February 2013

Location and Capacity Modeling of Network Interchanges

Aldo D. Fabregas
University of South Florida, fabregas@cutr.usf.edu

Location and Capacity Modeling of Multimodal Network Interchanges

by

Aldo Fabregas

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Industrial and Management Systems Engineering
College of Engineering
University of South Florida

Major Professor: Grisselle Centeno, Ph.D.
Tapas Das, Ph.D.
Pei-Sung Lin, Ph.D.
Beverly Ward, Ph.D.
Bo Zeng, Ph.D.

Date of Approval:
November 19, 2012

Keywords: Stackelberg games, bi-level programming, network design problem, traffic equilibrium, transportation planning

DEDICATION

To the memory of my father, Jose Fabregas.

ACKNOWLEDGMENTS

I would like to thank my major professor, Dr. Grisselle Centeno, for her support and words of encouragement during the past several years. I would also like to give a special thanks to Dr. Tapas Das for his advice from the beginning. He has been a role model as a researcher, teacher, and mentor.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Network design decisions, especially those pertaining to urban infrastructure, are made by a central authority or network leader, and taking into consideration the network users or followers. These network decision problems are formulated as non-linear bi-level programming problems. In this work, a continuous network design problem (CNDP) and discrete network design problem (DNDP) bi-level optimization programs are proposed and solved in the context of transportation planning. The solution strategy involved reformulation and linearization as a single-level program by introducing the optimality conditions of the lower level problem into the upper level problem. For the CNDP, an alternative linearization algorithm (modified least squares partitioning, MLSPA) is proposed. MLSPA takes into consideration the current arc capacity and potential expansion to find a reduced set of planes to generalize the flow-capacity surface behavior. The concepts of flow capacity surface was introduced as a way to model of congested network and capture the effect of capacity on travel time/cost. It was found that the quality of the linear approximation depends on the goodness of fit the bottleneck arcs. The proposed approach was tested with well-known benchmark problems in transportation which yielded promising results in terms of efficiency, without sacrificing solution quality.

CHAPTER 1: INTRODUCTION

A major challenge in today's society involves the redevelopment of existing systems to become sustainable. One of the aspects of system sustainability is closely related to the concept of achieving resource-efficient operations which is especially desirable in the case of urban infrastructure. Here, the term urban infrastructure refers to the physical systems that support the activities of a community; it comprises of water, sewer, electricity, communications, and transportation systems, among others. The National Academy of Engineering defines the engineering challenge for transportation systems as,

> "…the greater challenge will be engineering integrated transportation systems, making individual vehicle travel, mass transit, bicycling, and walking all as easy and efficient as possible." [1]

This challenge implies capital investments in infrastructure that enable switching from one transportation network (e.g. car or transit) to another. The place where this network switch occurs can be referred to as a multimodal network interchange. Decisions related to these interchanges have to account for their location and capacity. These decisions are associated with significant capital investments, highlighting the need of a systematic approach to transportation decision making. In that respect, the Environmental Protection Agency (EPA) considers the integration of land use and transportation decision-making as one of the key tools in its program named, Urban Sustainability and the Built Environment. This program is also aligned with other initiatives such as emissions reduction through the use of mass transit. This can be achieved through a sound transportation network decision-making practice, which will

provide the appropriate infrastructure with adequate capacity in a timely and efficient manner such that budget and operational goals are met.

From a decision modeling and operations research standpoint, the problem of communicating two or more networks falls into the network topology category, and is known as the discrete network design problem (DNDP). The scale and complexity of this network design problem is the group of constantly increasing, diverse network users, additional decision makers, and limited resources. This situation poses new challenges to current planning/decision support systems that must adapt to meet the needs of a constantly growing society.

It is necessary to construct flexible modeling frameworks that can operate with existing data warehousing systems and support a wide range of decision-making scenarios. Such decision may involve capacity allocation (parking spaces, intersection green time) and infrastructure location (new station, bus stop). The goal of this work is to contribute to the existing methodologies which will assist in network design decision-making in the transportation context by effectively addressing the following modeling key aspects:

- Central authority objective
- Network user's objectives
- Location modeling
- Capacity constraints
- Implementation considerations
- Multiple networks interaction

1.1 Intellectual Significance

Network decisions, especially those pertaining to urban infrastructure, are made by a central authority or network leader with system-wide objectives. In addition to the leader, other agents may be present in the network. These agents may adjust their behavior to adapt to the decisions of the leader. These agents could react in favor of the leader (system-optimal), selfishly (user-optimal), or against the central authority (pessimistic). These other agents are referred to as followers. This type of interaction between a central network authority, or leader, and network agents, or followers, is referred to as a Stackelberg game in the operations research literature, and is modeled via bi-level mathematical programming models.

Bi-level programs arise in different scenarios. In market economics, firms participating in a homogeneous product market can be modeled as bi-level mathematical programs. The market leader chooses his strategy first (produced quantities) and then the remaining, competing firms will adjust their strategies (production), pursuing their own benefit. In environmental economics, a government may establish a series of taxes to polluting firms. These firms, in turn, will adjust their strategies to minimize their environmental cost in a way that may not necessarily favor the government's objectives. In supply chain management, the facility location problem can be modeled, taking into consideration the changes in cost, demand, and price. In this case, the leader will attempt to find the best location for a new facility. As a result, changes in market prices and production levels may occur to accommodate the conditions imposed by the new facility. In the transportation context, a transportation agency may choose to implement a certain proposed network improvement such as an alternative transportation mode (network topology) or incentives to minimize emissions (network parameters). The

network users will react to these policies by finding the best paths to minimize their individual travel cost.

The class of problems aimed at modeling flow patterns in the context of transportation networks incorporate the non-linearity derived from congestion and the user's behavior. Such problems are associated with followers. The topology configuration problem is a binary problem and is associated with the leader. These problems constitute integer, non-linear, bi-level programming models, which in the transportation context are very likely to become large-scale.

The proposed research will contribute to the field of operations research and transportation decision-making by exploring a flexible modeling framework and proposing complexity reduction for a class of bi-level, non-linear, integer network design problems. The resulting solution framework could be further exploited in other fields related to network design such as supply chain design.

1.2 Societal Significance

The societal significance can be derived directly from the context of application in the proposed research. Transportation is one of the more challenging issues in today's society. Policy makers are looking into ways to encourage the use of mass public transportation to enhance the overall performance of the current transportation network and achieve environmental goals. These types of initiatives are expected to increase due to the emerging issues in transportation such as:

- Urban sprawl
- Increased demand/congestion
- Increasing gas prices

- Reduce emissions

Providing better transportation alternatives and infrastructure require major investments in capital improvements. Stations and facilities require decisions on the following subjects (relationship with mathematical programming in parentheses):

- Land acquisition (fixed charge cost, binary)
- Number, size, and length of stations (capacity, integer)
- Number of tracks or lanes (capacity, integer)
- Number and size of parking lots or garages (Capacity, integer or continuous)

Mathematical programming approaches can be used to model these and other decisions related to the problem of transportation infrastructure planning. By using operations research, the savings derived from capital expenses could be invested in additional projects. Therefore, there is a direct cost to society not only due to enhanced travel time, but in the investment portfolio for transportation funds. By using this type of approach, better transportations plans and alternatives can be formulated while making rational use of the available funds at the same time.

1.3 Industrial Significance

From the industry perspective the potential users for the outcome of this research could be companies developing and maintaining transportation planning decision support systems. Current issues in transportation planning are related to computational time, decision support capabilities, and solution accuracy.

At the core of every network modeling application there is a variation of a minimum-cost flow model. In the context of transportation, the core problem is the traffic

equilibrium assignment. The most advanced transportation models contain different transportation modes, several user classes, and may consider demand elasticity, among other features. Industrial-size models are implemented to take advantage of multi-core features of current computers. This feature has a direct influence in the computational time and convergence of the core algorithms running traffic assignment models. Convergence in transportation models may affect the confidence of the model in regards to decision-making. For instance, when evaluating a network improvement project such as a new arc (ramp or street), it is expected that the updated network flow, with the improvement, will differ significantly at the vicinity of the improvement. Such differences are expected to be less noticeable at certain distances from the network improvement. Inappropriate stopping criteria or certain heuristics used to compute traffic assignment may overlook these situations, inducing unexplained behaviors, attributable to the model, and thereby undermining the confidence in the decision support tool.

Transportation planning software companies could take advantage of some of the concepts related to this research by incorporating it into their current systems. Other industries with large enterprise resource planning (ERP) may also gain some benefits derived from the outcomes of the proposed research.

1.4 Dissertation Outline

In this research, location and capacity decisions for in the context of transportation network design problem are analyzed from a mathematical programming standpoint. The analysis starts with the basic multicommodity minimum cost network flow problem with linear cost and no congestion effects. The base problem was expanded adding more layers of complexity reaching a non-linear bi-level network design problem subject to congestion effects. Such problem reflects the situation of a transportation network

leader taking decisions on capacity allocation and new infrastructure taking in to consideration the network users' reaction. This dissertation addresses the modeling and solution of such network problems through a reformulation-linearization approach.

This dissertation is organized as follows: Chapter 2 presents the basic network model and the traffic assignment problem; Chapter 3 introduces the flow-capacity surface concept and linearization algorithm; Chapter 4 explains the Stackelberg games and their modeling via bi-level programming models. Chapter 5 presents the solution of selected network design problems and outlines the application of the proposed framework to multimodal networks. Chapter 6 summarized the findings and contributions of the proposed research and offer directions for future research.

CHAPTER 2: VARIANTS OF THE MINIMUM COST FLOW PROBLEM AND THE TRAFFIC EQUILIBRIUM PROBLEM

This section introduces the main network concepts used in this dissertation and the basic core problems in network design models. The main benchmarking network problems are also presented.

2.1 Network Representation and Notation

A great variety of real world situations can be modeled through a network representation. A network model may be used to represent water flows, energy, food chains, and the states of a manufacturing process, among others. The complexity of the network model depends on the nature of the situation being analyzed and the level of detail in the model abstraction. In this section, the network concepts and notations pertaining to the central topic of this dissertation are introduced. Network representation, cost functions, and notations are presented and explained.

In this dissertation, the network is represented by a directed graph $\mathcal{G}$ consisting of a set of nodes $\mathcal{N}$ and a set of arcs [2]. This network may be composed by different modal networks m (e.g., roadway network, transit network, rail network, etc.) denoted by $\mathcal{G}^m(\mathcal{N}_m, \mathcal{A}_m)$. The supersets for nodes and arcs are $\mathcal{N}$and $\mathcal{A}$ respectively. Similarly, the network formed by $\mathcal{N}$ and $\mathcal{A}$ is referred to as supernetwork.

It is assumed that the origin-destination information is available. The network demand is assumed to be concentrated in origin nodes ($\mathcal{P}$). Similarly, the destination nodes are a subset of nodes, denoted by $\mathcal{Q}$. The Cartesian product of origins and destinations generates the origin-destination matrix (O-D matrix). The O-D matrix is a set

formed by the ordered pairs (p, q). Each element $(p, q) \subset \mathcal{W}$ can be regarded as a commodity for certain problem formulations such as the multicommodity minimum cost flow problem [3].

A route or path r is an ordered sequence of arcs $a_1 \dots a_n \dots a_z$ such that the terminal node of a_n is the initial node of a_{n+1}. A path is said to connect the initial node of a_1 with the terminal node of a_z. No cycles are allowed in a path. Paths with the aforementioned characteristics are regarded as simple paths.

The pass of entities through the network is referred to as the flow. If the network element being referenced is an arc or link, the flow is regarded as an arc flow. If the element being referenced is a path, then the flow is regarded as path flow. Arc flows and path flows give origin to different problem formulations of network flow problems. Arc-flow based network problems are more granular than path-flow based network problems. Path flows have a unique representation through non-negative arc flows. However, arc flows are not uniquely represented by path flows [4].

The sign of network flows are determined using the following reasoning: a flow exiting the network (entering a node) will have a negative sign; conversely, a flow entering the network (exiting a node) will have a positive sign [2]. This convention will be adopted for conservation of flow equations throughout this document.

The arc cost function f_a represents the impedance of the arc a to the flow of entities. Such function is usually expressed in time units, but it can be translated to monetary costs through the appropriate conversion factors. Each arc has its own characteristics, leading to the existence of a variety of cost functions. For a network to be consistent, especially in transportation networks, these cost functions should be monotonically increasing with respect to the arc flow. The flow of an arc should be a function of all of the network arc flows to account for congestion and interaction effects

such as queue overflow. For an arc a this implies that $f_a = F_a(\boldsymbol{v})|a \in \mathcal{A}$ ($F_a : \mathbb{R}^{|\mathcal{A}|} \rightarrow \mathbb{R}$). These types of functions may be complicated to obtain and the effect of the flows on the surrounding arcs to the actual cost on arc a may be weak. The effect of this assumption is that the arc cost function in a only depends on the flow through arc a This can be expressed as $f_a = F_a(v_a)|a \in \mathcal{A}$. This assumption allows the cost function to be separable by arcs which facilitates the numerical treatment of network problems using mathematical programming.

The demand for transportation is represented by d_{pq} for O-D pair (p, q) for fixed demand case. For the variable demand case, or elastic demand, the demand is expressed as function of the cost of the shortest path (or less congested path) connecting O-D pair (p, q). This function will be denoted by $D(\pi)$ where π is the vector of path costs.

The notation is organized in sets, indexes, parameter and functions for clarity and to facilitate computational implementation.

2.1.1 Sets

Sets are the starting point of the mathematical model construction. Sets help define the types of elements and the size of the optimization problem. The network is represented by the sets of nodes and arcs. The number of origin destinations determines the number of commodities, which in turn, is also a measure of the complexity of the problem. The main sets related to the transportation network design problem are:

- $\mathcal{M}$: Set of modes indexed by m (car, transit, rail, etc.)
- $\mathcal{N}$: Set of all nodes (superset) for the multimodal network, indexed by i and j

- $\mathcal{A}$: Set of all arcs or pairs (i, j) for the network, indexed by a
- $\mathcal{N}_{\mathrm{m}}$: Index set of nodes for modal network m , $\mathcal{N}_m \subset \mathcal{N}$
- $\mathcal{A}_{\mathrm{m}}$: Index set of arcs for modal network m, $\mathcal{A}_m \subset \mathcal{A}$
- $\mathcal{P}$: Set of origin nodes, $\mathcal{P} \subset \mathcal{N}$
- $\mathcal{Q}$: Set of destination nodes, $\mathcal{Q} \subset \mathcal{N}$
- $\mathcal{W}$: Set of all the origin destination pairs (p, q) such that $\mathcal{W} \subseteq \mathcal{P} \times \mathcal{Q}$. For notation simplicity the subindex (p, q) may be replaced by w where required
- R_{pq}: Set of paths connecting the origin-destination pair $(p, q) \in \mathcal{W}$.
- R: Set of all the paths (superset) or routes in the network, $R = \bigcup_{(p,q)\in\mathcal{W}} R_{pq}$, indexed by r

2.1.2 Parameters

Parameters relate to the abstraction of the physical characteristics of the set of elements being modeled such as the transfer rate of a network arc or the price of traversing a network node. Parameters also are used to describe relationships between the set elements. For example, part of the network parameters are used to describe how the nodes and arcs are connected. The main parameters in the network problems are:

- d_{pq}: Demand for O-D pair $(p, q) \subset \mathcal{W}$. This parameter is used for fixed-demand problems
- δ_{pqra}: Arc-path incidence indicator; $\delta_{pqra} = 1$ if arc $a = (i, j)$ is part of the path r connecting O-D pair pq; 0 otherwise

- Δ: Matrix formed by the δ_{pqra} terms for graph $\mathcal{G}$. Its dimensions are |A| x |R|. Depending upon the context it may refer to a modal network by adding the subindex m

- γ_{pqr}: O-D pair-path indicator; γ_{pqr}=1 if origin-destination pair p, q is joined by path r; 0 otherwise

- Γ: Matrix formed by the γ_{pqr} terms. Its dimensions are |W| x |R|

- u_a: Upper bound on the flow for arc a or capacity

2.1.3 Functions

Functions describe special mathematical relationships between the parameters. The functions used in the modeling approach in this document are:

- c_r: Cost of using path r

- c: Vector of path costs

- t_a: Cost of traversing arc a. This cost is a function of the arc flow.

- $\boldsymbol{t}$: Vector of arc costs

- $D(\pi)$: Demand functions expressed in terms of the minimum cost between O-D pairs. This function is used for elastic-demand problems

- $\psi^+(i)$: Function that returns the set of out-arcs for node i, $i \in \mathcal{N}$

- $\psi^-(i)$: Function that returns the set of in-arcs for node i, $i \in \mathcal{N}$

2.1.4 Variables

This section describes the main decision variables in this research. The most relevant decision variables relate to flow, capacity, and equilibrium costs.

- h_{pqr}: Flow of commodity (p, q) on path r
- h: Vector of path flows
- f_a: Flow on arc a. Depending upon the context, a superindex m could be used to specify a modal network. f_a is a function of the path flows h_{pqr}
- f_{apq}: Flow on arc a due to O-D pair (p, q)
- x_a: Binary variable defined as 1 if arc a is implemented , 0 otherwise
- π_{pq}: Minimum cost path for O-D pair (p, q). This cost is derived from the shortest path (based on costs) connecting O-D pair (p, q)
- π_{ipq}: Equilibrium cost for node i for O-D pair (p, q). This will be used in the context of arc-node formulation
- π:Vector of minimum cost paths for all of the O-D pairs the type of problem formulation (arc-path or arc-node) will be based on the context
- y_a: Capacity increase on arc a
- $z_{aR_{pq}}$: Binary variable defined as 1 if arc a is used in an equilibrium solution, 0 otherwise.

2.2 Multicommodity Minimum Cost Flow Problem Formulations

The most basic network problem consisting of a single O-D pair is generally known as a single-commodity flow problem (each O-D pair can be referred to as a commodity). In this case, the set of O-D pairs $\mathcal{W}$ consist of only one element (p, q). In capacitated problems, the arc flows f_{ij} are subject to the capacity constraints denoted by k_{ij}. The objective function aims at minimizing the total transportation cost of transferring commodities from the origin p to the destination q. The arc flows are subject to flow conservation constraints and capacity constraints, expressed as upper bounds, on the decision variables, the arc flows f_{ij}. The cost of traversing an arc is denoted by the cost function t_{ij}. The arc cost function can be defined in different ways depending upon the nature of the problem being modeled. For example, in supply chain models transportation costs are proportional (linear) to the amounts being shipped, whereas in transportation networks the cost may be non-linear with respect to the flow which accounts for congestion effects. The minimum cost flow model can be formulated as a mathematical programming as follows [2]:

$$Min\ z(f_{pqij}) = \sum_{(p,q)\in\mathcal{W}} \sum_{(i,j)\in\mathcal{A}} t_{ij}(f_{pqij}) \tag{2.1}$$

subject to:

$$\sum_{j\,\in\,\psi^{+}(i)} f_{pqij} - \sum_{j\,\in\,\psi^{-}(i)} f_{pqij} = d_{pq}\ \forall\ i \in \mathcal{N} \tag{2.2}$$

$$d_{pq} = \begin{cases} -d_{pq}, if\ node\ i = p \\ +d_{pq}, if\ node\ i = q \\ 0, if\ node\ i \neq p \wedge i \neq q \end{cases} \tag{2.3}$$

$$\sum_{(p,q)\in\mathcal{W}} \sum_{a\in\mathcal{A}} f_{pqij} \leq u_{ij} \tag{2.4}$$

$$f_{pqa} \geq 0\ \forall\ (p,q) \in\ \mathcal{W}, (i,j) \in \mathcal{A} \tag{2.5}$$

It should be noted that in this case the arc subindex a was explicitly referred to as a pair (i,j) and the O-D pair (p,q) as w for explanation purposes. Problem 2.1 can be referred to as the minimum cost flow multicommodity flow problem (MCFMFP) in this problem; equation 2.1 represents the objective function to be minimized. For this problem, the cost functions are divisible and linear (no congestion). The function t_{ij} can be assumed as scalar function $t_{ij} = c_{ij} f_{pq}$ symbolizing the cost per flow-unit of commodity. Equation set 2.3 ensures that origin-demand requirements are satisfied. Constraint group 2.4 controls that the arc flow is kept at or below capacity. The non-negativity conditions for the decision variables are represented in constraint group 2.5. Formulation 2.1 is regarded as the arc-node formulation. In such formulation the number of variables is $|\mathcal{W}||\mathcal{A}|$ and the number of constraints is $|\mathcal{W}||\mathcal{A}| + |\mathcal{A}|$.

The arc-path formulation is an alternative method to the arc-node formulation. In the arc-path formulation the set of paths is assumed to be known and it uses one variable per path indicating path flows. The arc cost function is established by adding the flows on the path, using such arc, by means of the appropriate arc-path incidence coefficients (δ_{pqrij}). The number of variables in the arc-path formulation depends on the number paths which can grow exponentially with the size of the network. However, the constraint structure is simpler and allows the implementation of computationally-efficient solution algorithms (e.g. column generation) since only a fraction of the path is used. The arc-path formulation is as follows:

$$Min\ z(h_{pqr}) = \sum_{(i,j)\in\mathcal{A}} t_{(i,j)} \sum_{(p,q)\in\mathcal{W}} \sum_{r\in R_{pq}} \delta_{pqrij} h_{pqr} \tag{2.6}$$

subject to:

$$\sum_{r\in R_{pq}} h_{pqr} = d_{pq}\ \forall\ (p,q) \in \mathcal{W} \tag{2.7}$$

$$\sum_{(p,q)\in\mathcal{W}} \sum_{r\in R_{pq}} \delta_{pqrij} h_{pqr} \leq u_{ij} \; \forall \; (i,j) \in A \tag{2.8}$$

$$h_{pqr} \geq 0 \; \forall \; (p,q) \in \mathcal{W}, r \in R_{pq} \tag{2.9}$$

The arc flows (f_{ij}) can be included as definitional constraints for clarity. The definitional constrain is expressed in 2.10.

$$f_{ij} = \sum_{(p,q)\in\mathcal{W}} \sum_{r\in R_{pq}} \delta_{pqrij} h_{pqr} \tag{2.10}$$

The objective function and the capacity constraints can be rewritten in the following simplified form:

$$Min \; z\big(h_{pqr}, f_{ij}\big) = \sum_{(i,j)\in\mathcal{A}} t_{ij} f_{ij} \tag{2.11}$$

$$f_{ij} \leq u_{ij} \; \forall \; (i,j) \in A \tag{2.12}$$

The objective function 2.6 minimizes the path costs for the network. The demand requirement is handled by constraint group 2.7. Each path is assigned a flow h_{pqr} with the condition that all of the paths connecting the same origin-destination pair will match the corresponding demand d_{pq}. The capacity constraints are expressed in 2.8 by adding up the flows of all the paths using a particular arc. This expression can be simplified by the definitional arc flow variable f_{ij}.

The arc-path formulation has $|R|$number of variables and $|\mathcal{W}| + |\mathcal{A}|$ constraints. The simplified constraint structure is reached at the expense of an exponentially increasing number of variables. These types of problems can be handled efficiently by column generation methods.

The connection between multicommodity flows and linear programming can be made through the representation theorem [2]. This theorem states that any point x in a convex polyhedron (X) can be expressed as a convex combination of the extreme points

of X plus a nonnegative linear combination (conic combination) of the extreme rays of X. In the uncapacitated multicommodity flow, the extreme points correspond to simple paths and the extreme directions correspond to cycles [4]. Note that in the capacitated MCFP the capacity constraints form a compact polyhedral set and therefore no extreme directions (cycles) are present. Let y_i be and element of the set of extreme points (Y) and v_j an extreme direction of the set of extreme directions (V) of X.

$$x = \sum_{i \in Y} \lambda_i y_i + \sum_{j \in V} \mu_j v_j \tag{2.13}$$

$$\sum_{i \in Y} \lambda_i = 1 \tag{2.14}$$

$$\lambda_i, \mu_j \geq 0, \qquad \forall\, i \in Y\,, \forall\, j \in V \tag{2.15}$$

Let F^r be the feasible set of the arc-path formulation and let F^a be the feasible set of the arc-node formulation. Since only simple paths are included in the arc-path formulation, the set F^r can only be obtained by making $V = \emptyset$. This means that the arc-node formulation is composed of path flows and cycle flows and the path-flow formulation is a subset of the arc-flow formulation that excludes cycles. Letting $\mathcal{G}^*$ be the directed graph consisting of those arcs $a = (i, j)$ of $\mathcal{G}$ for which a positive flow f_{apq} exists in an optimal arc-flow problem solution ($f_{apq} > 0$). Let $\mathcal{W}^*$ be the O-D pairs (or commodities) and R_w^* the set of paths connecting the O-D pair w in for $\mathcal{G}^*$. The following pseudo-algorithm can be used to map the arc-path solution to an arc-node solution.

For each O-D pair pq in $\mathcal{W}^*$
For each path r in R_{pq}^*
Set $f_{rpq} \leftarrow Min\,\{f_{apq} : \delta_{arpq} = 1\}$
Subtract f_{rpq} from f_{apq} for all the arcs a with δ_{arpq}=1
Continue to next path r
Remove remaining flows since they should correspond to cycles
Continue to next O-D pair pq

Figure 1: Pseudo-Algorithm to Map an Arc-Node Solution to an Arc-Path Solution

Note that the previous algorithm does not guarantee uniqueness of the path solutions obtained due to the existence of cycles, as previously stated. Both arc-path and arc-node formulation will be the used as representations for the network problems in this dissertation.

2.3 Non-Linear Multicommodity Minimum Network Flow Problem

To more accurately represent certain network scenarios, the modeling approach needs to incorporate congestion effects. This is usually achieved by non-linear monotonically increasing cost functions. The effect of congestion can be modeled by these types of functions even in overflow conditions. An experimental test on an M/M/1 queuing system using simulated data and different traffic intensities (flow-to-capacity-ratio) was performed by Fabregas, Centeno, and Lin [5]. The authors presented closed-form expressions derived from simulated data that can be used to model congestion effects in service systems. Selected results of these tests are presented in Figure 2.

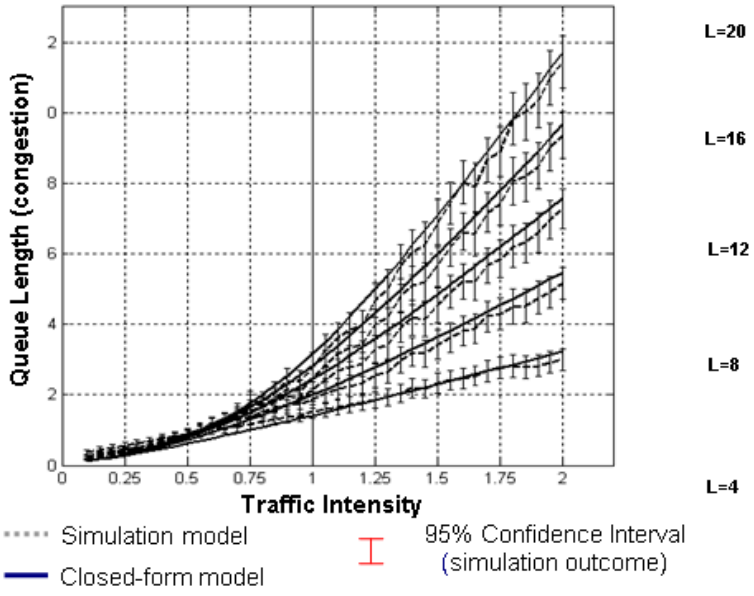

Figure 2: Congestion Cost Function in an M/M/1 Queuing System

It can be observed that under regular conditions the length of the queue (selected performance measure) tends to be a constant value. This is referred to as the free flow condition. As the congestion of the system increases, the time in the queue lengths increase linearly in overflow conditions. Similar cost functions are found in the transportation network literature to reflect congestion effects [6].

For transportation planning models, the widely adopted closed form to account for congestion in network problems is the Bureau of Public Roads curve (BPR). The BPR is presented in 2.16 as follows [7]:

$$t_a = t_{oa}[1 + B\left(\frac{f_a}{u_a}\right)^P] \tag{2.16}$$

where:

t_a: Average travel time on arc $a\ |a: (i,j) \in \mathcal{A}$

f_a: Traffic volume (flow) on arc $a\ |a: (i,j) \in \mathcal{A}$

t_{oqa}: Free-flow travel time on arc $a\ |a: (i,j) \in \mathcal{A}$

k_a: Capacity of arc $a\ |a: (i,j) \in \mathcal{A}$

B, p:Calibration parameters

2.4 Traffic Assignment Problem (TAP)

The traffic assignment problem is a variation of the minimum cost flow network problem and it is the core problem of the majority of transportation system models. A traffic assignment is a way to distribute (assign) the demand (O-D pairs) into a network following a predetermined assignment principle. The main elements of a traffic assignment problem are the network, the arc performance functions, and an O-D matrix. The result of a traffic assignment procedure consists of the flows on each network element and its corresponding cost.

The traffic assignment reflects the interaction between the supply and the demand for service on a network. The way these flows are computed depends on the rationality of the agents traversing the network. In a cooperative scheme these agents will seek to optimize a global performance measure such as total travel time. This type of traffic assignment is regarded as system-optimal. When each entity acts selfishly, seeking its own benefit (i.e. minimize its least-cost path), the resulting assignment is regarded as a user equilibrium problem (user-optimal).

2.4.1 Arc-Path Formulation of TAP

The deterministic user equilibrium (DUE) is reached when the travel cost among all the paths used, connecting the same O-D pair, is minimum.. In other words, any path connecting the same O-D pair experiencing an increased travel cost will be unused, or equivalently, will have a zero path flow. Such equilibrium conditions are known as Wardrop's first principle and this has been used extensively to reflect user behavior in transportation systems. Let h_{wr} be the flow on path r connecting O-D pair $w|\, w = (p, q) \in \mathcal{W}$. Let π_w be the minimum cost among all the paths connecting the O-D pair w. The equilibrium conditions are expressed mathematically in 2.17 and 2.18.

$$h_{wr} > 0 \Rightarrow c_{wr} = \pi_w, \forall\, r \in R_w \tag{2.17}$$

$$h_{wr} = 0 \Rightarrow c_{wr} \geq \pi_w, \forall\, r \in R_w \tag{2.18}$$

These conditions that can be expressed as constraints to a mathematical problem are given in 2.19.

$$\begin{gathered} c_{wr} - \pi_w \geq 0, \forall\, r \in R_w, \forall\, w \in \mathcal{W} \\ \sum_{r \in R_w} h_{wr} = d_w, \forall\, w \in \mathcal{W} \\ h_{wr} \geq 0, \forall\, r \in R_w, \forall\, w \in \mathcal{W} \\ \pi_w \geq 0, \forall w \in \mathcal{W} \end{gathered} \tag{2.19}$$

Under separable, symmetric, and monotonically increasing cost functions and letting $a|a:(i,j) \in \mathcal{A}$ be the deterministic traffic assignment problem it can be solved by solving the following equivalent non-linear programming problem:

$$Min\, T(f) = \sum_{a \in \mathcal{A}} \int_0^{f_a} t_a(s) ds \tag{2.20}$$

$$\sum_{r \in R_w} h_{wr} = d_w \,, \forall\, w \in \mathcal{W} \tag{2.21}$$

$$\sum_{w \in \mathcal{W}} \sum_{r \in R_w} \delta_{wra} h_{wr} = f_a, \forall\, a \in \mathcal{A} \tag{2.22}$$

The previous formulation is an equivalent mathematical form that its Lagrangian multipliers correspond to the minimum cost path at equilibrium. Furthermore, the equilibrium conditions of the optimization problems correspond to those defined in 2.19 [4]. Constraint 2.21 ensures that the demand for transportation is met. Constraint 2.22 is a definitional expression that relates path flows to arc flows. Network design problems using this formulation can be found in Ukkusuri et al. [8] , Boile & Spasovic [9], Gao & Wu [10] , and Patil & Ukkusuri [11] .

Another way to formulate the equilibrium assignment problem is through variational inequalities (VI). The formulation using VI can handle a wide range of situations such as asymmetry, non-separable/non-additively for arc costs/path costs.

$$\left(C\left(\bar{h}\right), h - \bar{h}\right) \geq 0 \tag{2.23}$$

$$\sum_{r \in w} h_{wr} = d_w \,, \forall\, w \in \mathcal{W} \tag{2.24}$$

$$h_{wr} \geq 0, \forall r \in R_w \in \mathcal{W} \tag{2.25}$$

The objective function states that since the cost flow is a non-negative quantity, the probable way in which the inner product 2.23 is non-negative is that some of the components in h are 0 for flows different than the equilibrium flows. Constraints 2.24 and

2.25 ensure that the set of feasible path flows is a polyhedral convex set. Models with VI formulation can be found in García & Marín [12] , Marín & Jaramillo [13], and Marín & García-Ródenas [14]. Both formulations are used extensively in the literature.

2.4.2 Arc-Node Formulation of TAP

The TAP can be formulated based on an arc-node formulation. In such cases the arc flow variables are defined as f_{aw} and can be interpreted as the flow on arc $a \in \mathcal{A}$ due to the O-D pair $w \mid w\colon (p,q) \in \mathcal{W}$. Note that a is the compact notation for the pair $(i,j) \in \mathcal{A}$. In the arc-node formulation the nodes are used explicitly, therefore a more granular notation was used. In this formulation the flow conservations conditions should be modeled explicitly as presented in equation 2.26. This modeling aspect is not required in the arc-path formulation since it is implicit in the path-building procedure.

$$\sum_{j \in \psi^-(i)} f_{ijw} - \sum_{j \in \psi^+(i)} f_{jiw} = d_{ijw}\,, \forall\, i \in \mathcal{N}, \forall\, w \in \mathcal{W} \qquad f_{ijw} \geq 0 \tag{2.26}$$

where,

$$d_{ijw} = \begin{cases} d_w\,, & if\ node\ i = p\ (\textbf{origin}\ of\ OD\ pair\ w = (p,q) \\ -d_w, & if\ node\ i = q\ (\textbf{destination}\ of\ OD\ pair\ w = (p.q) \\ 0, & otherwise \end{cases} \tag{2.27}$$

Flow conservation equations can be specified in compact form using the network structure. The network topology is described via the node-arc incidence matrix as defined in 2.28.

$$\phi_{ij} = \begin{cases} 1, if\ node\ i\ is\ the\ \textbf{source}\ node\ of\ link\ a = (i,j) \\ -1, \quad if\ node\ i\ is\ the\ \textbf{target}\ node\ link\ a = (i,j) \\ 0, \quad otherwise \end{cases} \tag{2.28}$$

The node-arc incidence matrix is denoted by Φ with elements ϕ_{ij}. The dimension of the matrix is$|\mathcal{N}| \times |\mathcal{A}|$. With the arc-node formulation the definitional constraints for the arc flows are as presented in equation 2.29.

$$f_a = \sum_{a \in \mathcal{A}} \sum_{w \in \mathcal{W}} f_{aw} \tag{2.29}$$

Based on the previous considerations, the arc-node formulation for the TAP is presented in 2.30 through 2.33.

$$Min\, T(f) = \sum_{a \in \mathcal{A}} \int_0^{f_a} t_a(s) ds \tag{2.30}$$

$$\sum_{j \in \psi^-(i)} f_{ijw} - \sum_{j \in \psi^+(i)} f_{jiw} = d_{aw}\,, \forall\, i \in \mathcal{N}, \forall\, w \in \mathcal{W} \tag{2.31}$$

$$f_a = \sum_{a \in \mathcal{A}} \sum_{w \in \mathcal{W}} f_{aw}\,, \forall\, a \in \mathcal{A} \tag{2.32}$$

$$f_{aw} \geq 0\ \forall\ a |\ a: (i,j) \in \mathcal{A}, \forall\, w | w: (p,q) \in \mathcal{W} \tag{2.33}$$

The equilibrium conditions for the arc-node formulation can be derived from the first-order conditions for the TAP arc-node problem [4].

$$f_{aw}(t_a(f_a) + \pi_{iw} - \pi_{jw}) = 0\,, \qquad \forall\, a \in \mathcal{A}, \forall\, w \in \mathcal{W} \tag{2.34}$$

$$t_a(f_a) + \pi_{iw} - \pi_{jw} \geq 0\,, \qquad \forall\, a \in \mathcal{A}, \forall\, w \in \mathcal{W} \tag{2.35}$$

The π_{iw} terms are the Lagrangean multipliers of the nodes resulting from the relaxation of the flow conservation constraint. Such multipliers can be interpreted as the minimum cost to deliver commodity or O-D pair $w = (p,q)$ from node i. Based on that definition, the path cost or the cost to deliver one unit of commodity (p,q) from p to q is the difference of the node cost between them ($c_w = \pi_{qw} - \pi_{pw}$). In equilibrium conditions, the difference between the cost potentials ($\pi_{iw} - \pi_{jw}$) for commodity w on an arc a should be equal to the cost of one unit of w traversing the arc $t_a(f_a)$. For arcs

where this condition is not satisfied, the arc should have flows of 0 for that commodity (f_{aw}=0).

2.5 Taxonomy of Traffic Assignment Problems

Traffic assignment problems comprise of several modeling components that can be modified to accommodate a wide variety of modeling situations. The inclusion of more components increases the complexity of the resulting traffic assignment model. In this section, a comprehensive taxonomy is provided. This taxonomy will be used later to lead the problem definition and scope.

When the traffic assignment process is performed, assuming that the traversing entities have perfect knowledge of the network costs and actions of the other entities, and this information is known and fixed, then the problem can be regarded as a deterministic traffic assignment problem. On the other hand, when variability of the perceived network cost is allowed, the problem is referred to as a stochastic assignment problem.

Another variation of the traffic assignment problem arises from the nature of the demand. If the demand in the origin-destination pair remains constant regardless of the network congestion then the problem is said to have a fixed demand. In cases where the demand is modeled so that it varies with the network performance, then the problem is said to have an elastic demand.

The time horizon of the model depends on the type of decision it supports. For general planning decisions, static models or time-dependent models are the suitable choices. In such models, the field parameters are assumed to be steady for a sufficient period of time. For a more detailed analysis at the operational level, a dynamic traffic assignment model can be used.

One of the most common decisions of transportation networks is related to network design. In capacitated networks, this problem consists of setting values to the maximum flow that can traverse a specific arc (capacity). This type of problem is modeled in most cases as a continuous problem. The connectivity of a network or topology problem consists of selecting the best set of new arcs (including any new terminal nodes) to be added to the network. This topology problem is referred to as a discrete network design problem due to its combinatorial nature. Combinations of the problem characteristics previously mentioned gives origin to a series of modeling approaches that can be used to describe different network scenarios, as can be observed in Table 1. The total number of variations for the traffic assignment problem according to the criteria in Table 1 is $3 \times 2^8 = 768$. This work addresses both continuous and discrete design network problems (mixed network design problem) in a multimodal setting with deterministic fixed demand for a single class of users. The cost functions are modeled as asymmetric functions with a user-optimal rationality.

Table 1: Taxonomy of Traffic Assignment Problems

Modeling Component	Values
Rationality	System optimal
	User optimal
User's perception	Deterministic
	Stochastic
Time horizon	Static
	Time-Dependent
	Dynamic
Demand Behavior	Fixed
	Elastic
Demand Uncertainty	Deterministic
	Stochastic
Mode	Unimodal
	Multimodal
User	Single Class
	Multi-Class
Link Cost Functions	Symmetric
	Asymmetric
Design Objective	Discrete
	Continuous

2.6 Benchmark Network Problems

Several test or benchmark networks have been studied in the traffic assignment problem. The most cited networks are the Friesz-Harker network and the Sioux Falls network. Optimal solutions for capacity (continuous network design) and topology (discrete network design) design formulations are available. The Friesz-Haker (F-H) network is a 16-link network introduced in 1974 and has been used repeatedly throughout the years. The F-H network has been in use by Suwansirikul et al. [15], Friesz et al. [16], Meng et al. [17], Gao & Wu [10], and more recently by Wang et al. [18], Farvaresh et al. [19] and Luathep et al. [20].

Parameters for the base F-H network are presented in Table 2. These parameters along with capacity improvements are used to test the results of the proposed solution methodology for the mixed network design problem. The graph corresponding to the F-H network is presented in Figure 3. Network design problems with the F-H network have been tested under three demand scenarios for light, moderate, and high demand. The continuous network design problem has multiple solutions as reported by Luathep et al. [20]. Results for the continuous network design problem for the F-H network for moderate demand are presented in Table 3.

Table 2: Base Data for the Friesz-Harker Network

Arc Number	Source node	Target node	t_{oa}	B	K	P
1	1	2	1	10	3	4
2	1	3	2	5	10	4
3	2	1	3	3	9	4
4	2	3	4	20	4	4
5	2	4	5	50	3	4
6	3	1	2	20	2	4
7	3	2	1	10	1	4
8	3	5	1	1	10	4

Table 2 (continued)

Arc Number	Source node	Target node	t_{oa}	B	K	P
9	4	2	2	8	45	4
10	4	5	3	3	3	4
11	4	6	9	2	2	4
12	5	3	4	10	6	4
13	5	4	4	25	44	4
14	5	6	2	33	20	4
15	6	4	5	5	1	4
16	6	5	6	1	4.5	4

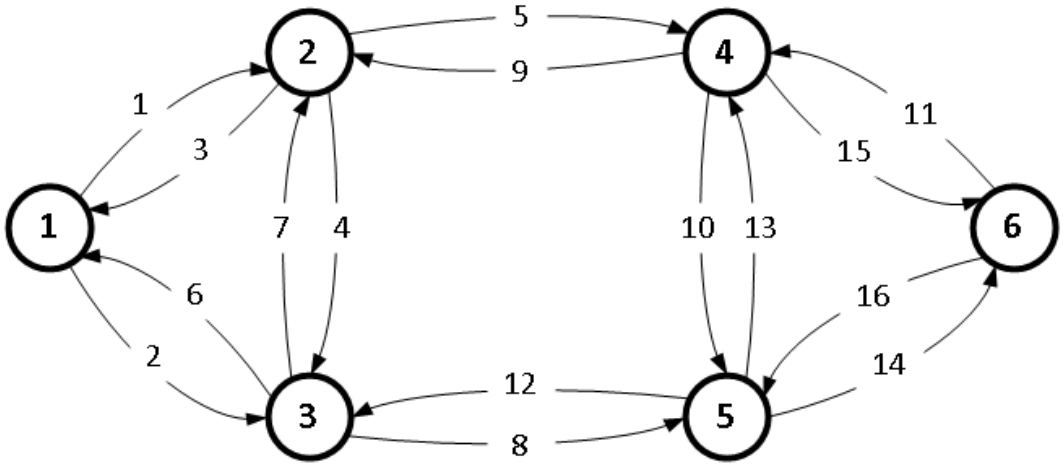

Figure 3: Friesz-Harker Network Graph

Table 3: Data for the Friesz-Harker Network for Moderate Demand

Arc Number	Source Node (i)	Target Node (j)	Initial Capacity (k_a)	Capacity Improvement (y_a)	Total Capacity (k_a+y_a)
1	1	2	3		
2	1	3	10		
3	2	1	9		
4	2	3	4		
5	2	4	3		
6	3	1	2	6.58	8.58
7	3	2	1		
8	3	5	10		
9	4	2	45		

Table 3 (continued)

Arc Number	Source Node (i)	Target Node (j)	Initial Capacity ($k_{a)}$	Capacity Improvement (y_a)	Total Capacity (k_a+y_a)
10	4	5	3		
11	4	6	2		
12	5	3	6		
13	5	4	44		
14	5	6	20		
15	6	4	1	7.01	8.01
16	6	5	4.5	0.22	0.472
Solver	**MINOS**			**Objective**	**211.25**

Another well-known network problem in the traffic assignment literature is the Sioux Falls (SF) network. Since its introduction by LeBlanc [21], the SF network has been used consistently as a benchmark for continuous and mixed network design problems. The data for the 24 nodes and 76 links comprising the SF base network are presented in Appendix A. Benchmarks for this problem for network design can be found in references [15], [16], [8] [10] , and more recently in refererecces [19], and [20]. The graph of the SF network is presented in Figure 4.

Two additional benchmark networks for transportation network design problems were introduced by Gao et al [10]. Gao's test network 1 (G1) is composed by 12 nodes and 17 links. The graph corresponding to network G1 is presented in Figure 5. The numbers in the arcs represent the free flow time. The dashed lines represent candidate arcs. The underlined number is the project number.

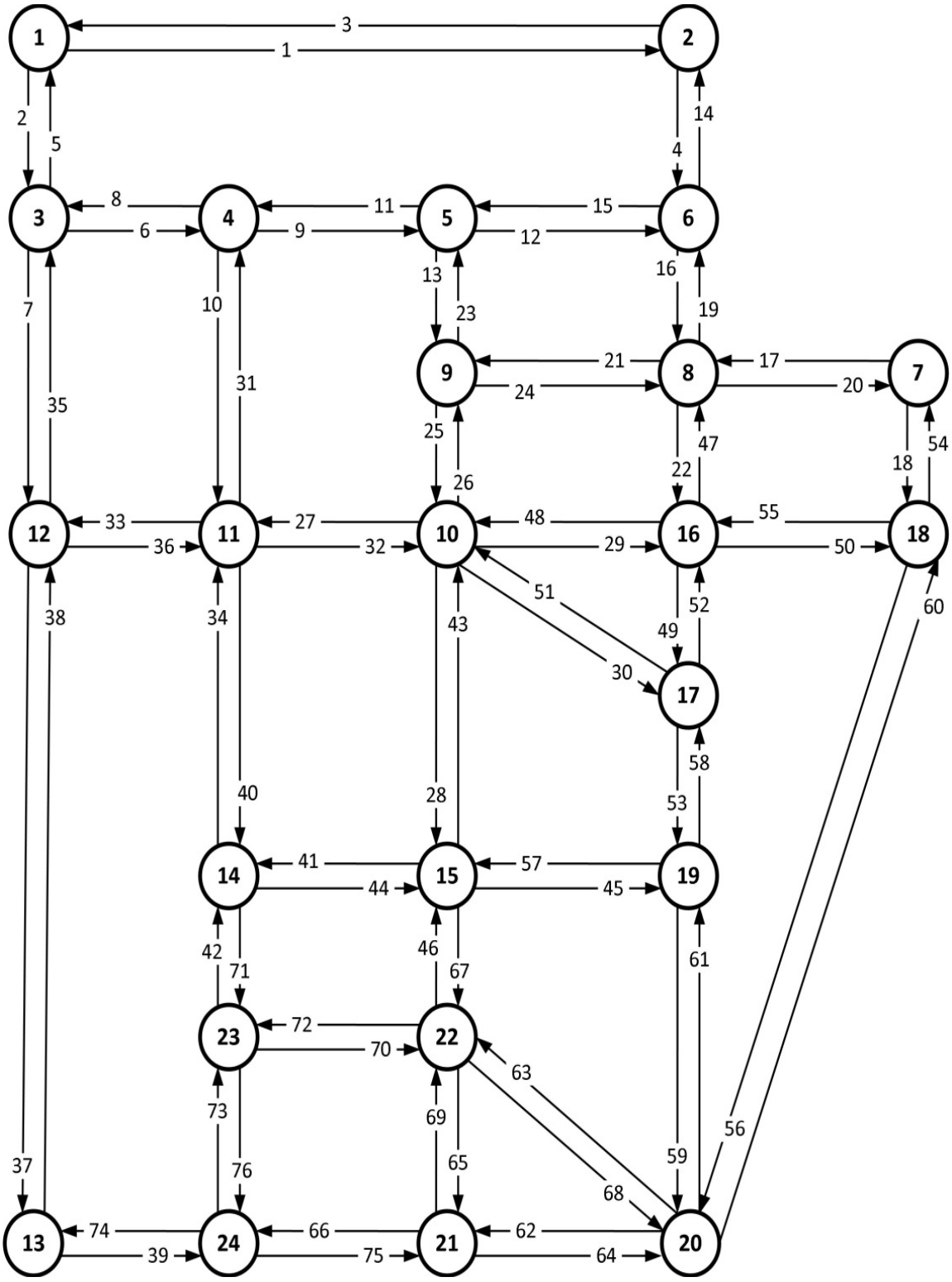

Figure 4: Sioux Falls Network Graph

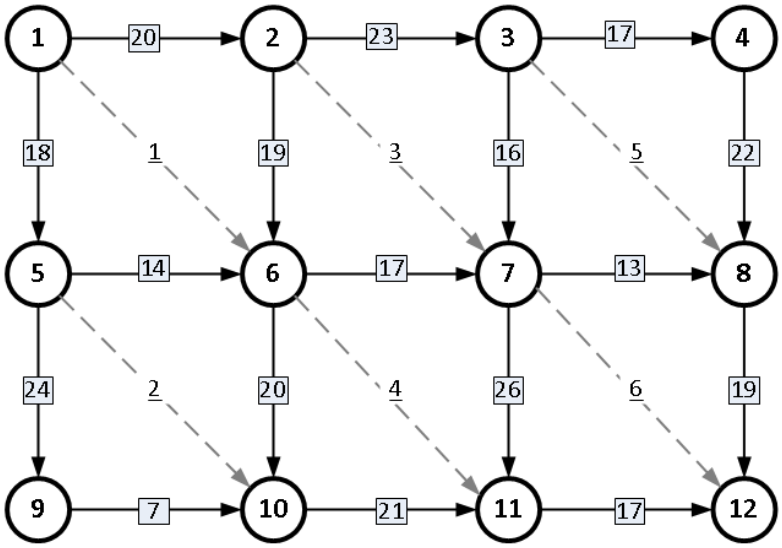

Figure 5: Gao's Test Network 1 (G1) Graph

Each improvement project is associated with a cost. The number of selected project depends on the available budget. The numerical parameters for problem G1 is presented in Table 4.

Table 4: Data for Network G1

	Project					
	1	**2**	**3**	**4**	**5**	**6**
(i, j)	(1,6)	(5,10)	(2,7)	(6,11)	(3,8)	(7,12)
t_{oij}	19	25	30	32	21	28
Project Cost	7	12	7	15	11	18

CHAPTER 3: MAX-AFFINE LINEARIZATION STRATEGY FOR CAPACITY MODELING IN NETWORK PROBLEMS

This chapter deals with arc capacity modeling in transportation networks. A convex piecewise linear fitting algorithm (least square partitioning algorithm) and its variants are introduced and tested. The proposed approaches can be used for linearization of capacity functions or it can be applied to fit piecewise linear functions to capacity measurements (raw data).

3.1 Sources of Non-Linearity

The main source of non-linearity is the link cost functions commonly represented by the Bureau of Public Roads (BPR) curve [7] . The BPR is expressed as follows:

$$t_a = t_{oa}[1 + B\left(\frac{f_a}{k_a}\right)^P] \tag{3.1}$$

where:

t_a: Average travel time on arc a

f_a: Traffic volume (flow) on arc a

t_{oa}: Free-flow travel time on arc a

k_a: Capacity of link a

B, p: Calibration parameters

An example of a BPR function with parameters 5, 1.5, 300, and 4 representing t_{oqa}, B, k_a, and p respectively is presented in Figure 6. The function represents a flow on a network arc such that when there is no congestion, the travel time (or travel cost)

associated is five units (e.g. minutes). As flow approaches the nominal capacity value (e.g. 300) the cost of traversing the arc starts to increase. For flow values exceeding the capacity, the rate of increase in cost becomes significantly larger. It is important to note that these functions act as penalty for unstable flows, but do not impose hard limits on the flow on an arc. This allows accounting for congestion effects for a realistic representation of network scenarios. The function can be calibrated with the parameters B and p.

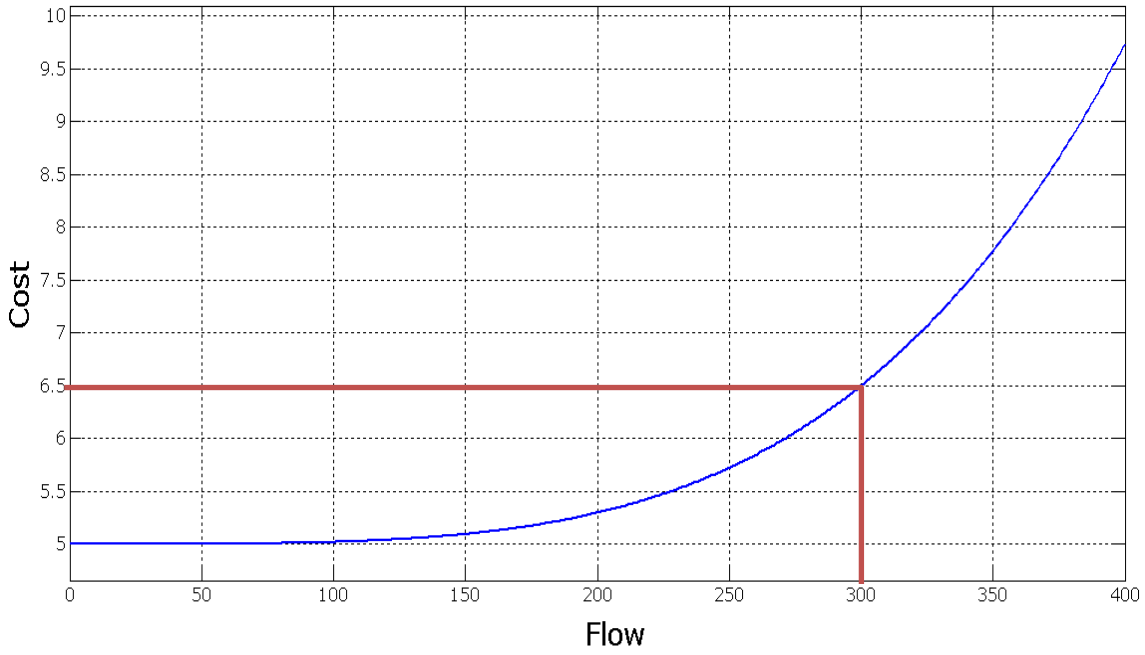

Figure 6: Example of the BPR Arc Cost Function

The BPR function is strictly monotonically increasing and convexing which helps in regards to the uniqueness of the solutions to the different network problems using this function [22]. Taking advantage of its convexity, the BPR function can be easily represented by a series of affine functions sorted by its slope in ascending order. The traffic assignment can be represented by such linear approximation and can be solved using standard linear programming solvers.

Another source of non-linearity in transportation problems arises from the modal-split procedure of the multimodal planning process. Marin & Garcia [14] proposed a multimodal network design model with the mode choice function as shown in 3.2.

$$p_w^m\left(u_\omega^a, u_\omega^b\right) = \frac{e^{-(\alpha^m + \beta\mu_\omega^m)}}{\sum_{m' \in \{a,b\}} e^{-(\alpha^{m'} + \beta\mu_\omega^{m'})}}, m \in \{a,b\}, \omega \in W \quad (3.2)$$

where:

$$g_\omega^m = p_w^m\left(u_\omega^a, u_\omega^b\right) g_\omega, \qquad m \in \{a,b\}, \omega \in W \quad (3.3)$$

The expression $\alpha^m + \beta\mu_\omega^m$ is referred to as the utility function of the mode m, ω represents a particular origin-destination pair, the set of all origin destination pairs is denoted by W, and g_ω represents the total demand for transportation in origin-destination pair ω. The modal split for network m at origin-destination pair ω is given g_ω^m.

The same function can be expressed as the cost differences (utilities) between modes $u_\omega = u_\omega^a - u_\omega^b$ obtaining the following simplified expression:

$$p_\omega^b(u_\omega) = \frac{1}{1 + e^{-(\alpha + \beta\mu)}}, \text{where } \alpha = \alpha^a - \alpha^b \quad (3.4)$$

The previous expression was linearized by the following polygonal function:

$$P(u) = \begin{cases} p_o + a_1(u - u_o) & u_o \le u < u_1 \\ p_1 + a_2(u - u_1) & u_1 \le u < u_2 \\ p_2 + a_3(u - u_2) & u_2 \le u < u_3 \\ \dots & \\ p_{n-1} + a_n(u - u_{n-1}) & u_{n-1} \le u < u_n \end{cases} \quad (3.5)$$

where:

$$a_i = \frac{p_i - p_{i-1}}{u_i - u_{i-1}}, i = 1,2, \dots n \quad (3.6)$$

This function can be modeled by introducing n continuous variables $\delta_1, \delta_2 \dots \delta_n$ and n binary variables $z_1, \dots z_n$ in $3n + 1$ constraints as presented in Equation 3.7.

The resulting model is a linear integer (binary) programming problem. Since this is required for each origin destination pair, a total of $|W|(3n+1)$ additional constraints and $|W|n$ binary variables and $|W|n$ linear variables have to be introduced in the formulation.

$$
\begin{aligned}
u &= u_o + \sum_{i=1}^{n} \delta_i \\
P(u) &= p_o + \sum_{i=1}^{n} a_i \delta_i \\
\Delta u_1 z_1 \le \delta_1 &\le \Delta u_1 \\
\Delta u_i z_i \le \delta_i &\le \Delta u_i z_{i-1}, \quad i = 2, \dots n-1 \\
0 \le \delta_n &\le \Delta u_n z_{n-1} \\
z_{i-1} &\le z_i \\
z_i &\in \{0,1\}
\end{aligned}
\tag{3.7}
$$

where:

$$\Delta u_i = u_i - u_{i-1}, \;\; i = 1, \dots, n$$

3.2 Linearization Techniques in Mathematical Programming

The most representative publications in linearization methodologies using binary piecewise representations can be found in references [23] and [24]. An alternative methodology based on the Maximum of Affine functions based on least squares partitions was developed by Magnani & Boyd [25]. The latter methodology provides an interval-free polygonal representation of a non-linear convex function. In this research, an adaptation of the ideas presented in Magnani & Boyd [25] will be applied to linearize the existing non-linear functions arising in transportation network design problems.

The problem of fitting an optimal polygonal representation of a non-linear function is a complex problem itself. An adequate and simple method of obtaining a linear representation is required to achieve the desired level of accuracy without significantly increasing the overall computational complexity.

3.3 Least-Squares Partitioning Algorithm (LPA)

The solution strategy adopted in this work for the continuous network design problem is based on a series of reformulations. First, the original non-linear bi-level program was transformed into a single-level non-linear program with equilibrium constraints. Next, the problem was transformed into a non-linear binary problem by substituting the equilibrium constraints for their equivalent formulation via binary variables (MIP). The remaining non-linear elements are related to the use of the BPR function to model arc congestion. This section deals with the linearization of such elements by means of a series of max-affine functions.

Several approaches have been proposed in recent literature to transform the bi-level, non-linear transportation network design problem into a mixed-problem. The complexity of the resulting linearization depends on the type of decision being modeled. For example, [18]modeled capacity decisions for fixed-topology network. The flow-capacity surface was approximated via mixed integer formulation adding additional binary variables and y constraints to the model. Similarly, [26] adopted a modeling approach to accommodate topology and capacity improvement for a transportation network. In their work, the flow-capacity surface was approximated via specially ordered sets, or SOS variables. The resulting model was too complex for conventional solvers and a relaxed version was used instead. Farvaresh, etal. [19] analyzed the mixed network design problem by using topology improvements with fixed capacity. In their approach they exploit the unimodularity of the formulation resulting from the linearization technique proposed by Padberg [23], and Bazaraa [2] by assuming fixed capacities for link improvements. In this work, an interval-free approach to model the flow-capacity surface is utilized. The technique was first presented by Magnani & Boyd [25]. The objective of the methodology was to fit piecewise linear approximations for convex (or

near convex) datasets. Since the methodology is a curve fitting procedure, it does not require strict properties of the underlying datasets. The approach is valid for multidimensional functions as long as the datasets have a reasonably convex form.

The procedure starts with a non-linear function H being approximated at the points $\dot{z}_i$ $(i = 1,..N)$ by a linear function. The domain of H is divided into k intervals with each one them having one linear approximation. The linear approximations are defined by the formula $\boldsymbol{a}_j^T + b_j$ where $\boldsymbol{a}_j^T$ is a row vector of n elements (number of independent variables), and b_j is the intercept for the j^{th} partition. The objective is to minimize the square of the deviations with respect to the original points for all the intervals as follows:

$$Min\ \Psi = \sum_{i=1}^{N} \left(\max_{j=1,\dots,k} \left(\boldsymbol{a}_j^T u_i + \beta_j \right) - \dot{z}_i \right)^2 \tag{3.8}$$

The decision variables for this optimization problem are the line coefficients and intercepts given by $a_1, a_2, \dots a_k \in R^n$ and $b_1, b_2, \dots, b_k \in R$respectively. The algorithm has two major alternating steps consisting of data partitioning and least-squares fitting to update the coefficients. In this particular application, the BPR function is first discretized to the desired resolution (e.g. 1,000 data points) and the resulting set of data points are indexed, i.e. $(1,..N)$. Let $G_j^{(v)}$ be the j^{th} partition of the data points at the v^{th} iteration i.e. $G_j^{(v)} \subseteq \{1, \dots N\}$ with the following conditions:

$$G_i^{(v)} \cap G_j^{(v)} = \emptyset\ for\ i \neq j \tag{3.9}$$

$$\bigcup_j G_j^{(v)} = \{1, \dots, m\} \tag{3.10}$$

At iteration$v$, the corresponding coefficients of the linear approximation for partition j are $\boldsymbol{a}_j^{(v)}$ and $\beta_j^{(v)}$. The main step for the algorithm consists of generating the next

values $\boldsymbol{a}_j^{(v+1)}$ and $\beta_j^{(v+1)}$ from the current partition $G_j^{(v)}$ using least-squares fitting. The algorithm is summarized in Figure 7.

Step 0:	Obtain and initial partition set $\boldsymbol{G}$ Set number of iterations $v = 0$
Step 1:	Update Coefficients For each partition $j = 1, \dots k$: Calculate linear approximation coefficients $\boldsymbol{a}_j^{(v)}$,$b_j^{(v)}$ Next partition j
Step 2:	Update partitions, assign point i to interval $G_j^{(v+1)}$ as follows: $\Psi^j(\dot{z}_i) = \max_{s=1,..,k}(\boldsymbol{a}_s^{(v)T}\dot{z}_i + \beta_s^{(v)})$, point i is added to interval $j = s$ $v = v + 1$
Step 3:	If $\boldsymbol{G}^{(v)} = \boldsymbol{G}^{(v+1)}$ or $v = max\ iterations$ then end , else goto step 1

Figure 7: Summary of Least-Squares Partition Algorithm

LSPA has several advantages, such as speed and simplicity which facilitates its implementation. Magnani & Boyd [25] recommended several trials with random seeds for the initial solution and selecting the best fit. While this is a good approach for a strict piecewise linear fitting problem, for a transportation network design problem the true performance of the fitting problem is only known after the linearized version problem is solved and compared to the solution of the non-linear version. The problem of fitting the best piecewise linear function to minimize the error between the linearized NDP and the non-linear NDP is itself a bi-level programming problem.

An example of the MB algorithm for piecewise linear fitting for the univariate case is presented in Figure 8. The example shows a five-piece approximation of the BPR function in equation 3.1. At iteration one, most of the lines were set to the same region, thus providing a poor approximation. As the algorithm evolves (iteration 5) the linear functions are reorganized so that the best estimate at flow x of the non-linear function (BPR) is the maximum of all the linear pieces evaluated at x. For the univariate case the algorithm converges after a few iterations.

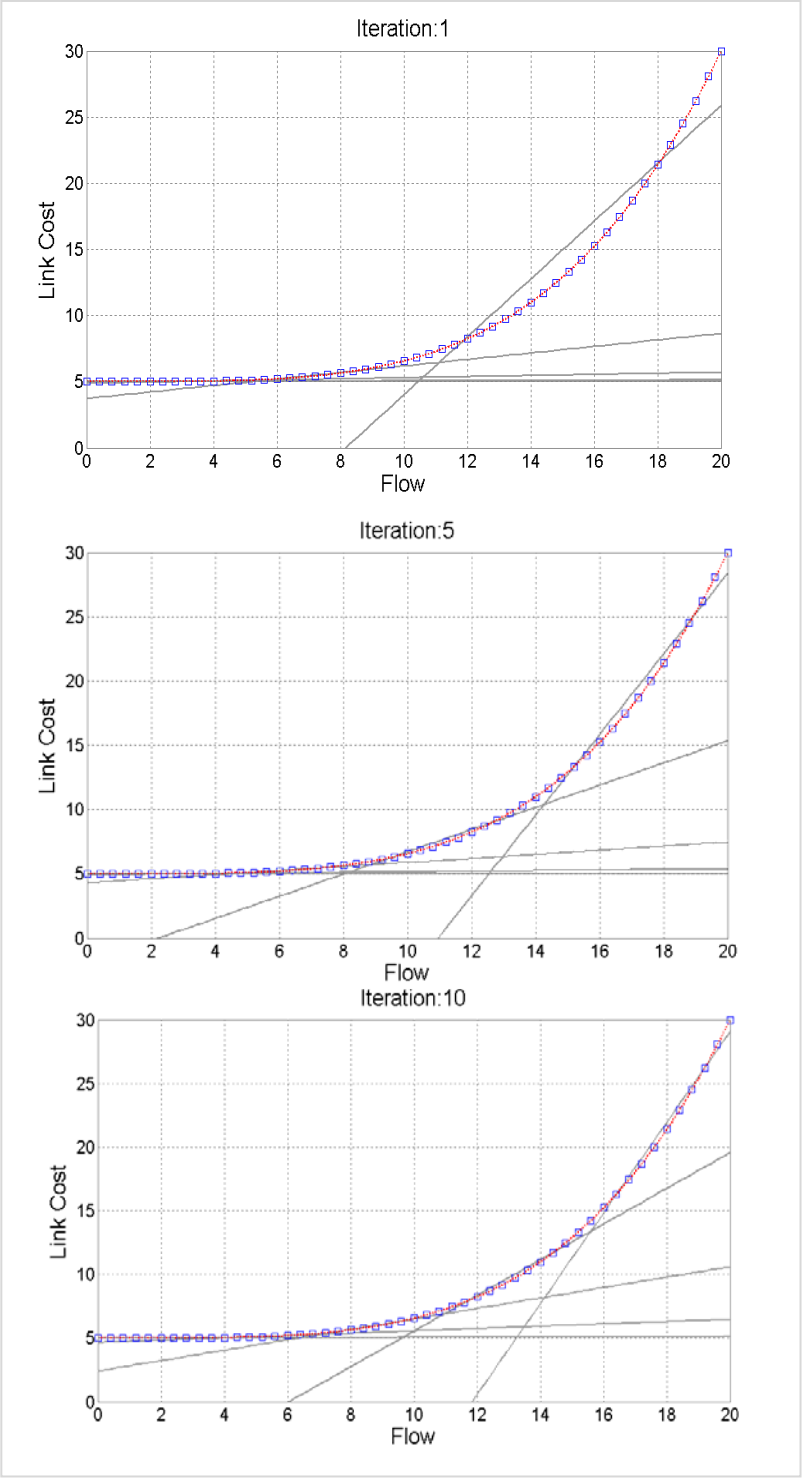

Figure 8: Least-Squares Partition Algorithm for the Univariate Case

The application of the MB algorithm to the bivariate case using a ten-plane linear fitting is presented in Figure 8. The upper left plot presents the discretization step of the

flow-capacity surface. At iteration one, the fitting is poor due to the concentration of the fitting planes in a confined region of the flow-capacity domain. As the algorithm evolves, it converges at iteration 115 to the solution presented in the lower left plot. The resulting linear approximation resulting of evaluating the maximum of all the planes at each point of the flow capacity domain is depicted in the lower right plot. The measure of error are the coefficient of determination or R^2 and the root mean square (RMS).

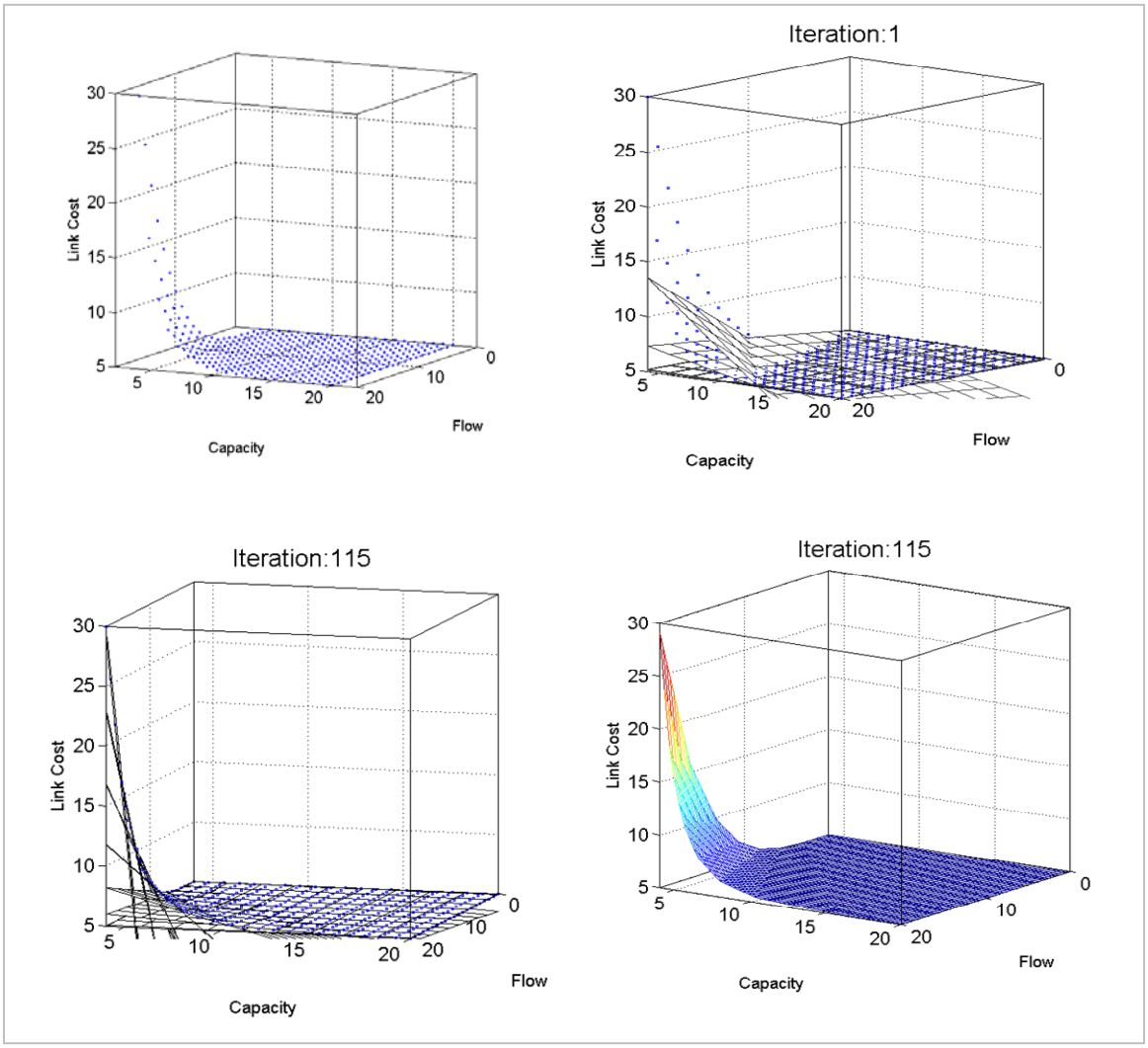

Figure 9: Least Squares Partition Algorithm for the Bivariate Case

The coefficient of determination or r-square (R^2) is a measure of the goodness of fit in a regression model. The calculation of R^2 starts by measuring the amount of variation of the dependent variable with respect to its average. This value is denoted as

the total sum of squares (SST). The variation with respect to a proposed model is calculated and it is denoted as SSM. The coefficient of determination is calculated as the percentage of the total variation explained by the proposed model. In addition to the coefficient of determination the root mean square (RMS) was used to assess the goodness of fit of the linear approximation. The expressions for R^2 and RMS are presented in equations 3.11 and 3.12 below.

$$SST = \sum_i (y_i - \bar{y})^2$$
$$SSM = \sum_i (y_i - \hat{y}_i)^2 \quad (3.11)$$
$$R^2 = 1 - \frac{SSM}{SST}$$

$$RMS = \sqrt{\frac{SST}{m}} \quad (3.12)$$

where:

y_i: Observation i

$\bar{y}$: Average of dependent variable

$\hat{y}_i$: Estimation of dependent variable at point i

As the number of intervals used to approximate the nonlinear function is increased, the goodness of fit measure increases. Several experiments varying the number of intervals were conducted. Examples of the result of such experiments for r-square and RMS are presented in Figures 10 and 11 respectively.

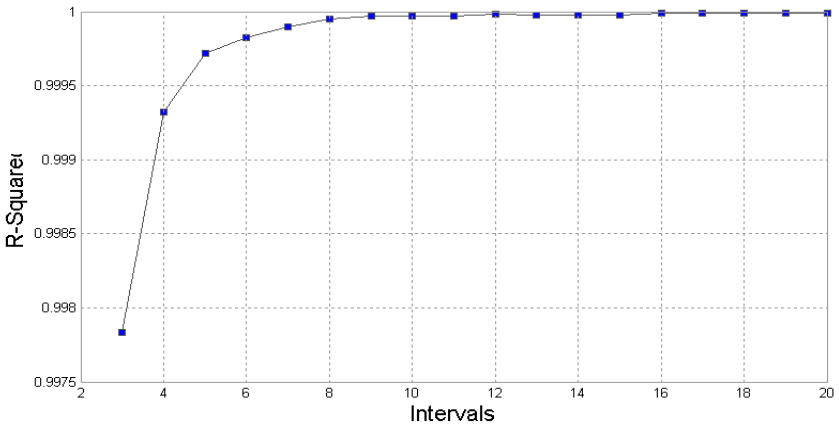

Figure 10: R-Square and Number of Intervals for the Linear Approximation Procedure

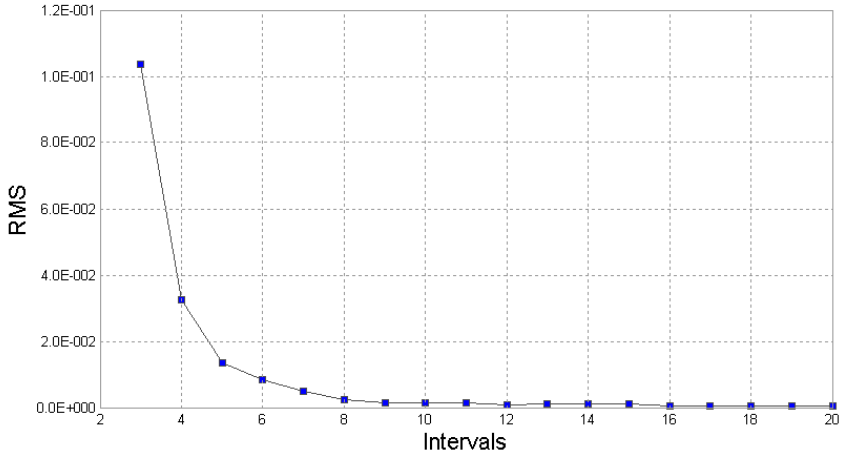

Figure 11: RMS and Number of Intervals for the Linear Approximation Procedure

3.4 Modified Least Square Partitioning Algorithm (MLSPA)

The capacity-flow surface can be partitioned in smaller subsets for a more accurate linear approximation. Luathep [20] divided the surface following a grid-type scheme. A similar approach was followed in Wang [18]. The main disadvantage of the grid partition approach is that it does not take into consideration the effect high flow-to-capacity ratios. The situation when flows equals capacity on an arc is referred to as a saturated condition and it can be represented as $f_a = (k_a + y_{a)}$. The cases in which the

arc flow f_a is less than the capacity of the arc are denoted as undersaturated conditions. Oversaturation occurs when the flow exceeds the capacity of the arc. A series of plots showing the concept of under and oversaturation are presented in Figure 12.

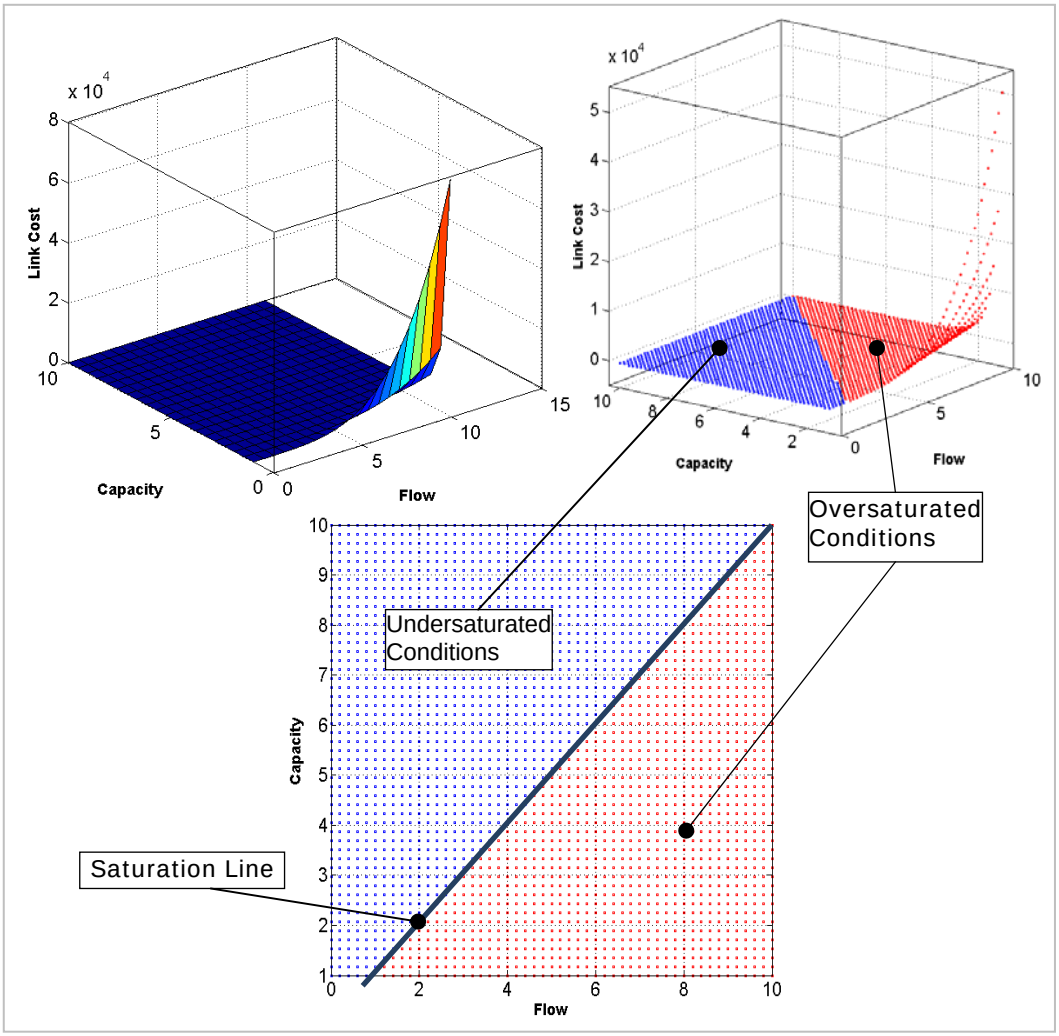

Figure 12: Under and Oversaturated Traffic Regions in the Flow-Capacity Surface

In Figure 12, it can be observed that most of the variation of the flow-capacity surface is concentrated in the oversaturated region. In a network with a moderate level of congestion the traffic tends to distribute among the different paths (divert) before assuming flow values in the oversaturated region. Optimal flows will tend to occur in the under saturated conditions. For that reason, it is reasonable to obtain a better function

representation for the approximation in the undersaturated region of the flow-capacity surface.

The modified least squares partitioning algorithm (MLPA) was devised to give different priorities to the under and oversaturated regions of the flow-capacity surface. First the algorithm acts upon the points of the undersaturated region ($f_a \leq y_a$) with a partition denoted by Gu, this set contains g_u partitions. The LSPA is applied to the set Gu. The same procedure is applied for the oversaturated conditions with a set denoted by Go. The final approximation is the union of the linear approximations of the two sets (see Figure 13 for summary).

Step 0:	Split original data into under and over saturated sets $\boldsymbol{Gu, Go}$
Step 1:	Apply LSPA to $\boldsymbol{Gu, Go}$ Obtain linear approximation ($\boldsymbol{a}_s^{(gu)T}\dot{z}_i + b_s^{gu}$) and ($\boldsymbol{a}_s^{(go)T}\dot{z}_i + b_s^{(go)}$) respectively
Step 2:	Construct final approximation set L as $L: \bigcup_{v \in \{Gu, Go\}} (\boldsymbol{a}_s^{(v)T}\dot{z}_i + b_s^{(v)})$

Figure 13: Modified Least Square Partitioning Algorithm (MLSPA) for the Flow-capacity Surface

The objective of the MLSPA is to obtain a better approximation, not in the goodness of fit of the function, but on the results of the optimization model. For instance, the fitting errors are reduced if a good fit is obtained in the oversaturated region. However, that fit is not useful in most of the cases since the optimization model will avoid the oversaturated region. Approximation tests based on this rationale will be presented in the following sections.

3.5 Max-Affine Formulation of TAP

The previous sections described the procedure to obtain a linear approximation of a convex function using a series of affine functions with a pointwise maximum

procedure. This approximation can be used to linearize a function without using binary variables. Let L_a be the linear approximation for arc a. The number of functions for L_a is given by G_a. The constant term of linear approximation L_a is denoted by α_a. Each linear term may have one or more variables. For ordinary arcs only one, the flow, is used as a predictor. For capacity improvement the arc cost is approximated using both, arc flow and arc capacity. Let $\mathcal{V}$ be the set of variables used for the linear approximation, the approximation of t_a is given by expression 3.13.

$$L_a \geq \alpha_{ag} + \sum_{v \in V} \beta_{vag} \;\; \forall\, g \in G_a \; \forall\, a \in \mathcal{A} \tag{3.13}$$

3.5.1 Arc-Path Linear Formulation of TAP

Based on the max-affine approximation 3.13 the arc-path formulation of TAP can be re-written as a pure linear program. The number of additional constraints depends depend on the number of intervals used to approximate the arc-cost function. For example if five functions are used ($G_a = 5\,, \forall\, a \in \mathcal{A}$) then there will be $|\mathcal{A}| \times 5$ new constraints and $|\mathcal{A}|$ number of linear variables. The arc-path formulation is presented by equations 3.14 through 3.17

$$Min\, T(f) = \sum_{a \in \mathcal{A}} V_a \tag{3.14}$$

$$V_a \geq \alpha_{ag} + \beta_{ag} f_a, \qquad \forall\, g \in G_a \; \forall\, a \in \mathcal{A} \tag{3.15}$$

$$\sum_{r \in R_{pq}} h_{pqr} = d_{pq}\,, \forall\, (p,q) \in \mathcal{W} \tag{3.16}$$

$$\sum_{(p,q) \in \mathcal{W}} \sum_{r \in R_{pq}} \delta_{pqra} h_{pqr} = f_a, \forall\, a \in \mathcal{A} \tag{3.17}$$

Note that in this case, the linear approximation depends only on the link flow (f_a). For capacity decisions the link cost will depend on the added capacity (y_a)

3.5.2 Arc-Node Linear Formulation of TAP

In a similar manner the arc-node formulation of TAP can be linearized with the inclusion of the linear approximation V_a. The original problem consists of a non-linear objective function optimized over linear feasible region. After linearizing the objective the resulting problem is a pure linear problem denoted by TAP-LP.

$$Min\, T(f) = \sum_{a \in \mathcal{A}} V_a \tag{3.18}$$

$$V_a \geq \alpha_{ag} + \beta_{ag} f_a, \qquad \forall\, g \in G_a\ \forall\, a \in \mathcal{A} \tag{3.19}$$

$$\sum_{\psi^-(i)} f_{ijpq} - \sum_{\psi^+(i)} f_{jipq} = d_{ijpq}\ , \forall\, i \in \mathcal{N}, \forall\, (p,q) \in \mathcal{W} \tag{3.20}$$

$$f_a = \sum_{a \in \mathcal{A}} \sum_{(p,q) \in \mathcal{W}} f_{apq}\ , \forall\, a \in \mathcal{A} \tag{3.21}$$

$$f_{apq} \geq 0\ \forall\, (i,j) \in \mathcal{A}, \forall\, (p,q) \in \mathcal{W} \tag{3.22}$$

3.6 Linearization Tests

For the linearization method implementation, several fitting parameters were tested. The main fitting parameters used to fine tune the algorithm are listed:

- Fitting method: Two fitting methods were considered, LSPA and MLSPA.
- Number of functions: The number of intervals or linear functions used to approximate the linear function where tested values of 5, 10, 20, and 30 were used.
- Function Distribution: This parameter was used in the MLSPA method only. It corresponds to the number of linear functions in the undersaturated region. For instance, a value of 0.6 with ten linear functions means that six functions will be used in the undersaturated region of the flow-capacity surface and four on the oversaturated region.

- Saturation Factor: This parameter was used in the MLSPA method only. This factor was used to vary the saturation line to provide a more flexible fitting process allowing the fitting process to extend either the oversaturated or the undersaturated regions of the flow-capacity surface.

For the LSPA method there are four levels of experimentation. For the MLSPA method there were 4x4x5=80 experiments. There were a total of 84 fitting model scenarios. The scenarios were tested with the three demand scenarios of the Friesz-Harker network. Each scenario was run for three replicates. The goodness of fit was tested by the R^2 and the RMS measures. An additional goodness of fit measure was tested by running a linearized version of the problem, constrained to a known solution for each scenario. The measure of the quality of the fitting process was based on how well the linearized model solution resembled that of the non-linear model. The fitting parameters for experimentation are summarized in Table 5.

Table 5: Fitting Parameter for Experimentation

Parameter	Values
Number of Functions (NF)	5,10,20,30
Fitting Method (FM)	LSPA, MLSPA
MLSPA- Function Distribution(FD)	0.4 0.5 0.6 0.7
MLSPA- Saturation Factor (SF)	0.8, 1.0, 1.2

Figure 14 presents an example with ten functions using a distribution of 0.6 (6 functions) for the undersaturated region and 0.4 (4 functions) for the oversaturated region. The different saturation limits based on the saturation factor are presented in Figure 14. A saturation factor of 0.8 has the effect of decreasing the slope of the saturation line, isolating the most congested section of the flow-capacity surface.

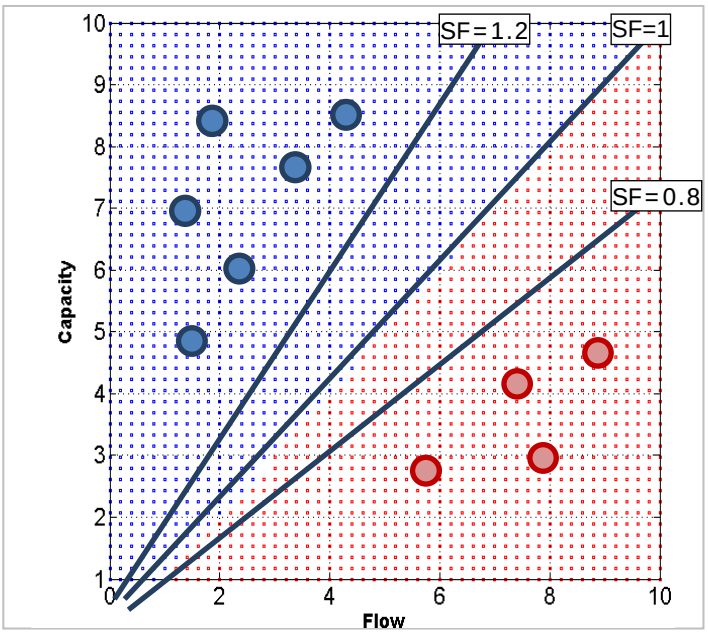

Figure 14: Flow-Capacity Fitting Parameters

In addition to the fitting parameters, the goodness of fit for any flow-capacity surface depends on the relationship between the current capacity with respect to the maximum capacity that could be potentially added to the arc. Let k_a be the current capacity for arc a and let u_a be the upper bound in the capacity increase. The potential maximum capacity is defined as $k_a + u_a$. The quality of the goodness of fit is related to the capacity ratio $k_a/(k_a + u_a)$. Arcs with low capacity ratios are very likely candidates to become system bottlenecks and could be selected for capacity improvement. The quality of the solution based on the approximated flow-capacity surface is also dependent on the goodness of fit on such critical links especially in the undersaturated region.

To verify this hypothesis an upper bound of ten capacity units was assumed for the Friesz-Harker network. The results for five functions using the LSPA method are presented in Figure 15.

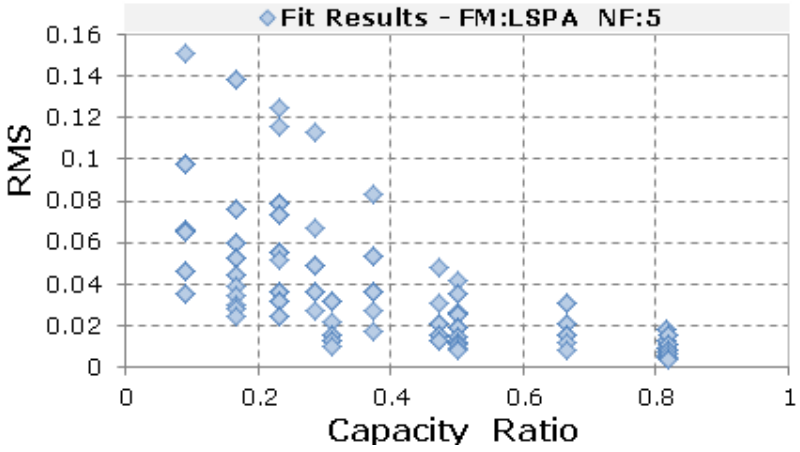

Figure 15: Fit Results and Capacity Ratio for the Friesz-Harker Network Using LSPA and 5 Functions

Figure 15 shows the hypothesized behavior of the goodness of fit results with respect to the capacity ratio. The overall goodness of fit is moderate and there is great variability in the performance at low capacity ratios. This can be compared with Figure 16 which shows the results for the same fitting method using ten functions. It can be observed that there was an overall reduction of the variability and an improved goodness of fit. Additional results using 20 and 30 functions are presented in Figure 17 and Figure 18 respectively.

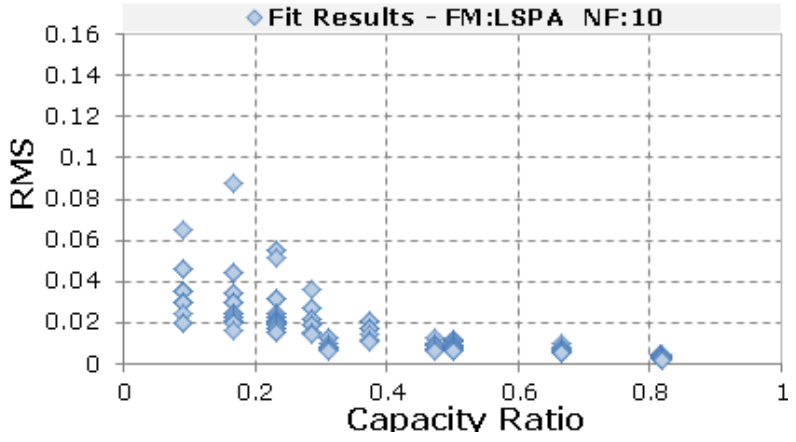

Figure 16: Fit Results and Capacity Ratio for the Friesz-Harker Network Using LSPA and 5 Functions

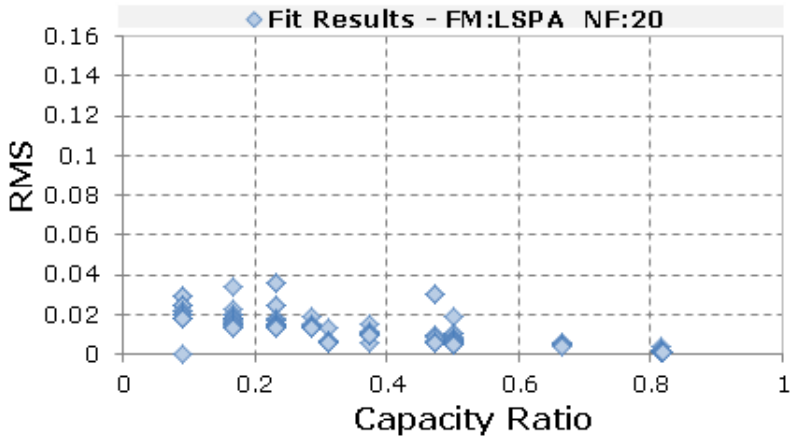

Figure 17: Fit Results and Capacity Ratio for the Friesz-Harker Network Using LSPA and 5 Functions

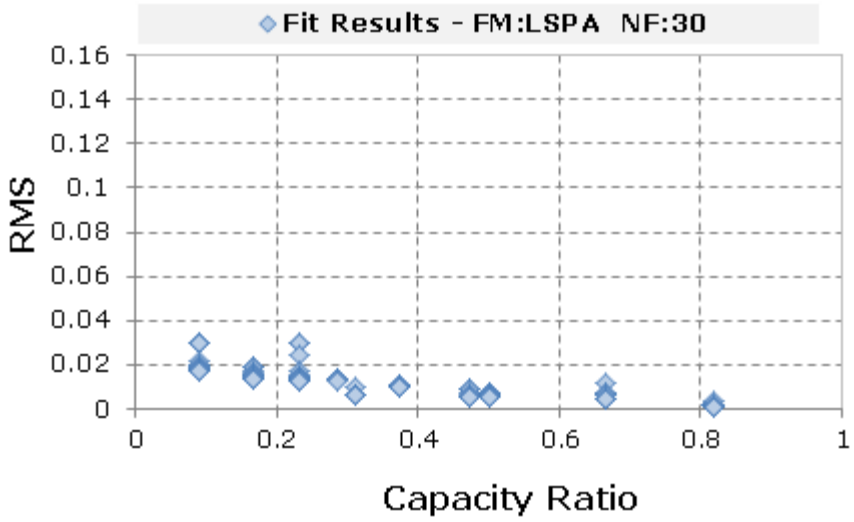

Figure 18: Fit Results and Capacity Ratio for the Friesz-Harker Network Using LSPA and 5 Functions

Results using the MLSPA and five functions are presented in Figure 19. The saturation factor was varied from 0.9 to 1.1 to test the sensitivity of the fitting method.

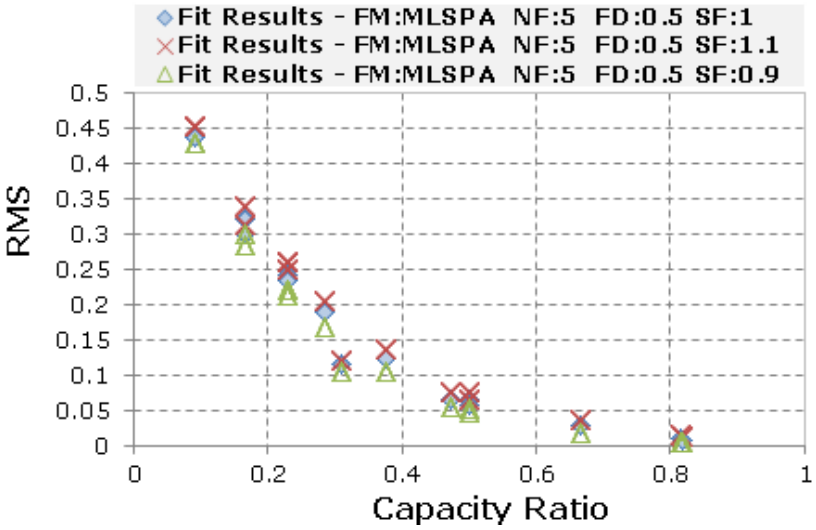

Figure 19: Fit Results and Capacity Ratio for the Friesz-Harker Network Using MLSPA, 5 Functions and Function Distribution 0.5

The MLSPA performs poorly for a low number of functions. However, when the number of functions was increased to ten, the performance is comparable to that of the LSPA with the additional benefit of reduced variability (see Figure 20). The final performance of both methods will be compared based on the quality of the optimization results. Similar results for 20 and 30-function linear piecewise fitting can be found in Appendix B.

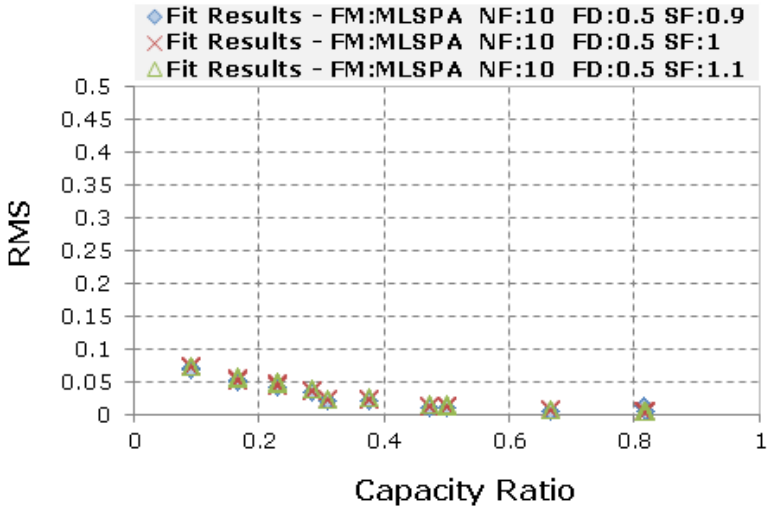

Figure 20: Fit Results and Capacity Ratio for the Friesz-Harker Network Using MLSPA, 10 Functions and Function Distribution 0.5

CHAPTER 4: BI-LEVEL OPTIMIZATION PROBLEMS IN TRANSPORTATION

In this section, the general bi-level optimization problem and its applications in transportations are introduced. Several bi-level optimization models for the continuous and discrete transportation network design problems are formulated and explained.

4.1 General Bi-Level Optimization Problem

In this section, the general bi-level programming problem is presented. The notations and terminology are adapted from those in Bard [27]. Bi-level programming models are motivated by leader-follower games called Stackelberg games. In bi-level programming, the set of decision variables is divided into two vectors x and y, with x representing the set of decision of the leader (upper level problem) and y representing the decisions of the follower (lower level problem).

The problem definition will be given for the general linear bi-level problem, further definitions for the discrete and nonlinear cases will be provided when appropriate. Let $x \in X \subset R^n$, $y \in Y \subset R^m$, $F: X \times Y \to R^1$ be the decision variables and objective function for the upper level problem. Let $f: X \times Y \to R^1$ be the objective function for the lower level problem. The linear bi-level programming problem (LL-BLPP) is shown in 4.1 through 4.5 below.

$$\min_{x \in X} F(x, y) = c_1 x + d_1 y \tag{4.1}$$

subject to:

$$A_1 x + B_1 y \leq b_1 \tag{4.2}$$

$$\min_{y \in Y} f(x, y) = c_2 x + d_2 y \tag{4.3}$$

$$A_2 x + B_2 \leq b_2 \tag{4.4}$$

where:

$$\begin{array}{llll} c_1, c_2 \in R^n & b_1 \in R^p & A_1 \in R^{p\times q} & B_1 \in R^{p\times m} \\ d_1, d_2 \in R^m & b_2 \in R^q & A_2 \in R^{q\times n} & B_2 \in R^{q\times m} \end{array} \quad (4.5)$$

Non-negativity constraints and bounds on the decision variables can be included in sets X and Y. Equations 4.1 to 4.5 represent the general bi-level linear programming problem. The objective function of the upper problem is shown in 4.1. Only x is in control of the upper level problem (leader) while the variable y is the decision variable of the lower level problem (followers' reaction), conditioned on the value of the decision variables of the upper level problem. Constraint group 4.2 represents the feasible region for the upper level problem. The objective function of the follower is represented by 4.3. It can be observed that once the leader has made a decision (x) the corresponding term in 4.3 becomes a constant and can be excluded from the objective function. Constraint set 4.59 corresponds to the feasible set of the lower level problem. Let S be the overall constraint region for the BLPP formed by the all the decision variable vectors defined as:

$$S = \{(x, y): x \in X, y \in Y\ , A_1x + B_1y \leq b_1, A_2x + B_2y \leq b_2\} \quad (4.6)$$

In order for the BLPP to be solved, the first assumption over S is that it should be non-empty and compact (i.e. bounded and closed).

In the problem logic, the upper level (leader) decides first, that is selects an x, conditioning the response of the lower level problem (follower). The feasible set for the follower for a fixed $x \in X$ is denoted by $S(x)$ and is given by:

$$S(x) = \{y \in Y : B_2y \leq b_2 - A_2x\} \quad (4.7)$$

The follower reacts to the decisions imposed by the leader by choosing a response, y. The response of the follower should be feasible not only on its own feasible

set, but in the joint feasible set for the leader. That is the projection of $S(x)$ onto the decision space of the leader. The set is denoted by $S(X)$ and is given by:

$$S(X) = \{x \in X : \exists\, y \in Y, A_1x + B_1y \leq b_1, A_2x + B_2y \leq b_2\} \quad (4.8)$$

In equation 4.8 it can be highlighted that $S(X)$ requires the existence of a feasible follower response y for both problems. The follower acts, seeking its own benefit, by selecting the best y for each x the leader select. Therefore the follower has a rational reaction set for $x \in S(X)$ denoted by $P(x)$ given by:

$$P(x) = \{y \in Y : y \in argmin\,[f(x, \hat{y}) : \hat{y} \in S(x)]\} \quad (4.9)$$

The set of leader choices and rational reactions of the follower receives the name of inducible region (IR) in the bi-level programming theory and is expressed as:

$$IR = \{(x, y) : (x.y) \in S, y \in P(x)\} \quad (4.10)$$

The logical assumption over the rational set of the follower is that it is non-empty (i.e. $P(x) \neq \emptyset$). In other words, the follower has some space to respond. The follower optimize its response over $P(x)$ and the leader over IR. Since IR contains the reaction of the follower then the BLPP can be expressed as follows:

$$\min\{F(x, y) : (x, y) \in IR\} \quad (4.11)$$

Another critical assumption for BLPP to be solvable is that $P(x)$ should be a point-to-point map of X onto Y. In other words, for each $x \in X$ the optimal solution of the lower level problem is unique. To illustrate these concepts the following example of a linear bi-level program should be considered:

$$\min_{x \in X} F(x, y) = x - 8y \quad (4.12)$$

subject to:

$$\min_{y \in Y} f(y) = y \tag{4.13}$$

$$-5x + 3y \leq 4 \tag{4.14}$$

$$x + 2y \leq 20 \tag{4.15}$$

$$4x - 5y \leq 2 \tag{4.16}$$

$$-x - 2y \leq -7 \tag{4.17}$$

The example problem has a simplified structure for illustration purposes. The feasible region S is represented by the shaded area in Figure 21.

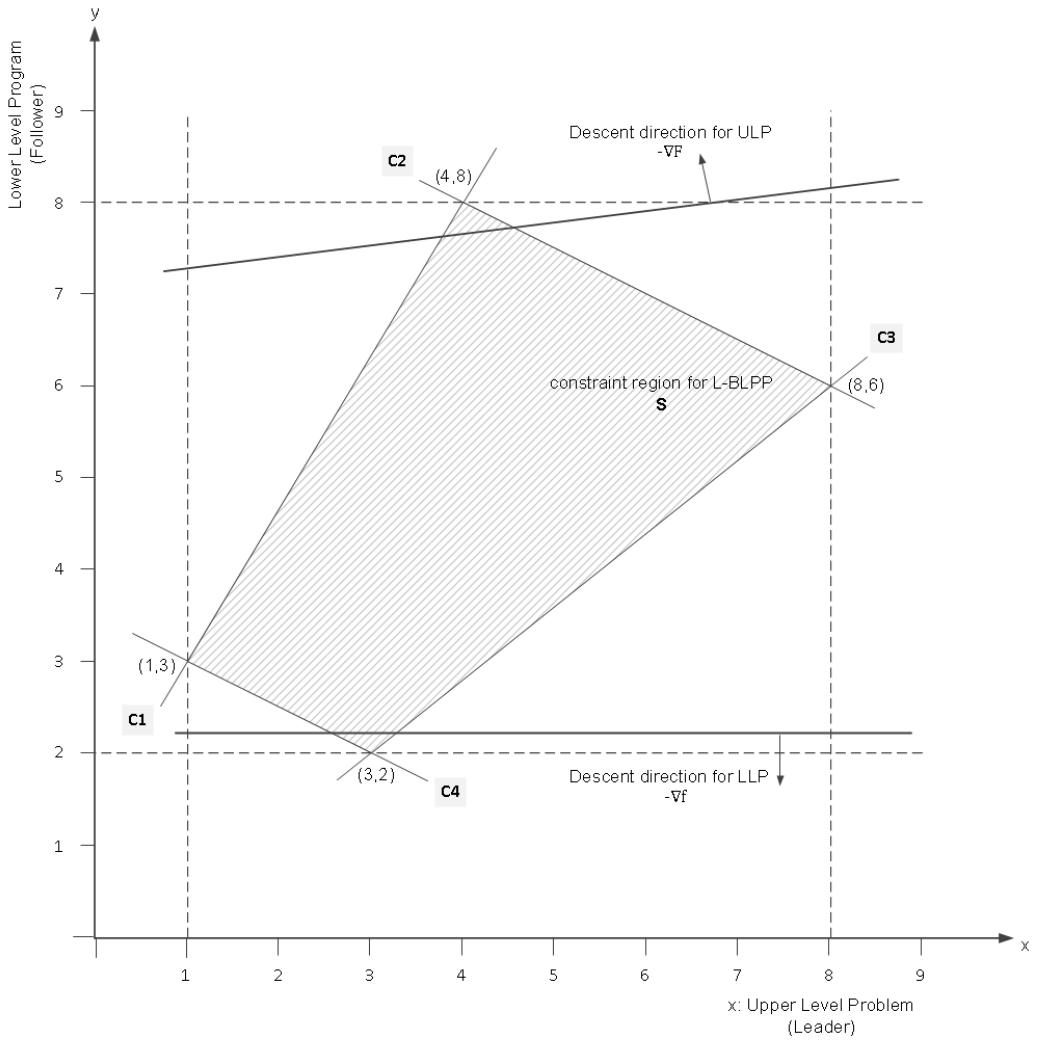

Figure 21: Feasible Region of Example L-BLPP

The x axis represents the upper level problem (UPL) or leader's problem. The y axis represents the lower level problem (LLP) or follower's problem. The descent directions for both objective functions are also depicted in Figure 21. It can be observed that the best strategy for the leader in this case is point (4,8) assuming total cooperation of the follower. Similarly, for the follower the best strategy is located in (3,2) without considering the leader's objective function. Since each of agents involved seek their own benefit, the best individual solutions for both problems will not coincide. In this case, the leader has certain degree of control over the environment and selects its strategy (x) first. The follower then prepares its reaction over the feasible set created by x. That feasible set is denoted by $S(x)$ and corresponds to the vertical line inside the feasible region where x is fixed. Since the follower minimizes its objective function based on x, the follower's optimal reaction in this case is the point in $S(x)$ where f is minimized (see Figure 22).

The collection of all points where the reaction of the follower is optimized for each decision of the leader is called the rational reaction set $P(x)$ and is represented in Figure 22 by the thick lines in the feasible region.

The induced region is defined $IR = \{(x, y): (x. y) \in S, y \in P(x)\}$ and depicted in Figure 23. It can be observed that it only suffices with optimizing over the inducible region to reach the optimum solution of the example L-BLPP. The optimum value of the example L-BLPP is the point (8,6) with a leader's objective of -40, whereas the follower's objective is six. If the leader attempts to improve its objective by selecting a different strategy its objective will be suboptimal due to the follower's reaction.

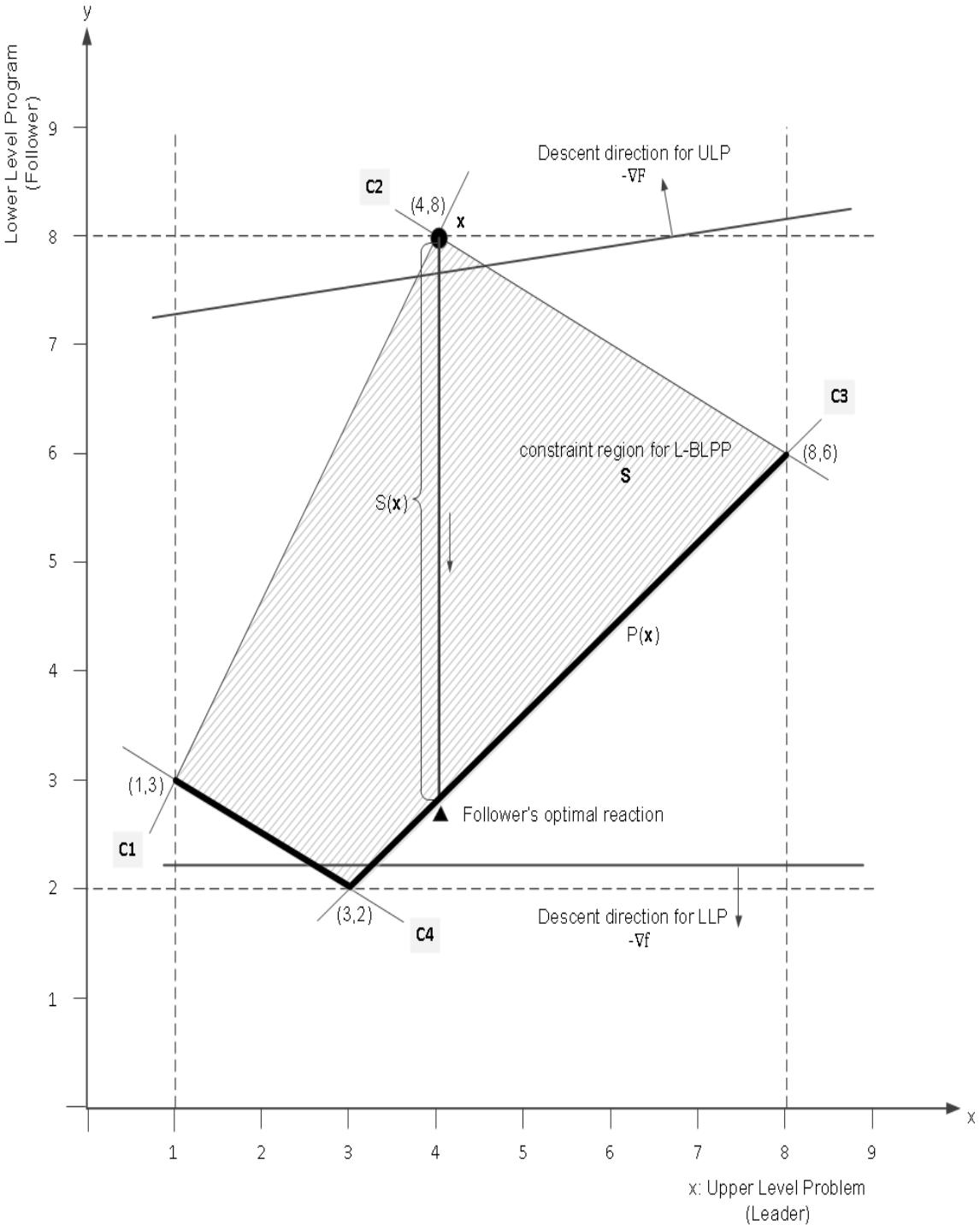

Figure 22: Rational Reaction Set for the Lower Level Problem

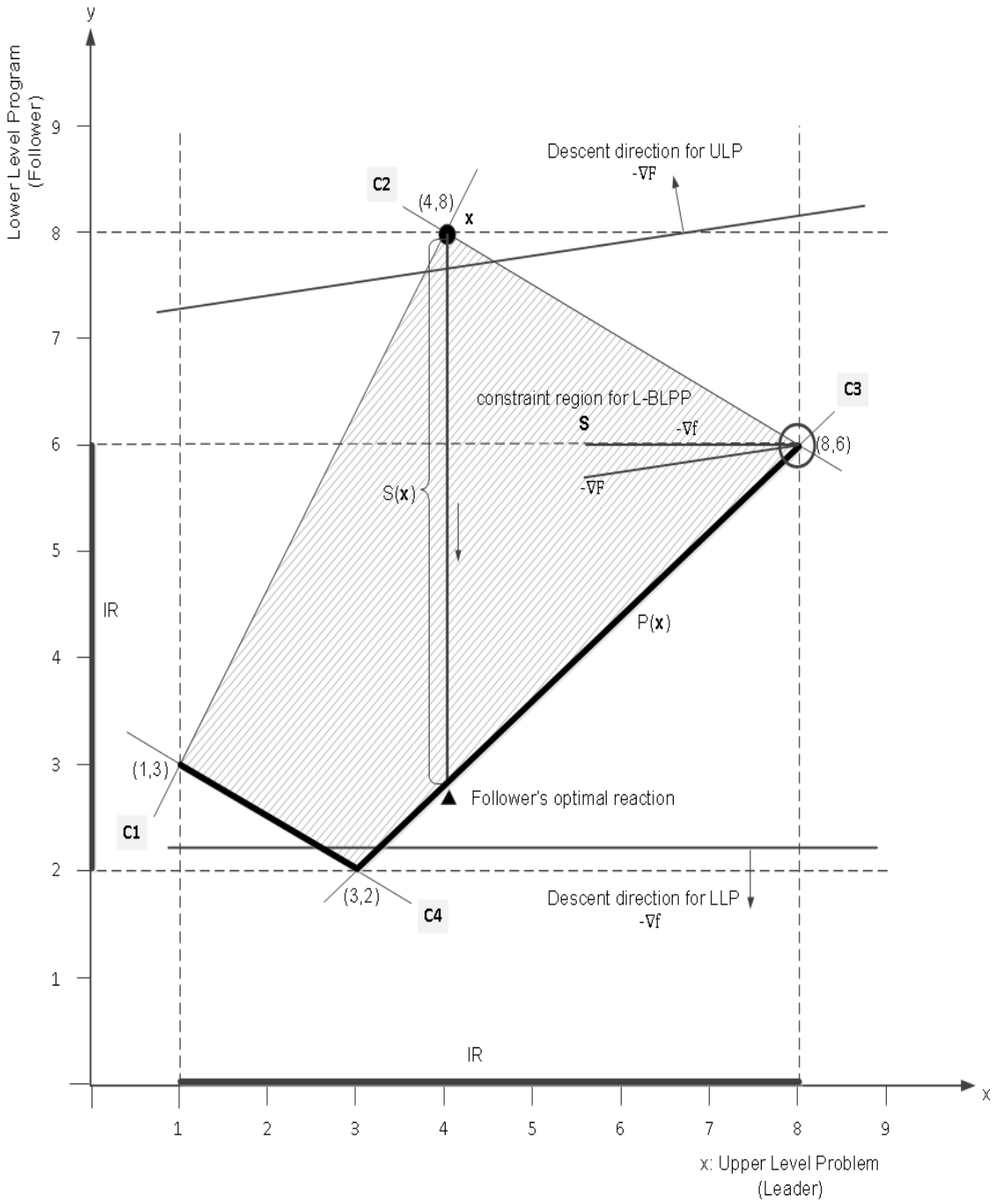

Figure 23: Inducible Region of the L-BLPP

Since the proposed problem has two objective functions and shares significant portions of the feasible region,it may be mistakenly associated with a multi-objective optimization problem. It is important to note that this is not true for all the cases. In a multi-objective optimization problem the different objective functions are related and

meet a criterion called Pareto efficiency. Pareto efficiency is related to the concept of dominance. This concept implies that there cannot be improvements in one objective without worsening other objective functions. In the context of leader-follower problems this means that neither the leader nor the follower can unilaterally improve their objectives. This concept is presented in Figure 24 where plot A shows the optimal solution for the leader and follower strategies and plot B presents the same points in the objective functions space. It can be observed that the optimal solution is dominated by all the points in the cone of the gradients of the objective functions. That means that Pareto optimality is not guaranteed unless the gradients of both the leader and follower are co-linear [28].

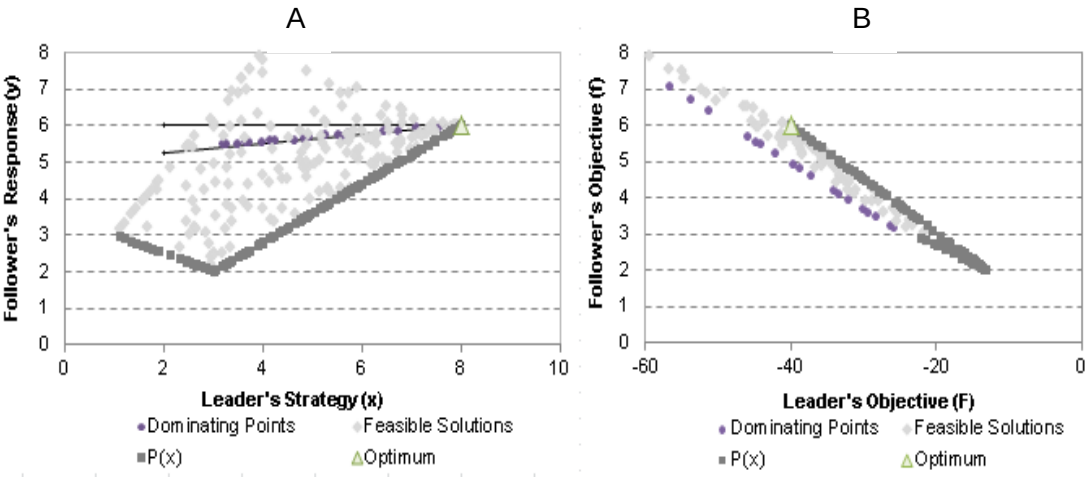

Figure 24: Design Space and Criterion Space for the L-BLPP

The overall complexity of the bi-level program depends on the modeling approach taken to represent the situation being analyzed. As an initial approach it is expected that the lower level model will be solved for each iteration/decision of the leader. This implies that a significant portion of the complexity of the solution can be reduced and there will be efficient re-formulation/solution of the lower level model.

In the context of transportation systems, any change in the network design (topology and parameters) will bring on changes in the flow patterns of the network. The flow patterns are modeled as a traffic equilibrium model to reflect the behavior or rationale of the followers (network users) with respect to the transportation network topology and parameters. Without considering the lower level model, the resulting network design may be optimal for the leader, but it may decrease the overall performance of the network. This situation is referred to as the Braess paradox [29].

4.2 Continuous Network Design Problem (CNDP)

The CNDP is associated with the capacity of the network arcs. Examples of capacity increase in transportation context are:

- Roadway widening
- Increase in green time at signalized intersection
- Increased size of a train
- Increase in bus headways

In multimodal networks the capacity is associated with the requirement of the necessary infrastructure to allow the transition from one transportation mode to another. An example could be the number of parking spaces at a train station.

Let $\mathcal{A}_0$ be the set of arcs that will not be modified. Let $\mathcal{A}_1$ be the set of arcs whose capacity will be modified. The CNDP for a unimodal network can be formulated as a bi-level mathematical programming model as presented in equations 4.18 through 4.21.

$$\min_{\boldsymbol{y}} Z(\boldsymbol{f}, \boldsymbol{y}) = \sum_{a \in A} \rho \cdot t_a(f_a, y_a) \cdot f_a + \sum_{a \in A_1} g_a(y_a) \tag{4.18}$$

subject to:

$$y_a^o \leq y_a \leq y_a^u, \forall\, a \in A_1 \tag{4.19}$$

$$g_a(y_a) \leq B \tag{4.20}$$

$$\boldsymbol{f} = TAP(\boldsymbol{y}) \tag{4.21}$$

The objective function (4.18) corresponds to the upper level problem or leader problem. In this case, the upper level function is minimizing the total travel and the capacity improvement costs. The term ρ is a factor that converts travel time into utility values (e.g. value of time). The capacity improvement is assumed continuous for this initial formulation and is optimized over a bounded region (4.19). Constraint 4.20 ensures that the available budget is not exceeded. This constraint may be optional since the capacity improvement cost is minimized in the objective function. The lower level problem corresponds to a deterministic traffic assignment problem (4.21). The TAP problem takes the new capacity as input $\boldsymbol{y}$ and gives the link flows corresponding to a traffic equilibrium problem following Wardrop's first principle (user equilibrium).

The extended CND in an extended arc-node formulation is presented in equations 4.22 through 4.29.

$$\min_{\boldsymbol{y}} Z(\boldsymbol{f}, \boldsymbol{y}) = \sum_{a \in A} \rho \cdot t_a(f_a, y_a) \cdot f_a + \sum_{a \in A_1} g_a(y_a) \tag{4.22}$$

subject to:

$$y_a^o \leq y_a \leq y_a^u, \forall\, a \in A_1 \tag{4.23}$$

$$g_a(y_a) \leq B \tag{4.24}$$

where f is the solution of:

$$Min\, T(f) = \sum_{a \in \mathcal{A}} \int_0^{f_a} t_a(s) ds \tag{4.25}$$

$$t_a = t_{oa}\left[1 + B\left(\frac{f_a}{k_a + y_a}\right)^P\right] \tag{4.26}$$

$$\sum_{j \in \psi^-(i)} f_{ijw} - \sum_{j \in \psi^+(i)} f_{jiw} = d_{ijw}\,, \forall\, i \in \mathcal{N}, \forall\, w \in \mathcal{W} \tag{4.27}$$

$$f_a = \sum_{a \in \mathcal{A}} \sum_{w \in \mathcal{W}} f_{aw}, \forall\, a \in \mathcal{A} \tag{4.28}$$

$$f_{aw} \geq 0 \;\forall\, a = (i,j) \in \mathcal{A}, \forall\, w = (p,q) \in \mathcal{W} \tag{4.29}$$

The problem expressed by 4.22 through 4.29 represents a CNDP in the arc-node formulation. The first term of the objective function (4.22) represents the total travel time experienced by the transportation network users. Such travel time accounts for congestion. The congested travel time per arc is given by (t_a). This is the travel time experienced by each transportation network user using arc a (f_a). The second term of the objective function seeks to minimize the investment cost in additional capacity (y_a). Constraints 4.23 and 4.24 provide bounds for the upper level decision variables. The lower level problem is a traffic assignment problem conditioned on the capacity decision made by the leader at the upper level. In this case, the leader makes capacity addition decisions that affect the travel cost. These decisions are captured in the definitional constraint 4.26 by the term y_a. Constraint group 4.27 represents the flow conservation constraints. Constraint group 4.28 defines the arc flow as the sum of the different commodities using the same arc. The non-negativity conditions for the arc flows are given by 4.29. Similarly, for the arc-path formulation the continuous network design problem can be expressed as follows:

$$\min_{\boldsymbol{y}} Z(\boldsymbol{f}, \boldsymbol{y}) = \sum_{a \in A} \rho \cdot t_a(f_a, y_a) \cdot f_a + \sum_{a \in A_1} g_a(y_a) \tag{4.30}$$

subject to:

$$y_a^o \leq y_a \leq y_a^u, \forall\, a \in A_1 \tag{4.31}$$

$$g_a(y_a) \leq B \tag{4.32}$$

where f is the solution of:

$$Min\, T(f) = \sum_{a \in \mathcal{A}} \int_0^{f_a} t_a(s)ds \tag{4.33}$$

$$t_a = t_{oa}[1 + B\left(\frac{f_a}{k_a + y_a}\right)^P] \tag{4.34}$$

$$\sum_{r \in R_w} h_{wr} = d_w, \qquad \forall\, w \in \mathcal{W} \tag{4.35}$$

$$f_a = \sum_{w \in \mathcal{W}} \sum_{r \in R_w} \delta_{wra} h_{wr} \;\; \forall\, a \in \mathcal{A} \tag{4.36}$$

$$h_{wr} \geq 0, \qquad \forall\, r \in R_w, \forall\, w \in \mathcal{W}$$
$$f_a \geq 0 \;\forall\, a = (i,j) \in \mathcal{A} \tag{4.37}$$

The arc-path upper level problem is the same as the arc-node formulation. The main differences can be observed in the traffic assignment problem. The upper level capacity decisions are incorporated via y_a variables in the definitional constraint group 4.34. The flow conservation constraints are automatically satisfied with the path formulation. Constraint group 4.35 ensures that the transportation demand is satisfied. The flow-definitional constraint is expressed by 4.36. The flow definitional constraint states that the flow in link a is the addition of all the paths for all the O-D pairs using such arc. Constraint group 4.37 represents the non-negativity constraints.

4.3 Discrete Network Design Model (DNDP)

In the discrete network design problem the goal is to modify the topology of the network by providing additional arcs to the network while minimizing the total costs. Let $\mathcal{A}_2 \subset \mathcal{A}$ be the subset of proposed arcs. Let E_a be the fixed cost of implementing arc $a \in \mathcal{A}_2$ and x_a a binary variable defined as 1 if arc $a \in \mathcal{A}_2$ is implemented, 0 otherwise. The DNDP in the arc-node formulation is expressed by 4.38 through 4.42.

$$\min_{\boldsymbol{y}} Z(\boldsymbol{f}, \boldsymbol{y}, \boldsymbol{x}) = \sum_{a \in A} \rho \cdot t_a(f_a, y_a) \cdot f_a + \sum_{a \in (A_1\, U\, A2)} g_a(y_a) + \sum_{a \in A_2} E_a x_a \tag{4.38}$$

subject to:

$$t_a = t_{oa}[1 + B\left(\frac{f_a}{k_a + y_a}\right)^P] \tag{4.39}$$

$$y_a^o \le y_a \le y_a^u, \forall\, a \in A \tag{4.40}$$

$$\boldsymbol{f} = TAP(\boldsymbol{y}, \boldsymbol{x}) \tag{4.41}$$

$$y_a \ge 0, x_a \in \{0,1\} \tag{4.42}$$

Where f is the solution of:

$$Min\, T(f) = \sum_{a \in \mathcal{A}} \int_0^{f_a} t_a(s)ds \tag{4.43}$$

$$t_a = t_{oa}[1 + B\left(\frac{f_a}{k_a + y_a}\right)^P] \tag{4.44}$$

$$\sum_{j \in \psi^-(i)} f_{ijw} - \sum_{j \in \psi^+(i)} f_{jiw} = d_{ijw}\,, \forall\, i \in \mathcal{N}, \forall\, w \in \mathcal{W} \tag{4.45}$$

$$f_a = \sum_{a \in \mathcal{A}} \sum_{w \in \mathcal{W}} f_{aw}\,, \forall\, a \in \mathcal{A} \tag{4.46}$$

$$f_a \le M x_a \;\forall\, a \in \mathcal{A}_2 \tag{4.47}$$

$$f_{aw} \ge 0 \;\forall\, a = (i,j) \in \mathcal{A}, \forall\, w = (p,q) \in \mathcal{W} \tag{4.48}$$

Formulations 4.38 through 4.42 present the DNDP in the arc-node form. Similar to the previous formulations, the leader's objective is to minimize the total travel time for the users. In addition, the leader seeks to minimize its investment cost in capacity (g_a) and new infrastructure ($E_a x_a$) for new links. The constraint structure is very similar to that of the CNDP. The additional constraint group, 4.46, ensures that any proposed arc must be implemented before sending any flow through it. The big number M in 4.46 can be the arc capacity. Constraint group 4.48 comprises of the non-negativity constraints for the arc flows. The DNDP for the arc-path formulation is presented in formulations 4.49 through 4.58.

$$\min_{\boldsymbol{y}} Z(\boldsymbol{f}, \boldsymbol{y}, \boldsymbol{x}) = \sum_{a \in A} \rho \cdot t_a(f_a, y_a) \cdot f_a + \sum_{a \in (A_1\, U\, A2)} g_a(y_a) + \sum_{a \in A_2} E_a x_a \tag{4.49}$$

subject to:

$$y_a^o \le y_a \le y_a^u, \forall\, a \in A_1 \tag{4.50}$$

$$g_a(y_a) + \sum_{a \in A_2} E_a x_a \leq B \tag{4.51}$$

$$y_a \geq 0, x_a \in \{0,1\} \tag{4.52}$$

Where f is the solution of:

$$Min\, T(f) = \sum_{a \in \mathcal{A}} \int_0^{f_a} t_a(s) ds \tag{4.53}$$

$$t_a = t_{oa}[1 + B\left(\frac{f_a}{k_a + y_a}\right)^P] \tag{4.54}$$

$$\sum_{r \in R_w} h_{wr} = d_w, \qquad \forall\, w \in \mathcal{W} \tag{4.55}$$

$$f_a = \sum_{w \in \mathcal{W}} \sum_{r \in R_w} \delta_{wra} h_{wr} \;\; \forall\, a \in \mathcal{A} \tag{4.56}$$

$$f_a \leq M x_a \;\forall\, a \in \mathcal{A}_2 \tag{4.57}$$

$$h_{wr} \geq 0, \qquad \forall\, r \in R_w, \forall\, w \in \mathcal{W}$$
$$f_a \geq 0 \;\forall\, a = (i,j) \in \mathcal{A} \tag{4.58}$$

Equations 4.49 through 4.58 present the arc-path formulation for the CNDP. The objective function 4.49 seeks to minimize the total travel time cost for the network users and the investment cost in capacity and infrastructure. Constraint 4.51 ensures that network improvements are feasible with respect to existing budget. The lower level problem is an arc-path traffic assignment problem on a network with location and capacity parameters x, y respectively. The effect on capacity is modeled via equation group 4.54. Constraint group 4.55 ensures that transportation demand is satisfied. The link definitional constraints are modeled by constraint group 4.56. Constraint group 4.57 ensures that arcs will be implemented before sending any flow through them. Constraint group 4.58 sets the non-negativity conditions for arc flows and path flows.

4.4 Multimodal Network Design Problem (MNDP)

This section outlines the multimodal network design problem. In the multimodal network design there are different networks with limited interchange points. In the

transportation context the different networks are referred to as modal networks. These networks represent transportation modes such as car, rail, bus, bicycle, etc. The point where a user can switch from one network to another is referred to as a multimodal interchange. The leader of the network may be interested in decreasing the overall transportation costs by providing adequate infrastructure that enables a user to switch between networks at specific points. Multimodal interchanges should be placed such that their utilization is maximized and this will depend on the user preferences.

Figure 25 presents an example of a multimodal network design scenario. The network leader or transportation authority seeks to minimize the total transportation time by promoting the use of combined transportation modes.

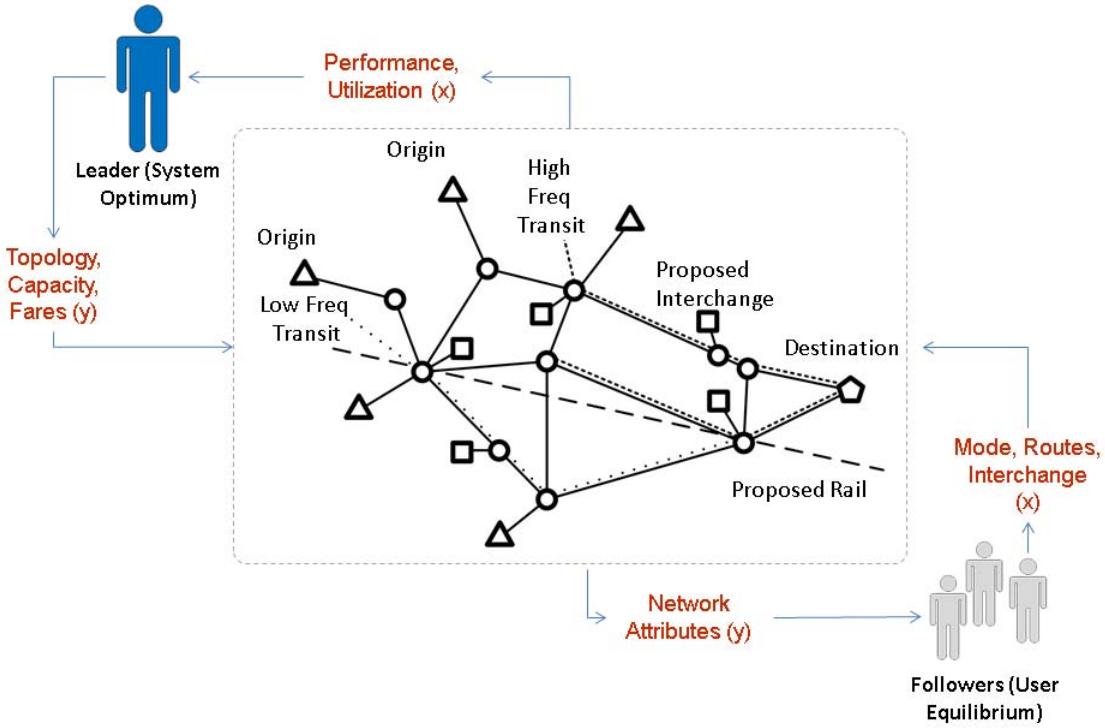

Figure 25: Overview of a Stackelberg Game in Transportation

The upper level has the ability to modify topology, set fares, capacity among others. Such network parameters define a network problem on which the users or followers will apply their strategies to reach their destination. This iterative procedure will

achieve the optimized setting of links and capacities that will minimize the total transportation cost, subject to budgetary and geographic constraints.

The problem of transportation network design has been studied from different standpoints throughout the decades. Dantzig et al. [30] formulated a continuous network design problem using decomposition. In their approach a piece-wise approximation model was assumed and the model was formulated as a single-level system-optimal optimization model. A bi-level model using a special type of linear approximation was formulated by LeBlanc & Boyce [31]. In their approach, the bi-level non-linear problem is converted to a bi-level linear problem. In the resulting linear problem the objective functions of both the lower level and the upper level problems are joined in a convex combination and solved using the solution procedure proposed in Bard [27]. An improvement to the formulation of LeBlanc & Boyce [31] is presented by Ben-Ayed, Boyce, & Blair [32] where a more general shape of cost functions was assumed and linearized to allow the reformulation of the resulting problem in one of the first comprehensive bi-level programs for the transportation network design problem. However, a solution methodology was not presented at that time.

An alternative view of the traffic equilibrium problem consists on formulating it as a system-optimal minimum cost flow model subject to equilibrium constraints. Such types of formulations are referred to as mathematical programming with equilibrium constraints (MPECs) and are re-formulated as variational inequality problems. Examples of network design models formulated as MPECs are presented in Friesz et al. [16]. In their approach, the authors proposed a simulated annealing algorithm to solve the resulting non-convex network design problem. Purely non-linear solution strategies are also suitable to solve certain classes of network design problems. A descend-type algorithm was proposed by Suh & Kim [33] . In their work, a comparative study of non-linear, bi-level programming models applied to the equilibrium network design model problem was

performed. Also, heuristic approaches to the network design problem based on Wardrop's equilibrium can be used to produce good solutions to the problem. An example of one of such heuristic is presented by Marcotte & Marquis [34].

Details regarding historical developments of the traffic network equilibrium models can be found in Boyce [35]. Additional aspects regarding software applications implementing traffic equilibrium algorithms are also provided. A comprehensive review of the network design models is presented in Yang & Bell [36]. Additional reviews regarding transportation network equilibrium and traffic congestion modeling are presented in Boyce [37].

Recently, a growing interest for a more sustainable transportation system has led to the development of multimodal design network problems. García & Marín [38] proposed a non-linear, bi-level programming model for the location of urban multimodal interchange. The proposed model was formulated as a bi-level programming model. The demand was represented by a nested logit model including decisions of mode choice, multimodal interchange and type of parking. Their formulation considered parking demand implicitly by means of a penalty function that increased exponentially as parking usage approached its capacity. The lower level problem on their formulation corresponded to a deterministic user equilibrium assignment model. The bi-level programming was primarily solved by simulated annealing. A similar work by the same authors can be found in García & Marín [39].

A deterministic network equilibrium model for multiple modes was proposed by García & Marín [12]. Their work presents a detailed modeling of the demand through a nested logit model formulation. The multimodal network is modeled through the use of hyperpaths and unified equilibrium conditions were derived combining the two modal networks. The resulting mathematical problem with equilibrium constraint was

formulated using variational inequalities and solved using the column generation/simplicial decomposition method.

Marín & García-Ródenas [14] considered the problem of location of infrastructure in urban rail network. Their objective was to maximize transit demand and minimize travel time. The problem was modeled initially as a non-linear integer bi-level program. The non-linearity arises from the use of a logit function for the modal split of the network users. To solve the resulting integer non-linear problem, a piecewise or polygonal function was used to represent the modal split. The resulting model was a linear integer and was modeled in GAMS,using CPLEX as the underlying solver. Patil & Ukkusuri [11] presented a network design problem formulation for a single-mode network with stochastic demand. Their work compared system optimal problem formulations against user optimal formulations. For their experimental cases they found that the difference between the two formulations does not exceed five percent. This finding allows the formulation of the network design problem without the user equilibrium constraints, making it more tractable. The authors extended the network design problem for the case of stochastic demand and solved the resulting non-linear constraint problem by introducing a new set of penalty functions. Alternative approaches to bi-level programming models have been considered. Farhan & Murray [40] formulated the problem of locating multimodal interchanges as a deterministic facility location problem from a system optimal viewpoint. Their approach was not based on traffic assignment; instead the authors modeled the preferences for the multimodal interchange location based on proximity to the demand. They used an exponential shaped function to represent a decay of preference for park-and-ride based on distance to populated areas.

In this work, the problem of locating infrastructure in the context of multiple transportation networks is considered. The initial model is a non-linear integer bi-level programming model. The model considers modal split and interchange selection by

means of logit functions. As part of the solution approach the non-linearity of the network will be addressed by an interval-free polygonal approximation of the logit functions. This will enable the solution method to have a linear induced lower level model while the binary variables remain in the upper level model.

This section presents an outline of the multimodal assignment model to be used as the lower level problem. The basis for the model is based on a model proposed by Boile & Spasovic [9] and slightly adapted to accommodate the model requirements for the current research. In the example below, the notation T_{pq}^{T} is introduced to denote the number of transit trips (e.g. bus, rail, intermodal, etc.) taking place, the O-D pair w. For a multimodal model, two different types of links have to be considered; roadway links and person links. For person link units, capacity and travel time are given in number of trips such as rail, sidewalks and bike lanes. Roadway links will handle buses and vehicles. For rail and bus service the quality of service can be reflected in the cost function and the preference function. The multimodal assignment problem can be formulated as follows:

$$Min\ L(f,T^{T}) = \sum_{a\in\mathcal{A}}\int_{0}^{f_a} t_a(s)ds + \sum_{w\in\mathcal{W}}\int_{0}^{T_w^T}\frac{1}{\beta}\ln\left(\frac{s}{d_w - s} + \alpha_{TA}\right)ds \tag{4.59}$$

$$d_w = \sum_{r\in R_v} h_{wr} + g_{mw}, \forall\, w \in \mathcal{W}\,; \tag{4.60}$$

$$T_w^T = \sum_{r\in R_T} h_w = \sum_{r\in R_e} h_{wr} + \sum_{r\in R_k} h_{wr}\,, \forall\, w \in \mathcal{W}; \tag{4.61}$$

$$f_a = \frac{1}{\gamma}\sum_{w\in\mathcal{W}}\sum_{r\in R_v}\delta_{wra}h_{wr} + \frac{1}{\gamma}\sum_{w\in\mathcal{W}}\sum_{r\in R_T}\delta_{wra}h_{wr}\ \,, \forall\, a \in A_v \cup A_T \tag{4.62}$$

$$f_a = \sum_{w\in\mathcal{W}}\sum_{r\in R_p}\delta_{wra}h_{wr}\,, \forall\, a \in A_e \cup A_w \tag{4.63}$$

$$f_a \le \kappa_a\ \forall\, a \in A_v \cup A_T \tag{4.64}$$

$$f_a \le \kappa_a\ \forall\, a \in A_e \cup A_w \tag{4.65}$$

$$h_{wrm} \ge 0\ \forall\, w \in \mathcal{W}, \forall\, r \in R_m, \forall\, m \in M \tag{4.66}$$

Equation 4.59 has no physical interpretation other than guiding the mathematical programming model to achieve the user equilibrium conditions at optimality. The first term could be thought as the cumulative link cost which increases with the flow. On the other hand, the second term is the decrease in link costs due to the shift to transit mode. Constraints 4.60 and 4.61 are travel demand satisfaction constraints based on the modal split. Constraints 4.62 and 4.63 are link-flow definition constraints and 4.64 and 4.65 are upper bound constraints for link capacities, and constrain group 4.66 is the non-negativity of route flows. The mode choice is obtained through the following logit model:

$$T_w^v = T_w \frac{e^{-(\alpha_v + \beta GC_w^v)}}{e^{-(\alpha_v + \beta GC_w^v)} + e^{-(\alpha_T + \beta GC_w^T)}}, \forall\, w \in \mathcal{W} \tag{4.67}$$

where α is the alternative specific constant and β is the coefficient of the generalized cost (GC) for the corresponding mode/O-D pair.

CHAPTER 5: SOLUTION OF THE PROPOSED NETWORK DESIGN PROBLEMS

The solution of the bi-level network design model can be performed by following two major solution principles, implicit enumeration and reformulation [28].

In reformulation approaches, the bi-level programming model is reformulated as single level program and the lower level model is replaced by its optimality conditions. The resulting single level program is referred to as a mathematical program with equilibrium constraint (MPEC) and it is a subject of continuous research in the transportation research arena [26]. The work of Gao et al. [10] is an example of implicit enumeration using Generalized Benders Decomposition (GBD).

The formulation presented in this dissertation corresponds to a mixed network design problem, being deterministic and user-optimal with asymmetric link cost functions for the fixed demand case. The problems are solved via reformulation of the bi-level program into a single-level mathematical program with equilibrium constraints (MPEC). The non-linear behaviors derived from the traffic equilibrium problem are represented by piece-wise approximations using a series of max-affine functions following the procedure introduced in Chapter 3. The MPEC bilinear terms are also linearized by means of binary variables. The results are compared with those of well-known network problems in the literature of transportation network design such as the Friesz-Harker, and G1. The solution approach is summarized in Figure 26.

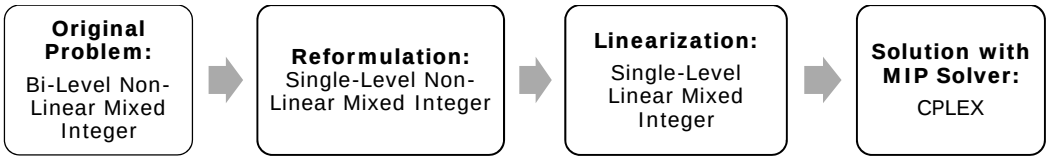

Figure 26: Summary of Solution Approach

5.1 Reformulation Approach

In this section, the optimality conditions for both, the arc-path and the arc-node version of the TAP are derived. Such optimality conditions are the first step to obtain the target single-level reformulation of the different network design problems.

5.1.1 Optimality Conditions for TAP

Recall the arc-path formulation expressed in Equations 2.20 through 2.22. The problem corresponds to a deterministic traffic equilibrium problem. The arc-path formulation is presented again for clarity purposes.

$$Min\, T(f) = \sum_{a \in \mathcal{A}} \int_0^{f_a} t_a(s) ds \tag{5.1}$$

$$\sum_{r \in R_{wq}} h_{wr} = d_w \,, \forall\, w \in \mathcal{W} \tag{5.2}$$

$$\sum_{w \in \mathcal{W}} \sum_{r \in R_w} \delta_{wra} h_{wr} = f_a, \forall\, a \in \mathcal{A} \tag{5.3}$$

The first order optimality conditions can be obtained by relaxing the demand constraints, in 5.2, and formulating the corresponding Lagrangean function. Let π_w be the Lagrange multipliers for constrain set 5.2. There are $|\mathcal{W}|$ constraints corresponding to each of the O-D pairs. The objective function depends of the link flows (f_a) which at the same time depend on the route flows (h_{wr}).

$$L(\boldsymbol{h},\boldsymbol{\pi}) = T\big(f(h)\big) + \sum_{w\in\mathcal{W}} \pi_w \left(d_w - \sum_{r\in R_w} h_{wr} \right) \tag{5.4}$$

The stationary point conditions for the Lagrangean $L(\boldsymbol{h},\boldsymbol{\pi})$ are be stated by 5.5 through 5.9 [4].

$$h_{wr}\frac{\partial L(\boldsymbol{h},\boldsymbol{\pi})}{\partial h_{wr}} = 0, \forall\, r \in R_w, \forall\, w \in \mathcal{W} \tag{5.5}$$

$$\frac{\partial L(\boldsymbol{h},\boldsymbol{\pi})}{\partial h_{wr}} \geq 0, \forall\, r \in R_w, \forall\, w \in \mathcal{W} \tag{5.6}$$

$$\frac{\partial L(\boldsymbol{h},\boldsymbol{\pi})}{\partial \pi_w} = 0, \forall\, r \in R_w, \forall\, w \in \mathcal{W} \tag{5.7}$$

$$h_{wr} \geq 0\ \forall\, r\ \in R_w, \forall\, w \in \mathcal{W} \tag{5.8}$$

$$\frac{\partial T(f(h))}{\partial h_{wr}} = \sum_{a\in A} \frac{\partial T}{\partial f_a}\frac{\partial f_a}{\partial h_{wr}}(f(h)) \tag{5.9}$$

The definitional constraint for the link flows and path flows can be expanded as presented in 5.10.

$$f_a = \delta_{11a}h_{11} + \delta_{12a}h_{12} + \cdots \delta_{1|R_1|a}h_{1|R_1|+\cdots}\delta_{21a}h_{21+\cdots} \tag{5.10}$$

Any derivative with respect to h_{wr} will be zero except for the term δ_{wra}. The term $\partial T/\partial f_a$ represents the derivative of the integral of the cost function. In this case, this will only be the cost function $t_a(f_a)$. In this case the sum over all the links gives the path cost since the term δ_{wra} is a link-path indicator.

$$\sum_{a\in A} \frac{\partial T}{\partial f_a}\frac{\partial f_a}{\partial h_{wr}}\big(f(h)\big) = \sum_{a\in A} \delta_{wra}t_a(f_a) = c_w \tag{5.11}$$

The derivative of the dualized constraint will be zero except for the term involving h_{wr} which results in a value of $-\pi_w$. The extended derivation of the equilibrium conditions for equation 5.5 are expressed in 5.12.

$$\frac{\partial L(\boldsymbol{h},\boldsymbol{\pi})}{\partial h_{wr}} = \frac{\partial}{\partial h_{wr}}\Big(T\big(f(h)\big)\Big) + \frac{\partial}{\partial h_{wr}}\left(\sum_{w\in\mathcal{W}} \pi_w \left(d_w - \sum_{r\in R_w} h_{wr}\right)\right)$$

$$\frac{\partial L(\boldsymbol{h},\boldsymbol{\pi})}{\partial h_{wr}} = \frac{\partial}{\partial h_{wr}}\left(T\big(f(h)\big) + \sum_{w\in\mathcal{W}} \pi_w \left(d_w - \sum_{r\in R_w} h_{wr}\right)\right) \tag{5.12}$$

$$\frac{\partial L(\boldsymbol{h},\boldsymbol{\pi})}{\partial h_{wr}} = \sum_{a\in A} \delta_{wra} t_a(f_a) - \pi_w$$

$$\frac{\partial L(\boldsymbol{h},\boldsymbol{\pi})}{\partial h_{wr}} = c_{wr} - \pi_w$$

The equilibrium expression based on the first-order condition for the arc-path formulation are shown below in 5.13 through 5.20.

$$h_{wr}\frac{\partial L(\boldsymbol{h},\boldsymbol{\pi})}{\partial h_{wr}} = 0, \Rightarrow h_{wr}(c_{wr} - \pi_w) = 0, \qquad \forall\, r \in R_w, \forall\, w \in \mathcal{W} \tag{5.13}$$

$$\frac{\partial L(\boldsymbol{h},\boldsymbol{\pi})}{\partial h_{wr}} \geq 0 \Rightarrow (c_{wr} - \pi_w) \geq 0, \qquad \forall\, r \in R_w, \forall\, w \in \mathcal{W} \tag{5.14}$$

$$\frac{\partial L(\boldsymbol{h},\boldsymbol{\pi})}{\partial \pi_w} = 0 \Rightarrow \sum_{r\in R_w} h_{wr} = d_w, \qquad \forall\, r \in R_w, \forall\, w \in \mathcal{W} \tag{5.15}$$

$$h_{wr} \geq 0\ \forall\, r\ \in R_w, \forall\, w \in \mathcal{W} \tag{5.16}$$

In a similar manner the optimality conditions for the arc-node formulation can be obtained. Equations 5.10 through 5.21 present the basic arc-node formulation of TAP.

$$Min\, T(f) = \sum_{a\in\mathcal{A}} \int_0^{f_a} t_a(s)ds \tag{5.17}$$

$$\sum_{j\in\psi^-(i)} f_{ijw} - \sum_{j\in\psi^+(i)} f_{jiw} = d_{ijw}\,, \forall\, i \in \mathcal{N}, \forall\, w \in \mathcal{W} \tag{5.18}$$

$$f_a = \sum_{a\in\mathcal{A}} \sum_{w\in\mathcal{W}} f_{aw}\,, \forall\, a \in \mathcal{A} \tag{5.19}$$

$$f_{aw} \geq 0\ \forall\, a = (i,j) \in \mathcal{A}, \forall\, w = (p,q) \in \mathcal{W} \tag{5.20}$$

The flow conservation constraint is dualized to obtain the first-order conditions as presented in 5.21.

$$L(\boldsymbol{f},\boldsymbol{\pi}) = T(\boldsymbol{f}_w) + \sum_{w\in\mathcal{W}}\sum_{i\in\mathcal{N}} \pi_{iw}\left(\sum_{j\in\psi^-(i)} f_{ijw} - \sum_{j\in\psi^+(i)} f_{jiw}\right) \tag{5.21}$$

The first order conditions for the arc-node formulation of TAP are summarized in equations 5.22 to 5.25 [4].

$$f_{ijw}\frac{\partial L(\sum_{w\in W}\boldsymbol{f}_w,\boldsymbol{\pi})}{\partial f_{ijw}} = 0 \Rightarrow f_{ijw}\left(t_{ij}(f_{ij}) + \pi_{iw} - \pi_{jw}\right) = 0 \quad \forall\,(i,j)\in\mathcal{A},(p,q)\in\mathcal{W} \tag{5.22}$$

$$\frac{\partial L(\sum_{w\in W}\boldsymbol{f}_w,\boldsymbol{\pi})}{\partial f_{ijw}} \geq 0 \Rightarrow t_{ij}(f_{ij}) + \pi_{iw} - \pi_{jw} \geq 0 \quad \forall\,(i,j)\in\mathcal{A}, w(p,q)\in\mathcal{W} \tag{5.23}$$

$$\frac{\partial L(\sum_{w\in W}\boldsymbol{f}_w,\boldsymbol{\pi})}{\partial \pi_{iw}} = 0 \Rightarrow \sum_{j\in\psi^-(i)} f_{ijw} - \sum_{j\in\psi^+(i)} f_{jiw} = d_{ijw} \quad \forall\, i\in\mathcal{N},\forall\, w\in\mathcal{W} \tag{5.24}$$

$$f_{ijw} \geq 0\ \forall\,(i,j)\in\mathcal{A}, w\in\mathcal{W} \tag{5.25}$$

5.1.2 Linearization of Equilibrium Conditions

The equilibrium conditions for both problems are expressed as non-linear terms. Such constraints can be linearized using binary variables. For the arc-path formulation the equilibrium conditions are presented in 5.26.

$$\begin{aligned} h_{wr}(c_{wr} - \pi_w) &= 0, \quad \forall\, r\in R_w,\forall\, w\in\mathcal{W} \\ c_{wr} - \pi_w &\geq 0, \quad \forall\, r\in R_w,\forall\, w\in\mathcal{W} \\ \sum_{r\in R_w} h_{wr} &= d_{pq}, \quad \forall\, w\in\mathcal{W} \\ h_{wr} &\geq 0, \quad \forall\, r\in R_w,\forall\, w\in\mathcal{W} \\ \pi_w &\geq 0, \quad w\in\mathcal{W} \end{aligned} \tag{5.26}$$

An additional variable z_{wr} is defined as 1 if path h_{wr} is used in the equilibrium, 0 otherwise. The new equilibrium conditions for the arc-path formulation are presented in 5.27.

$$\begin{gathered} h_{wr} \leq M_1 z_{wr} = 0, \qquad \forall\, r \in R_w, \forall\, w \in \mathcal{W} \\ c_{wr} - \pi_w \geq (1 - z_{wr}) M_2, \qquad \forall\, r \in R_w, \forall\, w \in \mathcal{W} \\ c_{wr} - \pi_w \geq 0, \qquad \forall\, r \in R_w, \forall\, w \in \mathcal{W} \\ \sum_{r \in R_w} h_{wr} = d_w, \forall\, w \in \mathcal{W} \\ h_{wr} \geq 0, \forall\, r \in R_w, \forall\, w \in \mathcal{W} \\ \pi_w \geq 0, \forall\, w \in \mathcal{W} \\ z_{wr} \in \{0,1\} \end{gathered} \tag{5.27}$$

The equilibrium conditions for the arc-node formulation can be linearized in a similar way. The original non-linear equilibrium conditions are presented in 5.28.

$$\begin{gathered} f_{ijw}(t_{ij}(f_{ij}) + \pi_{iw} - \pi_{jw}) = 0\,, \qquad \forall\, (i,j) \in \mathcal{A}, \forall\, w \in \mathcal{W} \\ t_{ij}(f_{ij}) + \pi_{iw} - \pi_{jw} \geq 0\,, \qquad \forall\, (i,j) \in \mathcal{A}, \forall\, w \in \mathcal{W} \\ \sum_{j \in \psi^-(i)} f_{ijw} - \sum_{j \in \psi^+(i)} f_{jiw} = d_{ijw}\,, \forall\, i \in \mathcal{N}, \forall\, w \in \mathcal{W} \\ f_{ijw} \geq 0, \pi_{ijw} \geq 0, \forall\, (i,j) \in \mathcal{A}, \forall\, w \in \mathcal{W} \end{gathered} \tag{5.28}$$

With the introduction of the binary variable z_{ijw} defined as 1 if link $(i,j) \in \mathcal{A}$ is used in the equilibrium solution, 0 otherwise. The linearized equilibrium conditions are re-written in 5.29.

$$\begin{gathered} f_{ijw} \leq M_1 z_{ijw}\,, \qquad \forall\, (i,j) \in \mathcal{A}, \forall\, w \in \mathcal{W} \\ t_{ij}(f_{ij}) + \pi_{iw} - \pi_{jw} \geq (1 - z_{ijw}) M_2, \qquad \forall\, (i,j) \in \mathcal{A}, \forall\, w \in \mathcal{W} \\ t_{ij}(f_{ij}) + \pi_{iw} - \pi_{jw} \geq 0\,, \qquad \forall\, (i,j) \in \mathcal{A}, \forall\, w \in \mathcal{W} \\ \sum_{j \in \psi^-(i)} f_{ijw} - \sum_{j \in \psi^+(i)} f_{jiw} = d_{ijw}\,, \forall\, i \in \mathcal{N}, \forall\, w \in \mathcal{W} \\ f_{ijw} \geq 0, \pi_{ijw} \geq 0, \forall\, (i,j) \in \mathcal{A}, \forall\, w \in \mathcal{W} \\ z_{ijw} \in \{0,1\} \end{gathered} \tag{5.29}$$

5.1.3 Reformulation of Objective Functions

The objective function of TAP seeks to minimize the total transportation time for all the network users. The implication of congestion modeling is that the unit cost of an arc is a function of a decision variable (flow) which is multiplied by another decision variable (flow). This poses additional challenges for the non-linearity of the objective function even in cases when the arc cost function can be linearized.

To cope with this situation, the optimality conditions can be used to re-formulate the objective function in way that does not involve a product of two decision variables. In the arc-path formulation the total travel time in the network is shown in 5.30.

$$T = \sum_{a \in \mathcal{A}} t_a(f_a) f_a \tag{5.30}$$

At the equilibrium, the transportation demands are satisfiedd_w, and the path cost for a given O-D pair is minimum and the same. Also, the only route flows h_w greater than 0 are those with route costs equal to the equilibrium cost π_w. The total travel time can be re-written as 5.31.

$$T = \sum_{w \in \mathcal{W}} \sum_{r \in R_w} h_{wr} \pi_w \tag{5.31}$$

A similar concept can be applied to the arc-node formulation, taking into consideration that the O-D pair cost is the difference in potentials between the origin and the destination nodes.

$$T = \sum_{w \in \mathcal{W}} d_w \left(\pi_{qw} - \pi_{pw}\right) \tag{5.32}$$

5.1.4 Linearized CNDP

The linearized version of the CNDP corresponds to network capacity modeling. The resulting model using the arc-path formulation is presented in 5.33 through 5.43.

$$\min_{\boldsymbol{y}} Z(\boldsymbol{f},\boldsymbol{y}) = \sum_{w\in\mathcal{W}} \rho \cdot \pi_w d_w + \sum_{a\in A_1} g_a(y_a) \tag{5.33}$$

subject to:

$$t_a \geq \alpha_{ag} + \beta_{ag} f_a + \theta_{ag} y_a, \qquad \forall\, g \in G_a\, \forall\, a \in \mathcal{A} \tag{5.34}$$

$$g_a(y_a) \leq B \tag{5.35}$$

$$y_a^o \leq y_a \leq y_a^u, \forall\, a \in A_1 \tag{5.36}$$

$$h_{wr} \leq M_1 z_{wr} \;\; \forall\, r \in R_w, \forall\, w \in \mathcal{W} \tag{5.37}$$

$$c_{wr} - \pi_w \leq (1 - z_{wr}) M_2, \qquad \forall\, r \in R_w, \forall\, w \in \mathcal{W} \tag{5.38}$$

$$c_{wr} - \pi_w \geq 0, \qquad \forall\, r \in R_w, \forall\, w \in \mathcal{W} \tag{5.39}$$

$$\sum_{r\in R_w} h_{wr} = d_w, \qquad \forall\, w \in \mathcal{W} \tag{5.40}$$

$$f_a = \sum_{w\in\mathcal{W}} \sum_{r\in R_w} \delta_{wra} h_{wr} \;\; \forall\, a \in \mathcal{A} \tag{5.41}$$

$$c_{wr} = \sum_{a\in A} \delta_{wra} t_a \;\; \forall\, r \in R_w, \forall\, w \in \mathcal{W} \tag{5.42}$$

$$\begin{gathered} h_{wr} \geq 0, \;\; \forall\, r \in R_w, \forall\, w \in \mathcal{W}, \\ \pi_w \geq 0, \qquad w \in \mathcal{W} \\ z_{wr} \in \{0,1\} \end{gathered} \tag{5.43}$$

The objective function (5.33) seeks to minimize the total travel time in the network. The parameter ρ can be thought as the value of time coefficient. It allows converting the network total travel time to a utility value or cost. The second term of the objective is related to the capacity investments. Constraint group 5.34 represents the linearized flow-capacity surface. There are $|G_a|$ functions per arc a, the arc travel time is denoted by t_a. Note how the travel time estimate is he maximum t_a for all the G_a functions for link a. Constraint group 5.35 ensures that the capacity improvements are within the available budget. Constraint group 5.36 set the search region or bounds on the additional capacity for arc a. Constraint group 5.37 is part of the linearized equilibrium conditions; it states that the flow on a route is zero if the route is not in use. The big number M_1 is the minimum between the demand for OD pair w or the route capacity (minimum capacity for ell the arcs in the route). Constraint group 5.38 is part of the equilibrium conditions

and states that when the path h_{wr} is used the path cost should be equal to the equilibrium cost or greater if the path is not used. Constraint group 5.39 ensures that the path cost is always greater or equal to the equilibrium path cost for OD pair w.Constraint group 5.40 ensures that the transportation demand is satisfied. The arc-flow definitional constraints are given by 5.41 which state that the arc flow is sum of the flows of the paths using that arc. Similarly, the path cost definitional constraints are given by constraint group 5.42 stating that the path cost is the sum of the cost of its arcs. Non negativity constraints and binary variables specifications are given by 5.43.

Similarly, the CNDP using the arc-node formulation is presented in 4.60 through 5.53.

$$\min_{\boldsymbol{y}} Z(\boldsymbol{f},\boldsymbol{y}) = \rho \sum_{w\in\mathcal{W}} d_w\left(\pi_{qw} - \pi_{pw}\right) + \sum_{a\in A_1} g_a(y_a) \tag{5.44}$$

subject to:

$$t_a \geq \alpha_{ag} + \beta_{ag} f_a + \theta_{ag} y_a, \qquad \forall\, g \in G_a\ \forall\, a \in \mathcal{A} \tag{5.45}$$

$$g_a(y_a) \leq B \tag{5.46}$$

$$y_a^o \leq y_a \leq y_a^u, \forall\, a \in A_1 \tag{5.47}$$

$$f_{ijw} \leq M_1 z_{ijw}\,, \qquad \forall\,(i,j) \in \mathcal{A}, \forall\, w \in \mathcal{W} \tag{5.48}$$

$$t_{ij}\left(f_{ij}\right) + \pi_{iw} - \pi_{jw} \geq \left(1 - z_{ijw}\right)M_2, \qquad \forall\,(i,j) \in \mathcal{A}, \forall\, w \in \mathcal{W} \tag{5.49}$$

$$t_{ij}\left(f_{ij}\right) + \pi_{iw} - \pi_{jw} \geq 0\,, \qquad \forall\,(i,j) \in \mathcal{A}, \forall\, w \in \mathcal{W} \tag{5.50}$$

$$\sum_{\psi^-(i)} f_{ijw} - \sum_{\psi^+(i)} f_{jiw} = d_{ijw}\,, \forall\, i \in \mathcal{N}, \forall\, w \in \mathcal{W} \tag{5.51}$$

$$f_a = \sum_{a\,\in\mathcal{A}} \sum_{w\in\mathcal{W}} f_{aw}\,, \forall\, a \in \mathcal{A} \tag{5.52}$$

$$\begin{gathered} f_{aw} \geq 0\ \forall\, a = (i,j) \in \mathcal{A}, \forall\, w = (p,q) \in \mathcal{W} \\ \pi_{ijw} \geq 0, \forall\,(i,j) \in \mathcal{A}, \forall\, w \in \mathcal{W} \\ z_{ijw} \in \{0,1\} \end{gathered} \tag{5.53}$$

The objective function seeks to minimize the total transportation cost and the capacity investment cost. Constraint group 5.45 represents the linearized flow-capacity surface. Budget constraints are represented by 5.46. Upper and lower bounds on the

capacity improvement are handled by constraint group 5.47. Constraint group 5.48 ensures that only selected arcs will have a positive O-D flow. Constraint group 5.49 is part of the equilibrium condition and states that when an arc has a positive O-D flow the difference in potentials between its terminal nodes is equal to the arc cost. Constraint group 5.50 guarantees that the arc cost is always greater than the difference in potential between its terminal nodes so that the equilibrium condition holds. Constraint group 5.51 deals with flow-conservation conditions and demand satisfaction. The arc flow definitional constraints are given by constraint group 5.52 which states that the arc flow is the sum of the arc flows due to the different O-D pairs w. Non-negativity constraints and binary variable definitions are given by 5.53.

5.1.5 Linearized DNDP

The linearized version of the DNDP is similar to the previously defined formulations with the addition of the arc creation binary variable. The arc-path formulation for the DNDP is given in 5.54 through 5.65

$$\min_{\boldsymbol{y}} Z(\boldsymbol{f},\boldsymbol{y}) = \sum_{w\in\mathcal{W}} \cdot \rho \cdot \pi_w d_w + \sum_{a\in(A_1\ U\ A2)} g_a(y_a) + \sum_{a\in A_2} E_a x_a \tag{5.54}$$

Subject to,

$$t_a \geq \alpha_{ag} + \beta_{ag} f_a + \theta_{ag} y_a, \qquad \forall\, g \in G_a\ \forall\, a \in \mathcal{A} \tag{5.55}$$

$$g_a(y_a) + \sum_{a\in A_2} E_a x_a \leq B \tag{5.56}$$

$$y_a^o \leq y_a \leq y_a^u, \forall\, a \in A_1 \tag{5.57}$$

$$h_{wr} \leq M_1 z_{wr}\ \ \forall\, r \in R_w, \forall\, w \in \mathcal{W} \tag{5.58}$$

$$c_{wr} - \pi_w \leq (1 - z_{wr}) M_2, \qquad \forall\, r \in R_w, \forall\, w \in \mathcal{W} \tag{5.59}$$

$$c_{wr} - \pi_w \geq 0, \qquad \forall\, r \in R_w, \forall\, w \in \mathcal{W} \tag{5.60}$$

$$\sum_{r\in R_w} h_{wr} = d_w, \qquad \forall\, w \in \mathcal{W} \tag{5.61}$$

$$f_a = \sum_{w\in\mathcal{W}} \sum_{r\in R_w} \delta_{wra} h_{wr}\ \ \forall\, a \in \mathcal{A} \tag{5.62}$$

$$c_{wr} = \sum_{a \in A} \delta_{wra} t_a \quad \forall\, r \in R_w, \forall\, w \in \mathcal{W} \tag{5.63}$$

$$f_a \leq M_3 x_a \;\forall\, a \in \mathcal{A} \tag{5.64}$$

$$\begin{gathered} h_{wr} \geq 0, \;\; \forall\, r \in R_w, \forall\, w \in \mathcal{W}, \\ \pi_w \geq 0, \qquad w \in \mathcal{W} \\ z_{wr} \in \{0,1\} \\ x_a \in \{0,1\} \end{gathered} \tag{5.65}$$

The objective function seeks to minimize the leader's investment cost and the total travel time for network users. The parameter ρ transforms the travel time units into monetary costs. The investments are composed by capacity and location of new infrastructure. The binary variable x_a determines whether the new arc $a \in A_2$ is implemented. Constraint 5.55 represents the linearized flow-capacity surface. Constraint 5.56 ensures that the network improvement does not exceed the available budget. Constraint group 5.57 establishes the bounds of the capacity improvements. Constraint group 5.58 is an equilibrium condition that guarantees that the only the selected paths can have a positive flow. Constraint group 5.59 ensures that path costs are equal to the equilibrium costs when the paths carry a positive flow. Constraint group 5.60 ensures that the path cost is grater or equal than the equilibrium cost. Constraint group 5.61 deals with demand satisfaction. Link definitional expressions are given by constraint group 5.62. 5.63 state that the path cost is the sum of the cost of its arcs. For arc addition, constraint group 5.64 establishes that only implemented links can have positive flow. Non-negativity constraints and binary variable definitions are given by constraint group 5.65.

The arc-node formulation for the linearized DNDP is presented in 5.66 through 5.76.

$$\min_{\boldsymbol{y}} Z(\boldsymbol{f}, \boldsymbol{y}) = \rho \sum_{w \in \mathcal{W}} d_w \left(\pi_{qw} - \pi_{pw}\right) + \sum_{a \in (A_1 \; U \; A2)} g_a(y_a) + \sum_{a \in A_2} E_a x_a \tag{5.66}$$

subject to:

$$t_a \geq \alpha_{ag} + \beta_{ag} f_a + \theta_{ag} y_a, \qquad \forall\, g \in G_a\, \forall\, a \in \mathcal{A} \tag{5.67}$$

$$g_a(y_a) + \sum_{a \in A_2} E_a x_a \leq B \tag{5.68}$$

$$y_a^o \leq y_a \leq y_a^u, \forall\, a \in A_1 \tag{5.69}$$

$$f_{ijw} \leq M_1 z_{ijw}\,, \qquad \forall\, (i,j) \in \mathcal{A}, \forall\, w \in \mathcal{W} \tag{5.70}$$

$$t_{ij}(f_{ij}) + \pi_{iw} - \pi_{jw} \geq (1 - z_{ijw}) M_2, \qquad \forall\, (i,j) \in \mathcal{A}, \forall\, w \in \mathcal{W} \tag{5.71}$$

$$t_{ij}(f_{ij}) + \pi_{iw} - \pi_{jw} \geq 0\,, \qquad \forall\, (i,j) \in \mathcal{A}, \forall\, w \in \mathcal{W} \tag{5.72}$$

$$\sum_{\psi^-(i)} f_{ijw} - \sum_{\psi^+(i)} f_{jiw} = d_{ijw}\,, \forall\, i \in \mathcal{N}, \forall\, w \in \mathcal{W} \tag{5.73}$$

$$f_a = \sum_{a \in \mathcal{A}} \sum_{w \in \mathcal{W}} f_{aw}\,, \forall\, a \in \mathcal{A} \tag{5.74}$$

$$f_a \leq M_3 x_a, \forall\, a \in \mathcal{A} \tag{5.75}$$

$$\begin{gathered} f_{aw} \geq 0\ \forall\, a = (i,j) \in \mathcal{A}, \forall\, w = (p,q) \in \mathcal{W} \\ \pi_{ijw} \geq 0, \forall\, (i,j) \in \mathcal{A}, \forall\, w \in \mathcal{W} \\ z_{ijw} \in \{0,1\} \\ x_a \in \{0,1\} \end{gathered} \tag{5.76}$$

In the preceding formulation the objective function seeks to minimize the total travel time and the investment cost in capacity and new infrastructure. The travel time between any origin p and destination q is given by the difference in their node potentials. This is equivalent to the equilibrium path cost in the arc-path formulation. The linearization of the flow-capacity surface is given by constraint group 5.67. Budget constraints are represented by 5.68. The bound on the capacity improvements are given by 5.69. Constraint group 5.70 represents the equilibrium conditions for O-D flow variables stating that only selected arcs can carry positive O-D flow. Constraint group 5.71 ensures that at the equilibrium conditions the difference in node potential between arc terminals is equal to the arc cost if the arc carries a positive flow. Constraint group 5.72 guarantees the consistency between cost and O-D flow for each arc. O-D flow conservation and demand satisfaction are handled by constraint group 5.73. Link flow definition constraints are given by 5.74. Constraint group 5.75 ensures that new links are

implemented before carrying any positive flow. Non-negativity constraints and binary variable definitions are given by constraint group 5.76.

5.2 Computational Approach

The computational framework was designed in Visual C#.NET. The LP/MIP solver of choice was CPLEX. Both solvers are accessed through academic licenses. For the non-linear benchmarks IPOPT was used. At early stages this research accessed solver via C#.NET. Later in the development it was decided to use GAMS to facilitate data input/output processes.

Additional computational tools employed in this research are QuickGraph which is an open-source library and data structures to handle graphs and graphs algorithms. In addition, several functions from the CSLapak library were compiled and made available as a library for the project. CSLapak is library for linear algebra manipulation translated from the Fortan library LAPACK to C#.NET. A graphical summary of the computational approach is presented in Figure 27.

The network problem or instance was set up in MS Excel. Also, a database to store the experiment data was created in MS Access. The database contained experiment parameters and methods to be tested on the problem instances.

The experiment configuration was created and stored in the database. Also the network instance was read from MS Excel and the corresponding network objects were generated. Once the input was processed all the linearization tests were performed in accordance to the given experiment parameters.

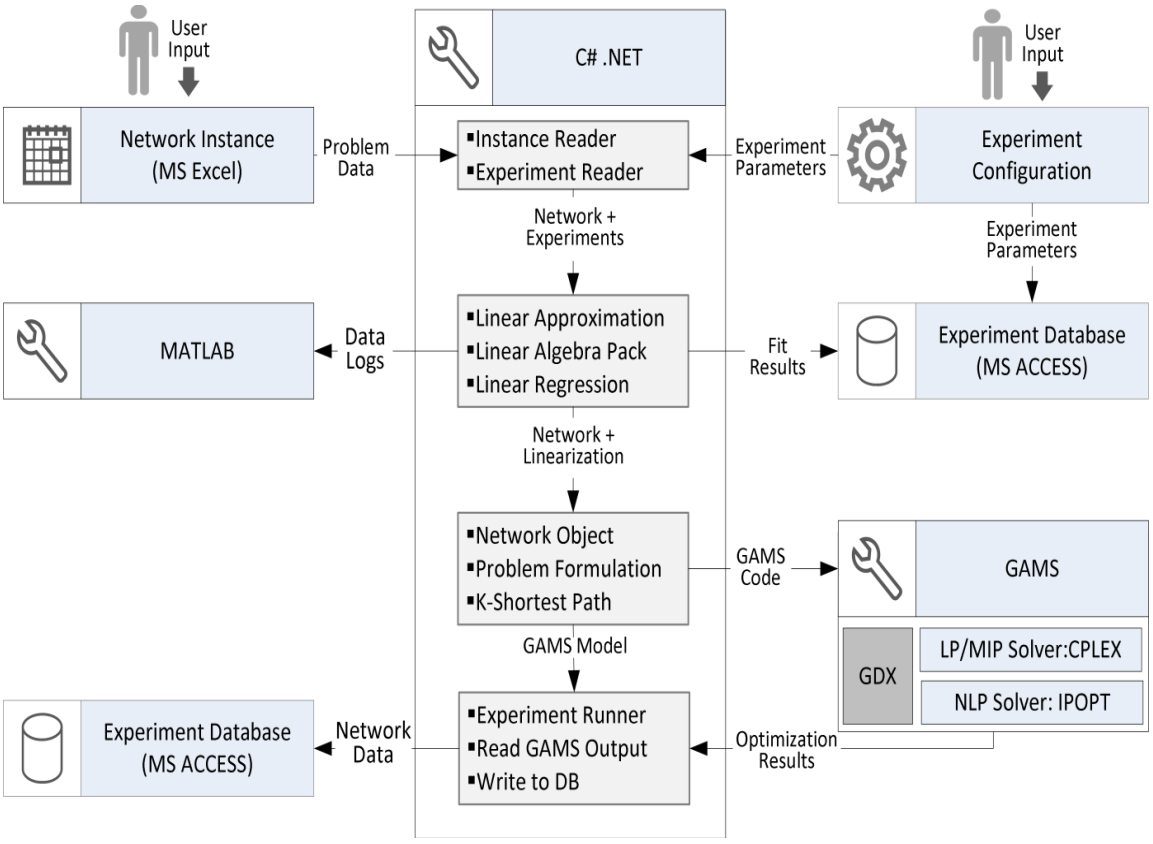

Figure 27: Overview of the Computational Approach

Note that the same model can be solved multiple times to test for the best subset of linearization parameters. The application then writes the GAMS code for the network problem providing instructions to produce a GDX dump file. The generated GAMS file is run from the main application via shell (DOS mode). Once the problem is solved, the application reads the GAMS output and transforms it into the desired data structure. The results are stored in the database via ADO.NET.

5.3 Computational Results for Capacity and Location Decisions

This section presents the computational results for the methods proposed in this dissertation for the capacity and location of infrastructure in unimodal networks.

The Friesz-Harker network was used to calibrate the models and algorithms. First, the linearization procedure parameters were adjusted to provide a good fit in general for

all arc costs, including those with low capacity expansion ratios (potential bottlenecks). For the CNDP, the non-linear version of the problem was solved first (NLP-Friesz). The NLP-Friesz solution was set as the target for performance evaluation of the linearization parameters. The problem is linearized and solved subject to the capacity values of the Modular in core Nonlinear System (MINOS) solution [16] [18]. Results from the Friesz-Harker network have been reported in the literature for more than two decades. The most accepted benchmarks are Equilibrium Decomposed Optimization (EDO), Hook and Jeeves (H-J) and Iterative optimization-assignment algorithm (IOA). CNDP solutions for the congested scenario are presented in Table 6. The baseline scenario for comparison is denoted as NLP-Friesz and it was solved using the non-linear solver IPOPT.

This step was used to select the best set of parameters for the linearization algorithm. The linearized problem is then solved using an LP-MIP solver (CPLEX) to obtain the capacity improvements. An additional step to obtain the equilibrium flows was performed using the obtained capacity as inputs to the NLP-Friesz problem. An overview of the evaluation approach is presented in Figure 28.

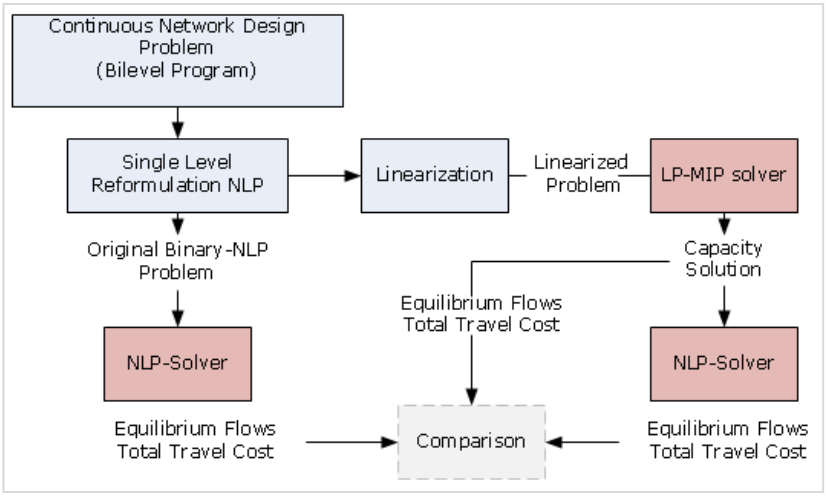

Figure 28: Overview of Performance Evaluation Approach

The best subset of parameter was chosen based on the minimum calibration difference and number of functions across all the network scenarios

Table 6: Constrained Calibration Test Problem for the Congested Scenario

NLP-Friesz	MINOS	H-J	EDO	IOA
4.61	4.61	5.4	4.88	4.55
9.86	9.86	8.18	8.59	10.65
7.71	7.71	8.1	7.48	6.43
			0.26	0
0.59	0.59	0.9	0.85	0.59
1.32	1.32	3.9	1.54	1.32
19.14	19.14	8.1	0.26	19.32
0.85	0.85	8.4	12.52	0.78
557.144	557.14	557.22	540.74	556.61

The best subset of a parameter was chosen based on the minimum calibration difference and number of functions across all the network scenarios. The selected linearization settings were 10 functions, with 0.5 distribution factor (equal number of linear functions for both undersaturated and oversaturated regions) and a saturation factor of 1.1. This means that the undersaturated region was reduced and is approximated with five functions while the oversaturated region was expanded and is approximated with five functions. These results are consistent with the modeling logic since most of the optimal solutions are expected to be in the undersaturated region. A comparison of the calibration results is presented inTable 7.

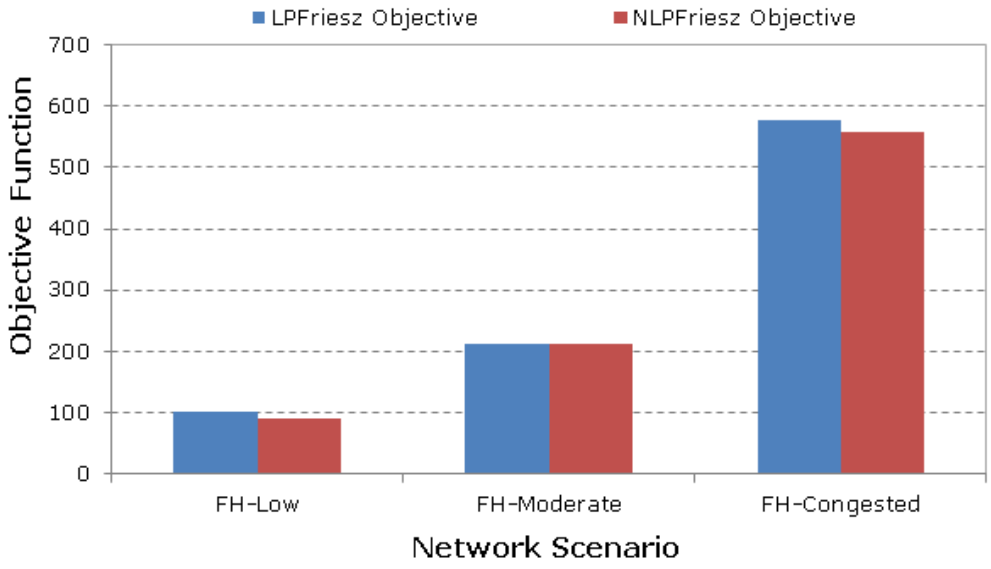

Figure 29: Comparison of Objective Functions for the Constrained Network Problem

The numerical results for the MIP solution for the constrained design problem are presented in Table 7. It can be observed that the maximum difference occurs in the low congestion scenario (6%). Wang at al. [18] performed similar experiments with a grid-type of partition using binary variables to model flow-capacity surface using a piece-wise linear representation. Such approach has the advantage of having a high level of accuracy at the expense of a computationally intensive model. For example, in [18] a total of 2338 variables including 336 binary variables were used to model the Friesz-Harker network. In the proposed approach, the same network was modeled using 64 variables of which 15 were binary.

The smallest experiment in [18] used a grid (5 x 5), the error between the objective and the solution of the linearized problem was 2.9 percent. For the congested scenario (Scenario II in [18]) the relative error was 3.68 percent. The computational time for the MIP modeling by Wang et al. [18] had a reported computational time of 1.5 min for a 5x5 discretization approach for the moderate congestion scenario. Using a 15 x 15

linearization approach, the reported computational time was 1.2 hours giving an error of 0.63 percent.

The results for the initial calibration test are presented in Table 7. The results were obtained with the selected linearization parameters in the calibration process. The results indicate a six percent difference in the low-congested scenario, one percent for the moderate-congested and 3.8 percent for the congested scenario. These results are for calibration purposes only and cannot be compared to those obtained in the literature.

Table 7: Numerical Results for MIP Solution

Scenario	LP Objective	NLP Objective	Difference (%)
FH-Low	98.29	92.10	6.07%
FH-Moderate	213.37	211.25	1.01%
FH-Congested	578.33	557.14	3.80%

The linearized CNDP was solved to obtain a full solution of the capacity expansion problem. This model is referred to as (LP-CAP-Friesz) and can be used to assess the solution quality of the proposed approximation. To finalize the evaluation of the proposed approach, the capacity results of the LP-CAP-Friesz are input as constraints to the capacity expansion variables y_a of the non-linear version of the problem. The purpose of this last step is to obtain the corresponding equilibrium flows for the improved capacity conditions. The results for all the aforementioned problems are presented in Figure 30.

To evaluate the quality of the solution, three performance measures indicators were devised, for calibration, application and equilibrium calculations. The calibration difference is the deviation of the L-CNDP with respect to the constrained version for parameter calibration. The application difference is the deviation of the solution of the linearized CNDP with respect to the baseline. The equilibrium difference is the difference between the baseline model and the equilibrium flows of the capacity vector obtained by L-CNDP. Tables 8 and 9 present such performance measures.

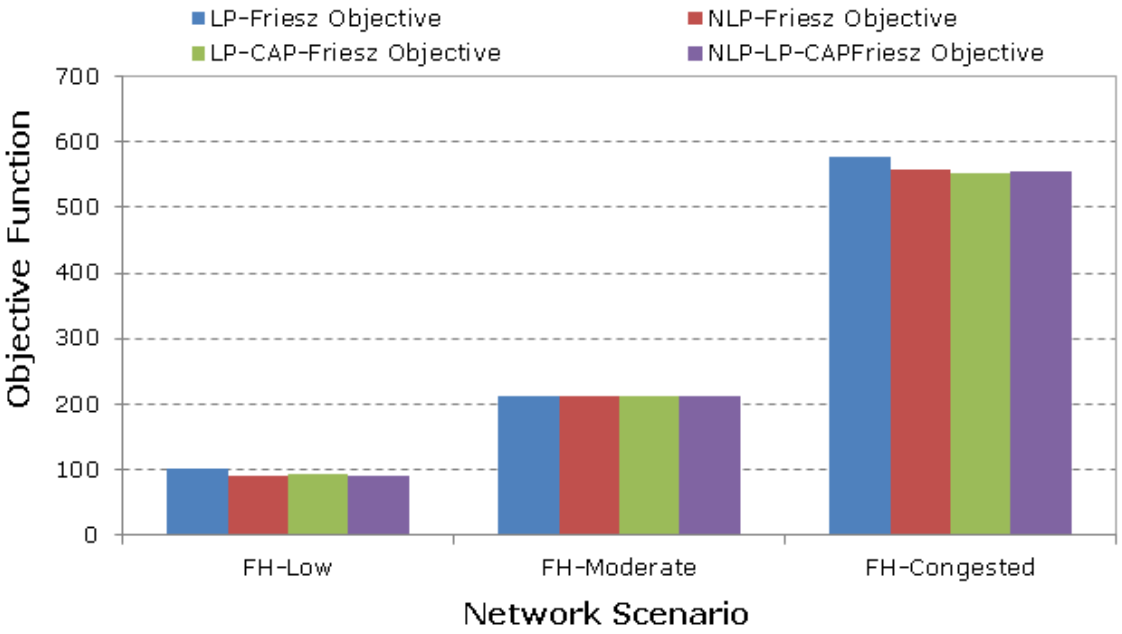

Figure 30: Objective Function Comparison for L-CNDP for the Friesz-Harker Network

Table 8: Objective Function Values for L-CNDP

Scenarios	LP-Friesz Objective	NLP-Friesz Objective	LP-CAP-Friesz Objective	NLP-LP-CAP Friesz Objective
FH-Low	98.28	92.09	93.08	90.73
FH-Moderate	213.37	211.24	212.73	211.61
FH-Congested	578.33	557.14	552.67	555.81

Table 9: Calibration, Application, and Equilibrium Differences for L-CNDP

Scenarios	Calibration Difference	Application Difference	Equilibrium Difference
FH-Low	6.723%	1.06%	1.50%
FH-Moderate	1.007%	0.70%	0.17%
FH-Congested	3.803%	0.81%	0.24%

It can be observed that for the application and equilibrium differences, the selected linearization parameters exhibit a competitive performance. The proposed models are

comparable or outperform the previously obtained results. In terms of computational time the problem L-CNDP running time is less than one second. This is due to the absence of binary variables for the linearization scheme. Figure 31 presents the elapsed time (compilation and execution) for L-CNDP. It can be observed that even when the number of functions increases that the execution time was not affected at exponential rates. This indicates that the model complexity is not heavily affected by the linearization approach. Comparing the maximum elapsed time for the L-CNDP with the minimum execution time of 90 seconds in [18], a reduction of 99.66% of the execution time was achieved.

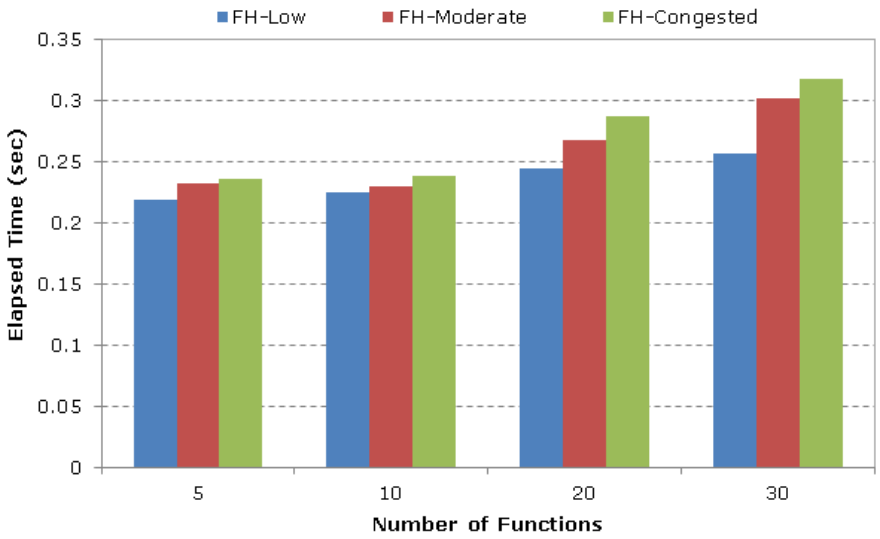

Figure 31: Elapsed Time for L-CNDP

The proposed linearization approach has many advantages. It is relatively simple to implement and it has strong generalization features. For instance, results with the selected parameters and only five functions are presented in Figure 32 and Table 10. Depending on the application type, the proposed capacity modeling can be us to simplify and generalize estimated behaviors and the error is less than six percent. For transportation application this can be a way to simplify non-linear behavior, taking advantage of the computational advances of linear programming solvers.

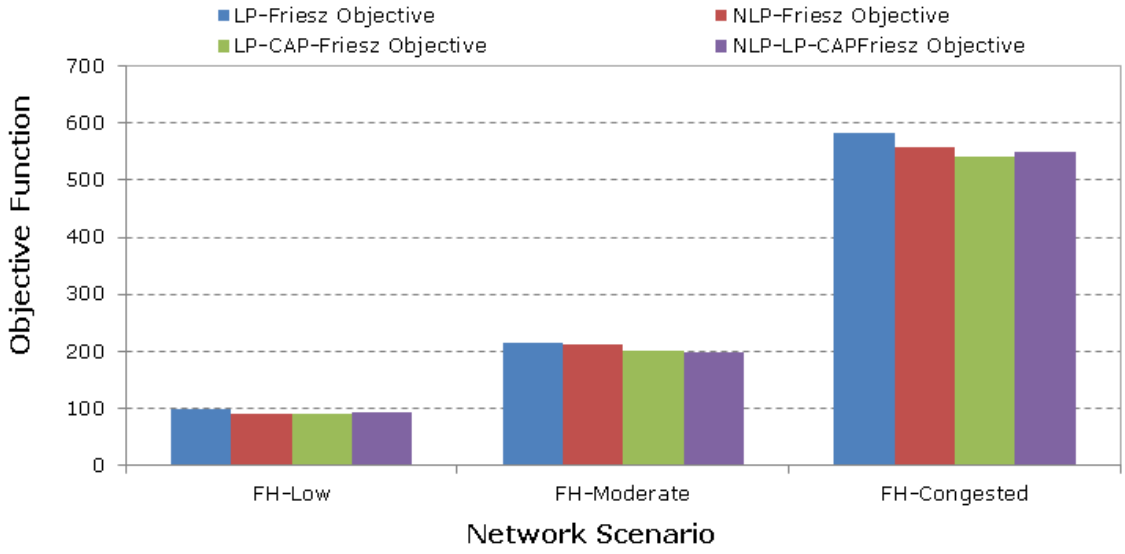

Figure 32: Results for L-CNDP for 5 Functions

Table 10: Calibration, Application, and Equilibrium Differences for L-CNDP

Scenarios	Calibration Difference	Application Difference	Equilibrium Difference
FH-Low	7.337%	1.04%	1.51%
FH-Moderate	1.766%	4.33%	5.73%
FH-Congested	4.427%	2.92%	1.51%

CHAPTER 6: CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

In this research a network design non-linear bi-level programming model decision problem was modeled and solved in the context of transportation network design. Location and capacity decisions were analyzed from a mathematical programming standpoint.

The transportation network design was analyzed describing its constituent elements. The analysis started with the basic multicommodity minimum cost network with linear cost and no congestion effects. The base problem was expanded adding more layers of complexity reaching a non-linear bi-level network design problem subject to congestion effects.

Solving the transportation network design problem is an ever evolving research topic in the transportation network modeling literature. Recently, there has been interest on solving the linearized version of the CNDP and the DNDP. It was observed that the most representative examples of the latest research work in transportation network design problems used a linear-mixed integer representation of the flow-capacity surface was used. Such representations gave origin to a great number of binary variables and constraints. In general, the linearization of a univariate function in N pieces will give origin to N additional binary variables, N continuous variables and 3N+1 constraints. For a bivariate function, these numbers are greatly increased. In this work, a data mining algorithm was used to estimate a linear representation of the flow-capacity surface taking advantage of its convexity. The proposed linearization approach was based onan existing data mining algorithm and consists of approximating the functions as max-affine

combinations of linear functions. In the proposed approach, the function is estimated by the maximum of a series of functions. For that reason, intervals and interval indicator variables (binary variables) are not required in the proposed linearization strategy. As a result, a non-linear model can be linearized to a pure linear model rather than to a mixed-integer program as in the case of binary representations. In terms of problem representation, a partition in N intervals will require N constraints and N continuous variables.

In addition to a decreased problem size, the proposed linearization of the flow-capacity surface was analyzed further to obtain additional insights on the effect of the linear approximation, capacity modeling, and solution quality. The capacity expansion ratio was defined as the ratio of the current arc capacity divided by the maximum potential capacity. A large value of this ratio means that the current arc capacity is close to the potential maximum capacity. On the other hand, small capacity expansion ratios are indication of capacity bottlenecks and high variability in the flow-capacity surface. It was found that when a good fit was obtained in the flow-capacity surface of the arcs with low capacity expansion ratios (e.g. less than 0.1), the solution of the linearized problem was within six percent to that of a non-linear version of the problem.

Additional capacity analyses were performed, dividing the flow-capacity plane in oversaturated and undersaturated regions depending on the flow-to-capacity ratio. In an oversaturated region the flow exceeds the capacity (flow-to-capacity ratio>1) leading to extreme values in the flow-capacity surface. On the other hand, the undersaturated region presents a more stable behavior. Such conditions were used in the linear approximation procedure to devise an algorithmic improvement, allocating more functions to the undersaturated region and fitting both regions separately. A detailed calibration procedure was performed to select adequate linearization parameters for the arc capacity modeling. The selected linearization settings were ten functions, distributed

evenly (5 for the unsaturated region and 5 for the oversaturated region) and a saturation factor of 1.1. This means that the undersaturated region was reduced and is approximated with five functions while the oversaturated region was expanded and is approximated with five functions. These results are consistent with the model logic since in most cases the minimum solution will tend to be in the undersaturated region.

The resulting algorithm produced linear approximations offering comparable performance in solution quality to those that are in the existing literature. The results obtained with five and ten functions are comparable or in some cases outperform some of the results for the Friesz-Harker benchmark network presented in the existing literature. For example, some authors solved the CNDP using a 5 x 5 grid (25 intervals) for the capacity-flow surface while with the proposed linearization approach the same results can be obtained with ten functions. Similar results were obtained for other benchmark networks such as G1 for the DNDP.

The computational time of the proposed approach for small and moderate networks outperformed some the existing result by a significant amount. Compared to linear mixed integer fallow-capacity representation, for three cases of the Fries-Harker benchmark network, the proposed approach was able to reduce the CPU time over 90 percent for the CNDP.

The problem of locating infrastructure was coupled with the proposed linearization approach to transform the original bi-level non-linear binary problem into a single-level mixed integer problem. Such problems can be solved by commercial linear and MIP solvers such as CPLEX for small and moderate network problems. The results obtained were compared to benchmark network G1. The results of the proposed approach are comparable to those obtained in the existing transportation network design literature. For larger networks it is recommended that specialized algorithms should be developed.

The problem of modeling and communicating multiple networks was treated as a larger instance of a single network problem using an arc-path formulation. This concept is known as the hyperpath approach. The original problem is a non-linear bi-level binary problem. The application of the proposed solution approach lead to the capacity and location selection of infrastructure in a multimodal network setting. If the candidate arc transfers units between two networks then the problem becomes location and capacity of a multimodal network interchange.

Additional detail on the modeling of a transportation network such as utility functions and congestion pricing can be integrated in the proposed approach and are left for future research. Also a more specialized algorithm for solving the resulting MIP can be expanded, using problem specific data to devise cuts to accelerate the convergence for large networks. These topics will be covered in research derived from this work.

In the context of transportation planning, network designs, and operations research, this dissertation contributes to the theory and practice of the following aspects:

- Modeled a decision-making process of a central agency with respect to an existing network with non-cooperative users.
- Provided an alternative methodology to analyze capacity in non-linear networks subject to congestion effects.
- Explored innovative linearization methods that provide good generalization power with competitive computational performance and solution quality.
- Introduced the concept of network capacity modeling using the flow-capacity surface, capacity expansion ratio and saturation conditions.
- Provided enhancement for a current convex piece-wise fitting method when applied to transportation system capacity analysis. The enhanced fitting

algorithm takes into consideration the existing capacity and the potential future capacity (upper boundary) to approximate the flow-capacity surface.

- Contributed to the current literature of solutions to benchmark transportation problems. In this work, the solution to the linearized CNDP significantly outperforms some of the recently published results in computational time and competitive solution quality.

- In this work, several mathematical programming models on the topic of transportation planning and traffic equilibrium were formulated. This by itself constitutes a contribution to the optimization/operations research field applied to transportation decision-making.

- Created a computational framework that can be systematically utilized and enhanced for future research projects. In addition to research, several byproducts related to optimization and network modeling were created. These are products that can be used for educational purposes.

The approach presented in this work can be expanded to other areas either within the transportation field or any application of network modeling.

- The proposed approach to capacity modeling does not require a functional form. It can be applied to raw data to fit a flow-capacity surface. The proposed linearization approach was applied to a bivariate function. Since it is a model fitting approach it can be applied to multivariate functions as long as the functions are or tend to exhibit a convex behavior. For example, there could be a flow-capacity-time function to schedule capacity changes in time.

- Multimodal freight networks: Topology decision on connectivity in multimodal networks can be adapted to modal networks for freight (e.g. rail, truck, air). Transshipment decisions occur at network interchanges, and capacity and location of such interchanges may have significant impact on the movement of goods in the country. The capacity modeling framework used in this work can be extended to other domains.

- Supply chains: network modeling in supply chains benefit with the result of this research since it expands the traditional location concept where the demand is fixed and known. Preference or utility functions for the facility location problem can be modeled and solved with the decision-making framework proposed in this research

- Telecommunications: With the increasing need of communication channels, leasing of optical fiber networks (dark fiber), and the inclusion of competing firms can make topology decisions on network/provider selection a critical issue in the near future.

REFERENCES

[1] National Academy of Science, "NAE Grand Challenges for Engineering: Restore and Improve Urban Infrastructure," 2012. [Online]. Available: http://www.engineeringchallenges.org/cms/8996/9136.aspx. [Accessed 03 10 2012].

[2] M. Bazaraa, J. Jarvis and H. Sherali, Linear Programming and Network Flows, Hoboken: John Wiley & Sons, 2010.

[3] D. Bertsekas, Network Optimization: Continuous and Discrete Models, Belmont, Massachusetts: Aetna Scientific, 1998.

[4] M. Patriksson, The Traffic Assignment Problem: Models and Methods, Utrecht, The Netherlands: VSP, 1994.

[5] A. Fabregas, G. Centeno and P.-S. Lin, "Extension of queuing models for signalized traffic intersections to manufacturing and service environments," in Proceedings of the the IIE Annual Conference, Orlando, FL, 2006.

[6] M. Bell and Y. Lida, Transportation Network Analysis, John WIley & Sons, 1997.

[7] Y. Sheffi, Urban Transportation Networks, Englewood Cliffs,N.J.: Prentice Hall, 1985.

[8] S. V. Ukkusuri, T. V. Mathew and S. T. Waller, "Robust Transportation Network Design Under Demand Uncertainty," Computer-Aided Civil and Infrastructure Engineering, vol. 22, no. 1, pp. 6-18, 2007.

[9] M. Boile and L. Spasovic, "An Implementation of the Mode-Split Traffic-Assignment Method," Computer-Aided Civil and Infrastructure Engineering, vol. 15, no. 4, pp. 293-307, 2001.

[10] Z. Gao, J. Wu and H. Sun, "Solution algorithm for the bi-level discrete network design problem," Transportation Research Part B: Methodological, vol. 39, no. 6, pp. 479-495, 2005.

[11] G. Patil and S. Ukkusuri, "tem-Optimal Stochastic Transportation Network Design," Transportation Research Record, pp. 80-86, 2007.

[12] R. García and A. Marín, "Network equilibrium with combined modes: models and solution algorithms," Transportation Research Part B: Methodological, vol. 39, no. 3, pp. 223-254, 2005.

[13] Á. Marín and P. Jaramillo, "Urban rapid transit network capacity expansion," European Journal of Operational Research, vol. 191, no. 1, pp. 43-58, 2008.

[14] Á. Marín and R. García-Ródenas, "Location of infrastructure in urban railway networks," Computers & Operations Research, vol. 36, no. 5, pp. 1461-1477, 2009.

[15] C. Suwansirikul, T. Friesz and R. Tobin, "Equilibrium Decomposed Optimization:A Heuristic for the Continuous Equilibrium Network Deign Problem," Transportation Science, vol. 21, no. 4, 1987.

[16] T. L. Friesz, H.-J. Cho, N. J. Mehta, R. L. Tobin and G. Anandalingam, "A Simulated Annealing Approach to the Network Design Problem with Variational Inequality Constraints," Transportation Science, vol. 26, no. 1, pp. 18-26, 1992.

[17] Q. Meng, H. Yang and M. G. H. Bell, "An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem," Transportation Research Part B: Methodological,, vol. 35, no. 1, 2001.

[18] D. Z. Wang and H. K. Lo, "Global optimum of the linearized network design problem with equilibrium flows," Transportation Research Part B, p. 482–492, 2010.

[19] H. Farvaresh and M. Mehdi, "A single-level mixed integer linear formulation for a bi-level discrete network design problem," Transportation Research Part E, p. 623–640, 2011.

[20] P. Luathep, S. Agachai, W. H. Lam, Z.-C. Li and H. K. Lo, "Global optimization method for mixed transportation network design problem: A mixed-integer linear programming approach," Transportation Research Part B, p. 808–827, 2011.

[21] L. Leblanc, "An algorithm for the discrete network design problem," Transportation Science, vol. 3, no. 9, p. 183–199, 1975.

[22] M. Smith, "The existence, uniqueness ans stability of traffic equilibria," Transportation Research Part B: Methodological, vol. 13, no. 4, pp. 295-304, December 1979.

[23] M. Padberg, "Approximating separable nonlinear functions via mixed zero-one programs," Operations Research Letters, pp. 1-5, 2000.

[24] J. P. Vielma, S. Ahmed and G. Nemhauser, "A Note on "A Superior Representation Method for Piecewise Linear Functions"," INFORMS Journal on Computing, pp. 493-497, 2010.

[25] A. Magnani and S. Boyd, "Convex piecewise-linear fitting," Optimization and Engineering, pp. 1-17, 2009.

[26] P. Luathep, A. Sumalee, W. Lam, Z.-C. Li and H. Lo, "Global optimization method for mixed transportation network design problem: A mixed-intger linear programming approach," Transportation Research Part B, pp. 808-827, 2011.

[27] J. F. Bard, "An Efficient Point Algorithm for a Linear Two-Stage Optimization Problem," Operations research, pp. 670-684, 1983.

[28] J. Bard, Practical Bilevel Optimization, Dordrecht, The Netherlands: Kluwer Academic , 1998.

[29] D. Braess, A. Nagurney and T. Wakolbinger, "On a Paradox of Traffic Planning," Transportation Science, pp. 446-450, 2005.

[30] G. B. Dantzig, R. P. Harvey, Z. F. Lansdowne, D. W. Robinson and S. F. Maier, "Formulating and solving the network design problem by decomposition," Transportation Research Part B: Methodological, vol. 13, no. 1, pp. 5-17, 1979.

[31] L. J. LeBlanc and D. E. Boyce, "A bilevel programming algorithm for exact solution of the network design problem with user-optimal flows," Transportation Research Part B: Methodological, pp. 259-265, 1986.

[32] O. Ben-Ayed, D. E. Boyce and C. E. Blair, "A general bilevel linear programming formulation of the network design problem," Transportation Research Part B: Methodological, vol. 22, no. 4, pp. 311-318, 1988.

[33] S. Suh and T. Kim, "Solving nonlinear bilevel programming models of the equilibrium network design problem: A comparative review," Annals of Operations Research, vol. 34, no. 1, pp. 203-218, 1992.

[34] P. Marcotte and G. Marquis, "Efficient implementation of heuristics for the continuous network design problem," Annals of Operations Research, vol. 34, no. 1-4, pp. 163-176, 1992.

[35] D. Boyce, "Forecasting Travel on Congested Urban Transportation Networks: Review and Prospects for Network Equilibrium Models," Networks and Spatial Economics, vol. 7, no. 2, pp. 99-128, 2007.

[36] H. Yang and M. G. Bell, "Models and algorithms for road network design: a review and some new developments," Transport Reviews, vol. 18, no. 3, p. 257, 1998.

[37] D. Boyce, "Future research on urban transportation network modeling," Regional Science and Urban Economics, vol. 37, no. 4, pp. 472-481, 2007.

[38] R. García and A. Marín, "Urban Multimodal Interchange Design Methodology," Mathematical Methods on Optimization in Transportation Systems, vol. 48, pp. 49-79, 2001.

[39] R. García and A. Marín, "Parking Capacity and Pricing in Park'n Ride Trips: A Continuous Equilibrium Network Design Problem," Annals of Operations Research, vol. 116, no. 1, pp. 153-178, 2002.

[40] B. Farhan and A. T. Murray, "Siting park-and-ride facilities using a multi-objective spatial optimization model," Computers & Operations Research, vol. 35, no. 2, pp. 445-456, 2008.

[41] H.-L. Li, H.-C. Lu, C.-H. Huang and N.-Z. Hu, "A Superior Representation Method for Piecewise Linear Functions," INFORMS Journal on Computing, pp. 314-321, 2009.

[42] C. Floudas, Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications, New York: Oxford University Press, 1995.

[43] H. Lo and W. Szeto, "Time-dependent transport network design under cost-recovery," Transportation Research Part B, pp. 43 (1),142-158, 2009.

APPENDICES

Appendix A: Sioux Falls Network Data

Table A: Sioux Fall Network Parameters

Arc Number	source	target	t_{oa}	B	k	P
1	1	2	0.06	0.009	25.900201	4
2	1	3	0.04	0.006	23.403473	4
3	2	1	0.06	0.009	25.900201	4
4	2	6	0.05	0.0075	4.9581809	4
5	3	1	0.04	0.006	23.403473	4
6	3	4	0.04	0.006	17.110524	4
7	3	12	0.04	0.006	23.403473	4
8	4	3	0.04	0.006	17.110524	4
9	4	5	0.02	0.003	17.782794	4
10	4	11	0.06	0.009	4.9088267	4
11	5	4	0.02	0.003	17.782794	4
12	5	6	0.04	0.006	4.9479955	4
13	5	9	0.05	0.0075	10	4
14	6	2	0.05	0.0075	4.9581809	4
15	6	5	0.04	0.006	4.947995	4
16	6	8	0.02	0.003	4.898588	4
17	7	8	0.03	0.0045	7.841811	4
18	7	18	0.02	0.003	23.40347	4
19	8	6	0.02	0.003	4.898588	4
20	8	7	0.03	0.0045	7.841811	4
21	8	9	0.1	0.015	5.050193	4
22	8	16	0.05	0.0075	5.045823	4
23	9	5	0.05	0.0075	10	4
24	9	8	0.1	0.015	5.050193	4
25	9	10	0.03	0.0045	13.91579	4

Appendix A (continued)

Table A (continued)

Arc Number	source	target	t_{oa}	B	k	P
26	10	9	0.03	0.0045	13.91579	4
27	10	11	0.05	0.0075	10	4
28	10	15	0.06	0.009	13.512	4
29	10	16	0.04	0.0075	4.854918	4
30	10	17	0.08	0.012	4.993511	4
31	11	4	0.06	0.009	4.908827	4
32	11	10	0.05	0.0075	10	4
33	11	12	0.06	0.009	4.908827	4
34	11	14	0.04	0.006	4.876508	4
35	12	3	0.04	0.006	23.40347	4
36	12	11	0.06	0.009	4.908827	4
37	12	13	0.03	0.0045	25.9002	4
38	13	12	0.03	0.0045	25.9002	4
39	13	24	0.04	0.006	5.0912562	4
40	14	11	0.04	0.006	4.8765083	4
41	14	15	0.05	0.0075	5.1275261	4
42	14	23	0.04	0.006	4.9247906	4
43	15	10	0.06	0.009	13.512002	4
44	15	14	0.05	0.0075	5.1275261	4
45	15	19	0.03	0.006	14.564753	4
46	15	22	0.03	0.006	9.5991806	4
47	16	8	0.05	0.0075	5.0458226	4
48	16	10	0.04	0.0075	4.8549177	4
49	16	17	0.02	0.003	5.2299101	4
50	16	18	0.03	0.0045	19.679897	4
51	17	10	0.08	0.012	4.9935107	4
52	17	16	0.02	0.003	5.2299101	4
53	17	19	0.02	0.003	4.823951	4
54	18	7	0.02	0.003	23.40347	4
55	18	16	0.03	0.0045	19.6799	4
56	18	20	0.04	0.006	23.40347	4
57	19	15	0.03	0.006	14.56475	4
58	19	17	0.02	0.003	4.823951	4
59	19	20	0.04	0.006	5.002608	4
60	20	18	0.04	0.006	23.40347	4

Appendix A (continued)

Table A (continued)

Arc Number	source	target	t_{oa}	B	k	P
61	20	19	0.04	0.006	5.002608	4
62	20	21	0.06	0.009	5.059912	4
63	20	22	0.05	0.0075	5.075697	4
64	21	20	0.06	0.009	5.059912	4
65	21	22	0.02	0.003	5.22991	4
66	21	24	0.03	0.0045	4.885358	4
67	22	15	0.03	0.006	9.599181	4
68	22	20	0.05	0.0075	5.075697	4
69	22	21	0.02	0.003	5.22991	4
70	22	23	0.04	0.006	5	4
71	23	14	0.04	0.006	4.924791	4
72	23	22	0.04	0.006	5	4
73	23	24	0.02	0.003	5.078508	4
74	24	13	0.04	0.006	5.091256	4
75	24	21	0.03	0.0045	4.885358	4
76	24	23	0.02	0.003	5.078508	4

Appendix B: Additional Flow-Capacity Surface Fitting Results

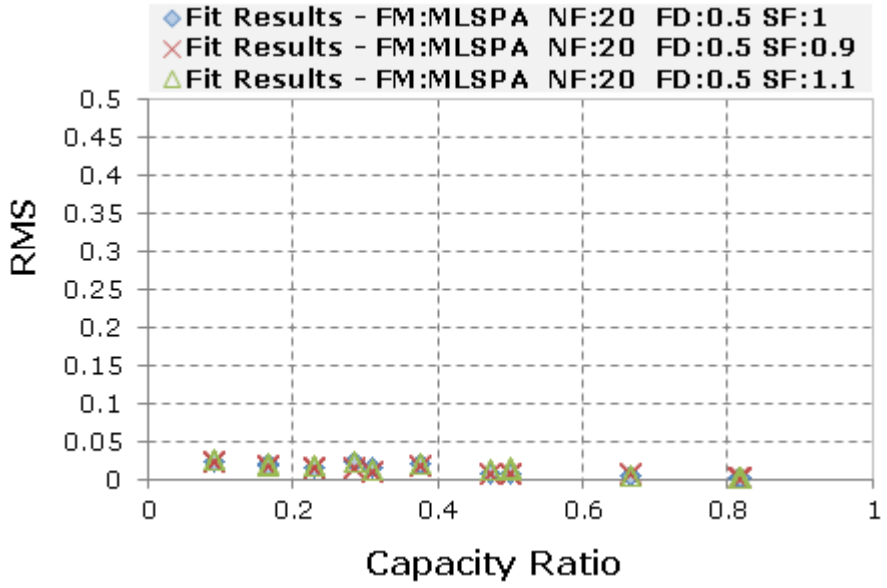

Figure A: Fit Results and Capacity Ratio for the Friesz-Harker Network Using MLSPA, 20 Functions and Function Distribution 0.5

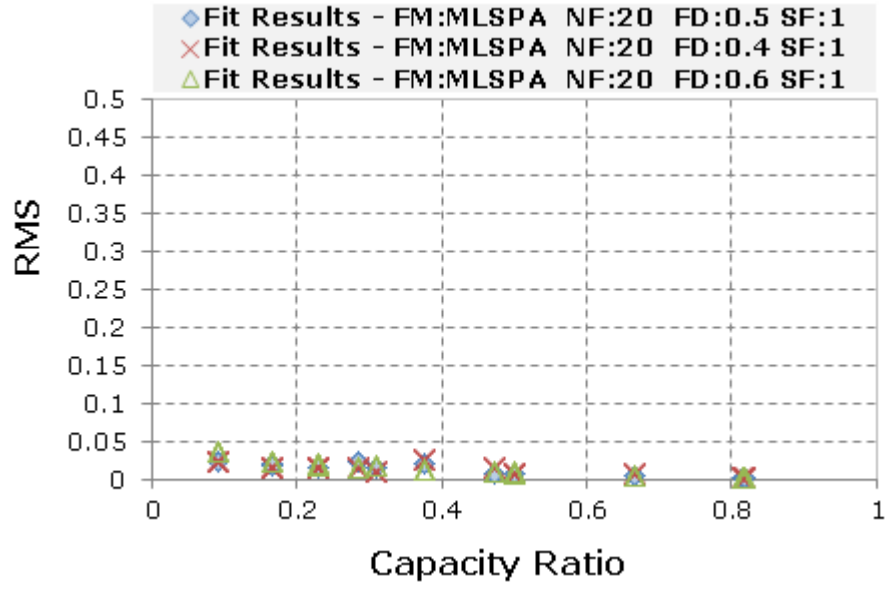

Figure B: Fit Results and Capacity Ratio for the Friesz-Harker Network Using MLSPA, 20 Functions Saturation Factor 1.0

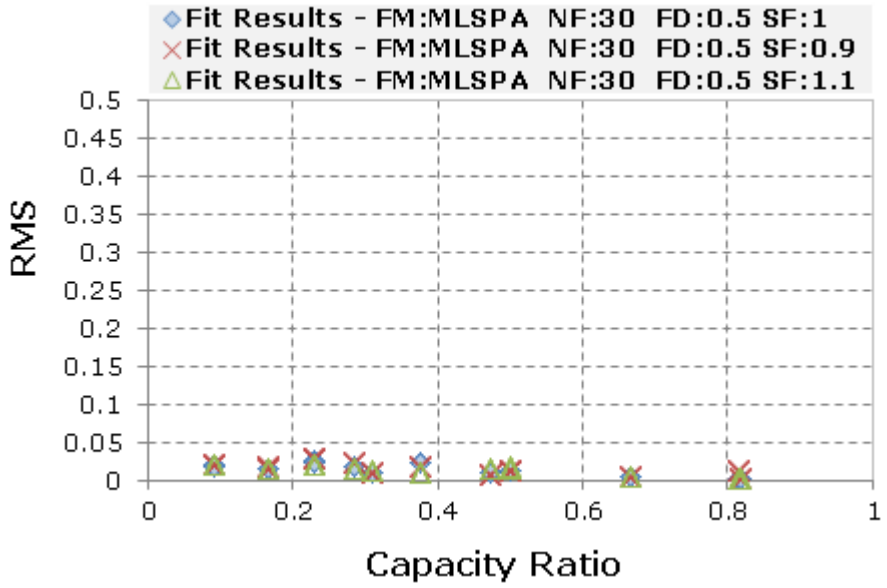

Figure C: Fit Results and Capacity Ratio for the Friesz-Harker Network Using MLSPA, 30 Functions and Function Distribution 0.5

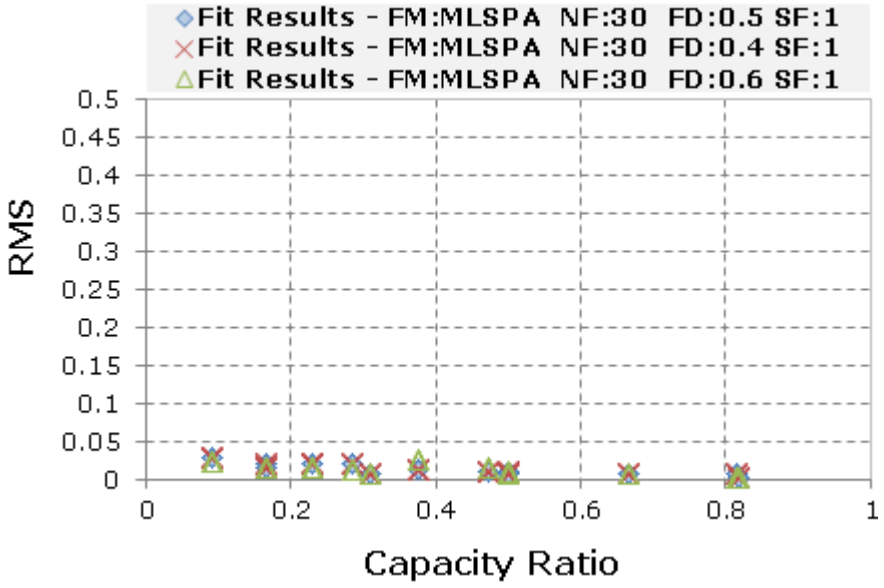

Figure D: Fit Results and Capacity Ratio for the Friesz-Harker Network Using MLSPA, 30 Functions Saturation Factor 1.0