The Dissertation Committee for Priyadarshan Nandkumar Patil
certifies that this is the approved version of the following dissertation:

Traffic assignment models – applicability and efficacy

Committee:

Stephen D. Boyles, Supervisor

John Hasenbein

Erhan Kutanoglu

Randy Machemehl

Traffic assignment models – applicability and efficacy

by

Priyadarshan Nandkumar Patil, M.S.E., B.Tech.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

July 2022

Dedicated to my *grandparents*

Abstract

Traffic assignment models – applicability and efficacy

Priyadarshan Nandkumar Patil, Ph.D.

The University of Texas at Austin, 2022

Supervisor: Stephen D. Boyles

This dissertation is concerned with the traffic assignment problem (TAP), an important tool in transportation planning. We first study a theoretical extension of TAP which incorporates symmetric interactions in the link travel costs. In particular, we prove the theoretical convergence of conventional solution methods for this problem, and analyze convergence behavior for these methods. We also show how a set of real world interactions such as merge models can be modeled using these type of symmetric interactions. Second, we apply these findings to a practical case study of railroad electrification, formulated as a rail network design problem. We solve this problem on a large network representing the North American railroad network, and analyze the solutions to provide policy recommendations. We model the interactions between diesel- and electric-goods flow as a symmetric congestion cost and a separable fuel/crew cost. Third, we study the empirical behavior of TAP under input uncertainty. Specifically, we analyze the effects of three types of input errors (uniform,

origin- or destination- specific, spatially correlated) on network metrics, such as total system travel time or congestion, at equilibrium. Empirical bounds for these output metrics are identified for various levels of input error. We apply these findings to a case study to demonstrate potential usage for planning purposes. Lastly, we conduct a comprehensive empirical study on the convergence behavior of traffic assignment convergence metrics. We analyze five commonly used network metrics at various convergence levels, for different solution algorithms, and identify concrete thresholds for convergence. We also show the relationship between different convergence criteria metrics, allowing for transfer of these thresholds to different metrics.

Table of Contents

List of Tables

List of Figures

xii

Chapter 1

Introduction

Traffic assignment is a common tool in transportation planning, and predicts how travelers will choose routes accounting for congestion effects. It is used in long-term planning, as the final step of the traditional four-step model, to assist in decision-making based on link flows, select link analysis, or shortest-path analysis. It also appears as a sub-problem in network design, toll-setting, and other related bi-level optimization problems. Despite many advances in dynamic traffic modeling, static assignment remains common in current practice. And despite advances in technology and algorithmic efficiency, computation times are still a relevant issue as agencies move to more detailed, multiclass models, or when assignment is a subproblem in a larger iterative scheme (feedback models, trip table estimation, network design, and so forth). This proposal document therefore focuses primarily on the static traffic assignment problem (TAP) as it is traditionally formulated.

1.1 Why is static traffic assignment still relevant?

Static traffic assignment has been studied for over five decades now, starting with the convex optimization formulation by Beckmann et al. (1956) and described at length in Patriksson (2015) and Boyles et al. (2020). With significant advances in traffic flow theory and traffic assignment in the interim, including the development of dynamic traffic assignment and micro-simulation techniques, it is worth asking whether the traditional traffic assignment problem is still worth studying. Despite the important roles that these other methods play in transportation analysis, there are still several settings where static assignment remains a valuable tool.

Well-known advantages of static assignment include a standard formulation, efficient and provably correct solution algorithms, and guarantees of equilibrium existence and uniqueness. The latter concerns are not strictly mathematical, but have important implications for practice — it is unclear how projects should be evaluated or ranked if multiple, potentially very different solutions exist, or none at all.

A less-appreciated advantage is its greater robustness to errors in input data, such as origin-destination matrices or link and node parameters. Many dynamic traffic assignment models feature queue spillback, which is a significant contributor to traffic congestion in the field. However, spillback introduces discontinuities into the assignment process, potentially amplifying any error or noise in the model inputs. Boyles and Ruiz Juri (2019) showed that when the error in the trip table is sufficiently large, a model *without* spillback actually produces a smaller absolute error in delay estimations than a model *with* spillback. Relatedly, static models are easier to calibrate; despite advances in travel demand modeling, forecasting a time-dependent trip table in a large network remains highly challenging.

Static assignment can also be solved in a shorter amount of time. Even as computational resources expand and more efficient algorithms are developed, in the amount of time required to run a single dynamic assignment it is possible to run multiple static assignments. In applications requiring hundreds or thousands of assignment runs — examples include Monte Carlo simulation to simulate distributions over input parameters (Waller et al., 2001; Zhao and Kockelman, 2002; Ukkusuri et al., 2007; Duthie et al., 2011), sensitivity analysis (Boyles, 2012; Jafari et al., 2017), trip table estimation (Yang, 1995; Lundgren and Peterson, 2008), network design (Yang, 1997; Yang and Bell, 1998; Josefsson and Patriksson, 2007), network pricing (Yang and Lam, 1996), and other bilevel optimization problems (Yin, 2000) — the computational advantages of static assignment are compelling, if for no other reason than a preliminary screening of alternatives to form a "shortlist" for more detailed modeling.

We lastly point out a recent line of research showing how a variety of static and dynamic models can be generalized into a single common framework (Bliemer et al., 2017; Bliemer and Raadsen, 2020), suggesting that research into one type of traffic assignment model may have relevance to the other as well.

For all the reasons above, static assignment remains a commonly-used tool in transportation planning practice. To be clear, none of this is to argue that static assignment should be universally applied. In applications where the input data are known with high precision, detailed congestion information is essential, and computation times are not constraining (e.g., present-day studies of work zone impacts), dynamic traffic assignment (DTA) or even

microsimulation are likely superior tools. Yet there remain applications where static assignment is preferred, as when inputs are poorly known, or when rapid assessment of a large number of alternatives is preferred to an in-depth assessment of a few (e.g., long-range regional planning, bilevel optimization), and it is such applications that this study has in mind.

1.2 Taking a fresh look at symmetric and asymmetric TAP

Symmetric TAP (S-TAP) and Asymmetric TAP (A-TAP) have had significant research interest in the 80s and 90s, but dwindled soon after. These models aimed to incorporate more realism in TAP, trying to model junction interactions, multi-class traffic, and traffic interactions. However, they were limited by the TAP advances and computational limitations of the time. A-TAP traded the advantages of TAP (convergence properties and uniqueness, optimization formulation, quick runtimes) while incorporating assumptions considered less realistic than DTA assumptions (queue spillback, merge models, etc.) This led to TAP and DTA being primary research foci.

TAP has since seen significant theoretical and computational advances, as well as widespread usage for planning. As we show, some of these advances are directly applicable to S-TAP and A-TAP, allowing retention of convergence properties and computational advances, while allowing modelers greater freedom. DTA has been the de-facto choice for "realistic" modeling, but studies show that input noise and queue spillback introduces significant errors (Zhang et al., 2013; Boyles and Ruiz Juri, 2019). Additionally, multiple applications require detailed modeling of select links/intersections only, where applying DTA for the entire network is excessive in terms of time, data, and resources required. This presents use-cases for S-TAP and A-TAP where these select links/intersections (freeways, on-ramps, flyovers, arterials, etc.) are modeled with significant interactions, and the rest (such as neighborhood roads and other minor links) are modeled at a high level.

TAP has also seen advances incorporating queueing behavior and traffic flow theory assumptions. Huntsinger and Rouphail (2011) conduct link performance function analysis based on queueing and bottleneck analysis. Bliemer et al. (2012) present a quasi-dymanic traffic assignment, i.e., static traffic assignment with queueing and hierarchical clustering analysis of historical data. Bliemer and Raadsen (2019) present TAP with residual queues and spillback, further adding traffic dynamics into TAP. These advances, in conjunction with link interactions, provide reason for reconsideration of S-TAP and A-TAP research.

1.3   Notation and Formulation

Consider a directed network with a set of links A, and a set of zones Z. For each link $(i, j) \in A$, let l_{ij} denote its physical length, and t_{ij} its travel time, assumed to be a function of its flow x_{ij} alone. For each origin $r \in Z$ and destination $s \in Z$, let d_{rs} denote the demand for travel between these zones, and let Π^{rs} denote the set of network paths connecting these zones. Further, let Π be the set of all network paths. For a given path π, the number of travelers choosing that path is given by h_π.

The classical formulation of TAP identifies a network state which reflects traveler behavior (all travelers choose a shortest path between their origin and destination) and congestion effects (these shortest paths depend on the choices made by other travelers). Under mild regularity assumptions, such a network state can be identified by solving the following convex program (Beckmann et al., 1956):

$$\min_{\mathbf{x}, \mathbf{h}} \quad \sum_{(i,j) \in A} \int_0^{x_{ij}} t_{ij}(x) dx \tag{1.1}$$

subject to:

$$x_{ij} = \sum_{\pi \in \Pi : (i,j) \in \pi} h_\pi \qquad \forall (i, j) \in A \tag{1.2}$$

$$\sum_{\pi \in \Pi^{rs}} h_\pi = d_{rs} \qquad \forall (r, s) \in Z^2 \tag{1.3}$$

$$h_\pi \geq 0 \qquad \forall \pi \in \Pi \tag{1.4}$$

If the link performance functions are strictly increasing, the objective function is strictly convex, and thus has a unique minimum solution in the link flows, which we denote by $\mathbf{x}^*$. This solution is called the user equilibrium (UE) state. We say that a path π is *used* in a solution to TAP if h_π is strictly positive, and define $\Pi_+(\mathbf{h})$ to be the set of used paths at a given solution.

Static TAP can also be formulated as a variational inequality in path flows. A path flow vector $\hat{h}$ satisfies the principle of user equilibrium *if and only if* it satisfies the following variational inequality:

$$c(\hat{\mathbf{h}}) \cdot (\hat{\mathbf{h}} - \mathbf{h}) \leq 0 \qquad \forall \mathbf{h} \in \Pi \tag{1.5}$$

In general, the UE path flow solution is not unique, since many path flow vectors $\mathbf{h}$ can generate the same link flow vector $\mathbf{x}$. The most likely path flows are the unique solution (denoted $\mathbf{h}^*$) to the following optimization problem, representing entropy maximization (Rossi et al., 1989):

$$\max_{\mathbf{h}} \quad -\sum_{(r,s)\in Z^2} \sum_{\pi\in\Pi^{rs}} h_\pi \log h_\pi \tag{1.6}$$

subject to:

$$\sum_{\pi\in\Pi:(i,j)\in\pi} h_\pi = x_{ij}^* \qquad \forall(i,j)\in A \tag{1.7}$$

$$\sum_{\pi\in\Pi^{rs}} h_\pi = d_{rs} \qquad \forall(r,s)\in Z^2 \tag{1.8}$$

$$h_\pi \geq 0 \qquad \forall\pi\in\Pi \tag{1.9}$$

Note the constraint that the path flows must generate the UE link flows $\mathbf{x}^*$. The most likely path flows use as many paths as possible given the user equilibrium state (Bar-Gera, 2006). Additionally, the formula for entropy calculation can be expressed in terms of link flows obtained from an origin-based assignment as follows:

$$E(\mathbf{x}) = -\sum_{p\in Z} \sum_{(i,j)\in A} x_{(i,j),p} \log\left(\frac{x_{(i,j),p}}{x_{j,p}}\right) \tag{1.10}$$

where, $x_{(i,j),p}$ is the flow on link (i,j) from origin p and $x_{j,p}$ is the flow through node j originating at p.

Now instead assume that the link performance functions depend on multiple links' flow. For generality, we write $t_a(\mathbf{x})$ to express dependence on (potentially) every other link flow, although in practice each link's travel time depends only on a few other links. If these functions are differentiable, and the Jacobian of $\mathbf{t}$ with respect to $\mathbf{x}$ is symmetric, equilibrium flows correspond to stationary points of the function

$$\min_{\mathbf{x},\mathbf{h}} \oint_0^{\vec{x}} t_i(\vec{s})d\vec{s} \tag{1.11}$$

relative to the same constraints as TAP, as can be seen by writing the optimality conditions. We will refer to this optimization problem as S-TAP. The symmetry requirement on the

Jacobian is critical. Otherwise, the line integral is path-dependent and the function as stated is not well-defined, and the model would need to be formulated and solved using variational inequality (VI) methods (Facchinei and Pang, 2003). We will refer to the asymmetric case as A-TAP.

If this Jacobian is additionally positive definite, then we say that the link performance functions are *monotone* (note that this is stronger than requiring each t_a to be monotone in each link flow separately). In such a case, the objective (1.11) is a strictly convex function (its Hessian is the Jacobian just described), and equilibrium again exists and is unique. If the link performance functions are not monotone, the function (1.11) is not convex, and multiple stationary points (and therefore equilibria) may exist, with the minima corresponding to stable equilibria. For example, see Figure 1.1. There are three equilibria corresponding to the flow vectors (0,10), (5,5), and (10,0); the first and third of these are stable, and the second is unstable. For derivations of the above results and more discussion, see the books by Patriksson (2015) and Boyles et al. (2020).
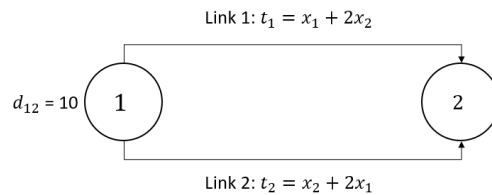
(a) Example network
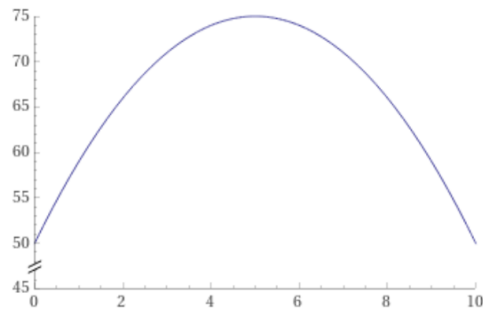
(b) Objective function visualization

Figure 1.1: S-TAP example with multiple extreme points

The optimization formulation in Equation 1.11 is not valid for A-TAP. As stated before, the line integral is path dependent in the absence of a symmetric Jacobian matrix,

invalidating the optimization formulation. Therefore, the following variation inequality formulation is used. The equilibrium flow vector x^* satisfies the following inequality for all feasible flow vectors x.

$$t(\mathbf{x^*}) \cdot (\mathbf{x^*} - \mathbf{x}) \leq 0 \qquad \forall \mathbf{x} \in \mathbf{X} \qquad (1.12)$$

Under conditions of positive definite Jacobian (strict monotonicity of cost vectors), this VI has a unique solution. S-TAP and A-TAP can also be formulated as a variational inequality in path flows, satisfying the formulation in Equation 1.5

1.4 Overview and contributions

The rest of the proposal document is organized as follows. Chapter 2 presents the state-of-the-art for static TAP, S-TAP, and A-TAP solution methods, and existing guidance about convergence criteria. Chapter 3 presents the proof of convergence for convex combination algorithms under symmetric link interactions, and the flow shift formula derivation for path flow equilibrating algorithms. This chapter also explores the computational behavior and effects of link interaction degree for S-TAP and A-TAP, and presents specific situations (merges) where DTA models are used to construct S-TAP/A-TAP models. Chapter 4 formulates the budget-constrained rail network electrification problem as a bi-level optimization problem, where the lower level problem is a symmetric traffic assignment problem for goods flow. This is a novel way of looking at the rail network electrification problem and connecting it to the traffic assignment problem. The costs and network parameters incorporate electrification costs; fuel, locomotive, and operational costs; and train resistance costs. The North American railroad network is used to demonstrate our heuristic and draw insights.

Chapter 5 characterizes of the effect of uniform, OD-specific, and spatially correlated demand errors on three common network metrics (equilibrium system travel time, system congestion, and vehicle miles traveled) for static TAP. We also observe the effects on a case study network, showing potential for planning purposes. Chapter 6 investigates the convergence behavior of five commonly used network performance metrics for static TAP, and we observe thresholds for "convergence" of various metrics. We also show results using these thresholds for various TAP scenarios (multi-class assignment, bi-level programs, etc.) as well as the relationship between three common convergence metrics. Lastly, we conclude with a summary of the key contributions and avenues for future work.

Chapter 2

Background

2.1 Static TAP

The convex programming formulation for TAP was proposed by Beckmann et al. (1956) based on Wardrop's user equilibrium principle (Wardrop and Whitehead, 1952). Adaptations of general convex program methods as well as evolution of specialized algorithms led to a variety of solution methods for TAP. Books by Patriksson (2015) and Boyles et al. (2020) are good resources for detailed reading on the evolution of the field as well as state-of-the-art. The major advances in the solution methods are detailed below.

The first class of solution methods for static TAP is link based methods. Given a current flow solution, the algorithms identify a target link flow solution, update flows using a step size, and then check for convergence. Until the convergence criteria is satisfied, the algorithms keeps operating. The advantages are low memory requirements and ease of implementation, but they require a long time to converge on large networks. The different algorithms differ in their choice of search direction and step size. MSA (Powell and Sheffi, 1982) chooses the current shortest path all-or-nothing assignment as target flows, with a fixed step size. Frank-Wolfe algorithm (Frank and Wolfe, 1956) chooses the step size adaptively, solving a restricted VI with feasible set as the line segment between current and target flows. Conjugate FW (Mitradjieva and Lindberg, 2013) algorithm improves the target flow selection by choosing the new search direction as conjugate (orthogonal) to previous search directions. Powell and Sheffi (1982) provide four necessary conditions for algorithm

convergence using predetermined step sizes in addition to the proof of convergence for MSA.

Algorithm 1: Framework for link-based algorithms

Generate an initial feasible link flow solution;

while *convergence criteria not met* **do**

> Generate a target link flow solution;
>
> Update the current link flow solution using some step size;
>
> Calculate the updated link travel times and convergence criteria;

end

Path-based algorithms evolved with improvements in computation power and memory, aiming to tackle the disadvantages of link-based methods. Two renowned algorithms are gradient projection (Jayakrishnan et al., 1994) and projected gradient (Florian et al., 2009). Path based methods track the used path sets, find the shortest paths every iteration, and shift flows within the path set for each origin-destination (OD) pair, repeating these steps till convergence. The gradient projection algorithm moves towards the negative of the gradient direction (direction of steepest ascent) and then uses projection to ensure feasibility. The projected gradient algorithm projects the gradient direction onto the solution set polyhedron, and then utilizes this search direction for target flows. Some other algorithms in this class are disaggregate simplicial decomposition algorithm (Larsson and Patriksson, 1992), path based FW (Chen et al., 2002), conjugate gradient projection (Lee et al., 2003), slope-based methods (Kumar and Peeta, 2010, 2014), and greedy path-based approach (Xie et al., 2018).

Algorithm 2: Framework for path-based algorithms

Initialize the working path set for each OD pair;

while *convergence criteria not met* **do**

> **for** *each OD pair* **do**
>
> > Find the shortest path from origin to destination. Add it to working path set if not already included;
> >
> > Shift flows within the working path set to get closer to equilibrium;
> >
> > Calculate the updated link travel times and convergence criteria;
> >
> > Drop zero flow paths from the set
>
> **end**

end

Bush-based (or origin-based) algorithms were developed to efficiently store path

information while improving computation time. They store network information in connected and directed acyclic graphs rooted at each zone, reducing memory requirements and allowing for easy decomposition into path flows. Algorithm B proposed by Dial (2006) selects the shortest and longest paths within a bush to each destination node and shifts flow using Newton's method, iterating over all bushes till convergence. Origin based assignment (Bar-Gera, 2002) tweaks the general structure of Algorithm B by allowing for flow shift across many paths simultaneously, as opposed to just the shortest and longest path. Nie (2012) demonstrated improvements to OBA by a different Hessian estimation. Linear user cost equilibrium (Gentile, 2014) is extremely similar to OBA, but adds a linearization and localization step when scanning nodes, trying to solve a local user equilibrium problem.

Algorithm 3: Framework for bush-based algorithms

Initialize the bushset with an initial bush for every zone (potentially the shortest path tree);

while *convergence criteria not met* **do**

Shift flows within the each bush to bring each origin closer to equilibrium;

Modify the bushes by adding links that reduce travel times and removing unused links;

Calculate the updated link travel times and convergence criteria;

end

The last algorithm to be discussed is titled "Traffic assignment by paired alternative segments" or TAPAS (Bar-Gera, 2010). It is a recent algorithm which both solves TAP and provides a path flow solution satisfying *proportionality*, a slightly weaker condition than entropy maximization. TAPAS differs from path-based and other bush-based algorithms by storing path segments used by multiple origins and then re-using information about PASs across iterations. PAS travel time equations also help concisely represent equilibrium conditions. TAPAS was further improved by Xie and Xie (2014) and presented as i-TAPAS. TAPAS, and related algorithms derived from it, are shown to be highly efficient (Xie and Xie, 2015). Xie and Xie (2015) also present an in-depth comparison among these bush-based algorithms (and a few others) and their computational performance.

2.2 S-TAP and A-TAP

Historically, Prager (1954) first mentioned the need to model traffic interactions on a two-way street. Dafermos (1971) and Dafermos (1972) were the first to formulate S-TAP and

A-TAP, showing equivalence with the multi-class TAP, and presenting an iterative flow update algorithm to obtain user equilibrium and system optimal flows. Sender and Netter (1970) used the fixed point theorem to show the existence of an equilibrium solution in the asymmetrically formulated multi-class TAP. Smith (1979) was the first to formulate A-TAP as a VI, presenting a set of conditions for existence (continuity of cost vectors) and uniqueness (strict monotonicity of cost vectors for all supply feasible vectors) of the equilibrium solution. Dafermos (1980) showed that these uniqueness conditions were equivalent to the Jacobian being positive definite, and proved the existence of an equilibrium solution. As checking positive definiteness can be cumbersome to test in practice, Heydecker (1983) proposed an easier test, based on diagonal dominance. This is a weaker condition than positive definiteness, thus showing that positive definiteness of the Jacobian is a sufficient but not necessary condition. Heydecker also discussed the existence of multiple (and unstable) equilibrium solutions when said conditions are violated. The equivalent requirement of positive definite Jacobian for the multi-class TAP solution existence was also shown by Braess and Koch (1979).

The VI formulation allowed for a number of different solution approaches to be proposed for A-TAP, including the non-linear Jacobi method (or diagonalization method), projection methods, and column generation methods. Dafermos (1982) proposed a relaxation method and proved its convergence. Fisk and Nguyen (1982) analyzed this approach, the projection method, and three other solution methods, concluding that the non-linear Jacobi method was most efficient. Nagurney (1984, 1986) extended this comparison of the relaxation and projection method for multi-modal problem with varying travel costs and A-TAP, concluding that no one method was uniformly superior. Smith (1983) modeled junction interactions using A-TAP, while proposing a new objective function measuring deviation of traffic distribution from equilibrium. Smith also provided a descent direction to monotonically reduce the new objective function and a path enumeration based algorithm for A-TAP.

Florian and Spiess (1982) provided a sufficient condition for diagonalization algorithm convergence. Nguyen and Dupuis (1984) proposed an iterative approach for A-TAP post-optimizing a linear sub-problem at each iteration. Lawphongpanich and Hearn (1984) used simplicial decomposition constructing the solution as a convex combination of all extreme points of the flow vector space. They compared their approach to the Nguyen and Dupuis (1984) approach on the networks proposed in Nguyen and Dupuis (1984) and Fisk and Nguyen (1982), concluding that the simplicial approach is competitive with the iterative

11

approach while storing a small number of extreme flow patterns. In the same vein, Gabriel and Bernstein (1997) and Bernstein and Gabriel (1997) solved A-TAP with non-additive route costs using column generation for nonlinear complementarity problem (NCP) gap minimization, equivalent to the VI formulation. The NCP formulation is needed when route-level interactions are modeled, such as non-additive route cost interactions. They also showed the existence and uniqueness of the solution if the function is separable. Lo and Chen (2000) also apply column generation to the reformulated NCP for route-specific tolls.

Mahmassani and Mouskos (1988) tested the diagonalization approach on three networks including the Texas highway network, and observed convergence of the algorithm despite violation of sufficient conditions presented in Dafermos (1982). Similar observations were made by Friesz et al. (1984) and Friesz (1985). Mahmassani and Mouskos also compared their implementations with Sheffi's streamlined implementation (Sheffi, 1985) with one FW iteration per subproblem and concluded that there is no single best choice for the number of FW iterations. Meneguzzer (1995) provided an overview of the advances in the field of diagonalization for A-TAP and convergence for explicit modeling of intersections. Marcotte and Guélat (1988) applied the modified Newton method to A-TAP, comparing its performance with cutting plane methods and diagonalization, observing their method performing better than other methods for complex asymmetric interactions. Dupuis and Darveau (1986) assessed the convergence conditions for A-TAP solutions using projection and diagonalization methods.

Hearn et al. (1984) drew a connection between convex programming formulation of A-TAP and the VI formulation. Marcotte and Wynter (2004) relaxed the monotonicity condition for interactions, proposing weaker convergence conditions for A-TAP with multiple modes. Wong et al. (2001) modeled A-TAP with simulation approach for the intersection delay. Panicucci et al. (2007) formulated the VI in terms of path flows and propose a column generation scheme based on Khobotov's method. Yook and Heaslip (2016) discussed ways to accelerate the convergence of double-projection method proposed by Panicucci et al. using the decomposable path flow VI structure. De Grange and Muñoz (2009) presented a method to equalize line integral paths for affine cost A-TAP. Chen et al. (2011) modeled A-TAP interactions using side constraints. Sancho et al. (2015) evaluated the performance of five projection methods for A-TAP and observe that a variation of Khobotov's method proposed by He et al. (2012) shows the fastest convergence. Patriksson (2015) and Yook (2014) provide good overviews of alternative solution approaches for A-TAP.

As suggested by the above review, there was significant focus on A-TAP research

in the 1980s and 1990s, but comparatively less since then. The broad explanation for this is the emergence of DTA as a serious modeling tool (Peeta and Ziliaskopoulos, 2001; Chiu et al., 2011), with better grounding in traffic science than link performance functions could ever hope to have, whether separable or with interactions.

2.3   Convergence criteria and network metrics

Algorithms for traffic assignment converge to an equilibrium solution in the limit, so a convergence criterion must be introduced to ensure output in finite time. Rose et al. (1988) considered the convergence behavior of the Frank-Wolfe algorithm on small networks (16 nodes). Their study used the relative gap metric (a common convergence criterion, defined below) based on the duality gap. They concluded that it was very difficult to obtain precise estimates of the equilibrium flows in networks which contain only a few O-D pairs with overlapping paths, and called for more research on convergence behavior. Boyce et al. (2004) found that link flows in the Philadelphia network stabilized once the relative gap (a different definition, also defined below) was below 10^{-4}.

The primary guidance to date is based on Boyce et al. (2004), and a relative gap level of 10^{-4} or 10^{-5} is common in current software as a default convergence criterion. The manual for Caliper's TransCAD software further suggests that "since traffic assignment problems vary in many dimensions, some experimentation is warranted to arrive at how much convergence is enough." (Caliper Corporation, 2018). While the study by Boyce et al. (2004) played a critical role in determining the necessary level of precision, its experiments were conducted on a single network and considered a single metric (freeway link flows).

Aggregate measures, such as total system travel time or vehicle-miles traveled, are used to capture the overall state of a network (Harrison et al., 2006; Weisbrod, 2008; Higgins, 2013; Moudon and Stewart, 2013). For instance, the North Carolina Department of Transportation strategic plan uses total system travel time to monitor network performance (North Carolina Department of Transportation, 2015), Litman (2016) uses vehicle-miles traveled as a sustainability indicator, Qian and Zhang (2012) use total system travel time and vehicle miles travelled as factors to compare interstate closure scenarios in Sacramento, and the California Department of Transportation uses a reduction in vehicle-miles traveled as a strategic target (California Department of Transportation, 2015). A few other examples include usage in analysis of delivery vehicle impact (Holguín-Veras et al., 2013), credit- or permit-based demand management (Lessan and Fu, 2019) and within network design

problems as project selection criteria (Shayanfar and Schonfeld, 2019; Gokalp et al., 2021; Patil et al., 2022). Disaggregate measures, such as link and path flows, more finely describe the impacts of projects on specific regions and populations. Such measures are commonly used by many practitioners and researchers (Bureau of Transportation Statistics, 2015; US Department of Transportation, 2016, 2017; Seattle Department of Transportation, 2016; Maryland Department of Transportation, 2018; Cherlow, 1981; Daniels et al., 1999; Astroza et al., 2017; Boyles et al., 2018).

2.4   Why are runtimes still an issue?

With advances in computing and solution algorithms, it is worth asking whether run times are still relevant in static assignment, particularly given the time frame of long-term planning. As discussed in the previous section, there are applications requiring large number of assignment runs, often with traffic assignment as a subproblem in an iterative scheme. For this reason, research continues in assessing and improving the computational performance of static assignment (Galligari and Sciandrone, 2019; Schneck and Nökel, 2020).

Furthermore, as computation power advances, network models have increased in scope and resolution. Regional planning models today commonly include tens of thousands of links and nodes, multiple user classes, and feedback to earlier modeling stages to ensure consistency. For large metropolitan areas, even using cutting-edge software and hardware, it is not uncommon for a single model run to take several hours. For a single scenario, this may be acceptable; as part of a bilevel trip table optimization requiring thousands of runs it is clearly not.

Chapter 3

Symmetric and asymmetric traffic assignment algorithm convergence

3.1 Introduction

Traffic assignment is a cornerstone of the urban planning and forecasting process, relating travel demand to a forecasted network loading, providing metrics such as link flows and travel times. Many traffic assignment models exist, as discussed in Chapter 2. Broadly speaking, there is a tension between the level of "realism" of a model (used here to refer to the level of detail in capturing traffic physics) and how "nice" the model is (in terms of mathematical analysis, computational efficiency, provably correct algorithms, and properties such as equilibrium existence, uniqueness, and stability). Dynamic link and node models are much more descriptive of traffic flow than the volume-delay functions typical of static assignment, a powerful argument in their favor. At the same time, dynamic equilibrium need not exist (Boyles et al., 2020, Section 11.3.1), or several may exist (Nie, 2010); and precisely the same features that make dynamic models more realistic may also make them more sensitive to any errors in input data, complicating calibration and possibly introducing more error than is saved by improving the traffic flow model (Boyles and Ruiz Juri, 2019). These are not merely theoretical concerns — if dynamic equilibrium may not exist at all, or if multiple equilibria exist, it is not at all clear how to use the results of such a model for alternatives analysis; and if a model is highly sensitive to parameters which are hard to estimate (such as a time-dependent OD matrix decades in the future) its practical utility is heavily limited. More extended versions of these issues are discussed in Bar-Gera (2010),

Boyles and Ruiz Juri (2019), and Patil et al. (2021); in short, some applications demand the realism of a dynamic model, while in other applications the advantages of static models outweigh their drawbacks, and careful researchers and practitioners select their tool based on the problem at hand.

This ongoing conversation serves as background for our investigation. Some researchers have attempted to find unifying frameworks for static and dynamic assignment; see Bliemer et al. (2012) and Bliemer and Raadsen (2019) for examples. An alternative approach is to improve the traffic model in static assignment. In the 1980s, there was an active line of research in the static assignment problem with *interactions* among links, rather than using separable link performance functions. (This background material is described at greater length in Chapter 2.) While research in this area has continued to this day, much of the community's attention shifted to dynamic assignment in the 1990s, with the advent of reasonable link and node models, such as the cell transmission model (Daganzo, 1994, 1995). In recent decades, our knowledge of good dynamic link and node models has advanced further still (Yperman, 2007; Tampère et al., 2011). Research has also advanced considerably in how the separable TAP is solved, with the discovery of path- and bush-based algorithms (Jayakrishnan et al., 1994; Bar-Gera, 2002; Dial, 2006; Bar-Gera, 2010; Xie and Xie, 2014; Chen et al., 2020).

We believe there are several reasons to take a fresh look at the static traffic assignment problem with symmetric link interactions, considering it at least as an alternative to static assignment, if not dynamic. As is known, with monotone cost functions, S-TAP retains most of the favorable properties of TAP, including formulating the equilibrium problem as a convex program, and the resulting features of solution existence, uniqueness, and algorithmic tractability. In this chapter, we accomplish the following:

1. As an example of how existing node models can be approximated by symmetric, monotone link performance functions, we develop such a representation of the Jin-Zhang merge model (Jin and Zhang, 2003). (Section 3.2)

2. We discuss solution algorithms for S-TAP, including classic convex combinations methods, but focusing mainly on more recent algorithms based on shifting flow between a pair of alternative segments. We show that the flow shift formula for S-TAP takes a familiar and simple form, and therefore existing algorithms for TAP can be easily adapted for S-TAP. (Section 3.3).

3. We implement these algorithms on standard test networks, and show that in most

cases, S-TAP actually converges *faster* than the separable TAP. Therefore, S-TAP can be considered as a serious alternative to TAP in planning practice. We also report preliminary results showing good performance of gradient projection for asymmetric link interactions, even though there is no convex objective function (and therefore no guarantees of convergence).

Section 3.6 summarizes our findings, and provides specific suggestions for further investigations on the utility of S-TAP.

3.2 A symmetric, monotone merge model

While much research has been done on theoretical properties of traffic assignment with interactions, we are not aware of specific guidance on exactly *how* interactions should be chosen to represent real-world scenarios. To further motivate our investigation, we will give an illustration of how symmetric, monotone link performance functions can be chosen to approximate a node model used in dynamic traffic assignment. The method we describe here is surely not the only way to do this, but a full study of how to develop such approximations (and to assess their quality on full-sized, realistic networks) is beyond the scope of this study. Our aim in this section is simply to demonstrate that S-TAP is a plausible model for certain applications.

We take as our starting point a simple network loading model, a network of point queues. In this model, the time required to travel each link is a fixed free-flow time (t^0), plus time spent waiting in a queue at the downstream end. Let $Q(\tau)$ denote the length of the queue at time τ; these queues are "point" queues in that they may grow arbitrarily long. Each link also has a uniform saturation flow u, perhaps proportional to the number of lanes. We will denote by $x(\tau)$ the *inflow* rate at the link's upstream end, and $y(\tau)$ the *outflow* rate at the link's downstream end. If x and y are time-invariant, we must have $x \geq y$, and the queue length at a given point in time will be $Q(\tau) = \tau(x - y)$. The travel time experienced for a vehicle entering the link at time t will be $t^0 + \tau(x/y - 1)$. If we assume that x is constant over an interval of length T (and zero otherwise), the average delay experienced by a vehicle on the link will be $t^0 + \frac{T}{2}(x/y - 1)$. If we choose units so that $T = 2$, the formula simplifies to $t^0 + (x/y - 1)$, which we will adopt for the remainder of the section.

Consider now a merge node with two upstream links (indexed 1 and 2) and one downstream link (indexed by 3). A "merge model" takes as input the demands from each

upstream link and the available supply from the downstream link, and produces the flow rates out of each upstream link into the downstream link. Several merge models have been proposed in the literature. (Daganzo, 1995; Lebacque, 1996; Jin and Zhang, 2003)

We now show that the Jin-Zhang merge model (Jin and Zhang, 2003) leads to a symmetric, monotone delay function, if the demands are interpreted as the inflows x.[1] With this interpretation, the model takes the following form. If $x_1 + x_2 \leq u_3$, there are no queues, and hence no delays: $y_1 = x_1$, $y_2 = x_2$, so $t_1 = t_1^0$ and $t_2 = t_2^0$. Otherwise, the Jin-Zhang model allocates flows proportionally to demands: $y_1 = u_3 x_1 / (x_1 + x_2)$, and hence $t_1 = t_1^0 + [(x_1 + x_2)/u_3 - 1]$. Likewise, $t_2 = t_2^0 + [(x_1 + x_2)/u_3 - 1]$. In either case, the Jacobian of $\mathbf{t}$ with respect to $\mathbf{x}$ is symmetric and positive semidefinite.

Alternative merge models, such as that of Daganzo (1995), do not directly lead to symmetric and monotone performance functions. However, it may be possible to create reasonable approximations to them that satisfy these conditions (indeed the Jin-Zhang model may be seen as such an approximation), at least in a region of demand and supply values the modeler believes to be likely at a particular junction. For instance, the Daganzo merge violates symmetry only in the exceptional case when a queue forms on one upstream link, not both. We believe it plausible that similar procedures or approximations can apply to other types of interactions between links, based either on node models from dynamic traffic assignment, formulas from the Highway Capacity Manual (2010) or similar literature, or regression from simulation, but we leave such investigation to future study.

3.3 Algorithms and convergence

This section discusses solution algorithms for S-TAP with monotone cost functions. To do so, it will be convenient to choose a specific integration path for the line integral in (1.11). If we choose the path $(0, 0, 0, \ldots, 0) \rightarrow (x_1, 0, 0, \ldots, 0) \rightarrow (x_1, x_2, 0, \ldots, 0) \rightarrow \cdots \rightarrow (x_1, x_2, \ldots, x_n)$, the line integral decomposes into a sum of ordinary integrals:

$$F(\mathbf{x}) = \sum_{i=1}^{n} \int_{(x_1, \ldots, x_{a-1}, 0, 0, \ldots, 0)}^{(x_1, \ldots, x_{a-1}, x_a, 0, \ldots, 0)} t_i(x) \, dx \, . \tag{3.1}$$

---

[1]This is an approximation to the actual dynamic model, since if a queue forms, the demand will increase to the saturation flow u. Indeed, one critique of the Jin-Zhang model is that it is unstable with respect to this transition, a violation of the "invariance principle," cf. Lebacque and Khoshyaran (2005).

Each link performance function is increasing for each independent flow variable, so each integral is convex, as is the sum.

With this representation, the gradient and Hessian of F take simple forms. Using the Leibniz rule, the derivative of F with respect to any link flow x_a is

$$\frac{\partial F}{\partial x_a} = t_a(x_1, \ldots, x_a, 0, \ldots, 0) + \sum_{i>a} \int_{(x_1,\ldots,x_{i-1},0,\ldots,0)}^{(x_1,\ldots,x_{i-1},x_i,0,\ldots,0)} \frac{\partial t_i}{\partial x_a}(x) \, dx .$$

But using the symmetry condition, $\partial t_i/\partial x_a = \partial t_a/\partial x_i$, and so we ultimately have $\partial F/\partial x_a = t_a$, from the fundamental theorem of calculus. That is, the gradient of F is simply the vector of link travel times, as it is in TAP (and indeed, as it should be for the optimality conditions to express equilibrium). The Hessian of F is then just the Jacobian of the link performance functions, that is, $(HF)_{ab} = \partial t_a/\partial x_b$.

3.3.1 Convex combinations algorithms

Convex combinations algorithms operate on the link flow vector $\mathbf{x}$, iteratively combining a current feasible solution with a "target" solution $\mathbf{x}^*$, with an update rule of the form $\mathbf{x} \leftarrow (1 - \lambda)\mathbf{x} + \lambda\mathbf{x}^*$. Typically $\mathbf{x}^*$ is an "all-or-nothing" solution obtained by placing all demand on shortest paths when the link costs are $\mathbf{t}(\mathbf{x})$. They are relatively naïve, but amenable to parallelization, and they typically make excellent progress in their initial iterations before the rate of convergence slows sharply (ultimately, to a sublinear rate).

The simplest possible convex combinations algorithm is the method of successive averages, where the step sizes λ are chosen *a priori* in a divergent sequence (but with $\sum \lambda^2$ finite); a common example is $\{1/2, 1/3, 1/4, \ldots\}$. Convergence of this method for S-TAP can be shown using the following result:

Proposition 3.3.1 (Powell and Sheffi, 1982) *Consider a twice-continuously differentiable convex function $F(\mathbf{x})$, and a sequence $\{\mathbf{x_0}, \mathbf{x_1}, \ldots\}$, where $\mathbf{x_0} \in X$ and $\mathbf{x_i} = (1-\lambda_i)\mathbf{x_{i-1}}+\lambda_i\mathbf{x_i^*}$ for $i \geq 1$, with $\mathbf{x_i^*} \in X$, $\lambda_i \in [0, 1]$, $\sum \lambda_i = \infty$, but $\sum \lambda_i^2 < \infty$. This sequence converges to a minimizer $\bar{\mathbf{x}}$ of F if the following conditions hold:*

1. *The inner product $(\nabla F(\mathbf{x_i}))^T (\mathbf{x_i^*} - \mathbf{x_i})$ is negative whenever $F(\mathbf{x_i}) > F(\bar{\mathbf{x}})$.*

2. *The values of $(\mathbf{x_i^*} - \mathbf{x_i})^T HF(\mathbf{x_i} + \psi\lambda_i(\mathbf{x_i^*} - \mathbf{x_i}))(\mathbf{x_i^*} - \mathbf{x_i})$ are bounded over all i and $\psi \in [0, 1]$.*

19

Each condition can be checked easily. If $\mathbf{x_i^*}$ is an all-or-nothing assignment to shortest paths, then $\mathbf{t}(\mathbf{x_i})^T (\mathbf{x_i^*} - \mathbf{x_i}) \leq 0$, with equality only if $\mathbf{x_i}$ is an equilibrium. Since $\nabla F(\mathbf{x}) = \mathbf{t}$, and since equilibria correspond to minima of F, the first condition is satisfied. Since X is compact, continuity of the link performance functions ensures that the elements of HF and that the magnitudes of $\mathbf{x_i^*} - \mathbf{x}$ are bounded, and the second condition is satisfied as well.

As a convex program, monotone S-TAP can also be solved by the Frank-Wolfe algorithm, which selects each $\lambda_i \in [0, 1]$ to minimize the value of the objective. A proof of convergence can be found in Section 2.2.2 of Bertsekas (2016).

3.3.2 Algorithms equilibrating paths

In many path- and bush-based algorithms, the fundamental operation involves equilibrating two paths: given a lower-cost path π_L and a higher-cost path π_U connecting the same origin and destination, shifting flow from π_U to π_L to either (approximately) equalize their costs, or to shift all flow onto π_L. Examples of such algorithms are the gradient projection method of Jayakrishnan et al. (1994), Algorithm B (Dial, 2006), and TAPAS (Bar-Gera, 2010).[2]

Given these paths, the question is how much flow Δx to shift from π_U to π_L. For TAP, Newton's method is commonly used to estimate the value of Δx that equalizes the path costs; this is also the value of Δx minimizing the Beckmann function. The same applies for S-TAP, although the scaling factor in the denominator must change to reflect link interactions, as we now show.

Let $\mathbf{x}(\Delta x)$ denote the link flows after Δx has been shifted away from π_U onto π_L. The only links whose flows will change are those in π_L or π_U, but not both; let A_L and A_U respectively denote the links only in π_L and π_U. Then

$$x_a(\Delta x) = \begin{cases} x_a + \Delta x & \text{if } a \in A_L \\ x_a - \Delta x & \text{if } a \in A_U \\ x_a & \text{otherwise} \end{cases} . \tag{3.2}$$

---

[2]TAPAS actually performs such shifts for multiple origin-destination pairs simultaneously, with alternative paths differing on the same segments of links; this point does not affect the discussion here.

To find the value of Δx minimizing the S-TAP objective F, we find where the derivative

$$\frac{dF}{d\Delta x} = \sum_a \frac{\partial F}{\partial x_a} \frac{dx_a}{d\Delta x} = \sum_{a \in A_L} t_a - \sum_{a \in A_U} t_a \qquad (3.3)$$

vanishes. As in the separable case, this is exactly when π_L and π_U have equal cost.

To apply Newton's method, we also need the second derivative of F with respect to Δx, in order to scale the step size properly. We calculate

$$\frac{d^2 F}{d(\Delta x)^2} = \sum_a \sum_{a'} \frac{\partial^2 F}{\partial x_a \partial x_{a'}} \frac{dx_a}{d\Delta x} \frac{dx_{a'}}{d\Delta x}$$

$$= \sum_{a \in A_L} \sum_{a' \in A_L} \frac{\partial t_a}{\partial x_{a'}} + \sum_{a \in A_U} \sum_{a' \in A_U} \frac{\partial t_a}{\partial x_{a'}} - \sum_{a \in A_L} \sum_{a' \in A_U} \frac{\partial t_a}{\partial x_{a'}} - \sum_{a \in A_U} \sum_{a' \in A_L} \frac{\partial t_a}{\partial x_{a'}}. \qquad (3.4)$$

Using the symmetry condition, this simplifies to

$$\frac{d^2 F}{d(\Delta x)^2} = \sum_a \sum_{a'} \frac{\partial^2 F}{\partial x_a \partial x_{a'}} \frac{dx_a}{d\Delta x} \frac{dx_{a'}}{d\Delta x}$$

$$= \sum_{a \in A_L} \sum_{a' \in A_L} \frac{\partial t_a}{\partial x_{a'}} + \sum_{a \in A_U} \sum_{a' \in A_U} \frac{\partial t_a}{\partial x_{a'}} - 2 \sum_{a \in A_L} \sum_{a' \in A_U} \frac{\partial t_a}{\partial x_{a'}}. \qquad (3.5)$$

If there are no interactions at all, this formula reduces to $\sum_{a \in A_L \cup A_U} t'_a$, as it must.

So, the Newton estimate for the flow shift equalizing path costs is the quotient of equations (3.3) and (3.5). The denominator is strictly positive, because HF is positive definite by monotonicity, and this ratio is well-defined. To preserve feasibility, the flow shift is truncated if any flow would become negative. This corresponds to the case when the longer path becomes unused after the shift.

This operation can be substituted for the flow shift step in gradient projection, Algorithm B, or TAPAS.

3.4 Numerical results

This section describes tests of the algorithms described in the previous section. The key questions concern the computation time needed when link interactions are considered, compared to the separable case. We focus on this issue since computational efficiency is one of the advantages static models have over dynamic ones, and if S-TAP is to be useful in

practice it should maintain this advantage. This section also considers asymmetric instances as well, using these algorithms as a heuristic.

We measure convergence using the relative gap, defined as the ratio between the total system travel time, and the total travel time of an all-or-nothing assignment to shortest paths (keeping the current link costs). Using κ_{rs} to reflect the shortest path cost between nodes r and s, the relative gap is calculated as

$$RG = \frac{\sum_a t_a x_a - \sum_{(r,s) \in Z^2} d_{rs} \kappa_{rs}}{\sum_{(r,s) \in Z^2} d_{rs} \kappa_{rs}} .$$

(3.6)

Additional details on the relative gap, its relationship with other gap measures, and with measures of effectiveness such as link flows and aggregate travel times, are discussed in Patil et al. (2021).

3.4.1 Motivation - Toy example

Consider the example network shown in Figure 3.1. We will use this network with different cost functions to show the effects of symmetric vs. asymmetric interactions, and the impact of how many other links affect the travel time of a given link. Consider the five cases (and the corresponding link costs) shown in Table 3.1.
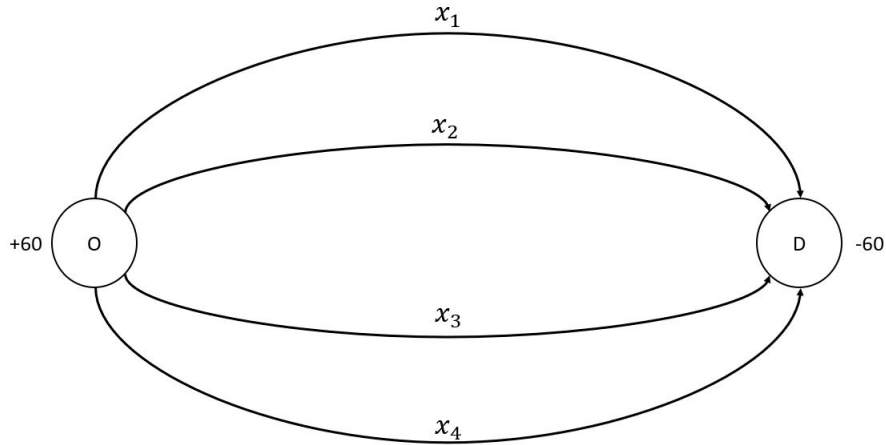
Figure 3.1: Toy network

Table 3.1: Toy network costs for various scenarios

	Separable	Symmetric-full	Asymmetric-full	Symmetric-partial	Asymmetric-partial
Link 1	$15 + x_1$	$15 + 0.5x_1 + 0.167(x_2 + x_3 + x_4)$	$15 + 0.5x_1 + 0.15x_2 + 0.167x_3 + 0.183x_4$	$15 + 0.75x_1 + 0.25x_2$	$15 + 0.75x_1 + 0.25x_2$
Link 2	$10 + x_2$	$10 + 0.5x_2 + 0.167(x_1 + x_3 + x_4)$	$10 + 0.5x_2 + 0.167x_1 + 0.183x_3 + 0.15x_4$	$10 + 0.75x_2 + 0.25x_1$	$10 + 0.75x_2 + 0.3x_1$
Link 3	$10 + x_3$	$10 + 0.5x_3 + 0.167(x_1 + x_2 + x_4)$	$10 + 0.5x_3 + 0.183x_1 + 0.167x_2 + 0.15x_4$	$10 + 0.75x_3 + 0.25x_4$	$10 + 0.75x_3 + 0.3x_4$
Link 4	$15 + x_4$	$15 + 0.5x_4 + 0.167(x_1 + x_2 + x_3)$	$15 + 0.5x_4 + 0.15x_1 + 0.183x_2 + 0.167x_3$	$15 + 0.75x_4 + 0.25x_3$	$15 + 0.75x_4 + 0.25x_3$

Table 3.2: Toy network convergence behavior

Relative gap					
Iteration	**Separable**	**Symmetric-full**	**Symmetric-partial**	**Asymmetric-full**	**Asymmetric-partial**
1	6.0000	1.0000	4.5000	0.9048	4.5000
2	1.6670	0.2444	0.3542	0.3150	0.7500
3	1.1806	0.1531	0.3399	0.1537	0.3632
4	0.4289	0.0733	0.1482	0.0828	0.2465
5	0.6968	0.0384	0.0828	0.0396	0.1591
Flows					
Iteration	**Separable**	**Symmetric-full**	**Symmetric-partial**	**Asymmetric-full**	**Asymmetric-partial**
1	[0.00,60.00,0.00,0.00]	[0.00,60.00,0.00,0.00]	[0.00,60.00,0.00,0.00]	[0.00,60.00,0.00,0.00]	[0.00,60.00,0.00,0.00]
2	[0.00,30.00,30.00,0.00]	[0.00,40.00,20.00,0.00]	[0.00,30.00,30.00,0.00]	[0.00,41.00,19.00,0.00]	[0.00,40.00,20.00,0.00]
3	[25.00,17.50,17.50,0.00]	[10.00,31.67,18.33,0.00]	[11.33,24.33,24.33,0.00]	[0.00,31.34,15.98,12.68]	[0.00,26.67,16.67,16.67]
4	[12.50,11.25,11.25,25.00]	[6.67,26.11,17.22,10.00]	[5.67,18.48,20.37,15.48]	[9.78,25.73,15.19,9.30]	[15.55,21.11,13.33,10.00]
5	[9.38,26.88,8.12,15.62]	[5.18,24.63,22.78,7.41]	[14.18,18.16,17.46,10.20]	[6.91,22.65,23.22,7.21]	[9.26,16.30,26.67,7.77]

The convergence results for the first few iterations of gradient projection are given in Table 3.2. More iterations for this toy example are omitted for brevity. The symmetric-full interaction scenario achieves a lower relative gap for the same number of iterations, followed by asymmetric-full interaction, symmetric-partial interaction, asymmetric-partial interaction, and lastly, separable scenario.

This is due to two effects. First, a higher level of interaction leads to faster convergence. We speculate that this occurs because each path flow equilibration affects more links in the network than just those in A_L and A_U, and therefore moves more of the network towards equilibrium at each step. Compare the symmetric-full and asymmetric-full scenarios to the symmetric-partial and asymmetric-partial scenarios in Table 3.2. The full-interaction cases perform better than partial interaction scenarios, which still perform better than the separable TAP scenario.

Second, symmetric scenarios tend to achieve (somewhat) faster convergence compared to asymmetric scenarios. This is attributed to the path equilibration step having accurate information about the rest of the network (by symmetric Jacobian effects) as opposed to approximate information in the asymmetric case. Consider the symmetric-partial and asymmetric-partial cases, with very minor link cost differences on links 2 and 3. The symmetric-partial relative gap is almost half of the asymmetric-partial gap by iteration 5, a trend that continues in additional iterations not shown in the detail.

These two effects can be quantified using the eigenvectors of the Hessian of the cost matrix. For our problem, this is represented by the weight matrix for linear cost functions, and approximated by it otherwise. For steepest descent methods, the condition number (ratio of largest eigenvector to smallest eigenvector) correlates to the rate of convergence (Bertsekas, 2016). A large condition number means the problem is ill-conditioned, i.e., the optimization variables are not relatively scaled well, and convergence will be slow. A small condition number (closer to 1) will have faster convergence behavior. For example, the condition numbers for the symmetric and asymmetric toy examples with full interaction in Table 3.1 are 3 and 3.154, respectively.

In our investigations, we observe that problem instances with differing condition numbers follow this behavior, and the cost matrix can be useful for predicting convergence behavior w.r.t. similar problem instances. For instance, see Figure 3.2. We generated problem instances with slightly different weight matrices (and therefore, condition numbers), and then allowed GP to solve the instances for 20 iterations. We can see that the instances with lower condition numbers generally show better convergence than instances with higher
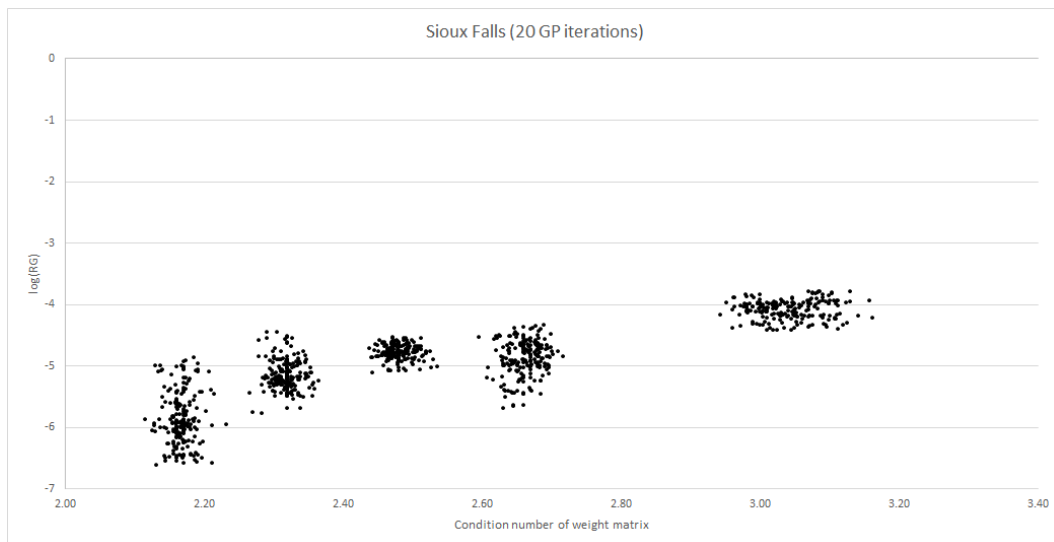
Figure 3.2: Condition number behavior of Sioux Falls problem instances

condition numbers.

The next subsection details our experiment design on significantly larger real-world networks to corroborate these observations, and our intuition about convergence for the symmetric/asymmetric and no interaction/full interaction cases.

3.4.2   Data and Experiment design

We test the method of successive averages (MSA), Frank-Wolfe (FW), and gradient projection (GP) on five standard networks, chosen for their varying size and congestion levels. The networks are obtained from the transportation problems test repository (Stabler, 2019). Our implementations of MSA, FW, and GP for S-TAP and A-TAP, as well as the testing framework, can be found on the first author's Github repository (Patil, 2021). Table 3.3 contains the network size details and average link volume over capacity as a proxy for congestion. These experiments used a relative gap of 10^{-6} as a convergence criterion.

We used the following procedure to generate cost functions with interactions that attempt to preserve the level of congestion in the original networks. The link performance functions in the original network are separable. In our experiments, we replace each x_a with a linear combination of the form $\sum_{a' \in A} w_{aa'} x_{a'}$, with $\sum_{a'} w_{aa'} = 1$ and each $w_{aa'} \geq 0$. If the weight matrix $\mathbf{W}$ is symmetric, then the interactions are approximately symmetric (but

25

Table 3.3: Description of networks used

Network name	Zones	Links	Nodes	Trips
SiouxFalls	24	76	24	360,600
Eastern-Massachusetts	74	258	74	65,576
Chicago-sketch	387	2950	933	1,260,907
Barcelona	110	2522	1020	184,679
Chicago-Regional	1790	39018	12982	1,360,427

not entirely so, since the link performance functions are nonlinear). The separable case is represented with $\mathbf{W} = \mathbf{I}$. We generate the weight matrices so that each link depends on a given number of other links (the number of "degrees of dependency," denoted N), and with $\mathbf{W}$ diagonally dominant to avoid cases with multiple equilibria.

The first set of experiments tested the convergence behavior of the three algorithms for TAP, S-TAP, and A-TAP on all networks. The best performing algorithm was then chosen for further convergence testing of S-TAP and A-TAP. The second set of experiments studied the effect of topographical link interactions. These experiments aim to understand impact the degree of dependency has on the convergence rate. The third set of experiments analyzed the effects of symmetry. We consider a smooth transition from asymmetry to symmetry to starting with an asymmetric matrix $\mathbf{W}$, and taking weighted averages with the associated symmetric matrix $\frac{1}{2}(\mathbf{W} + \mathbf{W}^T)$. Specifically, these experiments consider the Jacobian matrices $\lambda \mathbf{W} + (1 - \lambda)\frac{1}{2}(\mathbf{W} + \mathbf{W}^T)$ for $\lambda \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$.

3.4.3 Results

Figures 3.3 and 3.4 present the compiled results for Sioux Falls. The first observation from Figure 3.3(a) is the behavior of MSA and FW algorithms is extremely similar for TAP, S-TAP, and A-TAP, while GP outperforms them. The relative gap and number of iterations are linearly related on a logarithmic axis, consistent with prior literature (Xie et al., 2018). Based on these initial results, we use GP as the testing algorithm for further experiments.

An important observation here is the independence from implementation details and absolute computation time. Our experiments have been conducted on a basic GP implementation. Therefore, any performance gains achieved by parallelization or other implementation techniques are applicable to these results, helping speed up convergence. For instance, Chen et al. (2020) implement a parallel block coordinate descent algorithm

based on GP algorithm, and the absolute computation time gains would be applicable to our observations for S-TAP and A-TAP. Therefore, we do not emphasize absolute computation times, but focus on comparative behavior.

Figures 3.3(b) and 3.3(c) show the behavior of link cost dependency on link flows within N degrees of any given link for S-TAP and A-TAP, respectively. A higher degree of dependence leads to earlier convergence, as was observed in the motivating toy network. This is attributed to each flow equilibration step having more implicit information about the network state. Also, S-TAP is seen to converge marginally faster than A-TAP, as expected. The only exception is the case $N = 2$ case, though it still does not outperform $N = 5$ or $N = 6$ link interaction instances. Along similar lines, Figure 3.4 shows convergence behavior when the weight matrix proceeds from an asymmetric instance ($\lambda = 1$) to the corresponding symmetric instance ($\lambda = 0$). The results indicate as the matrix tends toward the symmetric version, more accurate information (about the remaining network) is available during each flow equilibration, leading to marginally faster convergence.

Figures 3.5–3.8 show the results for Eastern Massachusetts, Chicago Sketch, Barcelona, and Chicago Regional networks, respectively. The main observations from the Sioux Falls networks are seen to hold for these larger networks. The maximum N for these four networks is 8, 31, 30, and 112, respectively. These networks show no exception to these trends, unlike the Sioux Falls network. Also, the test instances with lower N values achieved better relative gap levels for larger networks; the highest N instances all reached a gap below 10^{-12}.
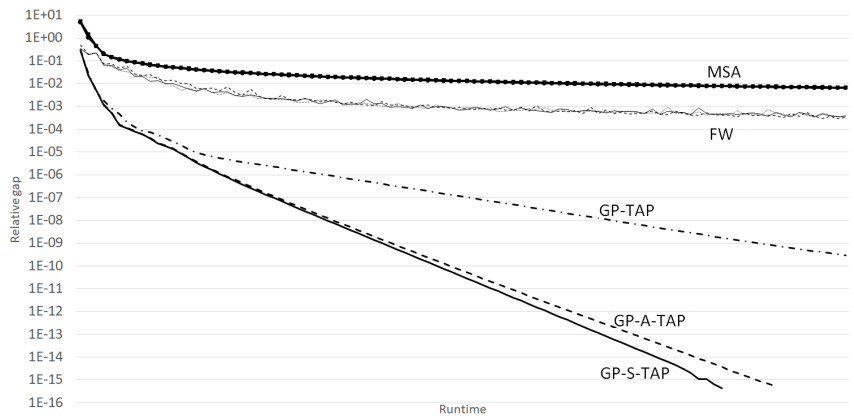
(a) Asymmetric to symmetric weight matrix convergence
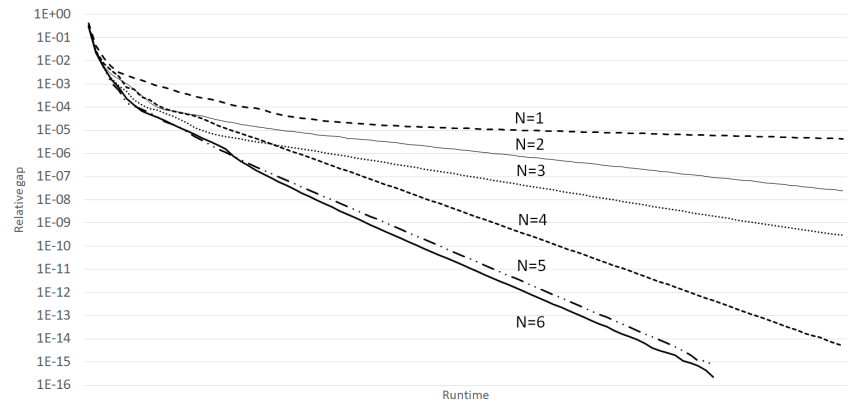
(b) Convergence for S-TAP GP with *N*-link cost dependency
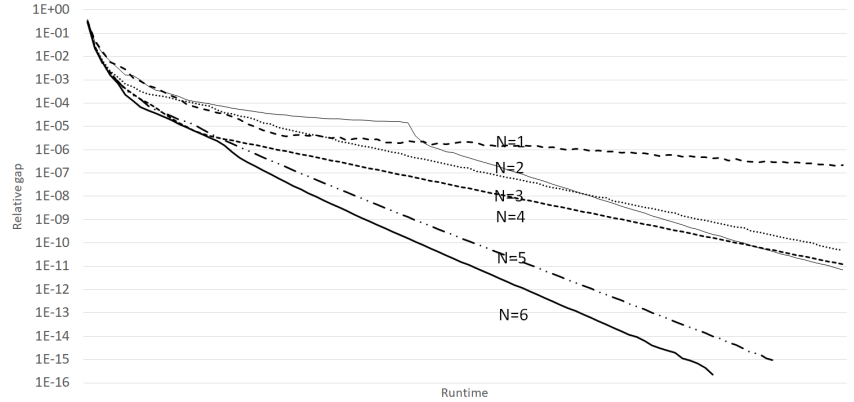
(c) Convergence for A-TAP GP with *N*-link cost dependency

Figure 3.3: Experimental results for Sioux Falls network
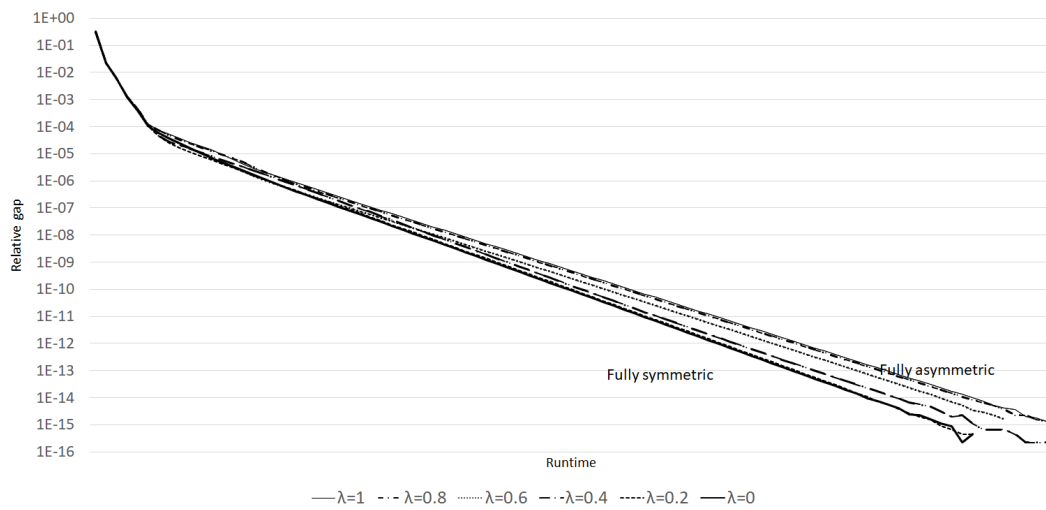
Figure 3.4: Sioux Falls asymmetric to symmetric weight matrix convergence
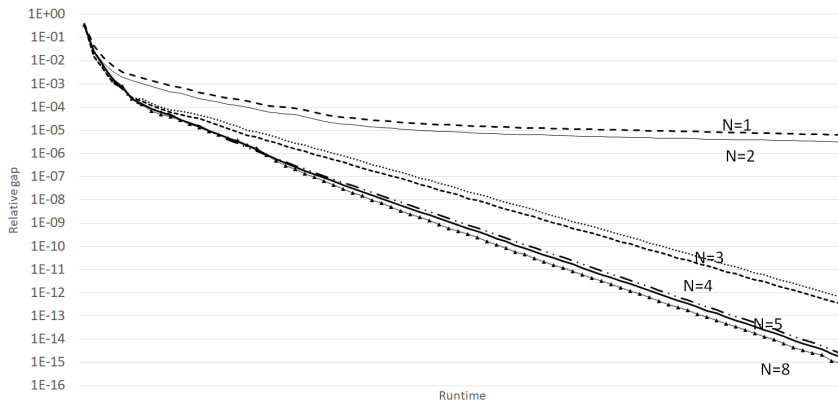
(a) Convergence for S-TAP GP with N-link cost dependency
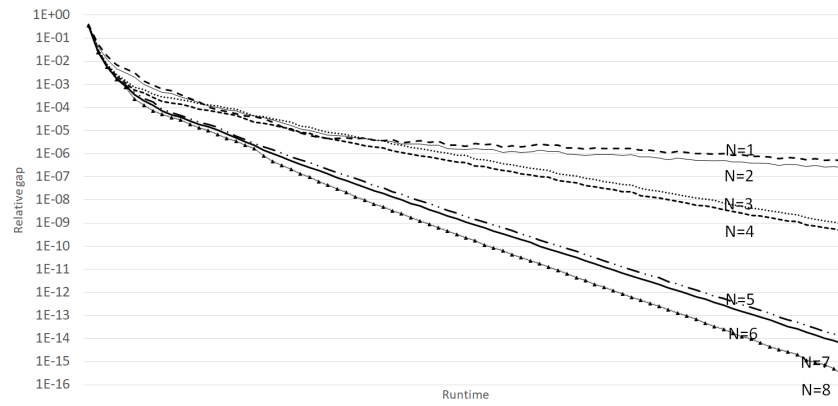
(b) Convergence for A-TAP GP with N-link cost dependency
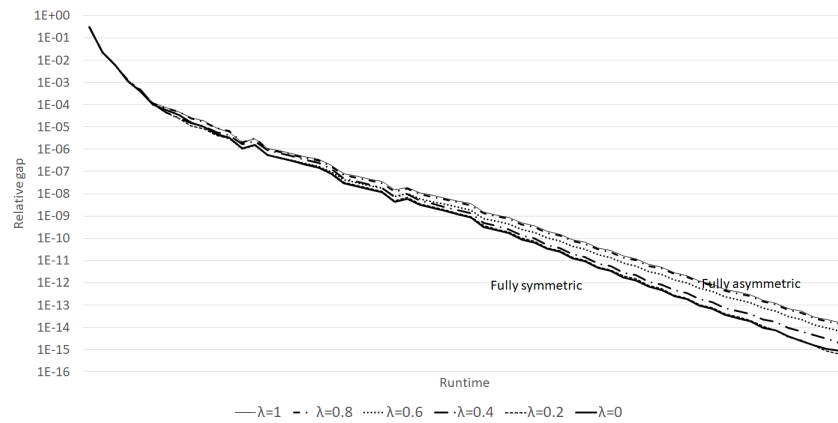
(c) Asymmetric to symmetric weight matrix convergence

Figure 3.5: Experimental results for Eastern Massachusetts network
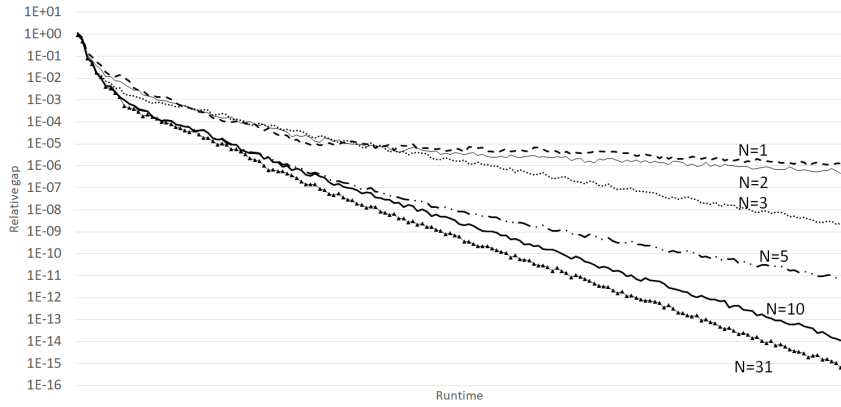
(a) Convergence for S-TAP GP with *N*-link cost dependency
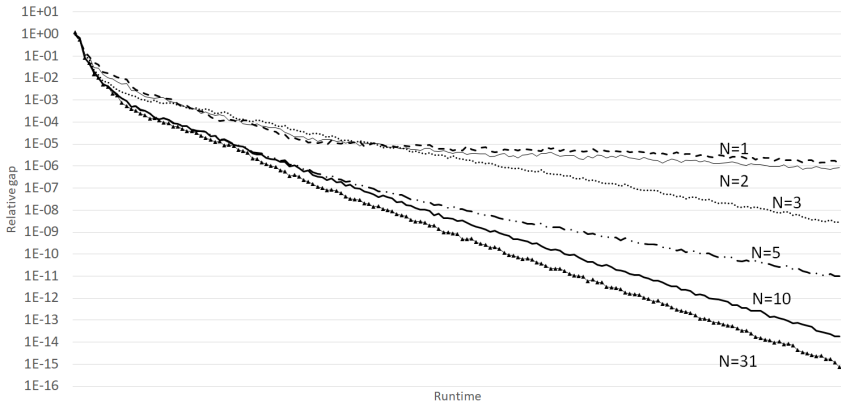
(b) Convergence for A-TAP GP with *N*-link cost dependency
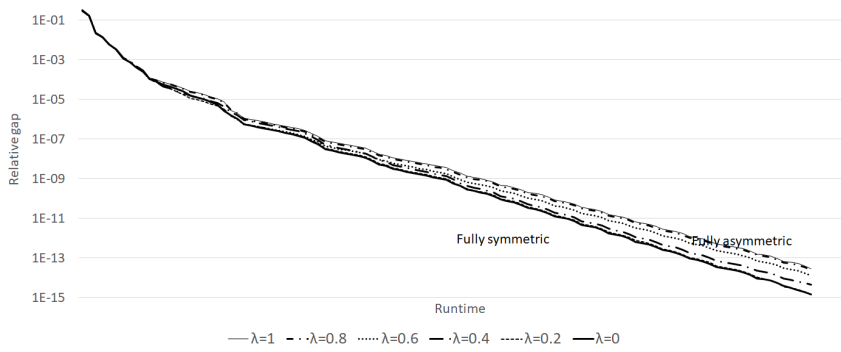
(c) Asymmetric to symmetric weight matrix convergence

Figure 3.6: Experimental results for Chicago-sketch network
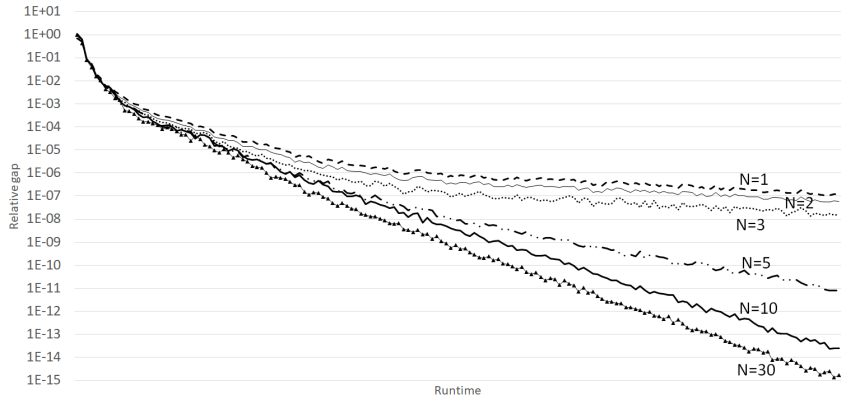
(a) Convergence for S-TAP GP with *N*-link cost dependency
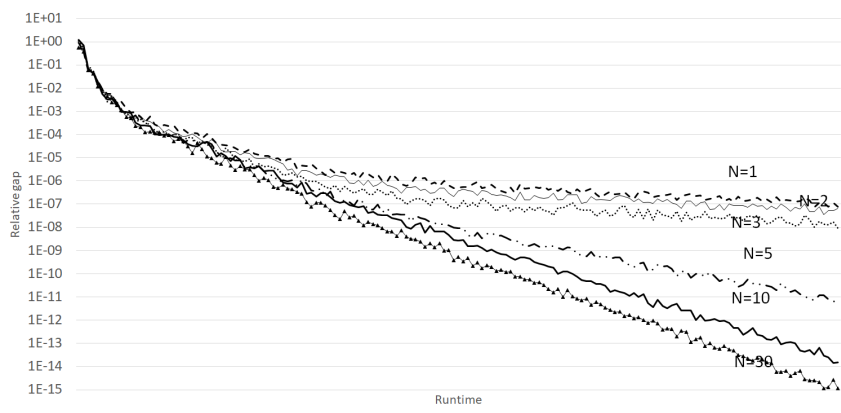
(b) Convergence for A-TAP GP with *N*-link cost dependency
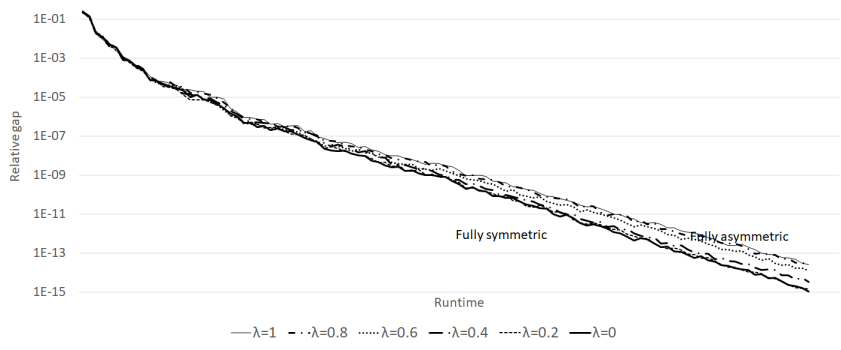
(c) Asymmetric to symmetric weight matrix convergence

Figure 3.7: Experimental results for Barcelona network
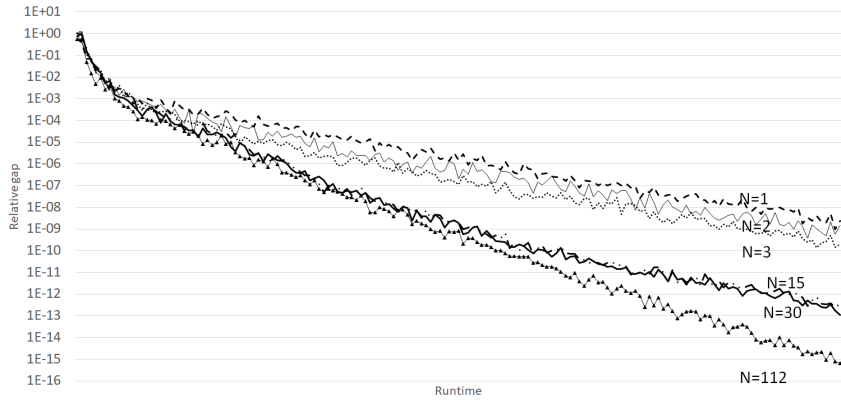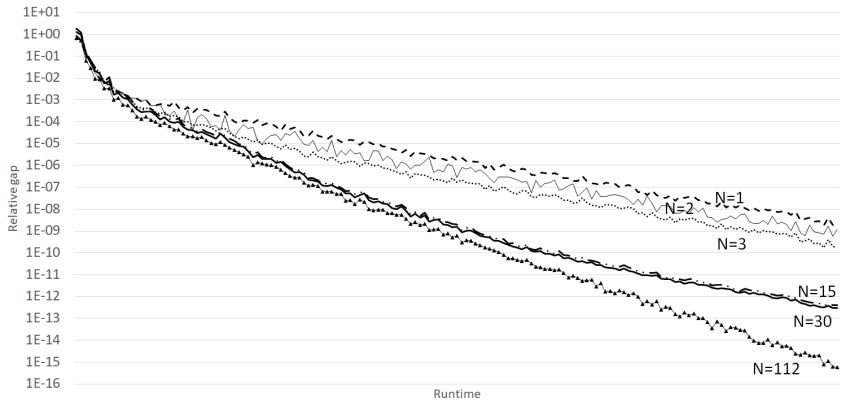
32

(a) Convergence for S-TAP GP with N-link cost dependency

(b) Convergence for A-TAP GP with N-link cost dependency
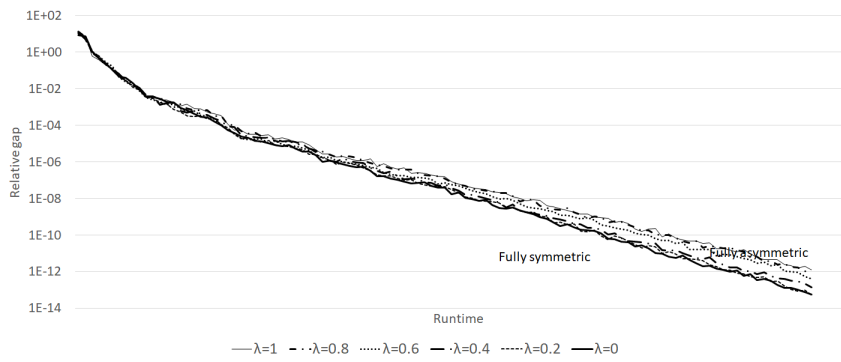
(c) Asymmetric to symmetric weight matrix convergence

Figure 3.8: Experimental results for Chicago-regional network

Table 3.4: Description of networks used

Network name	Zones	Links	Nodes	Trips	Average flow-to-capacity ratio
SiouxFalls	24	76	24	360,600	1.612
Eastern-Massachusetts	74	258	74	65,576	0.163
Anaheim	38	914	416	104,694	0.297
Chicago-sketch	387	2950	933	1,260,907	0.257
Berlin-Prenzlauerberg-Center	98	2184	975	23,648	0.121
Barcelona	110	2522	1020	184,679	1.137
Winnipeg	147	2836	1052	64,784	2.028
Terrassa	55	3264	1609	25,225,700	5.964
Austin	7388	18961	7388	739,351	0.875
Berlin-Center	865	28376	12981	168,222	0.092
Chicago-Regional	1790	39018	12982	1,360,427	0.522
Philadelphia	1525	40003	13389	18,503,872	0.949

3.5 Convergence experiments

This subsection focuses on the convergence behavior for S-TAP. As stated above, we primarily consider GP with two-link interactions. This limits our experiments to the most common use case, i.e., two-way roads. The networks studied in this chapter are shown in Table 3.4, all obtained from the Transportation Networks for Research repository (Stabler, 2019). For ease of reference, we categorize the networks roughly by size: Sioux Falls through Anaheim are designated as small, Chicago Sketch through Terrassa are designated as medium, and the remaining networks are designated as large. The last column in this table shows the average equilibrium flow-to-capacity ratios, excluding centroid connectors. We consider networks with ratios of less than 0.5 to be uncongested, with ratios between 0.5 and 1.0 to be semi-congested, and networks with ratios greater than 1.0 to be congested. The Terrassa network is a clear outlier in this regard, assigning over 25 million trips in a region whose current population is around 200,000, resulting in a flow-to-capacity ratio of almost 6. While such a demand level may not be realistic, we nevertheless include this network as a "stress test" to see whether consistent trends can be seen even in extremely congested networks. Lastly, we also conduct these experiments using Algorithm B (AlgB), a bush-based algorithm which is often faster in practice.

Given a feasible solution $(\mathbf{x}, \mathbf{h})$ to TAP, we select three metrics for analysis. The *total system*

travel time (TSTT) expresses the sum of each vehicle's travel time in the network:

$$TSTT(\mathbf{x}) = \sum_{(i,j) \in A} t_{ij} x_{ij} \tag{3.7}$$

Vehicle-miles traveled (VMT) expresses the total distance traveled by vehicles in the network:

$$VMT(\mathbf{x}) = \sum_{(i,j) \in A} l_{ij} x_{ij} \tag{3.8}$$

To measure convergence of these metrics, we calculate the relative difference between their values at the current solution $\mathbf{x}$ and the equilibrium solution $\mathbf{x}^*$:

$$\Delta TSTT(\mathbf{x}) = \frac{TSTT(\mathbf{x}) - TSTT(\mathbf{x}^*)}{TSTT(\mathbf{x}^*)} \tag{3.9}$$

$$\Delta VMT(\mathbf{x}) = \frac{VMT(\mathbf{x}) - VMT(\mathbf{x}^*)}{VMT(\mathbf{x}^*)} . \tag{3.10}$$

Both TSTT and VMT are aggregate metrics. To represent convergence of the specific link and path flows themselves, we measure the proportion of links within a given relative threshold ϵ of their equilibrium values. Let $A^*_\epsilon(\mathbf{x})$ denote the set of links with flows within this threshold:

$$A^*_\epsilon(\mathbf{x}) = \left\{ (i, j) \in A : \left| x_{ij} - x^*_{ij} \right| < \epsilon x^*_{ij} \right\} . \tag{3.11}$$

Using these sets, we define the *proportion of unconverged links (PUL)* as

$$PUL(\mathbf{x}, \epsilon) = 1 - \frac{|A^*_\epsilon(\mathbf{x})|}{|A|} , \tag{3.12}$$

These metrics — $\Delta TSTT$, ΔVMT, PUL — are directly related and used in practical applications and planning, and converge to zero at the equilibrium solution. Additional detail about these metrics can be found in Chapter 6. We track these metrics against relative gap (defined in Section 3.4), a convergence metric.

The full set of results can be found in Appendix A. The values of the three main metrics for GP experiments are shown in Figures 3.9–3.11 (presented according to each metric). The thin lines represent the values of each metric in one of the twelve networks tested, and the thick line represents the average value. Both sets of figures use logarithmic
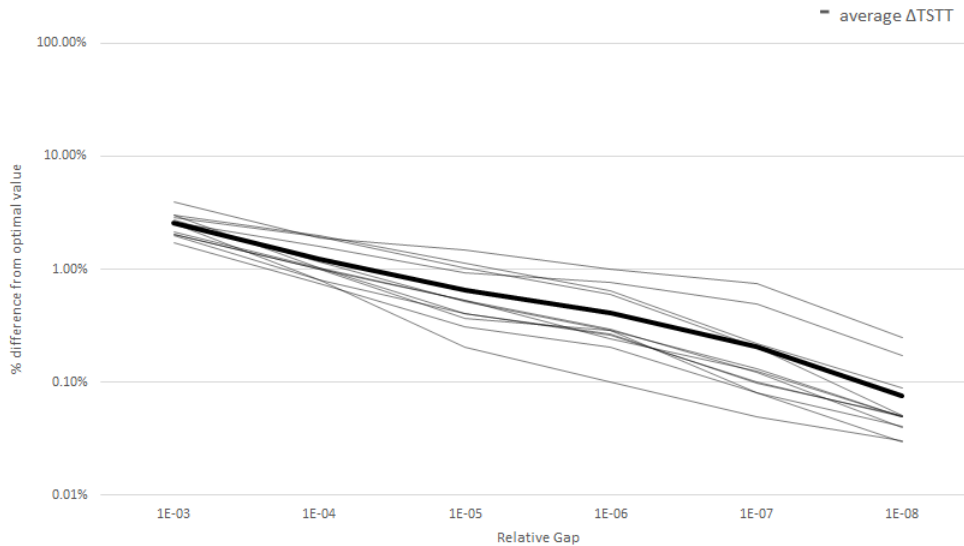
Figure 3.9: ΔTSTT trends for different gap levels

axes for the relative gap. All metrics converged at roughly similar rates, despite significant differences in the size and congestion level of the networks tested.

In all the networks, the aggregate metrics ($TSTT$ and VMT) are already very near stabilization at a relative gap of 10^{-3}. For the small and medium networks, these values are within 1% of the equilibrium values when the relative gap is 10^{-4}, and for the large networks they are within 2%. Both $\Delta TSTT$ and ΔVMT converge at roughly similar rates, but ΔVMT is usually slightly lower at a particular gap level. This behavior is in line with the metric behavior for TAP, as noted in Patil et al. (2021) and section 6.4 of this study.

The proportion of unconverged links was the metric originally studied by Boyce et al. (2004) for the Philadelphia regional network. They found that a gap of 10^{-4} was required to approach convergence for freeway links, defining convergence as a PUL of 1% or less. To achieve this level of convergence for arterial links as well as freeway links, a relative gap of 10^{-5} was needed. Our results show that this latter conclusion generally holds across the other networks tested, and that 99% of link flows are accurate to within 1% of equilibrium values at this gap level.

Next, we compare the GP convergence behavior to AlgB convergence behavior. The full data from these results are shown in Tables A.2 (raw data for Algorithm B) and A.3 (for a side-by-side comparison) in Appendix A. The trends are very similar between the two
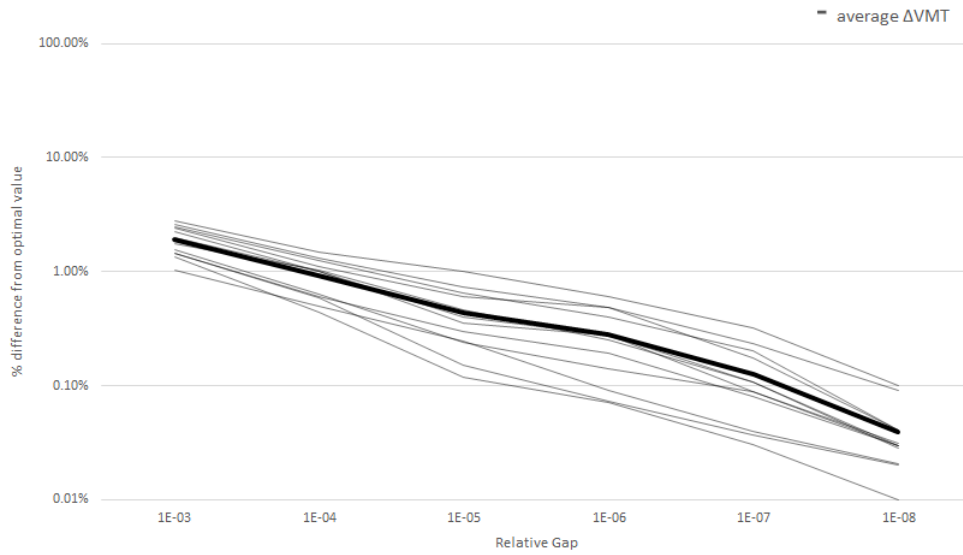
Figure 3.10: ΔVMT trends for different gap levels
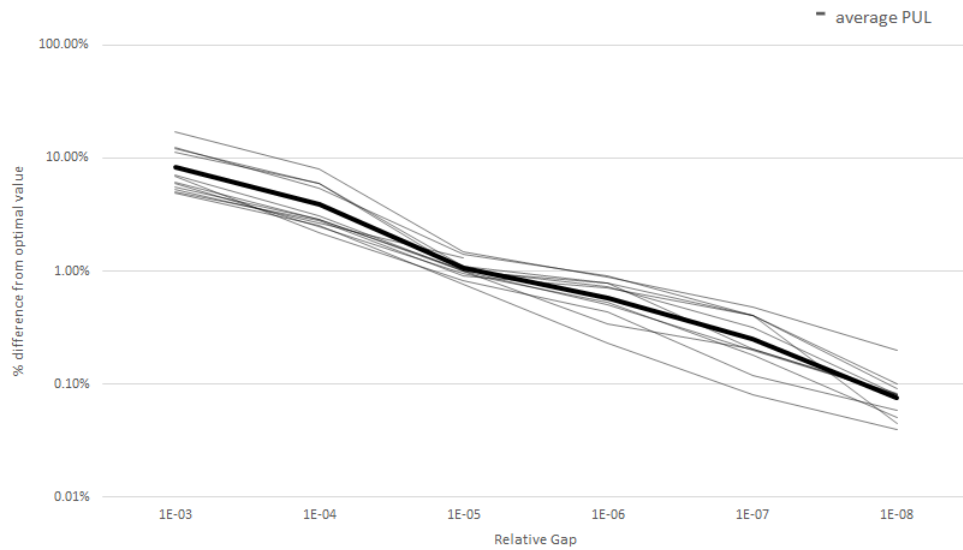
Figure 3.11: PUL trends for different gap levels

algorithms, and the values of each metric are always of the same order of magnitude, and almost always nearly identical numerically. This finding is encouraging, suggesting that the conclusions of the GP experiments are applicable to other algorithms, and that the relative gap is a good universal measure of convergence, regardless of the specific assignment algorithm. Again, these results are in line with prior results from Patil et al. (2021) as well as TAP results in section 6.4.

3.6 Conclusions

This chapter reconsiders the traffic assignment problems with interactions. We showed that merge models from dynamic traffic assignment can be approximated with symmetric, monotone cost functions; examined alternative algorithms, deriving the Newton shift formula with interactions; and considered the practical convergence rate of gradient projection and Algorithm B with such functions. We found that instances with interactions actually converge faster than separable instances of traffic assignment. All of this suggests that problems currently studied with static traffic assignment may benefit from considering interactions — there seems to be little computational difficulty (in fact, convergence was almost always faster), and the valuable properties of equilibrium existence and uniqueness are retained. Of course, there remain other problems where dynamic assignment is preferred.

Further research would be valuable along several lines. Further investigation of appropriate cost functions is needed, to derive them from other node models, and to consider the impacts of these approximations on network-wide flow. Additional research into algorithms for the non-monotone or non-symmetric cases is also needed; we show that gradient projection still functions as an acceptable heuristic, but comparisons with other algorithms from the variational inequality literature are needed.

Chapter 4

Rail network electrification problem

4.1 Introduction

Rail networks play a vital role in local and national economic structures. In many countries such as Japan and Switzerland, they constitute a significant share of passenger transport mode share (MLIT Ministry of Land, Infrastructure, Transport and Tourism, 2018; Statistics Bureau, Ministry of Internal Affairs and Communications, 2018; Federal Statistical Office - Section Mobility, 2019). Rail freight transit also accounts for a large portion of total freight transit, exceeding 50% modal share in large economies like Canada and Russia (Organisation for Economic Co-operation and Development, 2013; Business Standard, 2019; Eurostat, 2020). Therefore, a significant opportunity exists for government policies incentivizing rail network improvements to provide for larger economic, environmental and social returns.

Rail network electrification, and the accompanying transition from diesel-electric to fully electric locomotives, are important steps towards sustainable systems and renewable fuel sources. There are many studies on the impact and cost-benefit analysis of rail electrification (United States Department of Transportation - Federal Railroad Administration, 2015; U.S. Department of Transportation - Federal Railroad Administration, 2019). Advantages of electric locomotion include lower long-term energy and locomotive maintenance costs, lower noise and air pollution levels, faster acceleration, and more flexibility in the primary power source, leading to less volatility from fuel price fluctuations. These benefits must be balanced with significant upfront investment for infrastructure upgrades, higher infrastructure maintenance costs, vulnerability of overhead architecture, higher property tax obligations for private rail companies, and the general uncertainty in the investment (Walthall, 2019).

We consider a rail network design problem (RNDP) where parts of a freight rail network can be electrified, subject primarily to budget constraints. For readers unfamiliar with the bi-level optimization problem known as a network design problem (NDP), the literature review provides background reading. The RNDP we formulate aims to find the optimal subset of rail links for infrastructure upgrade with a given objective. In this chapter, the formulation uses the objective of minimizing private costs, laying a foundation that can be adapted to improve net social benefit and reflect where subsidies ought to be directed. Given that there are many rail operators, and that shippers can choose which operator to use (not necessarily aligning with net social benefit), we model the lower level problem as a user-equilibrium traffic assignment problem (TAP). Genetic algorithms (GAs) solve the higher level problem of finding the optimal links to electrify.

In this setting, rather than apportioning an electrification budget to each rail operator independently, a utilitarian schema allocates the electrification budget for specific link electrification in order to bring about the greatest possible cost reductions across the network. The rail operators and shippers then respond to these changes by altering their scheduling and flow patterns to minimize their individual costs. Given multiple operators, the lower level problem is a setting where flow is directed "selfishly," to minimize shipment costs. For the lower level problem, we assume that these activities lead to an equilibrium, where the costs of shipment flows cannot be lowered unilaterally. With shipment flow expressed in tons, as a continuous quantity over the long term, this problem satisfies the traffic assignment user equilibrium assumptions. This assumption is consistent with prior literature on the topic (Uddin and Huynh, 2015; Wang et al., 2018a). An alternative approach would be to model a smaller number of self-optimizing fleet owners, resulting in a Nash-Cournout equilibrium. However, Van Vuren and Watling (1991) show that the difference in the resulting assignments is less than 5% for aggregate measures in large networks, with this difference decreasing to zero as the number of owners increases (indeed, the user equilibrium is the limiting case of infinitely many owners.) Nash-Cournout equilibria are harder to compute, and require more data to calibrate; given the relatively small observed differences in the aggregate metrics of interest in this study, we opt to study only user equilibria.

4.1.1 Rail electrification and policy planning

The oil crisis of the 1970's spurred research into the economics of rail electrification. In 1977, Schwarm (1977) examined the detailed costs associated with rail electrification

infrastructure. Schwarm found three primary categories: power delivery, public works compatibility, and signaling systems compatibility. The power delivery category itself can be divided into the contact systems and the traction substations, the latter helping to determine each link's electrified capacity. Nearly ten years later, Kneschke (1986) detailed the design requirements of the traction substations, which affect the electrified lines' capacities. In the midst of these studies looking at the costs of electrification, and the trade-offs with regard to capacity, other researchers began looking at electrification's benefits. Whitford (1981) examined the energy savings from electrifying the high-density portions of the rail network, and Ditmeyer et al. (1985) would incorporate ancillary benefits and costs, such as those accruing to maintenance, reliability, and fuel handling. At the time, emissions were not a major concern. Additionally, computational costs and the relative youth of network optimization techniques meant that these analyses primarily used traffic density as a heuristic for link selection.

As computational resources and optimization algorithms progressed, the oil crisis abated and deregulation led to a decrease in rail network mileage. Interest in freight rail electrification only returned relatively recently. In 2012, Cambridge Systematics (2012) conducted an extensive study on electrification for the Southern California Association of Governments. RailTEC (2016) conducted a similar study for the California Air Resource Board in 2016. Both studies were primarily motivated by reducing rail emissions, particularly in dense urban areas. The latter study highlighted the problem of network connectivity: because rail freight routes are particularly long, and the costs of switching from electric to diesel-electric locomotives *en route* are high (in terms of delay and logistical costs of ensuring locomotives are available), confining electrification to relatively small regions is not an efficient way to limit emissions. Outside of regulatory constraints or internalized emissions costs, rail companies would likely send large numbers of diesel-electric trains through electrified portions of the rail network in order to limit overall route costs. We thus face an optimization problem impacted by policy: how should the initial links be selected for electrification, given budget constraints and a desire to reduce emissions from the rail network?

There are very few studies to draw on for link improvement under a budget constraint, such as Mishra et al. (2016), which applied numerical methods towards "[making] optimal investment decisions... in moderate and large transportation networks." Mishra utilized travel costs to road users, primarily consisting of temporal costs. Applying a similar framework to a rail network considers similar costs, although fuel costs for rail constitute a

relatively larger portion, and time costs a relatively smaller one.

4.1.2   Rail electrification of other national networks

While the European rail network is extensively electrified, most of it is used for passenger transport and comparisons to the US network are difficult. This can be attributed to deliberate prioritization of passenger rail efficiency over freight, as evidenced by limits on maximum permissibly train lengths, maximum axle loads, and vertical car height (Walthall, 2019). Similarly, the Japanese rail network is used primarily for passengers, accounting for less than 1% of total freight load.

India and China both offer examples of extensive electrification of freight rail systems relevant to the US network. China has extensively electrified both its passenger and freight networks in the past decades, with over 70% total electrification and about 25% of its network electrified between 1990 and 2007 (Lawrence et al., 2019). Much of this electrification has coincided with major construction to upgrade capacity and signaling infrastructure, especially in mountainous areas.

Similarly, India has extensively electrified its network (about 70%), prioritizing selected main lines and high density routes (Ministry of Railways (India), 2020). India has constructed high overhead contact systems in order to accommodate double-stacked freight. Similar construction would be required in the US, where double-stacked container freight is a large component of American railroads' revenues.

4.1.3   Contributions and overview

NDPs are widely studied in road networks. Rail NDPs vary in two significant ways. First, most of the freight rail network is privately owned by the user or contracted out for usage, which leads to non-socially-optimal usage restrictions. Rail electrification has highly uncertain, and possibly negative, rates of return when external benefits are not accounted for. This chapter provides a framework for future analysis of policy interventions to internalize the benefits to the private companies that own the rails and would be responsible for implementing electrification. Second, the characteristics of individual network links (track curvature and gradient) affect maintenance and operating costs more so than in road networks.

With these distinctions in mind, the main contributions of this chapter are as follows:

- We formulate the rail electrification NDP as a bi-level optimization problem incorporating electrification costs; fuel, locomotive, and operational costs; and train resistance (bearing, flange, air, grade, curve, braking, and inertia) costs.

- We present a heuristic for solving our formulation, using a general-purpose meta-heuristic (a genetic algorithm) based on problem-specific insights.

- We solve this problem on a large-scale network representing the North American rail system, and analyze the resulting solutions to draw insights and policy conclusions.

The rest of this chapter is organized as follows. We first describe the formulation for the rail electrification NDP and associated model components. We then provide background information and reading on NDPs and solution methods. We then describe the North American rail network dataset and demand data we use, and outline our experiment design. We follow this with a summary of the results from our experiments and draw practical insights. We conclude by summarizing our findings and suggest avenues for future work.

4.2 Model Formulation and literature review

Let $\mathbf{A}$ denote the set of rail links traversable by diesel-electric trains. Each link $i \in \mathbf{A}$ currently has a flow-dependent usage cost $c_i(x_i)$ per unit flow. The link can be electrified for a cost of c_i^e, changing the usage cost function to $c_i'(x_i)$ per unit flow. The usage costs for diesel and electrified links differ due to technological differences and fuel costs, and are lower for electrified links. The set of candidate links eligible for electrification is denoted by $\mathbf{A^E}$, a subset of $\mathbf{A}$. The flow on link i is given by x_i tons per day. The capacity of the link is denoted by u_i. Let Π denote the set of paths in the network, and h_π the flow on a particular path $\pi \in \Pi$.

The set of nodes (denoting stations, yards, interchanges, etc.) is denoted by $\mathbf{N}$. The demand between each origin r and destination s node is given by d_{rs}. The demand information between all origin-destination (OD) pairs is stored in the OD matrix $\mathbf{D}$. The total budget for upgrades is B. The decision variable y_i for $i \in \mathbf{A}$ equals 1 if link i is chosen for upgrades (electrification) and 0 otherwise. The cost for switching the mode of operations from diesel to electric (and vice versa) at yard i is w_i (and infinitely high at other nodes without switching facilities).

The practical interpretation of link electrification involves additional infrastructure allowing electric trains to use a link with different fuel costs while maintaining same total link

capacity. Therefore, a network transformation is needed to capture the separate diesel and electric flows of goods with interacting congestion costs and separate track resistance costs. Figure 4.1 shows this network transformation for each node in the network, incorporating switching costs w_i. The link cost (specified in the Cost Formulae sub-section) has two separable components, a congestion cost and track resistance. The link costs c_i are broken down into congestion costs c_i' dependent on the total diesel plus electric flow, and track resistance costs dependent only on mode-specific flow.
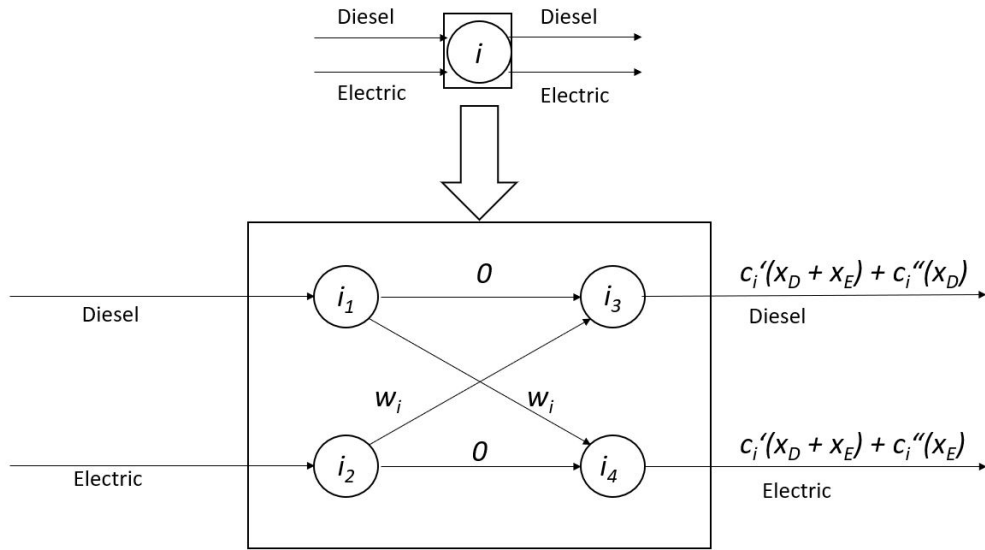
Figure 4.1: Network transformation

The RNDP is then formulated as follows:

$$\min \quad \sum_{i \in \mathbf{A}} c_i(x_i) x_i (1 - y_i) + c_i'(x_i) x_i y_i \tag{4.1}$$

subject to:

$$\sum_{i \in \mathbf{A}} c_i^e y_i \leq B \tag{4.2}$$

$$y_i \in \{0, 1\} \qquad \forall i \in \mathbf{A} \tag{4.3}$$

where y_i are found from the upper level problem and x_i values are found by solving the

44

following lower level assignment problem, in which M is a sufficiently large constant:

$$\min_{\mathbf{x},\mathbf{h}} \quad \oint_0^{\mathbf{x}} \mathbf{c}(\mathbf{s}) \cdot d\mathbf{s} \tag{4.4}$$

subject to:

$$x_i = \sum_{\pi \in \Pi : (r,s) \in \pi} h_\pi \qquad \forall (r,s) \in \mathbf{A} \tag{4.5}$$

$$x_i \leq M y_i \qquad \forall i \in \mathbf{A^E} \tag{4.6}$$

$$\sum_{\pi \in \Pi^{rs}} h_\pi = d_{rs} \qquad \forall (r,s) \in \mathbf{N}^2 \tag{4.7}$$

$$h_\pi \geq 0 \qquad \forall \pi \in \Pi \tag{4.8}$$

Equation 4.1 minimizes the total usage cost over electrified and non-electrified links. Equation 4.2 enforces the electrification budget constraint. Equation 4.3 indicates that y_i is a binary indicator variable. Equation 4.4 is the modified Beckmann function for symmetric link interactions, described more in the next sub-section. Equation 4.5 states that the relation between link and path flows, and Equation 4.6 ensures that the electrified links with flow are the ones actually chosen for construction. Equation 4.7 ensures that the demand is met across all used paths for each OD pair, and Equation 4.8 states flow non-negativity.

This lower level assignment problem is a generalized version of the static traffic assignment problem (TAP) formulated as a convex program by Beckmann et al. (1956). The rail traffic assignment can be expressed as static symmetric TAP if the flows refer to the tonnage of goods (Alliance Transportation Group, 2013; Uddin and Huynh, 2015; Wang et al., 2018a). We therefore assume the units of flow to be tons, allowing us to use existing solution methods to solve the lower level problem. With this formulation, this assignment problem follows Wardrop's first principle (Wardrop and Whitehead, 1952): "All used paths between each origin and destination must have equal and minimal cost."

The solution to static symmetric TAP is unique if the link performance functions are strictly monotone and the cost Jacobian is positive semidefinite. These conditions are satisfied by the cost functions we describe in the Cost Formulae section. This study uses an implementation of Algorithm B (Dial, 2006) to solve this assignment problem. The validity of the algorithm and flow shift formula for symmetric TAP is proved in the next section. The source code is available at the SPARTA lab github repository (Boyles, 2019) under the *pd-word* branch.

This formulation is presented with fixed demand. If elastic demand functions are known, the fixed-demand assumption can be relaxed by applying the Gartner network transformation (Gartner, 1980; Boyles et al., 2020). This transformation allows for the elastic-demand equilibrium problem to be solved as a fixed-demand problem, introducing new artificial links with specially-designed cost functions based on the demand functions. Since the elastic demand case can be handled by the same solution methods as the fixed demand case, for simplicity of exposition we do not discuss it further in our presentation of the formulation and solution methods.

It is also straightforward to calculate costs for multiple years, accounting for forecasted future changes by solving a lower-level traffic assignment problem for each future year and computing a total discounted cost. This approach would significantly increase computational requirements, however, and we do not pursue it in the experiments we report.

4.2.1 Network design problem background

Network design problems (like many other bi-level problems) are intractable to solve exactly. Such methods (such as branch-and-bound or Benders decomposition) have been proposed, as in Leblanc (1975), Chen and Alfa (1991), Drezner and Wesolowsky (1997), and Long et al. (2010). However, the largest network tested in any of these studies has 40 nodes and 99 links.

Heuristic methods are standard for this class of problems. Genetic algorithms, simulated annealing, and tabu search are examples of such methods that have been applied to traffic NDPs. For details on such methods, see the review papers by Farahani et al. (2013) and Iliopoulou et al. (2019). They study the urban transportation NDPs and provide an overview of the types of problem variations as well as solution methods and some applications. In our experiments, we use a genetic algorithm for a solution heuristic.

Alternatively, the discrete NDP can be transformed into a single level problem. Gao et al. (2005) transformed the problem into a nonlinear single level program utilizing support functions, which was solved by existing techniques. A third approach involves formulating the problem with equilibrium constraints, then using branch and bound or reduced gradient based methods (Lo and Szeto, 2004, 2009; Szeto and Lo, 2006, 2008; Szeto et al., 2010). Other methods, such as Lagrangian relaxation and column generation, have also been used as exact solution methods for small NDP instances (Meng et al., 2001; Borndörfer et al., 2008). As this approach is not viable for large instances, we focus on the non-exact methods,

which trade off guaranteed optimality for tractability.

In our experiments, we use a genetic algorithm as a solution heuristic. Katoch et al. (2021) provide a recent review of GAs with applications in transportation network design problems (road, transit, and multimodal network design), facility layout, inventory and scheduling, and other domains. Key advantages of genetic algorithms for our application are easy parallelization, only requiring objective function values (and not gradients or Hessians), and the ability to quickly obtain quality solutions in practice. These advantages help us exploit the sub-problem independence for parallelization. Additionally, this avoids first- and second-order information calculation for the line integral objective function. Other heuristics exploiting these properties can work well for this problem structure and can be used instead of GAs.

4.3 Cost Formulae

This section provides the cost calculation formulae for the link electrification costs, track resistance, and generalized link costs. Owing to a significant number of constants and rates within these formulae, they are omitted here for brevity, but can be found on the author's github repository (Patil, 2020). In order to update cost values for inflation, this study uses industry appropriate producer price indices from the U.S. Bureau of Labor Statistics (2020).

There are two main costs to be considered, roughly corresponding to the capital expenses and operational expenses. Electrification costs (or capital expenses) refer to all the costs associated with converting a link from standard operations to electric operations that we considered. The generalized link costs (operational expenses) refer to the costs of traversing a link, and incorporate both time and fuel (whether diesel or electricity) costs based on the link's speed and resistance, calculated on a per-ton basis.

4.3.1 Link Electrification Cost

The link electrification cost does not depend on the link's traffic/tonnage (costs associated with re-routing during construction are not considered), but rather the link's location and existing characteristics. This cost can be divided into five components:

- Overhead contact system (OCS) — c_{OCS}

- Electrical substations — c_{sub}

- Transmission lines — c_{trn}

- Public works — c_{pub}

- Signalling — c_{signal}

The first four cost components are positively correlated with the roughness of the link's terrain (e.g. the electrification infrastructure is more expensive to construct in mountainous segments than in plains segments). For each of those four categories, the cost calculations utilize two parameters: a low value for ideal conditions, and a high value for the worst possible condition. In order to create an upgrade cost estimate for all links in the network, the difficulty of the terrain is scaled linearly based on the ratio, α, between each link's actual length and its straight-line length. α has a minimum value of one, which represents a straight track. The cost of signaling depends on the complexity of replacing the link's existing signal systems, and is independent of terrain.

The cost of upgrading link i to be electrified is then written as:

$$
c_i^e = l_i \left(\left(\frac{\alpha_i - \alpha_{min}}{\alpha_{max} - \alpha_{min}} \right) (c_{OCS,max} + c_{sub,max} + c_{transmission,max} + c_{pub,max}) + \left(1 - \frac{\alpha_i - \alpha_{min}}{\alpha_{max} - \alpha_{min}} \right) (c_{OCS,min} + c_{sub,min} + c_{trn,min} + c_{pub,min}) + c_{signal,i} \right), \quad (4.9)
$$

where l_i is the length of link i, and α_{min} and α_{max} are the smallest and largest terrain difficulty values across the network, respectively. The parameters are derived from Schwarm (1977), Kneschke (1986), and Gattuso and Restuccia (2014).

4.3.2 Rail Link Delay Function

To assign freight flows to the network, a link performance function is required, accounting for the rising cost of using a link as link congestion increases. This subsection details the formulation of the function as well as the process used to obtain the coefficients and constants. The final form of the equation is provided in equation 4.12 at the end of this subsection.

Several studies have applied a rail link performance function similar in form to the Bureau of Public Roads (1964) function used in road assignment (Clarke, 1995). Uddin and

Huynh (2015) proposed a link delay function of the form:

$$t_i = t_i^0 \left(1 + \left(\frac{x_i}{u_i}\right)^{\beta}\right) \tag{4.10}$$

where x_i is the flow on link i, u_i is the link's capacity, and t_i^0 is the free-flow travel time. Uddin suggested a value of 4 for the parameter β. u_i is a function of the link's class, as well as the number of tracks in the link and the frequency of sidings and switches.

The link's total generalized cost is a function of the travel time, crew and cargo costs, as well as the fuel cost. The fuel cost is the product of the link travel time, cost per unit energy for the fuel source used, and power level required for the link divided by the efficiency of the locomotive. For this analysis, an average train unit is the basis of analysis, and the generalized cost is based on the cost for that train unit to traverse a link divided by that train unit's cargo mass. In this way, the network flows can be assigned as tons of cargo. The power level used on each link is calculated using the link's total resistance, as outlined in the next subsection. The generalized cost function can then be written as follows:

$$c_i \ (or \ c_i') = t_i \left(crewRate \ + \ cargoRate \ + \ \frac{P_i}{\eta} fuelCost\right) / (n_C m_{cargo}) \tag{4.11}$$

where $n_C m_{cargo}$ is the mass of the cargo hauled by a train unit used in the analysis. In order to separate the congestion cost and the resistance cost, we make the assumption that the fuel consumption on each link is fixed, meaning the generalized cost function changes as follows:

$$c_i \ (or \ c_i') = \left(t_i \ (crewRate \ + \ cargoRate \ ) + t_0 \left(\frac{P_i}{\eta} fuelCost\right)\right) / (n_C m_{cargo}) \tag{4.12}$$

4.3.3 Track resistance

In this chapter, we assume that trains have n_L locomotives and n_c railcars. Each railcar has a tare weight of m_c and a gross weight of $m_g = m_c + m_{cargo}$. Each locomotive has a weight of m_L. A loaded train then has a mass of $m_T = n_L m_L + n_c m_g$.

Time and energy exertions associated with switching between diesel-electric and electric locomotives are handled separately, and discussed in the switching costs section.

The resistance on the train is separated into bearing resistance, flange resistance, air

resistance, grade resistance, curve resistance, and brake resistance, and inertial resistance. Each of these quantities is discussed and specified below.

Bearing resistance: Assuming a relatively new train (less than fifty years old), the bearing resistance on each railcar will be $(a + (\frac{bN_{ax}}{m_g}))m_g$, where N_{ax} is the number of axles on the railcar. According to AREMA committee (Gillespie and Hayes, 2003), when m_g is measured in tons, $a = 2.9\frac{N}{ton}$ and $b = 97.3N$. Therefore, the total bearing resistance is calculated as $\sum_{k=1}^{n_L+n_C} (am_{g,k} + bN_{ax,k})$.

Flange resistance: The bearing resistance on each component of the train, according to the same CN examination, is $m(Bv)$, where v is the speed of the train relative to the track, and m is the mass of the railcar or locomotive. $B = 0.329 \frac{N \cdot s}{m \cdot ton}$ for locomotives(B_L), and $0.494 \frac{N \cdot s}{m \cdot ton}$ for railcars(B_C). The total flange resistance is $v(B_L n_L + B_C n_C)$. The flange resistance varies based on the track quality, so a factor $k_{f,i}$ can be applied to adjust the flange resistance for link i.

Air resistance: The air resistance (Hay, 1982) on each railcar is proportional to the square of the train's speed relative to the wind. For the purposes of this study, the wind speed is assumed to be zero, so that the train's speed relative to the track is equal to its speed relative to the air. The air resistance on each component of the train becomes Kv^2. When v is measured in $\frac{m}{s}$, K takes on a value of $1.56 \frac{N \cdot s^2}{m^2}$ for conventional equipment, or $2.06 \frac{N \cdot s^2}{m^2}$. The total air resistance is $\sum_{k=1}^{n_L+n_C} K_k v^2$. A factor $k_{a,i}$ can be used to adjust the air resistance based on the average air density for link i.

Grade resistance: The grade resistance is the resistance due to gravity. Unlike the other resistances discussed, grade resistance can be positive (upgrades) or negative (downgrades). The grade resistance for the train is given by $m_T g \sin(\theta)$, where g is the acceleration due to gravity and θ is the angle of incline. Rail inclines are small, allowing the small-angle approximation that $\sin(\theta)$ is approximately equal to the track grade gr.

Curve resistance: Curve resistance arises from the force of the track on the wheels within a curve. According to AREMA (Gillespie and Hayes, 2003), curve resistance is approximately equivalent to a grade of 0.04% per degree of curvature. That heuristic would put the curve resistance at $0.45836 m_T g \arcsin\left(\frac{30.48}{2r}\right)$, where r is the link's radius of curvature in meters (an average radius of curvature is assigned to each link based on the link's α value).

Brake resistance: The brake resistance is the force of the brakes applied to the train. This force is applied to maintain control of the train along down grades, and to stop the train. Trains have three braking systems, the most powerful of which, the air brake, takes large amounts of time to engage or disengage. Sometimes, the air brake is left on along flat terrain or small up grades to prevent delays accruing from recharging the pressure in the train tube. Electric locomotives have regenerative braking systems that allow them to recover some braking energy. The regenerative braking system is more reliable than a diesel-electric locomotive's rheostatic braking system, allowing electric locomotives to rely less so on the air brake.

The cost parameters for fuel efficiency and maintenance are derived from Whitford (1981), Fritz, S.G. (2000), and Nektalova (2008). Because we model trains as flows, it is beyond our scope to determine the actual brake usage on each link. We assume that on level terrain, links with an average positive grade, or links with a negative grade below a certain threshold, the incidental brake usage will be equivalent to the resistance of a 0.1% grade. A threshold is determined using the grade that would cause the train to exceed its desired speed along the link when utilizing the minimum throttle level. Beyond that threshold, the brake force is set to the level that will allow the train to reach its desired speed along the link while utilizing the minimum throttle position. The throttle is not set to zero because the grade utilized is only the average along the link, and even when a train is going downhill for a significant distance, the motors will sometimes be kept running to prevent the railcars from bunching together and reducing stability.

Inertial resistance: Inertial resistance is the positive or negative impedance from the train changing its velocity. Trains must use energy to accelerate, and much of that energy is not regained when slowing down due to friction or the need to slow down faster than would otherwise be necessary. The inertial resistance is $m_T a$. This study assumes that the positive and negative inertial resistances will cancel each other out. In reality, more energy is used in accelerating a train than the energy saved as a train decelerates, so this is one category where the total energy consumption is understated and future analysis could improve upon.

Total resistance: Combining the preceding formulae yields the total average resistance

along a link i as:

$$R_i = \sum_{k=1}^{n_L+n_C} (am_{g,k} + bN_{ax,k} + K_k k_{a,i} v_i^2) + v_i(B_L n_L + B_C n_C) k_{f,i}$$

$$+ m_T g(gr_i) + 0.4536 m_T g \arcsin\left(\frac{15.24}{r_i}\right) + R_{brake,i} + m_T a \tag{4.13}$$

where k refers to each rail vehicle in the train unit.

Power level and speed: The power level used along the link, P_i is a function of the link's resistance and the train's speed along the link:

$$P_i = R_i v_i \tag{4.14}$$

The train unit has a discrete number of possible power levels, which it will use to approach the desired speed along each link. Equations (4.13) and (4.14) are solved iteratively to determine each link's associated power level, P_i, and base travel time, $t_i^0 = l_i/v_i$. Those values can then be used in equation (4.12).

4.3.4 Switching costs

For many O-D pairs, the lowest-cost route for the shipper may involve switching between electric and diesel-electric operations en-route. This might occur when a small portion of the overall network is electrified and an electrified path does not exist between on O-D pair, or the electrified route is far enough away from the direct path that the savings from electric operations are not worth the cost in added time. If the network does not allow any switching, electric trains would only be assigned between O-D pairs that have fully electrified, mostly direct, paths between them, which is not realistic. The network allows flows to switch from electric to diesel-electric operations (and vice versa) at certain nodes in order to simulate more realistic routing.

461 nodes in the rail network have facilities adequate to allow for switching. At those nodes, links between the electric and diesel-electric links represent the cargo time, crew time, and energy costs associated with switching locomotives. The cost assumes an hour-and-a-half per switch, and that six employees will be involved (including the train

operators and the yard workers). Whenever the path switches between operations, each train accrues a cost of roughly $3800.

4.4 Solution methods

This section is divided into two parts. The first part discusses the genetic algorithm used to solve the upper-level optimization problem (selection of links to electrify). Within each generation, the "fitness level" of every solution must be calculated. This involves solving a symmetric TAP subproblem to calculate the link flows, and using these values to calculate the value of the upper-level objective function. Most TAP solvers assume separable link costs, and the second part of this section explains the changes we made to accommodate interactions.

4.4.1 Genetic algorithm

Procedure 4 describes how the genetic algorithm operates. GA is a heuristic and solution optimality is not guaranteed. Therefore, this set of experiments test the base problem for the best obtainable solution, across a set of GA inputs (and random starts) listed below:

- population size = 15

- generations = 150

- crossover probability = 0.1

- mutation probability $\in \{0.2, 0.4\}$

- elitism (best found solution each iteration carried to next iteration): True

- runtime: No limit

A GA solution is defined by a binary string of length $|A|$, each digit corresponding to a unique link and its electrification status. For the TAP sub-problem, the default settings from the repository were used, and the target relative gap level was set to 10^{-4}, as recommended by Patil et al. (2021) for comparisons based on total cost.

The GA base case solution was then compared with solutions obtained from an "engineering judgement" heuristic: sorting links in descending order of operational saving per cost of electrification $x_i(c_i' - c_i)/c_i^e$, and selecting links until the budget is exhausted.

Algorithm 4: Pseudocode for genetic algorithm

    Genetic algorithm (*N, A, B*)**:**

        Initialize population;

        Evaluation population;

        while *Termination criteria not met* **do**

            Select individuals with lowest total cost;

            Breed new individuals through crossover and mutation;

            Evaluate individual feasibility and fitness;

            Replace least fit population with new individuals;

        end

This heuristic implicitly considers track resistance in the improvement between c_i and c_i'. The heuristic depends on link flow, and therefore requires a single run of the rail TAP to obtain base link flows on all links. However, the heuristic overemphasized links that would benefit from improved capacity in general, rather than electrification in particular, and was therefore not selected.

4.4.2   Subproblem solution

Most algorithms for TAP are designed for the separable cost case. In this sub-section, we describe the changes that need to be made for our cost structure, and derive a new flow shift formula for this case. Our solution method is based on Algorithm B (Dial, 2006). The main points of this algorithm are disaggregating link flows by origin, where the sets of used links form acyclic subnetworks; shifting flows from longer-cost to shortest-cost segments within each subnetwork; and updating these subnetworks by removing unnecessary links and adding improving ones providing shorter paths.

It is the second component (flow shifting) whose original derivation relied on separable link costs, and which must be re-examined in light of interactions. We first show that the link interactions in our specific formulation satisfy the symmetry condition needed for the line integral formulation (4.4)–(4.8) to hold, and then re-derive the flow shift formula for Algorithm B.

As discussed in the Cost Formulas section, all flow continuing on the same fuel type has zero switching cost, whereas all flow changing fuel type has pre-defined switching costs. Note that only the parallel diesel-electric link pair has cost function dependent on the sum of the flows, and $t(x)$ is identical for both links. The Jacobian takes the form:

$$J_{\mathbf{x}} = \begin{pmatrix} \frac{\partial(c_1(\mathbf{x}))}{\partial x_1} & \frac{\partial(c_2(\mathbf{x}))}{\partial x_1} & \cdots & \frac{\partial(c_n(\mathbf{x}))}{\partial x_1} \\ \frac{\partial(c_1(\mathbf{x}))}{\partial x_2} & \frac{\partial(c_2(\mathbf{x}))}{\partial x_2} & \cdots & \frac{\partial(c_n(\mathbf{x}))}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial(c_1(\mathbf{x}))}{\partial x_n} & \frac{\partial(c_2(\mathbf{x}))}{\partial x_n} & \cdots & \frac{\partial(c_n(\mathbf{x}))}{\partial x_n} \end{pmatrix}$$

The off-diagonal elements are zero whenever the two links do not interact. In our case, the only non-zero off diagonal elements are for parallel diesel-electric pairs. These elements are calculated using the chain rule:

$$\frac{\partial t(x_D + x_E) + c_D}{\partial x_E} = \frac{\partial t(x_D + x_E)}{\partial(x_D + x_E)} \cdot \frac{\partial(x_D + x_E)}{\partial x_E} = \frac{\partial t(x_D + x_E)}{\partial(x_D + x_E)} \tag{4.15}$$

$$\frac{\partial t(x_D + x_E) + c_E}{\partial x_D} = \frac{\partial t(x_D + x_E)}{\partial(x_D + x_E)} \cdot \frac{\partial(x_D + x_E)}{\partial x_D} = \frac{\partial t(x_D + x_E)}{\partial(x_D + x_E)} \tag{4.16}$$

As can be seen, these two derivatives are identical, proving that the Jacobian is a symmetric matrix, and establishing the validity of the formulation (4.4)–(4.8).

We next show that the Algorithm B flow shift procedure (equalizing cost between routes) is valid even in the presence of link interactions. This result is more general, and can be applied to *any* instance of symmetric costs depending on at most two links, not just the ones we adopt in our study. We separate out the terms involving interactions as follows: $c_i = f_1(x_i, x_j) + f_2(x_i)$ and $c_j = g_1(x_i, x_j) + g_2(x_j)$, with $\frac{\partial f_1(x_i, x_j)}{\partial x_j} = \frac{\partial g_1(x_i, x_j)}{\partial x_i}$.

In this operation, we have identified two paths (a lower-cost path π_L and a higher-cost π_U), and wish to shift flow between them to minimize the objective. Recall that the line integral formulation can now be written as:

$$F(\mathbf{x}) = \oint_{\mathbf{0}}^{\mathbf{x}} \mathbf{c}(\mathbf{s}) \cdot d\mathbf{s} \tag{4.17}$$

Partition the arc set $\mathbf{A}$ into three subsets A_1, A_2, and A_3. Links in set A_3 have separable cost, depending only on their own flow. The travel time on the links in sets A_1 and A_2 depend on the flows on two links: the link itself, and exactly one link in the other set. Mathematically, there is a bijection between sets A_1 and A_2, with the notation $i(j)$ and $j(i)$ used to denote the counterparts of links in the other set. The notation for these cost functions lists the links own flow first, and its counterpart's flow second, that is, $t_i(x_i, x_{j(i)})$ and $t_j(x_j, x_{i(j)})$. Let $p = |A_1| = |A_2|$ and $m = |A_1| + |A_2| + |A_3|$. The index a will be used to denote a generic link, if it doesn't matter which set it's from.

As the line integral is path-independent, we choose the following integration path:

$$(0, 0, 0, \ldots, 0) \to (x_1, 0, 0, \ldots, 0) \to (x_1, x_2, 0, \ldots, 0) \to \cdots \to (x_1, x_2, \ldots, x_m) \, .$$

The line integral then decomposes into a sum of ordinary integrals:

$$F(\mathbf{x}) = \sum_{i=1}^{p} \int_{(x_1, \ldots, x_{i-1}, 0, 0, \ldots, 0)}^{(x_1, \ldots, x_{i-1}, x_i, 0, \ldots, 0)} t_i(\mathbf{s}) \cdot d\mathbf{s} + \sum_{j=p+1}^{2p} \int_{(x_1, \ldots, x_{j-1}, 0, 0, \ldots, 0)}^{(x_1, \ldots, x_{j-1}, x_j, 0, \ldots, 0)} t_j(\mathbf{s}) \cdot d\mathbf{s}$$

$$+ \sum_{k=2p+1}^{m} \int_{(x_1, \ldots, x_{k-1}, 0, 0, \ldots, 0)}^{(x_1, \ldots, x_{k-1}, x_k, 0, \ldots, 0)} t_k(\mathbf{s}) \cdot d\mathbf{s}$$

$$= \sum_{i=1}^{p} \int_{0}^{x_i} t_i(s, 0) \, ds + \sum_{j=p+1}^{2p} \int_{0}^{x_j} t_j(s, x_{i(j)}) \, ds + \sum_{k=2p+1}^{m} \int_{0}^{x_k} t_k(s) \, ds \, .$$

Let I_a be $+1$ if link a is on π_L, -1 if a is on π_U, and 0 otherwise. Then for a flow shift Δx from the longest to the shortest path, the change on each link's flow is given by $\Delta x_a = I_a \Delta x$ and $I_a = \partial(\Delta x_a)/\partial(\Delta x)$. (This latter equation will be used when differentiating via the chain rule below).

The objective in terms of a flow shift of size Δx is now

$$F(\Delta x) = \sum_{i=1}^{p} \int_{0}^{x_i + \Delta x_i} t_i(s, 0) \, ds + \sum_{j=p+1}^{2p} \int_{0}^{x_j + \Delta x_j} t_j(s, x_{i(j)} + \Delta x_{i(j)}) \, ds + \sum_{k=2p+1}^{m} \int_{0}^{x_k} t_k(s) \, ds \, ,$$

and when we differentiate with respect to Δx, we obtain the following (the second term splits into two terms from the Leibniz rule):

$$\frac{dF}{d(\Delta x)} = \sum_{i=1}^{p} t_i(x_i + \Delta x_i, 0) I_i + \sum_{j=p+1}^{2p} t_j(x_j + \Delta x_j, x_{i(j)} + \Delta x_{i(j)}) I_j$$

$$+ \sum_{j=p+1}^{2p} \int_{0}^{x_j + \Delta x_j} \frac{\partial t_j}{\partial x_{i(j)}}(s, x_{i(j)} + \Delta x_{i(j)}) I_{i(j)} \, ds + \sum_{k=2p+1}^{m} t_k(x_k + \Delta x_k) I_k \, .$$

We will show that $dF/d(\Delta x) = \sum_{a \in (A \cup A^E)} t_a(\mathbf{x} + \Delta \mathbf{x}) I_a$. The interpretation is that the derivative vanishes if Δx equalizes the cost difference on the longest and shortest segments, thus minimizing the Beckmann function. The second and fourth terms in the sum

56

are exactly what we need for the links in A_2 and A_3. We need to show

$$\sum_{i=1}^{p} t_i(x_i + \Delta x_i, 0)I_i + \sum_{j=p+1}^{2p} \int_0^{x_j + \Delta x_j} \frac{\partial t_j}{\partial x_{i(j)}}(s, x_{i(j)} + \Delta x_{i(j)})I_{i(j)} \, ds$$

$$= \sum_{i=1}^{p} t_i(x_i + \Delta x_i, x_{j(i)} + \Delta x_{j(i)})I_i \, .$$

Using the symmetry condition $\frac{\partial t_j}{\partial x_{i(j)}} = \frac{\partial t_i}{\partial x_{j(i)}}$, we have

$$\sum_{i=1}^{p} t_i(x_i + \Delta x_i, 0)I_i + \sum_{j=p+1}^{2p} \int_0^{x_j + \Delta x_j} \frac{\partial t_i}{\partial x_{j(i)}}(s, x_{i(j)} + \Delta x_{i(j)})I_{i(j)} \, ds \, .$$

Using the fundamental theorem of calculus and one-to-one correspondence of terms in the two summations, it is clear that it equals $\sum_{i=1}^{p} t_i(x_i + \Delta x_i, x_{j(i)} + \Delta x_{j(i)})I_i$, which is the required form. Therefore, $dF/d(\Delta x) = \sum_a t_a(\mathbf{x} + \Delta \mathbf{x})I_a$ and Algorithm B flow shifts are valid in our setting. Most implementations of Algorithm B (including ours) use a Newton method to equalize the costs on these links. The denominator in Newton's method can also be adjusted to reflect interactions, as discussed more in Patil and Boyles (2022), but even without this change existing implementations should still converge to the equilibrium.

4.5   Data sources

After obtaining solutions and validating the solution method for our formulation, the second set of experiments involves policy testing and sensitivity analysis. These experiments vary parameters such as total budget, demand data, electrification costs, electricity costs, crew/cargo costs, policy changes in the form of monetary incentives. The base electrification budget is \$30 billion, roughly equivalent to electrifying $65,000$ kilometers of track. This is varied by up to $\pm 20\%$, or \$24–36 billion. The analysis involves studying the return on investment (ROI) in the form of reduced costs, incentivizing policymakers and private stakeholders to upgrade infrastructure.

FAF4 provides demand data from 2010 and 2020, as well as forecasts for 2030 and 2040 (Bureau of Transportation Statistics, 2020). In addition, lowering the usage cost of rail freight allows for modal shift from trucking to rail. This potential demand

variation is considered by increasing the demand data by up to 25% from base values. Lastly, electrification costs and crew/cargo costs are also tested at increased values (+25%) to gauge effects on the electrification solution. Under the current experiment formulation, adjusting the electrification costs is equivalent to adjusting the budget for electrification, so that parameter was not explored.

The North American railroad network has been extracted from the statewide analysis module (Alliance Transportation Group, 2013) provided by the Texas Department of Transportation, originally based on the CTA rail network developed at Oak Ridge National Laboratory (Center for Transportation Analysis, 2014). The dataset includes 35,424 links, and has geographical information (length, latitude, longitude, grade category), ownership information, and other auxiliary information. There are 28,289 nodes connected by these links, denoting stations, yards and interchanges. Elevation data were obtained by overlaying the network on the North American elevation grid (U.S. Geological Survey, 2007) using ArcGIS (Figure 4.2). The demand data (in tons) has been obtained from the Freight Analysis Framework, version 4 (FAF4).

As specified, operational mode changes can only occur at yards. Therefore, corridor electrification is more sensible than individual link electrification. Candidate corridors were obtained by connecting each yard to the nearest neighboring yards. Specifically, the following method was employed to find the set of candidate corridors and results can be seen in Figure 4.3. This new shortest corridor network has tracks totaling $170,000$ kms of track available for electrification, down from over $305,000$ kms in the full network. The switching costs were obtained using the following assumptions for the locomotives: 1.5 hours to switch, throttle position 1 for diesel-electrics, 10% power for electrics, and 6 crew-equivalents manpower.

1. Calculate all pair shortest paths from all yards

2. Calculate all pair shortest paths from all yards, but stop exploring a branch if it reaches a yard.

3. Compare the two sets and keep the paths common to both sets
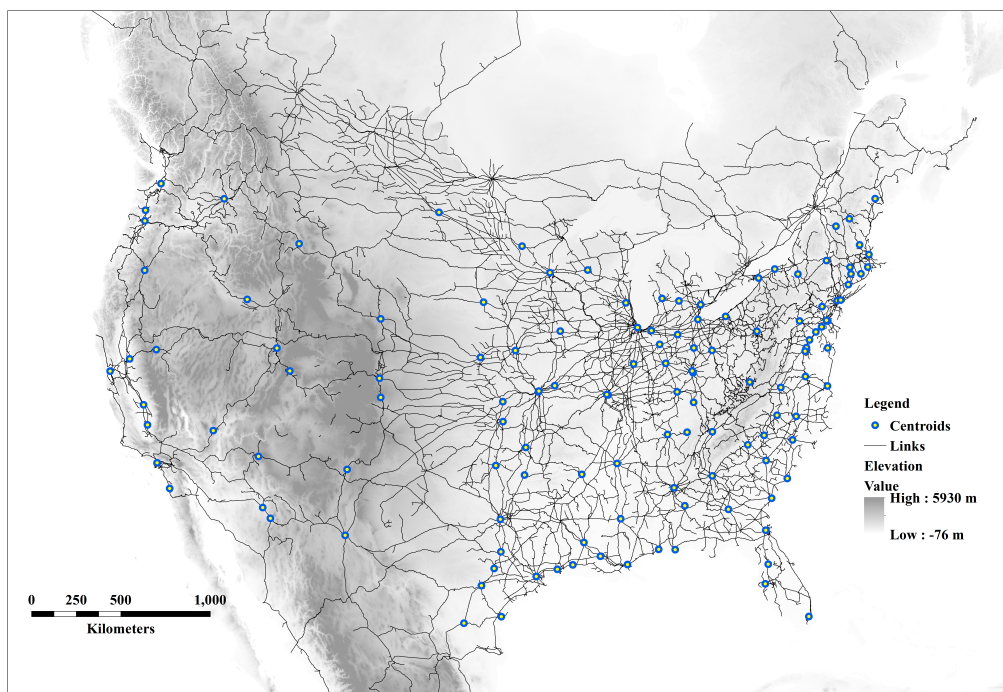
Figure 4.2: North American rail network representation with elevation data
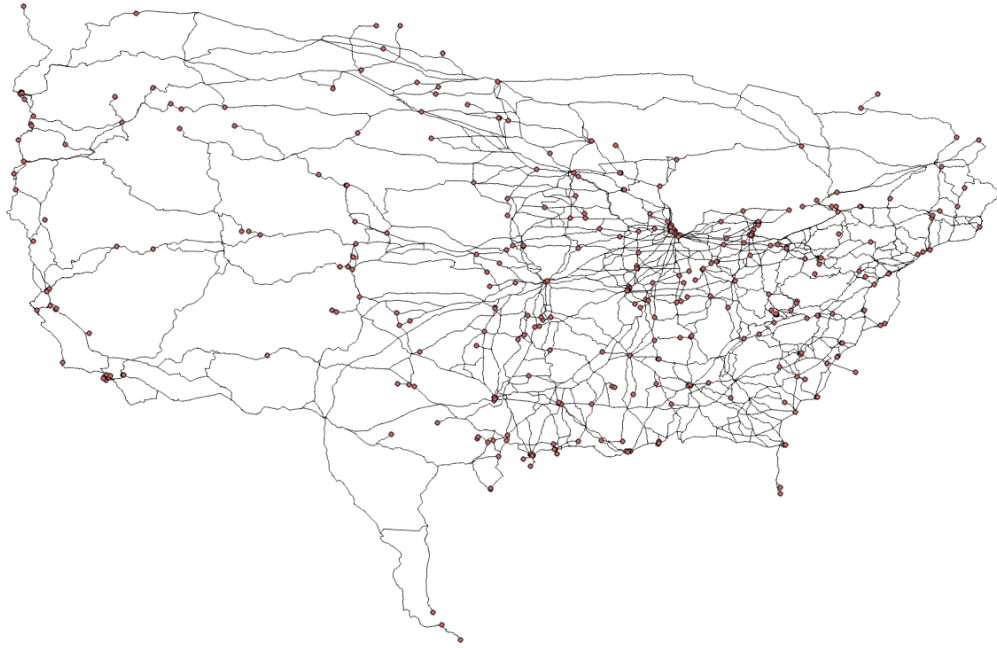
Figure 4.3: Candidate corridor visualization with yards

4.6 Results and Sensitivity analysis

All result visualizations are shown in Figures 4.4–4.7. The low budget, base case (medium budget), and high budget cases lead to electrification of about 49,000 (30,000), 61,000 (36,000), and 70,500 (42,000) kms (mi), respectively. Only the base case results are shown with the full network backdrop, all other results are shown on the shortest path corridor network for better visibility. Figures 4.5 - 4.7 show the results as the overlap of different scenarios. Figure 4.5 is a cumulative plot, where the medium budget (base case) and high budget results are shown as increments over the low budget case. Figures 4.6 and 4.7 pivot off the base case, therefore, we highlight the overlapping links as well as the differences.

The first observation is that the GA retained all corridors selected for the low budget case in the base case and high budget case. However, this changes for the increased demand and increased costs scenarios, given that the network OD matrix and costs change in the two scenarios. This allows us to identify the most impactful corridors and stations, such as all three of the major transcontinental routes (LA-Chicago, Oakland-Chicago, and Seattle-
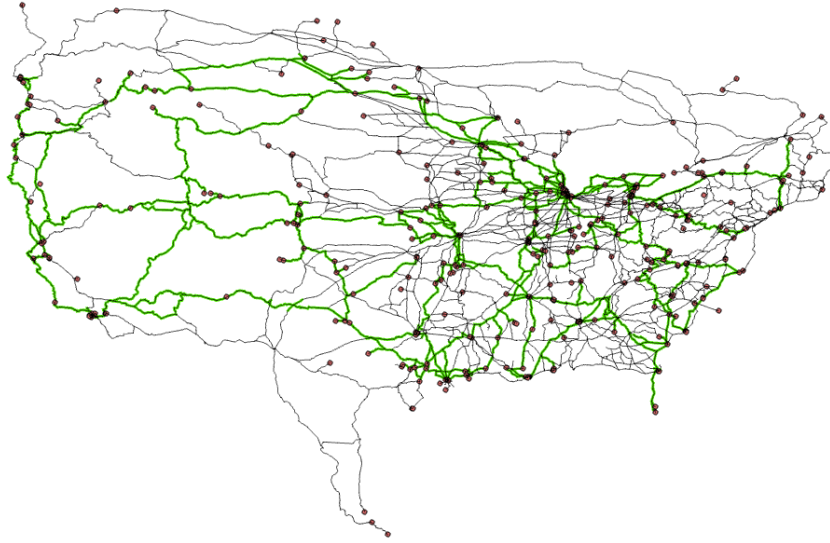
(a) Base case results with shortest path network
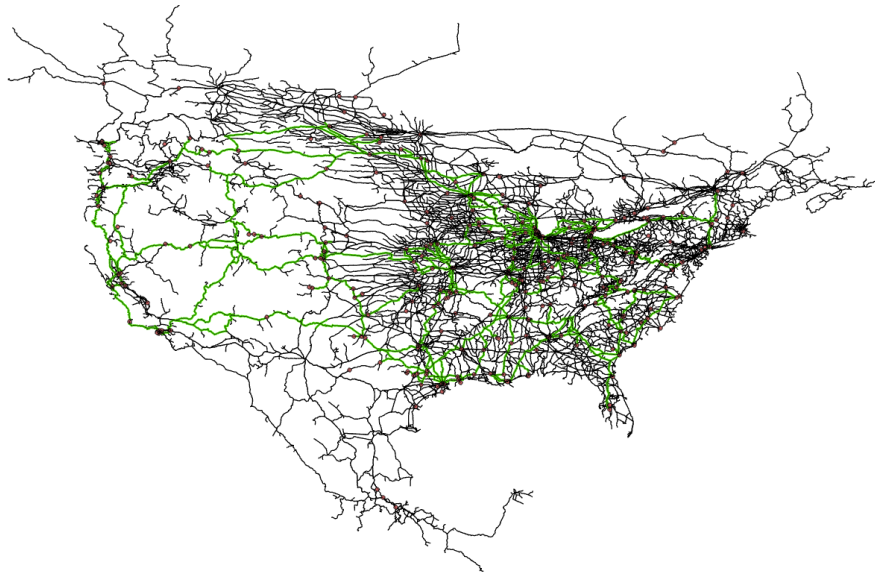
(b) Base case results with full network

Figure 4.4: Base case visualizations

61
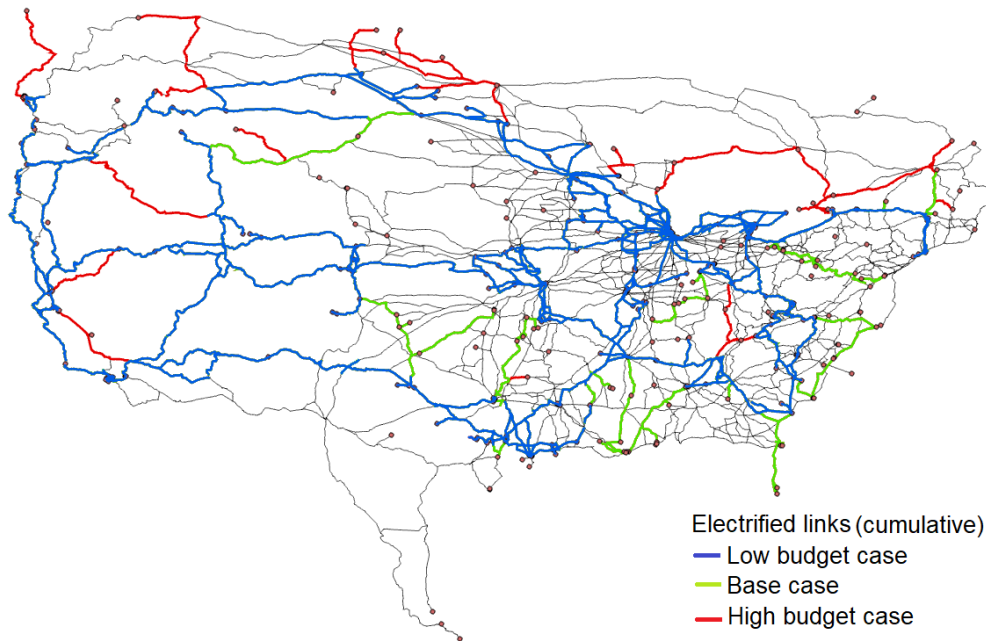
Figure 4.5: Results for varying electrification budget

Chicago), which the algorithm chooses for electrification in each scenario. There is a wide variety of corridors selected across the entire network. The algorithm selects quite a few mountainous routes where electrification provides the most benefits per train, but the algorithm also appears to select for links that provide connectivity throughout the network. This trend is best illustrated in figure 6, which seems to show that the increased demand case shifts the selection from the more mountainous west to the more populous east and gulf coasts. With higher demand, links that provide smaller benefits per train are more likely to provide greater benefits overall.

Examining the base case in further detail, the algorithm selects 13.2% of the network (by line-miles) for electrification, and predicts that 15.5% of final flow (by tonnage) will use electrical traction. This reflects that busier corridors benefit more from electrification. This also implies that most of the traffic along electrified links would be using electric rather diesel-electric locomotives.

The increased demand scenario increases electrified corridor connectivity to several new yards/stations. The increased cost scenario reduces selection from east and central US, and chooses corridors from western US. We hypothesize this selection occurs due to the

62
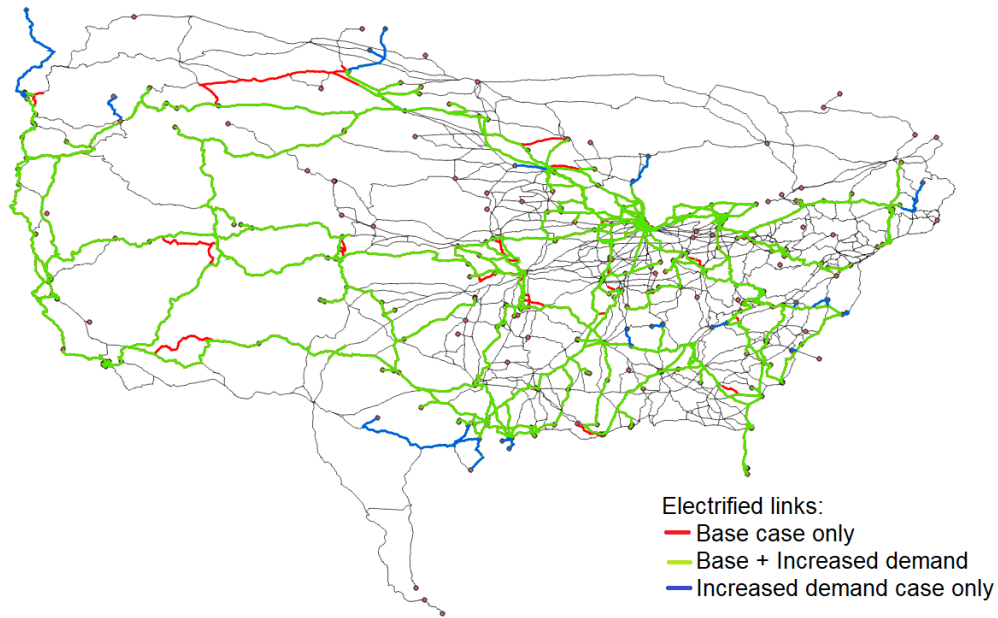
Figure 4.6: Results for increased demand case

higher elevations and grades for these tracks seeing benefit from electrification. The current formulation does not incorporate any regional variations in the wholesale cost of electricity or diesel.

An important question is the validation of the lower level flows. Uddin and Huynh (2015) have shown the comparison of static traffic assignment (w/o electrification) patterns for major links, given default link performance functions. Our model for the lower-level problem adapts the cost function parameters to incorporate additional cost factors. Aggregate metrics are easily available, and therefore can be compared to or model. We present a comparison between our aggregate assignment metrics for the lower-level problem and those provided by Bureau of Transportation Statistics data as well as Commodity Flow Survey data. The total rail freight ton-miles (in millions) from our model is 1,627,854 as compared to 1,712,567 for BTS data and 1,387,777 from the CFS data. Thus, it is within 5% of the BTS estimates and within 15% of the CFS estimates. Note that BTS and our model uses FAF4 demand with different assignment model, while CFS data relies on survey responses which can differ significantly.

Each sub-problem took 5 minutes to converge, and each GA instance took 36 hours
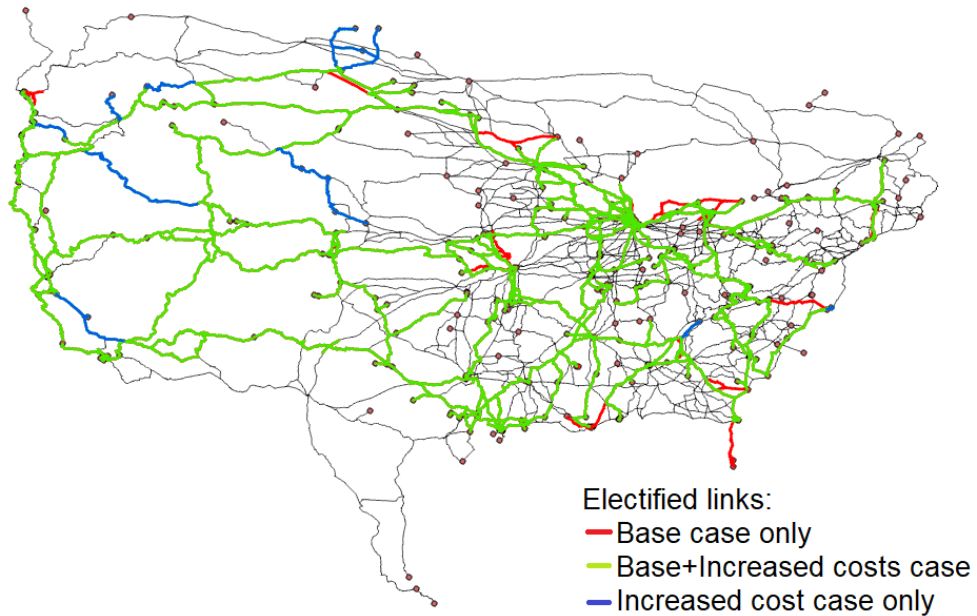
Figure 4.7: Results for increased operations cost case

to finish on a Linux PC with i7 Processor (2.6 GHz) and 16 GB RAM. Given the long term planning horizon, runtimes are not crucial, but still useful to scalability of the heuristic for a large network.

4.7 Conclusions and future directions

This study presents a novel way of looking at the rail network electrification problem using connections to the road traffic assignment problem. We formulate the problem as a bi-level network design problem, where the lower level problem is a symmetric traffic assignment problem, assigning goods flow instead of traffic. We then show the correctness of Algorithm B flow shift formula for the symmetric traffic assignment problem, using it to solve this sub-problem. The costs and network parameters incorporate electrification costs; fuel, locomotive, and operational costs; and train resistance costs.

The North American railroad network is used as a test network to demonstrate our method, show scalability for large networks, and draw insights. We observe and note some corridors chosen in all different testing scenarios (varying budgets, opex, and demand), as

well as provide policy insights for planning. The key observations are as follows:

- While there seems to be some evidence of corridors in more mountainous terrain producing better cost savings, overall there is a wide variety of corridors selected across the entire network. This implies that providing broad connectivity might be more important than electrifying the corridors with the best savings on a per-train basis.

- The results for the increased demand case, illustrated in figure 6, indicate a general shift in chosen corridors from the more mountainous west towards the more populous east and gulf coasts. This is probably a result of the increased demand causing those corridors to generate higher savings, even though they have fewer savings per train. This is also evidence that the highest priority corridors might be routes through rough terrain operating at or near capacity.

- Sensitivity to demand means long-term trends in the trucking industry could have a substantial influence on the highest priority corridors for electrification.

The main limitations of this work are: 1) static power levels (once calculated); 2) static brake usage assumption on links; 3) static electricity and diesel costs; and 4) lack of incorporation of track ownership restrictions on path selection. The first three limitations are relatively straightforward to tackle, although they increase problem complexity and computation time significantly by introducing another layer of iterative calculations.

An additional complication is partly based on the data available: because this analysis uses large zones with few centroids, the simulations might overestimate the benefit of electrifying well-traveled corridors that connect multiple zones. Studying how the selection changes based on varying switching costs might show whether the results are skewed by simulating a higher amount of connectivity than would actually exist. This problem could partially be ameliorated by applying a cost to electric trains reaching centroid connectors to reflect that some electric trains might require additional drayage over diesel-electric trains. Properly calibrating such a cost would remain a difficult problem.

The methods formulated in this study can be adapted to consider social benefits from changes in emissions, formulating the problem as one of social benefit maximization rather than private cost minimization. The composition of the power grid affects the emissions an electric locomotive causes, meaning the marginal social benefits of electrification and the marginal private cost savings can vary substantially from link to link. In some parts

of the power grid where most of the electricity is provided by coal, an electric train might even produce more emissions than an equivalent diesel-electric train (Walthall, 2019). Understanding social benefits is important because the high variability of private return-on-investment from electrification might necessitate public subsidies before capital construction becomes possible. Prior studies have formulated social benefit maximization problems as a single-level problems, thus reducing computational complexity.

Our formulation could also be generalized to reflect investments made at different points in time, along with variations in future demand. Such a dynamic model could present valuable insights about the proper timing of electrification, but poses challenges in estimating future demand, costs, and budgets.

Chapter 5

Effects of Origin-Destination Matrix Errors on User Equilibrium

5.1   Introduction and motivation

Network models play a central role in transportation planning and forecasting. The inputs to these models include both infrastructure and demand characteristics. To the latter end, public agencies often have strategic plans forecasting population and traffic patterns decades into the future (Texas Department of Transportation, 2015, 2019; State of California, 2018; Massachusetts Bay Transportation Authority, 2018). Forecasting demand this far in advance is challenging. For example, there may be substantial changes in technology, land use, policy, and demographics, which can significantly impact travel demand and its spatial distribution.

Given that demand matrices will likely have inaccuracies, despite best efforts at data collection and demand modeling, it is important to understand how any errors in demand matrices translate to errors in metrics used for planning (link flows, vehicle miles-traveled, etc.) . The answer depends critically on the magnitude of demand errors, and on its sources — systematic and random errors will have different effects.

Vehicle and transit OD matrices have traditionally been constructed from household travel demand survey data. With technological advances, sources such as origin-destination surveys, mobile phone data (Iqbal et al., 2014; Bonnel et al., 2018; Wang et al., 2018b), GPS data (Çolak et al., 2015), social media data (Lee et al., 2019; Cheng et al., 2020), fare collection data (Egu and Bonnel, 2020), etc. are now available for generating OD matrix data. These methods lead to different OD matrix values, prompting investigations

into similarities and differences between matrices developed from different methods (Day-Pollard and van Vuren, 2015; Caceres et al., 2020; Egu and Bonnel, 2020). Some notable observations from these studies are listed below:

- Household surveys no longer provide accurate data due to factors such as reporting errors, sampling biases, and increasing costs. Other data sources are proving to be cheaper, provide larger sample size, reduce time for collection and processing, and provide information on temporary mobility patterns. Therefore, household surveys produce random, uncorrelated errors in demand forecasting.

- Case studies show that household survey data uniformly underestimates trips (by up to 30%), attributed to lack of survey information for non-residents taking work trips, or survey and reporting biases.

- Traditional surveys have trouble capturing demand patterns from sparsely populated areas such as mass transportation facilities, industrial parks, hospitals, etc. where no households exist. This leads to spatial errors within specific geographical zones while forecasting demand.

Multiple studies cited above either call for or propose a hybrid approach to demand forecasting integrating several data sources. While promising in reducing forecasting errors, they cannot eliminate them entirely (the future is inherently unknowable), and it is therefore important to know what kind of improvement in decision-making capability can be obtained by reducing demand error to a lower level.

This study aims to address this question empirically, i.e., how do errors in demand forecasts propagate into errors in model outputs? A practical use case for this analysis would for a planner to account for historical inaccuracies in their demand data source to improve future planning. Furthermore, if a desired level of accuracy is required in model outputs for meaningful planning insights, we identify how much accuracy is needed in the demand forecast. The magnitude and impact of demand errors also have implications on model selection; when errors are higher, simpler models more robust to these errors can produce more accurate results than more sophisticated and "realistic" models (Boyles and Ruiz Juri, 2019)

Concretely, this chapter focuses on the impact of demand uncertainty on static traffic assignment under Wardropian user equilibrium (UE) (Beckmann et al., 1956). We form our understanding by performing numerical experiments on a variety of networks. We

specifically test three types of potential errors in demand: (a) uniform over- or under-prediction across all origin-destination pairs, representing a systematic error in forecasting; (b) noise in the specific entries of the demand matrix, representing random errors; and (c) spatial correlation in demand perturbations, representing errors in land use forecasting in specific parts of a network. These three error types have been chosen specifically due to observations and identification among prior studies in this area, and to cover a wide range of potential error sources. In practice, this chapter helps identify the level of accuracy necessary in the source OD matrix estimations for traffic forecasts.

The main contributions of this chapter are as follows:

1. Characterization of the effect of uniform, OD-specific, and spatially correlated demand errors on equilibrium system travel time, system congestion, and vehicle miles traveled.

2. Comparison of forecast and observed demand growth on a large case study network to demonstrate potential usage for planning purposes

Our hope is that practitioners can use these results to understand what level of accuracy they need in their demand forecasting, in order to produce acceptable model outputs for planning. As the specific sources of error in demand forecasting vary among regions and agencies, using our results in this way would require identifying which of our three error models best describes the types of errors encountered in a specific application context. To this end, our description of the three error models is paired with real-world examples that might lead to each type of error we model.

This chapter is organized as follows: the next section reviews the current literature on this topic, followed by a description of the network data and experiment design. The experiments' results are presented, and their insights applied to a case study in a practical network representing Austin, TX. We conclude with final takeaways, guidelines for practice, and avenues for future work.

5.2   Literature review

Both intuition and past research suggest that longer timescales for forecasting induce more forecasting error (Odeck and Welde, 2017; Odeck, 2013; Andersson et al., 2016; Cruz and Sarmento, 2019). Some reasons for the observed mismatch between forecasts and reality are:

1. Inaccuracy of forecasting models and/or data (Nicolaisen, 2012)

2. Land usage or demographic change (Naess et al., 2015)

3. Change in relevant infrastructure development or construction timeline (Parthasarathi and Levinson, 2010)

4. Optimism bias (Bain, 2009)

(Cruz and Sarmento, 2019) discuss other reasons and provide background reading on this issue. The error in demand forecasts naturally raises questions about the effect of input uncertainty on forecasting in terms of output uncertainty. De Jong et al. (2007) provide a review of previous studies which investigate the propagation of uncertainty, as well as provide confidence intervals for travel times and distances under uncertainty. Rasouli and Timmermans (2012) follow up on this work by surveying the literature for uncertainty propagation in travel demand modeling. Hartgen (2013) discusses how 20-year forecasts are likely off by a minimum of ±30%, while other estimates can be off by higher percentages for shorter durations. Flyvbjerg et al. (2005) discovered that about 90% of rail project forecasts were over the observed demand by an average of 106%. The same study also discussed the lack of substantial improvement in forecasts despite progress in corresponding forecasting models. Nicolaisen and Driscoll (2014) as well as Van Wee (2007) corroborate and discuss their observations of forecasting errors in infrastructure projects. Multiple studies discuss the effects of demand uncertainty on major infrastructure project applications such as evacuation planning, roadway pricing, network design problems, decision analysis for projects, and the appropriate choice of modeling tools (Duthie et al., 2011; Waller et al., 2001; Ng and Waller, 2010; Ukkusuri et al., 2007; Yin et al., 2009; Boyles et al., 2018; Gardner et al., 2011; Li et al., 2012; Patil et al., 2017; Boyles and Ruiz Juri, 2019; Venkatraman et al., 2021).

Zhao and Kockelman (2002) study the propagation of induced error across the four step model in terms of compounding versus damping trends for each step. Input errors were compounded in the trip generation, trip distribution, and mode choice steps, whereas traffic assignment seemed to dampen some of the noise amplification from the first three steps. Specifically considering traffic assignment, Waller et al. (2001) discuss the introduction of uncertainty in the origin-destination matrix and how variance should not be neglected, as using average results in overestimating network performance. The study recommended using Monte Carlo simulations or inflation factors to account for input variance, in addition to using multiple evaluation points. It provided motivation for multiple robust optimization

70

approaches and stochastic analysis in both the static and dynamic assignment realms (Duthie et al., 2011; Ng and Waller, 2010; Yin et al., 2009; Gardner et al., 2008; Do Chung et al., 2012), but there has been no comprehensive quantification of the impact of uncertainty on traffic assignment.

5.2.1 Error type discussion

This chapter discusses error types and their effect, and it is important to understand what situations can give rise to these three types of errors. A brief overview was provided in the introduction, this subsection delves deeper into these error types.

Spatial errors: Caceres et al. (2020) observe spatial errors in trips for zones associated with mass transportation facilities, university campuses, industrial parks and hospitals. These zones are often not included in traditional surveys due to lack of residential units, but are captured with mobile data or intercept surveys. Egu and Bonnel (2020) observe spatial errors in OD matrices due to survey data trips being skewed spatially towards peripheral zones. Janzen et al. (2018) observe spatial disparities between call record data and travel survey data, attributed to potential over-sampling of frequent callers and long-distance travelers, correlated with higher education and wealth levels. Iqbal et al. (2014) note that call record data is skewed toward certain zones such as mass transportation facilities, corroborating Cacares et al.'s observations. Osorio-Arjona and García-Palomares (2019) observe spatial disparities between twitter trip data and survey data, noting that central districts of Madrid are overestimated in the Twitter data, potentially due to data bias (Twitter users are predominantly 20-39 y.o.) or lower data from commercial/industrial zones and low-income areas.

Correlated errors: Caceres et al. (2020) also compare the OD matrices obtained from traditional survey data and mobile phone data. Using matrix sparsity analysis, they conclude that mobile data captures a higher percentage of mobility of all possible OD connections and demonstrate a strong linear relationship between trips generated according to survey data versus those generated according to phones. Egu and Bonnel (2020) observe systemic underestimation of trips forecast using survey data by up to 30% on an average weekday. They hypothesize that the cause is either lack of non-resident data in surveys (unlikely) or under-reporting of trips (validated using prior GPS data studies). Bonnel et al. (2015) observe uniform errors between survey data trips and mobile data trips with a near linear relation ($R^2 = 0.96$) with a slope close to 1. At the other extreme, Janzen et al. (2018)

find that survey data results in uniformly under-reported trips by a factor of 2. They note that this error is caused by inaccurate weighing constants in mathematical models, skewing the entire trip table uniformly.

Random errors: These types of errors are the hardest to quantify, solely due to the sheer variety of their possible sources. Surveyed studies observe this type of error in OD matrices due to survey biases, small sample sizes, inaccurate reporting, falling response rates, level of detail, and modelling assumptions.

5.3   Background and experiment design

The real-world networks, their characteristics, and the sample sizes we used to study their behavior are specified in Table 5.1. All networks (save for the one used in the case study) have been sourced from the Transportation Networks for Research repository (Stabler, 2019). For our experiments, we used an implementation of Algorithm B (Dial, 2006), a bush-based algorithm, to solve the traffic assignment problem. The source code for the following experiments is publicly available (Boyles, 2019; Liao, 2019).

As can be seen in Table 5.1, a variety of network sizes were chosen to provide better insight into error propagation. Networks were chosen for low computation time to enable Monte Carlo sampling of perturbed OD matrices, as well as having node coordinate information to model spatial correlation. The exact sample size used for Monte Carlo simulations n for each network is shown in the table, which also includes flow-weighted volume-capacity ratio (WVC) as a measure of network-wide congestion. (The exact definition of this metric is provided later in this section).

Table 5.1: Description of Networks Used

	Nodes	Links	Trips	WVC	n
Berlin-Friedrichshain	224	523	11,205	0.473	200
Berlin-Mitte-Center	398	871	11,482	0.333	200
Berlin-Mitte-Prenzlauerberg-Friedrichshain-Center	975	2,184	23,649	0.286	100
Berlin-Prenzlauerberg-Center	352	749	16,660	0.431	200
Berlin-Tiergarten	361	766	10,755	0.313	200
SiouxFalls	24	76	360,600	1.474	500

5.3.1 STA experiment design

In order to simulate different sources of prediction error, we perturbed OD matrices in three ways.

The first method is **uniform perturbation**, also referred to as correlated error. This method globally scales the demand values d^{rs} by a common value. This is intended to simulate systematic errors in forecasting network demand, perhaps due to an estimation error in trip generation parameters, or failing to anticipate technological changes with network-wide impacts on demand. Specifically, for some constant k, in uniform perturbation we replace each entry in the OD matrix with kd^{rs}.

The second method is **OD-specific perturbation**, also referred to as random error. This method replaces each OD matrix entry with $k^{rs}d^{rs}$, where k^{rs} is an OD-specific scaling factor. In our experiments, we generated k^{rs} by sampling from independent and identical normal distributions with a fixed mean $\mu = 1$ and standard deviation σ (truncated at zero if a negative value is sampled). This choice of mean implies that the OD matrix does not contain any systematic error, but there are errors in individual entries, perhaps due to sampling errors in the demand forecasting process, or latent noise in future predictions. Uniform perturbation is a special case of OD-specific perturbation with $\mu \neq 1$ and $\sigma = 0$.

We tested the following values of σ: $\{0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5\}$. The number of samples tested for each σ is given in Table 5.1, where each sample involved solving TAP with the original network and a newly generated σ-perturbed OD matrix. Monte Carlo simulations allowed us to describe the relationship between the variance of the perturbation noise and the outputs, as well as for the confidence intervals for the output metrics described later in this section.

The third method is **spatially-correlated perturbation**, also referred to as spatial error. This method uses network geographical data to perturb demand within a certain region, simulating heterogeneous error in OD demand sampling. The intent is to simulate errors in estimating demand to or from zones in a specific region, as might happen if specific, spatially clustered demographic groups are undersampled, or if there are errors forecasting future land use in specific regions.

We experimented with this method in two ways. In the first, we perturbed demand from nodes closest to the center of the network set in 25% increments p using the following steps:

1. Identify the geographic center by taking the maximum and minimum latitudes $\text{lat}_{\min}$, $\text{lat}_{\max}$

and longitudes $\text{lon}_{\text{min}}, \text{lon}_{\text{max}}$ from the nodes in the network

2. Take the midpoint between the aforementioned latitudes $\overline{\text{lat}} = \left(\frac{\text{lat}_{\text{min}} - \text{lat}_{\text{max}}}{2} \right)$ and longitudes $\overline{\text{lon}} = \left(\frac{\text{lat}_{\text{min}} - \text{lat}_{\text{max}}}{2} \right)$ as the center of the network c.

3. Calculate the distance between each node i and c as $d_i = \sqrt{\left(\text{lon}_i - \overline{\text{lon}} \right) + \left(\text{lat}_i - \overline{\text{lat}} \right)}$

4. Sort nodes by d_i and select a given proportion p of nodes by lowest d_i as a set S_p. Round down in case the exact proportion of nodes in the network is not an integer.

5. Scale demand associated with the zones found in S_p by a given factor k in the following three ways to generate three separate geographically perturbed OD matrices:

   (a) Scale only demand that originates at selected zones

   (b) Scale only demand that terminates at selected zones

   (c) Scale all demand that either originates or terminates at selected zones

6. Find the UE solution associated with all three matrices and collect results.

7. Repeat for several proportions p and perturbation factors k to generate trends and to characterize changes as demand shrinks or grows.

The second set of spatially-correlated perturbation experiments was meant to study the effect of inconsistent demand measurement error across an OD matrix. We tested for the difference between output metrics if demand scaling was concentrated on origins near the center of the network, or if the same total demand was distributed globally across all origins in the OD matrix using a uniform perturbation factor. This followed the same process as described above with three changes:

1. In Step 5, instead of scaling demand in three ways, we only scale demand that originates at selected zones as in (a).

2. We take the resulting OD matrix and find the associated total demand. We divide this total by the total demand of the unperturbed matrix to find a scalar factor which we use to create a uniformly scaled OD matrix with the same total demand.

3. We find two UE solutions instead of three for a given p and k pair, one associated with the partially perturbed matrix and one with the uniformly scaled matrix.

As before, both scaling factor and proportion of nodes scaled were varied to more fully characterize the potential effects of local over- and under-sampling.

Some of our analyses involve comparisons across these three types of error (uniform, OD-specific, and spatially-correlated). To control for the overall relative error in the OD matrix when making these comparisons, consider the following error metric:

$$R = \mathbb{E}\left[\sum_{r,s \in Z} \frac{(\tilde{d}^{rs} - d^{rs})^2}{(d^{rs})^2} \right]$$

where $\tilde{d}^{rs}$ is the perturbed demand between node r and node s and Z is the set of all zones. For each of the three methods, $\tilde{d}^{rs}$ is equivalent to

1. kd^{rs} for uniform perturbation, where k is a single random scalar sample drawn from a $N(1, \sigma_U)$ distribution that is fixed for all rs.

2. $k^{rs}d^{rs}$ for OD-specific perturbation, where k^{rs} is a random scalar sample drawn from a $N(1, \sigma_{OD})$ distribution per OD pair, where the distribution is independent and identical for all rs.

3. $k_{rs}d^{rs}$ for spatially-correlated perturbation, where k_{rs} is a single random scalar sample drawn from a $N(1, \sigma_{SC})$ distribution if rs is an OD pair that has been designated as being scaled and 0 otherwise. k_{rs} is fixed for all scaled OD pairs.

To hold average OD matrix error equal across all three perturbation methods' OD matrices, fix R for the three methods and solve for the σ of each perturbation method's distribution. We found that if $\sigma_U = \sigma_{OD} = p\sigma_{SC}$, the overall relative error in the OD matrices is the same.

For a given input error parameter σ (equivalent to σ_U, σ_{OD}, and $p\sigma_{SC}$), we characterized the expected UE output error from the three methods by finding the average result of the normal distribution describing the input OD matrix error of each. To find this average for each of the three methods, we followed this procedure:

1. For uniform perturbations, we take a weighted average of the normalized output metrics associated with each discrete measurement of k found in our earlier experiment, with a weighting for each k based on the area it covers in the $N(1, \sigma)$ distribution. This is effectively an interpolation of our discrete perturbation scaling results to a smooth function that we take a weighted average for.

2. For OD-specific perturbations, we take the value of the normalized output metrics associated with the given σ as found in the prior experiments.

3. For spatially-correlated perturbations, we follow a method similar to uniform perturbations except for each k weight we use the area each k covers under $N(1, p^{-1}\sigma)$ in order to account for the more concentrated input error spatially-correlated perturbations produce. For these experiments, we used a p of 25%.

Output metrics serve to quickly characterize the UEs generated by each OD matrix. The metrics used are total system travel time (TSTT), flow-weighted volume-capacity ratio (WVC), and total vehicle miles traveled (VMT). They are defined as follows:

$$\text{TSTT} = \sum_{i \in N} t_i v_i$$
$$\text{WVC} = \sum_{i \in N} \frac{v_i}{\sum_{j \in N} v_j} \frac{v_i}{c_i}$$
$$\text{VMT} = \sum_{i \in N} l_i v_i$$

where, N is the set of all links in the network,

v_i is the volume on link i,

c_i is the capacity on link i,

t_i is the travel time on link i, and

l_i is the length of link i.

As a proxy for congestion, WVC better scales the effect of highly congested, low capacity links (and vice versa) than if volume over capacity was not weighted by flow.

All experiments were conducted on a machine running Ubuntu 18.04 with an Intel i5 processor @ 3.30 GHz and 8 GB of memory. All UE solutions were calculated to a relative gap of 10^{-6}.

5.4 Results

5.4.1 Uniform perturbation

For uniform perturbations, TSTT, WVC, and VMT scale with the demand factor. As seen in Figure 5.1, WVC and VMC scale close to linearly, but for WVC the slope may change depending on the network. For example, given a unit increase in demand perturbation factor,

there is an increase of 0.3 in WVC for the Berlin-Mitte-Center network compared to 1.5 for the SiouxFalls network, but for both networks VMT holds at a steady 1:1 relative increase per unit of demand scaling factor.
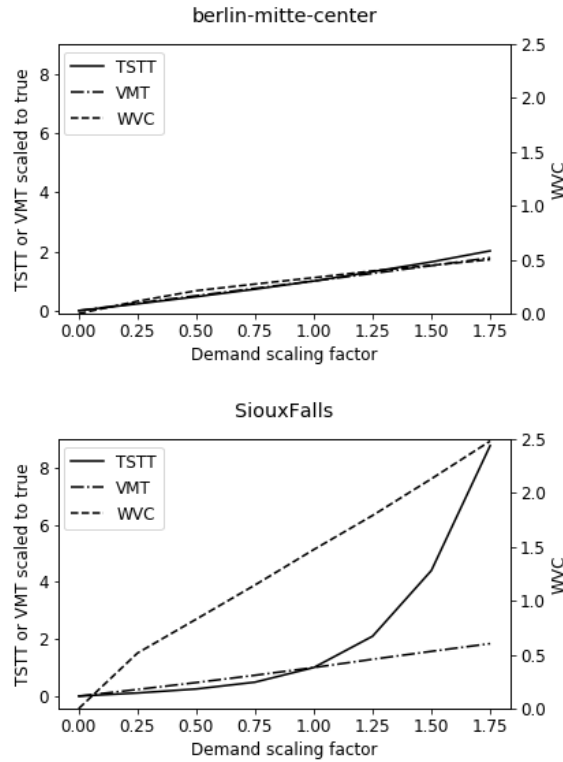
Figure 5.1: OD matrix scaling effect on TSTT and WVC

In contrast, TSTT shows a more complicated relationship with the demand factor. Comparing the two networks, we see that when a network's WVC is less than one (i.e., the network is uncongested), TSTT scales close to linearly with the perturbation factor, only increasing rapidly once WVC exceeds one and the network enters a congested regime. Overestimating demand will have a much larger impact on TSTT than underestimating demand, but it may not be as important if a traffic planner is more concerned about measuring congestion or miles traveled than total time spent traveling.

For a +25% error in input demand, the network with the greatest output error (SiouxFalls) shows a 200% larger TSTT, a 50% larger WVC, and a 30% larger VMT. That said, an uncongested network like Berlin-Mitte-Center only displays a 30% change across

the board, as does SiouxFalls when the input error is −25%, taking the network out of its congested state.

5.4.2 OD-specific perturbation

Because OD-specific perturbation involves noise, the resulting perturbed OD matrices are stochastic instead of deterministic like both uniform and spatially-correlated perturbation. Consequently we require Monte Carlo sampling to characterize the outcomes and for the results for this section to be displayed differently than for the two other perturbation methods. In addition, because fluctuations in the UE from OD-specific demand error tend to average out across all OD pairs, this subsection focuses mainly on describing the spread of output metrics for any given OD-specific perturbation scenario instead of drawing a clear trendline from lower to higher σ.

The degree to which output metrics vary under OD-specific perturbation depends on the average error in demand value. When observed demand levels deviate from their true values by an average of 2% or less with no bias, the resultant UE TSTT, WVC, and VMT all are within 1% of their true values. For larger demand errors, up to 10% error in OD matrix entries gives output metrics correct to within 5% of their true values for nearly all cases. For average error in demand up to 50%, TSTT, WVC, and VMT can still be expected to hold to within ±10% of their true values. As shown in Figure 5.2, these observations hold for every network we studied. For the same flow patterns, TSTT and WVC tend to exhibit larger changes than VMT due to link performance functions vis a vis link lengths. Table 5.2 summarizes the results of these trials.

5.4.3 Spatially-correlated perturbation

In the spatially-correlated perturbation tests, perturbing only origin or destination demand had almost identical effects on the output metrics, but perturbing both together—increasing total demand more than looking at origins or destinations individually—resulted in more extreme values as shown in Figure 5.3. Note the similarities between the trends in this figure with those of the Berlin-Mitte-Center uniform perturbation results in Figure 5.1. The trendlines are near identical in shape, with the only difference being that changes in response to the same demand scaling factor are dampened when the proportion of nodes affected is less than 100%. This dampening effect becomes more pronounced as the percentage of nodes perturbed is decreased (e.g., the change between 100% of nodes perturbed and 75% is
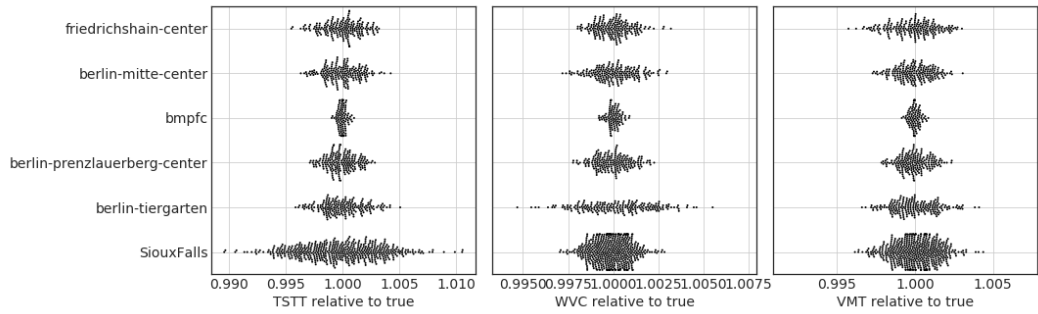
(a) $\sigma = 0.02$
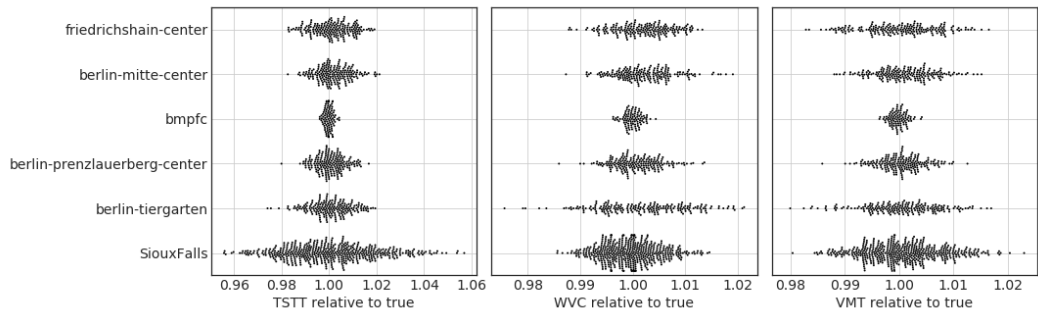
(b) $\sigma = 0.1$
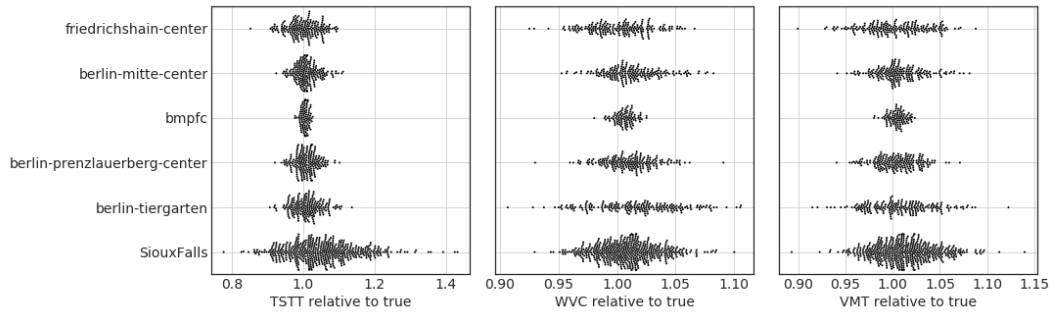
(c) $\sigma = 0.5$

Figure 5.2: Output metrics for all OD-specific perturbation trials

much less than between 50% and 25%). As expected, accurately determining UE solutions requires measuring demand at as many nodes as possible (equivalent to minimizing p in this experiment).
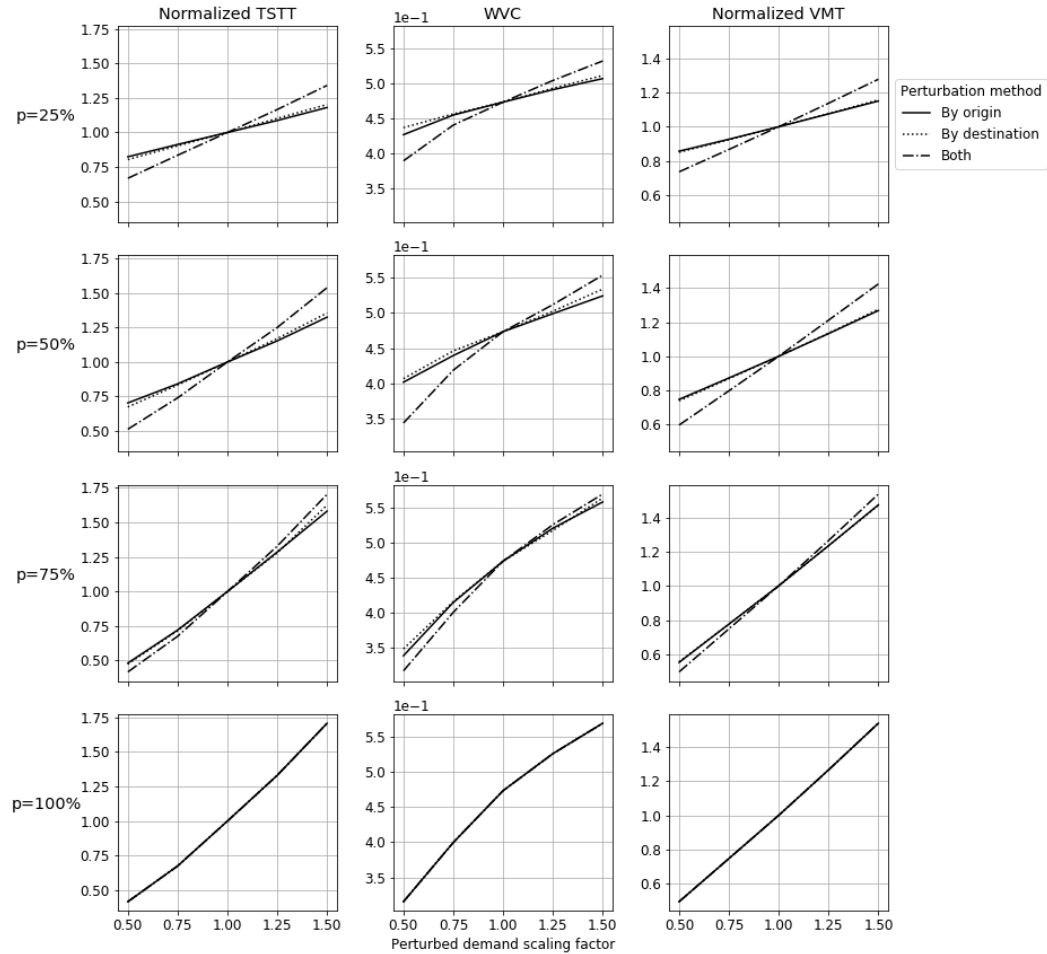
Figure 5.3: Effect of spatially-correlated perturbation methods on Berlin-Friedrichshain (rows denote the proportion of nodes perturbed p)

Under the conditions for this experiment, we see a 30% increase in TSTT and VMT when we perturb all demands by 25% (although WVC growth is slightly smaller than for Berlin-Mitte-Center), just as we do in the mean scaling scenario. There is almost no dampening effect when only 75% of nodes are perturbed, but the output metric error falls to 20%-25% and 10%-15% when only a half and a quarter of node demands are perturbed

respectively, depending on whether demands are perturbed by only one of or both origin and destination.

In the next experiment comparing the concentrated and global perturbation scenarios with fixed demand, the difference between the TSTTs of each scenario was within ±5% for demand scaling factors between 0.75 and 1.25. In fact, all networks studied were within this 5% error range, with half of them demonstrating even less variation. In line with our finding for uniform perturbation, TSTT diverges more at high scaling factors, but WVC sees its most extreme changes at low scaling factors as links become less congested. VMT behaves linearly in response to all scalings. Differences were minimized when a higher proportion of nodes were affected by the concentrated perturbation, as the demand situation approached that of the spatially invariant case. Full results from the Sioux Falls network can be seen in Figure 5.4. A generalized conclusion across networks implies that the proportion of nodes being perturbed has little impact on observed metrics, given that the magnitude of the perturbation is small (±25%). Barring major inaccuracies in demand projections, spatial effects cancel out due to the lack of correlation between errors across the OD matrix.
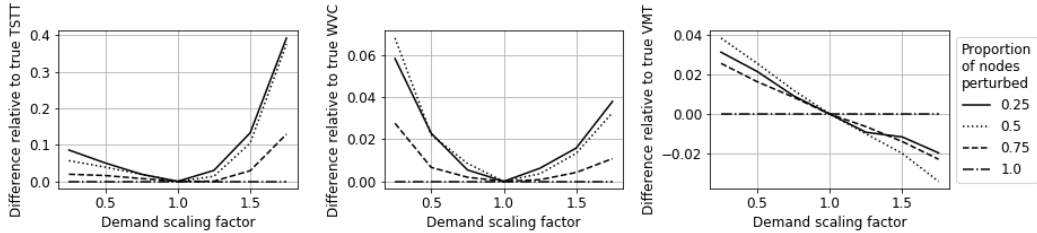
Figure 5.4: Effect of spatially-correlated vs. uniform demand perturbation on SiouxFalls

(Note that this last experiment has no analogue in the uniform and OD-specific perturbation experiments because here we are contrasting two different perturbed demand scenarios instead of perturbed scenarios with the unperturbed case.)

5.4.4 Parallel comparison

Figure 5.5 compares the results of our experiments across all three error types we identified given equal induced R for two representative networks, Berlin-Mitte-Prenzlauerberg-Friedrichshain-Center (bmpfc) and Sioux Falls. The metric values are reported relative to the unperturbed base case.

For $\sigma < 0.1$, the metrics do not change more than 1% even for the most congested

81

network, Sioux Falls. The largest σ we tested, 0.5, did not cause values in any network to deviate more than 5%, but more than doubled the TSTT in the Sioux Falls network under uniform perturbation. In general, uniform perturbations show the highest TSTT, VMT, and WVC deviations. In contrast, OD-specific perturbations result in relatively small errors, due to cancellation effects between OD pairs with overpredicted demand and those with underpredicted demand. The error due to spatial correlation fell between these two.
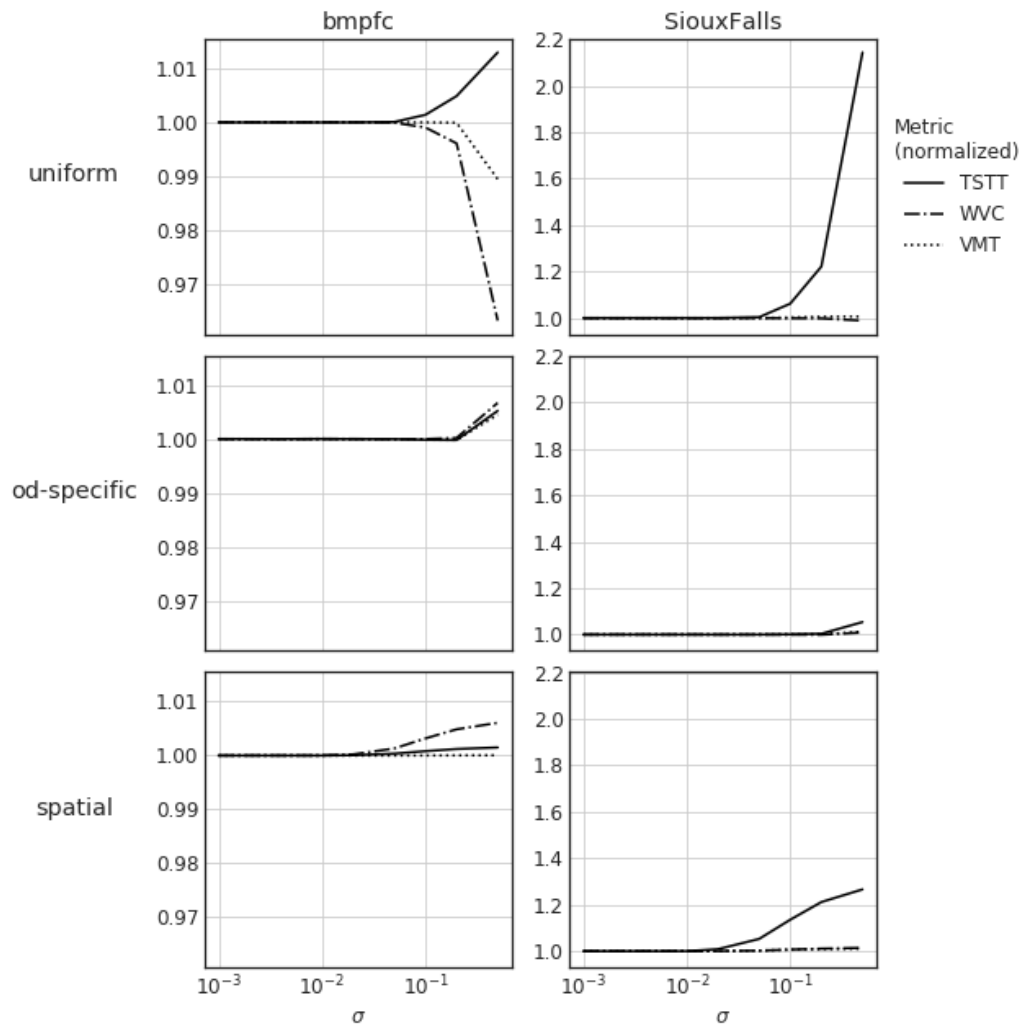
Figure 5.5: Parallel comparison between bmpfc and SiouxFalls

5.5   Case study: Austin

This section describes a case study on a network representing Austin, TX, based on network conditions from approximately 15 years ago. This is a much larger network than the ones tested in the previous section (7,466 nodes, 1,117 zones, 18,710 links, and 695,013 trips). Our experiments are based on observed population growth rates in this region since this network was created and the OD matrix was estimated. For the spatially-correlated perturbations, our experiments are informed by these observations: downtown Austin has grown significantly, as have certain neighborhoods in northern and southern suburbs due to new tech employment centers there. East Austin has gentrified, with higher-occupancy housing, while west Austin has changed less due partially to zoning laws.

This case study explores the differences between three variations on growth rate, specifically their effect on network performance as well as flows. We model these three cases as follows:

1. The demand growth rate is 2.08% annually, uniform across the network (this was the annual population growth rate between 2000 and 2010 (Bureau, 2010)).

2. The demand growth rate is 3% annually, uniform across the network (this was the observed population growth rate between 2010 and 2019).

3. The annual demand growth rates for north and south Austin are 3.25%, for east Austin 3%, and for the rest of the network, 2.75%. We call this the *heterogeneous growth* scenario below.

In particular, the first scenario would represent a forecast based on population growth rates at the time the network was created. In reality, the population grew faster than predicted, with the second scenario being a truer representation of what actually occurred. The third scenario is a more refined version of the second that accounts for spatial heterogeneity in population growth.

These factors are applied to trips in the OD matrix based on the origin node to account for home-to-work trips and extrapolated over 10 years. The resulting OD matrices correspond to demand forecasts for each of the three cases, which we used to find UE solutions along with link flows.

The topline results are shown in Table 5.3. The first thing to notice is the difference between the naive 2.08% forecast and the other two in terms of TSTT: the deviation is

over 9% even though the difference in annual growth rate was less than 1%. The second key observation is the lack of substantial differences in all metrics between the spatially distributed forecast and the uniform growth forecast. Figure 5.6 focuses on the volume-capacity ratio among all links in the Austin network, where we observe that the 3% growth rate and the regionally varied rate exhibit very similar link-by-link changes despite the regional variation in growth rate in the latter scenario. This is corroborated by the results seen in the fixed demand, globally scaled versus spatially concentrated demand experiments from the prior section.
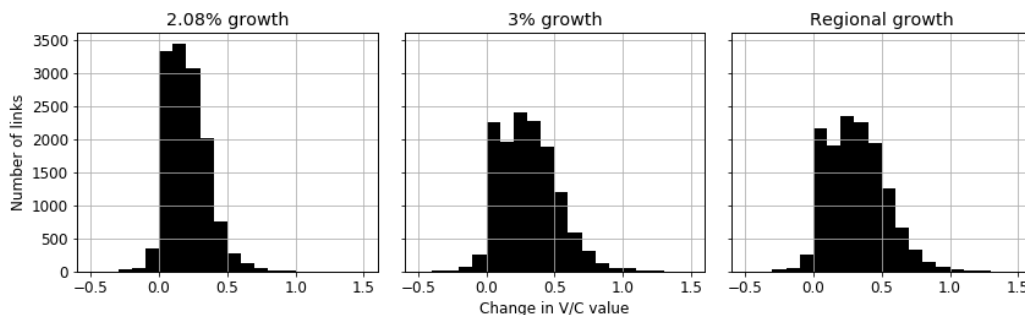
Figure 5.6: Change in V/C between growth rate assumptions relative to the no-growth scenario

5.6 Conclusions and future work

Inaccuracy in demand forecasting is difficult to avoid. To understand the impact of OD matrix input uncertainty on the traffic assignment problem, we studied the behavior of three key metrics—total system travel time, flow-weighted volume-capacity ratio, and total vehicle miles traveled—across three perturbation methods inspired by commonly observed errors in demand modeling. Uniform perturbation implies a global over- or underestimation, OD-specific perturbation implies individual OD-pair errors, and spatially-correlated perturbation covers systematic errors in a specific geographical region.

Under uniform perturbation, we see that WVC and VMT scale linearly with perturbation factors, but that TSTT begins exhibiting faster-than-linear growth once demand enters a congested regime. Consequently, the degree to which any metric changes under uniform perturbation (as well as spatially-correlated perturbation, to a lesser extent) is dependent

on the congestion level of the unperturbed network. We observed a 30% change in output metrics for a uniform 25% increase in demand for an uncongested network, and even larger changes for a congested network. Under OD-specific perturbation, to ensure output metric accuracy to within 10% of their true values, OD matrix demands need only be correct to within an average of ±50% of their true values provided there is no systematic over- or under-estimation. For accuracy within 5%, we can tolerate up to an average of ±10% error, and for accuracy within 1%, that tolerance drops to ±5%. Under spatially-correlated perturbation, we see a dampened version of the patterns observed in the uniform perturbation case as only a fraction of OD pairs experience the perturbation factor. Recasting this as a comparison between demand growth (or decay) being uniform or concentrated, we observe that for small changes in total demand ($\leq 25\%$), the output metrics do not change significantly (i.e., $\geq 5\%$) regardless of the proportion of nodes that experience the spatially-correlated perturbation.

Comparing the three perturbation methods for a fixed, unbiased OD matrix error, we see that uniform perturbation results in the largest average deviation in all output metrics, while both spatial and OD-specific perturbations exhibit errors in the same order of magnitude. (Note that unbiased demand estimates can still result in biased metrics, since they are related in a nonlinear way.) Nevertheless, if the forecast is unbiased, our experiments suggest that the standard error in demand would need to exceed 10% before the bias in the output metrics is greater than 1%.

Although our experiments have found strong patterns in the behavior of TAP in response to different assumptions of demand error and network congestion levels, we have not yet fully characterized the implications of these input errors. In a future study, we hope to examine the distribution of changes in link-by-link metrics in addition to system-wide metrics, as well as the interactions between uniform, OD-specific, and spatially correlated error to more accurately represent real-world error in demand sampling. We also plan to extend this analysis to multi-scenario analyses in which scenarios must be ranked and a preferred alternative selected.

Table 5.2: Summary of OD-specific Perturbation Trials

perturbation scenario		tstt		wvc		vmt	
net	σ	mean	std	mean	std	mean	std
friedrichshain-center	0.000	7.286e+05	na	0.473	na	1.731e+07	na
	0.001	7.286e+05	5.484e+01	0.473	0.000	1.731e+07	1.185e+03
	0.005	7.286e+05	2.927e+02	0.473	0.000	1.731e+07	6.190e+03
	0.010	7.286e+05	5.526e+02	0.473	0.000	1.731e+07	1.137e+04
	0.020	7.286e+05	1.186e+03	0.473	0.000	1.730e+07	2.512e+04
	0.050	7.287e+05	3.058e+03	0.474	0.001	1.730e+07	6.222e+04
	0.100	7.296e+05	5.762e+03	0.474	0.002	1.731e+07	1.191e+05
	0.200	7.296e+05	1.177e+04	0.474	0.004	1.731e+07	2.267e+05
	0.500	7.346e+05	2.942e+04	0.476	0.010	1.742e+07	5.527e+05
berlin-mitte-center	0.000	1.051e+06	na	0.333	na	2.178e+07	na
	0.001	1.051e+06	6.872e+01	0.333	0.000	2.178e+07	1.080e+03
	0.005	1.051e+06	3.501e+02	0.333	0.000	2.178e+07	5.347e+03
	0.010	1.051e+06	6.721e+02	0.334	0.000	2.178e+07	1.081e+04
	0.020	1.051e+06	1.328e+03	0.334	0.000	2.178e+07	2.048e+04
	0.050	1.051e+06	3.738e+03	0.334	0.001	2.178e+07	5.777e+04
	0.100	1.052e+06	6.542e+03	0.334	0.002	2.179e+07	1.049e+05
	0.200	1.053e+06	1.424e+04	0.335	0.004	2.179e+07	2.256e+05
	0.500	1.064e+06	3.644e+04	0.339	0.009	2.194e+07	5.787e+05
bmpfc	0.000	2.362e+06	na	0.286	na	5.714e+07	na
	0.001	2.362e+06	4.606e+01	0.286	0.000	5.714e+07	9.858e+02
	0.005	2.362e+06	2.093e+02	0.286	0.000	5.714e+07	4.534e+03
	0.010	2.362e+06	4.250e+02	0.286	0.000	5.714e+07	9.239e+03
	0.020	2.363e+06	8.660e+02	0.286	0.000	5.715e+07	1.864e+04
	0.050	2.363e+06	2.170e+03	0.286	0.000	5.715e+07	4.685e+04
	0.100	2.363e+06	4.515e+03	0.286	0.001	5.715e+07	9.777e+04
	0.200	2.363e+06	7.495e+03	0.286	0.001	5.715e+07	1.596e+05
	0.500	2.376e+06	2.072e+04	0.288	0.002	5.743e+07	4.504e+05
berlin-prenzlauerberg-center	0.000	1.400e+06	na	0.431	na	2.897e+07	na
	0.001	1.400e+06	8.810e+01	0.431	0.000	2.897e+07	1.324e+03
	0.005	1.400e+06	4.053e+02	0.431	0.000	2.897e+07	6.240e+03
	0.010	1.400e+06	8.464e+02	0.431	0.000	2.897e+07	1.261e+04
	0.020	1.400e+06	1.536e+03	0.431	0.000	2.897e+07	2.355e+04
	0.050	1.400e+06	3.794e+03	0.431	0.001	2.897e+07	5.877e+04
	0.100	1.401e+06	8.046e+03	0.431	0.002	2.898e+07	1.227e+05
	0.200	1.402e+06	1.810e+04	0.432	0.004	2.901e+07	2.591e+05
	0.500	1.416e+06	4.383e+04	0.436	0.010	2.914e+07	6.179e+05
berlin-tiergarten	0.000	7.168e+05	na	0.313	na	1.681e+07	na
	0.001	7.168e+05	6.717e+01	0.313	0.000	1.681e+07	1.179e+03
	0.005	7.169e+05	3.498e+02	0.313	0.000	1.681e+07	6.302e+03
	0.010	7.169e+05	6.256e+02	0.313	0.000	1.681e+07	1.115e+04
	0.020	7.168e+05	1.173e+03	0.313	0.001	1.681e+07	2.074e+04
	0.050	7.174e+05	3.095e+03	0.313	0.002	1.682e+07	5.537e+04
	0.100	7.171e+05	6.431e+03	0.313	0.003	1.681e+07	1.197e+05
	0.200	7.173e+05	1.306e+04	0.314	0.006	1.682e+07	2.459e+05
	0.500	7.212e+05	2.973e+04	0.316	0.012	1.684e+07	5.922e+05
SiouxFalls	0.000	7.480e+06	na	1.474	na	3.419e+06	na
	0.001	7.480e+06	1.298e+03	1.474	0.000	3.419e+06	2.354e+02
	0.005	7.480e+06	6.501e+03	1.474	0.000	3.419e+06	1.188e+03
	0.010	7.480e+06	1.236e+04	1.474	0.001	3.419e+06	2.247e+03
	0.020	7.479e+06	2.825e+04	1.474	0.002	3.420e+06	4.919e+03
	0.050	7.480e+06	6.561e+04	1.474	0.004	3.420e+06	1.141e+04
	0.100	7.483e+06	1.375e+05	1.473	0.008	3.422e+06	2.375e+04
	0.200	7.523e+06	2.579e+05	1.473	0.015	3.427e+06	4.377e+04
	0.500	7.838e+06	7.000e+05	1.486	0.038	3.459e+06	1.074e+05

Table 5.3: Results of Austin Forecasts after 10 Years

	Demand	TSTT	WVC	VMT
No growth	695,013	19,078,380	1.188	3.314e10
2.08% annual growth	853,885	38,900,203	1.422	4.120e10
3% annual growth	934,039	53,882,962	1.542	4.535e10
Heterogeneous growth	939,298	53,272,535	1.549	4.570e10

Chapter 6

Convergence behavior for traffic assignment characterization metrics

6.1 Introduction

Traffic assignment is a common tool in transportation planning, and predicts how travelers will choose routes accounting for congestion effects. Traffic assignment is used in long-term planning, as the final step of the traditional four-step model, to assist in decision-making based on link flows, select link analysis, or shortest-path analysis. It also appears as a sub-problem in network design, toll-setting, and other related bi-level optimization problems. Despite many advances in dynamic traffic modeling, static assignment remains common in current practice. And despite advances in technology and algorithmic efficiency, computation times are still a relevant issue as agencies move to more detailed, multiclass models, or when assignment is a subproblem in a larger iterative scheme (feedback models, trip table estimation, network design, and so forth). This chapter therefore focuses on the static traffic assignment problem (TAP) as it is traditionally formulated.

This chapter investigates the convergence behavior of other metrics — specifically total system travel time, vehicle-miles traveled, equilibrium path flows, and the set of used paths — on twelve standard networks. We thus aim to generalize the recommendations of Boyce et al. (2004) based on other networks and metrics.

The goal of this chapter is to provide guidance on the level of convergence needed, depending on the metric of interest. This allows computational resources to be used as effectively as possible, and not wasted on unnecessary precision beyond the requirements of

a particular application.

6.1.1 Contributions

The central question addressed in this chapter is the level of precision needed in the solution. An insufficiently-converged solution will not produce reliable estimates for planning. An overly-stringent convergence criterion, on the other hand, wastes computational resources that can be better spent on other model components or tasks. (For instance, examining additional alternatives in a Monte Carlo simulation or solutions in a bilevel program.) The appropriate convergence level depends on the application context: the specific network, the specific metrics of interest, and the decision being made. In particular, the appropriate convergence criterion when producing a single point prediction is likely different than that when selecting a preferred alternative among several. The latter problem introduces several complications relative to the former, so this chapter focuses primarily on the convergence level needed to stabilize a metric for a single modeling scenario.

The main contribution of this chapter is to identify the convergence behavior of five metrics on twelve different networks, thus generalizing the analysis of Rose et al. (1988) and Boyce et al. (2004). We examine the rates of convergence of these metrics compared to that of relative gap (the most common convergence metric), and identify trends based on network size and congestion level. We additionally describe the heterogeneity in convergence rates between different links within the same network. These analyses primarily have implications for choosing a convergence level for analysis of a particular scenario, and also lay the groundwork for future studies on appropriate convergence criteria for multi-scenario analyses. Our experiments also include tests of different traffic assignment algorithms, and examining a scenario with heterogeneous user classes. We also compare alternative gap functions in current use, suggesting how our results for one gap function can be translated to these alternatives.

The rest of this chapter is structured as follows: Section 6.2 provides mathematical specifications of the TAP metrics we study. We next describe our experimental structure, the networks we use, and the design of particular scenarios in Section 6.3. We next provide the results of these experiments, and our interpretation of these results in Section 6.4. We finally conclude with a summary of our findings and topics for future study.

6.2 Background

This study uses a combination of the TAPAS and Frank-Wolfe algorithms for its primary experiments, and also uses Algorithm B (a bush-based algorithm) to test transferability of the results. In our analysis, we do not test any path-based algorithms. To economize on memory, such algorithms employ "column dropping" rules to store as few paths as possible. However, such solutions have extremely low entropy (indeed, the most likely solution spreads flow over as many paths as possible), and thus the specific path flow solution is untrustworthy for further analysis; see the discussion and empirical results from Bar-Gera (2006) and Bar-Gera and Luzon (2007). Since we wish to examine convergence of the path flows in the solution, it is clearest to do so using an algorithm which converges to the (unique) entropy-maximizing path flow solution, rather than an arbitrary path flow equilibrium.

Given a feasible solution $(\mathbf{x}, \mathbf{h})$ to TAP, we select five metrics for analysis. (Since algorithms for TAP converge only in the limit, we do not demand optimal solutions to the above problems.) The *total system travel time* (TSTT) expresses the sum of each vehicle's travel time in the network:

$$TSTT(\mathbf{x}) = \sum_{(i,j) \in A} t_{ij} x_{ij} \tag{6.1}$$

Vehicle-miles traveled (VMT) expresses the total distance traveled by vehicles in the network:

$$VMT(\mathbf{x}) = \sum_{(i,j) \in A} l_{ij} x_{ij} \tag{6.2}$$

To measure convergence of these metrics, we calculate the relative difference between their values at the current solution $\mathbf{x}$ and the equilibrium solution $\mathbf{x}^*$:

$$\Delta TSTT(\mathbf{x}) = \frac{TSTT(\mathbf{x}) - TSTT(\mathbf{x}^*)}{TSTT(\mathbf{x}^*)} \tag{6.3}$$

$$\Delta VMT(\mathbf{x}) = \frac{VMT(\mathbf{x}) - VMT(\mathbf{x}^*)}{VMT(\mathbf{x}^*)} . \tag{6.4}$$

Both TSTT and VMT are aggregate metrics. To represent convergence of the specific link and path flows themselves, we measure the proportion of links (or paths) within a given relative threshold ϵ of their equilibrium values. Let $A_\epsilon^*(\mathbf{x})$ denote the set of links with flows

90

within this threshold:

$$A_\epsilon^*(\mathbf{x}) = \left\{ (i, j) \in A : \left| x_{ij} - x_{ij}^* \right| < \epsilon x_{ij}^* \right\} . \tag{6.5}$$

Let $\Pi_\epsilon^*(\mathbf{h})$ denote the set of paths whose flows are within this threshold:

$$\Pi_\epsilon^*(\mathbf{h}) = \left\{ \pi \in \Pi_+(\mathbf{h}^*) : \left| h_\pi - h_\pi^* \right| < \epsilon h_\pi^* \right\} , \tag{6.6}$$

where $\mathbf{h}^*$ is the (entropy-maximizing) solution to the most likely path flows problem at equilibrium. Note that $\Pi_\epsilon^*(\mathbf{h})$ is a subset of the used path set at equilibrium. Using these sets, we define the *proportion of unconverged links (PUL)* as

$$PUL(\mathbf{x}, \epsilon) = 1 - \frac{|A_\epsilon^*(\mathbf{x})|}{|A|} , \tag{6.7}$$

and the *proportion of unconverged paths (PUP)* as

$$PUP(\mathbf{h}, \epsilon) = 1 - \frac{|\Pi_\epsilon^*(\mathbf{h})|}{|\Pi_+(\mathbf{h}^*)|} . \tag{6.8}$$

Finally, we define the *path set deviation (PSD)* to represent how the set of used paths converges to set of equilibrium paths by defining

$$PSD(\mathbf{h}) = 1 - \frac{|\Pi_+(\mathbf{h}) \cap \Pi_+(\mathbf{h}^*)|}{|\Pi_+(\mathbf{h}^*)|} . \tag{6.9}$$

We thus have $PSD = 1$ if the set of currently used paths and the set of equilibrium paths is disjoint, and $PSD = 0$ if every equilibrium path is in the current set of used paths. As with the other metrics, it should decrease to zero over successive iterations.

Both *PUP* and *PSD* are calculated with respect to the used paths at equilibrium. Some restriction of the path set is necessary, since the number of paths grows exponentially with network size and the vast majority of these are unused. Such paths should not be considered in our metrics, and we decided to measure *PUP* and *PSD* relative to the equilibrium path sets to be consistent with the other metrics (which are measured relative to the equilibrium link flows and most-likely path flows). Defining *PUP* only based on the set of used paths at equilibrium is important because *PUP* is a relative error measure; any path for which $h_\pi^* = 0$ would thus appear "unconverged" even with an infinitesimal flow value. This is not a serious deficiency, because any solution placing positive flow on a non-equilibrium path must also place the "wrong" value on at least one path in $\Pi_+(\mathbf{h}^*)$ (by

flow conservation), which would be detected by PUP with an appropriate ϵ value.

These five metrics — $\Delta TSTT$, ΔVMT, PUL, PUP, and PSD — are directly related to practical applications of traffic assignment, and converge to zero as $\mathbf{x}$ and $\mathbf{h}$ approach $\mathbf{x}^*$ and $\mathbf{h}^*$, respectively. However, they are not suitable convergence criteria because they can only be evaluated if the equilibrium link flows or most-likely path flows are already known, and there would be no need to solve TAP if this were true. Therefore, in practice, convergence is measured using information available even at intermediate solutions.

The relative gap is one such measure. There are several definitions of relative gap in common practice; we use the following one for our experiments, and later discuss relationships with its alternatives. Let $\kappa_{rs}(\mathbf{x})$ denote the travel time on the shortest path from origin r to destination s using the link travel times corresponding to $\mathbf{x}$. The shortest path travel time (SPTT) can then give the total travel time we would expect if all vehicles were on shortest paths (as the UE condition requires):

$$SPTT(\mathbf{x}) = \sum_{(r,s) \in Z^2} \kappa_{rs}(\mathbf{x}) d_{rs} \qquad (6.10)$$

The gap and relative gap of a feasible solution, as defined in Rose et al. (1988), are:

$$gap(\mathbf{x}) = SPTT(\mathbf{x}) - TSTT(\mathbf{x}) \qquad (6.11)$$

and

$$RG(\mathbf{x}) = -\frac{gap(\mathbf{x})}{SPTT(\mathbf{x})} = \frac{TSTT(\mathbf{x})}{SPTT(\mathbf{x})} - 1 . \qquad (6.12)$$

Relative gap is non-negative, and equal to zero only at equilibrium solutions, and thus is a valid gap function. Other gap metrics used for convergence include alternative definitions of relative gap, average excess cost (AEC) and average total reduced cost. We next define a variant of relative gap, the one used by Boyce et al. (2004), and AEC, and discuss their relationship with the RG definition of equation (6.12). This section contains a brief mathematical discussion, and our results include an empirical comparison.

An alternate definition of relative gap (RG') normalizes the gap by a lower bound on the optimal value of the Beckmann function in equation (1.1). The lower bound calculated from a particular solution is given by

$$LB(\mathbf{x}) = \sum_{(i,j) \in A} \int_0^{x_{ij}(k)} t_{ij}(x) \, dx + gap(\mathbf{x}) , \qquad (6.13)$$

and the best lower bound is the tightest bound over the flow vectors $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_k}$ seen over successive iterations thus far:

$$BLB = \max_k \{LB(\mathbf{x_k})\}. \tag{6.14}$$

The relative gap is then given by

$$RG'(\mathbf{x}) = -\frac{gap(\mathbf{x})}{|BLB|}. \tag{6.15}$$

The average excess cost is defined as

$$AEC(\mathbf{x}) = -\frac{gap(\mathbf{x})}{\sum_{(r,s)\in Z^2} d_{rs}}. \tag{6.16}$$

Observe that the numerator of all three gap functions is the same, and they differ only in how they are normalized. The ratio of AEC and RG equals the ratio of $SPTT$ and $\sum_{(r,s)\in Z^2} d_{rs}$, which is the average travel time on the all-or-nothing assignment. As $SPTT$ stabilizes close to convergence, the ratio between AEC and RG will approach a constant representing average travel time.

To compare RG and RG', we compare LB and $SPTT$, and see that their ratio is

$$\frac{LB(\mathbf{x})}{SPTT(\mathbf{x})} = \frac{\sum_{(i,j)\in A} \int_0^{x_{ij}} t_{ij}(x)\, dx}{SPTT(\mathbf{x})} + \frac{SPTT(\mathbf{x}) - TSTT(\mathbf{x})}{SPTT(\mathbf{x})} \tag{6.17}$$

The second term in this equation is $-RG(\mathbf{x})$, which approaches zero as equilibrium is reached. To analyze the first term, we note that $SPTT(\mathbf{x})$ and $TSTT(\mathbf{x})$ become asymptotically equal, and so

$$\frac{\sum_{(i,j)\in A} \int_0^{x_{ij}} t_{ij}(x)\, dx}{SPTT(\mathbf{x})} \approx \frac{\sum_{(i,j)\in A} \int_0^{x_{ij}} t_{ij}(x)\, dx}{\sum_{(i,j)\in A} x_{ij} t_{ij}(x_{ij})}. \tag{6.18}$$

On the right-hand side of equation (6.18), both the numerator and denominator include a sum over links. For any specific link, the difference between its term in the numerator and its term in the denominator is illustrated in Figure 6.1, which shows a typical link performance function. The hatched area is the term in the numerator, whereas the area of the rectangle is its term in the denominator. We see visually that these two areas are approximately equal. To compare them numerically, consider a link for which $t_{ij} = t_{ij}^0 (1 + 0.15(x_{ij}/u_{ij})^4)$, where

93
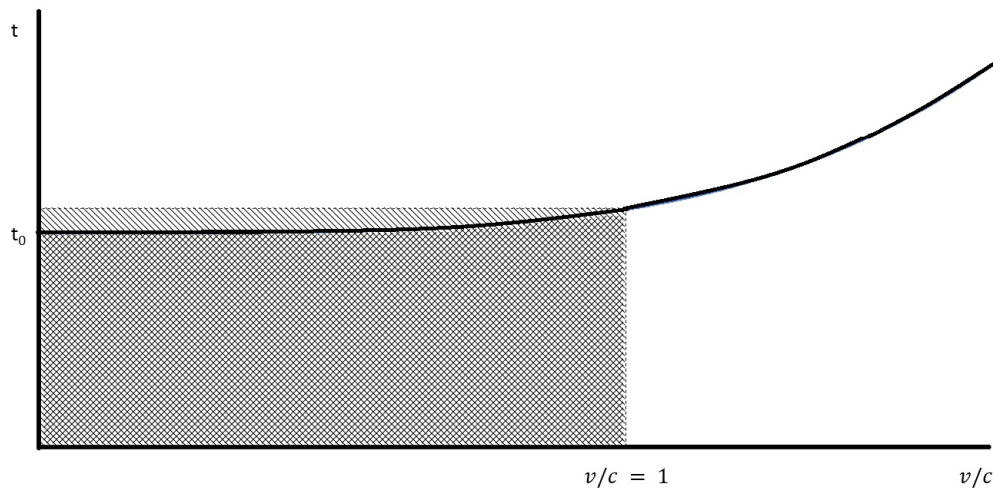
Figure 6.1: Visualization of RG and RG' ratio term

t_{ij}^0 and u_{ij} are its free-flow time and "practical capacity," respectively.[1] When the link is used relatively heavily ($x_{ij} = u_{ij}$), the ratio between the areas is 0.89. Even when the link is highly congested ($x_{ij} = 1.5u_{ij}$), the ratio between the areas is 0.65, and the areas are of the same order of magnitude. We thus expect the ratio $LB(\mathbf{x})/SPTT(\mathbf{x})$ to be fairly close to one, and thus RG and RG' to have a similar order of magnitude.

To summarize, we expect our results for RG (which will be given in terms of order of magnitude) to translate more or less directly to RG', although the specific numerical value may differ by up to 30%. To translate them to AEC, one must multiply by the average travel time in the network (whose order of magnitude can be estimated *a priori*). If travel times are reported in minutes and the network represents a typical metropolitan region, we would expect the AEC for a given solution to be one to two orders of magnitude higher than the RG, and our results can be adapted accordingly. Our experiments below validate this analysis numerically.

6.3 Data and Experiment Design

The main objective of this chapter is to determine the relationship between $\Delta TSTT$, ΔVMT, PUL, PUP, and PSD (which carry more practical meaning) and the corresponding RG level

---

[1]This is the commonly-used Bureau of Public Roads function with standard values for its shape parameters Bureau of Public Roads (1964).

(which can be calculated without knowing the equilibrium solution). We choose to index these results to relative gap, rather than iteration count or another measure of progress, because gap functions serve as an absolute measure of convergence that can be applied regardless of algorithm or parameter settings.

This section explains the procedures we used to determine the relationship between the five convergence metrics and relative gap. We first discuss the networks and algorithms used, and choices of specific parameters. We next discuss how we obtained solutions of a particular relative gap level for analysis.

The networks studied in this chapter are shown in Table 6.1 below, all obtained from the Transportation Networks for Research repository (Stabler, 2019). More details about these networks are provided in section 3.5.

Table 6.1: Description of networks used

Network name	Zones	Links	Nodes	Trips	Average flow-to-capacity ratio
SiouxFalls	24	76	24	360,600	1.612
Eastern-Massachusetts	74	258	74	65,576	0.163
Anaheim	38	914	416	104,694	0.297
Chicago-sketch	387	2950	933	1,260,907	0.257
Berlin-Prenzlauerberg-Center	98	2184	975	23,648	0.121
Barcelona	110	2522	1020	184,679	1.137
Winnipeg	147	2836	1052	64,784	2.028
Terrassa	55	3264	1609	25,225,700	5.964
Austin	7388	18961	7388	739,351	0.875
Berlin-Center	865	28376	12981	168,222	0.092
Chicago-Regional	1790	39018	12982	1,360,427	0.522
Philadelphia	1525	40003	13389	18,503,872	0.949

Calculating $\Delta TSTT$, ΔVMT, PUL, PUP, and PSD requires the equilibrium link flow and most-likely path flow solutions. Near-equilibrium link flows $\mathbf{x}^*$ and proportional path flows $\mathbf{h}^*$ were obtained using the TAPAS implementation by Perederieieva et al. (2015) (with default settings for TAPAS parameters used to determine cost-effective PAS and flow-effective PAS), setting a relative gap of 10^{-12} as the termination criterion. Our experiments

will cover solutions over a range of gap levels between 10^{-3} and 10^{-8}. Over this range, we expect the distinction between using our reference solution with a gap of 10^{-12}, and an exact equilibrium, to be small. For calculating the proportions of unconverged links and unconverged paths, we chose a threshold of $\epsilon = 0.01$, to be consistent with Boyce et al. (2004). (The results contain a sensitivity analysis with respect to this parameter.) All experiments were conducted on a machine with Ubuntu 18.04, 8 GB of memory and Intel i5 processor @ 3.30 GHz.

Our analysis involves solutions at six target gap levels: 10^{-3}, 10^{-4}, 10^{-5}, ..., 10^{-8}. To facilitate comparison between different networks, we obtained solutions on each network whose relative gap was within 10% of these levels (e.g., between 0.0009 and 0.0011 for 10^{-3}). Obtaining solutions on each network with such specific gap levels is not trivial, since algorithms for TAP are designed to reach equilibrium as rapidly as possible, and not aim for a specific nonzero gap level. Therefore, we used the procedure described below to generate solutions at a specified target gap level γ. This procedure involves a hybrid of the TAPAS implementation described above, and an implementation of Frank-Wolfe (Boyles, 2019) with ten bisection iterations per flow shift.

1. Run TAPAS with a termination criterion of γ as relative gap. If the solution is in the acceptable range $[0.9\gamma, 1.1\gamma]$, return the link flows **x** and path flows **h**, along with the values of $\Delta TSTT$, ΔVMT, PUL, PUP, and PSD.

2. If the returned solution has a gap level less than 0.9γ, examine the solution from the previous iteration to see if it is in the acceptable range $[0.9\gamma, 1.1\gamma]$. If so, return the link and path flows, and the five metrics, for that solution.

3. If neither of the above solutions is in the acceptable range, initialize Frank-Wolfe with the TAPAS solution from the previous iteration. Perform iterations of Frank-Wolfe until the gap is in the acceptable range $[0.9\gamma, 1.1\gamma]$, and return that solution and the corresponding metrics. Frank-Wolfe is used due to its relatively slow solution improvement, which leads to flow values within the desired gap range without "skipping over."

This process is repeated for each network and target gap level. This procedure worked for all but seven scenario-RG value combinations, due to the Frank-Wolfe algorithm in the last step jumping over the acceptable gap range. In these remaining cases, we repeated the last step, restarting Frank-Wolfe algorithm with the prior flow pattern, but fewer bisection iterations.

Our experiments are divided into the following analyses:

1. Identify the convergence rates of $TSTT$, VMT, and path/link flows to their equilibrium values using the procedure described above. We use both the base demand levels, and adjusted demand levels to study how congestion levels affect the convergence of these metrics. (This set of experiments is the most extensive, and is used as the basis for our core recommendations.)

2. Repeating the analysis using Algorithm B (Dial, 2006) to investigate transferability of results to algorithms besides TAPAS.

3. Repeating the analysis in a setting with toll roads and two user classes with distinct values of time.

4. Investigating the effect of different convergence levels when evaluating candidate solutions as a subproblem in network design, a bilevel program.

5. Comparison of the alternative gap measures RG, RG', and AEC so results can be translated appropriately. The intent is to numerically validate the mathematical analysis in the previous section, which involved several approximations.

For the second set of experiments, we used our own implementation of Algorithm B; the source code is available at Boyles (2019). Aggregate metrics (TSTT, VMT) and PUL were calculated from these experiments. The path-based metrics PUP and PSD were not calculated, since Algorithm B does not aim to maximize entropy or provide a proportional path flow solution.

For the third set of experiments, we introduced two user classes distinguished by their value of time ($15/hr and $30/hr), and used the toll values given in the network instances, where present. For networks without toll roads, we randomly selected 10% of links to be tolled. These experiments used Algorithm B, as our implementation supports multi-class assignment and the available TAPAS implementation does not.

For the fourth set of experiments, we formulated a network design problem, in which a discrete set of links had to be chosen for "upgrade." An upgrade increased a link's capacity by 50%, and had a cost proportional to its length. The budget allowed upgrading up to 5% of the total length of all links, and the objective is to minimize $TSTT$ subject to equilibrium constraints. This is a classic bilevel problem which is intractable to solve exactly. We thus solved it heuristically, using the genetic algorithm implementation in the pyeasyga library

(Remi-Omosowon, 2020). Such an algorithm involves solving a number of TAP instances as subproblems to evaluate fitness of candidate solutions. We used the default values for parameters in this library. For each network, we varied the RG level used for evaluating the TAP fitness function. Each combination of objective function and RG level was solved five times, and average performance reported.

For the fifth set of experiments, we calculate RG, RG', and AEC for the solutions obtained in the previous analyses, and conduct a linear regression to investigate whether $RG \approx RG'$, and that RG and AEC differ by a nearly constant multiplicative factor, as was suggested by the approximate mathematical analysis in the previous section.

6.4 Results

This section reports the results from the experiments described above. Each set of results is presented in its own subsection. Experiment 1 forms the core of our analysis, and is described in the greatest detail. The remaining experiments are described more briefly, highlighting key differences from the core analysis results. The Appendix for this chapter contains detailed results, separated for each of the twelve networks under consideration. The figures and tables in this section present summaries of this data, presenting the most important findings from each experiment.

6.4.1 Experiment 1: Network metric behavior results

The values of the main convergence metrics ($\Delta TSTT$, ΔVMT, PUL, PUP, and PSD) are shown in Figures 6.2–6.4 (presented according to network and size), and in Figures 6.5–6.9 (presented according to each metric). In the latter set of figures, the thin lines represent the values of each metric in one of the twelve networks tested, and the thick line represents the average value. Both sets of figures use logarithmic axes both for the relative gap, and for each metric, to focus on the orders of magnitude in these values. Table 6.2 shows the numerical values of the metric means, as well as the highest and lowest values seen at a particular level across all networks. The raw data, containing the specific values for over all twelve networks, is found in Table B.1 in Appendix B.
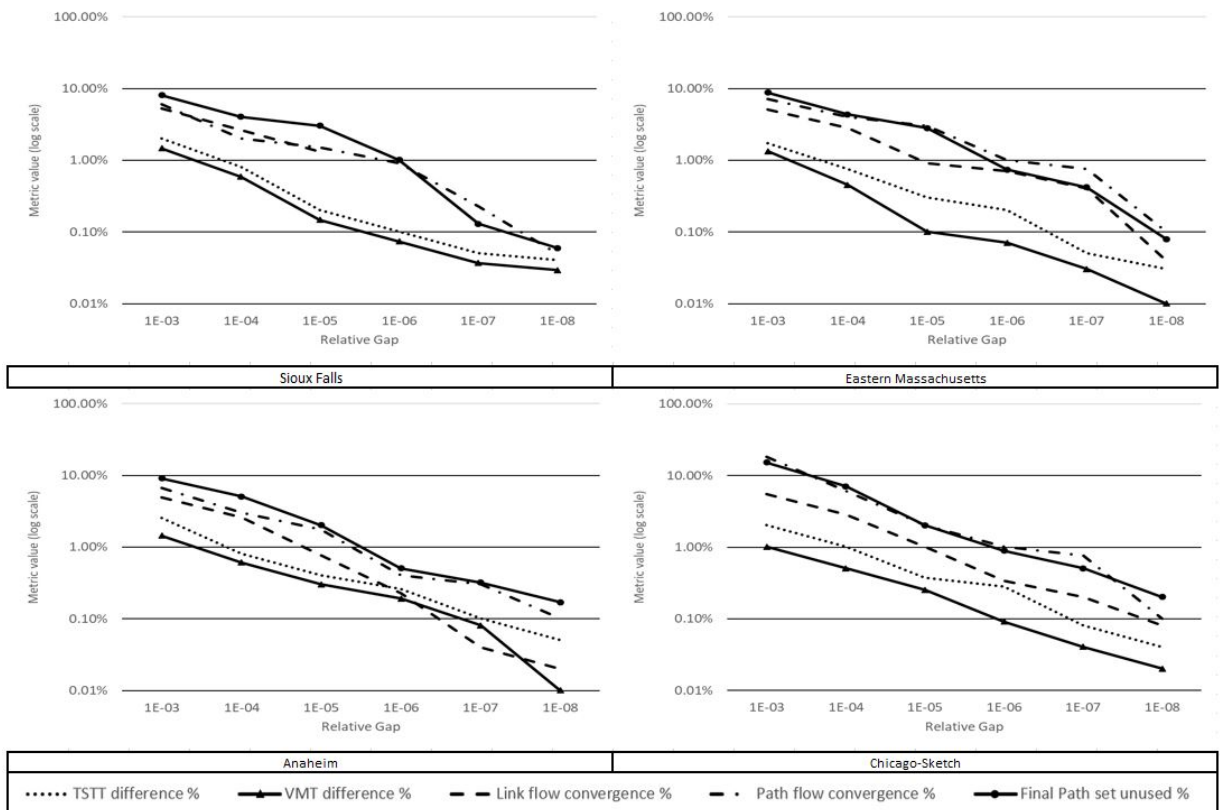
Figure 6.2: Stabilization behaviour of metrics at default demand, small networks
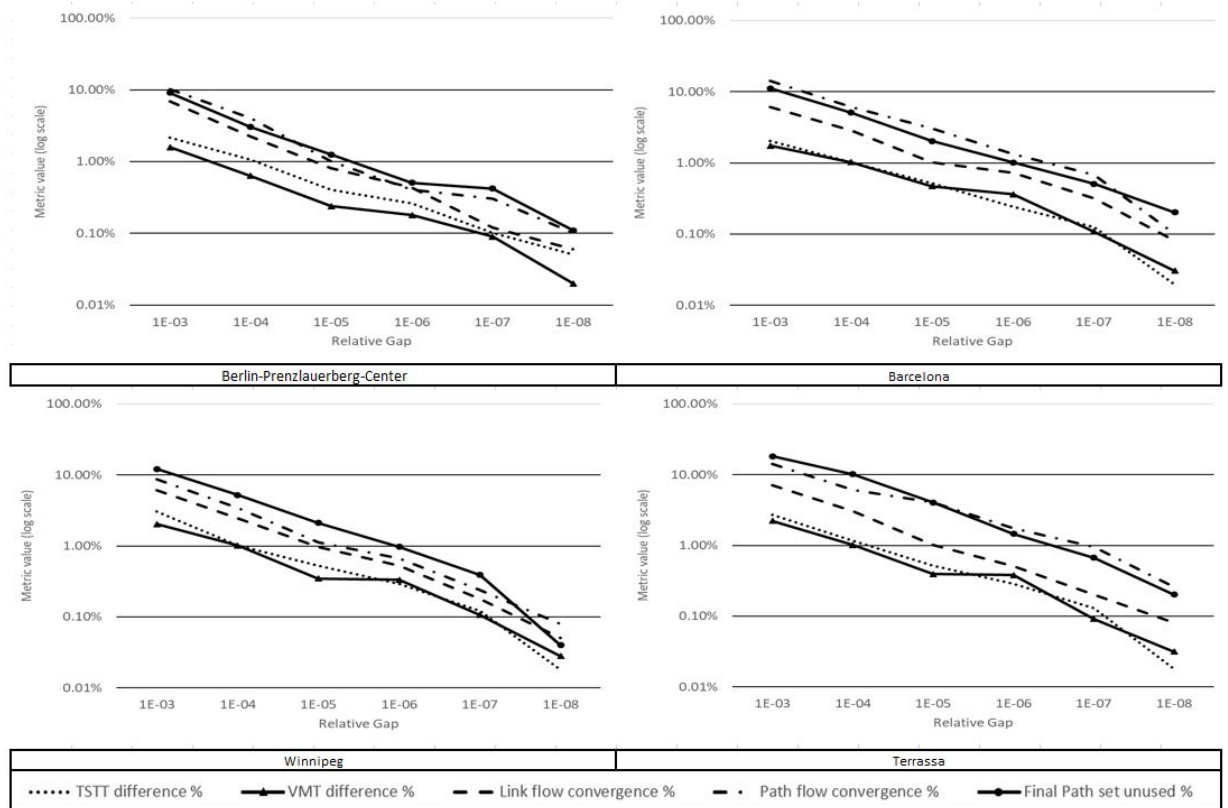
Figure 6.3: Stabilization behaviour of metrics at default demand, medium networks
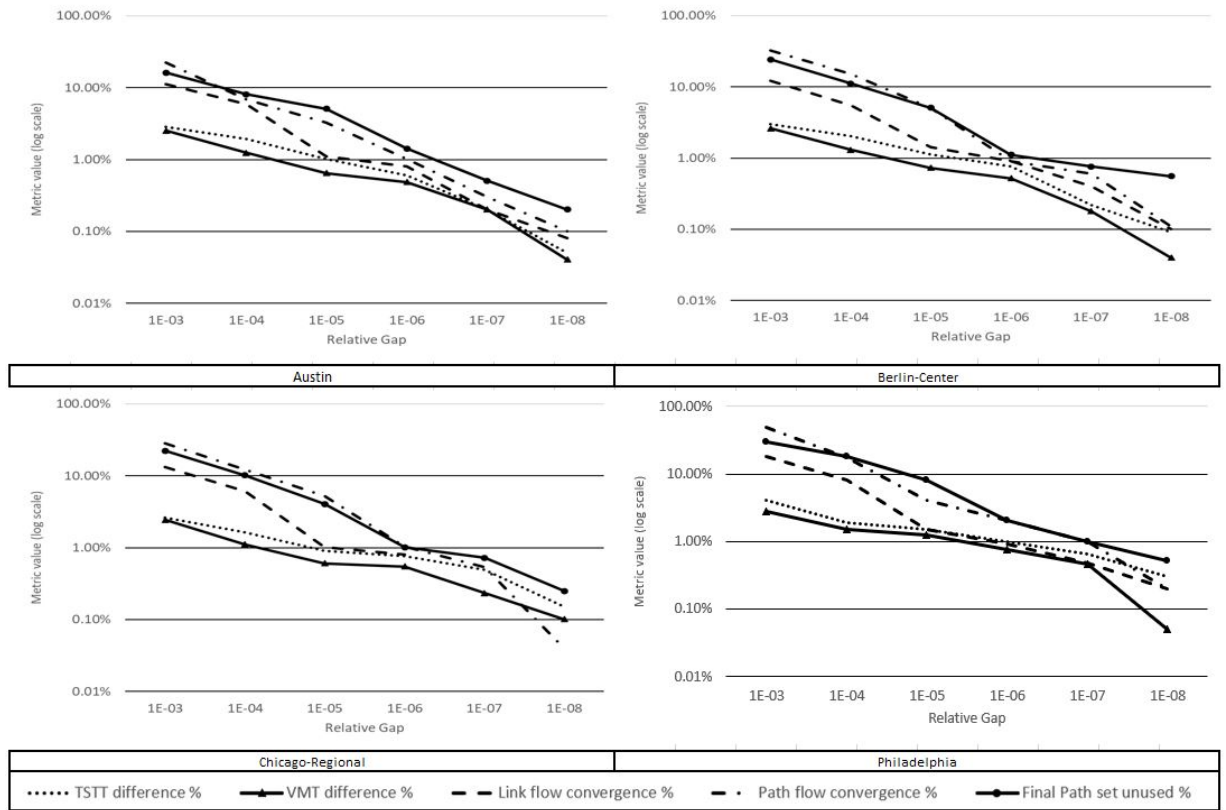
Figure 6.4: Stabilization behaviour of metrics at default demand, large networks
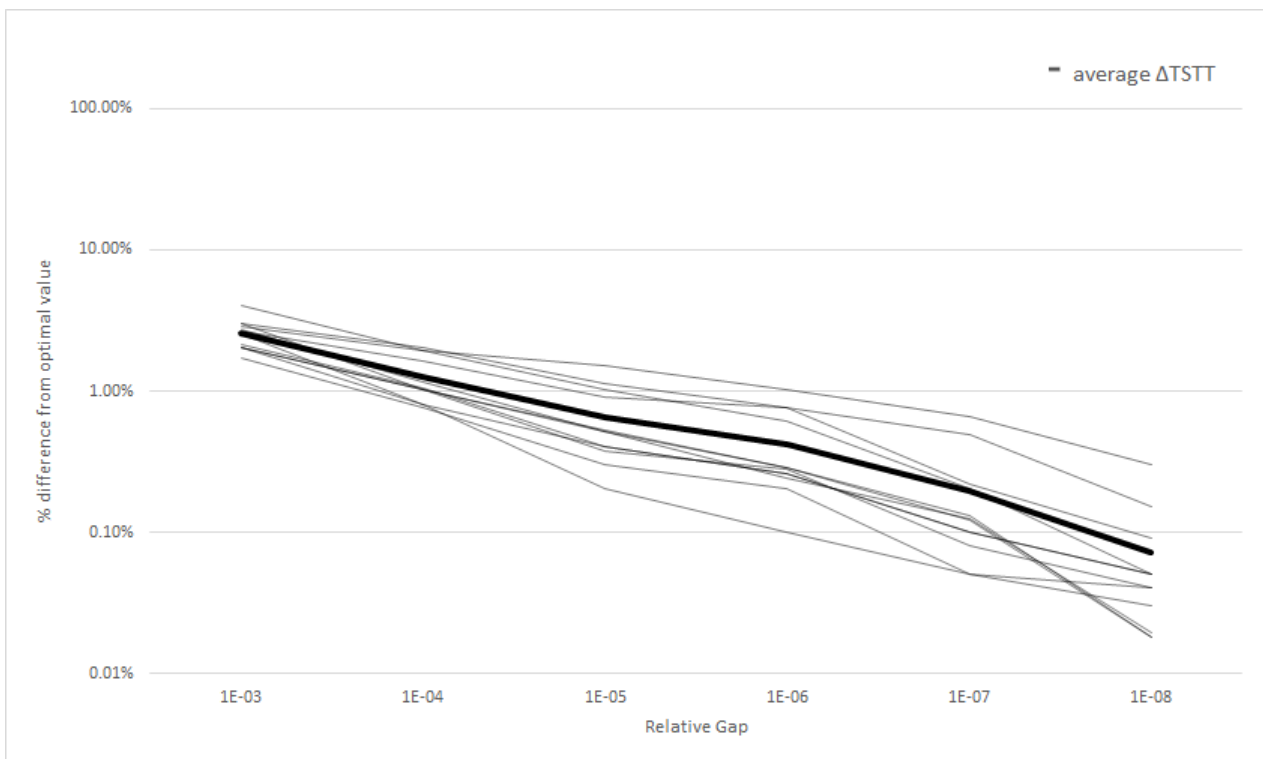
Figure 6.5: ΔTSTT trends for different gap levels
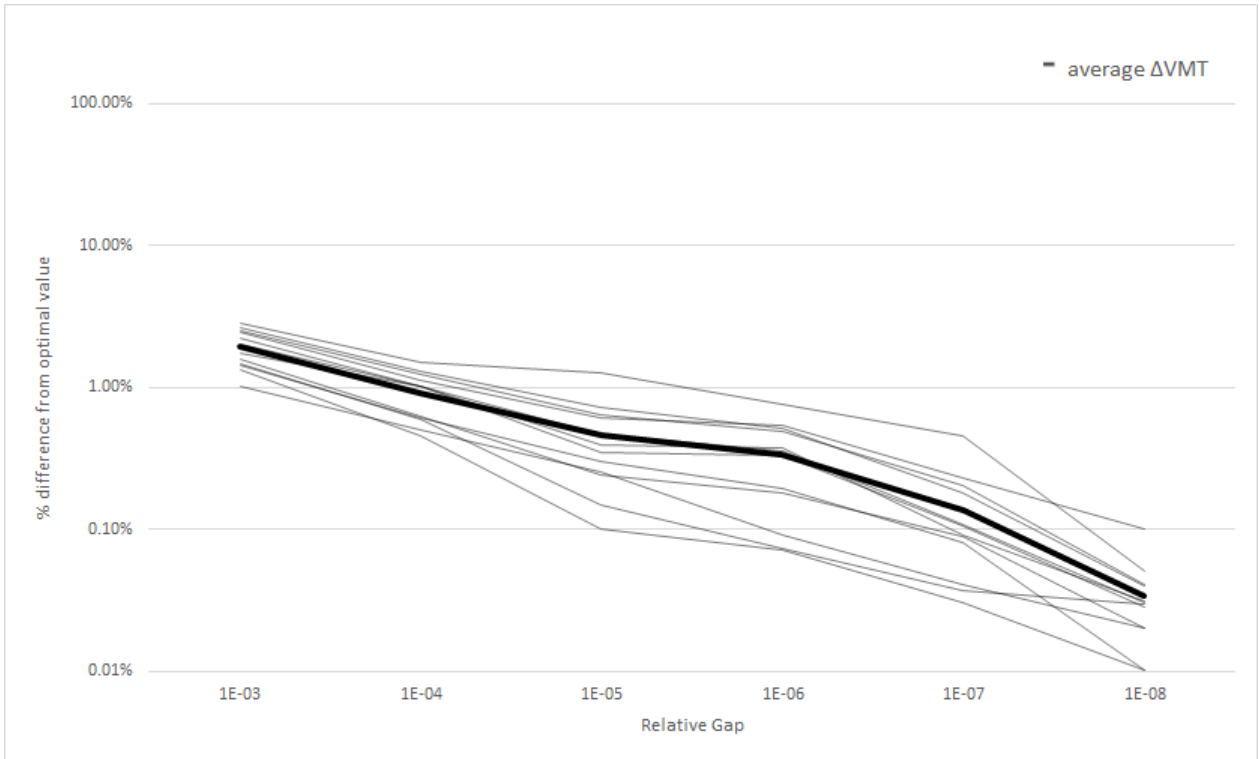
Figure 6.6: ΔVMT trends for different gap levels
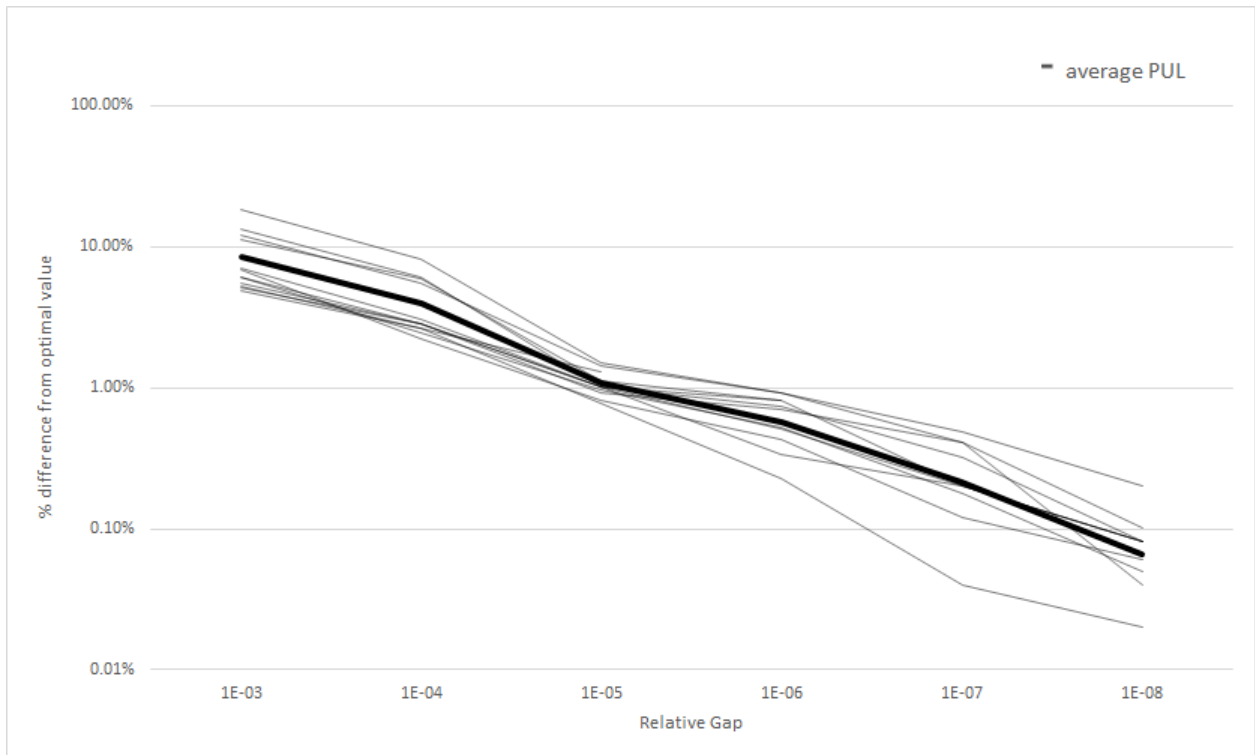
Figure 6.7: PUL trends for different gap levels
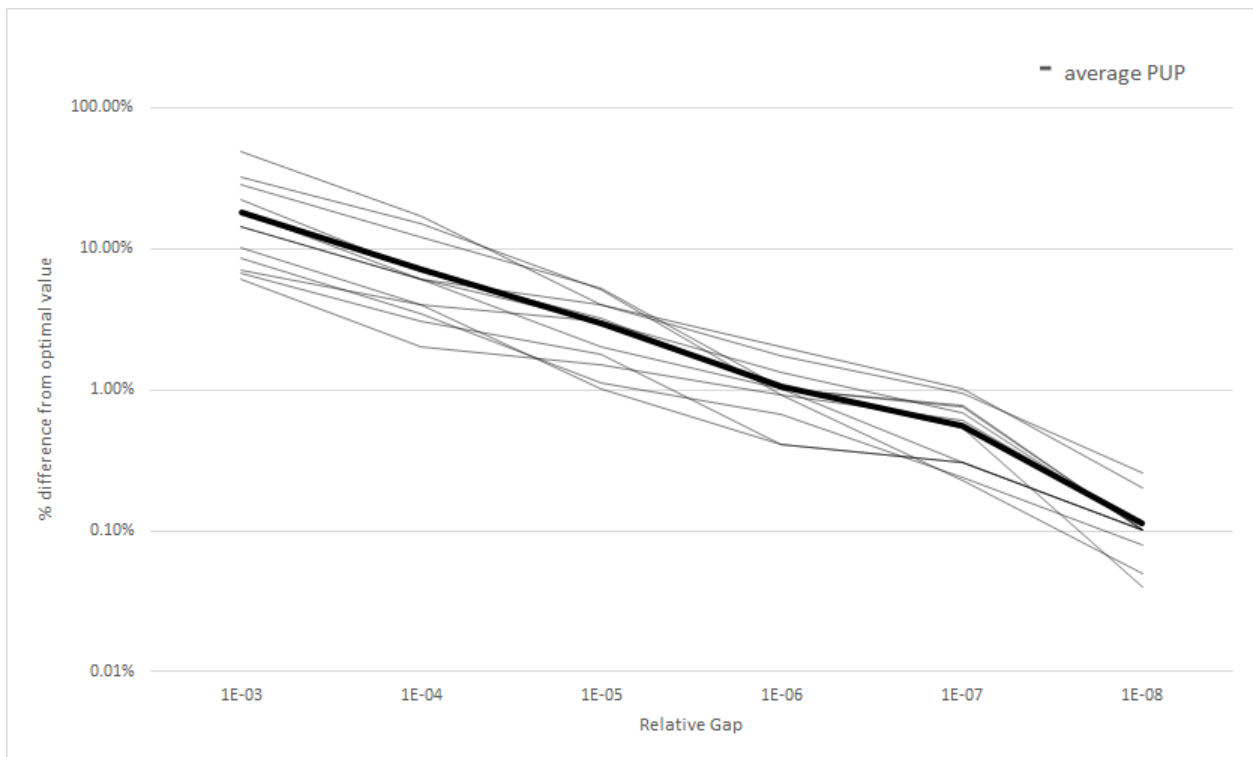
Figure 6.8: PUP trends for different gap levels
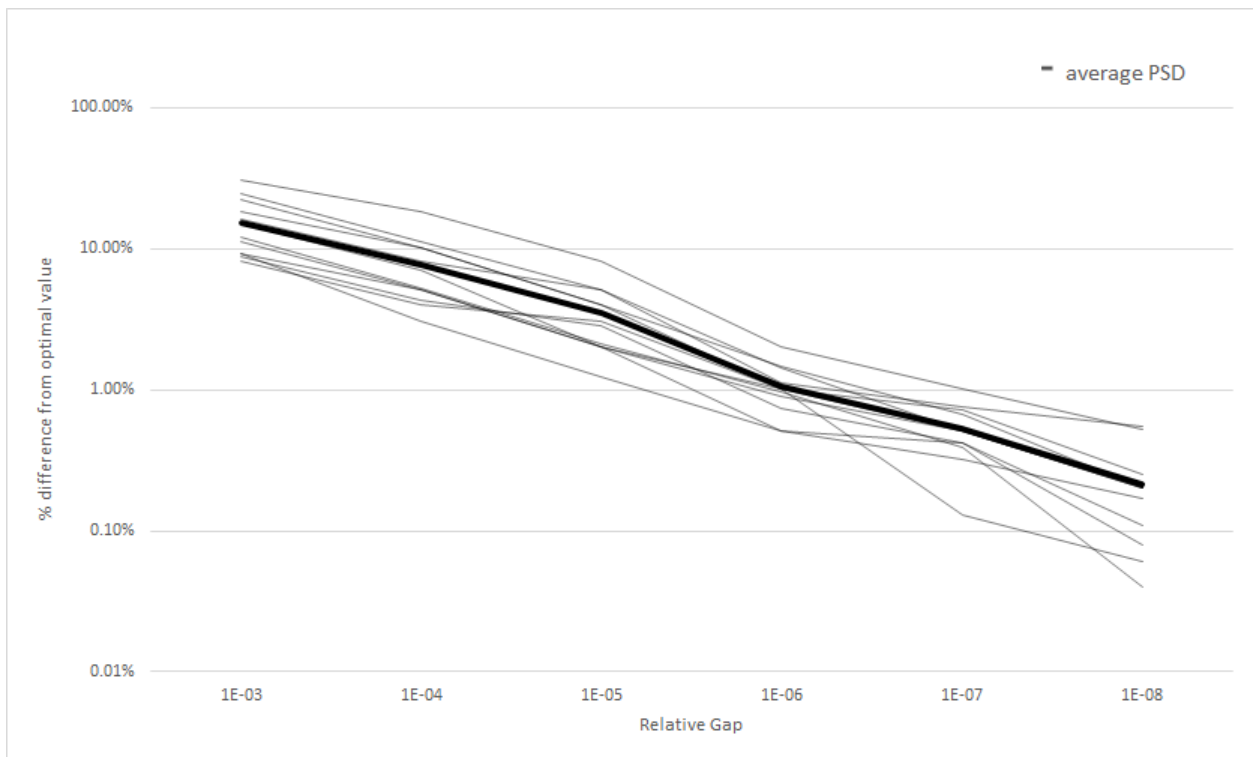
Figure 6.9: PSD trends for different gap levels

Table 6.2: Metric stabilization behavior data (average) using TAPAS

Gap Level	ΔTSTT			ΔVMT			PUL			Gap Level
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	
1E-03	1.72%	2.56%	4.05%	1.01%	1.93%	2.84%	4.89%	8.46%	18.23%	1E-03
1E-04	0.76%	1.26%	2.02%	0.46%	0.92%	1.52%	2.22%	3.92%	8.10%	1E-04
1E-05	0.20%	0.65%	1.52%	0.10%	0.46%	1.27%	0.77%	1.07%	1.52%	1E-05
1E-06	0.10%	0.42%	1.01%	0.07%	0.33%	0.76%	0.00%	0.58%	0.91%	1E-06
1E-07	0.05%	0.19%	0.66%	0.03%	0.14%	0.46%	0.00%	0.21%	0.49%	1E-07
1E-08	0.02%	0.07%	0.30%	0.01%	0.03%	0.10%	0.00%	0.07%	0.20%	1E-08

Gap Level	PUP			PSD			Gap Level
	Min	Mean	Max	Min	Mean	Max	
1E-03	6.08%	18.07%	48.60%	8.10%	15.41%	30.38%	1E-03
1E-04	2.03%	7.19%	17.21%	3.04%	7.63%	18.23%	1E-04
1E-05	1.01%	2.93%	5.26%	1.25%	3.47%	8.10%	1E-05
1E-06	0.41%	1.04%	2.03%	0.51%	1.06%	2.03%	1E-06
1E-07	0.23%	0.56%	1.01%	0.13%	0.53%	1.01%	1E-07
1E-08	0.04%	0.11%	0.26%	0.04%	0.22%	0.56%	1E-08

All five metrics converged at roughly similar rates, despite significant differences in the size and congestion level of the networks tested. This is encouraging from the standpoint of providing transferable, practical advice on convergence thresholds.

In all of the networks, the aggregate metrics ($TSTT$ and VMT) are already very near stabilization at a relative gap of 10^{-3}. For the small and medium networks, these values are within 1% of the equilibrium values when the relative gap is 10^{-4}, and for the large networks they are within 2%. Both $\Delta TSTT$ and ΔVMT converge at roughly similar rates, but ΔVMT is usually slightly lower at a particular gap level. We believe this is because the link lengths are constant, and thus only the flows are changing between iterations when calculating VMT. By contrast, the calculation of $TSTT$ involves flows and travel times, both of which are changing.

The proportion of unconverged links was the metric originally studied by Boyce et al. (2004) for the Philadelphia regional network. They found that a gap of 10^{-4} was required to approach convergence for freeway links, defining convergence as a PUL of 1% or less. To achieve this level of convergence for arterial links as well as freeway links, a relative gap of 10^{-5} was needed. Our results show that this latter conclusion generally holds across the other networks tested, and that 99% of link flows are accurate to within 1% of equilibrium values at this gap level.

Link flow behavior for multiple ϵ thresholds can be seen in Figure 6.10. Trends are similar within network size grouping, and variations therein are caused by differing congestion levels. For instance, Austin and Philadelphia show similar proportion of links in various relative error regimes, but Winnipeg and Chicago Sketch differ due to higher congestion on the Winnipeg network. This also relates to the relationship between congestion level and stabilization, explored later.

The remaining two metrics (proportion of unconverged paths and path set deviation) are the last to stabilize. Relative gap levels of 10^{-6} were needed before these metrics decreased to 1% or less. We believe this occurs because the number of used paths grows quickly with network size. For instance, in the Philadelphia network, the equilibrium solution uses over 300 million paths. Most of these paths necessarily have small flow, and changes in even a single link will change the flows across many paths.

Table 6.3 shows how the values of PUP vary across networks for different choices of ϵ, for a fairly converged solution of $RG = 10^{-6}$. This allows us to see the distribution of path convergence, similar to Figure 6.10 for links. We see that virtually all paths ($\approx$98%) are within 1% of their equilibrium values; almost all ($\approx$95% or more) are within 0.1% of their
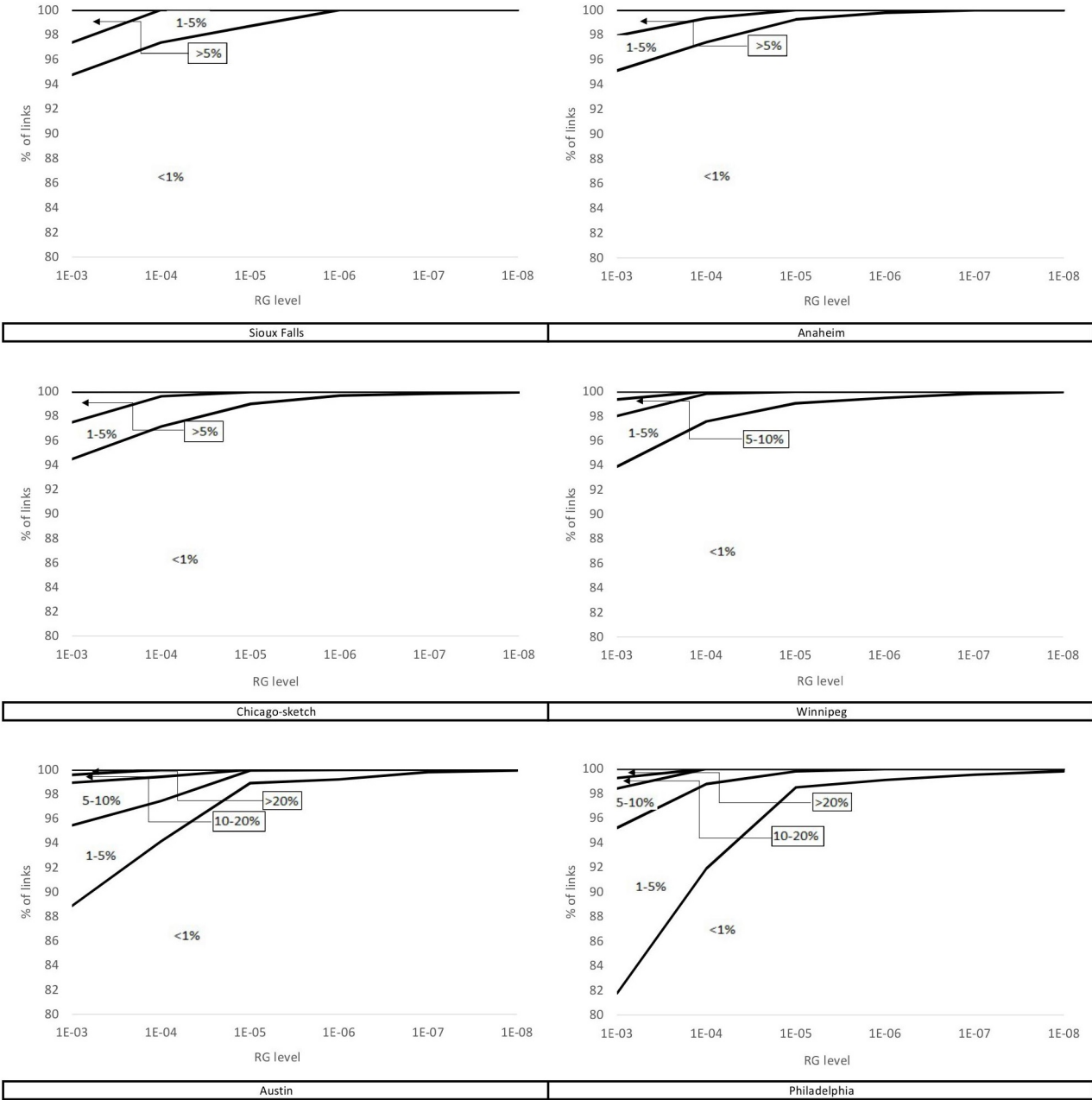
Figure 6.10: Link flow trends for various ϵ thresholds

Table 6.3: PUP sensitivity analysis w.r.t. ϵ

Sensitivity analysis for PUP(ϵ) (RG level of 10-6)				
ϵ Net	0.0001	0.001	0.01	0.1
Sioux Falls	7.02%	2.24%	0.91%	0.05%
Anaheim	4.53%	1.38%	0.41%	0.02%
Chicago Sketch	8.92%	3.30%	1.01%	0.03%
Winnipeg	5.42%	1.85%	0.67%	0.01%
Austin	9.18%	3.79%	1.01%	0.06%
Philadelphia	14.23%	5.41%	2.03%	0.17%

Table 6.4: Entropy values for varying RG values

	Chicago-Regional		Philadelphia	
Relative Gap	AlgB	TAPAS	AlgB	TAPAS
1E-03	370,682.16	850,027.30	2,609,915.27	4,797,572.35
1E-04	380,578.59	863,914.76	2,654,382.74	4,958,732.03
1E-05	389,116.51	885,997.08	2,724,551.75	5,054,152.48
1E-06	395,337.48	898,332.39	2,783,409.60	5,137,311.26
1E-07	397,353.30	908,074.29	2,793,543.55	5,189,488.01
1E-08	397,600.29	919,463.21	2,794,690.11	5,218,702.51
1E-12	397,654.25	920,159.60	2,794,750.93	5,226,184.24

equilibrium values; and the significant majority (> 85%) are within 0.01%. A negligible number of paths (roughly one in a thousand) remain more than 10% from their equilibrium values.

Table 6.4 provides the entropy values for Chicago-Regional and the Philadelphia networks at various RG levels. The entropy values show a clear increasing and convergent trend towards the final entropy value for each network-algorithm pair. Thus, as the network flow stabilizes, it tends to increase entropy, regardless of the algorithm used. As path flow patterns are intricately linked to entropy values, it stabilizes to within 1% of the convergence value at a RG level of 10^{-6} and below, in line with the observed behavior of PUP and PSD metrics.

We further investigate how path flows converge on the largest networks, by showing how the number of used paths stabilizes on the Chicago Regional and Philadelphia networks at convergence levels up to 10^{-12} relative gap. This is shown in Table 6.5. We see that the number of used paths increases with the solution precision, but that this number appears to

Table 6.5: Used paths (in millions) for various relative gap values

Gap Level	Chicago Regional	Philadelphia
1E-03	86.877	352.014
1E-04	88.819	358.008
1E-05	91.837	371.205
1E-06	91.482	368.607
1E-07	91.866	369.873
1E-08	91.895	369.998
1E-12	92.265	370.108

Table 6.6: Metric stabilization behavior data (average) for Algorithm B

Gap Level	ΔTSTT		ΔVMT		PUL	
	TAPAS	AlgB	TAPAS	AlgB	TAPAS	AlgB
1E-03	2.56%	2.38%	1.93%	1.73%	8.46%	7.57%
1E-04	1.26%	1.13%	0.92%	0.83%	3.92%	3.57%
1E-05	0.65%	0.56%	0.46%	0.38%	1.07%	0.98%
1E-06	0.42%	0.34%	0.33%	0.26%	0.58%	0.54%
1E-07	0.19%	0.14%	0.14%	0.09%	0.21%	0.21%
1E-08	0.07%	0.04%	0.03%	0.02%	0.07%	0.06%

converge, perhaps to the number of used paths at the exact equilibrium.

6.4.2 Experiment 2: Algorithm B comparison

The second experiment was performed using Algorithm B, to test transferability of the results to other traffic assignment algorithms. The results are summarized in Table 6.6, which compares the average values of each metric between TAPAS and Algorithm B.

The full data from these results are shown in Tables B.2 (raw data for Algorithm B) and B.3 (for a side-by-side comparison) in Appendix B. The path-based metrics are not computed or compared in this experiment, since Algorithm B does not attempt to provide a most likely path flow solution, and therefore its path flow results cannot be fairly compared to those of TAPAS (and indeed should not be used in practice, as with any other path flow solution which does not have high entropy).

The trends are very similar between the two algorithms, and the values of each metric are always of the same order of magnitude, and almost always nearly identical numerically. This finding is encouraging, suggesting that the conclusions of Experiment 1 are applicable

Table 6.7: Metric stabilization behavior data (average) for single-class and multi-class Algorithm B

Gap Level	ΔTSTT		ΔVMT		PUL	
	TAPAS	Multiclass	TAPAS	Multiclass	TAPAS	Multiclass
1E-03	2.56%	2.37%	1.93%	1.68%	8.46%	7.27%
1E-04	1.26%	1.01%	0.92%	0.77%	3.92%	3.37%
1E-05	0.65%	0.52%	0.46%	0.37%	1.07%	0.96%
1E-06	0.42%	0.32%	0.33%	0.26%	0.58%	0.52%
1E-07	0.19%	0.12%	0.14%	0.09%	0.21%	0.19%
1E-08	0.07%	0.04%	0.03%	0.03%	0.07%	0.06%

to other algorithms, and that the relative gap is a good universal measure of convergence, regardless of the specific assignment algorithm.

6.4.3 Experiment 3: Heterogeneous driver results

The third experiment divided the travel demand into two groups with different values of time, introducing tolls on 10% of the network links. Table 6.7 compares the values of each metric between the base case (Experiment 1) and this two-class setting. Raw data is shown in Table B.4 in Appendix B. Since these experiments were performed using our implementation of Algorithm B, path-based metrics are not computed or compared for the same reasons as in Experiment 2.

All three metrics behave extremely similar to single-class Algorithm B experiment metric behavior, indicating that the presence of multiple user classes does not significantly affect the convergence rates of these metrics.

6.4.4 Experiment 4: Network design application results

Our fourth experiment investigated the effects of subproblem precision in the network design problem, a bilevel program. In this experiment set, we varied the RG threshold used in the TAP solutions used in the lower level of this optimization problem, ranging from 10^{-3} to 10^{-8}), for minimizing TSTT. The resulting solutions at the end of the heuristic were then evaluated to a gap of 10^{-8} to compare their performance with a "benchmark" solution to the network design problem with solved all of its subproblems to a gap of 10^{-8}.

Figure 6.11 shows the gap between the objective function values with the subproblems solved at a looser gap to those with 10^{-8} (measured by the percentage difference), and
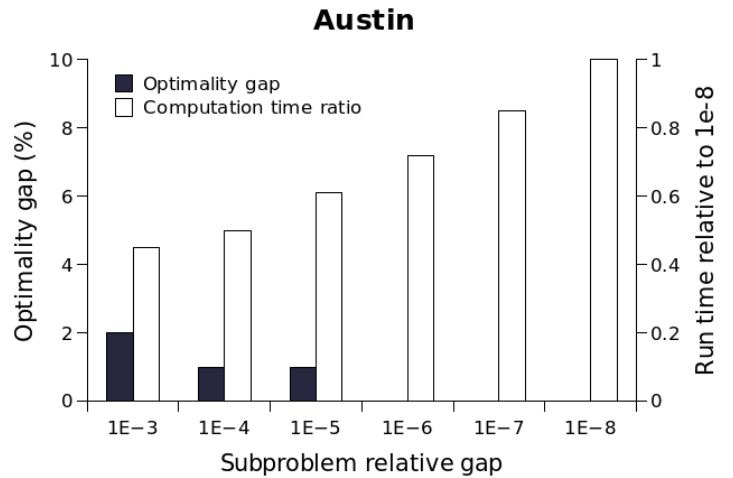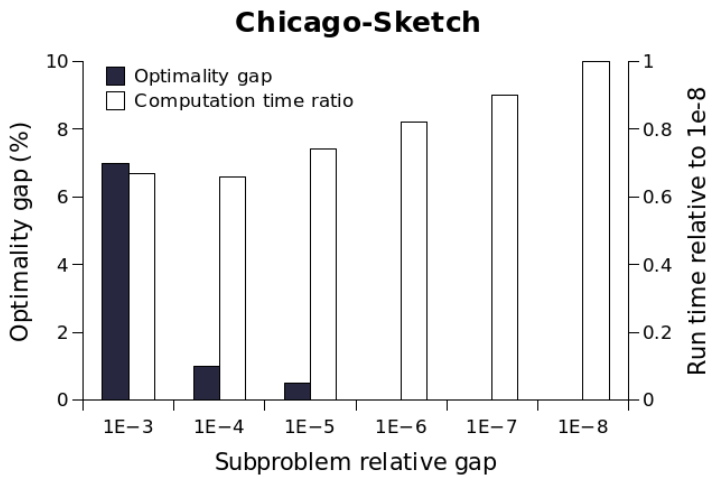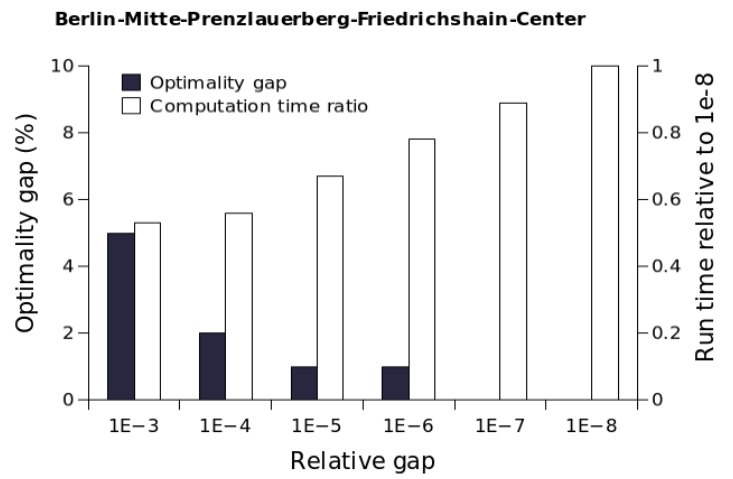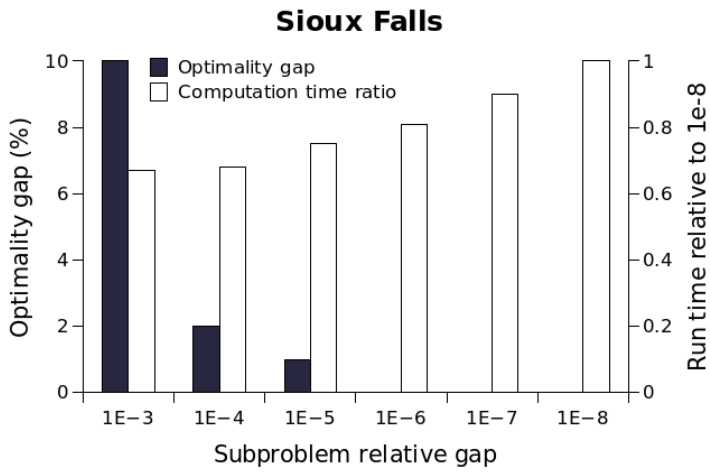
Figure 6.11: Network design performance with varying RG levels

Table 6.8: Regression of RG' and AEC with RG

Network	RG'			AEC		
	Coefficient	Intercept	R^2	Coefficient	Intercept	R^2
Sioux Falls	0.716	0.000	0.999	22.301	0.000	0.999
Anaheim	1.023	0.000	0.999	6.923	0.000	0.999
Chicago Sketch	0.968	0.000	0.999	27.565	0.000	0.999
Barcelona	0.913	0.000	0.999	8.935	0.000	0.995
Austin	1.008	0.000	0.999	7.819	0.000	0.998
Philadelphia	0.974	0.000	0.999	18.276	0.000	0.999

the computation times (reported as the fraction of time taken when solving all subproblems to 10^{-8}). The plotted values are averaged over five solutions of the genetic algorithm, which operates randomly.

For the higher convergence levels, there was no objective function gap, because the best-found solutions involved expanding the same set of links as in the solution for a gap of 10^{-8}. At these gap levels (around 10^{-6} or 10^{-7}), there was no advantage in solving the subproblems further. When the subproblems are solved to a relative gap of 10^{-4} or tighter, the objective function was within 2% of the benchmark value, and run times were decreased by 40–60%. This may be acceptable in certain applications, given the uncertainty in other components of the planning process (model specification, demand forecasting, etc.).

6.4.5 Experiment 5: Gap function comparison

The fifth experiment set compared the values of three gap functions for the solutions obtained in the previous experiments: RG, RG', and AEC. Linear regressions were performed on RG' vs. RG, and AEC vs. RG, with the results shown in Table 6.8.

The two definitions of relative gap (RG' and RG) are nearly identical, as shown by R^2 values greater than 0.999, an intercept of essentially zero, and a coefficient close to one. This confirms the analysis at the end of Section 6.2, and suggests that our conclusions can be equally applied regardless of which relative gap definition is being used.

We also observe that the ratio between AEC and RG is essentially constant within each network ($R^2 > 0.99$ and essentially zero intercept). As expected, this constant differs by network, as it reflects the average travel time on the shortest path available to travelers. For the sizes of networks used in common practice, and for the common choice of minutes as the unit for travel time, we see that AEC is roughly an order of magnitude larger than

RG. This suggests that our conclusions can be readily transferred to the *AEC* gap measure by translating them accordingly.

6.5 Conclusions

We studied the convergence rate of five metrics as the relative gap reduces over successive iterations of traffic assignment, in twelve networks of varying size and congestion levels. Across these networks, we observed trends for network metric behavior which are summarized below:

- The aggregate metrics (total system travel time and vehicle-miles traveled) were within 1% of their equilibrium values once the relative gap was below 10^{-4} (earlier for smaller networks)

- Link flows achieved stability (less than 1% of the links more than 1% away from equilibrium values) at a relative gap of 10^{-5}.

- Path flows and the sets of used paths stabilized later, at a relative gap of 10^{-6}.

- The above conclusions were seen whether TAPAS or Algorithm B was used to solve for equilibrium, and for both single and two-class assignments.

- In the network design problem, solving the subproblems to a gap level of 10^{-4} instead of 10^{-8} increased the objective function value by less than 2%, but decreased computation time by 40–60%.

- There are strong linear relationship between *RG*, *RG′* and *AEC* ($R^2 > 0.99$). This indicates the transferability of results between different gap metrics: *RG* and *RG′* can essentially be used interchangeably, whereas *AEC* differs from *RG* by a constant multiple representing average travel time.

The main limitations of the work are (1) that we propose no underlying theory to explain these findings, but present the analysis empirically; and (2) that we restrict our investigation to absolute levels of accuracy, as if a point prediction were sought for a single scenario in isolation. It is also unclear whether this guidance can be generalized to other traffic models, such as dynamic traffic assignment.

Future research should address all of these issues. In particular, regarding (1), the consistent convergence trends across very different networks (spanning several orders of

magnitude in both size and congestion level) suggest that there may be a more fundamental relationship between relative gap and these metrics. It may be possible to derive analytical relationships describing such a relationship, at least in stylized settings that roughly approximate practical traffic networks. While the current empirical results span a variety of network sizes and congestion levels to provide meaningful trends and useful guidelines to practitioners and researchers, theoretical bounds shall help generalize the findings of this study.

Regarding (2), another common application involves comparison of multiple alternatives or scenarios, where it is important to determine a stable ranking (or at least a preferred alternative). It would be valuable to see what gap levels are needed before project rankings become stable, although such a gap level would depend critically on how distinct the project impacts are, and a careful investigation is needed to account for this factor.

Chapter 7

Conclusion

This dissertation explored and extended the traffic assignment problem. Chapter 3 studied
the symmetric traffic assignment problem and convergence of solution methods. Chapter 4
applied these findings to a real world application of rail network electrification. Chapters 5
and 6 dealt with investigations of robustness and convergence behavior of classic TAP (i.e.,
without link interactions) using extensive empirical experiments.

7.1 Contributions

7.1.1 TAP with symmetric link interactions

We formulated the equilibrium problem as a convex program, and discussed the resulting
features of solution existence, uniqueness, and algorithmic tractability. Specifically, we
conducted the following experiments:

1. As an example of how existing node models can be approximated by symmetric,
   monotone link performance functions, we developed a representation of the Jin-Zhang
   merge model (Jin and Zhang, 2003) from the dynamic traffic assignment literature
   using symmetric, monotone link performance functions.

2. We discussed solution algorithms for S-TAP, including classic convex combinations
   methods, but focused primarily on more recent algorithms based on shifting flow
   between a pair of alternative segments. We showed that the flow shift formula for
   S-TAP takes a familiar and simple form, and therefore existing algorithms for TAP can
   be easily adapted for S-TAP while retaining their theoretical convergence guarantees.

3. We implemented these algorithms on standard test networks, and showed that in most cases, S-TAP actually converged *faster* than the separable TAP. Therefore, S-TAP can be considered as a serious alternative to TAP in planning practice. We also reported results showing good performance of gradient projection for asymmetric link interactions, even though there is no convex objective function (and therefore no guarantees of convergence).

4. We studied the behavior of network metrics for different convergence levels, and identified thresholds for practical algorithm convergence.

These findings were then used to formulate a rail network electrification problem, where the interaction between electrified and non-electrified flow was modeled using S-TAP. The problem was formulated as a bi-level optimization problem, and we presented a solution heuristic for the problem based on findings from Chapter 3. We also solved this problem on a large-scale network representing the North American rail system, and analyzed the results to identify crucial corridors and recommend corridor electrification.

7.1.2 Classic TAP

TAP is a well-studied problem. We first studied the behavior of TAP under input uncertainty. Specifically, we concluded the following:

1. We characterized the effect of three types of demand errors (uniform, OD-specific, and spatially correlated errors) on equilibrium system travel time, system congestion, and vehicle miles traveled.

2. We then used these findings to compare forecast and observed demand growth on a large case study network to demonstrate potential usage for planning purposes

Second, we studied the convergence behavior of important network metrics. We considered network level metrics (TSTT, VMT) as well as link/path level metrics (link set and path set convergence metrics). We examined the rates of convergence of these metrics compared to that of relative gap, a common convergence metric, and identified trends based on network size and congestion level. Our experiments also included tests of different traffic assignment algorithms, and examining a scenario with heterogeneous user classes. We also compared alternative gap functions in current use, suggesting how our results for one gap function can be translated to these alternatives.

7.2   Future Work

S-TAP is a step towards generalizing static TAP and bringing it closer to the general nature of dynamic TAP and micro-simulation approaches. While our work is a good first step towards showing the feasibility of this idea, a lot of exploration is needed in the areas of appropriate cost functions, integration of other interaction types, solution methods for asymmetric TAP, and related error bounds. Additionally, the robustness of these problems and solution methods towards input uncertainty as well as demand elasticity should be investigated. Similarly, we show but one practical application of S-TAP. Wider adoption of S-TAP for different applications will automatically give rise to future research directions in those domains.

Our contributions to the TAP literature would also benefit from an alternative approach, i.e., theoretical bounds on the errors in the results as well as convergence levels. Our work is primarily empirical in nature, with some theoretical backing. However, work in the realm of error bounds, especially using VI theory, would be greatly beneficial towards solidifying our results and advancing the field.

Appendix A

Convergence criteria experiments for STAP - raw data

Table A.1: Metric stabilization behavior data using Gradient Projection for S-TAP

Austin

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.39%	2.18%	9.55%
1E-04	1.59%	1.09%	5.38%
1E-05	0.85%	0.59%	0.90%
1E-06	0.58%	0.39%	0.72%
1E-07	0.19%	0.19%	0.19%
1E-08	0.05%	0.03%	0.08%

Berlin Center

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.58%	2.29%	10.54%
1E-04	1.87%	1.13%	4.61%
1E-05	0.99%	0.65%	1.20%
1E-06	0.64%	0.44%	0.79%
1E-07	0.19%	0.16%	0.33%
1E-08	0.08%	0.03%	0.08%

Chicago-Regional

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.24%	1.51%	11.58%
1E-04	1.24%	0.84%	5.07%
1E-05	0.61%	0.43%	0.94%
1E-06	0.46%	0.34%	0.77%
1E-07	0.18%	0.08%	0.27%
1E-08	0.04%	0.02%	0.04%

Philadelphia

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	3.83%	2.43%	15.82%
1E-04	1.62%	1.24%	7.53%
1E-05	1.05%	0.79%	1.26%
1E-06	0.63%	0.35%	0.82%
1E-07	0.36%	0.14%	0.37%
1E-08	0.10%	0.03%	0.18%

Berlin-Mitte-Prenzlauerberg-Friedrichshain-Center

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.04%	1.42%	6.24%
1E-04	0.98%	0.58%	1.95%
1E-05	0.36%	0.23%	0.78%
1E-06	0.23%	0.17%	0.44%
1E-07	0.09%	0.08%	0.10%
1E-08	0.05%	0.02%	0.05%

Barcelona

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	1.78%	1.65%	5.52%
1E-04	0.94%	0.93%	2.54%
1E-05	0.45%	0.43%	0.91%
1E-06	0.23%	0.31%	0.67%
1E-07	0.12%	0.11%	0.29%
1E-08	0.02%	0.03%	0.08%

Winnipeg

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.84%	1.81%	5.49%
1E-04	0.94%	1.01%	2.20%
1E-05	0.47%	0.30%	0.91%
1E-06	0.26%	0.30%	0.52%
1E-07	0.11%	0.10%	0.16%
1E-08	0.02%	0.03%	0.05%

Terrassa

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.55%	2.16%	6.51%
1E-04	0.98%	0.96%	2.77%
1E-05	0.52%	0.36%	0.98%
1E-06	0.29%	0.37%	0.48%
1E-07	0.13%	0.08%	0.20%
1E-08	0.02%	0.03%	0.08%

Sioux Falls

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.13%	1.48%	5.26%
1E-04	0.89%	0.65%	2.63%
1E-05	0.35%	0.18%	1.31%
1E-06	0.07%	0.03%	0.00%
1E-07	0.02%	0.01%	0.00%
1E-08	0.00%	0.00%	0.00%

Eastern Massachusetts

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	1.80%	1.25%	4.50%
1E-04	0.75%	0.44%	2.66%
1E-05	0.29%	0.10%	0.87%
1E-06	0.20%	0.06%	0.72%
1E-07	0.05%	0.03%	0.41%
1E-08	0.03%	0.01%	0.04%

Anaheim

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.47%	1.41%	4.77%
1E-04	0.77%	0.57%	2.58%
1E-05	0.41%	0.31%	0.72%
1E-06	0.25%	0.18%	0.24%
1E-07	0.10%	0.08%	0.04%
1E-08	0.05%	0.01%	0.02%

Chicago Sketch

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	1.88%	0.94%	4.87%
1E-04	0.89%	0.45%	2.49%
1E-05	0.38%	0.25%	0.96%
1E-06	0.29%	0.08%	0.32%
1E-07	0.08%	0.04%	0.18%
1E-08	0.04%	0.02%	0.07%

Table A.2: Metric stabilization behavior data using Algorithm B for S-TAP

Austin

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.87%	2.47%	11.24%
1E-04	1.91%	1.24%	5.90%
1E-05	1.02%	0.66%	1.11%
1E-06	0.61%	0.49%	0.79%
1E-07	0.20%	0.20%	0.20%
1E-08	0.05%	0.04%	0.08%

Berlin Center

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.97%	2.59%	12.29%
1E-04	1.98%	1.32%	5.44%
1E-05	1.13%	0.74%	1.41%
1E-06	0.77%	0.52%	0.91%
1E-07	0.22%	0.18%	0.40%
1E-08	0.09%	0.04%	0.10%

Chicago-Regional

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.61%	2.38%	13.07%
1E-04	1.59%	1.10%	5.97%
1E-05	0.92%	0.61%	1.00%
1E-06	0.76%	0.54%	0.79%
1E-07	0.50%	0.23%	0.00%
1E-08	0.15%	0.10%	0.00%

Philadelphia

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	3.95%	2.80%	17.88%
1E-04	1.89%	1.50%	7.91%
1E-05	1.48%	1.26%	1.49%
1E-06	1.01%	0.76%	0.89%
1E-07	0.64%	0.45%	0.49%
1E-08	0.30%	0.05%	0.20%

Berlin-Mitte-Prenzlauerberg-Friedrichshain-Center

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.11%	1.56%	6.90%
1E-04	1.03%	0.64%	2.21%
1E-05	0.40%	0.24%	0.82%
1E-06	0.26%	0.18%	0.43%
1E-07	0.10%	0.09%	0.12%
1E-08	0.05%	0.02%	0.06%

Barcelona

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.02%	1.74%	6.10%
1E-04	1.00%	1.03%	2.88%
1E-05	0.52%	0.46%	1.02%
1E-06	0.24%	0.36%	0.73%
1E-07	0.13%	0.11%	0.32%
1E-08	0.02%	0.03%	0.08%

Winnipeg

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.97%	1.99%	6.02%
1E-04	1.01%	1.00%	2.45%
1E-05	0.53%	0.35%	0.95%
1E-06	0.29%	0.33%	0.52%
1E-07	0.12%	0.11%	0.18%
1E-08	0.02%	0.03%	0.05%

Terrassa

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.74%	2.23%	7.07%
1E-04	1.15%	1.00%	3.07%
1E-05	0.52%	0.40%	1.01%
1E-06	0.29%	0.38%	0.51%
1E-07	0.13%	0.09%	0.20%
1E-08	0.02%	0.03%	0.08%

Sioux Falls

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.00%	1.44%	5.26%
1E-04	0.80%	0.59%	2.63%
1E-05	0.20%	0.15%	1.31%
1E-06	0.10%	0.07%	0.00%
1E-07	0.05%	0.04%	0.00%
1E-08	0.04%	0.03%	0.00%

Eastern Massachusetts

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	1.73%	1.34%	5.03%
1E-04	0.75%	0.44%	2.82%
1E-05	0.31%	0.10%	0.92%
1E-06	0.21%	0.07%	0.70%
1E-07	0.05%	0.03%	0.40%
1E-08	0.03%	0.01%	0.04%

Anaheim

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.56%	1.45%	4.93%
1E-04	0.80%	0.60%	2.55%
1E-05	0.40%	0.30%	0.77%
1E-06	0.26%	0.19%	0.23%
1E-07	0.10%	0.08%	0.04%
1E-08	0.05%	0.01%	0.02%

Chicago Sketch

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.03%	1.02%	5.55%
1E-04	0.99%	0.50%	2.85%
1E-05	0.37%	0.25%	0.99%
1E-06	0.29%	0.09%	0.34%
1E-07	0.08%	0.04%	0.20%
1E-08	0.04%	0.02%	0.08%

Table A.3: Metric stabilization behavior comparison between Algorithm B and Gradient Projection

Winnipeg

Gap Level	ΔTSTT		ΔVMT		PUL	
	Alg-B	GP	Alg-B	GP	Alg-B	GP
1E-03	2.97%	2.84%	1.99%	1.81%	6.02%	5.49%
1E-04	1.01%	0.94%	1.00%	1.01%	2.45%	2.20%
1E-05	0.53%	0.47%	0.35%	0.30%	0.95%	0.91%
1E-06	0.29%	0.26%	0.33%	0.30%	0.52%	0.52%
1E-07	0.12%	0.11%	0.11%	0.10%	0.18%	0.16%
1E-08	0.02%	0.02%	0.03%	0.03%	0.05%	0.05%

Austin

Gap Level	ΔTSTT		ΔVMT		PUL	
	Alg-B	GP	Alg-B	GP	Alg-B	GP
1E-03	2.87%	2.39%	2.47%	2.18%	11.24%	9.55%
1E-04	1.91%	1.59%	1.24%	1.09%	5.90%	5.38%
1E-05	1.02%	0.85%	0.66%	0.59%	1.11%	0.90%
1E-06	0.61%	0.58%	0.49%	0.39%	0.79%	0.72%
1E-07	0.20%	0.19%	0.20%	0.19%	0.20%	0.19%
1E-08	0.05%	0.05%	0.04%	0.03%	0.08%	0.08%

Philadelphia

Gap Level	ΔTSTT		ΔVMT		PUL	
	Alg-B	GP	Alg-B	GP	Alg-B	GP
1E-03	3.95%	3.83%	2.80%	2.43%	17.88%	15.82%
1E-04	1.89%	1.62%	1.50%	1.24%	7.91%	7.53%
1E-05	1.48%	1.05%	1.26%	0.79%	1.49%	1.26%
1E-06	1.01%	0.63%	0.76%	0.35%	0.89%	0.82%
1E-07	0.64%	0.36%	0.45%	0.14%	0.49%	0.37%
1E-08	0.30%	0.10%	0.05%	0.03%	0.20%	0.18%

Sioux Falls

Gap Level	ΔTSTT		ΔVMT		PUL	
	Alg-B	GP	Alg-B	GP	Alg-B	GP
1E-03	2.00%	2.13%	1.44%	1.48%	5.26%	5.26%
1E-04	0.80%	0.89%	0.59%	0.65%	2.63%	2.63%
1E-05	0.20%	0.35%	0.15%	0.18%	1.31%	1.31%
1E-06	0.10%	0.07%	0.07%	0.03%	0.00%	0.00%
1E-07	0.05%	0.02%	0.04%	0.01%	0.00%	0.00%
1E-08	0.04%	0.00%	0.03%	0.00%	0.00%	0.00%

Anaheim

Gap Level	ΔTSTT		ΔVMT		PUL	
	Alg-B	GP	Alg-B	GP	Alg-B	GP
1E-03	2.56%	2.47%	1.45%	1.41%	4.93%	4.77%
1E-04	0.80%	0.77%	0.60%	0.57%	2.55%	2.58%
1E-05	0.40%	0.41%	0.30%	0.31%	0.77%	0.72%
1E-06	0.26%	0.25%	0.19%	0.18%	0.23%	0.24%
1E-07	0.10%	0.10%	0.08%	0.08%	0.04%	0.04%
1E-08	0.05%	0.05%	0.01%	0.01%	0.02%	0.02%

Chicago-Sketch

Gap Level	ΔTSTT		ΔVMT		PUL	
	Alg-B	GP	Alg-B	GP	Alg-B	GP
1E-03	2.03%	1.88%	1.02%	0.94%	5.55%	4.87%
1E-04	0.99%	0.89%	0.50%	0.45%	2.85%	2.49%
1E-05	0.37%	0.38%	0.25%	0.25%	0.99%	0.96%
1E-06	0.29%	0.29%	0.09%	0.08%	0.34%	0.32%
1E-07	0.08%	0.08%	0.04%	0.04%	0.20%	0.18%
1E-08	0.04%	0.04%	0.02%	0.02%	0.08%	0.07%

Appendix B

Convergence criteria experiments - raw data

Table B.1: Metric stabilization behavior using TAPAS

Austin

Gap Level	ΔTSTT	ΔVMT	PUL	PUP	PSD
1E-03	2.84%	2.51%	11.14%	22.28%	16.20%
1E-04	1.92%	1.25%	5.87%	6.89%	8.10%
1E-05	1.01%	0.65%	1.11%	3.24%	5.06%
1E-06	0.61%	0.49%	0.81%	1.01%	1.42%
1E-07	0.20%	0.20%	0.20%	0.30%	0.51%
1E-08	0.05%	0.04%	0.08%	0.10%	0.20%

Berlin Center

Gap Level	ΔTSTT	ΔVMT	PUL	PUP	PSD
1E-03	2.98%	2.60%	12.15%	32.40%	24.30%
1E-04	2.02%	1.31%	5.47%	15.19%	11.14%
1E-05	1.13%	0.73%	1.42%	5.06%	5.06%
1E-06	0.77%	0.52%	0.91%	0.91%	1.11%
1E-07	0.22%	0.18%	0.41%	0.61%	0.76%
1E-08	0.09%	0.04%	0.10%	0.11%	0.56%

Chicago-Regional

Gap Level	ΔTSTT	ΔVMT	PUL	PUP	PSD
1E-03	2.63%	2.43%	13.16%	28.35%	22.28%
1E-04	1.62%	1.11%	6.08%	12.15%	10.13%
1E-05	0.91%	0.61%	1.01%	5.26%	4.05%
1E-06	0.76%	0.55%	0.81%	1.03%	1.01%
1E-07	0.50%	0.23%	0.00%	0.54%	0.72%
1E-08	0.15%	0.10%	0.00%	0.04%	0.25%

Philadelphia

Gap Level	ΔTSTT	ΔVMT	PUL	PUP	PSD
1E-03	4.05%	2.84%	18.23%	48.60%	30.38%
1E-04	1.92%	1.52%	8.10%	17.21%	18.23%
1E-05	1.52%	1.27%	1.52%	4.05%	8.10%
1E-06	1.01%	0.76%	0.91%	2.03%	2.03%
1E-07	0.66%	0.46%	0.49%	1.01%	1.01%
1E-08	0.30%	0.05%	0.20%	0.20%	0.52%

Berlin-Mitte-Prenzlauerberg-Friedrichshain-Center

Gap Level	ΔTSTT	ΔVMT	PUL	PUP	PSD
1E-03	2.15%	1.58%	6.87%	10.13%	9.11%
1E-04	1.06%	0.63%	2.22%	4.05%	3.04%
1E-05	0.41%	0.24%	0.81%	1.01%	1.25%
1E-06	0.26%	0.18%	0.43%	0.41%	0.51%
1E-07	0.10%	0.09%	0.12%	0.30%	0.42%
1E-08	0.05%	0.02%	0.06%	0.10%	0.11%

Barcelona

Gap Level	ΔTSTT	ΔVMT	PUL	PUP	PSD
1E-03	2.03%	1.74%	6.08%	14.18%	11.14%
1E-04	1.01%	1.02%	2.85%	6.08%	5.06%
1E-05	0.52%	0.47%	1.01%	3.04%	2.03%
1E-06	0.24%	0.36%	0.73%	1.34%	1.01%
1E-07	0.13%	0.11%	0.32%	0.68%	0.51%
1E-08	0.02%	0.03%	0.08%	0.10%	0.20%

Winnipeg

Gap Level	ΔTSTT	ΔVMT	PUL	PUP	PSD
1E-03	3.04%	2.03%	6.12%	8.67%	12.15%
1E-04	1.01%	1.01%	2.45%	3.42%	5.23%
1E-05	0.53%	0.35%	0.97%	1.12%	2.11%
1E-06	0.29%	0.33%	0.53%	0.67%	0.97%
1E-07	0.12%	0.11%	0.18%	0.24%	0.39%
1E-08	0.02%	0.03%	0.05%	0.08%	0.04%

Terrassa

Gap Level	ΔTSTT	ΔVMT	PUL	PUP	PSD
1E-03	2.70%	2.23%	7.09%	14.18%	18.23%
1E-04	1.16%	1.01%	3.04%	6.08%	10.13%
1E-05	0.52%	0.39%	1.01%	4.05%	4.05%
1E-06	0.29%	0.38%	0.51%	1.76%	1.46%
1E-07	0.13%	0.09%	0.20%	0.95%	0.67%
1E-08	0.02%	0.03%	0.08%	0.26%	0.20%

Sioux Falls

Gap Level	ΔTSTT	ΔVMT	PUL	PUP	PSD
1E-03	2.02%	1.47%	5.26%	6.08%	8.10%
1E-04	0.81%	0.59%	2.63%	2.03%	4.05%
1E-05	0.20%	0.15%	1.31%	1.52%	3.04%
1E-06	0.10%	0.07%	0.00%	0.91%	1.01%
1E-07	0.05%	0.04%	0.00%	0.23%	0.13%
1E-08	0.04%	0.03%	0.00%	0.05%	0.06%

Eastern Massachusetts

Gap Level	ΔTSTT	ΔVMT	PUL	PUP	PSD
1E-03	1.72%	1.34%	5.06%	7.09%	8.75%
1E-04	0.76%	0.46%	2.84%	4.05%	4.36%
1E-05	0.30%	0.10%	0.91%	3.04%	2.82%
1E-06	0.20%	0.07%	0.71%	1.01%	0.74%
1E-07	0.05%	0.03%	0.41%	0.76%	0.42%
1E-08	0.03%	0.01%	0.04%	0.10%	0.08%

Anaheim

Gap Level	ΔTSTT	ΔVMT	PUL	PUP	PSD
1E-03	2.53%	1.44%	4.89%	6.69%	9.11%
1E-04	0.81%	0.61%	2.61%	3.04%	5.06%
1E-05	0.41%	0.30%	0.77%	1.77%	2.03%
1E-06	0.26%	0.19%	0.23%	0.41%	0.51%
1E-07	0.10%	0.08%	0.04%	0.30%	0.32%
1E-08	0.05%	0.01%	0.02%	0.10%	0.17%

Chicago Sketch

Gap Level	ΔTSTT	ΔVMT	PUL	PUP	PSD
1E-03	2.03%	1.01%	5.53%	18.23%	15.19%
1E-04	1.01%	0.51%	2.86%	6.08%	7.09%
1E-05	0.37%	0.25%	1.01%	2.03%	2.03%
1E-06	0.28%	0.09%	0.34%	1.01%	0.89%
1E-07	0.08%	0.04%	0.20%	0.77%	0.51%
1E-08	0.04%	0.02%	0.08%	0.10%	0.20%

Table B.2: Metric stabilization behavior data using Algorithm B

Austin

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.39%	2.18%	9.42%
1E-04	1.60%	1.09%	5.50%
1E-05	0.85%	0.58%	0.92%
1E-06	0.58%	0.40%	0.71%
1E-07	0.20%	0.19%	0.19%
1E-08	0.05%	0.04%	0.08%

Berlin Center

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.57%	2.29%	10.77%
1E-04	1.90%	1.12%	4.71%
1E-05	0.99%	0.65%	1.21%
1E-06	0.63%	0.45%	0.79%
1E-07	0.19%	0.16%	0.33%
1E-08	0.08%	0.03%	0.09%

Chicago-Regional

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.21%	1.54%	11.51%
1E-04	1.24%	0.86%	5.12%
1E-05	0.62%	0.43%	0.95%
1E-06	0.47%	0.35%	0.76%
1E-07	0.18%	0.08%	0.28%
1E-08	0.04%	0.02%	0.04%

Philadelphia

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	3.85%	2.49%	15.61%
1E-04	1.63%	1.22%	7.59%
1E-05	1.05%	0.78%	1.27%
1E-06	0.63%	0.35%	0.82%
1E-07	0.37%	0.14%	0.37%
1E-08	0.10%	0.03%	0.18%

Berlin-Mitte-Prenzlauerberg-Friedrichshain-Center

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.03%	1.45%	6.39%
1E-04	0.98%	0.58%	1.93%
1E-05	0.37%	0.23%	0.77%
1E-06	0.23%	0.17%	0.43%
1E-07	0.09%	0.08%	0.10%
1E-08	0.05%	0.02%	0.05%

Barcelona

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	1.79%	1.64%	5.63%
1E-04	0.94%	0.94%	2.55%
1E-05	0.44%	0.43%	0.91%
1E-06	0.23%	0.31%	0.66%
1E-07	0.12%	0.10%	0.28%
1E-08	0.02%	0.03%	0.08%

Winnipeg

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.83%	1.84%	5.47%
1E-04	0.95%	1.01%	2.19%
1E-05	0.47%	0.31%	0.92%
1E-06	0.26%	0.30%	0.52%
1E-07	0.11%	0.10%	0.16%
1E-08	0.02%	0.03%	0.05%

Terrassa

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.56%	2.16%	6.49%
1E-04	0.99%	0.98%	2.82%
1E-05	0.52%	0.37%	1.00%
1E-06	0.29%	0.37%	0.48%
1E-07	0.13%	0.08%	0.20%
1E-08	0.02%	0.03%	0.08%

Sioux Falls

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.12%	1.52%	5.26%
1E-04	0.90%	0.66%	2.63%
1E-05	0.35%	0.18%	1.31%
1E-06	0.07%	0.03%	0.00%
1E-07	0.02%	0.01%	0.00%
1E-08	0.00%	0.00%	0.00%

Eastern Massachusetts

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	1.80%	1.28%	4.59%
1E-04	0.77%	0.43%	2.64%
1E-05	0.29%	0.10%	0.86%
1E-06	0.20%	0.06%	0.73%
1E-07	0.05%	0.03%	0.42%
1E-08	0.03%	0.01%	0.04%

Anaheim

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.49%	1.40%	4.85%
1E-04	0.78%	0.56%	2.63%
1E-05	0.40%	0.30%	0.71%
1E-06	0.25%	0.19%	0.24%
1E-07	0.10%	0.08%	0.04%
1E-08	0.05%	0.01%	0.02%

Chicago Sketch

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	1.91%	0.94%	4.84%
1E-04	0.91%	0.45%	2.53%
1E-05	0.37%	0.25%	0.97%
1E-06	0.29%	0.08%	0.32%
1E-07	0.08%	0.04%	0.19%
1E-08	0.04%	0.02%	0.07%

Table B.3: Metric stabilization behavior comparison between TAPAS and Algorithm B

Sioux Falls

Gap Level	ΔTSTT		ΔVMT		PUL	
	TAPAS	Alg-B	TAPAS	Alg-B	TAPAS	Alg-B
1E-03	2.02%	2.12%	1.47%	1.52%	5.26%	5.26%
1E-04	0.81%	0.90%	0.59%	0.66%	2.63%	2.63%
1E-05	0.20%	0.35%	0.15%	0.18%	1.31%	1.31%
1E-06	0.10%	0.07%	0.07%	0.03%	0.00%	0.00%
1E-07	0.05%	0.02%	0.04%	0.01%	0.00%	0.00%
1E-08	0.04%	0.00%	0.03%	0.00%	0.00%	0.00%

Winnipeg

Gap Level	ΔTSTT		ΔVMT		PUL	
	TAPAS	Alg-B	TAPAS	Alg-B	TAPAS	Alg-B
1E-03	3.04%	2.83%	2.03%	1.84%	6.12%	5.47%
1E-04	1.01%	0.95%	1.01%	1.01%	2.45%	2.19%
1E-05	0.53%	0.47%	0.35%	0.31%	0.97%	0.92%
1E-06	0.29%	0.26%	0.33%	0.30%	0.53%	0.52%
1E-07	0.12%	0.11%	0.11%	0.10%	0.18%	0.16%
1E-08	0.02%	0.02%	0.03%	0.03%	0.05%	0.05%

Anaheim

Gap Level	ΔTSTT		ΔVMT		PUL	
	TAPAS	Alg-B	TAPAS	Alg-B	TAPAS	Alg-B
1E-03	2.53%	2.49%	1.44%	1.40%	4.89%	4.85%
1E-04	0.81%	0.78%	0.61%	0.56%	2.61%	2.63%
1E-05	0.41%	0.40%	0.30%	0.30%	0.77%	0.71%
1E-06	0.26%	0.25%	0.19%	0.19%	0.23%	0.24%
1E-07	0.10%	0.10%	0.08%	0.08%	0.04%	0.04%
1E-08	0.05%	0.05%	0.01%	0.01%	0.02%	0.02%

Austin

Gap Level	ΔTSTT		ΔVMT		PUL	
	TAPAS	Alg-B	TAPAS	Alg-B	TAPAS	Alg-B
1E-03	2.84%	2.39%	2.51%	2.18%	11.14%	9.42%
1E-04	1.92%	1.60%	1.25%	1.09%	5.87%	5.50%
1E-05	1.01%	0.85%	0.65%	0.58%	1.11%	0.92%
1E-06	0.61%	0.58%	0.49%	0.40%	0.81%	0.71%
1E-07	0.20%	0.20%	0.20%	0.19%	0.20%	0.19%
1E-08	0.05%	0.05%	0.04%	0.04%	0.08%	0.08%

Chicago-Sketch

Gap Level	ΔTSTT		ΔVMT		PUL	
	TAPAS	Alg-B	TAPAS	Alg-B	TAPAS	Alg-B
1E-03	2.03%	1.91%	1.01%	0.94%	5.53%	4.84%
1E-04	1.01%	0.91%	0.51%	0.45%	2.86%	2.53%
1E-05	0.37%	0.37%	0.25%	0.25%	1.01%	0.97%
1E-06	0.28%	0.29%	0.09%	0.08%	0.34%	0.32%
1E-07	0.08%	0.08%	0.04%	0.04%	0.20%	0.19%
1E-08	0.04%	0.04%	0.02%	0.02%	0.08%	0.07%

Philadelphia

Gap Level	ΔTSTT		ΔVMT		PUL	
	TAPAS	Alg-B	TAPAS	Alg-B	TAPAS	Alg-B
1E-03	4.05%	3.85%	2.84%	2.49%	18.23%	15.61%
1E-04	1.92%	1.63%	1.52%	1.22%	8.10%	7.59%
1E-05	1.52%	1.05%	1.27%	0.78%	1.52%	1.27%
1E-06	1.01%	0.63%	0.76%	0.35%	0.91%	0.82%
1E-07	0.66%	0.37%	0.46%	0.14%	0.49%	0.37%
1E-08	0.30%	0.10%	0.05%	0.03%	0.20%	0.18%

Table B.4: Metric stabilization behavior data for multi-class assignment

Austin

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.43%	1.90%	9.32%
1E-04	1.13%	0.86%	4.53%
1E-05	0.76%	0.60%	1.03%
1E-06	0.47%	0.36%	0.62%
1E-07	0.16%	0.12%	0.15%
1E-08	0.04%	0.03%	0.06%

Berlin Center

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.69%	1.93%	9.28%
1E-04	1.22%	0.94%	4.21%
1E-05	0.84%	0.68%	1.19%
1E-06	0.52%	0.41%	0.77%
1E-07	0.19%	0.14%	0.37%
1E-08	0.08%	0.06%	0.09%

Chicago-Regional

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.20%	1.53%	10.54%
1E-04	1.15%	0.86%	5.13%
1E-05	0.69%	0.48%	0.77%
1E-06	0.42%	0.31%	0.74%
1E-07	0.13%	0.10%	0.00%
1E-08	0.04%	0.03%	0.00%

Philadelphia

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	3.46%	2.63%	14.76%
1E-04	1.24%	0.98%	6.77%
1E-05	0.80%	0.60%	1.25%
1E-06	0.58%	0.42%	0.84%
1E-07	0.20%	0.16%	0.36%
1E-08	0.11%	0.08%	0.15%

Berlin-Mitte-Prenzlauerberg-Friedrichshain-Center

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.08%	1.54%	6.96%
1E-04	1.02%	0.61%	2.17%
1E-05	0.38%	0.24%	0.81%
1E-06	0.25%	0.17%	0.41%
1E-07	0.10%	0.08%	0.12%
1E-08	0.05%	0.02%	0.06%

Barcelona

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.00%	1.73%	5.78%
1E-04	0.96%	1.04%	2.73%
1E-05	0.51%	0.46%	0.96%
1E-06	0.23%	0.35%	0.71%
1E-07	0.12%	0.11%	0.31%
1E-08	0.02%	0.03%	0.08%

Winnipeg

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	3.05%	1.93%	5.81%
1E-04	1.03%	0.98%	2.41%
1E-05	0.51%	0.33%	0.95%
1E-06	0.29%	0.34%	0.53%
1E-07	0.12%	0.10%	0.17%
1E-08	0.02%	0.03%	0.05%

Terrassa

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.60%	2.17%	6.86%
1E-04	1.14%	0.97%	2.89%
1E-05	0.52%	0.40%	1.01%
1E-06	0.29%	0.37%	0.51%
1E-07	0.13%	0.09%	0.20%
1E-08	0.02%	0.03%	0.08%

Sioux Falls

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.05%	1.36%	3.94%
1E-04	0.87%	0.56%	1.97%
1E-05	0.24%	0.12%	0.98%
1E-06	0.07%	0.03%	0.00%
1E-07	0.03%	0.00%	0.00%
1E-08	0.00%	0.00%	0.00%

Eastern Massachusetts

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	1.68%	1.24%	4.92%
1E-04	0.76%	0.46%	2.93%
1E-05	0.30%	0.10%	0.90%
1E-06	0.18%	0.05%	0.67%
1E-07	0.04%	0.02%	0.35%
1E-08	0.02%	0.00%	0.03%

Anaheim

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.73%	1.55%	4.83%
1E-04	0.81%	0.64%	2.48%
1E-05	0.41%	0.31%	0.75%
1E-06	0.26%	0.20%	0.23%
1E-07	0.11%	0.08%	0.04%
1E-08	0.05%	0.01%	0.02%

Chicago Sketch

Gap Level	ΔTSTT	ΔVMT	PUL
1E-03	2.14%	0.96%	5.68%
1E-04	1.03%	0.52%	2.97%
1E-05	0.39%	0.25%	1.05%
1E-06	0.28%	0.09%	0.35%
1E-07	0.08%	0.04%	0.20%
1E-08	0.04%	0.02%	0.09%

Bibliography

Alliance Transportation Group (2013). Statewide analysis model - third version (sam-v3). Technical report.

Andersson, M., Brundell-Freij, K., and Eliasson, J. (2016). Validation of reference forecasts for passenger transport (no. 2016: 15). Technical report.

Astroza, S., Patil, P. N., Smith, K. I., and Bhat, C. R. (2017). Transportation planning to accommodate needs of wind energy projects. *Transportation Research Record*, 2669(1):10–18.

Bain, R. (2009). Error and optimism bias in toll road traffic forecasts. *Transportation*, 36(5):469–482.

Bar-Gera, H. (2002). Origin-based algorithm for the traffic assignment problem. *Transportation Science*, 36(4):398–417.

Bar-Gera, H. (2006). Primal method for determining the most likely route flows in large road networks. *Transportation Science*, 40(3):269–286.

Bar-Gera, H. (2010). Traffic assignment by paired alternative segments. *Transportation Research Part B: Methodological*, 44(8-9):1022–1046.

Bar-Gera, H. and Luzon, A. (2007). Differences among route flow solutions for the user-equilibrium traffic assignment problem. *Journal of transportation engineering*, 133(4):232–239.

Beckmann, M., McGuire, C. B., and Winsten, C. B. (1956). Studies in the economics of transportation. Technical Report RM-1488-PR, RAND corporation.

Bernstein, D. and Gabriel, S. A. (1997). Solving the nonadditive traffic equilibrium problem. In *Network optimization*, pages 72–102. Springer.

Bertsekas, D. P. (2016). *Nonlinear Programming*. Athena Scientific, Cambridge, MA.

Bliemer, M., Brederode, L., Wismans, L. J. J., and Smits, E. (2012). Quasi-dynamic traffic assignment: static traffic assignment with queueing and spillback. In *91st Transportation Research Board (TRB) Annual Meeting 2012*, pages 1–24. Transportation Research Board (TRB).

Bliemer, M. C. and Raadsen, M. P. (2019). Static traffic assignment with residual queues and spillback. *Transportation Research Procedia*, 38:647–667.

Bliemer, M. C. and Raadsen, M. P. (2020). Static traffic assignment with residual queues and spillback. *Transportation Research Part B: Methodological*, 132:303–319.

Bliemer, M. C., Raadsen, M. P., Brederode, L. J., Bell, M. G., Wismans, L. J., and Smith, M. J. (2017). Genetics of traffic assignment models for strategic transport planning. *Transport reviews*, 37(1):56–78.

Bonnel, P., Fekih, M., and Smoreda, Z. (2018). Origin-destination estimation using mobile network probe data. *Transportation Research Procedia*, 32:69–81.

Bonnel, P., Hombourger, E., Olteanu-Raimond, A.-M., and Smoreda, Z. (2015). Passive mobile phone dataset to construct origin-destination matrix: potentials and limitations. *Transportation Research Procedia*, 11:381–398.

Borndörfer, R., Grötschel, M., and Pfetsch, M. E. (2008). Models for line planning in public transport. In *Computer-aided systems in public transport*, pages 363–378. Springer.

Boyce, D., Ralevic-Dekic, B., and Bar-Gera, H. (2004). Convergence of traffic assignments: how much is enough? *Journal of Transportation Engineering*, 130(1):49–55.

Boyles, S., Patil, P., Pandey, V., and Yahia, C. (2018). Beyond political boundaries: Constructing network models for megaregion planning. Technical Report CM2-11, The consortium of Cooperative Mobility for Competitive Megaregions (CM2).

Boyles, S. D. (2012). Bush-based sensitivity analysis for approximating subnetwork diversion. *Transportation Research Part B: Methodological*, 46(1):139–155.

Boyles, S. D. (2019). tap-b implementation. *https://github.com/spartalab/tap-b*. Accessed: 2019-05-25.

Boyles, S. D., Lownes, N. E., and Unnikrishnan, A. (2020). *Transportation Network Analysis*, volume 1. 0.85 edition.

Boyles, S. D. and Ruiz Juri, N. (2019). Queue spillback and demand uncertainty in dynamic network loading. *Transportation Research Record*, 2673(2):38–48.

Braess, D. and Koch, G. (1979). On the existence of equilibria in asymmetrical multiclass-user transportation networks. *Transportation Science*, 13(1):56–63.

Bureau, U. S. C. (2010). United States Census Bureau decennial census of population and housing. *https://www.census.gov/programs-surveys/decennial-census/data/datasets.2010.html*. Accessed: 2019-06-25.

Bureau of Public Roads (1964). *Traffic Assignment Manual for Application with a Large, High Speed Computer*, volume 37. US Department of Commerce, Bureau of Public Roads, Office of Planning.

Bureau of Transportation Statistics (2015). Travel time index. *https://www.bts.gov/content/travel-time-index*. Accessed: 2019-06-25.

Bureau of Transportation Statistics (2020). Freight analysis framework version 4. *https://faf.ornl.gov/fafweb/*. Accessed: 2020-07-25.

Business Standard (2019). Indian railways targets at least 50 of freight traffic by 2030. *https://www.business-standard.com/article/economy-policy/indian-railways-targets-at-least-50-of-freight-traffic-by-2030-119040201291_1.html*. Accessed: 2020-07-27.

Caceres, N., Romero, L., and Benitez, F. G. (2020). Exploring strengths and weaknesses of mobility inference from mobile phone data vs. travel surveys. *Transportmetrica A: Transport Science*, 16(3):574–601.

California Department of Transportation (2015). CalTrans Strategic Management Plan 2015-2020. *https://www.dot.ca.gov/perf/library/pdf/Caltrans_Strategic_Mgmt_Plan_033015.pdf*. Accessed: 2019-05-25.

Caliper Corporation (2018). *TransCAD version 8 User Guide*. Caliper Corporation, Newton, MA.

Cambridge Systematics (2012). Task 8.3: Analysis of freight rail electrification in the scag region. `http://www.freightworks.org/DocumentLibrary/CRGMSAIS%20-%20Analysis%20of%20Freight%20Rail%20Electrification%20in%20the%20SCAG%20Region.pdf`. Accessed: 2020-07-27.

Center for Transportation Analysis (2014). Cta railroad network. Accessed: 2018-10-03.

Chen, A., Jayakrishnan, R., and Tsai, W. K. (2002). Faster frank-wolfe traffic assignment with new flow update scheme. *Journal of Transportation Engineering*, 128(1):31–39.

Chen, A., Zhou, Z., and Ryu, S. (2011). Modeling physical and environmental side constraints in traffic equilibrium problem. *International Journal of Sustainable Transportation*, 5(3):172–197.

Chen, M. and Alfa, A. S. (1991). A network design algorithm using a stochastic incremental traffic assignment approach. *Transportation Science*, 25(3):215–224.

Chen, X., Liu, Z., Zhang, K., and Wang, Z. (2020). A parallel computing approach to solve traffic assignment using path-based gradient projection algorithm. *Transportation Research Part C: Emerging Technologies*, 120:102809.

Cheng, Z., Jian, S., Rashidi, T. H., Maghrebi, M., and Waller, S. T. (2020). Integrating household travel survey and social media data to improve the quality of od matrix: A comparative case study. *IEEE Transactions on Intelligent Transportation Systems*, 21(6):2628–2636.

Cherlow, J. R. (1981). Measuring values of travel time savings. *Journal of Consumer Research*, 7(4):360–371.

Chiu, Y.-C., Bottom, J., Mahut, M., Paz, A., Balakrishna, R., Waller, S., and Hicks, J. (2011). Dynamic traffic assignment: A primer (transportation research circular e-c153).

Clarke, D. B. (1995). *An examination of railroad capacity and its implications for rail-highway intermodal transportation*. PhD thesis, University of Tennessee - Knoxville.

Çolak, S., Alexander, L. P., Alvim, B. G., Mehndiratta, S. R., and González, M. C. (2015). Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities. *Transportation Research Record*, 2526(1):126–135.

Cruz, C. O. and Sarmento, J. M. (2019). Traffic forecast inaccuracy in transportation: a literature review of roads and railways projects. *Transportation*, pages 1–36.

Dafermos, S. (1980). Traffic equilibrium and variational inequalities. *Transportation science*, 14(1):42–54.

Dafermos, S. (1982). Relaxation algorithms for the general asymmetric traffic equilibrium problem. *Transportation Science*, 16(2):231–240.

Dafermos, S. C. (1971). An extended traffic assignment model with applications to two-way traffic. *Transportation Science*, 5(4):366–389.

Dafermos, S. C. (1972). The traffic assignment problem for multiclass-user transportation networks. *Transportation science*, 6(1):73–87.

Daganzo, C. F. (1994). The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B*, 28B(4):269–287.

Daganzo, C. F. (1995). The cell transmission model, part II: network traffic. *Transportation Research Part B*, 29B(2):79–93.

Daniels, G., Ellis, D. R., and Stockton, W. R. (1999). *Techniques for manually estimating road user costs associated with construction projects*, volume 3. Texas Transportation Institute, College Station, TX.

Day-Pollard, T. and van Vuren, T. (2015). When are origin-destination matrices similar enough? Paper presented at the 94th annual meeting of Transportation Research Board, Washington D.C.

De Grange, L. and Muñoz, J. C. (2009). An equivalent optimization formulation for the traffic assignment problem with asymmetric linear costs. *Transportation planning and technology*, 32(1):1–25.

De Jong, G., Daly, A., Pieters, M., Miller, S., Plasmeijer, R., and Hofman, F. (2007). Uncertainty in traffic forecasts: literature review and new results for the netherlands. *Transportation*, 34(4):375–395.

Dial, R. B. (2006). A path-based user-equilibrium traffic assignment algorithm that obviates path storage and enumeration. *Transportation Research Part B: Methodological*, 40(10):917–936.

Ditmeyer, S., Martin, J., Olson, P., Rister, M., Ross, B., Schmidt, J., and Sjokvist, E. (1985). Railway electrification and railway productivity: A study report. *Transportation Research Record*, 1029.

Do Chung, B., Yao, T., Friesz, T. L., and Liu, H. (2012). Dynamic congestion pricing with demand uncertainty: a robust optimization approach. *Transportation Research Part B: Methodological*, 46(10):1504–1518.

Drezner, Z. and Wesolowsky, G. O. (1997). Selecting an optimum configuration of one-way and two-way routes. *Transportation Science*, 31(4):386–394.

Dupuis, C. and Darveau, J.-M. (1986). The convergence conditions of diagonalization and projection methods for fixed demand asymmetric network equilibrium problems. *Operations Research Letters*, 5(3):149–155.

Duthie, J. C., Unnikrishnan, A., and Waller, S. T. (2011). Influence of demand uncertainty and correlations on traffic predictions and decision. *Computer-Aided Civil and Infrastructure Engineering*, 26(1):16–29.

Egu, O. and Bonnel, P. (2020). How comparable are origin-destination matrices estimated from automatic fare collection, origin-destination surveys and household travel survey? an empirical investigation in lyon. *Transportation Research Part A: Policy and Practice*, 138:267–282.

Eurostat (2020). Freight transport statistics - modal split. *https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Freight_transport_statistics_-_modal_split#Modal_split_in_the_EU*. Accessed: 2020-07-27.

Facchinei, F. and Pang, J. S. (2003). *Finite Dimensional Variational Inequalities and Complementarity Problems*. Springer Verlag.

Farahani, R. Z., Miandoabchi, E., Szeto, W. Y., and Rashidi, H. (2013). A review of urban transportation network design problems. *European Journal of Operational Research*, 229(2):281–302.

Federal Statistical Office - Section Mobility (2019). Passenger transport performance. https://www.bfs.admin.ch/bfs/en/home/statistics/mobility-transport/passenger-transport/performance.html. Accessed: 2020-07-27.

Fisk, C. and Nguyen, S. (1982). Solution algorithms for network equilibrium models with asymmetric user costs. *Transportation Science*, 16(3):361–381.

Florian, M., Constantin, I., and Florian, D. (2009). A new look at projected gradient method for equilibrium assignment. *Transportation Research Record*, 2090(1):10–16.

Florian, M. and Spiess, H. (1982). The convergence of diagonalization algorithms for asymmetric network equilibrium problems. *Transportation Research Part B: Methodological*, 16(6):477–483.

Flyvbjerg, B., Skamris Holm, M. K., and Buhl, S. L. (2005). How (in) accurate are demand forecasts in public works projects?: The case of transportation. *Journal of the American planning association*, 71(2):131–146.

Frank, M. and Wolfe, P. (1956). An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110.

Friesz, T. L. (1985). Transportation network equilibrium, design and aggregation: key developments and research opportunities. *Transportation Research Part A: General*, 19(5-6):413–427.

Friesz, T. L., Harker, P. T., and Tobin, R. L. (1984). Alternative algorithms for the general network spatial price equilibrium problem. *Journal of Regional Science*, 24(4):475–507.

Fritz, S.G. (2000). Diesel fuel effects on locomotive exhaust emissions. Technical Report 08.02062, Southwest Research Institute, San Antonio, TX.

Gabriel, S. A. and Bernstein, D. (1997). The traffic equilibrium problem with nonadditive path costs. *Transportation Science*, 31(4):337–348.

Galligari, A. and Sciandrone, M. (2019). A computational study of path-based methods for optimal traffic assignment with both inelastic and elastic demand. *Computers & Operations Research*, 103:158–166.

Gao, Z., Wu, J., and Sun, H. (2005). Solution algorithm for the bi-level discrete network design problem. *Transportation Research Part B: Methodological*, 39(6):479–495.

Gardner, L. M., Boyles, S. D., and Waller, S. T. (2011). Quantifying the benefit of responsive pricing and travel information in the stochastic congestion pricing problem. *Transportation Research Part A: Policy and Practice*, 45(3):204–218.

Gardner, L. M., Unnikrishnan, A., and Waller, S. T. (2008). Robust pricing of transportation networks under uncertain demand. *Transportation Research Record*, 2085(1):21–30.

Gartner, N. H. (1980). Optimal traffic assignment with elastic demands: a review part ii. algorithmic approaches. *Transportation Science*, 14(2):192–208.

Gattuso, D. and Restuccia, A. (2014). A tool for railway transport cost evaluation. *Procedia-Social and Behavioral Sciences*, 111:549–558.

Gentile, G. (2014). Local user cost equilibrium: a bush-based algorithm for traffic assignment. *Transportmetrica A: Transport Science*, 10(1):15–54.

Gillespie, A. J. and Hayes, H. I. (2003). Practical guide to railway engineering. *AREMA committee-24, USA*.

Gokalp, C., Patil, P. N., and Boyles, S. D. (2021). Post-disaster recovery sequencing strategy for road networks. *Transportation research part B: methodological*, 153:228–245.

Harrison, R., Schofield, M., Loftus-Otway, L., Middleton, D., and West, J. (2006). Freight performance measures guide. Project progress report no. 0-5410-P3, Texas Department of Transportation.

Hartgen, D. T. (2013). Hubris or humility? accuracy issues for the next 50 years of travel demand modeling. *Transportation*, 40(6):1133–1157.

Hay, W. W. (1982). *Railroad engineering*, volume 1. John Wiley & Sons.

He, B., Liao, L.-Z., and Wang, X. (2012). Proximal-like contraction methods for monotone variational inequalities in a unified framework i: effective quadruplet and primary methods. *Computational Optimization and Applications*, 51(2):649–679.

Hearn, D. W., Lawphongpanich, S., and Nguyen, S. (1984). Convex programming formulations of the asymmetric traffic assignment problem. *Transportation Research Part B: Methodological*, 18(4-5):357–365.

Heydecker, B. G. (1983). Some consequences of detailed junction modeling in road traffic assignment. *Transportation Science*, 17(3):263–281.

Higgins, S. F. (2013). Estimating economic impacts from transportation investments using the texas statewide analysis model and tredis. B.S. thesis, The University of Texas at Austin, Austin, TX.

Highway Capacity Manual (2010). Hcm2010. *Transportation Research Board, National Research Council, Washington, DC*, 1207.

Holguín-Veras, J., Cruz, C. A. T., and Ban, X. (2013). On the comparative performance of urban delivery vehicle classes. *Transportmetrica A: Transport Science*, 9(1):50–73.

Huntsinger, L. F. and Rouphail, N. M. (2011). Bottleneck and queuing analysis: calibrating volume–delay functions of travel demand models. *Transportation research record*, 2255(1):117–124.

Iliopoulou, C., Kepaptsoglou, K., and Vlahogianni, E. (2019). Metaheuristics for the transit route network design problem: a review and comparative analysis. *Public Transport*, 11(3):487–521.

Iqbal, M. S., Choudhury, C. F., Wang, P., and González, M. C. (2014). Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, 40:63–74.

Jafari, E., Pandey, V., and Boyles, S. D. (2017). A decomposition approach to the static traffic assignment problem. *Transportation Research Part B: Methodological*, 105:270–296.

Janzen, M., Vanhoof, M., Smoreda, Z., and Axhausen, K. W. (2018). Closer to the total? long-distance travel of french mobile phone users. *Travel Behaviour and Society*, 11:31–42.

Jayakrishnan, R., Tsai, W. T., Prashker, J. N., and Rajadhyaksha, S. (1994). A faster path-based algorithm for traffic assignment. Technical report, University of California Transportation Center.

Jin, W. and Zhang, H. M. (2003). On the distribution schemes for determining flows through a merge. *Transportation Research Part B: Methodological*, 37(6):521–540.

Josefsson, M. and Patriksson, M. (2007). Sensitivity analysis of separable traffic equilibrium equilibria with application to bilevel optimization in network design. *Transportation Research Part B: Methodological*, 41(1):4–31.

Katoch, S., Chauhan, S. S., and Kumar, V. (2021). A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, 80(5):8091–8126.

Kneschke, T. (1986). Simple method for determination of substation spacing for ac and dc electrification systems. *IEEE Transactions on Industry Applications*, IA-22(4). Accessed: 2020-07-27.

Kumar, A. and Peeta, S. (2010). Slope-based multipath flow update algorithm for static user equilibrium traffic assignment problem. *Transportation research record*, 2196(1):1–10.

Kumar, A. and Peeta, S. (2014). Slope-based path shift propensity algorithm for the static traffic assignment problem. *International Journal for Traffic and Transport Engineering*, 4(3):297–319.

Larsson, T. and Patriksson, M. (1992). Simplicial decomposition with disaggregated representation for the traffic assignment problem. *Transportation Science*, 26(1):4–17.

Lawphongpanich, S. and Hearn, D. W. (1984). Simplical decomposition of the asymmetric traffic assignment problem. *Transportation Research Part B: Methodological*, 18(2):123–133.

Lawrence, M., Bullock, R., and Ziming, L. (2019). China's High-Speed Rail Development. Accessed: 2021-10-20.

Lebacque, J. P. (1996). The Godunov scheme and what it means for first order traffic flow models. In *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, pages 647–678, London.

Lebacque, J. P. and Khoshyaran, M. M. (2005). First-order macroscopic traffic flow models: intersection modeling, network modeling. In *Proceedings of the 16th International Symposium on Transportation and Traffic Theory*, pages 365–386, College Park, MD.

Leblanc, L. J. (1975). An algorithm for the discrete network design problem. *Transportation Science*, 9(3):183–199.

Lee, D.-H., Nie, Y., and Chen, A. (2003). A conjugate gradient projection algorithm for the traffic assignment problem. *Mathematical and computer modelling*, 37(7-8):863–878.

Lee, J. H., Davis, A., McBride, E., and Goulias, K. G. (2019). Statewide comparison of origin-destination matrices between california travel model and twitter. In *Mobility Patterns, Big Data and Transport Analytics*, pages 201–228. Elsevier.

Lessan, J. and Fu, L. (2019). Credit-and permit-based travel demand management state-of-the-art methodological advances. *Transportmetrica A: Transport Science*, pages 1–24.

Li, Z.-C., Lam, W. H., Wong, S., and Sumalee, A. (2012). Environmentally sustainable toll design for congested road networks with uncertain demand. *International Journal of Sustainable Transportation*, 6(3):127–155.

Liao, C. (2019). od-perturbation. https://github.com/spartalab/od-perturbation. Accessed: 2019-06-15.

Litman, T. (2016). *Well measured*. Victoria Transport Policy Institute.

Lo, H. K. and Chen, A. (2000). Traffic equilibrium problem with route-specific costs: formulation and algorithms. *Transportation Research Part B: Methodological*, 34(6):493–513.

Lo, H. K. and Szeto, W. (2004). Planning transport network improvements over time. *Urban and regional transportation modeling: Essays in honor of David Boyce*, pages 157–176.

Lo, H. K. and Szeto, W. (2009). Time-dependent transport network design under cost-recovery. *Transportation Research Part B: Methodological*, 43(1):142–158.

Long, J., Gao, Z., Zhang, H., and Szeto, W. Y. (2010). A turning restriction design problem in urban road networks. *European Journal of Operational Research*, 206(3):569–578.

Lundgren, J. T. and Peterson, A. (2008). A heuristic for the bilevel origin–destination-matrix estimation problem. *Transportation Research Part B: Methodological*, 42(4):339–354.

Mahmassani, H. S. and Mouskos, K. C. (1988). Some numerical results on the diagonalization algorithm for network assignment with asymmetric interactions between cars and trucks. *Transportation Research Part B: Methodological*, 22(4):275–290.

Marcotte, P. and Guélat, J. (1988). Adaptation of a modified newton method for solving the asymmetric traffic equilibrium problem. *Transportation Science*, 22(2):112–124.

Marcotte, P. and Wynter, L. (2004). A new look at the multiclass network equilibrium problem. *Transportation Science*, 38(3):282–292.

Maryland Department of Transportation (2018). Maryland Department of Transportation Strategic Plan. `https://www.roads.maryland.gov/OPPEN/2018_MDOT_TSMO_Strategic_Plan.pdf`. Accessed: 2019-05-25.

Massachusetts Bay Transportation Authority (2018). Focus 40: The 2040 investment plan for the mbta. Accessed on 06/17/2019.

Meneguzzer, C. (1995). An equilibrium route choice model with explicit treatment of the effect of intersections. *Transportation Research Part B: Methodological*, 29(5):329–356.

Meng, Q., Yang, H., and Bell, M. G. (2001). An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem. *Transportation Research Part B: Methodological*, 35(1):83–105.

Ministry of Railways (India) (2020). Railway Electrification. Accessed: 2021-10-20.

Mishra, S., Kumar, A., Golias, M. M., Welch, T., Taghizad, H., and Haque, K. (2016). Transportation investment decision making for medium to large transportation networks. *Transportation in Developing Economies*. Accessed: 2020-07-27.

Mitradjieva, M. and Lindberg, P. O. (2013). The stiff is moving—conjugate direction frank-wolfe methods with applications to traffic assignment. *Transportation Science*, 47(2):280–293.

MLIT Ministry of Land, Infrastructure, Transport and Tourism (2018). Statistics. `https://www.mlit.go.jp/road/road_e/statistics.html`. Accessed: 2020-07-27.

Moudon, A. V. and Stewart, O. (2013). Tools for estimating VMT reductions from built environment changes. *Washington State Department of Transportation, Olympia, Washington.*

Naess, P., Andersen, J. A., Nicolaisen, M. S., and Strand, A. (2015). Forecasting inaccuracies: a result of unexpected events, optimism bias, technical problems, or strategic misrepresentation? *Journal of Transport and Land Use*, 8(3).

Nagurney, A. (1986). Computational comparisons of algorithms for general asymmetric traffic equilibrium problems with fixed and elastic demands. *Transportation Research Part B: Methodological*, 20(1):78–84.

Nagurney, A. B. (1984). Comparative tests of multimodal traffic equilibrium methods. *Transportation Research Part B: Methodological*, 18(6):469–485.

Nektalova, T. (2008). *Energy Density of Diesel Fuel*. Accessed: 2020-07-30.

Ng, M. and Waller, S. T. (2010). Reliable evacuation planning via demand inflation and supply deflation. *Transportation Research Part E: Logistics and Transportation Review*, 46(6):1086–1094.

Nguyen, S. and Dupuis, C. (1984). An efficient method for computing traffic equilibria in networks with asymmetric transportation costs. *Transportation Science*, 18(2):185–202.

Nicolaisen, M. S. (2012). *Forecasts: fact or fiction?: Uncertainty and inaccuracy in transport project evaluation*. PhD thesis, Aalborg University.

Nicolaisen, M. S. and Driscoll, P. A. (2014). Ex-post evaluations of demand forecast accuracy: A literature review. *Transport Reviews*, 34(4):540–557.

Nie, Y. (2012). A note on bar-gera's algorithm for the origin-based traffic assignment problem. *Transportation Science*, 46(1):27–38.

Nie, Y. M. (2010). Equilibrium analysis of macroscopic traffic oscillations. *Transportation Research Part B*, 44:62–72.

North Carolina Department of Transportation (2015). North Carolina Department of Transportation Strategic Plan F.Y. 2018-2022. `https://www.ncdot.gov/about-us/our-mission/Performance/Documents/StrategicPlan_2015_2017.pdf`. Accessed: 2019-05-25.

Odeck, J. (2013). How accurate are national road traffic growth-rate forecasts?—the case of norway. *Transport policy*, 27:102–111.

Odeck, J. and Welde, M. (2017). The accuracy of toll road traffic forecasts: An econometric evaluation. *Transportation Research Part A: Policy and Practice*, 101:73–85.

Organisation for Economic Co-operation and Development (2013). Recent developments in rail transportation services. `www.oecd.org/daf/competition/Rail-transportation-Services-2013.pdf`. Accessed: 2020-07-27.

Osorio-Arjona, J. and García-Palomares, J. C. (2019). Social media and urban mobility: Using twitter to calculate home-work travel matrices. *Cities*, 89:268–280.

Panicucci, B., Pappalardo, M., and Passacantando, M. (2007). A path-based double projection method for solving the asymmetric traffic network equilibrium problem. *Optimization Letters*, 1(2):171–185.

Parthasarathi, P. and Levinson, D. (2010). Post-construction evaluation of traffic forecast accuracy. *Transport Policy*, 17(6):428–443.

Patil, P., Walthall, R., and Boyles, S. D. (2022). Budget-constrained rail electrification modeling using symmetric traffic assignment: A north american case study. *Journal of Infrastructure Systems*, 28(2):04022007.

Patil, P. N. (2020). Rail electrification data repository. `https://github.com/PriyadarshanPatil/Rail_electrification_GA`. Accessed: 2020-07-25.

Patil, P. N. (2021). A-tap implementation repository. `https://github.com/PriyadarshanPatil/A-TAP-implementation`. Accessed: 2021-06-07.

Patil, P. N. and Boyles, S. D. (2022). A fresh look at symmetric traffic assignment and algorithm convergence. In *Proc., 101th Annual Meeting*. Transportation Research Board.

Patil, P. N., Dubey, S. K., Pinjari, A. R., Cherchi, E., Daziano, R., and Bhat, C. R. (2017). Simulation evaluation of emerging estimation techniques for multinomial probit models. *Journal of choice modelling*, 23:9–20.

Patil, P. N., Ross, K. C., and Boyles, S. D. (2021). Convergence behavior for traffic assignment characterization metrics. *Transportmetrica A: Transport Science*, 17(4):1244–1271.

Patriksson, M. (2015). *The traffic assignment problem: models and methods*. Courier Dover Publications.

Peeta, S. and Ziliaskopoulos, A. K. (2001). Foundations of dynamic traffic assignment: the past, the present, and the future. *Networks and Spatial Economics*, 1:233–265.

Perederieieva, O., Ehrgott, M., Raith, A., and Wang, J. Y. (2015). A framework for and empirical study of algorithms for traffic assignment. *Computers & Operations Research*, 54:90–107.

Powell, W. B. and Sheffi, Y. (1982). The convergence of equilibrium algorithms with predetermined step sizes. *Transportation Science*, 16(1):45–55.

Prager, W. (1954). *Problems* of traffic and transportation.

Qian, Z. and Zhang, M. H. (2012). Full closure or partial closure? Evaluation of construction plans for the I-5 closure in downtown Sacramento. *Journal of Transportation Engineering*, 139(3):273–286.

RailTEC (2016). Transitioning to a zero or near-zero emission line-haul freight rail system in california: Operational and economic considerations. `https://ww3.arb.ca.gov/railyard/docs/uoi_rpt_06222016.pdf`. Accessed: 2020-07-27.

Rasouli, S. and Timmermans, H. (2012). Uncertainty in travel demand forecasting models: literature review and research agenda. *Transportation letters*, 4(1):55–73.

Remi-Omosowon, A. (2020). Pyeasyga repository. `https://github.com/remiomosowon/pyeasyga`. Accessed: 2020-05-25.

Rose, G., Daskin, M. S., and Koppelman, F. S. (1988). An examination of convergence error in equilibrium traffic assignment models. *Transportation Research Part B: Methodological*, 22(4):261–274.

Rossi, T. F., McNeil, S., and Hendrickson, C. (1989). Entropy model for consistent impact-fee assessment. *Journal of urban planning and development*, 115(2):51–63.

Sancho, E. C., Ibáñez Marí, G., and Bugeda, J. B. (2015). Applying projection-based methods to the asymmetric traffic assignment problem. *Computer-Aided Civil and Infrastructure Engineering*, 30(2):103–119.

Schneck, A. and Nökel, K. (2020). Accelerating traffic assignment with customizable contraction hierarchies. *Transportation research record*, page 0361198119898455.

Schwarm, E. G. (1977). Capital and maintenance costs for fixed railroad electrification facilities. *Transportation Research Board Special Report*, 180.

Seattle Department of Transportation (2016). Seattle Department of Transportation Strategic Plan F.Y. 2018-2022. *https://www.seattle.gov/Documents/Departments/SDOT/TechnologyProgram/ITSStrategicPlan20102020.pdf*. Accessed: 2019-05-25.

Sender, J. G. and Netter, M. (1970). Equilibre offre-demande et tarification sur un réseau de transport, modele astarte (application de systèmes tarifaires à un réseau de transport : trafics et tarifs d'équilibre).

Shayanfar, E. and Schonfeld, P. (2019). Selecting and scheduling interrelated road projects with uncertain demand. *Transportmetrica A: Transport Science*, 15(2):1712–1733.

Sheffi, Y. (1985). *Urban transportation networks*, volume 6. Prentice-Hall, Englewood Cliffs, NJ.

Smith, M. J. (1979). The existence, uniqueness and stability of traffic equilibria. *Transportation Research Part B: Methodological*, 13(4):295–304.

Smith, M. J. (1983). The existence and calculation of traffic equilibria. *Transportation Research Part B: Methodological*, 17(4):291–303.

Stabler, B. (2019). Transportation networks. *https://github.com/bstabler/TransportationNetworks*. Accessed: 2019-05-25.

State of California (2018). California transportation plan 2050. Accessed on 06/17/2019.

Statistics Bureau, Ministry of Internal Affairs and Communications (2018). Statistical handbook of japan. *https://www.stat.go.jp/english/data/handbook/pdf/2018all.pdf*. Accessed: 2020-07-27.

Szeto, W. Y., Jaber, X., and O'Mahony, M. (2010). Time-dependent discrete network design frameworks considering land use. *Computer-Aided Civil and Infrastructure Engineering*, 25(6):411–426.

Szeto, W. Y. and Lo, H. K. (2006). Transportation network improvement and tolling strategies: the issue of intergeneration equity. *Transportation Research Part A: Policy and Practice*, 40(3):227–243.

Szeto, W. Y. and Lo, H. K. (2008). Time-dependent transport network improvement and tolling strategies. *Transportation Research Part A: Policy and Practice*, 42(2):376–391.

Tampère, C. M. J., Corthout, R., Cattrysse, D., and Immers, L. H. (2011). A generic class of first order node models for dynamic macroscopic simulation of traffic flows. *Transportation Research Part B*, 45:289–309.

Texas Department of Transportation (2015). Texas transportation plan 2040. Accessed on 06/17/2019.

Texas Department of Transportation (2019). Texas transportation plan 2050. Accessed on 06/17/2019.

Uddin, M. M. and Huynh, N. (2015). Freight traffic assignment methodology for large-scale road–rail intermodal networks. *Transportation Research Record*, 2477(1):50–57.

Ukkusuri, S. V., Mathew, T. V., and Waller, S. T. (2007). Robust transportation network design under demand uncertainty. *Computer-Aided Civil and Infrastructure Engineering*, 22(1):6–18.

United States Department of Transportation - Federal Railroad Administration (2015). Cost-benefit analysis of rail electrification for next generation freight and passenger rail transportation. *https://cms8.fra.dot.gov/sites/fra.dot.gov/files/fra_net/19061/Cost%20Benefit%20Analysis%20of%20Rail%20Electrification.pdf*. Accessed: 2020-07-27.

U.S. Bureau of Labor Statistics (2020). Producer price indexes. *https://www.bls.gov/ppi*. Accessed: 2020-07-30.

US Department of Transportation (2016). US Department of Transportation Value of Time Memorandum. *https://www.transportation.gov/sites/dot.gov/files/docs/Value%20of%20Travel%20Time%20Memorandum.pdf*. Accessed: 2019-05-25.

US Department of Transportation (2017). US Department of Transportation Strategic Plan F.Y. 2018-2022. `https://www.transportation.gov/sites/dot.gov/files/docs/mission/administrations/office-policy/304866/dot-strategic-planfy2018-2022508.pdf`. Accessed: 2019-05-25.

U.S. Department of Transportation - Federal Railroad Administration (2019). Federal-state partnership for state of good repair grant program (fy 2019). `https://railroads.dot.gov/grants-loans/competitive-discretionary-grant-programs/federal-state-partnership-state-good-repair-0`. Accessed: 2020-05-25.

U.S. Geological Survey (2007). North america elevation 1-kilometer resolution grid. `https://www.sciencebase.gov/catalog/item/4fb5495ee4b04cb937751d6d`. Accessed: 2020-07-27.

Van Vuren, T. and Watling, D. (1991). A multiple user class assignment model for route guidance. *Transportation research record*, pages 22–22.

Van Wee, B. (2007). Large infrastructure projects: a review of the quality of demand forecasts and cost estimations. *Environment and Planning B: Planning and Design*, 34(4):611–625.

Venkatraman, R., Boyles, S. D., James, R., Unnikrishnan, A., and Patil, P. N. (2021). Adaptive routing behavior with real-time information under multiple travel objectives. *Transportation Research Interdisciplinary Perspectives*, 10.

Waller, S. T., Schofer, J. L., and Ziliaskopoulos, A. K. (2001). Evaluation with traffic assignment under demand uncertainty. *Transportation Research Record*, 1771(1):69–74.

Walthall, R. (2019). Rail electrification's potential for emissions abatement in the freight industry : a case study of a transcontinental rail corridor. Technical report.

Wang, H., Nozick, L., Xu, N., and Gearhart, J. (2018a). Modeling ocean, rail, and truck transportation flows to support policy analysis. *Maritime Economics & Logistics*, 20(3):327–357.

Wang, Z., He, S. Y., and Leung, Y. (2018b). Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society*, 11:141–155.

Wardrop, J. G. and Whitehead, J. I. (1952). Correspondence. some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1(5):767–768.

Weisbrod, G. (2008). Models to predict the economic development impact of transportation projects: historical experience and new applications. *The Annals of Regional Science*, 42(3):519–543.

Whitford, R. K. (1981). *Railroad Electrification: An Alternative for Petroleum Saving*. Automotive Transportation Center, Purdue University.

Wong, S. C., Yang, C., and Lo, H. K. (2001). A path-based traffic assignment algorithm based on the transyt traffic model. *Transportation Research Part B: Methodological*, 35(2):163–181.

Xie, J., Nie, Y., and Liu, X. (2018). A greedy path-based algorithm for traffic assignment. *Transportation Research Record*, 2672(48):36–44.

Xie, J. and Xie, C. (2014). An improved tapas algorithm for the traffic assignment problem. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 2336–2341. IEEE.

Xie, J. and Xie, C. (2015). Origin-based algorithms for traffic assignment: algorithmic structure, complexity analysis, and convergence performance. *Transportation Research Record*, 2498(1):46–55.

Yang, H. (1995). Heuristic algorithms for the bilevel origin-destination matrix estimation problem. *Transportation Research Part B: Methodological*, 29(4):231–242.

Yang, H. (1997). Sensitivity analysis for the elastic-demand network equilibrium problem with applications. *Transportation Research Part B: Methodological*, 31(1):55–70.

Yang, H. and Bell, M. G. H. (1998). Models and algorithms for road network design: a review and some new developments. *Transport Reviews*, 18(3):257–278.

Yang, H. and Lam, W. H. (1996). Optimal road tolls under conditions of queueing and congestion. *Transportation Research Part A: Policy and Practice*, 30(5):319–332.

Yin, Y. (2000). Genetic-algorithms-based approach for bilevel programming models. *Journal of transportation engineering*, 126(2):115–120.

Yin, Y., Madanat, S. M., and Lu, X.-Y. (2009). Robust improvement schemes for road networks under demand uncertainty. *European Journal of Operational Research*, 198(2):470–479.

Yook, D. (2014). Models and solution algorithms for asymmetric traffic and transit assignment problems.

Yook, D. and Heaslip, K. (2016). Acceleration of double-projection method in asymmetrically formulated traffic assignment. *Journal of Computing in Civil Engineering*, 30(6):04016025.

Yperman, I. (2007). *The Link Transmission Model for Dynamic Newtork Loading*. PhD thesis, Katholieke Universiteit Leuven, Belgium.

Zhang, H., Nie, Y., and Qian, Z. (2013). Modelling network flow with and without link interactions: the cases of point queue, spatial queue and cell transmission model. *Transportmetrica B: Transport Dynamics*, 1(1):33–51.

Zhao, Y. and Kockelman, K. M. (2002). The propagation of uncertainty through travel demand models: an exploratory analysis. *The Annals of Regional Science*, 36(1):145–163.

Vita

Priyadarshan Nandkumar Patil was born in Kolhapur, Maharashtra, India on February 17, 1993, to Sucheta and Nandkumar Patil. After graduating from Loyola High School and MIT Jr. College in Pune, Maharashtra in 2008 and 2010, respectively, he enrolled at the Indian Institute of Technology Madras in Chennai, Tamil Nadu, where he received a Bachelor of Technology degree in civil engineering, graduating in 2015. In August 2015, he entered The Graduate School at The University of Texas at Austin under the supervision of Dr. Chandra R. Bhat. He received a Master of Science degree in civil engineering in January 2017, with an emphasis in transportation engineering. In August 2017, he started as a doctoral student in the Operations Research and Industrial Engineering program at The University of Texas at Austin, under the supervision of Dr. Stephen D. Boyles.

Permanent Address: Flat 17, Bldg. 18B, Ravikiran Society
Sakalnagar, Baner Road, Aundh
Pune, Maharashtra, India
Zip code: 411007