is not difficult to show that the additivity assumption of travel time, coupled with a logit model of destination choice, leads to both the entropy and gravity models of trip distribution.

aggregation procedure (Subsection 2.2.1) are invalid, the expected cost over the chosen hyperpath cannot be shown to be monotonic (see Subsection (2.2.3) and the solution properties developed with respect to the hypernetwork assignment formulation of the problem do not hold. The justification of this assumption is identical to the abovementioned argument with respect to the additivity assumption.

This concludes the discussion of the algorithm and the properties of the equilibrium solution. The next subsection includes an example of a hypernetwork formulation of a contrived problem, its solution using the algorithms presented in this section and comparison with traditional approaches for the same problem.

3.3.3 Example

The hypernetwork approach and the algorithm, presented in this thesis, are applied, in this subsection, to a hypothetical example problem. The purposes of the example are to demonstrate the following issues:

a. Aggregation with MNP models,

b. Execution of the algorithm,

c. The accuracy of the results, and

d. The failure of naive techniques to converge.

In order to achieve all these goals, the problem chosen is the (classical) binary mode choice. The basic network is modeled as one link only, hence finding the skim tree, the access table, and the loading of the basic network are all trivial steps which the reader can follow without need for too many calculations. There is only one destination and one origin hyperzone spanned by the mode choice disutility functions' compon-

ents and only traditional centroids are used (i.e., no intrazonal access time is modeled). For this problem, the analysis can be performed manually and the exact equilibrium flow pattern can be found directly, enabling a comparison of the solutions.

Figure 3-8 displays the hypernetwork that corresponds to this example. Link AD stands for the basic network and the inherent disutility of the car mode is shown as link OA. The transit alternative is similarly represented by links BD and OB. The link disutilities are given in the figure for each hyperlink and the corresponding error terms distribution at the bottom of the figure.

The supply equations for each mode are the following:

$$T_{car} = \frac{10}{1 - x} \text{ minutes}$$

$$T_{tr} = 15 \text{ minutes}$$

where T_{car} and T_{tr} are the travel times by car and transit, respectively, and x is the volume on hyperpath OAD.

The disaggregate MNP model (assumed to be estimted prior to the analysis) is given by the disutility functions:

$$U_{car} = 10-5 \cdot INC + T_{car} + \xi_{car} \qquad [3.23a]$$

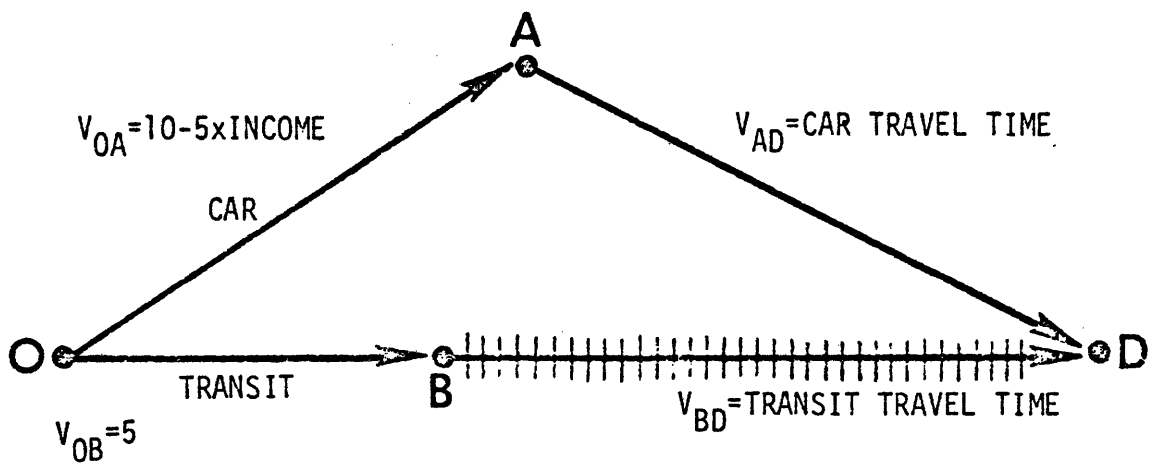
$$U_{tr} = 5 + T_{tr} + \xi_{tr} \qquad [3.23b]$$

and

$$\begin{pmatrix} \xi_{car} \\ \xi_{tr} \end{pmatrix} \sim BVN \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 50 & 0 \\ 0 & 75 \end{pmatrix} \right] \qquad [3.23c]$$

Figure 3-8

Hypernetwork for the Numerical Example



$\xi_{OA} \sim N(0;50)$	$\xi_{AD} \sim N(0;0)$
$\xi_{OB} \sim N(0;75)$	$\xi_{BD} \sim N(0;0)$
ALL MUTUALLY INDEPENDENT	

where U_{car} and U_{tr} are the disutilities associated with the car and transit mode, respectively and INC is an income variable.

These equations are not ready for use since they are not functions of the alternatives' attributes only, i.e., they have to be aggregated with respect to the socio-economic characteristics. Assuming that the income is normally distributed across the population with mean and variance, say, 4; i.e., INC ~ N(4,4), the aggregate disutilities are given by:

$$\begin{pmatrix} \bar{U}_{car} \\ \bar{U}_{tr} \end{pmatrix} = \begin{pmatrix} U_{car} \\ U_{tr} \end{pmatrix} - 5 \cdot \begin{pmatrix} INC \\ 0 \end{pmatrix}$$

and the joint density function of the aggregate disutilities is:

$$\begin{pmatrix} \bar{U}_{car} \\ \bar{U}_{tr} \end{pmatrix} \sim BVN \left[\begin{pmatrix} T_{car} - 10 \\ T_{tr} + 5 \end{pmatrix}, \begin{pmatrix} 150 & 0 \\ 0 & 75 \end{pmatrix} \right] \qquad [3.24]$$

Before applying the algorithm, the exact solution is obtained by solving the equilibrium equations directly and the non-convergence of traditional methods demonstrated. This is done below.

Writing explicitly the equilibrium equations for this example (see Eqs. [3.21]) one gets:

Demand:

$$\frac{x}{q} = P_r \, (\bar{U}_{car} \leq \bar{U}_{tr}) \qquad [3.25a]$$

where q is the total O-D trip rate interchange,

Supply:

$$T_{car} = \frac{10}{1 - x} \text{ minutes}$$

$$T_{tr} = 15 \text{ minutes .} \qquad [3.25b]$$

Substituting Eq. [3.24] in the demand function, one gets:

$$\frac{x}{q} = \Phi \left(\frac{15+T_{tr} - T_{car}}{15} \right)$$

Assuming, without loss of generality, that q=1 (i.e., the trip rate interchange is measured in units of q), and upon substituting the supply equations in the last expression, one gets a single equilibrium equation:

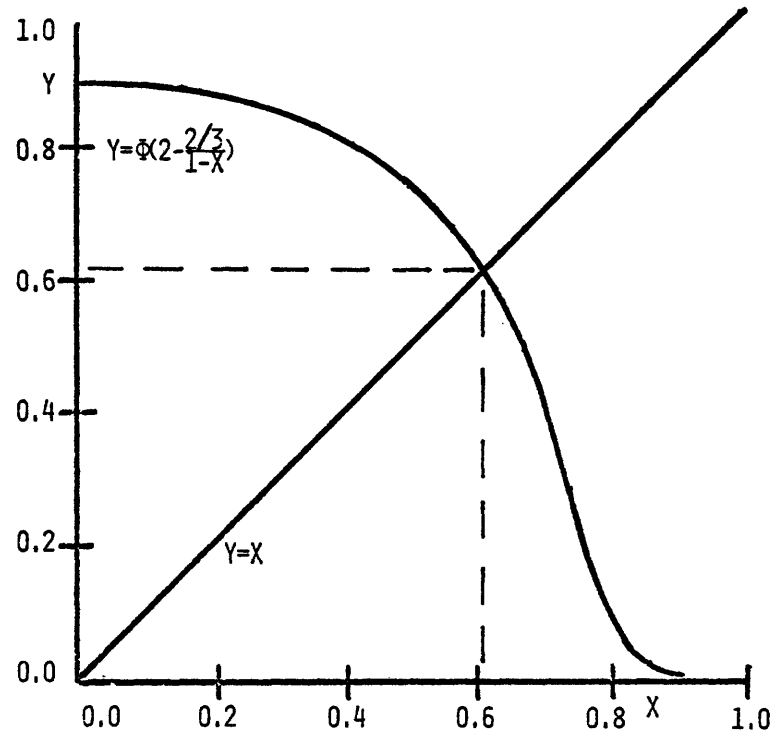
$$x = \Phi \left(2 - \frac{2/3}{1-x}\right) \qquad [3.26]$$

A graphical solution of Eq. [3.26] is illustrated in Figure 3-9, yielding $x = 0.61$ (numerical solution yield $x = 0.6116$) for the equilibrium flow pattern. This is the solution that any equilibrium algorithm should achieve for this example.

An (often used) naive equilibration procedure (involving feedback loops) consists in solving Equations [3.25a] and [3.25b] alternatively. In other words, given a flow over the basic network, the supply equations are solved. Using the current level of service, the modal split is determined by solving the demand equations, and the new flows assigned to the basic network to serve as the basis for computing the level of service for the next iteration.

Figure 3-9

Graphical Solution of the Example



In the example under consideration, and to better illustrate the point, let the initial starting value for the abovementioned procedure be X = 0.62, (a value which is very close to the equilibrium solution). Table 3-1 displays the results obtained with this procedure. The same iterative scheme could have been carried out graphically, as shown in Figure 3-10, to yield a divergence pattern resembling the well known cobweb model.

Table 3.1

DIVERGENCE PATTERN OF TRADITIONAL EQUILIBRATION SCHEME

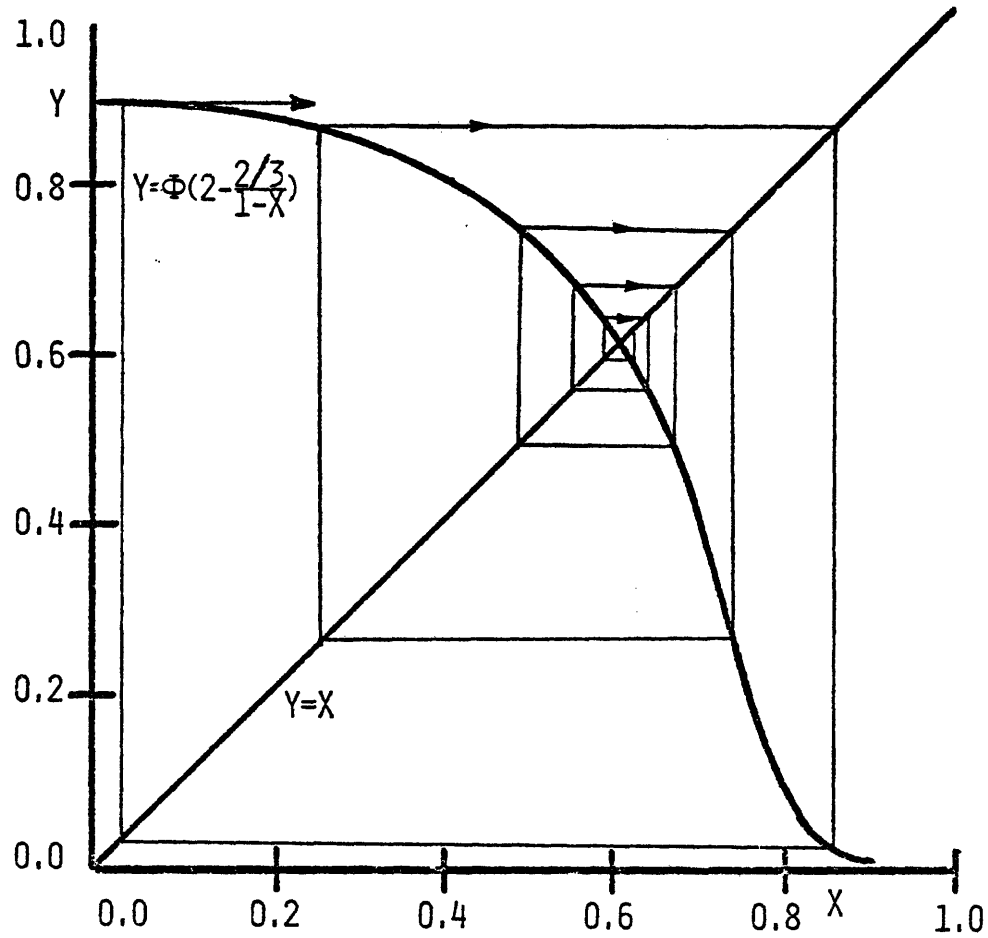
Iteration	X	T_{car} min	Y split
0	0.6200	26.3158	0.5970
1	0.5970	24.8139	0.6352
2	0.6352	27.4123	0.5685
3	0.5685	23.1750	0.6754
4	0.6754	30.8071	0.4786
5	0.4786	19.4791	0.7646
6	0.7646	42.4809	0.2027
7	0.2027	12.5423	0.8778
8	0.8778	81.8331	0.0003
9	0.0003	10.0030	0.9087
10	0.9087	109.5290	0
11	0	10.000	0.9088
12	0.9088	109.6491	0
.	.	.	.
.	.	.	.
.	.	.	.

It is evident that the naive approach diverges, for this example, even when started from an excellent initial solution.

The algorithm described in Subsection 3.3.1 is applied now to the same example problem. All the expressions needed in the course of executing the algorithm are given explicitly below, so that the algorithmic steps can be followed easily. These expressions are:

Figure 3-10

Divergence Pattern of Traditional Equilibration Method



a) The hyperpath assignment (Step 1, part b.1) formula:

The hyperpath assignment formula determines the trip table entries, which in this case are equivalent to the (aggregate) modal split.

$$q''_{AD} = q \cdot P_r\,(\bar{U}_{car} < \bar{U}_{tr}) \qquad [3.27]$$

$$= \Phi\left(\frac{30 - T_{car}}{15}\right).$$

b) The access disutility (Step 1, part b.3) formula:

This formula is obtained using Clark's formulae (see Subsection 2.1.2):

$$AU_x = E[\min(\bar{U}_{car}, \bar{U}_{tr})] - T_{car} \cdot x.$$

Since Clark's formulae apply to the maximum of two normal variates, the first expression on the RHS of the last equation should be expressed as:

$$E[\min(\bar{U}_{car}, \bar{U}_{tr})] = -E[\max(-U_{car}, -U_{tr})].$$

Following Eqs. [2.5] and substituting in Eq. [3.27], the expression for the access disutility becomes:

$$AU_x = 20 - 30X - 15\Phi\left(\frac{30 - T_{car}}{15}\right) \qquad [3.28]$$

c) The optimal step size (Step 3):

The optimal step size, α^*, is the solution of the program:

$$\min_{0 \le \alpha \le 1} \left[\alpha(AU_y - AU_x) + \int_0^{x+\alpha(y-x)} 10/(1-\omega)\,d\omega\right].$$

The upper bound of α is $\bar{\alpha} = 1$, since in this case the total trip rate, q=1 and the intermediate flow, y cannot exceed the basic link's capacity, which is 1. Letting α' be the unconstrained optimal step size, it can easily be verified that:

$$\alpha' = \frac{1 - x}{y - x} + \frac{10}{AU_y - AU_x} \qquad [3.29a]$$

and the optimal step size is given by:

$$\alpha^* = \begin{cases} 0 & \text{for } \alpha' \leq 0 \\ \alpha' & \text{for } 0 \leq \alpha' \leq 1 \\ 1 & \text{for } 1 \leq \alpha'. \end{cases} \qquad [3.29b]$$

Now the algorithmic steps can be easily followed. The algorithm is initialized at the natural "empty basic network" value, $T_{car}(0) = 10$ minutes. The first iteration of the algorithm is summarized in Figure 3-11 below. Table 3-2 displays the convergence pattern (note that since standard normal tables where used in parts b.1 and b.2 of Step 1, the displayed results are subject to some round-off error).

The table illustrates the convergence of the algorithm with respect to the skim tree, the total access disutility, the flows, and in particular the test quantity. The output of the algorithm includes all these quantities at the ε-optimal point.[45]

[45] Note that if the percentage change in flow over the hypernetwork is used as a stopping rule, the flow on access hyperlinks has to be counted for as well, not only the flow on the basic network.

Figure 3-11

FIRST ITERATION OF THE ALGORITHM FOR THE EXAMPLE PROBLEM

STEP 0:

$T_{car}(0) = 10$ minutes

$X = 0.9071$

$AU_x = -9.6730$ minutes

STEP 1:

a. $T_{AD} = 107.6426$ min. (the skim tree)

b. 1) $q''_{AD} = 0$ (the access tree)

2) $Y = 0$

3) $AU_y = 20$ min.

STEP 2:

$|AU_x - AU_y + (y-x)T_{AD}| = 67.9696$

STEP 3:

$\alpha^* = 0.2346$

STEP 4:

$X^{next} = 0.6943$

$AU_x^{next} = -2.7117$

$T_{AD}^{next} = 32.7118$

Table 3-2

THE CONVERGENCE OF THE HYPERNETWORK EQUILIBRIUM ALGORITHM

Iteration	T (skim tree)	y (hyperpath assignment)	AU_y (Access disutility) min	Test Quantity	α^*	X^{next}	AU_x^{next} min	T^{next} min
0	10					0.9071	-9.6730	107.6426
1	107.6426	0	20	67.9696	0.2346	0.6943	-2.7117	32.7118
2	32.7118	0.4283	1.2650	4.7246	1	0.4283	1.2650	17.4917
3	17.4917	0.7978	-8.1610	2.9628	0.4863	0.6080	-3.3189	25.5102
4	25.5102	0.6176	-4.2502	0.6864	1	0.6176	-4.2502	26.1502
5	26.1502	0.6013	-3.8290	0.0040	0.2816	0.6130	-4.1316	25.8398
6	25.8398	0.6092	-4.0345	0.0011	1	0.6092	-4.0345	25.5885
7	25.5885	0.6157	-4.2017	0.0009	0.3145	0.6112	-4.0871	25.7202
8	25.7202	0.6124	-4.1185	0.0008	1	0.6124	-4.1185	25.7998
9	25.7998	0.6103	-4.0635	0.0008				

In general, several more quantities can be recorded for policy analysis and decision-making purposes. For example, the total automobile-kilometers traveled, the transit operator's revenue, etc.

This concludes Chapter 3 which includes the major results of the research of this thesis. Section 3.1 described a modification of the algorithm for the assignment with spatial aggregation (reviewed in Chapter 2) based on some computational efficiency consideration. In Section 3.3, the transportation market is formulated as a hypernetwork, based on the hypercentroid and hyperzone concepts. The hypernetwork assignment algorithm is given and exemplified in Section 3.4.

The next chapter concludes the thesis. It is divided into two sections. In the first section, the research is summarized and the main modelling assumptions and approximations reviewed. In the second section, some applications of the hypernetwork approach are illustrated and some extensions of the methodology discussed.

CHAPTER 4

SUMMARY AND APPLICATIONS

4.1 SUMMARY OF THE APPROACH AND MODELLING ASSUMPTIONS

This section summarizes the thesis and reviews the assumptions and limitations of the modelling approach and the hypernetwork concept.

The main contribution of this thesis is in identifying and defining the conditions of equilibrium in the passenger transportation market, formulating the equilibration problem, and introducing an efficient algorithm for its solution. The approach consists in reducing the equilibration problem to a network assignment problem (with fixed demands) over a modified network -- the hypernetwork. All choices faced by tripmakers (e.g., taking a trip, mode, destination, route, etc.) are viewed as choice of path (hyperpath) through the hypernetwork. The travel disutility associated with each hyperpath is assumed to be MVN distributed at the disaggregate and aggregate level (i.e., the disaggregate demand models are all MNP and the components of the measured disutility functions are MVN distributed). Thus, the equilibration approach utilizes disaggregate demand models as an integral part of the transportation equilibrium problem.

The use of the MNP models eliminates theoretical shortcomings that are inherent in some other modelling approaches and enables analytical aggregation which is one of the keys to the network formulation of the problem.

The use of the normal approximation to eliminate the intrazonal

spatial aggregation bias introduces the concept of access node choice. Since the associated intrazonal travel time is viewed as MVN distributed disutility, it can be convoluted with the other dimensions of choice to produce the general MNP model representing the aggregate demand side of the problem. Furthermore, the reduction of the spatial aggregation bias makes the use of larger zones possible. This implies the use of fewer zones and fewer links and nodes in the basic network, thereby reducing the total cost of analysis.

The availability of an equilibration method also suggests that less variables may be used in the estimation phase of the analysis since by analytical aggregation and equilibration, one is making better use of the information contained in the variables, thereby compensating a possible specification error.

The hypernetwork concept can be used at almost any level of analysis, i.e., for detailed urban passenger transportation planning or as a sketch planning tool with very few zones and crudely aggregated network. Thus, the hypernetwork is more of an approach to problems rather than a rigid model. Some other applications of the hypernetwork approach are reviewed in the following section.

The approximations involved in the modelling approach include the deterministic treatment of congested hyperlinks[46] and the normal approximation of all the explatory variables (excluding impedance of congested hyperlinks) comprising the disutility of the access hyperpaths.

---

[46] For uncongested settings, stochastic effects cannot be ignored and a different approach should be used (see Subsection 2.3.1 and 2.3.2). Note, however, that in such cases the equilibration problem is obviated.

Due to the graph formulation of the problem, the flow-dependent (equilibrated) part of the hypernetwork (e.g., travel time over basic links) has to enter the disutility functions of all choice models involved in an additive form. Conservation of flow constraints have also to be met, implying that car pooling models cannot be incorporated in a straightforward fashion in the hypernetwork.[47]

The covariance matrix of all MNP models involved is assumed to be independent of the associated mean disutility. This implies that taste variations across the population cannot be conveniently modeled within a supply-demand equilibrium framework, with the approach presented in this thesis.[48]

As mentioned in the introductory chapter, the hypernetwork approach deals with static equilibrium and "steady state" behavior. This means that issues such as dynamic route selection, trip chaining and time-dependent demand are not handled within the analytic framework suggested in this thesis.

The hypernetwork methodology does provide a unified approach to the transportation planning by modelling all choices with MNP models integrated into an efficient equilibrium algorithm that is proven to

---

[47]Existing car occupancy levels can be modeled through a deflating factor applied to the basic links' congestion curves.

[48]Intrinsically different population groups can be modeled, though, through the introduction of additional origin hyperzones, with the implication of a significant increase in the analysis costs. Modelling such different groups independently is equivalent to discreticizing and segmenting the taste variations across the total population by "taste groups."

converge. Some applications and extensions of the hypernetwork concept are suggested in the next section.

4.2 APPLICATIONS AND EXTENSIONS OF HYPERNETWORK METHODOLOGY

As mentioned in Section 4.1 above, the hypernetwork approach is not limited to a given scale of application, thus hypernetworks such as the ones used in Chapter 3 can be applied to an urban setting as detailed planning tool or at the sketch planning level.

The hypernetwork methodology is applicable to determining the consequences of altering the capacity of components of a transportation system. Thus, for example, it can be used as a design tool in determining the location and capacity of parking lots along a transit line. Such parking lots are intended for "Park and Ride" mode of transportation and within a hypernetwork framework this mode can be modelled in a natural way. The usage of any suggested scheme of lots can be determined solely by a choice model, the physical characteristics of the lots, the transit line and the basic network (aggregated to reflect mainly congestion delays along competing routes of the basic network and parking lots' capacities). Hyperlinks associated with the disutility of mode transfer or parking fees can be added to the hypernetwork in a trivial fashion. The results of such analysis would be theoretically sound and obtainable at a cost that is comparable with traditional methodologies.

Another example of a hypernetwork analysis is Dial-a-Ride (DAR) systems sketch modelling and design. Using DAR supply functions[49], one

[49] Such as the ones recently developed by Daganzo, Hendrickson and Wilson (1977) for the many-to-one case.

can create a one hyperzone (covering the area of the DAR operation), one basic link (with supply curve given by the DAR performance curve) hypernetwork, where all intrazonal trip impedances are MVN approximated - using the guidelines and techniques mentioned in Subsection 2.2.2. The equilibrium flow pattern in such a market can be obtained by a manual application of the algorithm, analogous to the solution of the example of Subsection 3.3.3. Furthermore, such hypernetwork can be readily incorporated within a larger hypernetwork, for example, in designing an integrated DAR-fixed route system.

The abovementioned applications are only two simple examples of the usefulness and simplication implied by the hypernetwork approach.

Further research in transportation equilibrium modelling might be directed in two main directions. The first one is in developing "Engineering Wisdom" with regard to model specification. This includes continuing research concerning good specifications for the demand models used and in particular parametrizations of covariance matrices of MNP models. Tied to this is the development of hypernetwork representations applicable to many of the planning and design issues faced by the transportation analyst.

The second line of further research is in developing better computational techniques designed to solve the problem as posed in the thesis, i.e., more efficient demand estimation and network equilibration methods. Coupled with this, one might think of computational and theoretical developments with regard to network equilibration, such as the equilibration of several measures, all flow dependent, simultaneously; an efficient equilibration of a network exhibiting stochastic and flow-

dependent impedances; equilibration with interdependent link impedances; equilibration when route selection is dynamic and depends on traffic situations, and other transient solutions to equilibration problems.

REFERENCES

Albright, R., S. Lerman and C. Manski (1977), Report on the Development of An Estimation Program for the Multinomial Probit Model, Report prepared by Cambridge Systematic, Inc., for FHWA, Cambridge, MA.

Avriel, M. (1976), Non-Linear Programming Analysis and Methods. Prentice-Hall, Englewood Cliffs, N.J.

Alonzo, W. (1968), "Predicting Best with Imperfect Data", Journal of the American Inst. of Planners 34, 248-255.

Beckman, M., C. McGuire and C. Winsten (1956), Studies in the Economics of Transportation. Yale University Press, New Haven, Conn.

Beilner, H. and F. Jacobs (1972), "Probabilistic Aspects of Traffic Assignment", in G. Newel (Ed.), Traffic Flow and Transportation, pp. 183-194. Elsevier, New York, NY.

Ben-Akiva, M. (1974), "Structure of Passenger Travel Demand Models", Trans. Res. Record 526, 26-42, Washington, DC.

Ben-Akiva, M. (1976), "A Note on Random Utility Models, Multinomial Logit and Consumer Benefits", memorandum to E. Weiner and B. Spear, FHWA, (12/2).

Ben-Akiva, M. and S. Lerman (1977), "Disaggregate Travel and Mobility Choice Models and Measures of Accessibility", paper presented at the 3rd International Conference on Travel Modelling. Adelaide, Australia.

Bolan, R. (1970), "New Rules for Judging Analytical Techniques in Urban Planning", in ASPO, Analytical Techniques, pp. 75-84.

Bouthelier, F. (1978), An Efficient Methodology to Estimate and Predict with Multinomial Probit: Applications to Transportation Problems, Ph.D. Dissertation, Dept. of Ocean Eng., M.I.T., Cambridge, MA.

Bouthelier, F. and C. Daganzo (1978), "Aggregation with Multinomial Probit and Estimation of Disaggregate Models with Aggregate Data: A New Methodological Approach", paper presented at the 57th TRB Meeting, Washington, D.C.

Brand, D. (1972), "Travel Demand Forecasting: Some Foundations and a Review", Highway Res. Board, Special Report 143, pp. 239-282, Washington, D.C.

Burrell, J. (1968), "Multipath Route Assignment and its Application to Capacity Restraint", Proceedings, 4th International Symposium on the Theory of Traffic Flow. Karlsruhe, Germany.

Burrell, J. (1976), "Multipath Route Assignment: A Comparison of Two Methods", in Traffic Equilibrium Methods: Lecture Notes in Economics and Mathematical Systems, Vol. 118. Springer-Verlag, New York, NY.

Charles River Ass., Inc. (1972), A Disaggregate Behavioral Model of Urban Travel Demand, Report to FHWA, USDOT, Washington, D.C.

Clark, C. (1961), "The Greatest of Finite Set of Random Variables", Operation Research 9, 145-162.

Dafermos, S. (1976), "Integrated Equilibrium Flow Models for Transportation Planning", in Traffic Equilibrium Methods: Lecture Notes in Economics and Mathematical Systems, Vol. 118. Springer Verlag, New York, NY.

Daganzo, C. (1977a), "On the Traffic Assignment Problem with Flow Dependent Costs-I", Trans. Res. 11, 433-437.

Daganzo, C. (1977b), "An Equilibrium Algorithm for the Spatial Aggregation Problem of Traffic Assignment", Working Paper 7701, Stochastic Traffic Assignment Project, Inst. for Trans. Studies, University of California, Berkeley, CA.

Daganzo, C. (1977c), "Network Representation, Continuum Approximations and a Solution to the Spatial Aggregation Problem of Traffic Assignment", Working Paper 7702, Stochastic Traffic Assignment Project, Inst. of Trans. Studies, University of California, Berkeley, CA.

Daganzo, C. (1977d), "Some Remarks on Traffic Assignment Methodology Selection", Working Paper 7703, Stochastic Traffic Assignment Project, Inst. of Transp. Studies, University of California, Berkeley, CA.

Daganzo, C., F. Bouthelier and Y. Sheffi (1977a), "An Efficient Approach to Estimate and Predict with Multinomial Probit Models", paper presented at the 56th TRB Meeting. Washington, D.C.

Daganzo, C., F. Bouthelier and Y. Sheffi (1977b), "Multinomial Probit and Qualitative Choice: A Computationally Efficient Algorithm", Trans. Sci. 11, No. 4 (in press).

Daganzo, C. and Y. Sheffi (1977), "On Stochastic Models of Traffic Assignment", Trans. Sci. 11, 253-274.

Daganzo, C., C. Hendrickson and N. Wilson (1977), "An Approximate Analytical Model of Many-to-One Demand Responsive Transportation Systems", Proc., 7th Int. Symp. on Transp. and Traffic Theory, Kyoto, Japan, 743-772.

Danzig, G., F. Maier and Z. Landsdowne (1976), The Application of Decomposition to Transportation Network Analysis. Control Analysis Corp., Report DOTOTSC-76-26.

Dial, R. (1971), "A Probabilistic Multipath Traffic Assignment Model which Obviates Path Enumeration", Trans. Res. 5, 83-111.

Domencich, T. and D. McFadden (1975), Urban Travel Demand: A Behavioral Approach. North-Holland, Amsterdam, The Netherlands.

Ferland, J., M. Floridan and C. Achim (1975), "On Incremental Methods for Traffic Assignment", Trans. Res. 9, 237-239.

Fisk, C. (1978) "Some Developments in Equilibrium Traffic Assignment Methodology", a paper submitted to Trans. Res.

Florian, M. and B. Fox (1976), "On the Probabilistic Origin of Dial's Multipath Assignment Model", Trans. Res. 10, 339-341.

Florian, M. and S. Nguyen (1976), "An Application and Validation of Equilibrium Trip Assignment Methods", Trans. Sci. 10, 374-391.

Frank, M. and P. Wolfe (1956), "An Algorithm for Quadratic Programming", Naval Res. Logistics Quar. 3, 95-110.

Gartner, N. (1977), "Analysis and Control of Transportation Networks by Frank-Wolfe Decomposition", Proc. of the 7th Int. Symp. on Trans. and Traffic Theory, pp. 591-624. Kyoto, Japan.

Gunnarson, S. (1972), "An Algorithm for Multipath Traffic Assignment", Proceedings, PTRC, Urban Traffic Model Research Seminar. London, England.

Harris, A. and J. Tunner (1972), Transport Demand Models Based on Personal Characteristics, TRRL Report 65. UC., Department of Environment. U.K.

Housman, J. and D. Wise (1978), "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogenous Preferences", Econometrica 46, 403-426.

Jacobson, J. (1977), Case Study Comparison of Alternative Urban Travel Forecasting Methodologies, S.M. Dissertation, Department of Civ. Eng., M.I.T., Cambridge, MA.

Jorgenson, N. (1963), "Some Aspects of the Urban Traffic Assignment Problem", ITTE Graduate Report. University of California, Berkeley, CA.

Koppelman, F. (1976), "Guidelines for Aggregate Travel Predictions Using Disaggregate Choice Models", Trans. Res. Record 610. Washington, DC.

Lave, C. (1969), "A Behavioral Approach to Modal Split Forecasting", Trans. Res. 3, 463-480.

Leblanc, L., E. Morlok and W. Pierskalla (1975), "An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem", Trans. Sci. 9, 309-318.

Lee, D. (1973), "Requiem to Large Scale Models", Journal of the American Inst. of Planners 39, 163-178.

Lovelock, H. (1975), "A Market Segmentation Approach to Transit Planning", Proc., Trans. Res. Forum 16, 247-258.

Luce, R. (1959), Individual Choice Behavior. Wiley, New York, NY.

Manheim, M. (1973), "Proposed Direct Equilibrium Approach Using Exponential Functions". Memorandum to R. Dial, UMTA, (9/24).

Manheim, M. (1978), Fundamentals of Transportation Systems Analysis. M.I.T. Press, Cambridge, MA. (in press).

Manheim, M. and E. Ruiter (1970), "A Decision-Oriented Computer Language for Analysis of Multi-Mode Transportation Systems", Highway Res. Record 314, 135-163.

Manski, C. (1973), The Analysis of Qualitative Choice. Ph.D. Dissertation, Department of Economics, M.I.T., Cambridge, MA.

Manski, C. and S. Lerman (1977), "Estimation of Choice Probabilities from Choice-Based Samples", Econometrica, 45, 1977-1988.

Mayberry, J. (1970), "Structural Requirements for Abstract-Mode Models of Passenger Transportation" in R. Quandt (Ed.), The Demand for Travel-Theory and Measurement. Heath, Lexington, MA.

McFadden, D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior". In Zarembeka, P. (Ed.), Frontiers of Econometrics. Academic Press, New York, NY.

McFadden, D. (1977), "A Closed-Form Multinomial Choice Mode Without the Independence from Irrelevant Alternatives Restrictions", Working Paper No. 7703, Urban Travel Demand Forecasting Project, Inst. of Transportation Studies, University of California, Berkeley, CA.

McFadden, D. and F. Reid (1975), "Aggregate Travel Demand Forecasting From Disaggregate Demand Models", Trans. Res. Record 534, 24-37.

McGillivray, R. (1972), "Some Problems with Urban Transportation Planning Models", Traffic Quarterly 26, 547-559.

McLynn, J. (1976), "The Simulation of Travel Choice Behavior without the Independence of Irrelevant Alternatives Hypothesis". A paper presented at the Summer Simulation Conf., Washington, DC.

Moavenzadeh, F., Y. Sheffi and B. Brademeyer (1977), The Road Investment Analysis Model: General Framework for Network Evaluation. TAP Report 77-6, M.I.T., Cambridge, MA.

Nguyen, S. (1974), "An Algorithm for the Traffic Assignment Problem", Trans. Sci. 8, 203-216.

Nicolaidis, G., M. Wachs and T. Golob (1977), "Evaluation of Alternative Market Segmentations for Transportation Planning". Paper presented at the 56th TRB Meeting, Washington, DC.

Peat, Marwick, Mitchell and Co. (1973), A Review of Operational Urban Transportation Models; Final Report. Report No. DOT-TSC-496, Cambridge, MA.

Reichman, S. and P. Stopher (1971), "Disaggregate Stochastic Models of Travel Mode Choice", Highway Res. Record 369, 91-103.

Richards, M. (1974), "Disaggregate, Simultaneous Urban Travel Demand Models: A Brief Introduction", Transportation 3, 335-342.

Richards, M. and M. Ben-Akiva (1975), A Disaggregate Travel Demand Model. Saxon House, Westmead, England.

Ruiter, E. and M. Ben-Akiva (1978), "A System of Disaggregate Travel Demand Models: Structure, Component Models and Application Procedures". Paper presented at the 57th TRB Meeting. Washington, DC.

Scarf, H. (1973), "The Computation of Economic Equilibrium", Cowels Foundation Monograph No. 24.

Schneider, M. (1973), "Probability Maximization in Networks", Proceedings Int. Con. on Trans. Res. Bruges, Belgium, 748-755.

Sheffi, Y. (1977a), "Measures of Network Performance". Paper presented at the Int. Symp. on Travel Supply Models, Montreal, Canada.

Sheffi, Y.(1977b), The Road Network Analysis Model , S.M. Thesis, Dept. of Civ. Eng., M.I.T., Cambridge, MA.

Sheffi, Y. (1978a), "Estimating Choice Probabilities among Integer Alternatives", paper submitted to Trans. Res.

Sheffi, Y. (1978b), "A Note on the Turn and Arrival Likelihood Algorithms of Traffic Assignment", paper submitted to Trans. Res.

Sheffi, Y. and C. Daganzo (1978a), "A Unified Approach to Transportation Forecasting: Hypernetworks and Supply-Demand Equilibrium with Disaggregate Demand Models", paper presented at the 57th TRB Meeting, Washington, DC.

Sheffi, Y. and C. Daganzo (1978b), "Another Paradox of Traffic Flow", Trans. Res. 12 , 43-46.

Talvitie, A. and I. Hasan (1977), "An Equilibrium Model System for Transportation Corridors and Its Applications", paper presented at the Int. Symp. on Travel Supply Models, Montreal, Canada.

Tobin, R. (1977), "An Extension of Dial's Algorithm Utilizing a Model of Tripmakers' Perceptions", Trans. Res. 11, 337-342.

U.S. DOT/UMTA (1974), UMTA Transportation Planning System (UTPS) Reference Manual. Washington, DC.

Von Falkenhausen, H. (1966), "Traffic Assignment by a Stochastic Model", Proc. 4th Int. Conf. on Operational Science, 415-421.

Wardrop, J. (1952), "Some Theoretical Aspects of Road Traffic Research", Proc, Inst. of Civ. Eng., Part II, 325-378.

Westin, R. (1974), "Prediction from Binary Choice Models", Journal of Econometrics, 2, 1-16.

Wildermuth, B. (1972), "The Use of a Multiple Routing Technique for One-Pass Capacity Restraint Assignment", Traffic Quarterly 26, 247-265.

Williams, H. (1977), "On the Formation of Travel Demand Modles and Economic Evaluation Measures of User Benefits", Environment and Planning A 9, 285-344.

Wilson, N. (1973), "Travel Demand Forecasting: Achievements and Problems", Highway Res. Board Special Report 143, Washington, DC.

Yagar, S. (1976), "Emulation of Dynamic Equilibrium in Traffic Networks", in Traffic Equilibrium Methods: Lecture Notes in Economics and Mathematical Systems, Vol. 118, Springer-Verlag, New York, NY.

Zangwill, W. (1969), Non-linear Programming: A Unified Approach. Prentice-Hall, Englewood Cliffs, NJ.

APPENDIX†

THE CONVEX COMBINATIONS METHOD OF USER EQUILIBRATION

This appendix reviews in more detail the solution of the user-equilibrium problem of traffic assignment, using the method of convex combinations. It also includes Daganzo's algorithm for the user-equilibrated multi-centroid problem.

As mentioned in Section 2.4, the user equilibrium (U-E) problem of traffic assignment can be formulated as a mathematical program. Consider a network (directed graph) with a set J of nodes, a set of C of centroids (the special nodes where traffic originates and/or terminates), and a set N of non-centroid nodes ($U \cup C = J$; $N \cap C = \phi$). Let L be the set of links (denoted by their end nodes, i.e., $ij \varepsilon L$ if there is a link from i to j). Denote the link flow by x_{ij} and the trip rate interchange by q_{rs} ($r,s \varepsilon C$). The flow-cost (non-negative, increasing and with continuous derivatives) curve associated with each link is denoted $T_{ij}(x_{ij})$ and C_{ij} denotes the link's capacity.

Beckman et al. (1956) have shown the user-equilibrium flow pattern is the solution of the following mathematical program:

$$\text{Min}F(x) = \text{Min} \sum_{ij \varepsilon L} \int_0^{x_{ij}} T_{ij}(\omega)d\omega \qquad [A.1]$$

s.t.

$$\sum_{\forall j} x_{ij} - \sum_{\forall k} x_{ki} = \begin{cases} \sum_{\forall r} q_{ri} - \sum_{\forall s} q_{is} & \text{if } i \varepsilon C \\ 0 & \text{if } i \varepsilon N \end{cases} \qquad [A.2]$$

$$0 \leq x_{ij} < C$$

†The references mentioned in the Appendix appeared already in the thesis and are included in the main REFERENCE section.

As shown in the abovementioned reference, the equilibrium flow pattern is also unique if the flow-cost curves are strictly increasing.

Frank and Wolfe (1956) described an iterative algorithm for quadratic programming with linear constraints that [as shown by LeBlanc (1975) and Daganzo (1977a)] can be applied to solve the U-E program. The method is a feasible-direction with optimal step size one, where, in each iteration, a good direction of descent is found through a linear approximation of the objective function at the current solution. The steps of the algorithm are given in Figure A-1 below.

Given a current solution $x^k = (\ldots, x^k_{ij}, \ldots)$, Step 1 is merely a solution to a linearized U-E problem since x is assumed fixed. The direction of descent is given by the vector $(y^k - x^k)$ where y^k is determined through a minimization of a first order approximation of the objective function [A.1] at x^k. The linearized problem, $F_L(y)$ is:

$$\text{Min } F_L(y) = \text{Min}[F(x^k) + \nabla F(x^k)(y - x^k)] \qquad [A.3]$$

s.t. constaints [A.2] in y.

Not that since x^k is fixed in this LP, the term $[F(x^k) - \nabla F(x^k)x^k]$ can be discarded and the objective function of [A.3] becomes:

$$\text{Min } F_L(y) = \text{Min } \nabla F(x^k) \cdot y \qquad [A.4]$$

which is an "All-or-Nothing" problem since the links' costs, $\nabla F(k) = T(x)\Big|_{x=x^k}$ are independent of y.

The stopping test (Step 2) is based on the fact that the solution of the subproblem [A.4] is a lower bound the optimal value objective function [A.1] at each iteration. To see this, denote this optimal

value by x*. By convexity:

$$F(x^*) \geq F(x^k) + \nabla F(x^k)(x^*-x^k).$$

Since y^k minimizes $\nabla F(x^k)$ for every feasible y, we also have:

$$F(x^k) + \nabla F(x^k)(x^*-x^k) \geq F(x^k) + \nabla F(x^k)\cdot(y^k-x^k)$$

Therefore, $[F(x^k) + \nabla F(x^k)\cdot(y^k-x^k)]$ is a lower bound on $F(x^*)$ for every k. The algorithm, therefore, terminates when the current solution is within a given ε of this lower bound, i.e., when:

$$|F(x^k) + \nabla F(x^k)\cdot(y^k-x^k) - F(x^k)| \leq \varepsilon$$

or:

$$|T(x^k)\cdot(y^k-x^k)| \leq \varepsilon .$$

Step 3 consists in a one-dimensional search to find α^* that is the solution of:

$$\text{Min } F[x^k + \alpha(y^k - x^k)] \qquad [A.5]$$

Subject to:

$$\alpha \leq \alpha_{max} = \underset{x^k_{ij}<y^k_{ij}}{\text{Min}} \{(C_{ij} - x^k_{ij})/(y^k_{ij} - x^k_{ij})\} \qquad [A.6]$$

$$\alpha \leq 1 \qquad [A.7]$$

$$\alpha \geq 0 \qquad [A.8]$$

In the original revision of LeBlanc's algorithm, the constraint [A.6] was

not included, imposing a continuity requirement on the flow-cost curves. Daganzo (1977a) generalized this algorithm to be applicable to a problem including capacity constraints as well (i.e., the introduction of constraints of the form: $x_{ij} \leq C_{ij}$; $\forall ij$, in the Program [A.1, A.2], where C_{ij} are link capacities (which can be set to $C_{ij}=\infty$). The original algorithm cannot handle such a problem, since, given a feasible flow pattern x^k, the result of the direction finding step can be a flow pattern, y^k, which is infeasible, and for some links $x_{ij}^{k+1} = x_{ij}^k + \alpha^*(y_{ij}^k - x_{ij}^k)$ might exceed capacity (i.e., $x_{ij}^{k+1} > C_{ij}$). Thus, to ensure the feasibility of x^{k+1}, the step size is restricted by Eq. [A.6].

Note that the presence of link capacities requires an initialization procedure that would guarantee an initial feasible solution. Such a procedure is suggested by Daganzo (1977a) who also proved the convergence of this algorithm.

The line search of Step 3 can be accomplished by standard techniques such as Golden Section or Fibonacci search (see for example Zangwill (1969) or Avriel (1976)].

The updating phase (Step 4) consists in moving from the current solution x^k, to the next one, x^{k+1} (denoted x^{next} in Figure A-1) along the feasible descent line $(y^k - x^k)$ by the linear optimal amount $\alpha^*(y^k - x^k)$.

Nguyen (1974) suggests a special adaptation of the Convex Simplex for the same problem (Eqs. [A.1] and [A.2]). The steps of Nguyen's approach are conceptually similar except that in Step 1 (see Figure A-1) the shortest spanning trees are not recomputed from scratch but rather revised from the previous iteration. Thus, his algorithm is

Figure A-1

THE FRANK-WOLFE ALGORITHM OF TRAFFIC ASSIGNMENT

Step 0. INITIALIZATION

Determine an initial link flow pattern $\{x_{ij}\}$ and the assoicated links costs $\{T_{ij}\}$.

STEP 1. DIRECTION FINDING

Perform an "All-or-Nothing" assignment using the current T_{ij}'s; label the resulting flow pattern $\{y_{ij}\}$.

STEP 2. STOPPING TEST

If $\left| \sum_{ij \varepsilon L} (y_{ij}-x_{ij})T_{ij} \right| \leq \varepsilon$ STOP; otherwise, CONTINUE.

STEP 3. STEP SIZE DETERMINATION

Find α^* that solves: $$\min_{0 \leq \alpha \leq \bar{\alpha}} \sum_{ij \varepsilon L} \int_0^{x_{ij}+\alpha(y_{ij}-x_{ij})} T_{ij}(\omega)\,d\omega$$

where $\bar{\alpha} = \min\{1, \alpha_{max}\}$, and $\alpha_{max} = \min_{x_{ij}<y_{ij}} \{(C_{ij}-x_{ij})/y_{ij}-x_{ij})\}$

STEP 4. UPDATING

$$x_{ij}^{next} = x_{ij} + \alpha^*(y_{ij}-x_{ij})$$

$$T_{ij}^{next} = T_{ij}(x_{ij}^{next})$$

Go to Step 1.

faster but requires larger core (see Nguyen (1976) for computational comparisons). Gartner (1977) gives an excellent review of the F-W algorithm and many of its variants.

As mentioned in Section 2.3.4, it is Step 1 that limits the size of the problem that can be solved with the F-W algorithm. Daganzo (1976b), in trying to reduce the aggregation bias, represented each zone by several centroids and developed a streamlined version of the F-W algorithm to moderate the associated increase in computational cost. To explain the decomposition of the algorithm, it is applied below to the traditional single-centroid representation.

Daganzo's decomposition takes advantage of the flow-independence property of the costs over the links connecting the zone centroids to the basic network (the access links). Denoting the set of access links by L' (i.e., $ij \varepsilon L'$ if i or j ε C) and the complementary set of basic links by $\bar{L}'$, the objective function of the U-E problem (Eq. [A.1]) can be written as:

$$\text{Min } F(x) = \text{Min}\left[\sum_{ij \varepsilon L'} T_{ij} \cdot x_{ij} + \sum_{ij \varepsilon \bar{L}'} \int_0^{x_{ij}} T_{ij}(\omega)\, d\omega\right] \qquad [A.9]$$

Since for the access links, $\int_0^{x_{ij}} T_{ij}(\omega)\, d\omega = T_{ij} \cdot x_{ij}$, problem [A.1] naturally decomposes to an assignment over the access links and an assignment over the basic links. The advantages of this decomposition become apparent as one increases the number of centroids representing each zone. (Subsection 2.3.4 included a discussion of the multicentroid problem and Section 3.2 covered some of the related computational considerations.) Using the notation introduced in Subsection 2.3.4 the original version of

of Daganzo's algorithm for the multicentroid problem is given in Figure A-2. (This algorithm corresponds to the F-W algorithm description given in Subsection 2.3.4.)

Figure A-2

DAGANZO'S ALGORITHM FOR THE MULTICENTROID PROBLEM

STEP 0. INITIAILIZATION

Step 0a. Define some link costs and obtain the minimum travel cost between each access node pair.

Step 0b. For each subcentroid pair, find the access nodes that result in the least travel cost from subcentroid to centroid, allocate the trip interchanges, q_{rs}, to such access node pair and obtain the total access cost. Repeat the process for all subcentroid pairs to obtain the access table q''_{ij} and the total access cost, L_x.

Step 0c. Load the access table onto the network by either using the shortest paths obtained in Step 0a or performing another "All-or-Nothing" assignment. This yields a set of flows $x = (\ldots, x_{ij}, \ldots)$.

STEP 1. COST UPDATING

Define a new set of link costs by entering the recently obtained flow pattern, x, into the flow-cost curves $T_{ij}(\cdot)[ij \varepsilon \bar{L}']$.

STEP 2. DIRECTION FINDING

Repeat Step 0 using the link costs obtained in Step 1. The result is a flow pattern, y, and a different total access cost, L_y.

STEP 3. INTERPULATION

Obtain α^* by solving Eq. [A.5] and obtain the new set of link flows, x^{next}, and the new total access cost, L_x^{next}, by:

$$x^{next} = x + \alpha^*(y-x)$$

$$L_x^{next} = L_x + \alpha^*(L_y - L_x)$$

STEP 4. STOPPING TEST

If convergence has not been achieved go to Step 1. Otherwise, terminate.