The Dissertation Committee for Man Wo Ng Certifies that this is the approved version of the following dissertation:

Travel Time Reliability Assessment Techniques for Large-Scale Stochastic Transportation Networks

Committee:

S. Travis Waller, Supervisor

Kara M. Kockelman

Zhanmin Zhang

John J. Hasenbein

David P. Morton

Travel Time Reliability Assessment Techniques for Large-Scale Stochastic Transportation Networks

by

Man Wo Ng, B.Sc., M.Sc.

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

May, 2010

Acknowledgements

I would like to take this opportunity to express my gratitude to the members of the dissertation committee. First and foremost, I would like to thank my advisor, Dr. S. Travis Waller, for providing me the chance to enter the world of transportation – after spending my first year at UT in the Operations Research and Industrial Engineering (ORIE) program. I am also grateful for his advice, flexibility, and help at critical moments in my educational career. My gratitude also goes to Dr. Kara M. Kockelman who is always willing to help and to provide suggestions and who always promptly responds to my emails. I am grateful to Dr. Zhanmin Zhang for his encouragement and advice in our joint work. My educational experience at UT has also been enriched by Dr. David. P. Morton. I have very much enjoyed learning optimization techniques from him, via his exemplary teaching style. Finally, I am grateful to Dr. John J. Hasenbein (who would have been my advisor if I had stayed in ORIE) for serving on multiple of my committees.

Outside the dissertation committee, I am indebted to Libbie A. Toler for her administrative support and fellow students in the transportation engineering program that have made the Ph.D. experience more enjoyable, with special mention for David R. Suescun (also famously known as “Mr. Suescun”).

Travel Time Reliability Assessment Techniques for Large-Scale Stochastic Transportation Networks

Publication No.______________

Man Wo Ng, Ph.D.
The University of Texas at Austin, 2010

Supervisor: S. Travis Waller

Real-life transportation systems are subject to numerous uncertainties in their operation. Researchers have suggested various reliability measures to characterize their network-level performances. One of these measures is given by travel time reliability, defined as the probability that travel times remain below certain (acceptable) levels. Existing reliability assessment (and optimization) techniques tend to be computationally intensive. In this dissertation we develop computationally efficient alternatives. In particular, we make the following three contributions.

In the first contribution, we present a novel reliability assessment methodology when the source of uncertainty is given by road capacities. More specifically, we present a method based on the theory of Fourier transforms to numerically approximate the probability density function of the (system-wide) travel time. The proposed methodology takes advantage of the established computational efficiency of the fast Fourier transform.

In the second contribution, we relax the common assumption that probability distributions of the sources of uncertainties are known explicitly. In reality, this distribution may be unavailable (or inaccurate) as we may have no (or insufficient) data to calibrate the distributions. We present a new method to assess travel time reliability that is distribution-free in the sense that the methodology only requires that the first N moments (where N is any positive integer) of the travel time to be known and that the travel times reside in a set of known and bounded intervals. Instead of deriving exact probabilities on travel times exceeding certain thresholds via computationally intensive methods, we develop analytical probability inequalities to quickly obtain upper bounds on the desired probability.

Because of the computationally intensive nature of (virtually all) existing reliability assessment techniques, the optimization of the reliability of transportation systems has generally been computationally prohibitive. The third and final contribution of this dissertation is the introduction of a new transportation network design model in which the objective is to minimize the unreliability of travel time. The computational requirements are shown to be much lower due to the assessment techniques developed in this dissertation. Moreover, numerical results suggest that it has the potential to form a computationally efficient proxy for current simulation-based network design models.

Table of Contents

Chapter 1: Introduction

1.1 MOTIVATION

Travel time has long been recognized as an important criterion in the route choice selection process of travelers. More recently, empirical studies have demonstrated that, besides the (mean) travel time, travel time variability is another important criterion (Knight, 1974; Jackson and Jucker, 1981; Abdel-Aty et al., 1995; Noland and Small, 1995; Bates et al., 2001; Brownstone and Small, 2005; Asensio and Matas, 2008). Studies on the valuation of reliability are typically based on the theory of discrete choice, using either revealed preference data (Lam and Small, 2001) or, much more commonly, stated preference data (e.g., Abdel-Aty et al., 1995; Bates et al., 2001; Asensio and Matas, 2008). Several studies have even concluded that the variability of travel time is a more important criterion in route choice than the mean travel time value (e.g., Bates et al., 2001; Liu et al., 2004, Lemp and Kockelman, 2010). Hence, questions that naturally arise are how reliable travel times are and how one can improve reliability. In this dissertation, we will address these questions at a network level, rather than at a link level or from the perspective of an individual traveler (although the techniques developed can easily be adapted to examine the reliability assessment problem from a single individual's perspective). To date, only a limited number of studies have examined the transportation network reliability problem from an aggregate perspective (see Section 1.2 for a comprehensive literature review). These studies tend to be computationally intensive and/ or are based on restrictive modeling assumptions. Thus, there is a need for a new generation of reliability assessment techniques that are computationally light and, at the same time, are premised on more realistic modeling assumptions. This dissertation presents significant progress in this direction by complementing existing work and relaxing frequently found assumptions.

1.2 LITERATURE REVIEW

It is widely accepted that the reliability of a transportation system is critical for society at large (Berdica, 2002). In terms of transportation networks, the reliability literature has its roots in the study of what has become known as **connectivity or terminal reliability** (Wakabayashi and Iida, 1992; Bell and Iida, 1997; Asakura et al., 2003). This type of study examines the probability that specific Origin-Destination (OD) pairs in a network remain connected when links are subject to complete failures. Because of the binary character of the link performance (they are either in service or not, or more generally, provide an acceptable level of service or not), connectivity reliability tends to be more appropriate for extreme events (e.g., earthquakes).

Travel time reliability relates to the probability that travel times remain below acceptable levels. The earliest studies in this area used extensive computer simulation to determine the reliability of travel times (e.g., Asakura and Kashiwadani, 1991), while later studies employed sensitivity analysis to reduce the computational burden (Du and Nicholson, 1997; Bell et al., 1999). In an attempt to further improve on the computational efficiency, Sumalee and Watling (2003) proposed an approach to obtain bounds on the reliability by only considering a subset of all possible scenarios. However, as the authors have noted, the approximation scheme is efficient only in situations where a large fraction of the probability mass is concentrated on a relatively small number of scenarios. Recently, the same authors proposed a novel approach based on the partitioning of the sample space to obtain bounds on the reliability of travel time (Sumalee and Watling, 2008).

The output of network-level reliability studies is typically a single scalar performance index (e.g., the probability that the system travel time exceeds a *predetermined* threshold) as a summary of the overall system performance. Clark and Watling (2005) departed from this philosophy and constructed the entire Probability Density Function (PDF) of the system-wide travel time. With the entire PDF it becomes possible to evaluate the probability that the travel time exceeds *any* given value. Bell (2000) and Bell and Cassir (2002) also assumed a novel approach by taking a more pessimistic view in the assessment of travel time reliability. They proposed models based on game theory to assess the worst case performance of a network in terms of its travel time.

Chen et al. (1999) introduced the notion of **capacity reliability**, which is defined as the probability that the transportation system can accommodate a given demand level at an acceptable level of service. A comprehensive simulation-based framework to assess this particular form of reliability was presented and discussed in Chen et al. (2002). More recently, Sumalee and Karauchi (2006) used the concept of capacity reliability to evaluate network reliability in the wake of a major disaster.

If there were no uncertainty in a transportation network, the above probabilistic reliability measures would either be zero or one. In the current literature, the sources of uncertainty in a transportation network are often categorized as **demand uncertainty** (e.g., Asakura and Kashiwadani, 1991; Waller et al., 2001; Clark and Watling, 2005; Lam et al. 2008) or **capacity uncertainty** (e.g., Wakabayashi and Iida, 1992; Chen et al. 2002; Sumalee and Watling, 2003; Sumalee and Watling, 2008; Ng and Waller, 2009a). A major theme in these studies is the assumption of statistical independence. Unless the study is simulation-based (Chen et al., 2002 assumed that capacity degradations were correlated continuous random variables; for a comprehensive review of this correlation-based approach, see Ng et al., 2010), it is often necessary to make the independence assumption for mathematical tractability. However, as pointed out in Lo and Tung (2003), while it is true that for certain situations the modeling of dependencies is crucial for the validity of the reliability assessment (e.g., during floods and earthquakes when certain areas are simultaneously affected), for other situations, the assumption of independence might be justifiable (e.g., traffic accidents, parking violations). Apart from this, the validity of the independence assumption might also depend on the geographical region in which the network is located as certain regions are more prone to incidents that would give rise to dependent capacity reductions in a road network. Because of the challenging nature, only a very limited number of studies have attempted to relax the assumption of independence. For instance, by postulating a multivariate normal distribution for the link flows in a network, Clark and Watling (2005) were able to model dependencies. Sumalee and Watling (2003, 2008) used a cause-based failure framework to introduce correlations in the capacity degradations.

A final criterion to characterize the transportation network reliability assessment literature is travel behavior under uncertainty. Three assumptions can be found in the literature. The first possibility assumes that travelers exhibit **non-adaptive** behavior, i.e, they do not change their paths as a function of the unfolding scenario. One motivation for such behavior is that travelers simply do not have enough time to react because of the unpredictability of events (Clark and Watling, 2005). On the other extreme, Lo and Tung (2003) suggested that people have learned about the possible scenarios, based on which they have settled in a single, fixed, long-term equilibrium pattern accounting for the uncertainties. An immediate consequence of the non-adaptivity assumption is that only a single, representative run of traffic assignment (e.g., based on some nominal levels of demand and capacity) is needed to predict travel patterns. Another possibility hypothesizes that only the travelers on affected routes (i.e, routes whose capacities have reduced) have the ability to change their paths while travelling. The behavior underlying this type of **partial adaptivity** has been coined partial user equilibrium (Sumalee and Watling, 2003). The third and final possibility allows for **fully adaptive** behavior, i.e, for every single scenario travelers decide on a new path to follow (e.g., Chen et al, 2002). Clearly, there is still little consensus on travel behavior under uncertainty. Various network equilibrium models – with different behavioral assumptions – have been proposed in the literature to model this behavior (Yin and Ieda, 2001; Watling, 2002; Lo and Tung, 2003; Yin et al., 2004; Lo et al., 2006; Shao et al., 2006; Siu and Lo, 2008; Szeto et al., 2006; Zhou and Chen, 2008; Lam et al., 2008; Ng and Waller, 2009b).

1.3 PROBLEM DEFINITION

In order to be more precise about the contribution of this dissertation, some mathematical notation is needed. Let A denote the set of links in a transportation network (with some abuse of notation, we shall also refer to the cardinality of this set as A; from the context it should be clear what is meant). And let v_a, c_a and t_a^f denote the traffic volume, capacity and free-flow speed on link $a \in A$, respectively. In the transportation network reliability literature (e.g., Chen et al. 2002; Lo and Tung, 2003; Ng and Waller, 2009a), the travel time t_a on link a is often assumed to be given by the Bureau of Public Roads (BPR) volume-delay function (Bureau of Public Roads, 1964):

$$t_a \equiv t_a^f \left(1 + \alpha \left(\frac{v_a}{c_a} \right)^{\beta} \right).$$

For notational convenience, the parameters α and β are assumed to be link independent, which can easily be relaxed by the addition of a subscript. Uncertainties in capacity (e.g., Lo and Tung, 2003; Ng and Waller, 2009a) and link flows (e.g., Clark and Watling, 2005) translate into uncertainties in the travel times according to the above nonlinear relation.

Throughout this dissertation, the tilde (~) on a lower case letter is used to denote its random counterpart. For example, if the road capacity is random, we write $\tilde{c}_a$ instead of c_a. On occasions, capital letters are also used to denote random variables. For instance, the total random travel time on link a is denoted as T_a. From the context, it should be clear what variables are random. The random variable of interest in network-level reliability assessment is the total system travel time (TSTT):

$$T \equiv \sum_{a=1}^{A} T_a ,$$

In particular, from a reliability perspective, we are mainly interested in the unreliability function $R(x)$ of T at x (here and in the subsequent we use Pr(A) to denote the probability of event A), which is defined as:

$$R(x) \equiv \Pr\left(T > x\right).$$

Equivalently, one can examine the PDF of T as it uniquely determines $R(x)$. Direct estimation (i.e, via simulation) of the unreliability function is computationally prohibitive. Computationally efficient methodologies are needed to complement and supplement existing techniques in the current literature. They will also form the starting to optimize reliability in a transportation network.

Let us close this section with two remarks. First, notice that in the probability literature, $R(x)$ is oftentimes referred to as the *reliability* function (e.g., Ross, 2002). However, unlike in the transportation reliability literature – where small values of $R(x)$ are desirable – in the probability literature, high values of $R(x)$ typically indicate "good" system performance (where T would for example correspond to the lifetime of a component of a machine). For this reason, we have chosen to call $R(x)$ the *unreliability* function. Second, we want to note that most techniques developed in this dissertation can be readily modified to analyze travel time reliability at the level of a single individual (as opposed to an entire network), simply by redefining T_a as the travel time experienced by a single traveler (e.g., in a BPR-framework, one would have $T_a \equiv \tilde{t}_a$). However, in light of existing work in the travel time reliability assessment literature (that all focused on the

TSTT), we take the system-level perspective in this dissertation. The reader should bear in mind that the results can be readily modified for a more disaggregate analysis.

1.4 Outline of Dissertation

This remainder of this dissertation is organized as follows. Chapter 2 presents a novel methodology to assess travel time reliability in a transportation network, when the source of uncertainty is given by random road capacities. Specifically, we present a method based on the theory of Fourier transforms to numerically approximate the PDF of the system-wide travel time. Any common continuous or discrete probability distribution can be used to model capacity uncertainty. Theoretical bounds on the approximation errors are formally derived, both for general distributions as well as for the specific instance of normally distributed capacities. These bounds provide valuable insights into the structure of the approximation errors and suggest ways to reduce them. From a practical point of view, we propose a procedure based on successively refining the computational grid in order to guarantee accurate approximations. The proposed methodology takes advantage of the established computational efficiency of the fast Fourier transform.

An assumption that pervades the current transportation system reliability assessment literature is that probability distributions of the sources of uncertainty are known explicitly. However, this distribution may be unavailable (inaccurate) in reality as we may have no (insufficient) data to calibrate the distribution. Chapter 3 relaxes this assumption and presents a new method to assess travel time reliability that is distribution-free in the sense that the methodology only requires that the first N moments (where N is any positive integer) of the travel time to be known and that the travel times reside in a set of bounded and known intervals. Because of our modeling approach, all sources of uncertainty are automatically accounted for, as long as they are statistically independent. Furthermore, the results do not make explicit reference to any specific volume-delay function. Instead of deriving exact probabilities on travel times exceeding certain thresholds via computationally intensive methods, we develop analytical probability inequalities to quickly obtain upper bounds on the desired probability.

In Chapter 4 we introduce a new transportation network design model that is a direct application of the travel time reliability assessment technique developed in Chapter 3. Unlike in traditional network design models (where the objective is typically to minimize system travel time), the proposed model aims to minimize the unreliability of transportation networks. The potential of the new model as a computationally efficient alternative to existing simulation-based models is examined and discussed. Comparisons with traditional network design models aimed at improving system travel time will also be made.

Chapter 5 concludes this dissertation with a summary of its main contributions, the main conclusions and the identification of critical future research directions.

Chapter 2: Reliability Assessment with the Fast Fourier Transform

2.1 INTRODUCTION

This focus of this chapter is to present a computationally efficient approximation scheme to obtain estimates of the PDF of the system travel time (which in turn can be used to derive the unreliability function). To this end, we examine the case of travel time unreliability induced by independent road capacity variations under non-adaptive travel behavior. This work can be seen as a complement to some of the results presented in Lo and Tung (2003) and Lo et al., 2006 in which the Central Limit Theorem (CLT) was invoked to guarantee normality of the system travel time (by assuming strictly positive capacities and the exclusion of pathological cases). However, in the more general case, it is not always feasible to rely on the CLT. In this chapter we derive the PDF of the system-wide travel time without relying on the CLT (Section 2.6 provides examples in which the CLT fails to hold, whereas the proposed methodology remains valid), which leads us to a different positioning of the current work: a complement to Clark and Watling (2005) who obtained the PDF of the travel time under stochastic demand. Apart from the source of uncertainty, the proposed methodology is tremendously different. Clark and Watling assumed the multivariate normal distribution based on which they were able to obtain analytical expressions for the moments of the random travel time. The moments were subsequently used to obtain the PDF of the system travel time by fitting a parametric family of PDFs known as Jonhson curves (Jonhson, 1949). The underlying principles of our approach are based on the theory of Fourier transforms. Furthermore, we obtain a numerical approximation of the PDF rather than an analytical expression. Finally, while we assume independence, the proposed methodology is able to accommodate virtually every continuous or discrete distribution (except for pathological cases) in the modeling of capacities. In this chapter we shall develop the framework assuming a continuous distribution; modifications to the methodology in the discrete case are readily made.

Following existing literature in this area (e.g., Chen et al., 2002; Lo and Tung, 2003), we assume that the link travel time t_a on link $a \in A$ is given by the Bureau of Public Roads (BPR) volume-delay function (Bureau of Public Roads, 1964):

$$t_a = t_a^f \left(1 + \alpha \left(\frac{v_a}{c_a} \right)^{\beta} \right). \tag{2.1}$$

Conditional on v_a, we assume that capacity is the single source of uncertainty in this chapter. The objective of this chapter is then to derive a numerical approximation of the PDF of system travel time T under the above stated assumptions.

The remainder of this chapter is organized as follows. In Section 2.2 we present a brief review of the theory of Fourier transforms, focusing only on the essential elements needed for the understanding of the work in this chapter. Section 2.3 presents the approximation scheme. Error bounds on the resulting errors are derived. From a more practical point of view, we present a successive refinement scheme in Section 2.4 to guarantee accurate approximations. In Section 2.5 we consider the special case when capacities are normally distributed random variables. Numerical results for this special case are reported in Section 2.6 based on the well-known Nguyen-Dupuis test network.

2.2 FOURIER TRANSFORMS

In this section we present a brief review of Fourier analysis, only focusing on the essential elements that aid the reader in the understanding of the work in this chapter. For more details and related topics, we refer the reader to the numerous comprehensive textbooks available (e.g., Goldberg, 1961; Briggs and Henson, 1995; Howell, 2001). Let us start by being more precise about the continuous Fourier transform (henceforth simply referred to as the Fourier transform). In the following, we use $\Re$ and $\mathbb{C}$ to denote the set of real and complex numbers, respectively.

Definition 2.1 Let *f*: $\Re \to \Re$ be a continuous and absolutely integrable function, i.e,

$$\int_{-\infty}^{\infty} |f(t)|\, dt < \infty \, .$$

The Fourier transform of *f* is then given by the complex-valued function $\hat{f} : \Re \to \mathbb{C}$

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} f(t) dt$$

where $\omega \in \Re$ and $i = \sqrt{-1}$. The function *f* can be recovered from its Fourier transform $\hat{f}$, provided that $\hat{f}$ is continuous and absolutely integrable, in which case we define the Inverse Fourier Transform as:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} \hat{f}(\omega) d\omega \, . \qquad \square$$

The domain on which f is defined is frequently referred to as the spatial or time domain, whereas we say that the Fourier transform $\hat{f}$ is defined on the frequency domain. We want to note that the above definition for the Fourier transform pair is not unique. Slight variations can be found in the literature (e.g., see Howell, 2001 for a discussion). However, they all share the same fundamental properties that make them useful in practice. In this dissertation we adopt Definition 2.1 as *the* Fourier transform.

This chapter is heavily built on the convolution theorem (Briggs and Henson, 1995). Let f and g be two functions defined in the time domain that satisfy the properties stated in Definition 2.1. The convolution of f and g (denoted as $f * g$) is the function h given by

$$h(x) = \int_{-\infty}^{\infty} f(s)g(x-s)ds\,. \tag{2.2}$$

The convolution theorem states that the Fourier transform of h equals the product of the Fourier transforms of f and g, i.e,

$$\hat{h}(\omega) = \hat{f}(\omega)\hat{g}(\omega)\,.$$

While f and g can be any arbitrary, "nice" functions (in the sense of Definition 2.1), in this chapter we are interested in the special case when they represent PDFs (in this case, the Fourier transforms are also known as characteristic functions, Ushakov (1999); however, we will employ the name Fourier transform here). Consequently, h can be interpreted as the PDF of the sum of two independent random variables with PDFs f and g (Feller, 1966). Theoretically, one can apply the inverse Fourier transform to recover h (that is, instead of the direct evaluation of the convolution integral (2.2) in the time domain, we have reduced the problem to one of multiplication in the frequency domain). Unfortunately, closed-form expressions for Fourier transforms are only available for relatively simple functional forms. Hence we need an approximation. This approximation is naturally given by the discrete Fourier transform (DFT), as we will demonstrate in Section 2.3. For the same reason as in the continuous case, a definition is in order here. Let y' denote the transpose of the vector y.

Definition 2.2 Let $z = (z_1, z_2, ..., z_N)'$ be an N-dimensional complex-valued vectors in $\mathbb{C}^N$. Furthermore, define $\vartheta_N = e^{-2\pi i/N}$, then the DFT of z is given by $Z = (Z_1, Z_2, ..., Z_N)' \in \mathbb{C}^N$ where

$$Z_k = \sum_{n=1}^{N} z_n \vartheta_N^{(n-1)(k-1)}$$

and k =1, 2,…, N. The vector z can be recovered from Z via the inverse discrete Fourier transform (IDFT):

$$z_n = \frac{1}{N} \sum_{k=1}^{N} Z_k \vartheta_N^{-(n-1)(k-1)}$$

where n =1, 2,…, N. □

Based on the matrix representations of the DFT and its inverse (Briggs and Henson, 1995), it is not hard to see that the number of arithmetic operations for each of the transforms is proportional to N^2, i.e,, the number of arithmetic operations is $O(N^2)$. Although polynomial, the DFT and its inverse would not have been as ubiquitous as it is today if there were not more efficient algorithms. This efficient class of algorithms is collectively known as the fast Fourier transform (FFT), which has a computational complexity of $O(N \log N)$. To give a hint of why this complexity can be achieved, we briefly consider one class of FFT algorithms known as the decimation-in-time algorithms (Van Loan, 1992). Without loss of generality, assume that N is some power of 2, i.e, $N = 2^L$ for some positive integer L (the fact that this assumption is always without loss of generality is due to a technique called zero-padding, e.g., see Briggs and Henson, 1995). Define $y_n \equiv z_{2n-1}$ and $w_n \equiv z_{2n}$, n = 1, 2,…, N/2 that are the odd and even subsequences of z_n, n =1, 2,…, N. The DFT of the vector z_n can now be re-expressed as (cf. Definition 2.2):

$$Z_k = \sum_{n=1}^{N/2} \left(y_n \vartheta_N^{(2n-2)(k-1)} + w_n \vartheta_N^{(2n-1)(k-1)} \right)$$

where k =1, 2,…, N. By observing that

$$\vartheta_N^{(2n-2)(k-1)} = \vartheta_{N/2}^{(n-1)(k-1)}$$

we can rewrite the above DFT as

$$Z_k = \sum_{n=1}^{N/2} y_n \vartheta_{N/2}^{(n-1)(k-1)} + \vartheta_N^{k-1} \sum_{n=1}^{N/2} w_n \vartheta_{N/2}^{(n-1)(k-1)} = Y_k + \vartheta_N^{k-1} W_k .$$

In particular,

$$\begin{aligned} Z_k &= Y_k + \vartheta_N^{k-1} W_k \\ Z_{k+N/2} &= Y_{k+N/2} + \vartheta_N^{k-1+N/2} W_{k+N/2} \end{aligned}$$

where k =1, 2,…, N/2. Since $Y_{k+N/2} = Y_k$, $W_{k+N/2} = W_k$ and $\vartheta_N^{N/2} = -1$, one gets the so-called butterfly relations

$$Z_k = Y_k + \vartheta_N^{k-1} W_k$$
$$Z_{k+N/2} = Y_k - \vartheta_N^{k-1} W_k$$

where k =1, 2,…, N/2. That is, we have shown that it is possible to express the DFT of the vector z_n in terms of the DFTs of its odd and even subsequences (each of length N/2). Clearly, we can apply this splitting-idea to the odd and even subsequences of y_n and w_n, until we have sequences of length one (whose DFTs are trivially computed: they are the sequences themselves). Using a series of butterfly relations, the DFT of the original sequence can then be recovered.

The discovery of the FFT is often attributed to Cooley and Tukey (1965), although the technical birth date of the algorithm can be traced back to 1805 when Gauss devised the FFT for his calculations in astronomy (Heideman et al., 1985). However, only in 1965, did the FFT become widespread. The importance of the FFT cannot be overstated (see Brigham, 1988, for a long list of applications). In this dissertation we have to confine ourselves to the limited view that the FFT is an efficient algorithm to evaluate the DFT (and IDFT). For more details on the FFT, we refer to the reader to Van Loan (1992).

2.3 The Approximation Scheme and its Error Bounds

Recall that the aim of this chapter is to obtain an approximation of the PDF of the system-wide travel time T, denoted as f_{TSTT}. Before we attempt to make any statements at the system level, let us present some results at the link level. Throughout this section we shall assume that the volume-delay function is given by the BPR function (2.1).

Lemma 2.1 Let the PDF of the capacity of link $a \in A$ be given by f_{ca}. Furthermore, let $x > v_a t_a^f$. Then the PDF of T_a is given by

$$f_a(x) = -f_{ca}(c^*)\frac{c^*}{\beta(v_a t_a^f - x)},$$

where

$$c^* = \frac{v_a}{\left(\frac{1}{\alpha}\left(\frac{x}{v_a t_a^f} - 1\right)\right)^{1/\beta}}.$$

Proof Since the BPR function is strictly decreasing as a function of the capacity of the link, we have

$$\Pr[T_a \le x] = \Pr[\tilde{c}_a \ge c^*] = 1 - \Pr[\tilde{c}_a \le c^*] = 1 - F_{ca}(c^*)$$

where c^* and F_{ca} are the solution to the equation $t_a v_a = x$ and the cumulative distribution function of $\tilde{c}_a$, respectively. The PDF of T_a can now be obtained by differentiating with respect to x:

$$\frac{d\Pr\left[T_a \le x\right]}{dx} = -f_{ca}(c^*)\frac{dc^*}{dx} = -f_{ca}(c^*)\frac{c^*}{\beta(v_a t_a^f - x)},$$

which completes the proof. **Q.E.D.**

From probability theory it is known that the PDF of a sum of independent random variables is given by the convolution of the PDFs involved. Using the current notation, this can be stated as

$$f_{TSTT}(x) = f_1 * f_2 * \ldots * f_A(x).$$

Note that if, in addition, the capacities are assumed to have identical distributions, the goal of this chapter is tremendously simplified as the CLT (Feller, 1966) ensures that the TSTT follows some normal distribution (assuming that a typical transportation network has a sufficiently large number of links). That is, the task would have reduced to the derivation of the mean and variance of the TSTT as they uniquely determine a normal distribution. In this chapter, random capacities are allowed to have non-identical probability distributions so that the CLT does not necessarily hold. Now, by the convolution theorem (cf. Section 2.2) we have

$$\hat{f}_{TSTT}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t}(f_1 * f_2 * ... * f_A)dt = \prod_a \hat{f}_a(\omega). \tag{2.3}$$

Theoretically, Fourier inversion can be invoked to recover f_{TSTT}. Practically, however, this might pose some serious challenges. The main difficulty in the application of the convolution theorem stems from the absence of closed-form expressions for the Fourier transforms of f_a. Note that although we have not assumed any specific probability distribution for the random capacities yet, we can anticipate that a closed-from expression for the Fourier transform is challenging, if not impossible, to obtain for general distributions as it involves complex integrations of non-trivial functions in the complex domain. Hence we have to resort to approximations.

The Fourier transform of f_a is given by

$$\hat{f}_a(\omega) = \int_{v_a t_a^f}^{\infty} e^{-i\omega x} f_a(x)dx.$$

In order to numerically approximate this integral, we need to truncate the interval of integration to a finite range. In particular, we replace the interval of integration with the finite interval $(v_a t_a^f, U]$. Proposition 2.1 provides an upper bound on the resulting truncation error. Note that in Proposition 2.1 we only require that f_a be continuous: the condition of being absolutely integrable follows directly from the fact that f_a is a PDF (cf. Definition 2.1).

Proposition 2.1 (Truncation Error) Let f_a be a continuous PDF. For the truncation error E_T, there holds

$$E_T \equiv \left| \int_U^\infty f_a(t) e^{-i\omega t} dt \right| \leq \Pr\left[\tilde{c}_a \leq \frac{v_a (\alpha v_a t_a^f)^{1/\beta}}{(U - v_a t_a^f)^{1/\beta}} \right].$$

Proof We have

$$\left| \int_U^\infty f_a(t) e^{-i\omega t} dt \right| \leq \int_U^\infty \left| f_a(t) e^{-i\omega t} \right| dt = \int_U^\infty f_a(t) dt = \Pr[T_a > U] = \Pr\left[\tilde{c}_a \leq \frac{v_a (\alpha v_a t_a^f)^{1/\beta}}{(U - v_a t_a^f)^{1/\beta}} \right].$$

Q.E.D.

Clearly, the truncation error can be reduced by increasing *U*. The truncated integral will next be approximated at a discrete set of points. To this end, let $\omega_k = (k-1)\Delta\omega, k = 1,2,...,N, \Delta\omega \geq 0$ and $x_n = v_a t_a^f + (n-1)\Delta x, n = 1,2,...,N, \Delta x \geq 0$. The trapezoidal rule for numerical integration (Burden and Faires, 2004) gives

$$\begin{aligned} \hat{f}_a(\omega_k) &\approx \int_{v_a t_a^f}^{U} e^{-i\omega_k x} f_a(x) dx \\ &\approx e^{-i\omega_k v_a t_a^f} \Delta x \left(\sum_{n=1}^{N} f_a(x_n) e^{-i(k-1)(n-1)\Delta\omega\Delta x} - \frac{1}{2}\left(f_a(x_1) e^{-i\omega_k v_a t_a^f} + f_a(x_N) e^{-i\omega_k v_a t_a^f} e^{-i(k-1)(N-1)\Delta\omega\Delta x} \right) \right). \end{aligned} \quad (2.4)$$

If we impose the relation:

$$\Delta\omega\Delta x = \frac{2\pi}{N} \tag{2.5}$$

one can recognize the DFT of the vector $\{f_a(x_n)\}_{n=1}^{N}$ in (2.4). In other words, the approximation (2.4) can be rapidly evaluated using the FFT algorithm! Relation (2.5) is known as the reciprocity relation: For a given number of sample points *N*, a finer grid in the spatial domain (i.e, a small value of $\Delta\omega$) necessitates a coarser grid in the frequency domain (i.e, a large value of Δx), and vice versa. Equivalently, for a given value of *N*, the length of the interval of integration in the spatial domain is inversely related to the length of the interval of integration in the frequency domain. We shall refer to the DFT approximation (2.4) as $\hat{f}_a^{DFT}(\omega_k)$. The following proposition characterizes the discretization error in the approximation.

Proposition 2.2 (Discretization Error) Let f_a be a twice continuously differentiable PDF. For the discretization error E_D in the DFT approximation (2.4) we have

$$E_D \equiv \left|\hat{f}_a(\omega_k) - \hat{f}_a^{DFT}(\omega_k)\right| \le \frac{U - v_a t_a^f}{12}(\Delta x)^2 \left|\max_x \frac{d^2\left[\cos(x) f_a(x)\right]}{dx^2} + \max_x \frac{d^2\left[\sin(x) f_a(x)\right]}{dx^2}\right|.$$

Proof Since we have simply used the trapezoidal rule in (2.4), the standard bound on the discretization error for the trapezoidal rule applies (Burden and Faires, 2004), provided that we make a slight modification. More precisely, we have to consider the errors in the real and imaginary components of the approximation separately. Let Re[*z*] and Im[*z*] denote the real and imaginary parts of the complex number $z=a+bi$, respectively (that is, $\mathrm{Re}[z]=a$ and $\mathrm{Im}[z]=b$). Then, for the error in the real part of the approximation we have

$$\mathrm{Re}[\hat{f}_a(\omega_k)]-\mathrm{Re}[\hat{f}_a^{DFT}(\omega_k)]=\frac{U-v_a t_a^f}{12}(\Delta x)^2\left.\frac{d^2\left[\cos(x)f_a(x)\right]}{dx^2}\right|_{x=\varsigma_{1a}} \tag{2.6}$$

for some $\varsigma_{1a}\in(v_a t_a^f,U]$. Likewise, for the imaginary part we have

$$\mathrm{Im}[\hat{f}_a(\omega_k)]-\mathrm{Im}[\hat{f}_a^{DFT}(\omega_k)]=\frac{U-v_a t_a^f}{12}(\Delta x)^2\left.\frac{d^2\left[\sin(x)f_a(x)\right]}{dx^2}\right|_{x=\varsigma_{2a}} \tag{2.7}$$

for some $\varsigma_{2a}\in(v_a t_a^f,U]$. By taking the sum of (2.6) and (2.7) and bounding the second derivatives by their maximum values we get the desired result. **Q.E.D.**

From the result in Proposition 2.2 we can make at least three important observations. First, there is a trade-off between the truncation and discretization error: The larger the truncation error, the smaller the discretization error, *ceteris paribus* (and vice versa). Second, in order to make exact statements about the discretization error, we need to bound some second order derivatives. Depending on the choice of the PDF, this might not be analytically tractable (of course, one can use numerical methods to obtain such bounds for each link). Third, the result tells us that the DFT approximation is an $O((\Delta \mathrm{x})^2)$ method. For instance, reducing the grid spacing Δx by ½ would reduce the discretization error by a factor of 4, again everything else being equal. Before we proceed, we want to note that the magnitudes of the derived error bounds are relative to the absolute value of the Fourier transform which is bounded by one:

$$\left|\hat{f}_a(\omega_k)\right| = \left|\int_{v_a t_a^f}^{\infty} f_a(t) e^{-i\omega_k t} dt\right| \leq \int_{v_a t_a^f}^{\infty} \left|f_a(t) e^{-i\omega_k t}\right| dt = \int_{v_a t_a^f}^{\infty} f_a(t) dt = 1 . \qquad (2.8)$$

Once $\hat{f}_a^{DFT}(\omega_k)$ have been evaluated for each link in the transportation network, point-wise multiplication is performed as a proxy for (2.3). As the terms $\hat{f}_a^{DFT}(\omega_k)$ are approximations themselves, the estimate of $\hat{f}_{TSTT}(\omega_k)$ will be subject to errors. To derive a bound on this estimate, note that we can write

$$\prod_{a=1}^{A} \hat{f}_a(\omega_k) - \prod_{a=1}^{A} \hat{f}_a^{DFT}(\omega_k)$$

$$= \left(\hat{f}_A(\omega_k) - \hat{f}_A^{DFT}(\omega_k)\right) \prod_{a=1}^{A-1} \hat{f}_a(\omega_k) + \hat{f}_A^{DFT}(\omega_k) \left(\prod_{a=1}^{A-1} \hat{f}_a(\omega_k) - \prod_{a=1}^{A-1} \hat{f}_a^{DFT}(\omega_k) \right).$$

By the triangle inequality, we get

$$\left| \prod_{a=1}^{A} \hat{f}_a(\omega_k) - \prod_{a=1}^{A} \hat{f}_a^{DFT}(\omega_k) \right|$$

$$\leq \left| \left(\hat{f}_A(\omega_k) - \hat{f}_A^{DFT}(\omega_k)\right) \prod_{a=1}^{A-1} \hat{f}_a(\omega_k) \right| + \left| \hat{f}_A^{DFT}(\omega_k) \left(\prod_{a=1}^{A-1} \hat{f}_a(\omega_k) - \prod_{a=1}^{A-1} \hat{f}_a^{DFT}(\omega_k) \right) \right|$$

so that

$$\left| \prod_{a=1}^{A} \hat{f}_a(\omega_k) - \prod_{a=1}^{A} \hat{f}_a^{DFT}(\omega_k) \right|$$

$$\leq \left| \hat{f}_A(\omega_k) - \hat{f}_A^{DFT}(\omega_k) \right| \left| \prod_{a=1}^{A-1} \hat{f}_a(\omega_k) \right| + \left| \hat{f}_A^{DFT}(\omega_k) \right| \left| \left(\prod_{a=1}^{A-1} \hat{f}_a(\omega_k) - \prod_{a=1}^{A-1} \hat{f}_a^{DFT}(\omega_k) \right) \right|.$$

Thus, if we assume that $\left|\hat{f}_a^{DFT}(\omega_k)\right| \leq 1$, we get the following result which bounds the "propagation error" in terms of the discretization error.

Proposition 2.3 (Error propagation by multiplication) Suppose that $\left|\hat{f}_a^{DFT}(\omega_k)\right| \le 1$, then

$$\left|\prod_{a=1}^{A}\hat{f}_a(\omega_k) - \prod_{a=1}^{A}\hat{f}_a^{DFT}(\omega_k)\right| \le \sum_{a=1}^{A}\left|\hat{f}_a(\omega_k) - \hat{f}_a^{DFT}(\omega_k)\right|. \tag{2.9}$$

Furthermore, if f_a is a twice continuously differentiable PDF, then

$$\sum_{a=1}^{A}\left|\hat{f}_a(\omega_k) - \hat{f}_a^{DFT}(\omega_k)\right| \le (\Delta x)^2 \sum_{a=1}^{A}\frac{(U - v_a t_a^f)}{12}\left|\max_x \frac{d^2\left[\cos(x) f_a(x)\right]}{dx^2} + \max_x \frac{d^2\left[\sin(x) f_a(x)\right]}{dx^2}\right| \tag{2.10}$$

Proof By assumption, we have $\left|\hat{f}_a^{DFT}(\omega_k)\right| \le 1$, so that

$$\left|\prod_{a=1}^{A}\hat{f}_a(\omega_k) - \prod_{a=1}^{A}\hat{f}_a^{DFT}(\omega_k)\right| \le \left|\hat{f}_A(\omega_k) - \hat{f}_A^{DFT}(\omega_k)\right| + \left|\left(\prod_{a=1}^{A-1}\hat{f}_a(\omega_k) - \prod_{a=1}^{A-1}\hat{f}_a^{DFT}(\omega_k)\right)\right|$$

where we have used (2.8) to deduce that $\left|\prod_{a=1}^{A-1}\hat{f}_a(\omega_k)\right| \le 1$. By an induction argument, we establish (2.9). Now, suppose that f_a is a twice continuously differentiable PDF, then inequality (2.10) follows directly from Proposition 2.2. **Q.E.D.**

As with the discretization error, we see that the "propagation error" reduces at a quadratic rate. The product $\prod_{a=1}^{A}\hat{f}_a^{DFT}(\omega_k)$ will next be deconvoluted via the IDFT. Before embarking on formulae, we want to comment on the way we have defined ω_k in (2.4). Recall that $\omega_k = (k-1)\Delta\omega, k = 1, 2, ..N$, so that we only obtain approximations of the Fourier transforms at non-negative frequencies ω_k. This might seem unnatural given our definition of the inverse Fourier transform in Section 2.2 (which requires that ω assumes both non-positive as well as non-negative values). Next we show that this is without loss of generality. To establish this, we need the following lemma (we want to note that Lemma 2.2 remains valid for *arbitrary* complex-valued functions; we have stated the lemma as is, given the focus of the current chapter).

Lemma 2.2 The Fourier transform $\hat{f}(\omega)$ satisfies the equality

$$e^{-i\omega x}\overline{\hat{f}(\omega)} = \overline{e^{i\omega x}\hat{f}(\omega)} \tag{2.11}$$

where $\overline{z}$ denotes the complex conjugate of z.

Proof Let us write $\hat{f}(\omega) = g(\omega) + ih(\omega)$, where $g(\omega)$ and $h(\omega)$ are real-valued functions. Using Euler's formula (Howell, 2001), the left hand side of (2.11) can then be rewritten as:

$$\begin{aligned} e^{-i\omega x}\overline{\hat{f}(\omega)} &= \left(\cos(\omega x) - i\sin(\omega x)\right)\left(g(\omega) - ih(\omega)\right) \\ &= g(\omega)\cos(\omega x) - h(\omega)\sin(\omega x) - i\left(g(\omega)\sin(\omega x) + h(\omega)\cos(\omega x)\right). \end{aligned}$$

Since

$$\begin{aligned} e^{i\omega x}\hat{f}(\omega) &= \left(\cos(\omega x) + i\sin(\omega x)\right)\left(g(\omega) + ih(\omega)\right) \\ &= g(\omega)\cos(\omega x) - h(\omega)\sin(\omega x) + i\left(g(\omega)\sin(\omega x) + h(\omega)\cos(\omega x)\right) \end{aligned}$$

the equality follows. **Q.E.D.**

Now, as

$$f_{TSTT}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\hat{f}_{TSTT}(\omega)d\omega = \frac{1}{2\pi}\int_{-\infty}^{0} e^{i\omega t}\hat{f}_{TSTT}(\omega)d\omega + \frac{1}{2\pi}\int_{0}^{\infty} e^{i\omega t}\hat{f}_{TSTT}(\omega)d\omega,$$

By the change of variable ω = -s, we get

$$\int_{-\infty}^{0} e^{i\omega t}\hat{f}_{TSTT}(\omega)d\omega = \int_{0}^{\infty} e^{-ist}\hat{f}_{TSTT}(-s)ds = \int_{0}^{\infty} e^{-ist}\overline{\hat{f}_{TSTT}(s)}ds = \int_{0}^{\infty}\overline{e^{ist}\hat{f}_{TSTT}(s)}ds = \overline{\int_{0}^{\infty} e^{ist}\hat{f}_{TSTT}(s)ds},$$

where the second equality follows from the fact that $\hat{g}(-\omega) = \overline{\hat{g}(\omega)}$ for real-valued functions g (e.g., see Howell, 2001). Lemma 2.2 yields the third equality; the last equality follows from the linearity property of complex integrals. Therefore, we have shown that

$$f_{TSTT}(t) = \frac{1}{\pi}\mathrm{Re}\left\{\int_{0}^{\infty} e^{i\omega t}\hat{f}_{TSTT}(\omega)d\omega\right\}.$$

Using this result, Fourier inversion can now be realized via the trapezoidal rule as follows. Define $t_0 \equiv \sum_a v_a t_a^f$ and let $t_n = t_0 + (n-1)\Delta x, n = 1,2,..,N$, then

$$f_{TSTT}(t_n)$$

$$\approx \frac{\Delta\omega}{\pi}\mathrm{Re}\left\{\sum_{k=1}^{N} e^{i\omega_k t_0}\hat{f}_{TSTT}(\omega_k)e^{i(k-1)(n-1)\Delta\omega\Delta x} - \frac{1}{2}\left(\hat{f}_{TSTT}(\omega_1) + \hat{f}_{TSTT}(\omega_N)e^{i\Delta\omega(N-1)t_0}e^{i(N-1)(n-1)\Delta\omega\Delta x}\right)\right\}$$

$$\approx \frac{\Delta\omega}{\pi}\mathrm{Re}\left\{\sum_{k=1}^{N} e^{i\omega_k t_0}\hat{f}_{TSTT}^{DTF}(\omega_k)e^{i(k-1)(n-1)\Delta\omega\Delta x} - \frac{1}{2}\left(\hat{f}_{TSTT}^{DTF}(\omega_1) + \hat{f}_{TSTT}^{DTF}(\omega_N)e^{i\Delta\omega(N-1)t_0}e^{i(N-1)(n-1)\Delta\omega\Delta x}\right)\right\}$$

$$\equiv \dot{f}_{TSTT}(t_n)$$

where we have defined $\hat{f}_{TSTT}^{DTF}(\omega_k) \equiv \prod_{a=1}^{A}\hat{f}_a^{DFT}(\omega_k)$. Assuming that the reciprocity relation (2.5) holds, one can recognize (up to a scaling factor of 1/N) the IDFT of the vector $\left\{\hat{f}_{TSTT}^{DTF}(\omega_k)e^{i\omega_k t_0}\right\}_{k=1}^{N}$ in the above approximation. Thus we can again use the FFT to rapidly evaluate the approximation. The next proposition characterizes the error involved in this inversion procedure. To simplify notation, we have assumed that the reciprocity relation (2.5) holds.

Proposition 2.4 (Inversion error) Let f_a be a twice continuously differentiable PDF. Define $f_{TSTT}^{RE}(t,\omega) \equiv \mathrm{Re}[e^{i\omega t}\hat{f}_{TSTT}(\omega)]$, $f_{TSTT}^{IM}(t,\omega) \equiv \mathrm{Im}[e^{i\omega t}\hat{f}_{TSTT}(\omega)]$ and assume that they are twice continuously differentiable (as a function of ω). Let $\Omega \equiv (N-1)\Delta\omega$, then

$$\left| f_{TSTT}(t_n) - \dot{f}_{TSTT}(t_n) \right| \le E_{trapezoidal} + E_{propagation} \tag{2.12}$$

where

$$E_{trapezoidal} = \frac{\Omega}{12}(\Delta\omega)^2 \left| \max_{\omega} \frac{d^2 f_{TSTT}^{RE}(t_n,\omega)}{d\omega^2} + \max_{\omega} \frac{d^2 f_{TSTT}^{IM}(t_n,\omega)}{d\omega^2} \right|$$

and

$$E_{propagation} = \Delta\omega(\Delta x)^2 \frac{(N+1)}{\pi} \sum_{a=1}^{A} \frac{(U - v_a t_a^f)}{12} \left| \max_x \frac{d^2\left[\cos(x) f_a(x)\right]}{dx^2} + \max_x \frac{d^2\left[\sin(x) f_a(x)\right]}{dx^2} \right|$$

$$\approx 2\Delta x \sum_{a=1}^{A} \frac{(U - v_a t_a^f)}{12} \left| \max_x \frac{d^2\left[\cos(x) f_a(x)\right]}{dx^2} + \max_x \frac{d^2\left[\sin(x) f_a(x)\right]}{dx^2} \right|.$$

Proof Define $\hat{f}_{TSTT}^{DFT}(\omega_k) = \hat{f}_{TSTT}(\omega_k) + \delta_{TSTT}(\omega_k)$, where $\delta_{TSTT}(\omega_k)$ is the error in the approximation of $\hat{f}_{TSTT}(\omega_k)$. Using this relation, we can write

$$f_{TSTT}(t_n) - \dot{f}_{TSTT}(t_n) = f_{TSTT}(t_n) - f_{trapezoidal}(t_n) - f_{error}(t_n)$$

where

$f_{trapezoidal}(t_n)$

$$= \frac{\Delta\omega}{\pi} \mathrm{Re}\left\{ \sum_{k=1}^{N} e^{i\omega_k t_0} \hat{f}_{TSTT}(\omega_k) \vartheta_N^{-(k-1)(n-1)} - \frac{1}{2}\left(\hat{f}_{TSTT}(\omega_1) + \hat{f}_{TSTT}(\omega_N) e^{i\Delta\omega(N-1)t_0} e^{i(N-1)(n-1)\Delta\omega\Delta x} \right) \right\}$$

and

$f_{error}(t_n)$

$$= \frac{\Delta\omega}{\pi} \mathrm{Re}\left\{ \sum_{k=1}^{N} e^{i\omega_k t_0} \delta_{TSTT}(\omega_k) \vartheta_N^{-(k-1)(n-1)} - \frac{1}{2}\left(\delta_{TSTT}(\omega_1) + \delta_{TSTT}(\omega_N) e^{i\Delta\omega(N-1)t_0} e^{i(N-1)(n-1)\Delta\omega\Delta x} \right) \right\}$$

That is, $f_{trapezoidal}(t_n)$ is the trapezoidal rule-based approximation of $f_{TSTT}(t_n)$ using the *exact* values of $\hat{f}_{TSTT}$ (which are unknown) and $f_{error}(t_n)$ is an error term that comes into existence as we use approximate values of $\hat{f}_{TSTT}$ in the inversion process. From the triangle inequality it now follows that

$$\left| f_{TSTT}(t_n) - \dot{f}_{TSTT}(t_n) \right| \le \left| f_{TSTT}(t_n) - f_{trapezoidal}(t_n) \right| + \left| f_{error}(t_n) \right| .$$

Analogous to Proposition 2.2, we have the bound:

$$\left| f_{TSTT}(t_n) - f_{trapezoidal}(t_n) \right| \le \frac{\Omega}{12} (\Delta\omega)^2 \left| \max_{\omega} \frac{d^2 f_{TSTT}^{RE}(t_n,\omega)}{d\omega^2} + \max_{\omega} \frac{d^2 f_{TSTT}^{IM}(t_n,\omega)}{d\omega^2} \right| .$$

For $f_{error}(t_n)$, we have

$$\left| f_{error}(t_n) \right| \le \frac{\Delta\omega}{\pi} \bullet$$
$$\left\{ \sum_{k=1}^{N} \left| \mathrm{Re}[e^{i\omega_k t_0} \delta_{TSTT}(\omega_k) \vartheta_N^{-(k-1)(n-1)}] \right| + \frac{1}{2} \left| \mathrm{Re}[\delta_{TSTT}(\omega_1) + \delta_{TSTT}(\omega_N) e^{i\Delta\omega(N-1)t_0} e^{i(N-1)(n-1)\Delta\omega\Delta x}] \right| \right\}$$
$$\le \frac{\Delta\omega}{\pi} \left\{ \sum_{k=1}^{N} \left| e^{i\omega_k t_0} \delta_{TSTT}(\omega_k) \vartheta_N^{-(k-1)(n-1)} \right| + \frac{1}{2} \left| \delta_{TSTT}(\omega_1) + \delta_{TSTT}(\omega_N) e^{i\Delta\omega(N-1)t_0} e^{i(N-1)(n-1)\Delta\omega\Delta x} \right| \right\}$$

$$\le \frac{\Delta\omega}{\pi} \left\{ \sum_{k=1}^{N} \left| \delta_{TSTT}(\omega_k) \right| + \frac{1}{2} \left| \delta_{TSTT}(\omega_1) \right| + \frac{1}{2} \left| \delta_{TSTT}(\omega_N) \right| \right\}$$

$$\le \frac{\Delta\omega}{\pi}\left\{\sum_{k=1}^{N}\sum_{a=1}^{A}\left|\hat{f}_a(\omega_k)-\hat{f}_a^{DFT}(\omega_k)\right|+\frac{1}{2}\sum_{a=1}^{A}\left|\hat{f}_a(\omega_1)-\hat{f}_a^{DFT}(\omega_1)\right|+\frac{1}{2}\sum_{a=1}^{A}\left|\hat{f}_a(\omega_N)-\hat{f}_a^{DFT}(\omega_N)\right|\right\}$$

$$\le \Delta\omega(\Delta x)^2\,\frac{(N+1)}{\pi}\sum_{a=1}^{A}\frac{(U-v_a t_a^f)}{12}\left|\max_x \frac{d^2\left[\cos(x) f_a(x)\right]}{dx^2}+\max_x \frac{d^2\left[\sin(x) f_a(x)\right]}{dx^2}\right|$$

where we have used Proposition 2.3 to arrive at the last two inequalities. Using the reciprocity relation and noting that $N >> 1$ completes the proof. **Q.E.D.**

Let us now take a step back and interpret the bound on the inversion error. First, note that the error bound involves the maximization of some functions consisting of the second order derivative of $\hat{f}_{TSTT}$ (cf. Propositions 2.2 and 2.3). Unfortunately, since $\hat{f}_{TSTT}$ is not known, we are not able to evaluate the right-hand side of (2.12). However, Proposition 2.4 does provide valuable insights into the approximation scheme. The first observation is that the error can be decomposed into two parts. The first part $E_{\text{trapezoidal}}$ is the discretization error introduced by the trapezoidal rule *as if we knew the exact values* of $\hat{f}_{TSTT}$. The size of this error component can be reduced by decreasing $\Delta\omega$. Furthermore, the proposition tells us that the bound improves at a quadratic rate as a function of $\Delta\omega$. The second part of the error bound $E_{\text{propagation}}$ is a result of the propagation of errors. That is, this error term is introduced because of the truncation (Proposition 2.1), discretization (Proposition 2.2) and multiplication (Proposition 2.3) of the Fourier transform. This error term decreases (approximately) at a linear rate as a function of Δx.

Although it is not possible to evaluate the right-hand side of (2.12), we believe that the bound in Proposition 2.4 (and in fact, all other bounds derived in this section) are too conservative. That is, the accuracy is often much higher than can be expected based on the bounds alone. This phenomenon was also observed in the authoritative work by Abate and Whitt (1992) who recognized that in practice it is often necessary to apply a successive refinement procedure to aid determining the accuracy of DFT approximations. In order words, an approximation is declared to be sufficiently accurate when subsequent refinements of the computational parameters do not yield noticeable differences in the approximated values. Before we present such a scheme in the next section, we have summarized the approximation procedure outlined in this section in Figure 2.1, where we have used $\dot{f}_{TSTT}$ to denote the resulting approximation of f_{TSTT}. The pseudo-code associated with the procedure is also provided below. Note that we have not explicitly included $\Delta\omega$ as an input parameter, the rationale being that once N and Δx are fixed, the reciprocity relation uniquely determines $\Delta\omega$. Furthermore, we want to emphasize that link flows v_a are part of the input to the procedure (cf. Lemma 2.1). These can, for example, be obtained via the solution of some traffic assignment problem, as is done in the numerical case study in Section 2.6.

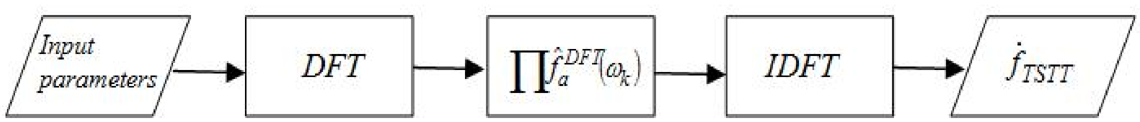

Figure 2.1: A summary of the Fourier-based approximation scheme.

Pseudo-code Fourier-based Approximation Scheme

INPUT $N, \Delta x, f_{ca}, \alpha, \beta, v_a, t_a^f$.

OUTPUT $\dot{f}_{TSTT}(t_n)$, $t_n = t_0 + (n-1)\Delta x, n = 1,2,..,N$, where $t_0 \equiv \sum_a v_a t_a^f$.

Step 1

Evaluate $\omega_k = (k-1)\Delta\omega, k = 1,2,..N$ where $\Delta\omega = 2\pi/N\Delta x$. For each link $a \in A$, use the FFT to evaluate (2.4) to obtain the approximations $\hat{f}_a^{DFT}(\omega_k),\ \omega_k = (k-1)\Delta\omega, k = 1,2,..N$.

Step 2

Evaluate $\hat{f}_{TSTT}^{DTF}(\omega_k) \equiv \prod_{a=1}^{A} \hat{f}_a^{DFT}(\omega_k)\ \omega_k = (k-1)\Delta\omega, k = 1,2,..N$.

Step 3

Use the FFT to evaluate

$$\dot{f}_{TSTT}(t_n)$$
$$\equiv \frac{\Delta\omega}{\pi} \mathrm{Re}\left\{ \sum_{k=1}^{N} e^{i\omega_k t_0} \hat{f}_{TSTT}^{DTF}(\omega_k) e^{i(k-1)(n-1)\Delta\omega\Delta x} - \frac{1}{2}\left(\hat{f}_{TSTT}^{DTF}(\omega_1) + \hat{f}_{TSTT}^{DTF}(\omega_N) e^{i\Delta\omega(N-1)t_0} e^{i(N-1)(n-1)\Delta\omega\Delta x} \right) \right\}$$

where $t_n = t_0 + (n-1)\Delta x, n = 1,2,..,N$.

As a final remark, we want to note that although we have presented our results assuming the BPR function, the methodology in this section is much more general and remains applicable as long as the volume-delay function yields a PDF for the link travel time (cf. Lemma 2.1) for which the Fourier transform is defined (cf. Definition 2.1). Of course, the specifics of the error analysis will change, depending on the volume-delay function chosen.

2.4 ACCURACY CHECKING VIA SUCCESSIVE REFINEMENT

The theoretical error bounds derived in the previous section provide valuable insights into the structure of the error. Moreover, they suggest ways to improve on the accuracy. However, as we have noted, they tend to be too conservative. Following Abate and Whitt (1992), in this section we provide a successive refinement scheme to empirically guarantee the accuracy of the approximations.

Given a transportation network, suppose that we want to assess whether the pair $(N, \Delta x)$ yields a sufficiently accurate approximation of f_{TSTT}. In the following, let us use $\dot{f}_{TSTT}(N, \Delta x)$ to denote the result of the approximation scheme depicted in Figure 2.1 with N and Δx being the input parameters. Intuitively (and based on Proposition 2.2), if we increase the number of sampling points N by a factor $K > 1$ and simultaneously decrease Δx by the same factor – leading to the new approximation $\dot{f}_{TSTT}(KN, \Delta x / K)$ – the discretization error should reduce. If this reduction in error is relatively small, it is reasonable to believe that $\dot{f}_{TSTT}(N, \Delta x)$ is an accurate approximation, *given the current interval of integration* (which has a length of $(N-1)\Delta x$). In this latter case, given Δx, the only way the approximation might be improved is by enlarging the interval of integration, while maintaining the same grid spacing, which leads us to evaluate $\dot{f}_{TSTT}(KN, \Delta x)$. To ensure that the grid spacing (in light of the discretization error) is sufficiently small on this larger interval, we compare $\dot{f}_{TSTT}(KN, \Delta x)$ and $\dot{f}_{TSTT}(K^2N, \Delta x / K)$ – note that the range of integration is the same in both cases. If these two approximations do not differ significantly *and* if the difference between $\dot{f}_{TSTT}(N, \Delta x)$ and $\dot{f}_{TSTT}(KN, \Delta x)$ is also small, we declare that the parameters $(N, \Delta x)$ yield a sufficiently accurate approximation of f_{TSTT}. Let us summarize this successive refinement accuracy checking procedure ($\varepsilon > 0$ denotes a pre-specified tolerance level):

<u>Pseudo-code Accuracy Checking via Successive Refinement</u>

INPUT $N, \Delta x, K, \varepsilon$ **.**

OUTPUT Decision whether $\dot{f}_{TSTT}(N, \Delta x)$ is an accurate approximation.

Step 1

Evaluate $\dot{f}_{TSTT}(N, \Delta x)$ and $\dot{f}_{TSTT}(KN, \Delta x / K)$.

IF $\max\left|\dot{f}_{TSTT}(KN, \Delta x / K) - \dot{f}_{TSTT}(N, \Delta x)\right| > \varepsilon \max\left|\dot{f}_{TSTT}(N, \Delta x)\right|$, then

STOP and OUTPUT: $\dot{f}_{TSTT}(N, \Delta x)$ is not a sufficiently accurate approximation.

ELSE GoTo Step 2.

Step 2

Evaluate $\dot{f}_{TSTT}(KN, \Delta x)$ and $\dot{f}_{TSTT}(K^2 N, \Delta x / K)$.

IF $\max\left|\dot{f}_{TSTT}(K^2 N, \Delta x / K) - \dot{f}_{TSTT}(KN, \Delta x)\right| > \varepsilon \max\left|\dot{f}_{TSTT}(KN, \Delta x)\right|$, then

STOP and OUTPUT: $\dot{f}_{TSTT}(N, \Delta x)$ is not a sufficiently accurate approximation.

ELSE GoTo Step 3.

Step 3

IF $\max\left|\dot{f}_{TSTT}(N,\Delta x)-\dot{f}_{TSTT}(KN,\Delta x)\right|<\varepsilon\max\left|\dot{f}_{TSTT}(N,\Delta x)\right|$, then

STOP and OUTPUT $\dot{f}_{TSTT}(N,\Delta x)$ is a sufficiently accurate approximation

ELSE STOP and OUTPUT: $\dot{f}_{TSTT}(N,\Delta x)$ is not a sufficiently accurate approximation.

Remarks

1. The proposed accuracy checking procedure requires the specification of *K*. For obvious reasons, *K* must be such that *KN* and K^2N are integers. However, note that the computational burden increases with *K* since we multiply the number of grid points by this value. Finally, *K* must be such that successive refined estimates have grid points in common (see next remark).
2. Note that successively refined estimates of the PDF cannot be compared at all grid points on which they are defined. For instance, in Step 1 we evaluate $\dot{f}_{TSTT}(N,\Delta x)$ and $\dot{f}_{TSTT}(KN,\Delta x/K)$. If *K*=5/4, the computational grids in the evaluation of $\dot{f}_{TSTT}(N,\Delta x)$ and $\dot{f}_{TSTT}(KN,\Delta x/K)$ are given by

$$\boldsymbol{0},\ \Delta x,\ 2\Delta x,\ 3\Delta x,\ \boldsymbol{4\Delta x},\ 5\Delta x,\ 6\Delta x,\ 7\Delta x,\ \boldsymbol{8\Delta x},\ldots,\ (N-1)\Delta x,$$

and

$$\boldsymbol{0},\ 4/5\Delta x,\ 8/5\Delta x,\ 12/5\Delta x,\ 16/5\Delta x,\ \boldsymbol{4\Delta x},\ 24/5\Delta x,\ 28/5\Delta x,$$
$$32/5\Delta x,\ 36/5\Delta x,\ \boldsymbol{8\Delta x},\ldots,\ 4/5(5/4N-1)\Delta x,$$

respectively. Hence, we can only compare $\dot{f}_{TSTT}(N,\Delta x)$ and $\dot{f}_{TSTT}(KN,\Delta x/K)$ at grid points that are a multiple of $4\Delta x$ and that are less or equal to $(N-1)\Delta x$. However, we do not believe that this is a serious issue as N tends to be large in practice, ensuring a sufficiently large number of common grid points. In light of this observation, the maxima between differences in the accuracy checking procedure are to be taken among *common* grid points.

3. There are several ways to obtain an initial guess of $(N, \Delta x)$. One possibility is by trial and error. Due to the computational efficiency of the FFT, one can guess a set of values and rapidly evaluate its performance by the visual inspection of the resulting PDF. When the PDF looks convincing (from experience, we can say that it is very easy to recognize inaccurate approximations), one can execute the above successive refinement scheme to check the accuracy more thoroughly.

2.5 The Special Case of Normally Distributed Capacities

Until now, we have not made any specific assumptions on the distribution of the random capacities. Except for pathological cases, *any* continuous or discrete probability distribution could have been used in the proposed methodology (with appropriate modifications in the discrete case, e.g., for discrete distributions we would start with the DFT instead of the continuous Fourier transform). In this and the next section we illustrate – without loss of generality – the proposed methodology based on the normal distribution, i.e, we assume that the capacity of link a has a normal distribution with mean μ_a and standard deviation σ_a, denoted as $\tilde{c}_a \sim N(\mu_a, \sigma_a)$, where μ_a can be taken as the practical (i.e, *not* the maximum) capacity of the link (Sheffi, 1985). Other probability distributions have been used in the literature. For example, Chen et al., (2002) and Lo and

Tung (2003) adopted the uniform distribution in order to illustrate their work (cf. Chapter 4).

Proposition 2.5 Let $\tilde{c}_a \sim N(\mu_a, \sigma_a)$, then

$$f_a(x) = \frac{v_a(\alpha v_a t_a^f)^{1/\beta}}{\sigma_a \sqrt{2\pi}\beta(x - v_a t_a^f)^{1+1/\beta}} \exp\left(-\frac{1}{2\sigma_a^2}\left[\frac{v_a(\alpha v_a t_a^f)^{1/\beta}}{(x - v_a t_a^f)^{1/\beta}} - \mu_a \right]^2 \right). \qquad (2.13)$$

Moreover, f_a achieves its maximum at

$$x^* \equiv v_a t_a^f + \frac{(2v_a)^\beta \alpha v_a t_a^f}{\left(\mu_a + \sqrt{\mu_a^2 + 4\sigma_a^2(1+\beta)}\right)^\beta}$$

and it is non-decreasing when $v_a t_a^f < x \le x^*$ and non-increasing when $x \ge x^*$.

Proof Since $\tilde{c}_a \sim N(\mu_a, \sigma_a)$, from Lemma 2.1 it follows that f_a is given by (2.13). To prove the second part of the proposition, note that

$$\frac{df_a(x)}{dx} = \frac{v_a^2 y_a^3 - \mu_a v_a y_a^2 - \sigma_a^2(1+\beta) y_a}{\sqrt{2\pi}\sigma_a^3 \beta^2 (v_a t_a^f - x)^2} \exp\left(\frac{-(v_a y_a - \mu_a)^2}{2\sigma_a^2} \right)$$

where

$$y_a = \left(\frac{1}{\alpha}\left(\frac{x}{v_a t_a^f} - 1 \right) \right)^{-1/\beta}.$$

A finite real-valued solution to the equation $df_a(x)/dx = 0$ can be obtained by solving the equation

$$y_a^2 v_a^2 - \mu_a v_a y_a - \sigma_a^2(1+\beta) = 0 \qquad (2.14)$$

as the exponential function does not contain roots on the real line. The solutions to (2.14) are given by

$$y_{+,-} = \frac{\mu_a \pm \sqrt{\mu_a^2 + 4\sigma_a^2\left(1+\beta\right)}}{2v_a}.$$

In terms of the x variable we have

$$x_{+,-} = v_a t_a^f + \frac{(2v_a)^\beta \alpha v_a t_a^f}{\left(\mu_a \pm \sqrt{\mu_a^2 + 4\sigma_a^2\left(1+\beta\right)}\right)^\beta}.$$

Without loss of generality, one can assume that $\sigma_a^2\left(1+\beta\right) > 0$, so that

$$\sqrt{\mu_a^2 + 4\sigma_a^2\left(1+\beta\right)} > \sqrt{\mu_a^2} = \mu_a.$$

Therefore, the only root in the range $x > v_a t_a^f$ is given by

$$x_+ \equiv v_a t_a^f + \frac{(2v_a)^\beta \alpha v_a t_a^f}{\left(\mu_a + \sqrt{\mu_a^2 + 4\sigma_a^2\left(1+\beta\right)}\right)^\beta}$$

Next we argue that the only stationary point x_+ must be a (global) maximum of f_a. To see this, note that

- f_a must be decreasing after some sufficiently large value of x as it is a PDF on the support $(v_a t_a^f, \infty]$, i.e, it has to integrate to one.
- We have $\lim_{x \downarrow v_a t_a^f} f_a(x) = 0$.

From these two observations it is clear that x_+ cannot be a point of inflection or a global minimum, as that would lead to contradictions of the above observations. Therefore, it must be a global maximum. **Q.E.D.**

From Proposition 2.5 we can see one important implication of a normally distributed capacity: the PDF (2.13) decreases very slowly, i.e, its right tail behaves like $1/x^{1+1/\beta}$ as $x \to \infty$. In other words, travel times can potentially be very high (with relatively large probabilities) on links whose capacities are normal random variables. In light of Proposition 2.1, we have

Proposition 2.6 Let $\tilde{c}_a \sim N(\mu_a, \sigma_a)$, then

$$E_T \le \Phi\left(\frac{1}{\sigma_a}\left(\frac{v_a(\alpha v_a t_a^f)^{1/\beta}}{(U - v_a t_a^f)^{1/\beta}} - \mu_a\right)\right).$$

Proof This follows directly from Proposition 2.1 and the fact that

$$\Pr[T_a > U] = \Phi\left(\frac{1}{\sigma_a}\left(\frac{v_a(\alpha v_a t_a^f)^{1/\beta}}{(U - v_a t_a^f)^{1/\beta}} - \mu_a\right)\right)$$

when $\tilde{c}_a \sim N(\mu_a, \sigma_a)$. **Q.E.D.**

From Proposition 2.2, we know that the bound on the discretization error in the DFT approximation is $O((\Delta x)^2)$. However, in order to calculate a specific value for this bound based on the normality assumption, we need to be able to maximize a function involving the second derivative of (2.13) which is not particularly attractive from an analytical point of view (cf. Proposition 2.2). Of course, one can resort to numerical techniques, but we won't pursue such a strategy here since we rely on the successive refinement procedure to guarantee the accuracy of the approximations. However, to illustrate the analytical approach, we derive an explicit bound that is of lower order, but that does not involve any derivatives at all.

Proposition 2.7 Let $\tilde{c}_a \sim N(\mu_a, \sigma_a)$, then

$$E_D \le 4\Delta x\left(2 f_a(x^*) - f_a(x_N)\right).$$

Proof As in Proposition 2.2, we first derive bounds on the discretization error for the real and imaginary parts separately. Consider the DFT approximation of $\mathrm{Re}[\hat{f}_a(\omega_k)]$ on the interval $[x_n, x_{n+1}]$, i.e,

$$\int_{x_n}^{x_{n+1}} f_a(t)\cos(\omega t)dt \approx \frac{f_a(x_n)\cos(x_n\omega) + f_a(x_{n+1})\cos(x_{n+1}\omega)}{2}\Delta x.$$

Let us focus on intervals $[x_n, x_{n+1}]$ for which $x_{n+1} \le x^*$ (Note that by choosing a sufficiently small value for Δx we can always ensure that the range $(v_a t_a^f, x^*]$ is covered by an integer number of non-overlapping intervals of the form $[x_n, x_{n+1}]$. In the following we shall assume that the choice of Δx is such that this is always the case). From Proposition 2.5, we know that f_a is non-decreasing when $x \le x^*$. Since $|\cos(\omega t)| \le 1$, the discretization error in the interval $[x_n, x_{n+1}]$ can be bounded by

$$2\Delta x\left(f_a(x_{n+1}) - f_a(x_n)\right).$$

Let $M = \max\{n \mid x_n \le x^*\}$, then we have for the total discretization error in the range $x \le x^*$:

$$2\Delta x\sum_{n=1}^{M-1}\left(f_a(x_{n+1}) - f_a(x_n)\right) = 2\Delta x\left(f_a(x_M) - f_a(x_1)\right) = 2\Delta x f_a(x^*). \qquad (2.15)$$

Let us now examine the intervals $[x_n, x_{n+1}]$ for which $x_n \geq x^*$. As f_a is non-increasing for $x \geq x^*$, the discretization error in the interval $[x_n, x_{n+1}]$ can now be bounded by

$$2\Delta x\left(f_a(x_n) - f_a(x_{n+1})\right),$$

so that the total discretization error in the range $x \geq x^*$ is bounded above by

$$2\Delta x\sum_{n=M}^{N-1}\left(f_a(x_n) - f_a(x_{n+1})\right) = 2\Delta x\left(f_a(x_M) - f_a(x_N)\right) = 2\Delta x\left(f_a(x^*) - f_a(x_N)\right). \qquad (2.16)$$

The sum of (2.15) and (2.16) yields an upper bound on the discretization error for the DFT approximation of $\mathrm{Re}[\hat{f}_a(\omega_k)]$. Using the same approach, one can easily show that the same bound applies to $\mathrm{Im}[\hat{f}_a(\omega_k)]$. Invoking the triangle inequality completes the proof. **Q.E.D.**

It is clear that analogs of Propositions 2.3 and 2.4 can also be derived based on the normality assumption. However, as we rely on the successive refinement scheme, we chose not to do so. Instead, let us examine some numerical results.

2.6 NUMERICAL DEMONSTRATION

In this section we present a numerical case study based on the Nguyen-Dupuis test network (Nguyen and Dupuis, 1984). This test network has been extensively used in the transportation literature, including for reliability studies (e.g., Lo and Tung, 2003). The results in this section are obtained using the FFT algorithm as implemented in MATLAB. For a more detailed discussion of the specifics of the algorithm, we refer the reader to the official MATLAB documentation.

Consider the network given in Figure 2.2. For our analysis, we will use the data for this network as reported in Lo and Tung (2003) with a number of slight modifications. For example, we randomly lowered the nominal link capacities and demand values as we prefer to work with smaller, manageable numbers. Table 2.1 summarizes the data for the analysis (recall that for link (i, j), i is the tail and j is the head of the link). The last column contains the randomly generated standard deviations of the link capacities. Following Lo and Tung, we set $\alpha = 1$ and $\beta = 2$ throughout this case study. For the OD demand, we have rescaled their values to 4 units (from node 1 to node 2 and from node 4 to 3) and 3 units (from node 1 to node 3 and from node 4 to 2). We shall henceforth refer to these values as the base values.

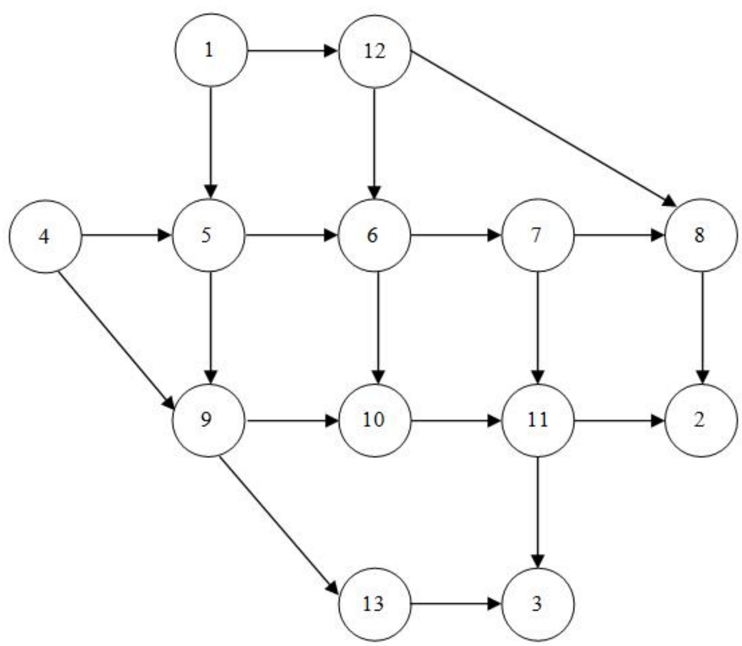

Figure 2.2: Nguyen-Dupuis test network.

Table 2.1: Data for the Nguyen-Dupuis network used in the current case study

Link	Tail	Head	Free flow travel time	Nominal capacity	Standard deviation
1	8	2	10	5	1
2	11	2	10	4	0.4
3	11	3	10	5	1.3
4	13	3	10	4	0.7
5	1	5	10	3	0.8
6	4	5	10	3	0.7
7	5	6	10	4	0.9
8	12	6	10	3	0.8
9	6	7	10	5	0.6
10	7	8	10	4	0.5
11	12	8	30	2	0.5
12	4	9	20	5	0.8
13	5	9	10	3	0.5
14	6	10	10	3	0.5
15	9	10	10	3	0.3
16	7	11	10	4	0.7
17	10	11	10	5	0.7
18	1	12	10	5	1.4
19	9	13	20	4	1.1

Based on a procedure of trial and error, we have set K=5/4 and $\varepsilon = 0.001$. All PDFs in this section were obtained with N=2^{16}, Δx=0.05 and $\Delta\omega$=0.0019, unless stated otherwise. We have checked the accuracy of the resulting approximations with the successive refinement procedure described in Section 2.4. As conjectured in the literature, theoretical error bounds tend to be too conservative in practice. To confirm this in our specific instance, we evaluated the truncation error (Proposition 2.6) and discretization error (Proposition 2.7) for the above combination of parameters that were found to be typically of order 10^{-4} and 10^{-3}, respectively (note that these values are absolute, rather than relative errors which makes them relatively large).

We first analyze the impact of demand on the travel time distribution. To this end, we assume that the change in demand values is such that they do not influence the nominal capacities and their standard deviations (see Lo and Tung, 2003 for a motivation of this assumption). Three demand levels are considered. For each demand level, we solve for a new deterministic user equilibrium (Wardrop, 1952) assuming the nominal capacities in Table 2.1 (we want to note that any other traffic assignment model could have been used; one could even simply use traffic detector data to obtain link volumes, if available). The resulting link flows are then used as input in the proposed Fourier-based methodology to derive the PDF of the system travel time. Figure 2.3 shows the PDFs for a range of different demand values – ranging from 0.75 times the base demand to 1.5 times as large. From the figure, it is clear that the increase in demand introduces more variance in the TSTT which makes higher travel times possible. Furthermore, note that the mode of the PDF is shifting to the right when demand is increased. We want to emphasize that in the above we have assumed non-adaptive behavior (to the unfolding scenario) since for each demand level we obtain a single set of link flows that is based on a single capacity value (i.e, the nominal capacity). Clark and Watling (2005) adopted the same approach in their numerical case study.

Our second experiment is to examine the effect of changes in the standard deviations of the capacities on the PDF of the travel time. Three levels of standard deviation are considered. We assume that the demand level is given by the "base demand" and that the capacities are the nominal capacities reported in Table 2.1. Using these values, we solve for a deterministic user equilibrium, which yields one set of link flows. For each level of standard deviation considered (using the *same* set of link flows for each level of standard deviation), we apply the proposed methodology to obtain a PDF. Figure 2.4 shows the results. As can be anticipated, the increase in standard deviations leads to more variation, rendering both smaller as well as larger travel times more likely (the values given in Table 2.1 serve as the base standard deviations, denoted as base SD in the figure). However, this change in travel time is not symmetric, in the sense that the right tail of the probability distribution becomes much heavier than the left tail. Note that, as opposed to Figure 2.3 (and Figure 2.5 below), in order to arrive at Figure 2.4, we have only solved for a *single* set of link flows to obtain the *three* different PDFs shown (in Figures 2.3 and 2.5 three sets of link flows were necessary). The reason for this is our implicit assumption that the set of link flows remains unchanged as a function of the standard deviations of road capacities, which is merely a consequence of the fact that we used the conventional deterministic user equilibrium assignment (where the standard deviation of capacity is not a model parameter) to obtain the link flows. Nonetheless, we believe that this assumption is appropriate to illustrate the methodology. Furthermore, note that we have still assumed non-adaptivity: to obtain each of the PDFs in Figure 2.4, we only rely on one scenario of the capacities (the nominal capacities).

Finally, we investigate the impact of a change in the nominal capacities. Here we assume that changes in the nominal capacity do not affect the magnitudes of the standard deviations. Three levels of nominal capacities are considered. For each of the three levels of nominal capacities considered, we solve for a new deterministic user equilibrium using the base demand values and the capacities' standard deviations reported in Table 2.1, resulting in three different sets of link flows (note that we have again assumed non adaptive travel behavior since each set of link flows is obtained based on a single realization of the capacities). These three sets of link flows are then used as input to obtain the three depicted PDFs in Figure 2.5. From the BPR function in (2.1) it is clear that when the volume-to-capacity ratio increases, the travel time increases as well. Moreover, it increases more rapidly when this ratio exceeds unity. This observation can be recognized in Figure 2.5: When the nominal capacity is at 75% of their base values (the maximum volume-to-capacity ratio in this case was found to be 1.3), uncertainty can lead to relatively extreme realizations of the travel time (note the PDF displays a lot of variance). Reducing the capacities to their nominal values (the maximum volume-to-capacity ratio in this case was found to be 1) decreases the variance of the travel time, so that a smaller range of travel times are possible. The further reduction in variance yields the further concentration of the probability mass around a smaller range of values.

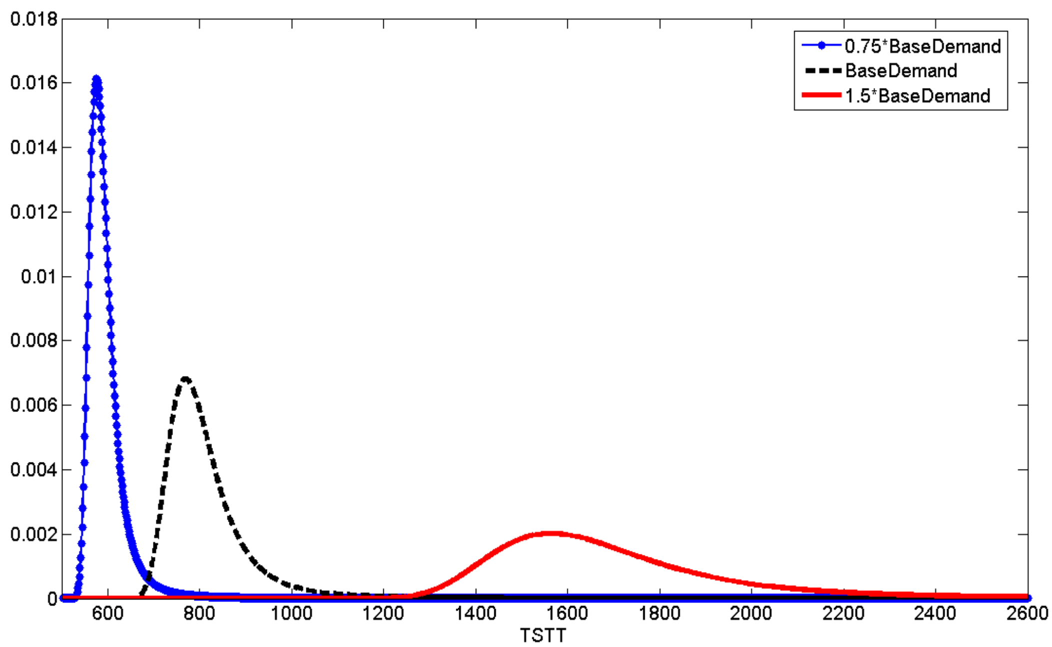

Figure 2.3: Impact demand on travel time reliability.

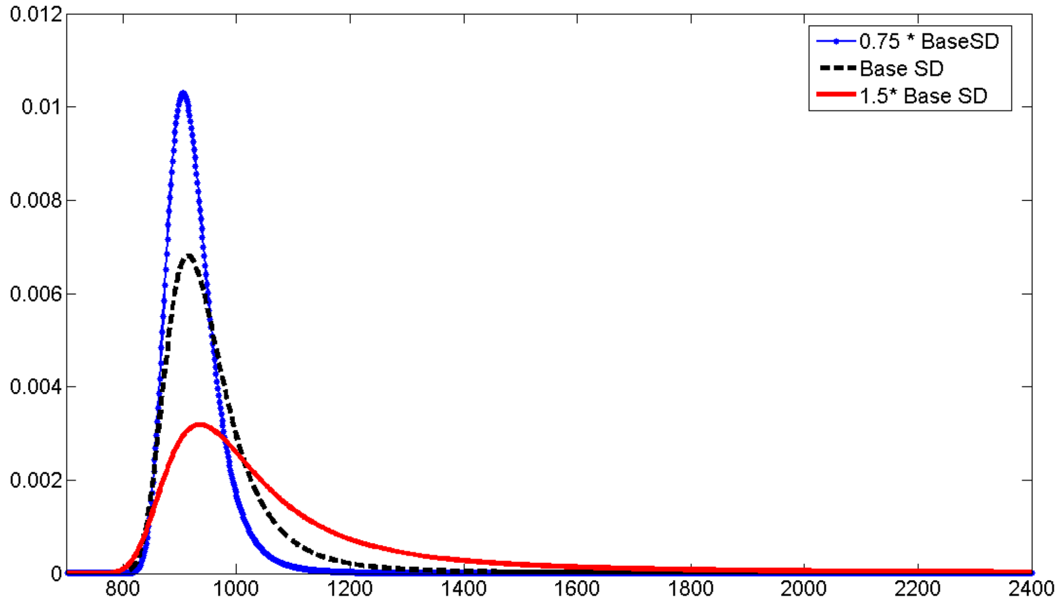

Figure 2.4: Impact of capacities' standard deviations on travel time reliability, assuming base demand values.

The computational time in the above numerical experiments was quite modest: Given N=2^{16}, Δx=0.05 and $\Delta\omega$=0.0019, it took less than 3 seconds to construct the PDF on a Linux machine with an Intel® 3.00GHz Xeon™ CPU and 32 GB of memory (of course, three more successively refined runs were needed to declare this set of parameters to be sufficiently accurate. The longest computational time among these runs was less than 5 seconds). In another computational test, we used the Sioux-Falls network (e.g., see Lam et al., 2008) that consists of 76 links. With $N = 2^{17}$, $\Delta x = 0.001$ and $\Delta \omega = 0.047$ we recorded a computational time of about 21 seconds (the longest computation in the accuracy checking procedure was recorded to be about 34 seconds, Figure 2.6).

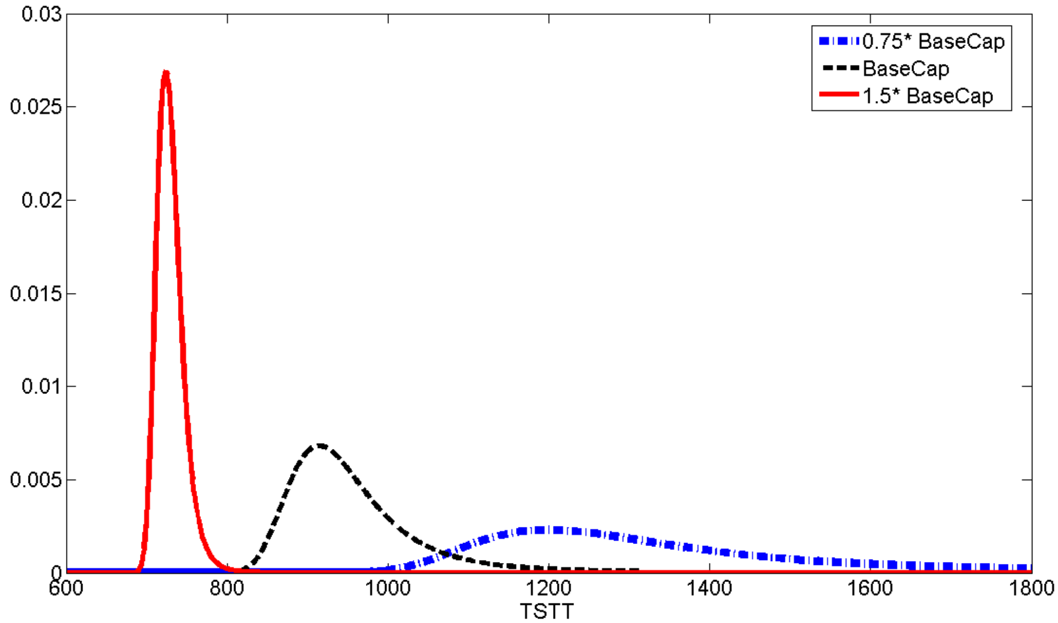

Figure 2.5: Impact of capacities on travel time reliability, assuming base demand values.

Despite this computational efficiency, it is always good practice to explore opportunities for further improvement, especially when analyzing large networks that consists of thousands of links. Before we can suggest any such strategies, let us focus on potential difficulties in the proposed methodology. One such difficulty is potentially presented by travel times with small variances (e.g., due to capacities with small variations or small volume-to-capacity ratios) that manifest themselves as PDFs that are "peaked" (i.e, a large fraction of the probability mass is concentrated in a small range of travel times, e.g., see Figure 2.5). Such peaked PDFs are difficult to integrate numerically, often requiring an overly fine grid. In fact, certain PDFs in the above figures could have been obtained using a finer grid (for example, the PDF for the case of 1.5 times the base demand in Figure 2.3 could have been obtained using $N = 2^{14}$, $\Delta x = 0.2$), but for ease of presentation, we have used a larger N so that all PDFs could be constructed using a common set of parameters. One possible remedy to deal with such links is to treat them as being deterministic and simply discard them in the reliability analysis. In order to obtain f_{TSTT} including the discarded links, one can always right-shift the resulting PDF by the amount of discarded "deterministic travel time". The same trick can be applied to links with relatively low traffic volumes that are not expected to contribute much to the overall travel time. This will be especially the case when the variance of the capacity is small too. An alternative of discarding the links would be to use different grids for each of the links in the network: Links with small variances are analyzed with a finer grid, whereas links characterized by a higher uncertainty are analyzed with a coarser grid. However, this might not be always feasible because of the reciprocity relation. To see this, from the reciprocity relation (2.5) we know that once N and Δx are fixed, $\Delta\omega$ and the range of integration in the frequency domain are fixed as well. Since the DFTs are to be multiplied point-wise for a set of common frequency values ω_k (cf. Figure 2.1), it is not always possible to choose a different set of N and Δx for each link such that the resulting set of ω_k values coincide (or that a sufficiently large set of common ω_k values exists). Much of the above complications can be attributed to the requirement of the FFT for uniform grid spacings. That is, Δx and $\Delta\omega$ are required to be the same throughout the intervals of integration (whereas, ideally, we would like a finer grid in a range where the integrand fluctuates more heavily and a coarser grid when the integrand does not change that much). This leads us to the idea to use the non-uniform DFT (Dutt and Rokhlin, 1993) which might be more attractive when there is a very diverse set of PDFs. Another strategy would be to use higher order quadrature rules (e.g., the Simpsons rule, Press et al., 2007) instead of the trapezoidal rule. The number of gird points could potentially be further reduced.

As a final remark, from the figures in this section it is clear that for independent *but not identically distributed* capacities, the CLT does not apply as the resulting PDFs tend to be skewed to the right. That this is not due to the relatively small number of links in the Nguyen-Dupuis network is corroborated by the Sioux-Falls network. Figure 2.6 shows the resulting PDF for the system travel time based on the network data reported in Suwansirikul et al. (1987) – we have doubled the reported capacities in order to get volume-to-capacity ratios closer to one – and a set of randomly generated standard deviations. As before, the PDF is asymmetric, with a fat right tail, contradicting a normal distribution.

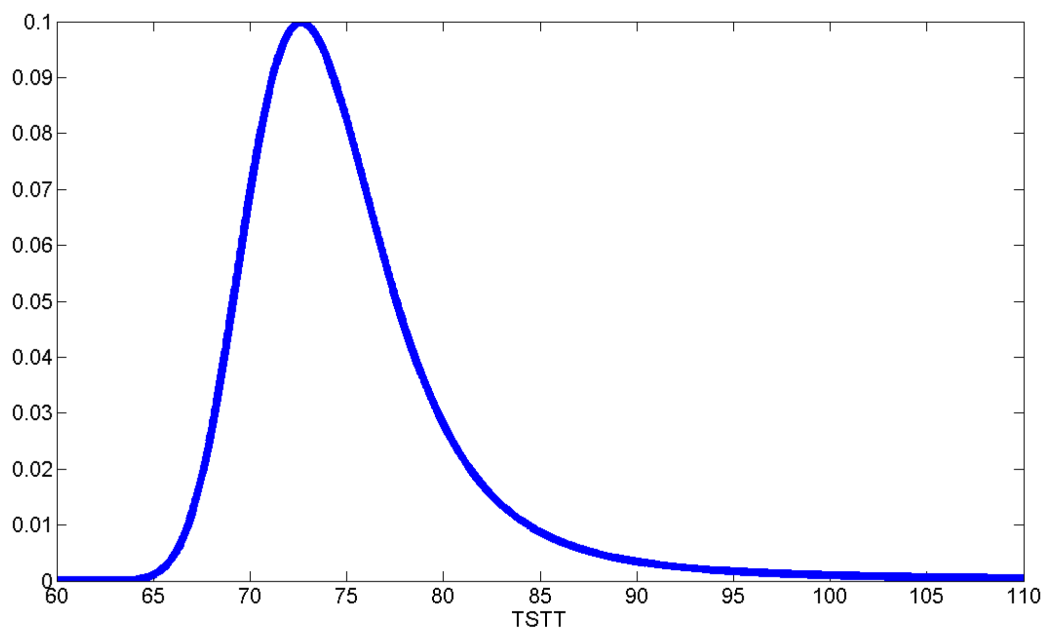

Figure 2.6: PDF for TSTT in the Sioux-Falls network. Clearly, the PDF is skewed to the right.

Chapter 3: Distribution-free Reliability Assessment

3.1 INTRODUCTION

An assumption that pervades the current transportation system reliability assessment literature is that probability distributions of the sources of uncertainty are known explicitly. However, this distribution may be unavailable (inaccurate) in reality as we may have no (insufficient) data to calibrate the distribution. In this chapter we relax this assumption and present a new method to assess travel time reliability that is distribution-free in the sense that the methodology only requires that the first N moments (where N is any positive integer) of the travel time to be known and that the travel times reside in a set of bounded and known intervals. Instead of deriving exact probabilities on travel times exceeding certain thresholds via computationally intensive methods, we develop analytical probability inequalities to quickly obtain upper bounds on the desired probability.

Following recent work in this field (such as Lo and Tung, 2003; Lo et al., 2006), we assume that T_a are independent random variables. However, as opposed to the above cited work – where the source of uncertainty was restricted to road capacities – we do not impose any assumptions on the underlying source of uncertainty, provided that the uncertainties exhibit the property of statistical independence. Of course, for concreteness, the reader is free to think of the random travel times as results of independent capacity variations due to, for example, minor traffic accidents, Lo and Tung (2003). However, at the same time, the reader should bear in mind that the results in this chapter are much more general. Furthermore, notice that the results do make explicit reference to any volume-delay function. In this chapter we derive nontrivial upper bounds on the unreliability function for a range of interesting values.

The remainder of this chapter is organized as follows. In Section 3.2 we collect some inequalities that are used repeatedly in this work. Bounds using first order moments only are formally derived in Section 3.3. In Section 3.4 we develop a new class of bounds that can potentially include moments up to order *N*. Some convexity results with regard to the bounds are discussed in Section 3.5. A numerical case study using the Sioux Falls test network is conducted in Section 3.6.

3.2 SOME USEFUL INEQUALITIES

In this section we collect a number of inequalities that will be used repeatedly in Sections 3.3 and 3.4 to arrive at the main results of this work.

Lemma 3.1 (Chernoff, 1952) Let $T_1, T_2, \ldots, T_A$ be non-negative random variables and let $T = \sum_{a=1}^{A} T_a$. Then for $\lambda > 0$ we have

$$\Pr\left(T > t\right) \le \inf_{\lambda>0}\left[\exp(-\lambda t)E\left[\exp(\lambda \textstyle\sum_{i=1}^{A} T_i)\right]\right]$$

where $E[X]$ denotes the mathematical expectation of the random variable *X*.

Proof Since $T \ge 0$ with probability one and $\lambda > 0$, we have

$$\Pr\left(T > t\right) = \Pr\left(\exp(\lambda T) > \exp(\lambda t)\right) \le \exp(-\lambda t)E\left[\exp(\lambda T)\right] = \exp(-\lambda t)E\left[\exp(\lambda \textstyle\sum_{i=1}^{A} T_i)\right] \tag{3.1}$$

where we have applied Markov inequality (Ross, 2002) to obtain the inequality. As $\lambda > 0$ was arbitrary, the tightest possible bound is obtained by taking the infimum. **Q.E.D.**

We want to emphasize that due to the minimization operation, the upper bound in Lemma 3.1 is always at least as tight as when the conventional Markov inequality is employed (which corresponds to the special case of $\lambda = 1$).

Another inequality that will be of fundamental importance in the current chapter is due to Madansky (1959).

Lemma 3.2 (Madansky, 1959) Let *f*(.) be a convex function and let *Y* denote a random variable with bounded support, i.e, there exists real numbers a and b such that $a \le Y \le b$ with probability one. Then

$$E\left[f(Y)\right] \le \frac{b - E[Y]}{b-a} f(a) + \frac{E[Y]-a}{b-a} f(b).$$

The inequalities in Lemmas 3.1 and 3.2 consist of first order moments only. In the numerical case study in Section 3.6, we will see that the inclusion of higher order moments can substantially improve the bounds. The next lemma is invaluable in the development of these higher order bounds.

Lemma 3.3 Let $\varsigma > 0$ and suppose that X is a non-negative random variable that is bounded from above, i.e, there exists a real number *b* such that $X \le b$ with probability one. Then

$$E\left[\exp(\varsigma X)\right] \le \sum_{k=0}^{N-1} \frac{\varsigma^k E[X^k]}{k!} + \frac{E[X^N]}{b^N}\left(\exp(\varsigma b) - \sum_{k=0}^{N-1} \frac{\varsigma^k b^k}{k!}\right).$$

Proof Using the Taylor series expansion of the exponential function and the linearity property of the mathematical expectation, we have

$$E\left[\exp(\varsigma X)\right] = E\left[\sum_{k=0}^{\infty} \frac{(\varsigma X)^k}{k!}\right]$$

$$= \sum_{k=0}^{\infty} \frac{E[(\varsigma X)^k]}{k!} = \sum_{k=0}^{N-1} \frac{E[(\varsigma X)^k]}{k!} + \sum_{k=N}^{\infty} \frac{\varsigma^k E[X^N X^{k-N}]}{k!}.$$

Since $X \le b$ with probability one, we have $X^N X^{k-N} \le X^N b^{k-N}$ with probability one. Hence,

$$E[X^N X^{k-N}] \le E[X^N b^{k-N}] = b^{k-N} E[X^N]$$

so that

$$\sum_{k=N}^{\infty} \frac{\varsigma^k E[X^N X^{k-N}]}{k!} \le \sum_{k=N}^{\infty} \frac{\varsigma^k b^{k-N} E[X^N]}{k!}$$

$$= E[X^N] \sum_{k=N}^{\infty} \frac{\varsigma^k b^{k-N}}{k!} = \frac{E[X^N]}{b^N}\left(\exp(\varsigma b) - \sum_{k=0}^{N-1} \frac{\varsigma^k b^k}{k!}\right)$$

which completes the proof. **Q.E.D.**

Two immediately corollaries are:

Corollary 3.1 $(N = 1)$ Let $\varsigma > 0$ and suppose that X is a non-negative random variable that is bounded from above, i.e, there exists a real number b such that $X \le b$ with probability one. Then

$$E\left[\exp(\varsigma X)\right] \le 1 + \frac{E[X]}{b}\left(\exp(\varsigma b) - 1\right).$$

Corollary 3.2 $(N = 2)$ Let $\varsigma > 0$ and suppose that X is a non-negative random variable that is bounded from above, i.e, there exists a real number b such that $X \leq b$ with probability one. Then

$$E\left[\exp(\varsigma X)\right] \leq 1 + \varsigma E[X] + \frac{E[X^2]}{b^2}\left(\exp(\varsigma b) - 1 - \varsigma b\right).$$

3.3 An Upper Bound using First Order Moments

In this section we derive an upper bound on the unreliability function of T. Contrary to existing research in this area (e.g., Chen et al., 2002; Sumalee and Watling, 2008), we do not require that the probability distributions are known. We only assume that the mean travel times are given and that the travel times lie within known and finite intervals. The main result of this section is summarized in Proposition 3.1. Its proof is inspired by Hoeffding (1963).

Proposition 3.1 (Upper bound using first order moment) Suppose that $t_{al} \leq T_a \leq t_{au}$ with probability one. Moreover, assume that T_a are independent, then

$$\Pr\left(T > t\right) \leq \inf_{\lambda > 0}\left[\exp(-\lambda t)\left(\frac{1}{A}\sum_{a=1}^{A}\left(\frac{E[T_a](\exp(\lambda t_{au}) - \exp(\lambda t_{al}))}{t_{au} - t_{al}} + \frac{t_{au}\exp(\lambda t_{al})}{t_{au} - t_{al}} - \frac{t_{al}\exp(\lambda t_{au})}{t_{au} - t_{al}}\right)\right)^{A}\right].$$

Proof By independence and Madansky inequality, (3.1) reduces to

$$\Pr\left(T > t\right) \leq \exp(-\lambda t)\prod_{i=1}^{A} E\left[\exp(\lambda T_i)\right]$$

$$\leq \exp(-\lambda t)\prod_{a=1}^{A}\left(\frac{t_{au} - E[T_a]}{t_{au} - t_{al}}\exp(\lambda t_{al}) + \frac{E[T_a] - t_{al}}{t_{au} - t_{al}}\exp(\lambda t_{au})\right).$$

Applying the arithmetic mean- geometric mean (AM-GM) inequality we have

$$\left(\prod_{a=1}^{A}\left(\frac{t_{au}-E[T_a]}{t_{au}-t_{al}}\exp(\lambda t_{al})+\frac{E[T_a]-t_{al}}{t_{au}-t_{al}}\exp(\lambda t_{au})\right)\right)^{1/A}$$

$$\leq \frac{1}{A}\sum_{a=1}^{A}\left(\frac{t_{au}-E[T_a]}{t_{au}-t_{al}}\exp(\lambda t_{al})+\frac{E[T_a]-t_{al}}{t_{au}-t_{al}}\exp(\lambda t_{au})\right)$$

$$=\frac{1}{A}\sum_{a=1}^{A}\left(\frac{E[T_a](\exp(\lambda t_{au})-\exp(\lambda t_{al}))}{t_{au}-t_{al}}+\frac{t_{au}\exp(\lambda t_{al})}{t_{au}-t_{al}}-\frac{t_{al}\exp(\lambda t_{au})}{t_{au}-t_{al}}\right).$$

Hence

$$\Pr\left(T>t\right)\leq \exp(-\lambda t)\left(\frac{1}{A}\sum_{a=1}^{A}\left(\frac{E[T_a](\exp(\lambda t_{au})-\exp(\lambda t_{al}))}{t_{au}-t_{al}}+\frac{t_{au}\exp(\lambda t_{al})}{t_{au}-t_{al}}-\frac{t_{al}\exp(\lambda t_{au})}{t_{au}-t_{al}}\right)\right)^{A} \tag{3.2}$$

Since $\lambda > 0$ was arbitrary, the result follows. **Q.E.D.**

We conclude this section with two remarks. First, one has no guarantee that the above derived bound (and the bounds in the next section) always assumes values in the natural interval [0,1]. However, to follow standard practice – for instance, nothing guarantees that Markov inequality (Ross, 2002) yields bounds valued between 0 and 1 (it often does not) – we have chosen to state the bounds as above. In numerical studies, one can always adjust the bounds to ensure that they lie in the natural interval (see Section 3.6). Second, one could further bound the right-hand side of (3.2), for instance, by using the BPR-function. However, the resulting bounds will become less tight, which was part of our motivation to work with travel times directly.

3.4 BOUNDS USING THE FIRST N MOMENTS

In the previous section, we have derived bounds that were based on first order moments only. As such, they might not be as tight as one wants in practice. In this section we develop a more general framework to develop bounds that potentially can incorporate moments of order up to N, provided that they exist. A careful inspection of the proof of Proposition 3.1 reveals that the fundamental reason that higher order moments cannot be incorporated into the bounds developed in Section 3.3 is that Madansky inequality (cf. Lemma 3.2) do not account for these higher order moments. Hence, a fundamentally different approach is needed. Lemma 3.3 forms the foundation of this new approach. For notational convenience, but also from a practical perspective (where N cannot be an arbitrarily large number), in this section we present results for the cases $N \le 2$ only. Of course, when there is a reason to believe that on top of the mean and variance of travel times, higher order moments are known as well, Lemma 3.3 can be readily used to extend the results presented in this section.

In order to illustrate the importance of the inclusion of higher order moments, we start with an alternative upper bound that only relies on mean values (cf. Proposition 3.1).

Proposition 3.2 (Alternative upper bound using first order moment) Suppose that $T_a \le t_{au}$ with probability one. Moreover, assume that T_a are independent, then

$$\Pr\left(T > t\right) \le \inf_{\lambda > 0}\left[\exp(-\lambda t)\left(\frac{1}{A}\sum_{a=1}^{A}\left(1+\frac{E[T_a]}{t_{au}}\left(\exp(\lambda t_{au})-1\right)\right)\right)^{A}\right].$$

Proof By independence and Corollary 3.1, (3.1) reduces to

$$\Pr\left(T>t\right)\le \exp(-\lambda t)\prod_{i=1}^{A} E\left[\exp(\lambda T_i)\right]\le \exp(-\lambda t)\prod_{a=1}^{A}\left(1+\frac{E[T_a]}{t_{au}}\left(\exp(\lambda t_{au})-1\right)\right).$$

As in the proof of Proposition 3.1, the AM-GM inequality yields

$$\left(\prod_{a=1}^{A}\left(1+\frac{E[T_a]}{t_{au}}\left(\exp(\lambda t_{au})-1\right)\right)\right)^{1/A}\le \frac{1}{A}\sum_{a=1}^{A}\left(1+\frac{E[T_a]}{t_{au}}\left(\exp(\lambda t_{au})-1\right)\right).$$

Hence

$$\Pr\left(T>t\right)\le \exp(-\lambda t)\left(\frac{1}{A}\sum_{a=1}^{A}\left(1+\frac{E[T_a]}{t_{au}}\left(\exp(\lambda t_{au})-1\right)\right)\right)^{A}.$$

Since $\lambda>0$ was arbitrary, the result follows. **Q.E.D.**

The inclusion of variance information yields Proposition 3.3.

Proposition 3.3 (Upper bound using second order moment) Suppose that $T_a \le t_{au}$ with probability one. Moreover, assume that T_a are independent, then

$$\Pr\left(T>t\right)\le \inf_{\lambda>0}\left[\exp(-\lambda t)\left(\frac{1}{A}\sum_{a=1}^{A}\left(1+\lambda E[T_a]+\frac{E[T_a^2]}{t_{au}^2}\left(\exp(\lambda t_{au})-1-\lambda t_{au}\right)\right)\right)^{A}\right].$$

Proof By independence and Corollary 3.2, (3.1) reduces to

$$\Pr\left(T>t\right)\le \exp(-\lambda t)\prod_{i=1}^{A} E\left[\exp(\lambda T_i)\right]$$

$$\le \exp(-\lambda t)\prod_{a=1}^{A}\left(1+\lambda E[T_a]+\frac{E[T_a^2]}{t_{au}^2}\left(\exp(\lambda t_{au})-1-\lambda t_{au}\right)\right).$$

As in the proof of Proposition 3.1, the AM-GM inequality yields

$$\left(\prod_{a=1}^{A}\left(1+\lambda E[T_a]+\frac{E[T_a^2]}{t_{au}^2}\left(\exp(\lambda t_{au})-1-\lambda t_{au}\right)\right)\right)^{1/A}$$

$$\leq \frac{1}{A}\sum_{a=1}^{A}\left(1+\lambda E[T_a]+\frac{E[T_a^2]}{t_{au}^2}\left(\exp(\lambda t_{au})-1-\lambda t_{au}\right)\right).$$

Hence

$$\Pr\left(T>t\right)\leq \exp(-\lambda t)\left(\frac{1}{A}\sum_{a=1}^{A}\left(1+\lambda E[T_a]+\frac{E[T_a^2]}{t_{au}^2}\left(\exp(\lambda t_{au})-1-\lambda t_{au}\right)\right)\right)^{A}.$$

Since $\lambda > 0$ was arbitrary, the result follows. **Q.E.D.**

Three remarks are in order. First, a sufficient condition for the Central Limit Theorem (CLT) to hold, i.e, for T to converge in distribution to some normal random variable – note that in such a case the development of bounds on the unreliability function would be simplified – is that T_a are bounded and $\sum_a Var[T_a]\to\infty$ (Billingsley, 1995). Here we have used Var[*X*] to denote the variance of the random variable *X*. The former assumption is by definition satisfied in the current chapter; the assumption that $\sum_a Var[T_a]\to\infty$ is a much stronger assumption that we do not impose here (cf. Lo et al., 2006). Consequently, our results hold in much more generality. Second, note that the bound in Proposition 3.1 is an explicit function of the left end point t_{al} of the support of T_a, in Propositions 3.2 and 3.3 the bounds only depend on the right end point t_{au} of the support. From the proofs of Propositions 3.2 and 3.3, it is clear that T_a can be bounded from below too, i.e, there can exist a value $t_{al} > 0$ such that $t_{al} \leq T_a$ with probability one. But since the bound will not involve this lower end point, we have chosen not to explicitly state this bounded-from-below-assumption in the propositions. Third, notice that we have not explicitly stated the assumption that the moments are finite since the boundedness assumption on T_a implies that all its moments exist.

3.5 SOME PROOFS OF CONVEXITY

The bounds derived in the previous sections all involve the minimization of some univariate function in λ. Clearly, any feasible choice for λ results in a valid bound. In particular, one can set $\lambda = 1$, so that the standard Markov inequality is obtained in (3.1). By solving the minimization problems, tighter bounds can be found. To ensure global optimality, convexity of the objective functions is desired. In this section we prove that the functions in Propositions 3.1 and 3.2 are indeed convex via second order characterizations. Although no formal proof is provided, we conjecture that the function in Propositions 3.3 is also convex. Some empirical evidence to support this conjecture is provided in the next section. In any case (i.e, convex or not), we want to emphasize that the bounds are valid for *any* feasible choice for λ. Moreover, as will be seen in Section 3.6, irrespective of whether we have convexity, the bounds are useful. In other words, convexity is a luxury, not a necessity, in the current chapter. The following two results are of crucial importance in establishing convexity in the current section **(**Boyd and Vandenberghe, 2004**)**.

Lemma 3.4 If f is convex and non-decreasing and if h is convex, then $g(\lambda) = f(h(\lambda))$ is also convex.

Lemma 3.5 Let H_i, i =1,2,…A be convex functions. Moreover, let γ_i be non-negative real numbers, then $\sum_i \gamma_i H_i$ is also convex.

Proposition 3.4 The function

$$g_1(\lambda) \equiv \exp(-\lambda t)\left(\frac{1}{A}\sum_{a=1}^{A}\left(\frac{E[T_a](\exp(\lambda t_{au})-\exp(\lambda t_{al}))}{t_{au}-t_{al}}+\frac{t_{au}\exp(\lambda t_{al})}{t_{au}-t_{al}}-\frac{t_{al}\exp(\lambda t_{au})}{t_{au}-t_{al}}\right)\right)^A$$

in Proposition 3.1 is convex.

Proof Note that we can rewrite $g_1(\lambda)$ as

$$g_1(\lambda) = f_1(h_1(\lambda))$$

where

$$f_1(x) = x^A$$

and

$$h_1(\lambda) = \exp\left(\frac{-\lambda t}{A}\right)\frac{1}{A}\sum_{a=1}^{A}\left(\frac{E[T_a](\exp(\lambda t_{au})-\exp(\lambda t_{al}))}{t_{au}-t_{al}}+\frac{t_{au}\exp(\lambda t_{al})}{t_{au}-t_{al}}-\frac{t_{al}\exp(\lambda t_{au})}{t_{au}-t_{al}}\right).$$

Thus if we can show that $h_1(\lambda)$ is convex, we have proved the claim as $f_1(x)$ is clearly convex and non-decreasing for $x \geq 0, A \geq 1$. Indeed, $h_1(\lambda)$ is convex since

$$\begin{aligned}
&\frac{d^2}{d\lambda^2}\exp\left(\frac{-\lambda t}{A}\right)\left(\frac{E[T_a](\exp(\lambda t_{au})-\exp(\lambda t_{al}))}{t_{au}-t_{al}}+\frac{t_{au}\exp(\lambda t_{al})}{t_{au}-t_{al}}-\frac{t_{al}\exp(\lambda t_{au})}{t_{au}-t_{al}}\right) \\
&= \exp\left(\frac{-\lambda t}{A}\right)\frac{1}{A^2}\frac{((t-At_{au})^2(E[T_a]-t_{al})\exp(\lambda t_{au})-(t-At_{al})^2(E[T_a]-t_{au})\exp(\lambda t_{al}))}{t_{au}-t_{al}} \\
&\geq 0
\end{aligned}$$

By Lemma 3.5, the result follows. **Q.E.D.**

With the aid of a computer algebra package, one can establish the next convexity result.

Proposition 3.5 The function

$$g_2(\lambda) \equiv \exp(-\lambda t)\left(\frac{1}{A}\sum_{a=1}^{A}\left(1+\frac{E[T_a]}{t_{au}}\left(\exp(\lambda t_{au})-1\right)\right)\right)^{A}$$

in Proposition 3.2 is convex.

Proof As above, we can rewrite $g_2(\lambda)$ as

$$g_2(\lambda) = f_2(h_2(\lambda))$$

where

$$f_2(x) = x^A$$

and

$$h_2(\lambda) = \exp\left(\frac{-\lambda t}{A}\right)\frac{1}{A}\sum_{a=1}^{A}\left(1+\frac{E[T_a]}{t_{au}}\left(\exp(\lambda t_{au})-1\right)\right).$$

If $h_2(\lambda)$ is convex, the claim follows directly from Lemma 3.4. To prove convexity of $h_2(\lambda)$, by Lemma 3.5, it suffices to show that

$$\frac{d^2}{d\lambda^2}\exp\left(\frac{-\lambda t}{A}\right)\left(1+\frac{E[T_a]}{t_{au}}\left(\exp(\lambda t_{au})-1\right)\right) \geq 0 .$$

With some tedious algebra, one can show that

$$\frac{d^2}{d\lambda^2}\exp\left(\frac{-\lambda t}{A}\right)\left(1+\frac{E[T_a]}{t_{au}}\left(\exp(\lambda t_{au})-1\right)\right) = \frac{1}{A^2 t_{au}}\exp\left(\frac{-\lambda t}{A}\right)(at^2+bt+c)$$

where

$$a = t_{au} + E[T_a]\left(\exp(\lambda t_{au}) - 1\right)$$
$$b = -2At_{au}E[T_a]$$
$$c = A^2 t_{au}^2 E[T_a]\exp(\lambda t_{au})$$

One can verify that the discriminant of the quadratic equation $at^2 + bt + c$ is given by

$$4E[T_a]t_{au}^2 A^2 \exp(\lambda t_{au})(E[T_a] - t_{au})$$

which is non-positive since $E[T_a] \le t_{au}$. This implies that $at^2 + bt + c \ge 0$, completing the proof. **Q.E.D.**

It is easy to see that proving convexity requires numerous tedious algebraic manipulations, especially when higher order moments are involved. However, given the usefulness of the bounds (see next section), we believe that convexity is a pure luxury in the current chapter. Hence, let us simply conclude this section with a conjecture that the function

$$g_3(\lambda) \equiv \exp(-\lambda t)\left(\frac{1}{A}\sum_{a=1}^{A}\left(\exp\left(\lambda E[T_a] + \frac{E[T_a^2]}{t_{au}^2}\left(\exp(\lambda t_{au}) - 1 - \lambda t_{au}\right)\right)\right)\right)^A \tag{3.3}$$

is convex. In the next section, we will present some empirical evidence to support this conjecture.

3.6 NUMERICAL DEMONSTRATION

In this section we use the well-known Sioux Falls network (Figure 3.1) to illustrate the bounds derived in the previous sections (note that for the Sioux Falls network $A = 76$). The network data reported in Suwansirikul et al. (1987) provide the starting point of our numerical analysis. Without loss of generality, we solved an instance of the deterministic user equilibrium assignment (Sheffi, 1985) based on the data in Suwansirikul et al. (1987). The resulting total travel times on the links are assumed to be the expected values $E[T_a]$. Upper and lower bounds on the travel times are obtained by the multiplication of their mean values with factors q_{al} and q_{au} where $0 < q_{al} < 1 < q_{au}$, i.e, $t_{al} = q_{al}E[T_a]$ and $t_{au} = q_{au}E[T_a]$. Likewise, the second moments are obtained by the multiplication of $\left(E[T_a]\right)^2$ with $c \geq 1$ (since $0 \leq Var[T_a] = E[T_a^2] - \left(E[T_a]\right)^2$) such that $c\left(E[T_a]\right)^2 \leq (t_{au} - t_{al})^2 / 4 + (E[T_a])^2$. This bound on the largest possible second moment can, for example, be found in Theorem 4.1 of Seaman and Odell (1985). It follows from the fact that a Bernoulli random variable (Ross, 2002) has the highest possible variance among all random variables on [0,1].

To evaluate the bounds derived in Propositions 3.1 through 3.3, we employ the “fminbnd” routine available in MATLAB which is an optimization procedure based on golden section search and parabolic interpolation (Press et al., 2007). The interval in which to search for an optimal value for λ was constrained to be $(0, 50]$ that was empirically found to contain a minimizer for the cases considered. Computational tests demonstrated the effectiveness and computational efficiency of this optimization technique (e.g., Figure 3.2 was obtained within 1 second on an HP laptop computer with a 4.00 GHz AMD Turion™ Dual-Core Processor and 4 GB RAM). Note that from the convexity results in Section 3.5 it is known that the resulting bounds in Proposition 3.1 and 3.2 are global minima, whereas it is not guaranteed that the bound from Proposition 3.3 is minimized with respect to λ. However, based on our computational experience, we conjecture that they *are* in fact global minima. Some empirical evidence to support this conjecture will be presented at the end of this section.

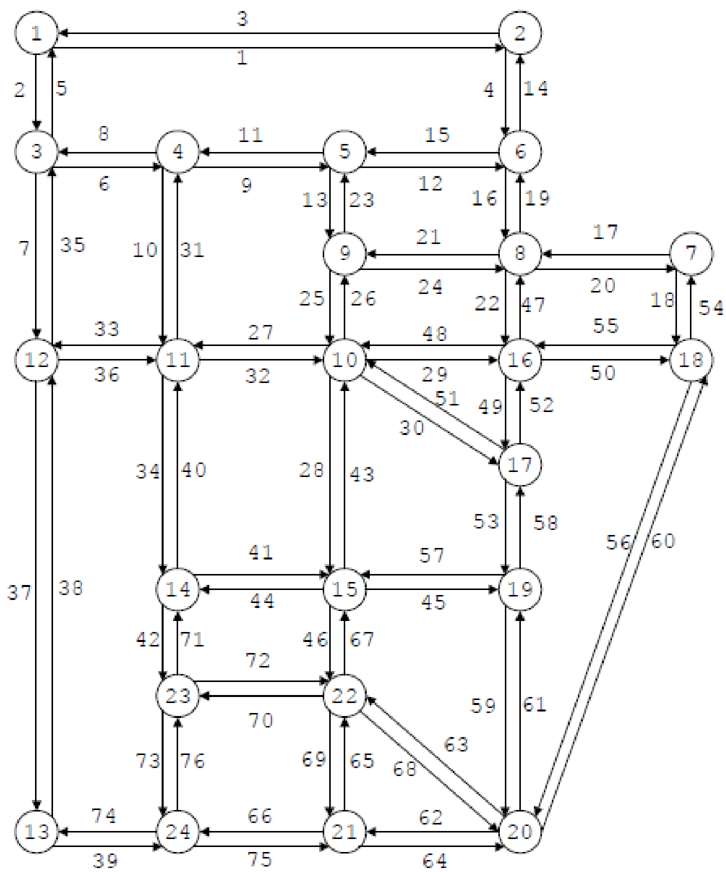

Figure 3.1: Sioux Falls test network with node and link numbers.

Unless stated otherwise, we set $q_{al} = 0.2$ and $q_{au} = 3$ in the following numerical experiments. Figure 3.2 depicts the upper bounds from Propositions 3.1 through 3.3 for the case $c = 1.1$ as a function of the TSTT (using the demand values reported in Suwansirikul et al., 1987). While it was not necessary in our case study, one can always adjust the predicted bounds so that the resulting value lies between 0 and 1 (cf. the remark at the end of Section 3.3). That is, suppose that the *predicted* upper bound (i.e, the theoretical bounds derived in Sections 3.3 and 3.4) is given by z_u, then the adjusted upper bound would be given by $\min\{1, z_u\}$.

From Figure 3.2, we can make several interesting observations. A theoretical upper bound on the TSTT is given by the TSTT associated with the case when $T_a = q_{au}E[T_a]$, henceforth referred to as the worst case TSTT (Note that the worst case TSTT is an bound on the TSTT. They are not to be confused with the bounds derived in Sections 3.3 and 3.4 that were bounds on the *right-tail probability* of the TSTT). The worst case TSTT for the instance examined in Figure 3.2 is equal to 302 hours. All upper bounds are able to capture this fact very accurately: they predict a probability of zero that the worst case TSTT is exceeded. In fact, they predict that a much larger range of TSTT values are not possible (see Figure 3.2). From the figure, we can also that in this specific instance, the bound in Proposition 3.3 is tighter than the bound in Proposition 3.1 which is turn is tighter than the bound in Proposition 3.2. Since the bounds in Proposition 3.1 and Proposition 3.3 are developed using fundamentally different approaches (e.g., in Proposition 3.1 we relied on Madansky inequality and in Proposition 3.3 we did not), it is not possible to assess the importance of the inclusion of variance information based on these two bounds. To do this, we need to consider the bounds from Propositions 3.2 and 3.3. From Figure 3.2 it is clear that the incorporation of variance information results in substantially tighter bounds in this case. For example, for TSTT = 130 hours, the bound from Proposition 3.2 predicts that this TSTT value is exceeded with a probability of at most 0.43, whereas, the bound from Proposition 3.3 tells us that the probability is at most 0.09 (which is a reduction in uncertainty of about 80%!). Finally, while not really relevant from a reliability perspective (since we are mostly interested in the likelihood that relatively high TSTT values are exceed), we note that for TSTT values less than about 102 hours, the bounds are trivial (the mean TSTT in this case was 101 hours, i.e, $\sum E[T_a] = 101$).

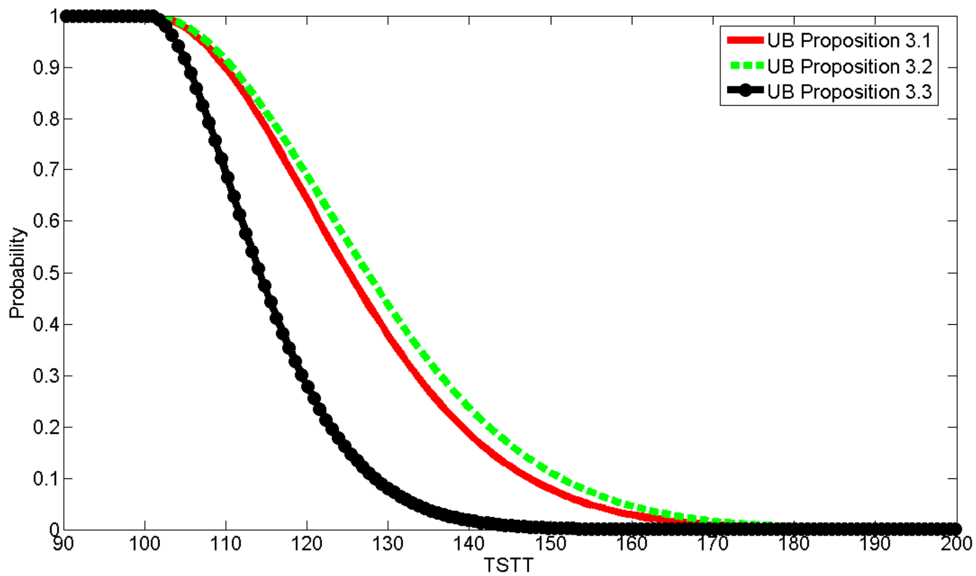

Figure 3.2: Upper bounds on the unreliability function for $c = 1.1$.

Before we proceed to the next numerical experiment, let us first take a step back and examine in a thought experiment under what circumstances variance/ second order moment information is likely to result in tighter bounds. Clearly, the higher the variance, the less information it actually gives us. Hence it seems to be logical that when the variance increases (due to an increase in the second order moment), *ceteris paribus*, the difference between the bounds from Propositions 3.2 and 3.3 will get smaller. Figure 3.3 ($c = 2.2$) confirms this reasoning. It shows that the bound from Proposition 3.3 has moved much closer to the bound from Proposition 3.1 (which obviously has not changed as it does not depend on the second order moment). In the limiting case when $E[T_a^2]=(t_{au}-t_{al})^2/4+(E[T_a])^2$ (cf. Seaman and Odell, 1985), the bound in Proposition 3.3 becomes less tight than its counterpart from Proposition 3.1, see Figure 3.4. At first, it might seem contradictory that a bound that contains variance information is less tight than one that does not, this is perfectly fine since the bounds were derived using very different approaches. Moreover, the bound from Proposition 3.3 almost coincides with the bound in Proposition 3.2. Note that we have not given any theoretical guarantee on the ordering of the bounds. This would require a discussion of the sharpness of the bounds, i.e, whether there exists random travel times such that the bounds are attained, which we have left as an important future research topic. In light of these observations, for now we suggest to use the minimum of the upper bounds as *the* bound in reliability studies.

We want to emphasize that our bounds are distribution-free. Hence, we believe that any (simulation) exercise to evaluate the bounds based on specific distributional assumptions (of course, if distributions were known, one would not adopt the proposed distribution-free approach to begin with!) would be misleading since the quality of the bounds can be made better or worse, depending on the choice of the underlying probability distributions. However, for a *very* rough idea of how the bounds compares to specific distributions, we have used Monte Carlo simulation to estimate a number of points of the unreliability function based on the normal, gamma and extreme value (type 1) distributions, see Figure 3.5. Each estimate was obtained based on 10000 samples, using the same mean and variance as the instance in Figure 3.2. As can be seen, the closeness to the bound is a function of the probability distribution chosen.

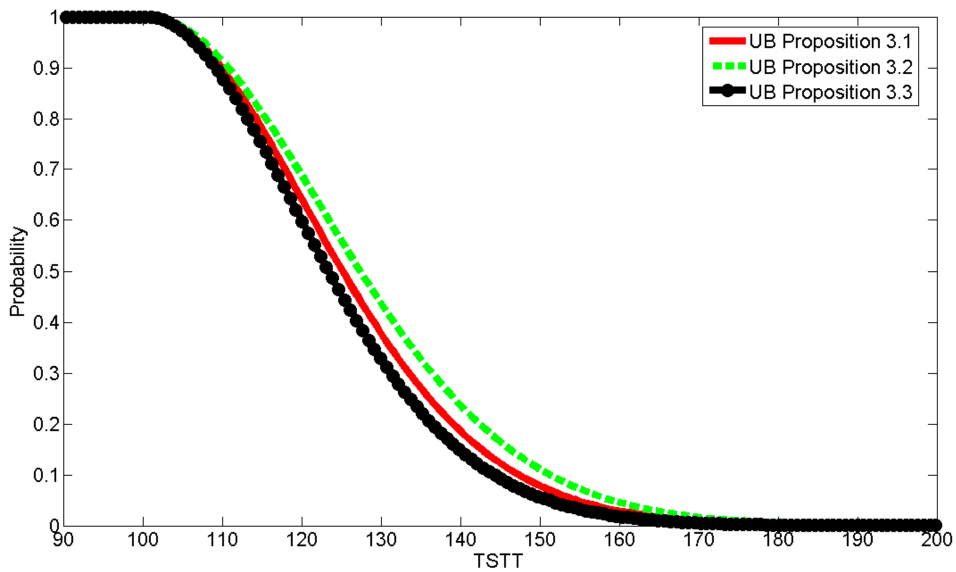

Figure 3.3: Upper bounds on the unreliability function for c = 2.2.

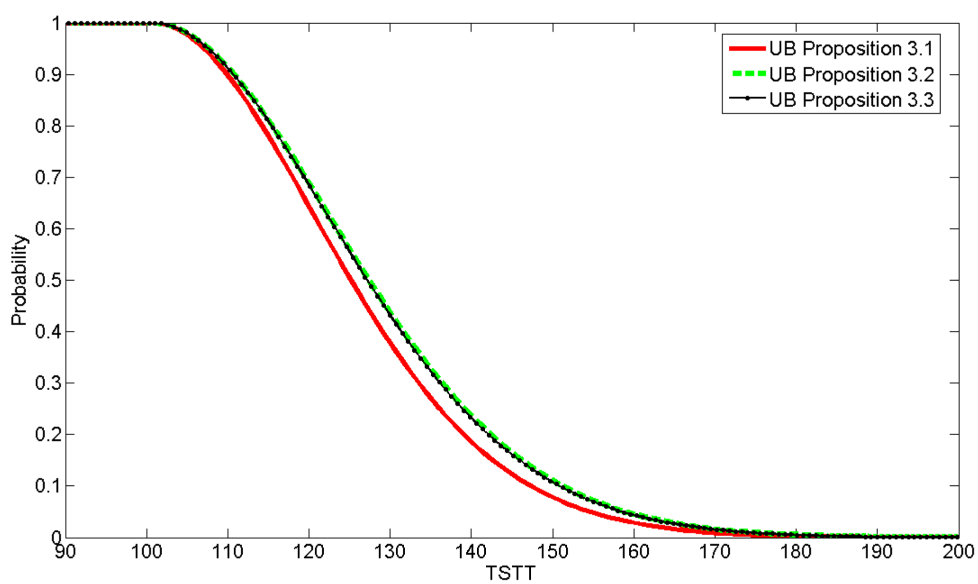

Figure 3.4: Upper bounds on the unreliability function under maximum variance.

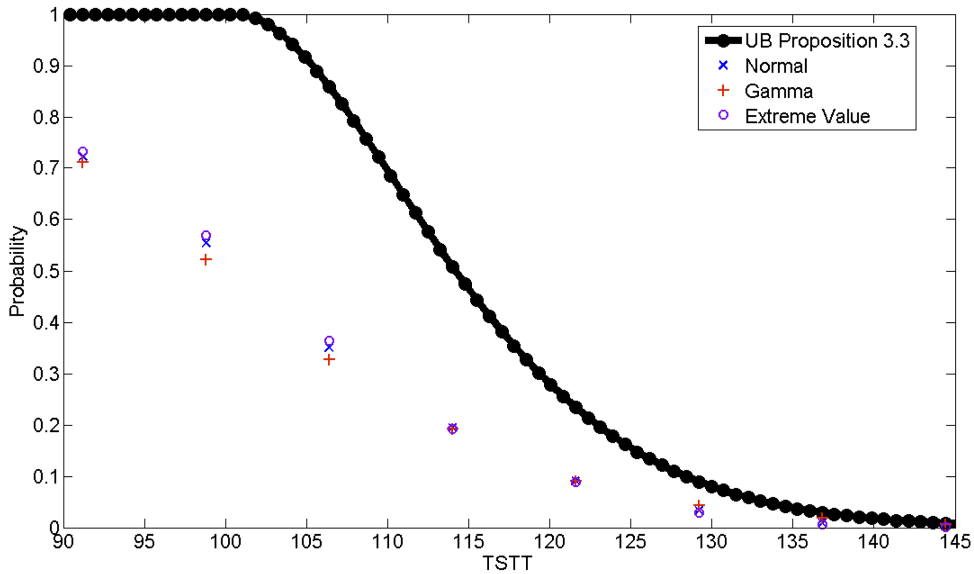

Figure 3.5: Monte Carlo estimates of the unreliability function.

Before we conclude our numerical case study, we present some observations to support our conjecture that the function (3.3) is convex. One indication is that all bounds shown in the above figures are continuous. If the function had multiple local minima (i.e, non-convex), discontinuities would likely appear at certain points. Another indication is provided by a visual inspection of (3.3). Figure 3.6 shows three (t = 114, 122 and 129) typical graphs of (3.3) using the Sioux Falls network data and $c = 1$. Clearly, the functions are nicely convex in the range considered (it can be verified that outside this range the functions simply increase without bound). However, we are fully aware that these observations, by no means, constitute a proof of convexity. On the other hand, we have demonstrated that the bounds are extremely useful since they enable us to make nontrivial probabilistic statements on the system travel time (whether the functions involved are convex or not) using rather minimal assumptions.

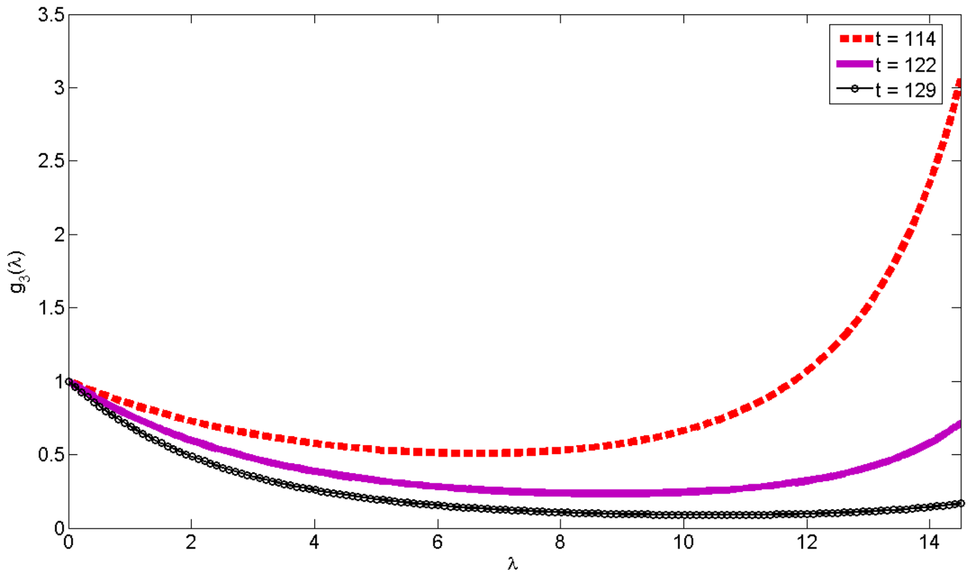

Figure 3.6: Three typical convex graphs of $g_3(\lambda)$.

Chapter 4: Transportation Network Design for Travel Time Reliability

4.1 INTRODUCTION

The problem of judiciously selecting links for capacity expansions in transportation networks is known as the transportation network design problem (NDP). Virtually all network design models examined in the literature have the objective of minimizing the (expected) total system travel time. For example, Abdulaal and LeBlanc (1979), Suwansirikul et al. (1987), Marcotte (1983), Meng et al. (2001), Friesz et al. (1992), LeBlanc and Boyce (1986), Patriksson and Rockafellar (2002), Mouskos (1991) and Ng and Waller (2009c) minimized the TSTT based on the user optimal static traffic assignment paradigm. Other authors suggested to use the system optimal traffic assignment problem as a computationally viable proxy for the otherwise NP-hard problem (e.g., LeBlanc (1975), Hoang (1982), LeBlanc and Abdulaal (1979), Dantzig et al. (1979), LeBlanc and Abdulaal (1984) and, more recently, Ng and Waller, 2009a). For a comprehensive overview of these NDP problems, we refer to Magnanti and Wong (1984), or more recently, Yang and Bell (1998). Models based on dynamic traffic assignment (DTA) have been introduced to alleviate concerns regarding the time-invariant nature of classical static models (e.g., Janson (1995), Waller et al. (2006), Li et al. (2003), Waller and Ziliaskopoulos (2001) and Karoonsoontawong and Waller, 2005). However, the increased modeling realism has substantially increased the computational burden. As such, DTA-based NDP models are typically limited to small networks (Ng et al., 2009).

Alternative objective functions (other than system travel time) has barely been considered in the NDP literature. Chen et al. (2003) formulated an NDP with the objective of maximizing the mean and minimizing the variance of profits in the context of a Build-Operate-Transfer scheme. Sumalee et al. (2006) presented an NDP to maximize the probability that the system travel time remains below certain threshold levels, using the travel time reliability assessment technique developed by Clark and Watling (2005). In Chen et al. (2007), the objective was to minimize the budget requirements so that it was guaranteed that the TSTT remains below a pre-specified level. The fact that the minimization of travel time unreliability has barely been considered is most likely due to the (virtually) non-existence of (computationally efficient) reliability assessment techniques. For instance, Chen et al. (2009) simply used Monte Carlo simulation to evaluate travel time reliability. With the new computationally efficient methodologies introduced in this dissertation, the situation has changed. In this chapter we present a new network design model that is aimed at improving the reliability of travel time and which is substantially less computationally intensive than current simulation-based models. The model uses the distribution-free approach presented in Chapter 3 as the basis for reliability assessment.

The remainder of this chapter is organized as follows. Before we present the new network design model in Section 4.3, we first present a simple example to give us some insights into how designing for reliability differs from designing for system travel time (Section 4.2). Numerical experiments demonstrate the computational performance of the proposed model in Section 4.5. Comparisons with traditional network design models will be made throughout this chapter.

4.2 Design for Reliability versus System Travel Time

In this section we consider the the 2-link, 2-node test network shown in Figure 4.1. Node O generates d units of travel demand destined for node D. The nodes are connected by two links whose link-performance functions are given by

$$\tilde{t}_1 = 1 + 0.15\left(\frac{v_1}{\tilde{c}_1 + z_1}\right)^2,\ t_2 = 1 + 0.15\left(\frac{v_2}{c_2 + z_2}\right)^2$$

where z_1 and z_2 are non-negative integers, denoting the amount of capacity additions to link 1 and 2, respectively. (Note that the integrality assumption is totally without loss of generality as long as the amount of capacity additions is countable, which is always the case in practice). Notice that the capacity/ travel time on link 2 is fully deterministic, i.e, $Var[t_2] = 0$, whereas the travel time on link 1 is uncertain. Furthermore, as in Chapter 2, we make the assumption that the variance of the capacities do not change due to capacity additions. (Note that the variance of the travel times will generally change due to the rerouting of traffic). While the network in Figure 4.1 is very simple, it allows us to gain a number of insights into the reliability optimal NDP (RO-NDP).

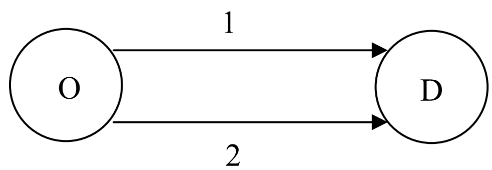

Figure 4.1: Two-Link Test Network

Let $t^* \geq v_1 + v_2$, from Lemma 2.1, we have

$$\Pr\left(T_1 + T_2 > t^*\right) = \Pr\left(T_1 > t^* - T_2\right)$$

$$= 1 - \Pr\left(T_1 \le t^* - T_2\right) = 1 - \Pr\left(\tilde{c}_1 > \frac{v_1}{\left(\frac{1}{\alpha_1}\left(\frac{t^* - T_2}{v_1 t_1^f} - 1\right)\right)^{1/\beta_1}}\right) \tag{4.1}$$

To determine the link flows, we (analytically) solve for the deterministic user equilibrium (Sheffi, 1985) using the nominal capacity values. Once the distribution of $\tilde{c}_1$ is specified, (4.1) can be readily evaluated. Here we assume that $\tilde{c}_1$ is normally distributed with mean c_1 and standard deviation σ_1. Suppose that $z_1 + z_2 = K$ for some positive integer *K*. Since we have assumed that z_1 and z_2 are non-negative integers, we can easily enumerate all feasible solutions of the RO-NDP (our goal here is to examine properties of the reliability optimal network design problem rather than devising solution algorithms). In the following, we examine a number of instances of the problem (assuming $d = 60$), illustrating the various scenarios that might occur.

Scenario 1 (Travel time optimal and reliability optimal network design solutions are very different) Consider the instance $c_1 = 20, c_2 = 25, \sigma_1 = 6, K = 20$. The upper part of Figure 4.2 depicts the travel time unreliability (4.1) for $t^* = 70$ versus the amount of capacity expansion z_1 (recall that z_1 uniquely determines z_2 via the relation $z_2 = K - z_1$). The lower figure shows the conventional TSTT (obtained assuming nominal capacity values) versus z_1. The figure demonstrates that any addition of capacity to link 1 will increase system's reliability. Furthermore, the more capacity is added, the more reliable the network becomes. This might seem intuitive as one might argue that since the travel time on link 1 is random (and the travel time on link 2 is fully deterministic), one can increase link 1's capacity to increase the reliability of the network. Below (scenario 3), we will demonstrate that such reasoning is fundamentally flawed. In this specific instance, the optimal solution to the reliability optimal network design problem is given by $z_1 = 20, z_2 = 0$. This is clearly diametrically opposite to the solution of the conventional system optimal network design problem, as the TSTT increases with z_1.

Scenario 2 (Travel time optimal and reliability optimal network design solutions coincide) Consider the instance $c_1 = 10, c_2 = 25, \sigma_1 = 3, K = 20, t^* = 75$. Figure 4.3 shows the resulting reliability (upper figure) and TSTT values (lower figure) for this instance. As opposed to Figure 4.2, increasing capacity on link 1 is optimal both in terms of reliability as well as system travel time.

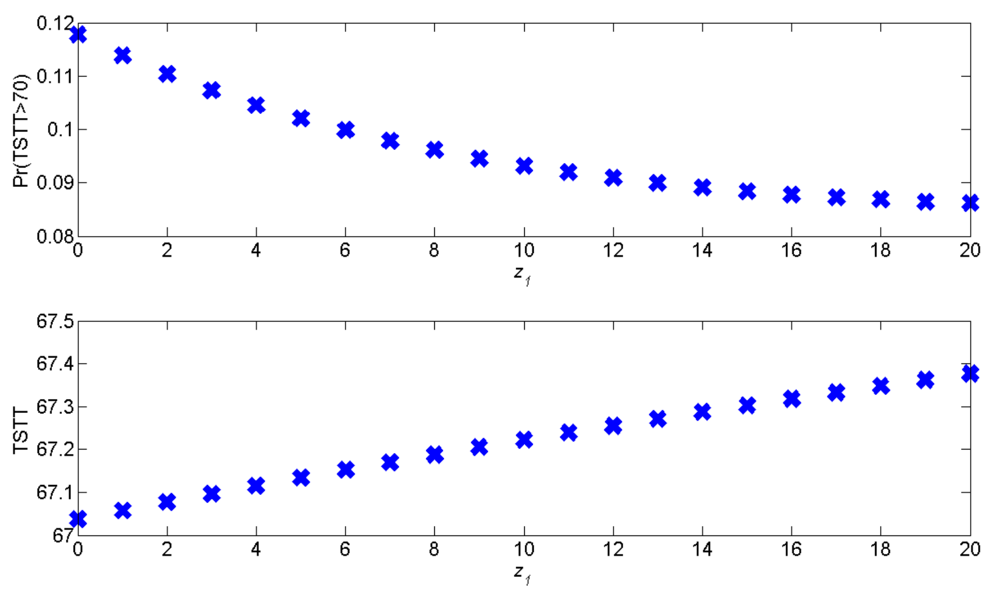

Figure 4.2: System optimal and reliability optimal network design solutions are diametrically opposite.

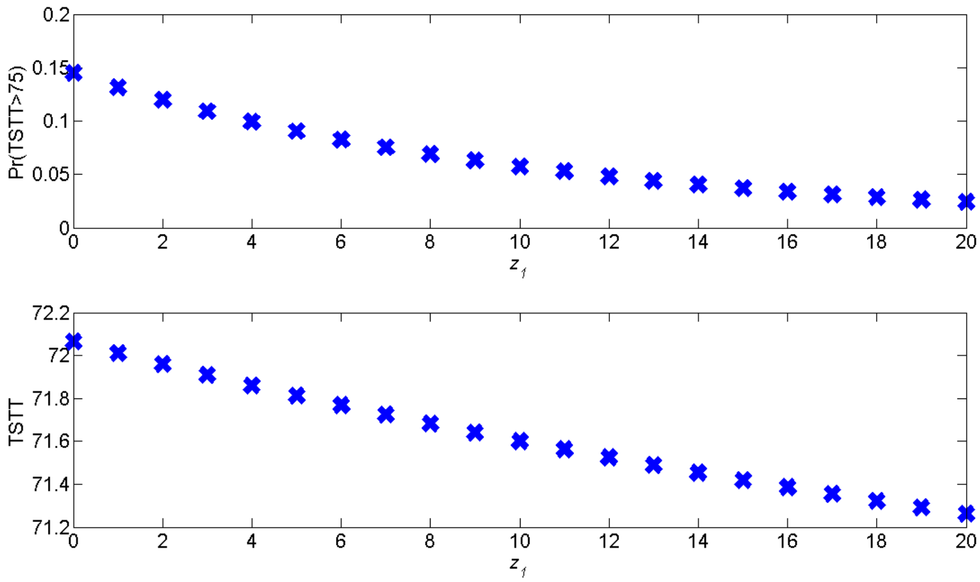

Figure 4.3: System optimal and reliability optimal network design solutions coincide.

Scenario 3 (Capacity expansion on uncertain link can lead to an increase in unreliability) Consider the instance $c_1 = 25, c_2 = 10, \sigma_1 = 6, K = 30$. The scenarios above might have given the reader the impression that any capacity additions to links with random travel times might be beneficial for the reliability of a stochastic transportation network. Figure 4.4 shows that this reasoning does not necessarily hold. Consider the case $t^* = 69$. Figure 4.4 shows that after a certain amount of capacity additions ($z_1 = 12$ in this case), any additional capacity on link 1 would be detrimental for the reliability of the network! The upper figure in Figure 4.5 shows the PDFs of the total travel time on link 1 associated with three different capacity expansions decisions ($z_1 = 5, z_1 = 12, z_1 = 30$). By expanding the capacity on link 1, the link becomes more attractive for travelers, whereas link 2 becomes less attractive. (One might argue that this it is simply a consequence of the DUE assumption where travelers do not consider travel time variability in their route choice. However, in real-life situations, we believe that it is reasonable to anticipate more travelers on a link when its capacity has been increased.) Due to the increased travel volume on link 1, the PDF is shifting to the right as z_1 increases. Moreover, the PDF becomes slightly flatter. The PDF of the TSTT is shown in the lower part of the figure (that can be simply obtained by a translation of the PDFs in the upper figure). The figure shows that certain ranges of the TSTT become more likely (e.g., 69-70 hours) with an increasing value of z_1, whereas the higher TSTT values become less likely due to the more rapid decay of the PDFs when z_1 is larger. It is not difficult to imagine that one can construct examples in which any addition of capacity to link 1 results in a less reliable system. Figure 4.5 shows such an instance ($c_1 = 30, c_2 = 10, \sigma_1 = 6, K = 25, t^* = 69$).

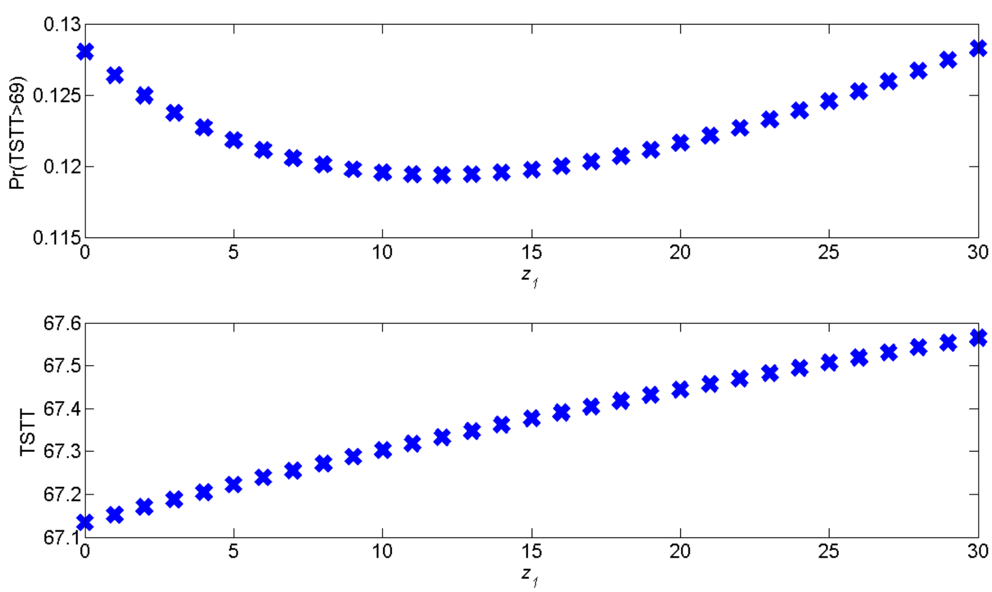

Figure 4.4: Capacity expansion on the random link can increase unreliability.

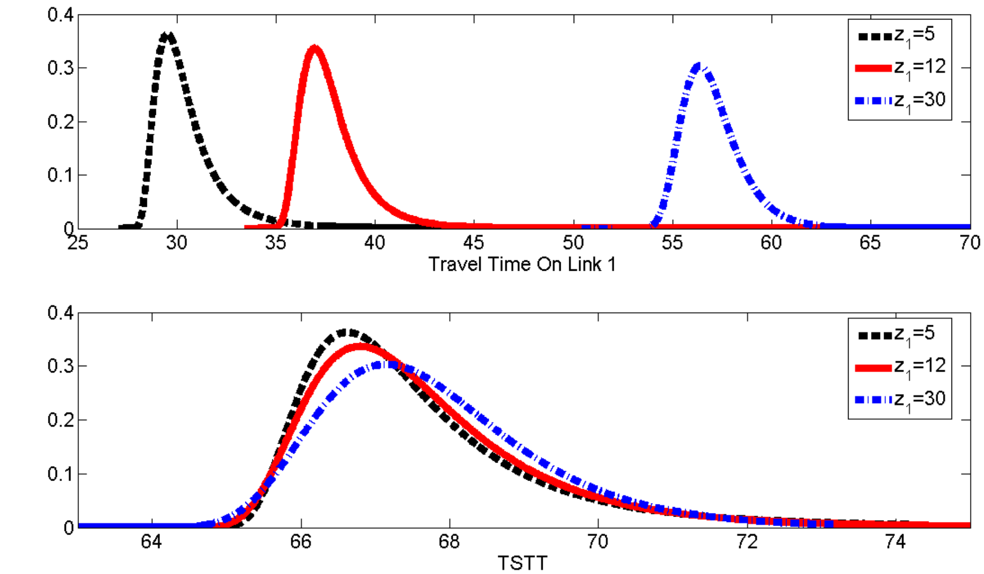

Figure 4.5: Upper: PDFs of the total travel time on link 1 for three different capacity expansion decisions. Lower: PDFs of TSTT for three different capacity expansion decisions.

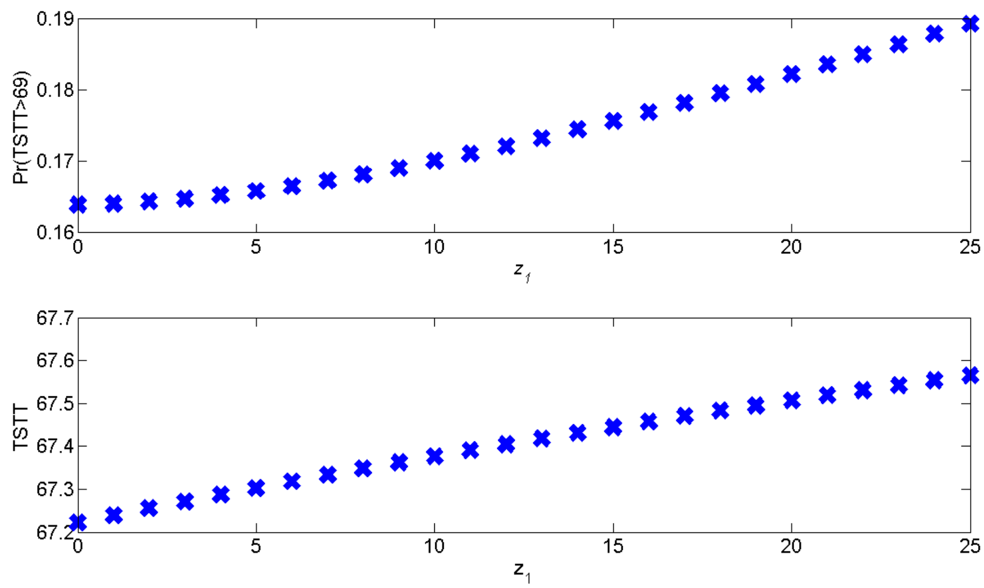

Figure 4.6: Any Capacity addition on the stochastic link decreases system reliability.

4.3 MODEL FORMULATION

To introduce the new network design model RO-NDP, some notation is in order. Let us collect some old and introduce some new notation:

V = set of all nodes in a transportation network.

A = set of all links in a transportation network.

$f(\cdot)$ = capacity expansion cost function.

B = total budget for capacity expansion.

v_a^{rs} = traffic volume on link *a* from travelers going from $r \in V$ to $s \in V$.

v_a = total traffic volume on link *a*.

q^{rs} = travel demand for travel from $r \in V$ to $s \in V$.

c_a = nominal capacity on link *a*.

y_a = amount of capacity to be added to link *a*.

u_a = maximum level of additional capacity on link *a*.

$\Gamma(i)$ = the set of links emanating from node $i \in V$.

$\Gamma^{-1}(i)$ = the set of links incident to node $i \in V$.

Traditional network design models typically assume that future link flows (and hence future travel times) can be predicted with certainty. Here we relax this assumption and assume that only "mean values" can be predicted together with a range in which the actual travel times lie. The objective of our model is to minimize an upper bound on the unreliability of the system travel time. Following Chapter 3, we do not assume that we know exact probability distributions assumptions, so that this is the best we can do. In particular, the objective is to minimize the function (cf. Proposition 3.1):

$$UB_1(y_a)$$
$$\equiv \inf_{\lambda>0}\left[\exp(-\lambda t^*)\left(\frac{1}{A}\sum_{a=1}^{A}\left(\frac{E[T_a](\exp(\lambda t_{au})-\exp(\lambda t_{al}))+t_{au}\exp(\lambda t_{al})-t_{al}\exp(\lambda t_{au})}{t_{au}-t_{al}}\right)\right)^{A}\right]$$

where we have – for notational convenience – suppressed the dependence of $E[T_a]$, t_{au} and t_{al} on v_a and y_a. Of course, other objective functions are possible (e.g., one that includes variance information, cf. Proposition 3.3). While in the assessment methodology in Chapter 3 the user is free to estimate $E[T_a]$ using every possible way (s)he can think of, in this chapter we assume that $E[T_a]$ is predicted by the solution of the traditional user optimal traffic assignment problem (Sheffi, 1985). That is, we equate travel times resulting from the solution of a static traffic assignment problem (using the nominal parameter values) to $E[T_a]$. In order to capture the uncertainties in these predictions, we specify a set of factors q_{al} and q_{au}, $0<q_{al}<1<q_{au}$ (cf. Chapter 3) to describe the uncertainty intervals. The RO-NDP can now be stated as the following bi-level optimization problem (Bard, 1998) to capture the sequential nature of the decision making process (transportation planners determine capacity expansions, road users react to these by finding new alternative routes):

$$\min_{y_a} \; UB_1(y_a) \tag{4.2}$$

subject to

$$\sum_a f(y_a) \le B \tag{4.3}$$

$$0 \le y_a \le u_a \qquad \forall a \in A \tag{4.4}$$

$$\min_{v_a} \sum_a \int_0^{v_a} t_a(\omega, y_a)\, d\omega \tag{4.5}$$

subject to

$$\sum_{a \in \Gamma(r)} v_a^{rs} - \sum_{a \in \Gamma^{-1}(r)} v_a^{rs} = q^{rs} \qquad \forall r, s \in V, r \ne s \tag{4.6}$$

$$\sum_{a \in \Gamma(s)} v_a^{rs} - \sum_{a \in \Gamma^{-1}(s)} v_a^{rs} = -q^{rs} \qquad \forall r, s \in V, r \ne s \tag{4.7}$$

$$\sum_{a \in \Gamma(i)} v_a^{rs} - \sum_{a \in \Gamma^{-1}(i)} v_a^{rs} = 0 \qquad i \in V \setminus \{r, s\} \tag{4.8}$$

$$\sum_{r,s} v_a^{rs} = v_a \qquad \forall a \in A \tag{4.9}$$

$$v_a^{rs} \ge 0 \qquad \forall a \in A, \forall r, s \tag{4.10}$$

where

$$t_a(v_a, y_a) \equiv t_a^f \left(1 + \alpha \left(\frac{v_a}{c_a + y_a} \right)^{\beta} \right)$$

and

$$E[T_a] = v_a t_a(v_a, y_a)$$

In (4.2) the objective of minimizing the upper bound on the unreliability function is expressed. Constraint (4.3) imposes a budget constraint, while constraint (4.4) ensures that the expansion of road capacities has a (physical) limit. The assumed route choice behavior – the traditional deterministic user equilibrium (Sheffi, 1985) – is specified in (4.5) to (4.10). Bi-level problems are known to be among the most challenging optimization problems (Bard, 1998). One solution method is the use of meta-heuristics.

4.4 SOLUTION METHOD: GENETIC ALGORITHM

To address the computational challenges of the RO-NDP, we employ genetic algorithms (GAs) that have a demonstrated record of solving challenging optimization problems and, in particular, NDPs (e.g., see Ng et al., 2009 and the references therein). A GA is an iterative procedure aimed at finding the global optimum in complex optimization problems (Goldberg, 1989). At each generation, it maintains a population of candidate solutions (also known as chromosomes) that evolves according to principles from natural selection (survival of the fittest) and genetics (crossover and mutation). For this work, we have implemented a GA in the Java programming language. A candidate solution is encoded as a string of non-negative real numbers. Therefore, the length of the string is equal to the number of links present in the network under consideration. Each number in the chromosome corresponds to a value of the expansion decision y_a. Without loss of generality, here we assume that y_a can assume five different values:

$$y_a \in \left\{0, \frac{c_a}{4}, \frac{c_a}{2}, \frac{3c_a}{4}, c_a\right\}. \tag{4.11}$$

That is, we can either "do nothing", add a quarter of the existing nominal capacity, add a half of the existing nominal capacity, add three quarters of the existing nominal capacity or double the existing nominal capacity. The fitness values (i.e, the unreliability) of the chromosomes are obtained by evaluating (4.2). Notice that in order to evaluate (4.2), one must solve the traffic assignment problem (4.5)-(4.10) and a uni-dimensional optimization problem to determine the optimal value for λ. We employed the Frank-Wolfe algorithm (1956) and golden section search (Press et al., 2007) for these purposes. If the budget constraint (4.3) is violated, the unreliability of the chromosome is automatically set to 1 in order to discourage infeasibilities in future iterations. Chromosomes for reproduction are selected based on the roulette wheel selection algorithm. We applied single point crossover (which occurs with probability p_c) and random mutations (which occurs with probability p_m) to maintain diversity in the pool of candidate solutions.

4.5 NUMERICAL CASE STUDY

As noted above, in the absence of specific probability distributions describing the sources of uncertainty, one can at best minimize an upper bound on the unreliability. On the other hand, one can also think of the bound as a proxy for the exact unreliability since the relative (and not the absolute) performances of capacity expansion decisions are the most critical. Some empirical evidence to support such a hypothesis was already presented in Figure 3.5. Here we will further examine this approximation using the Nguyen-Dupuis test network introduced in Chapter 2, Figure 2.2. For the OD demand, we have rescaled their values to 6 units (from node 1 to node 2 and from node 4 to 3) and 4 units (from node 1 to node 3 and from node 4 to 2) in order to ensure higher volume-to-capacity ratios. Furthermore, here we have set $\alpha = 0.15,\ \beta = 4$. The nominal capacities and link flows in this network have been repeated in Table 4.1 for the reader's convenience. We will assume that the sources of uncertainties are the road capacities and that they vary independently. Following Chen et al. (2009), we use the following procedure to estimate the unreliability of the system travel time.

INPUT $S,\ t^*$

SET $S^* = 0$

FOR $k = 1$ to S

Sample from the random capacities.

Evaluate the system travel time based on the realized capacities, denote it by $TSTT_k$.

IF $TSTT_k > t^*$ **THEN** $S^* = S^* + 1$

END

OUTPUT: S^*/S

The fraction S^*/S (where S is a positive integer denoting the number of replications) is then used as an estimate of the unreliability of the system at t^* (Feller, 1966). Beyond small networks, the above Monte-Carlo procedure is computationally prohibitive. Next we investigate whether the (much more computationally efficient) RO-NDP can serve as a good proxy for the NDP where the objective function is evaluated based on the above Monte-Carlo procedure (Chen et al., 2009). To this end, we will solve two NDPs: RO-NDP and the NDP proposed in Chen et al. (2009). We set $t^* = \sum_a v_a t_a(v_a, 0) = 3024$ minutes, the current "nominal" system travel time (i.e, assuming nominal capacity values).

First, we examine the simulation-based based NDP. We assume that the capacities are uniform random variables (following Lo and Tung, 2003 and Chen et al., 2002). More specifically, we assume that that they are given by

$$\tilde{c}_a = y_a + c_a(1 + U(-0.5, 0.5)) \tag{4.12}$$

where $U(-0.5, 0.5)$ denotes a uniform random variable on the interval [-0.5,0.5]. Following Chen et al. (2009), to initiate the GA we randomly generated a set of r = 32 initial candidate capacity expansion plans by randomly choosing links (until the entire budget of $B = 15$ units is used) and assigning one of the five possible values of y_a to it (also in a random fashion). The cost function that will be used is given by $f(y_a) = 4y_a / c_a$. Again following Chen et al. (2009), we use $p_c = 0.3$, $p_m = 0.2$ in our calculations. The maximum number of generations is limited to 400 and the number of replications is S = 2000. The resulting optimal budget allocation decisions are given in Table 4.1 in the column named "$f(y_a)$ (MC)" (note that the entire budget has been used). The resulting estimate of the probability that the TSTT exceeds $t*$ is 0.21, while the initial unreliability was 0.89. A computation time of about 50 minutes were needed to complete the 400 iterations on a Linux machine, equipped with an Intel 3.00 GHz Xeon CPU and 32 GB of memory.

To use the proposed RO-NDP model, the parameters q_{al} and q_{au} need to be specified. In order to estimate the highest possible travel times on a link, a traffic assignment is solved with the lowest possible capacities, i.e, we set $\tilde{c}_a = 0.5c_a$ with probability one, see (4.11) and (4.12). The lowest possible travel times can be estimated by setting $\tilde{c}_a = 2.5c_a$ (i.e., $y_a = c_a$). Comparing to the base case ($y_a = 0, \tilde{c}_a = c_a$), estimates of q_{au} and q_{al} can be obtained, see Table 4.1 under the columns q_{au} and q_{al}. With these estimates (and the same GA-parameters as above), we can solve the RO-NDP model. The proposed budget allocation decisions are given in Table 4.1 in the column named "$f(y_a)$ (RO-NDP)". The computational time was negligible (less than 10 seconds on a standard laptop computer). This solution, if implemented in the above simulation-based NDP (assuming uniform capacities), gives us a probability that the TSTT exceeds 3024 minutes of 0.27. Hence we see that using the new RO-NDP model, one is able to get quite reasonable approximations (here within 6 percentage points) of the computationally intensive, simulation-based NDP. Moreover, while the computation time was around 50 minutes using the Monte-Carlo based approach, the computation time using the RO-NDP model was minimal. For large-scale networks, it is clear that Monte-Carlo simulation is computationally prohibitive.

In the last column of Table 4.1 we have also shown the best capacity expansion decision found in a conventional NDP (C-NDP) where the objective is to minimize the *expected* system travel time. The optimal value was found to be 2712 minutes. This solution if implemented in the RO-NDP model would yield a bound on the unreliability of 0.98, whereas the optimal RO-NDP solution implemented in C-NDP would give an expected TSST of 2759 minutes. It is interesting to note that the optimal solution to the C-NDP problem tends to prescribe capacity expansions to links that are not considered for expansion in the RO-NDP model.

Figure 4.7 presents some insights (in the form of histograms) into the variability (in terms of the system travel time) of the different solutions reported in Table 4.1, using 2000 samples. Notice that the optimal solution of RO-NDP gives a fatter right tail than other solutions, again suggesting that multi-objective optimization is appropriate.

A special case of the NDP is project ranking. In order to investigate how the distribution-free reliability assessment method performs in ranking projects, a number of capacity expansion projects (with three choices in each instance) on the Nguyen-Dupuis network is examined, see Table 4.2. The projects will be ranked using Monte-Carlo simulation – assuming (4.12) – and the distribution-free paradigm. The project resulting in the smallest unreliability is chosen (here we set $t^* = 4082$ minutes). For example, in the first instance (see Table 4.2), project 1 proposes to allocate 4 units (of the total budget of 20 units) to each of the links 1, 2, 3, 4, 5, project 2 proposes allocate to 4 units to each of the links 6, 7, 8, 9, 10, and project 3 proposes to allocate 4 units to links 15, 16, 17, 18, 19. Using Monte-Carlo simulation, it can be determined that the project that results in the smallest unreliability is project 2. The same conclusion can be reached using the much more computationally efficient distribution-free method (see last column of Table 4.2). In fact, in all instances that have been examined, the project rankings were identical, suggesting that the distribution-free methodology can be used as a computationally efficient proxy for the simulation-based methods when projects need to be ranked.

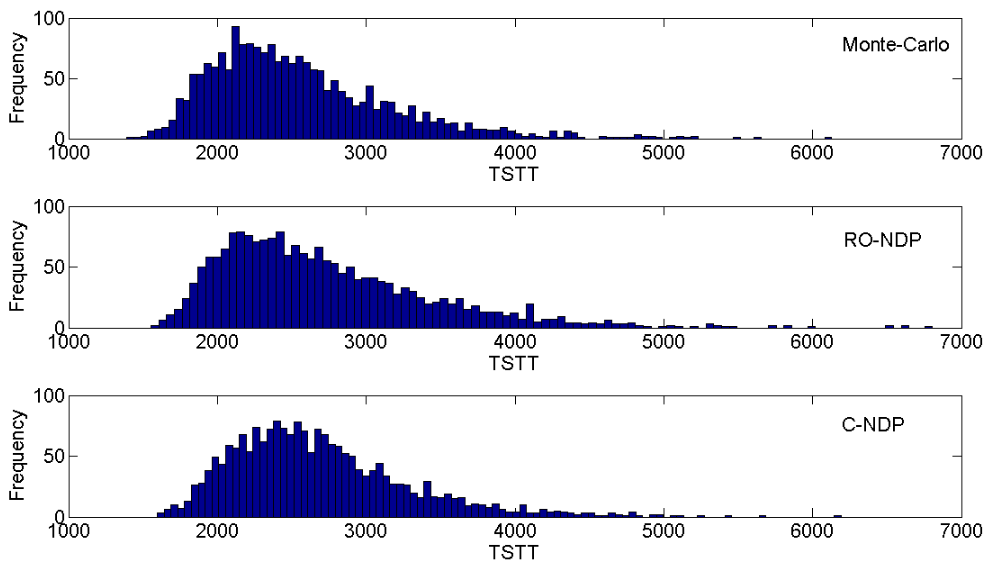

Figure 4.7: Upper: Histogram TSTT Monte-Carlo simulation-based solution. Middle: Histogram TSTT RO-NDP solution. Lower: Histogram TSTT C-NDP solution.

Table 4.1: Summary optimal budget allocations Nguyen- Dupuis test network

Link	c_a	v_a	q_{au}	q_{al}	**$f(y_a)$ (MC)**	**$f(y_a)$ (RO-NDP)**	**$f(y_a)$ (C-NDP)**
1	5	5.54	10.31	0.42	0	0	0
2	4	4.46	9.75	0.40	0	0	2
3	5	5.17	8.92	0.48	3	0	0
4	4	4.83	11.31	0.34	0	0	2
5	3	3.75	12.53	0.31	3	4	1
6	3	4.05	13.93	0.26	0	0	2
7	4	5.17	10.92	0.28	1	0	0
8	3	3.88	11.03	0.28	0	1	0
9	5	6.33	12.49	0.30	1	0	1
10	4	3.16	5.38	0.73	0	0	0
11	2	2.37	11.24	0.35	3	0	0
12	5	5.95	10.23	0.34	4	4	0
13	3	2.63	10.00	0.70	0	0	0
14	3	2.72	4.56	0.55	0	0	0
15	3	3.75	14.10	0.32	0	0	4
16	4	3.17	5.62	0.74	0	0	1
17	5	6.47	11.20	0.28	0	0	1
18	5	6.25	11.13	0.31	0	3	0
19	4	4.83	11.31	0.34	0	3	1

Table 4.2: Project Ranking based on Monte-Carlo Simulation and the distribution-free approach

Project 1	**Project 2**	**Project 3**	**Monte-Carlo**	**Distribution-free**
$f(y_1) = f(y_2) = f(y_3) = f(y_4) = f(y_5) = 4$	$f(y_6) = f(y_7) = f(y_8) = f(y_9) = f(y_{10}) = 4$	$f(y_{15}) = f(y_{16}) = f(y_{17}) = f(y_{18}) = f(y_{19}) = 4$	Project 2	Project 2
$f(y_1) = f(y_3) = f(y_5) = f(y_7) = f(y_9) = 4$	$f(y_2) = f(y_4) = f(y_6) = f(y_8) = f(y_{10}) = 4$	$f(y_{10}) = f(y_{12}) = f(y_{14}) = f(y_{16}) = f(y_{18}) = 4$	Project 1	Project 1
$f(y_2) = f(y_3) = f(y_4) = f(y_5) = 5$	$f(y_7) = f(y_8) = f(y_9) = f(y_{10}) = 5$	$f(y_{16}) = f(y_{17}) = f(y_{18}) = f(y_{19}) = 5$	Project 2	Project 2
$f(y_3) = f(y_5) = f(y_7) = f(y_9) = 5$	$f(y_4) = f(y_6) = f(y_8) = f(y_{10}) = 5$	$f(y_{12}) = f(y_{14}) = f(y_{16}) = f(y_{18}) = 5$	Project 1	Project 1
$f(y_2) = f(y_3) = f(y_4) = 6.67$	$f(y_7) = f(y_8) = f(y_9) = 6.67$	$f(y_{16}) = f(y_{17}) = f(y_{18}) = 6.67$	Project 2	Project 2
$f(y_3) = f(y_5) = f(y_7) = 6.67$	$f(y_4) = f(y_6) = f(y_8) = 6.67$	$f(y_{12}) = f(y_{14}) = f(y_{16}) = 6.67$	Project 1	Project 1

Chapter 5: Summary, Conclusions and Extensions

Real-life transportation systems are subject to numerous uncertainties in their operation. Researchers have studied the resulting unreliability from different perspectives. In this dissertation we examine reliability at the network level. One of the measures to characterize network reliability is given by travel time reliability, which is defined as the probability that the (system) travel time remains below certain (acceptable) threshold levels. Existing reliability assessment (and optimization) techniques tend to be computationally intensive. In this dissertation we developed computationally efficient alternatives.

In Chapter 2 we presented a new travel time reliability assessment methodology assuming independent, random capacity variations and non-adaptive behavior. As such, the methodology is most useful when capacity uncertainty is caused by, for example, minor traffic incidents and in transportation planning problems where travelers are assumed to have settled in a long-term equilibrium pattern in face of these uncertainties (Lo and Tung, 2003). Chapter 2 complements parts of the work in Lo and Tung (2003) and Lo et al. (2006) in the sense that we do not rely on the CLT in deriving the PDF of the system travel time. The CLT oftentimes requires restrictive and hard to verify conditions (Feller, 1966). Numerical examples demonstrated that the CLT does not necessarily hold in the case of independent but not identically distributed normal road capacities, whereas the proposed methodology remains valid. Alternatively, the work in Chapter 2 can be seen as a complement to Clark and Watling (2005) who derived the

PDF for the travel time in a network characterized by uncertain demand, using a fundamentally different approach.

The proposed methodology has the following features:

- The methodology provides a numerical approximation of the PDF of the system-wide travel time.
- Any common continuous or discrete probability distribution can be used to model the uncertain capacities (with slight modifications in the discrete case).
- The methodology relies on the established computational efficiency of the fast Fourier transform.

Theoretical bounds on the approximation errors were formally derived, both for general probability distributions as well as for the special case of normally distributed capacities. These bounds give us important insights into the structure of the errors as well as how errors might be reduced. Unfortunately, these bounds tend to be too conservative in practice. Therefore, we followed the authoritative work by Abate and Whitt (1992), and proposed a successive refinement scheme to guarantee the accuracy of the approximations.

A numerical case study based on the assumption of normally distributed capacities demonstrated that the results of the methodology are consistent with intuition: increased demand levels, lower nominal capacities and higher capacity uncertainty shift the probability mass to the right, i.e, make higher travel times more likely. The numerical examples have shown that the resulting PDFs are skewed to the right with relatively fat right tails, an indication that the CLT is not applicable. In other words, the CLT might lead to overly optimistic conclusions on travel time reliability and it is therefore of crucial importance to rigorously justify the application of the CLT, which, except under specific assumptions (Lo et al., 2006) is not always mathematically tractable. The case study also demonstrated the computational efficiency of the approximation scheme. For instance, the PDF for the Sioux Falls network that consists of 76 links was constructed in about 21 seconds, excluding the accuracy checking procedure.

Despite the computational efficiency, it is always good practice to explore opportunities for further improvement, especially when analyzing large networks that consist of thousands of links. To this end, we have suggested various strategies to improve on the computation time that are, however, left as future work:

- Discarding links with small uncertainties and small volume-to-capacity ratios from the reliability analysis (which basically reduces the network size).
- The use of the non-uniform DFT to handle a diverse range of capacity variations in a network. From a numerical integration perspective, integrals associated with links with small uncertainties are substantially different from those characterized by large uncertainties. The former typically requires a finer computational grid in specific regions of the domain of integration.
- The use of higher order quadrature rules, e.g., the Simpsons rule. Higher order quadrature rules tend to achieve similar accuracy levels with a coarser computational grid, i.e, are less computationally intensive.

A final interesting extension of this work would be to incorporate some form of dependence in the capacities, or at least, measure the impact of such dependence on the PDF quantitatively. Unfortunately, we believe that such an extension is not feasible within the current Fourier framework. To see this, simply note that the fundamental result that underlies the proposed methodology is the convolution theorem. Without the assumption of statistical independence, the convolution theorem is no longer applicable. A fundamentally new approach is required.

In Chapter 3, we relaxed the common assumption that the exact probability distributions of the sources of uncertainty are known explicitly. In reality, these distributions may be unavailable (inaccurate) because we may have no (insufficient) data to calibrate the distribution. We presented a travel time reliability assessment technique based on probability inequalities. Instead of the specification of the probability distributions, the methodology only requires the specification of moments (up to order N) and a set of bounded intervals in which the random quantities are expected to reside. The price to pay for this relaxation is that we only obtain upper bounds on the unreliability function as opposed to exact probabilities of events (that are only as good as the underlying assumptions on the probability distributions). We also departed from previous modeling paradigms in that we directly worked with travel time rather than, for example, road capacities (we argue that it is much easier to answer the question what the variance of travel time is than what the variance of road capacity is). The only assumption is that the travel times are independent across links (Lo and Tung, 2003; Lo et al., 2006).

A numerical case study using the well-known Sioux Falls test network revealed a number of important properties of the bounds. First, the bounds were found to be nontrivial for the most interesting (in terms of reliability assessment) subset of the feasible region, i.e, the higher travel times. Second, first order moment-based bounds can potentially be significantly improved by the inclusion of higher order moments. The improvement is larger when the additional information is more informative. For example, a smaller second order moment (i.e, smaller variance) is more informative than larger values (i.e, larger variances). Third, the proposed methodology is extremely computationally efficient.

The introduction of the distribution-free paradigm is without doubt a major contribution to the transportation systems reliability literature. While we have developed bounds that are extremely useful (until now, not much – if anything – could be said about transportation system reliability given only information regarding moments and supports of random quantities), there is a substantial amount of work left for future research. For example, the sharpness of the bounds needs to be examined: are there travel times with given moments that attain the bounds? The independence assumption is consistent with the state-of-the-art in the current reliability assessment literature. However, it is straightforward to imagine situations where stochastic dependencies cannot be ignored (e.g., because of correlations in link flows). Fundamentally new statistical techniques are needed to account for such dependencies.

The final contribution of this dissertation (Chapter 4) is the introduction of a new network design model that aims at minimizing the unreliability of travel time (while virtually all existing network design models have the objective of minimizing system travel time). As opposed to the handful of existing publications in this area, the proposed model (RO-NDP) – which is based on the distribution-free reliability assessment technique developed in Chapter 3 – does not assume that exact probability distributions are known and is computationally feasible. While the model is primarily intended for situations where one is not able to specify exact distributions, we have demonstrated that it also has the potential to serve as a (computationally efficient) proxy for an existing simulation-based network design model that *does* require the specification of exact distributions (Chen et al., 2009). We showed that for the relatively small Nguyen-Dupuis test network the computation time can already reach 50 minutes, whereas the RO-NDP is much less computationally demanding due to the semi-analytical nature of the objective function. Results also suggested that the distribution-free reliability assessment methodology can be safely used in ranking capacity expansion projects that are aimed at improving the reliability of a transportation system. Numerical experiments demonstrated that network design solutions resulting from the RO-NDP model can be significantly different from capacity expansion decisions resulting from conventional network design models aiming at minimizing the (expected) system travel time. Hence, it would be interesting to develop bi-objective network design models that can provide a rational trade-off between the two objectives. The assumption of independence has been inherited from the reliability assessment technique developed in Chapter 3. The development of reliability assessment methodologies that are able to incorporate stochastic dependencies can thus also lead to network design models in which dependencies are accounted for. On the other hand, it will be interesting to investigate how well the current model can serve as a proxy for the case when dependencies are present. However, this is left as future work. A final extension of this work would be the formulation and analysis of NDPs that employs higher order moments in characterizing network reliability.

In line with the state-of-the-art in this area, in this dissertation the time-dependency of unreliability has not been accounted for. Clearly, in addition to the modeling of spatial dependencies, another critical long-term research effort would be to incorporate temporal dependencies in measuring and optimizing the reliability of transportation networks.

References

Abate, J., Whitt, W., 1992. The Fourier series method for inverting transforms of probability distributions. *Queueing Systems* 10, 5–88.

Abdel-Aty, M., Kitamura, R., Jovanis, P., 1995. Investigating effect of travel time variability on route choice using repeated measurement stated preference data. *Transportation Research Record: Journal of the Transportation Research Board* 1493, 39–45.

Abdulaal, M., LeBlanc, L.J., 1979. Continuous equilibrium network design models. *Transportation Research B* 13, 19–32.

Asakura, Y., Hato, E., Kashiwadani, M., 2003. Stochastic network design problem: an optimal link improvement model for reliable network. *In: Proceedings of the 1st INSTR*, Kyoto, Japan.

Asakura, Y., Kashiwadani, M., 1991. Road network reliability caused by daily fluctuation of traffic flow. *In: Proceedings of the 19th PTRC Summer Annual Meeting*, Brighton, 73–84.

Asensio, J., Matas, A. Commuters' valuation of travel time variability. *Transportation Research Part E* 44, 1074-1085.

Bard, J.F., 1998. Practical Bi-level Optimization algorithms and applications. Kluwer, Academic Publishers.

Bates, J., Polak, J., Jones, P., Cook, A., 2001. The valuation of reliability for personal travel. *Transportation Research Part E* 37, 191–229.

Bell, M.G.H., 2000. A game theory approach to measuring the performance reliability of transport networks. *Transportation Research B* 34, 533–546.

Bell, M.G.H., Cassir, C., 2002. Risk-averse user equilibrium traffic assignment: an application of game theory. *Transportation Research B* 36, 671–682.

Bell, M.G.H., Cassir C., Iida, Y., Lam, W.H.K., 1999. A sensitivity-based approach to network reliability assessment, *Proceedings 14th ISTTT*, Jerusalem, 283-300.

Bell, M.G.H., Iida, Y., 1997. *Transportation network analysis*, John Wiley & Sons, Chichester, UK.

Berdica, K., 2002. An Introduction to Road Vulnerability: What has been done, is done and should be done. *Transport Policy* 9, 117–127.

Billingsley, P., 1995. *Probability and measure*, John Wiley and Sons, Inc., New York.

Boyd, S., Vandenberghe, L., 2004. *Convex optimization*. Cambridge Univ. Press, Cambridge, U.K.

Briggs, W. L., Henson, V. E., 1995. *The DFT: An owner's manual for the discrete Fourier transform*, SIAM.

Brigham, E. O., 1988. *The fast Fourier transform and its applications*, Prentice Hall Signal Processing Series, Englewood Cliffs, NJ.

Burden, L. R., Faires, J. D., 2004. *Numerical analysis*, Brooks-Cole Publishing.

Bureau of Public Roads, 1964. *Traffic assignment manual*. U.S. Dept. of Commerce, Urban Planning Division, Washington D.C.

Chen, A, Kim J, Zhou Z, Chootinan P., 2007. Alpha reliable network design problem. *Transportation Research Record: Journal of the Transportation Research Board* 2029, 49-57.

Chen, A., Kim, J., Lee, S., Choi, J., 2009. Models and algorithm for stochastic network designs, *Tsinghua Science & Technology* 14, 341-351.

Chen, A., Subprasom, K., Ji, Z.W., 2003. Mean-variance model for the build–operate–transfer scheme under demand uncertainty. *Transportation Research Record: Journal of the Transportation Research Board* 1857, 93–101.

Chen, A., Yang, H., Lo, H.K., Tang, W.H., 1999. A capacity related reliability for transportation networks. *Journal of Advanced Transportation* 33, 183–200.

Chen, A., Yang, H., Lo, H.K, Tang, W.H., 2002. Capacity reliability of a road network: an assessment methodology and numerical results. *Transportation Research B* 36, 225–252.

Chernoff, H., 1952. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Annals of Mathematical Statistics* 23, 493-507.

Clark, S. D., Watling, D. P., 2005. Modelling network travel time reliability under stochastic demand. *Transportation Research B* 39, 119-140.

Cooley, J. W. & Tukey, J. W., 1965. An algorithm for the machine computation of the complex Fourier series, *Mathematics of Computation* 19, 297-301.

Dantzig, G. B., Harvey, R. P., Landsowne, Z. F., Robinson, D. W., Maier, S. F., 1979. Formulating and solving the network design problem by decomposition. *Transportation Research B* 13, 5-17.

Du, Z.P., Nicholson, A.J., 1997. Degradable transportation systems: sensitivity and reliability analysis. *Transportation Research B* 31, 225–237.

Dutt, A., Rokhlin, V., 1993. Fast Fourier transforms for non-equispaced data, SIAM Journal on Scientific Computing 14, 1368–1393.

Feller, W.,1966. *An introduction to probability theory and its applications*, Volume 2, John Wiley & Sons, Inc.

Frank, M., Wolfe, P., 1956. An algorithm for quadratic programming, *Naval Research Logistics Quarterly* 3, 95—110.

Friesz, T. L., Cho, H., Mehta, N., Tobin, R., Anandalingam, G., 1992. A simulated annealing approach to the network design problem with variational inequality constraints. *Transportation Science* 26, 18-26.

Goldberg, D.E., 1989. *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley, Reading, MA.

Goldberg, R. R., 1961. *Fourier transforms*, Cambridge University Press.

Heideman, M.T, Johnson, D. H., Burrus, C. S., 1985. Gauss and the history of the fast Fourier transform, *Archive for History of Exact Sciences* 34, 265-277.

Hoeffding, W., 1963. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58, 13–30.

Hoang, H. H., 1982. Topological optimization of networks: A Non-Linear Mixed Integer Model Employing Generalized Benders Decomposition. *IEEE Transactions on Automated Control* 27, 164-169.

Howell, K.B., 2001. *Principles of Fourier analysis* (Studies in Advanced Mathematics), CRC Press

Janson, B. N., 1995. Network design effects of dynamic traffic assignment. *Journal of Transportation Engineering* 121, 1-13.

Johnson, N.L., 1949. Systems of frequency curves generated by methods of translation. *Biometrika* 36, 149–176.

Karoonsoontawong, A., Waller, S.T., 2005. A comparison of system- and user-optimal stochastic dynamic network design models using Monte Carlo bounding techniques. *Transportation Research Record: Journal of the Transportation Research Board* 1923, 91-102.

Lam, T.C., Small, K.A., 2001. The value of time and reliability: measurement from a value pricing experiment. *Transportation Research Part E* 37, 231–251.

Lam, W.H.K., Shao, H., and Sumalee, A., 2008. Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply, *Transportation Research Part B* 42, 890-810.

LeBlanc, L. J., 1975. An algorithm for the discrete network design problem. *Transportation Science* 9, 183-199.

LeBlanc, L. J., and M. Abdulaal, 1979. An efficient dual approach to the urban road network design problem. *Computers and Mathematics with Applications* 5, 11-19.

LeBlanc, L. J., and M. Abdulaal, 1984. A comparison of the user-optimum versus system-optimum traffic assignment in transportation network design. *Transportation Research B* 18, 115-121.

LeBlanc, L. J., Boyce, D., 1986. A bi-Level programming algorithm for the exact solution of the network design problem with user-optimal traffic flows, *Transportation Research B* 20, 259-265.

Lemp, J., Kockelman, K., 2010. Empirical investigation of the continuous logit for departure time choice using Bayesian methods. *Transportation Research Record: Journal of the Transportation Research Board*, forthcoming.

Li, Y., Waller, S. T., Ziliaskopoulos, A. K., 2003. A decomposition scheme for system-optimal dynamic traffic assignment Models. *Networks and Spatial Economics* 3, 441-455.

Liu, H.X., Recker, W., Chen, A., 2004. Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. *Transportation Research Part E* 38, 435-453.

Lo, H., Luo, X.W., Siu, B., 2006. Degradable transport network: Travel time budget of travelers with heterogeneous risk aversion. *Transportation Research Part B* 40, 792–806.

Lo, H.K., Tung, Y.K., 2003. Network with degradable links: capacity analysis and design. *Transportation Research Part B* 37, 345–363.

Madansky, A., 1959. Bounds on the expectation of a convex function of a multivariate random variable, *The Annals of Mathematical Statistics* 30, 743-746.

Magnanti, T. L., Wong, R. T., 1984. Network design and transportation planning: models and algorithms. *Transportation Science* 18, 1-55.

Marcotte, P., 1983. Network optimization with continuous control parameters, *Transportation Science* 17, 181-197.

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

Mouskos, K., 1991. *A Taboo-search based approach for network design.* Ph.D. dissertation, The University of Texas at Austin.

Ng, M.W., Waller, S.T., 2009a. Reliable system optimal network design: a convex mean-variance-type model with implicit chance constraints, *Transportation Research Record: Journal of the Transportation Research Board* 2090, 68-74.

Ng, M.W., Waller, S.T., 2009b. A dynamic route choice model in face of uncertain capacities, *Networks and Spatial Economics*, in review.

Ng, M.W., Waller, S.T., 2009c. The evacuation optimal network design problem: model formulation and comparisons, *Transportation Letters: The International Journal of Transportation Research* 1, 111-119.

Ng, M.W., Lin, D-.Y, Waller, S.T., 2009. Optimal long-term infrastructure maintenance planning accounting for traffic dynamics. *Computer-Aided Civil and Infrastructure Engineering* 24, 459- 469.

Ng, M.W., Kockelman, K., Waller, S.T., 2010. Relaxing the multivariate normality assumption in the simulation of transportation system dependencies, *Transportation Letters: The International Journal of Transportation Research*, in press.

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

Patriksson, M., Rockafellar, R. T., 2002. A mathematical model and descent algorithm for bilevel traffic management. *Transportation Science* 36, 271-291.

Press, W. H., Tuekolsky, S. A., Vetterling, W. T., Flannery, B. P., 2007. *Numerical recipes: the art of scientific computing*, Cambridge University Press.

Ross, S., 2002. *Introduction to probability models*. Academic Press.

Seaman, J., Odell, P., 1985. On Goldstein's variance bound. *Advances in Applied Probability* 17, 679-681.

Shao, H., Lam, W.H.K, Meng, Q., Tam, M.L., 2006. Demand-driven traffic assignment problem based on travel time reliability. *Transportation Research Record: Journal of the Transportation Research Board* 1985, 220–230.

Sheffi, Y., 1985. *Urban transportation networks: equilibrium analysis with mathematical programming methods*, Prentice-Hall, Englewood Cliffs, NJ.

Siu, B.W.Y., Lo, H.K, 2008. Doubly uncertain transportation network: Degradable capacity and stochastic demand. *European Journal of Operational Research* 191, 166–181.

Sumalee, A., Kurauchi, F., 2006. Network capacity reliability analysis considering traffic regulation after a major disaster. *Networks and Spatial Economics*, 6, 205-219.

Sumalee, A., Watling, D. P., 2003. Travel time reliability in a network with dependent link modes and partial driver response. *Journal of Eastern Asia Society for Transportation Studies*, 5, 1687-1701.

Sumalee, A., Watling, D.P., 2008. Partition-based approach for estimating travel time reliability with dependent failure probability. *Journal of Advanced Transportation*, 42, 213-238.

Sumalee, A., Watling, D. P., Nakayama, S., 2006. Reliable network design problem: case with uncertain demand and total travel time reliability. *Transportation Research Record: Journal of the Transportation Research Board* 1964, 81–90.

Suwansirikul, C., Friesz, T. L., Tobin, R. L., 1987. Equilibrium decomposed optimization: a heuristic for the continuous equilibrium network design problem. *Transportation Science* 21, 254–263.

Szeto, W.Y., O'Brien, L, O'Mahony, M., 2006. Risk-averse traffic assignment with elastic demands: NCP formulation and solution method for assessing performance reliability. *Networks and Spatial Economics* 6, 313–332.

Ushakov, N.G., 1999. *Selected topics in characteristic functions* (Modern Probability and Statistics), VSP B.V.

Van Loan, C., 1992. *Computational frameworks for the fast Fourier transform*, SIAM.

Wakabayashi, H., Iida, Y., 1992. Upper and lower bounds of terminal reliability of road networks: an efficient method with Boolean algebra. *Journal of Natural Disaster Science* 14, 29-44.

Waller, S. T., Schofer, J. L., Ziliaskopoulos, A. K., 2001. Evaluation with traffic assignment under demand uncertainty. *Transportation Research Record: Journal of the Transportation Research Board* 1771, 69–74.

Waller, S.T., Mouskos, K.C., Kamaryiannis, D., Ziliaskopoulos, A.K., 2006. A linear model for continuous network design problem. *Computer-Aided Civil and Infrastructure Engineering* 21, 334-345.

Waller, S. T., Ziliaskopoulos, A. K., 2001. Stochastic dynamic network design problem. *Transportation Research Record: Journal of the Transportation Research Board* 1771, 106-113.

Wardrop, J. G., 1952. Some theoretical aspects of road traffic research. *In: Proceedings of the Institute of Civil Engineers*, Pt. II, Volume 1, 325–378.

Watling, D.P., 2002. A second order stochastic network equilibrium model. *Transportation Science* 36 (2), 149–183.

Yang, H., Bell., M. G. H., 1998. Models and algorithms for road network design: a review and some new developments. *Transport Reviews* 18(3), 257-278.

Yin, Y., Ieda, H., 2001. Assessing performance reliability of road networks under non-recurrent congestion. *Transportation Research Record: Journal of the Transportation Research Board* 1771, 148–155.

Yin, Y., Lam, W.H.K., Ieda, H., 2004. New technology and the modelling of risk taking behaviour in congested road networks. *Transportation Research Part C* 12, 171–192.

Zhou, Z., Chen, A, 2008. Comparative analysis of three user equilibrium models under stochastic demand, *Journal of Advanced Transportation* 42, 239-263.

Vita

Man Wo Ng (also written as ManWo Ng) was born and raised in Rotterdam, The Netherlands. In 2002, he received his Bachelor of Science in Mechanical Engineering from Delft University of Technology. Upon graduation, he entered the Applied Mathematics program at the same school and obtained his Master of Science in 2005 (cum laude). In Fall 2005, he entered the Graduate School at The University of Texas at Austin, pursuing his doctoral degree, where he has been the recipient of two of the most prestigious fellowships at the University (University Pre-emptive Fellowship and University Continuing Fellowship).

Email: mn562.mail@gmail.com

This dissertation was typed by the author.