An Algorithm for Reliable Shortest Path Problem with Travel Time Correlations

A Thesis

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Yufeng Zhang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Alireza Khani

February, 2019

Acknowledgements

I cannot begin to express my deepest gratitude to my academic advisor, Dr. Alireza Khani, who has been giving me relentless support since the first day I joined the Transit Lab. The completion of the thesis would not have been possible without his patience, knowledge, and encouragement. He has shown me by his passionate attitude and diligence what a good researcher and person should be.

I would also like to extend my thanks to the rest of my committee, Professor Gary Davis and Professor Zhi-Li Zhang, for their invaluable advice and practical suggestions. My thanks also go to those anonymous reviewers for their insightful comments and contribution to the work.

I would also like to thank my fellow lab mates with whom I discussed research questions, conducted research projects, traveled for conferences, and happily talked about stuff beyond our papers. I must also thank my friends here in the U.S. and those in China for their emotional support, endless encouragement and remote caring.

Last but not least, my profound gratitude goes to my Mom and Dad for their unconditional love and unparalleled support that are always with me for each and every day. Nobody has ever been more important than them in the pursuit of my academic goals and I dedicate this thesis to them.

Abstract

Reliable shortest path (RSP) problem reflects the variability of travel time and is more realistic than standard shortest path problem which considers only the average travel time. This thesis describes an algorithm for solving the mean-standard deviation RSP problem considering link travel time correlations. The proposed algorithm adopts the Lagrangian substitution and covariance matrix decomposition technique to deal with the difficulty resulting from non-linearity and non-additivity of the Mixed Integer Non-Linear Program (MINLP). The problem is decomposed into a standard shortest path problem and a convex optimization problem whose optimal solution is proved and the Lagrangian multipliers ranges are related to the eigenvalues of the covariance matrix to further speed up the algorithm. The complexity of the original problem is notably reduced by the proposed algorithm such that it can be scaled to large networks. In addition to the sub-gradient Lagrangian multiplier updating strategy integrated with projection, a novel one based on the deep-cut ellipsoid method is proposed as well. Numerical experiments on large-scale networks show the efficiency of the algorithm in terms of relative duality gap and computational time. Besides, there is evidence showing that, though having longer computational time, the ellipsoid updating method tends to obtain better solutions compared with the sub-gradient method. The proposed algorithm outperforms the existing one-to-one Lagrangian relaxation-based RSP algorithms and the exact Outer Approximation method in the literature.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

In transportation, shortest path algorithms have wide applications in problems such as vehicle routing, traffic assignment, and network design. To some extent, the shortest path problem is the most fundamental component and lays the foundation of many complex and advanced transportation studies.

The most widely applied shortest path algorithm only solves the problem with respect to the mean travel time, which ignores the variability of the traffic conditions in real networks. However, in reality, travel time is not constant but is subject to change due to congestions, demand variations, traffic accidents, and so on. Mutable link travel times also have interactions with each other. Thus, path travel time reliability is significantly influenced by the link travel time variations and correlations. Nowadays, with the requirements of real-time travel information and travel models shifting from static to dynamic, the shortest path problem with travel time variability, or the so-called "reliable shortest path problem" (RSP), has started to receive more attention.

In the literature, the stochasticity of travel time is usually described either with prior travel time distributions or with the means and standard deviations of link travel times. Depending on how travel time stochasticity is embodied, reliable shortest path problem is addressed with two mainstream approaches - maximizing the on-time arrival probabilities or path reliability and minimizing the path travel time. Approaches in the first category explicitly deal with travel time distributions for development of probability-based conditions, in contrast, methods that belong to the second category are more or less characterized by optimization techniques. In our study, we adopt the latter approach using mean and standard deviation to formulate an optimization problem, which is straightforward in the sense that data required for the problem can be directly obtained from real data and relieves the need for either finding a good distribution or making assumptions on variable travel times.

The RSP problem in this study is addressed as minimizing the sum of mean and standard deviation of path travel time. Due to the non-additive and non-linear property, the problem cannot be solved by standard labeling algorithms because Bellman's optimality principle does not hold any longer. Including the correlation between link travel times significantly increases the complexity of the problem due to the variable dimension increase and the inclusion of the quadratic term of more cross-multiplications of integer variables.

These difficulties make the problem harder yet intriguing and interesting. The goal of this study is to find the one-to-one reliable shortest path, defined as the path with the minimum sum of mean and standard deviation travel time, taking into account link travel time correlations.

1.2 Literature Review

Reliable shortest path problem can be addressed from various perspectives. In this section, a summary of problem formulations and solution approaches to the reliable shortest path problem is provided, together with research contributions and their limitations.

One mainstream approach to tackling RSP problem is to define the stochasticity of path travel times by the summation of average travel times and standard deviation travel times with or without correlations. Uchida and Iida (1993) are believed to be the pioneers of studying the mean-standard deviation defined shortest path problem [1].

To model the correlated link travel times defined by the aforementioned objectives, the following mentioned work contributed to solution algorithms. Having a mean travel time minimization problem as part of the optimization problem formulated, it is very natural for the researchers to make use of the standard shortest path problem algorithm for solving the RSP problem. In this category, the solution algorithms are characterized by solving a standard shortest path sub-problem. Xing and Zhou (2011) modeled the correlations of stochastic link travel times by dealing with samples of historical traffic data [2]. Formulating a relatively simpler form of the correlations in their objective function, they proposed a Lagrangian relaxation approach to solve the problem. Their dual function can be decomposed into three sub-problems: a standard shortest path problem, a convex optimization problem, and a constrained univariate concave optimization problem, all of which can be easily solved. Along similar lines, Zeng et. al. (2015) proposed a Cholesky decomposition based Lagrangian relaxation method for finding RSP [3]. They constructed a concave sub-problem by decomposing the variance-covariance matrix using Cholesky decomposition technique, which differs from Xing and Zhou's implicit covariance method. However, unlike labeling algorithms for standard shortest path finding a shortest path tree, both of these two approaches only gives the shortest path between a single pair of origin and destination in each run. Besides, the duality gap of their method may result from the fact that a non-convex sub-problem exists in their formulations. The most recent work related to mean-standard deviation RSP problem is Zhang et al.'s study (2017) [4]. Using the Lagrangian relaxation based method as is used by Xing and Zhou, they adopted a Lagrangian relaxation method based on Cholesky decomposition, and proposed two Lagrangian multipliers updating methods. It is shown that their methods converge to the optimal solution with smaller duality gaps compared with Xing and Zhou's method. The method proposed by Khani and Boyles (2015), though designed for an uncorrelated link travel time case, is an exact algorithm for solving an RSP without link travel correlations [5]. They showed that potentially optimal solutions to the mean-variance problem are potentially optimal to the mean-standard deviation problem. Based on this finding, they proposed an exact one-to-all solution algorithm for the mean standard deviation problem. Numerical experiments showed that their algorithm took less computational time in comparison with other algorithms in the literature. Nonetheless, the restricted applicability only to networks with uncorrelated link travel times was a limitation of their study. Having mentioned the mean-variance defined RSP problem, it is worth noting that a recent study by Rostami et al.(2016) on the mean-variance RSP problem developed a branch-and-bound algorithm in exploring the upper and lower bounds of the shortest path [6]. The solution algorithm works by iteratively branching on the neighbors of the current node and solving shortest path sub-problems. In addition, the study showed that restricting correlations within adjacent/consecutive links does not ease the complexity of the problem. Following the results of this work, Hu and Sotirov(2017) studied some special cases of the mean-variance RSP problem that can be solved in polynomial time [7].

A one-to-all exact algorithm for correlated link travel times is proposed by Shahabi et al. (2013, 2015) [8, 9]. The mixed-integer nonlinear programming (MINLP) was reformulated into a mixed-integer conic quadratic problem. The Outer Approximation (OA) based algorithm iterates between solving a mixed integer linear problem (MILP) and a continuous non-linear problem (NLP). The conic constraint is approximated by the sole inclusion of the elements of the covariance matrix that make up the shortest paths, making it applicable to huge covariance matrix cases but less computationally expensive. Under the problem structure, the NLP sub-problem can be easily evaluated by the solutions obtained from MILP. Shahabi et al. (2015) later developed a robust optimization strategy for uncertain link travel time [9]. This formulation differs from their previous work in the expression of the uncertainty of travel times, but outer approximation method still forms the backbone of their solution algorithm. Table 1.1 summarizes the research work where the reliable shortest path problem is defined as a combination of mean travel time and travel time standard deviation (or variance).

Research Work	**With correlations?**	**Solution algorithm**	**Methodology summary**
Khani et. al. (2015)	No	Exact; Labelling algorithm; One-to-all	Set up the relationship between uncorrelated travel time shortest paths of mean-variance and mean-standard deviation problems; Developed an efficient labelling algorithm in finding efficient frontier in additive mean-variance problem.
Shahabi et al. (2013)	Yes	Exact; Outer approximation; One-to-all	Proposed the outer approximation method where in the master mixed integer problem cutting planes are only generated using existing shortest paths which avoids the computation burden in large covariance matrix.
Xing et. al. (2011)	Yes	Lagrangian substitution; One-to-one	Proposed Lagrangian substitution method; Decomposed the problem into standard shortest path problem, a univariate convex problem and a univariate concave problem; Proposed sub-gradient method in updating Lagrangian multipliers.
Zeng et. al.(2015)	Yes	Lagrangian substitution; One-to-one	Proposed Cholesky decomposition method to deal with covariance matrix, using variable splitting techniques the problem is decomposed into three sub-problems of the same types as Xing's.
Zhang et. al (2017)	Yes	Lagrangian substitution; One-to-one	Applied Cholesky decomposition method, the problem is decomposed into a standard shortest path problem and a convex problem; Lagrangian multipliers updating method using sub-gradient projection and by solving a quadratic constrained convex problem were brought forward.
Proposed method	Yes	Lagrangian substitution; One-to-one	Proposed eigendecomposition method to transform the problem into a standard shortest path problem and a convex problem; Proposed Lagrangin multipliers updating methods using sub-gradient and deep-cut ellipsoid method with projections.

Table 1.1: Algorithm comparisons of closely related studies in the literature

The variability of travel times in the preceding category is embodied in their variances and covariances. On the other hand, the RSP problem can be modeled by utilizing the travel time distributions properties. In this line of research, concepts such as “travel time budget”, “on-time arrival probability”, “α-reliable path” are proposed. Normal distribution of path travel times assumption is commonly adopted by researchers to derive optimal conditions and solution algorithms. Seshadri and Srinivasan (2010) developed a sufficient condition for the optimal path. Combined with Monte-Carlo path generation approach, K-shortest expected travel time paths were found to find the optimal solution [10]. Chen et al. (2012, 2013) studied a RSP problem where only neighboring links' travel times are correlated. They developed dominant conditions and an algorithm based on these conidtions [11,12], and later a k-reliable shortest path problem was studied which extended the idea of deviation path algorithm in K shortest paths [13]. In dominance theory approach, studies that focus on developing stochastic dominance theory for reliable shortest paths received attention as well [14,15]. Fan et al. (2005) proposed a method to maximize the probability of arriving to the destination on time or sooner based on a model applying Bellman's optimality principle [16]. Later in the work of Nie and Fan (2006), they improved the computational efficiency of the stochastic on-time arrival solution algorithm [17]. Along similar lines, Nie and Wu (2009) provided a label-correcting method and extended the problem to time-dependent traffic network [18]. Reliable shortest path problems are also studied with a link to route choice model and traffic assignment problem in these studies, Lo et al., Walting , Wang et al. and Shao et al.. Lo et al. (2006) brought up a concept named “travel time budget” and defined a within travel budget reliability to reflect the probability that actual travel time is within travel time budget [19]. Walting (2006) addressed the travelers' aversion of late arrival for uncertain travel in a user equilibrium traffic assignment study [20]. The last two studies were followed by the research on a bi-objective UE problem, where travelers were assumed to minimize their expected travel time and travel time budget at the same time [21]. Shao et al. (2006) formulated a Variation Inequality problem for a traffic assignment problem where travel time reliability is considered and is assumed to be normally distributed [22].

Among other solution approaches, some heuristic and simulation-based algorithms were also explored in the literature [23–25]. Interested readers are referred to these paper for details.

1.3 Proposed Method

There are two main challenges for solving the mean-standard deviation defined RSP problem. One is dealing with link travel time correlations, albeit there is an exact solution algorithm for mean-standard deviation problem without correlations [5]. The other concern is the computational complexity and efficiency of the solution algorithm, because efficient algorithms that can generate solutions fast are always favored for solving real traffic problems. Our work aims to tackle these two significant difficulties in RSP problem.

Variance-covariance matrix decomposition and Lagrangian substitution are the two main breakthroughs that greatly ease the complexity of the problem. Utilizing variable substitution method of Lagrangian relaxation and decomposed matrix, we introduced an auxiliary variable to the problem and by dualizing the linking constraints, the original mixed integer quadratic problem is decomposed into two separable sub-problems. One sub-problem is a standard shortest path problem, and the other is a convex problem of the auxiliary variable. Solving them separately is much easier than solving the original one. In addition, we proved the optimal solution to the the second sub-problem which further reduces computational costs. This research incorporates the efficient standard labeling algorithm into the proposed solution algorithm for the RSP problem, which greatly enhances the algorithm efficiency. The contribution of our work is briefly presented below.

Novel decomposition approach. Matrix decompostion method, such as Cholesky decompositon, has been used in the literature in solving RSP problems [3,4,8]. It is believed, however, that the eigen-decomposition technique which is studied in this thesis has not been used in the same research field. The proposed eigen-decomposition method relates the Lagrangian multipliers and the eigenvalues of travel time correlation matrix together. Eigen-decomposition is a more general method in that a Cholesky decomposition can be easily converted from the eigendecomposition.

Sub-problem convexity. Convexity of the resulting problems has been proved and the optimal value and solution of the sub-problem are found. This further reduces computational time without relying on optimization solver package for getting the optimal solution. Furthermore, the ranges that Lagrangian multipliers should vary within are exploited and related to the the eigenvalues of the matrix, and a projection technique is proposed in the Lagrangian multiplier updating method.

Computational efficiency. Comparisons with the existing Outer Approximation method [8] and one-to-one Lagrangian substituion solution algorithms, namely the sampling based method [2] and sub-gradient projection based method [4], were carried out. Tests reveals that the proposed Lagrangian substitution based sub-gradient updating method takes at most about 1.5 seconds for Barcelona Network and about 7 seconds by deep-cut ellipsoid method. Comparing these figures with the OA method, the computational time decrease at least 95% on average. It is also shown that our algorithm outperforms sampling based method in terms of duality gap. Comparing with the sub-gradient projection based method, it is found that our algorithms converge to a smaller gap, and is faster in terms of computational time.

Deep-cut ellipsoid updating strategy. Another contribution of this work is that we compared and explored two Lagrangian multiplier updating methods, the sub-gradient method and deep-cut ellipsoid method. Sub-gradient method enjoys its popularity due to its extreme simplicity and ease of implementation. However, it comes at the cost of a potentially bad updating direction [26]. Our results reveal evidence on this and show that even though deep-cut ellipsoid method usually takes longer to convergence, it is capable of giving better solutions, whereas sub-gradient method may converge to a local optimum.

The remainder of the thesis is organized as follows. Chapter 2 formulates the minimum mean-standard deviation path problem with link travel time correlations. Chapter 3 provides the detailed Lagrangian substitution based solution algorithm and two Lagrangian updating methods — the sub-gradient and deep-cut ellipsoid method. Computational experiment results on an illustrative example and large-scale networks, and comparisons with existing RSP solution algorithms are summarized in Chapter 4, together with evaluations of the proposed algorithm. Finally, Chapter 5 concludes the thesis by providing major contributions, limitations of this study and future work.

Chapter 2

Problem Statement

This chapter first introduces notations that are used throughout the thesis and proposes the mathematical formulation of the mean-standard deviation form of RSP problem with travel time correlations. Then, the reformulation of the RSP problem is presented.

2.1 Notation Preliminaries

Problem-related

$\mathbb{N}$: Set of nodes, $|\mathbb{N}| = n$

$\mathbb{A}$: Set of links, $|\mathbb{A}| = m$

i, j: Node index with r representing the origin and s representing the destination, $i, j \in \mathbb{N}$

ij, kl: Link index, $ij, kl \in \mathbb{A}$

c_{ij}: Avereage link travel time of link ij

c: m-dimensional link travel time vector

σ_{ij}: Standard deviation link travel time of link ij

$Cov(ij, kl)$: Covariance of travel times of link ij and link kl

Σ: $m \times m$ variance-covariance matrix of link travel times

x_{ij}: Binary decision variable, equals 1 if link ij is on the shortest path connecting origin and destination

x: m-dimensional decision variable vector

y: m-dimensional auxiliary variable vector

η: Reliability coefficient

Algorithm-related

Z: Objective value for the RSP problem with link time correlations

Z_{UB}: Upper bound of the objective value Z

Z_{LB}: Lower bound of the objective value Z

μ: m-dimensional Lagrangian multiplier vector

$L(x, y, \mu)$: Lagrangian function of decision variables and Lagrangian multipliers

$L(\mu)$: Lagrangian dual function

$L_x(\mu)$: Lagrangian dual sub-problem of x

$L_y(\mu)$: Lagrangian dual sub-problem of y

$L_x^\star(\mu)$: Optimal value of $L_x(\mu)$ for a given μ

$x^\star$: Optimal solution of $L_x(\mu)$ for a given μ

ϵ: Duality gap threshold

s: Iteration number index

K: Maximum iteration number

2.2 Problem Formulation

The RSP problem to be addressed is defined as a mean-standard deviation shortest path problem with correlated link travel times.

Let $G(\mathbb{N}, \mathbb{A})$ represent a directed transportation network. For each link ij, its travel time is stochastic, which is characterized by its average link travel time c_{ij} and the standard deviation travel time, σ_{ij}. Let x_{ij} be a binary variable denoting the decision of whether link ij is included in the reliable shortest path from the origin r to the destination s. x_{ij} takes the value of 1 if and only if ij is a component of the reliable shortest path, otherwise, it takes the value of 0. Thus, a mean-standard deviation RSP problem which does not consider link travel time correlations can be formulated as a minimization program with the objective function,

$$\sum_{ij\in\mathbb{A}} c_{ij}x_{ij} + \eta\sqrt{\sum_{ij\in\mathbb{A}} \sigma_{ij}^2 x_{ij}}, \tag{2.1}$$

and constraints on decision variables of flow conservation, $\sum_{j:ij\in\mathbb{A}} x_{ij} - \sum_{j:ji\in\mathbb{A}} x_{ji} = v$, where $v = 1$ if $i = r$, $v = -1$ if $i = s$, and otherwise $v = 0$. In the formulation, η captures the risk aversion towards travel time, which we will refer to as "reliability coefficient" in the sequel. We treat this parameter as a known constant, interested readers can find more on parameter calibration from the work [27, 28].

Compared with a standard shortest path problem, where the objective function only contains the term that decisions are only involved in average path travel time, the RSP problem increases its complexity in that it introduces a non-linear term that captures the variability of path travel time. Non-additivity of the RSP problems makes the Bellman's optimality condition not valid any longer, which forms the backbone of solution algorithms of the standard shortest path problem.

An RSP problem defined above assumes that there are no interactions between link travel times. However, this may not be true in the real world. The links adjacent to a heavily congested link, for example, are likely to have longer travel times due to the higher density of surrounding links. Thus, a more realistic model would be considering the travel time correlations among links. To extend the preceding defined problem to a more general case, we formulate the following objective function, where $Cov(ij, kl)$ denotes the link travel time covariance of link ij and kl.

$$\sum_{ij\in\mathbb{A}} c_{ij}x_{ij} + \eta\sqrt{\sum_{ij\in\mathbb{A}} \sigma_{ij}^2 x_{ij} + \sum_{ij,kl\in\mathbb{A}} Cov(ij, kl)x_{ij}x_{kl}}. \tag{2.2}$$

Noticing that the decision variables are binary, a formulation of the RSP problem in a compact form is proposed as follows:

$(\mathbb{P})$

$$\text{Min} \quad Z = c^T x + \eta\sqrt{x^T \Sigma x} \tag{2.3}$$

$$\text{s.t.} \quad \sum_{j:ij\in\mathbb{A}} x_{ij} - \sum_{j:ji\mathbb{A}} x_{ij} = \begin{cases} -1 & i = r \\ 0 & i \in \mathbb{N} - \{r, s\} \\ 1 & i = s \end{cases} \tag{2.4}$$

$$x_{ij} \in \{0, 1\} \tag{2.5}$$

The RSP problem with correlated link travel times formulated in $\mathbb{P}$ is a mixed integer nonlinear problem. The objective function is a combination of the standard shortest path problem (minimizing the mean travel time) and a variance-covariance matrix characterizing the path travel time variability. The complexity of the problem lies in that the efficient labeling algorithms cannot be used to find the optimal path due to the failure of Bellman's optimality conditions. A MINLP is typically solved by outer approximation techniques, but in this study, noting that the variance-covariance matrix, Σ, is always symmetric and positive semi-definite, exploring and making the best of this property makes it possible to use methods like Lagrangian relaxation.

2.3 Problem Reformulation

Given a square matrix, the decomposition technique, eigendecomposition, can be applied to decompose the matrix into a diagonal matrix, Λ, whose elements are eigenvalues, and a matrix V whose columns are corresponding normalized eigenvectors. The decomposition takes the following form:

$$\Sigma = V\Lambda V^{-1}. \tag{2.6}$$

A variance-covariance matrix is always positive semi-definite [29]. An imporant property of a positive (semi-)definite matrix is that it can be decomposed in such a way that its eigenvectors are pairwise orthogonal [30, 31]. This property makes it possible to store all its eigenvectors in one orthogonal matrix, and this implies that $V^T = V^{-1}$. Thus, the eigendecomposition for the variance-covariance matrix, Σ, can be expressed equivalently as the following:

$$\Sigma = V\Lambda V^{T}. \tag{2.7}$$

Substituting Σ with Λ and V and defining a new continuous auxiliary variable y, the objective function defined in $\mathbb{P}$ is reformulated as equation (2.8) with a link constraint defined in equation (2.11). Together with the constraint sets equation (2.4) and (2.5), we get an equivalent RSP problem $\mathbb{P}'$.

$(\mathbb{P}')$

$$\text{Min} \qquad c^T x + \eta \sqrt{y^T \Lambda y} \tag{2.8}$$

$$\text{s.t.} \qquad \sum_{j:ij\in\mathbb{A}} x_{ij} - \sum_{j:ji\in\mathbb{A}} x_{ij} = \begin{cases} -1 & i = r \\ 0 & i \in \mathbb{N} - \{r, s\} \\ 1 & i = s \end{cases} \tag{2.9}$$

$$x_{ij} \in \{0, 1\} \tag{2.10}$$

$$y = V^T x \tag{2.11}$$

In the following chapter, a solution algorithm that works directly to find the optimal solution for the problem $\mathbb{P}'$ is proposed, the solution of which is also the optimal solution for the problem $\mathbb{P}$.

Chapter 3

Matrix Decomposition-based Lagrangian Substitution Solution Method

This chapter first introduces the construction of a Lagrangian relaxation problem, then explores the convexity of the sub-problem, and discusses two Lagrangian multiplier updating strategies, and finally presents the complete solution algorithm for the link travel time correlated RSP problem.

3.1 Lagrangian Relaxation Problem Construction

Lagrangian relaxation is a commonly adopted technique to solve constrained optimization problems ([2, 32, 33]). By introducing Lagrangian multipliers, the method imposes penalties when inequality constraints are violated, hence forces improvement of the optimal value in the direction in accordance with the constraints.

Defining Lagrangian multipliers μ for the equality constraint, equation (2.11), and moving them to the objective function, we construct a new Lagrangian function, $L(x, y, \mu)$:

$$L(x, y, \mu) = c^T x + \eta\sqrt{y^T \Lambda y} + \mu^T (V^T x - y). \tag{3.1}$$

Now, we can construct the Lagrangian dual function, $L(\mu)$ which is a function of the Lagrangian multiplier μ. The Lagrangian dual function is rearranged according to decision variables, and together with the other two original constraints the complete Lagrangian dual function is presented as following:

$$L(\mu) = \inf_{x,y}\{c^T x + \eta\sqrt{y^T \Lambda y} + \mu^T (V^T x - y) : (2.4)(2.5)\} \tag{3.2}$$

$$= \inf_{x,y}\{(c + V\mu)^T x + \eta\sqrt{y^T \Lambda y} - \mu^T y : (2.4)(2.5)\} \tag{3.3}$$

$$= L_x(\mu) + L_y(\mu) \tag{3.4}$$

Where, $L_x(\mu)$ and $L_y(\mu)$ are defined as follows:

$$L_x(\mu) = \inf_x \left\{ (c + V\mu)^T x \,\middle|\, \sum_{j:ij \in \mathbb{A}} x_{ij} - \sum_{j:ji \in \mathbb{A}} x_{ij} = \begin{cases} -1, & \text{if} \quad i = r \\ 0, & \text{if} \quad i \in \mathbb{N} - \{r, s\} \\ 1, & \text{if} \quad i = s \end{cases}, \quad x_{ij} \in \{0, 1\} \right\} \tag{3.5}$$

$$L_y(\mu) = \inf_y \{\eta \sqrt{y^T \Lambda y} - \mu^T y\} \tag{3.6}$$

After regrouping the variables, the Lagrangian dual function can be regarded as a combination of two new sub-problems equation (3.5) and (3.6). The first sub-problem, $L_x(\mu)$, is a standard shortest path problem considering fixed penalties on some of the links, which can be solved efficiently with labeling algorithms. The penalty is determined by the eigenvector matrix V and the current Lagrangian multipliers μ. The other sub-problem, $L_y(\mu)$, is an unconstrained problem of auxiliary variable y. In the next subsection, we will prove the convexity and show the solution for the problem.

Before the convexity proof, we present the Lagrangian dual problem for the sake of completeness of solution algorithm. The Lagrangian dual problem is important in Lagrangian relaxation method because its solution provides a lower bound to the primal problem. A series of increasing lower bounds approaches the primal optimum gradually. With decreasing upper bounds, the difference between lower bound and upper bound decreases until an optimal solution is found.

The dual problem is a constrained maximization problem, with decision variables being the Lagrangian multipliers. Interested readers are referred to the Lagrangian relaxation tutorial by Guignard [34]. They referred to this Lagrangian relaxation method as dualizing "linking constraint", where the linking constraint refers to the constraint $V^T x = y$ that were introduced to link the original variables x and auxiliary variables y. Note that the Lagrangian multipliers do not have to be non-negative since the linking constraint is an equality.

($\mathbb{P}2$)

$$\max_{\mu \in R^m} \quad L(\mu) = L_x(\mu) + L_y(\mu) \tag{3.7}$$

3.2 Sub-problem Convexity and Solution Proof

This section first proves the convexity of the second sub-problem, gives its optimal solution and also develops the ranges that the Lagrangian multipliers should vary within so as to obtain optimal value.

Lemma 1 *Function $h(y) = \eta\sqrt{y^T \Lambda y} - \mu^T y$ is convex.*

To prove the convexity of the function $g(y)$, it suffices to prove the convexity of $f(y) = \sqrt{y^T \Lambda y}$ with the default setting that $\eta > 0$. Take any two vectors $u, v \in R^n$ and a scalar $k \in R$, then the following equation always holds.

$$f(k) = \sqrt{(u+kv)^T \Lambda (u+kv)} = \sqrt{(v^T \Lambda v k^2 + 2v^T \Lambda u k + u^T \Lambda u)}$$

Recall that the matrix Λ is a diagonal matrix, whose elements are the eigenvalues of the variance-covariance matrix, so Λ is also a PSD matrix. Viewing the above as a function with respect to variable k, thus the inequality $v^T \Lambda v k^2 + 2v^T \Lambda u k + u^T \Lambda u \geq 0$ holds for any k. And this indicates that $(v^T \Lambda u)^2 - (v^T \Lambda v)(u^T \Lambda u) \leq 0$.

Then take the second derivative of the function $f(k)$, and check that

$$f(k)'' = \frac{-(v^T \Lambda u)^2 + (v^T \Lambda v)(u^T \Lambda u)}{(v^T \Lambda v k^2 + 2v^T \Lambda u k + u^T \Lambda u)^{\frac{3}{2}}} \geq 0$$

Thus, $f(k)$ is convex with respect to k, and consequently the convexity of $f(y) = \sqrt{y^T \Lambda y}$ is proved.

Lemma 1 proves the convexity of the problem, the following lemma shows the optimal solution of the problem. Having this proved, there is no need for using an optimization solver to solve the problem.

Lemma 2 *The optimal value of the second sub-problem, $L_y(\mu) = \eta\sqrt{y^T \Lambda y} - \mu^T y$, is 0, and $y = 0$ is one optimal solution for the problem.*

The optimal value of the Lagrangian dual problem, L_μ, servers as a lower bound to the original problem, thus the optimal value of the sub-problem, $L_y(\mu) = \eta\sqrt{y^T \Lambda y} - \mu^T y$, should be bounded.

Let's transform variables y from Cartesian coordinates $(y_1, ..., y_n)$ to n-spherical coordinates $(r, \theta_1, ..., \theta_{n-1})$ through the following equations:

$$y_i = r f_i(\theta) \; ; \forall i$$

where $r \geq 0$ is the radius from the reference point, in this case the origin point in the Cartesian coordinates, θ is the vector of angles with reference directions/planes, and $f_i(\theta)$ are the functions to determine the direction of the vectors, e.g. in two-dimensional space, $f_1 = cos(\theta)$,$f_2 = sin(\theta)$ where θ is the polar angle. The transformed subproblem will be:

$$\begin{aligned}
L_y(\mu) &= \eta\sqrt{\sum_i \lambda_i y_i^2} - \sum_i \mu_i y_i \\
L_{r,\theta}(\mu) &= \eta\sqrt{\sum_i \lambda_i r^2 f_i^2(\theta)} - \sum_i \mu_i r f_i(\theta) \\
L_{r,\theta}(\mu) &= r(\eta\sqrt{\sum_i \lambda_i f_i^2(\theta)} - \sum_i \mu_i f_i(\theta))
\end{aligned}$$

where λ_i is the element in $i-th$ column and $i-th$ row of the diagonal matrix Λ. In other words, λ_i's are the eigenvalues of the travel time variance-covriance matrix.

Let's define $C = \eta\sqrt{\sum_i \lambda_i f_i^2(\theta)} - \sum_i \mu_i f_i(\theta)$, which is only a function of vector θ. Then

$$L_{r,\theta}(\mu) = rC$$

It is clear that for $C < 0$, the optimal value is $L^*_{r,\theta}(\mu) = -\infty$ at $r^* = \infty$, for $C > 0$, the optimal value is $L^*_{r,\theta}(\mu) = 0$ at $r^* = 0$, and for $C = 0$, the optimal value is $L^*_{r,\theta}(\mu) = 0$ for any value of r including $r = 0$.

Because we are interested in a lower bound for the dual problem, the case with $L^*_{r,\theta}(\mu) = -\infty$ is not acceptable, therefore, we adopt $L^*_y(\mu) = L^*_{r,\theta} = 0$. In this case, $r = 0$ indicates that the optimal point is indeed the origin point:

$$y_i^* = 0 \; ; \forall i \tag{3.8}$$

which proves the lemma.

Lemma 3 *If the second sub-problem* $L_y(\mu) = \eta\sqrt{y^T\Lambda y} - \mu^T y$ *is bounded, then the Lagrangian multipliers* μ_i *are constrained by* $-\eta\sqrt{\lambda_i} \leq \mu_i \leq \eta\sqrt{\lambda_i}$.

From Lemma 2, $L_y(\mu)$ is bounded when $C \geq 0$, therefore:

$$C \geq 0 \rightarrow \eta\sqrt{\sum_i \lambda_i f_i^2(\theta)} - \sum_i \mu_i f_i(\theta) \geq 0$$

For n=1, $f_1 = \pm 1$. Therefore, $-\beta\sqrt{\lambda_1} \leq \mu_1 \leq \beta\sqrt{\lambda_1}$.

For n=2, $f_1 = cos(\theta), f_2 = sin(\theta)$. Therefore,

$$\eta\sqrt{\lambda_1 cos^2(\theta) + \lambda_2 sin^2(\theta)} \geq \mu_1 cos(\theta) + \mu_2 sin(\theta)$$

While it is difficult to find a relationship for each μ_i in terms of other parameters, some can be achieved by considering the extreme cases as following:

- Case I: $\theta = 0$. In this case, λ_2 and μ_2 disappear and the remaining term is $\mu_1 \leq \eta\sqrt{\lambda_1}$.
- Case II: $\theta = \frac{\pi}{2}$. In this case, λ_1 and μ_1 disappear and the remaining term is $\mu_2 \leq \eta\sqrt{\lambda_2}$.
- Case III: $\theta = \pi$. In this case, λ_2 and μ_2 disappear and the remaining term is $\mu_1 \geq -\eta\sqrt{\lambda_1}$.
- Case VI: $\theta = \frac{3\pi}{2}$. In this case, λ_1 and μ_1 disappear and the remaining term is $\mu_2 \geq -\eta\sqrt{\lambda_2}$.

For $n = 3, f_1 = sin(\theta_1)cos(\theta_2), f_2 = sin(\theta_1)sin(\theta_2), f_3 = cos(\theta_1)$ and therefore the following inequality should hold:

$$\eta\sqrt{\lambda_1 sin^2(\theta_1)cos^2(\theta_2) + \lambda_2 sin^2(\theta_1)sin^2(\theta_2) + \lambda_3 cos^2(\theta_1)} \geq$$
$$\mu_1 sin(\theta_1)cos(\theta_2) + \mu_2 sin(\theta_1)sin(\theta_2) + \mu_3 cos(\theta_1).$$

One can still use extreme cases to find some bounds on Lagrangian multipliers: the case where $\theta_1 = \theta_2 = 0$ and $\theta_1 = \theta_2 = \pi$ leads to $-\eta\sqrt{\lambda_3} \leq \mu_3 \leq \eta\sqrt{\lambda_3}$; the case where $\theta_1 = \theta_2 = \frac{\pi}{2}$ and

$\theta_1 = \theta_2 = \frac{3\pi}{2}$ leads to $-\eta\sqrt{\lambda_2} \leq \mu_2 \leq \eta\sqrt{\lambda_2}$; the case where $\theta_1 = \frac{\pi}{2}, \theta_2 = 0$ and $\theta_1 = \frac{3\pi}{2}, \theta_2 = 0$ leads to $-\eta\sqrt{\lambda_1} \leq \mu_1 \leq \eta\sqrt{\lambda_1}$.

One can continue extending these to higher dimensions, therefore, some bounds on μ are as following:

$$-\eta\sqrt{\lambda_i} \leq \mu_i \leq \eta\sqrt{\lambda_i}\ ; \forall i \tag{3.9}$$

which proves the lemma.

Now, the intractable nonlinear mixed integer part, $\sqrt{x^T\Sigma x}$, has been converted into an unconstrained convex problem of y whose optimal solutions have been proved. In terms of the range of the Lagrangian multipliers, they will be incorporated using projection in the Lagrangian multiplier updating strategy that will be briefly discussed in the complete algorithm.

Thanks to the matrix decomposition technique and the variable substitution, a complicated MINLP is successfully decomposed to two convex problems that brings a great simplification in computational efforts, which will be shown in Chapter 4. One taxonomy of decomposition method-based algorithms in the literature that were designed for RSP problems, is the convexity of the decomposed subproblems. The algorithm proposed by Xing and Zhou [2] results in sub-problems involving a concave minimization problem as well as convex minimization problems. In contrast, the Cholesky decomposition method proposed by Zhang et al. [4] admits two convex minimization problems and it was proved that their method has a smaller gap. Our method indeed falls into the second category.

It is noteworthy to point out that the proposed eigen-decomposition encompasses the Cholesky decomposition as a special case when the diagonal matrix Λ is the identity matrix. Take the square-root of the matrix of Λ and we can perform a QR factorization to get a Cholesky decomposition. Therefore, the method proposed in Zhang et al. [4] can also be applied for the convexity proof.

3.3 Lagrangian Multiplier Updating Strategies

The Lagrangian dual problem is provided in $\mathbb{P}2$. By duality theory, the solutions of the Lagrangian dual problem provides a lower bound to the original problem $\mathbb{P}$. In an iterative process, an increasing trend for the lower bound is favorable because it reduces the duality gap. To this end, the goal of solving the Lagrangian dual problem is to find a solution that increases the objective value.

This section provides two methods for updating the Lagrangian multipliers: sub-gradient method and deep-cut ellipsoid method. The sub-gradient method has been widely used to find a direction for solution updating. Its simplicity and good practicality make itself a popular method when there is no necessity for getting an exact direction and step size. Ellipsoid method, the first polynomial-time algorithm to solve a linear problem, is efficient in theory with proved convergence and slow but steady in practice. The ellipsoid method to be discussed in this thesis is the "deep-cut ellipsoid method", compared to the basic ellipsoid method, deep-cut version cuts more volume at each iteration and hence improves its performance. This stimulates us to compare these two methods in an RSP problem setting. To our best knowledge, this study is the first one comparing ellipsoid method and sub-gradient method in an RSP problem setting, which is one of the contributions of the work.

3.3.1 Sub-gradient method

The search direction is given by the sub-gradient of the Lagrangian dual problem L_μ.

$$\nabla L(\mu) = V^T x - y = V^T x \tag{3.10}$$

The second equality follows from the Lemma 2 that the optimal solution of $L_y(\mu)$ is zero.

Define a scalar θ, which controls the step size towards the alteration to the direction indicated by the sub-gradient. The step size is controlled by the duality gap ($Z_{UB} - Z_{LB}$). Following Fisher's method [35], the final updating scheme is as follows, where Z_{UB} is defined as the best upper bound (Z in equation (2.3)) found so far, Z_{LB} is the best lower bound found by the Lagrangian dual problem $L(\mu^{(s)})$ and the coefficient β is a positive adjusting factor.

$$\mu^{(s+1)} = \mu^{(s)} + \theta^{(s)} \nabla L(\mu) = \mu^{(s)} + \beta \frac{Z_{UB} - Z_{LB}}{\|V^T x^{(s)}\|^2} (V^T x^{(s)}). \tag{3.11}$$

where, s is used as the iteration index, and $x^{(s)}$ is the solution of the sub-problem obtained from iteration s.

3.3.2 Deep-cut ellipsoid method

The ellipsoid method was first developed in 1970's by Soviet Union mathematicians and it is the first proved solution algorithm for solving linear program in polynomial time. It has been observed that the solution algorithm suffers from slow convergence in practice but very steady. In addition to application in solving linear program, the ellipsoid method has also been used in solving Variational Inequalities (VI) problems, combinatorial optimization problems, and etc. Interested readers can refer to the related work [36–38] for understanding the mechanism and the developement of the algorithm, but specific applications in transportation field have not been aware of by the authors of the thesis.

The basic idea of the method is present as follows. The ellipsoid method generates a sequence of ellipsoids with decreasing volumes and in each iteration it is guaranteed that the optimum value is within the ellipsoid. Ellipsoid method also applies the sub-gradient to obtain a half-space that is guaranteed not to contain any optimal point. Given a series cutting-planes, it can sequentially reduce the volume of ellipsoids so as to get the optimal point.

Any ellipsoid in $\mathbb{R}^n$, $\mathbb{E}$, can be represented in the following mathematical form with the vector, o, being the center of the ellipsoid and matrix, A, depicting the shape of the parameters, whose eigenvalues are the lengths of principal semi-axes.

$$\mathbb{E} = \{x| \quad (x - o)^T A^{-1} (x - o) \leq 1\} \tag{3.12}$$

Take minimization problem with the objective function h as an example. By taking the sub-gradient $g^T \in \partial h$ at the center, the half-space where the objective value goes down can be determined, that is, the space defined by the following,

$$\{x| \quad g^T(x - o) \leq 0\}. \tag{3.13}$$

A half-ellipsoid which contains the minimizer can the be defined as the intersection of the original ellipsoid and the half-space:

$$\mathbb{E} \cap \{x| \quad g^T(x - o) \leq 0\}. \tag{3.14}$$

A new ellipsoid is then generated by forming an minimum volume ellipsoid that contains the half-ellipsoid. For the brevity of the purpose, the proof for the following is left out only showing the results. Interested readers can refer to the work by [38]. The new ellipsoid center and shape matrix is updated as follows:

$$o^{(s+1)} = o^{(s)} - \frac{1}{1+n} A^{(s)} \tilde{g^{(s)}} \tag{3.15}$$

$$A^{(s+1)} = \frac{n^2}{n^2-1} (A^{(s)} - \frac{2}{n+1} A^{(s)} \tilde{g^{(s)}} \tilde{g^{(s)}}^T A^{(s)}) \tag{3.16}$$

$$\tilde{g}^{(s+1)} = \frac{1}{\sqrt{g^{(s)T} A^{(s)} g^{(s)}}} g^{(s)} \tag{3.17}$$

where, $\tilde{g^{(s)}}$ is the normalized sub-gradient and n is the dimension of the ellipsoid. Iteration index s is presented to make the updating method clear. One can interpret the updating method in equation (3.15) as an sub-gradient alike method with fixed step-size of $\frac{1}{n+1}$.

Deep-cut ellipsoid method can be derived in a similar way, it but instead excludes a half-space bigger than the space excluded in the basic version. The excluded space is defined by $\{x | g^T(x - o) \geq u, u \leq 0\}$. Since u is a non-positive real number, the excluded space defined precedingly is bigger than $\{x | g^T(x - o) \geq 0\}$, which is the complement of the space defined in equation (3.13). Now we discuss how the real number u should be chosen. By definition of the sub-gradient, the following inequality, $h(y) \geq h(o) + g^T(y - o)$, holds for any y where $g^T \in \partial h(o)$. Thus, it is true for the optimizer, say y^*. Then we have the following: $h^* = h(y^*) \geq h(o) + g^T(y^* - o)$. Note that the ellipsoid method does not guarantee a monotonically decreasing objectives, in other words, the current objective value $h(o^{(s)})$ may be greater than the best upper bound $h^{(s)*}$ that has been generated so far. In addition, note that $h^* \leq h(o^{(s)*})$, therefore, the negative number can be chosen as the difference of the current objective value and the best objective value found so far, which gives the following space.

$$\{x| \quad g^T(x - o) \geq h(o^{(s)*}) - h(o^{(s)})\}. \tag{3.18}$$

It is because of cutting off a larger plane every iteration that the deep-cut ellipsoid method performs better than its basic version. Center and shape matrix updating method for deep-cut ellipsoid method should also be adjusted accordingly as follows with same updating method for the normalized sub-gradient as is in equation (3.17).

$$o^{(s+1)} = o^{(s)} - \frac{1+n\alpha}{1+n} A^{(s)} \tilde{g^{(s)}} \tag{3.19}$$

$$A^{(s+1)} = \frac{n^2(1-\alpha^2)}{n^2-1}(A^{(s)} - \frac{2(1+n\alpha)}{(n+1)(1+\alpha)} A^{(s)} \tilde{g^{(s)}} \tilde{g^{(s)}}^T A^{(s)}) \tag{3.20}$$

$$\alpha = \frac{h(o^{(s)}) - h(o^{(s)*})}{\sqrt{g^{(s)T} A^{(s)} g^{(s)}}} \tag{3.21}$$

The convergence of the ellipsoid method largely relys on the constant decreasing rate of ellipsoid volume that depends on n. Theoretically, the number of iterations needed to achieve a $\epsilon-$optimal solution is $\mathcal{O}(n^2)$ [38]. For the sake of brevity, we refer readers to the work [38] for the proof of the convergence.

To put the deep-cut ellipsoid method in the context of our problem which is maximizing the Lagrangian dual problem, i.e, $max \quad L_\mu$. Thus, when using the sub-gradient to define the half space, the space that should be excluded is the complement to the space introduced above. This explains the sign change for updating the ellipsoid center presented in Algorithm 1. The bigger half-space is obtained by comparing the best optimal value found so far (Z_{LB} in our case) and the objective value at the current solution ($L(\mu^{(s)})$).

The complete deep-cut ellipsoid method is described as follows. Given an initial ellipsoid centered at $\mu^{(0)}$ with initial lengths of principal axes depicted by $E^{(0)}$, ($E^{(0)}$, $\mu^{(0)}$), and the best lower bound (Z_{LB}) found so far, the deep-cut ellipsoid method for updating the Lagrangian multipliers is as follows. Note that the presented deep-cut ellipsoid method is different from the standard one, where there is an iterative process for determining the optimal solutions for the optimization problem. Since in our problem setting, the Lagrangian multiplier updating is just for the current solution, and this solution may not already be the optimal solution, thus there is no necessity to obtain the optimal Lagrangian multiplier for the intermediate solutions. In this regards, the deep-cut ellipsoid method is only performed once in each iteration of the complete algorithm to simply get a direction for updating the Lagrangian multipliers. This also benefits the algorithm in the way that it saves computational time.

To initialize the deep-cut ellipsoid method, one can begin with a ball centered at the origin. One want to have relatively longer principal axes so as to avoid the problem that optimizer is excluded in the initial ellipsoid. The authors have carried out some experiments, although the results show that the computational cost are not very sensitive to the initial lengths of semi-axis, it is recommended to take larger value for the initial lengths of principal axes.

Algorithm 1: Deep-cut ellipsoid method for updating Lagrangian multipliers

Input: $(E^{(0)}, \mu^{(0)}), Z_{LB}$

Result: Updated Lagrangian multipliers μ

begin

Compute the sub-gradient $h^{(s)} = V^T x^{(s)}$

Normalize the sub-gradient $\tilde{h^{(s)}} = \frac{h^{(s)}}{\sqrt{h^{(s)T} E^{(s)} h^{(s)}}}$

Calculate $\alpha = \frac{L(\mu^{(s)}) - Z_{LB}}{\sqrt{h^{(s)T} E^{(s)} h^{(s)}}}$

Update Lagrangian multiplier $\mu^{(s+1)} = \mu^{(s)} + \frac{1+m\alpha}{m+1} E^{(s)} \tilde{h^{(s)}}$

Update ellipsoid shape matrix

$E^{(s+1)} = \frac{m^2}{m^2-1}(1-\alpha^2)(E^{(s)} - \frac{2(1+m\alpha)}{(m+1)(1+\alpha)} E^{(s)} \tilde{h^{(s)}} \tilde{h^{(s)}}^T E^{(s)})$

3.3.3 Dealing with negative link costs

During the process of updating the Lagrangian multiplers, it is possible that with some certain updated μ, the adjusted link travel costs, $c + V\mu$, in the standard shortest path problem become negative. Though the label correcting algorithm can still deal with the negative link costs if there is no negative cycle formed, the safest way to avoid generating negative cycles is proportionally adjusting the stepsize until none of the adjusted link costs is non-positive. This same method has been used in Xing and Zhou's paper [2].

To be specific, for the sub-gradient updating strategy, when a negative link cost occurs, the second term in equation (3.11), $\eta \frac{Z_{UB} - Z_{LB}}{\|V^T x^{(s)}\|^2}(V^T x^{(s)})$ is proportionally adjusted until there is no more negative link costs. For the deep-cut ellipsoid updating strategy, similarly, non-negative link costs can be obtained by adjusting the term $\frac{1+m\alpha}{m+1}E^{(s)}\tilde{h^{(s)}}$.

3.4 Complete Solution Algorithm

Having introduced the decomposition of the original problem into two sub-problems, proved convexity and optimal solution for the sub-problem of y, and discussed two Lagrangian multiplier updating strategies, this subsection provides the complete solution algorithm for the RSP problem with link travel time correlations. The major idea of the following presented Lagrangian substitution based method is generating upper bound and lower bound series until sufficient small gap is achieved. The lower bound can be obtained by maximizing the Lagrangian dual function, equation (3.7), and the Lagrangian dual function has already been decomposed to two sub-problems for which sub-problem of y has an analytically proved optimal solution. Thus the lower bound reduces to solving the sub-problem of x. An upper bound can be evaluated by equation (2.3) using the solution of sub-problem of x. Figure 3.1 presents the logic flow of the algorithm in brevity.

Starting with solving a standard shortest path problem, this provides an initial feasible solution and an upper bound for the optimal path's standard deviation. Apply the standard labeling setting algorithm to solve sub-problem $L_x(\mu)$. The optimal solution to the second sub-problem, $L_y(\mu)$ is proved to be zero, thus the lower bound $Z_{LB} = L_x(\mu)$ can be found. Evaluating the solutions obtained by solving the first sub-problem with equation (2.3), a candidate of upper bound (Z_{UB}) is available. A series of decreasing upper bounds is favorable, thus if there is no improvement of upper bound we keep the previous solution and upper bound from the previous iteration. In other words, the solution and the upper bound are updated only when the value evaluated by equation (2.3) decreases. The Lagrangian multipliers are updated in an iterative process until a convergence criterion is satisfied. Be noted that if, by any one of the Lagrangian multiplier updating strategies, the updated μ is outside the range derived in equation (3.9), the corresponding element μ should be projected back to the feasible range. The algorithm for solving the RSP problem with link travel

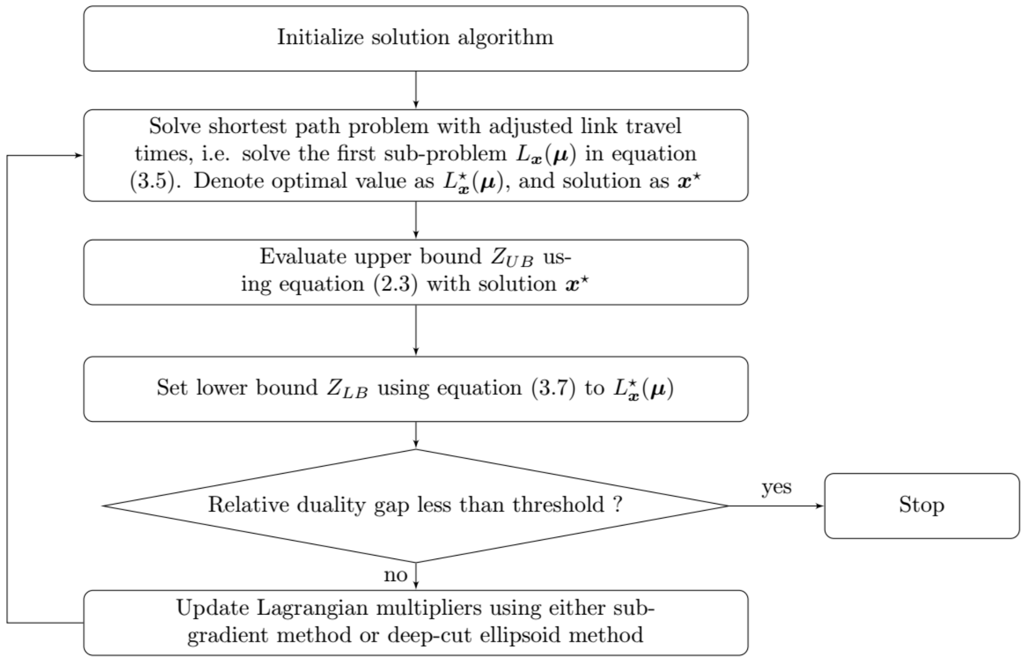


Figure 3.1: Major steps of the solution algorithm

time correlations is summarized in Algorithm 2.

Algorithm 2: Decomposition based Lagrangian substitution solution algorithm

Input: Relative duality gap threshold ϵ, maximum number of iteration K

Result: Reliable shortest path solution x, and duality gap

begin

Step 1: Variance-covariance matrix decomposition

Decompose the variance-covariance matrix, Σ, into the diagonal matrix Λ and V by eigendecompostion.

Step 2: Initialization

Set iteration number $s = 0$ and initialize lower bound $(Z_{LB}) = -\infty$;

Choose an initial Lagrangian multiplier $\mu^0 \geq 0$;

Solve a standard shortest path, compute Z as the upper bound (Z_{UB});

Step 3: Solve Lagrangian relaxation problems

Solve $L_x(\mu)$, denoting solutions as x and optimal value as $L_x^\star(\mu)$;

Compute Z with solution $x^\star$, set it as $\tilde{Z_{UB}}$;

if $Z_{UB} \leq \tilde{Z_{UB}}$ **then**

  Set $Z_{UB} = \tilde{Z_{UB}}$

  Set $x^{(s)} = x^\star$

else

  Set $x^{(s)} = x^{(s-1)}$

Set $L(\mu) = L_x^\star(\mu)$;

if $L(\mu) \geq Z_{LB}$ **then**

  Set $Z_{LB} = L(\mu)$

Step 4: Update Lagrangian multiplier

Update Lagrangian multipliers using either sub-gradient method (equation (3.11)) or deep-cut ellipsoid method using Algorithm 1;

Project Lagrangian multipliers back to the feasible region using equation (3.9);

Step 5: Termination check

if $\frac{Z_{UB}-Z_{LB}}{Z_{LB}} \leq \epsilon \; or \; s \geq K$ **then**

  Terminate the algorithm

  Report $x^{(s)}$ as the RSP solution and duality gap $\frac{Z_{UB}-Z_{LB}}{Z_{LB}}$

else

  Go to step 3

Chapter 4

Numerical Experiments

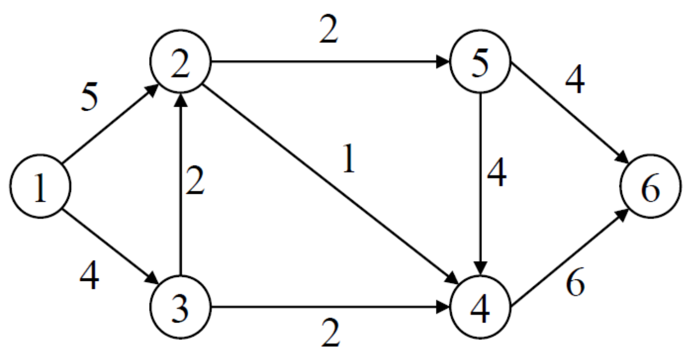
4.1 Illustrative Example Computational Tests

For the purpose of demonstrating the algorithm steps and its characteristics, the results of a small illustrative network borrowed from Shahabi et. al. [8] are presented in this section. Figure 4.1 shows the network topology and the variance-covariance matrix of link travel times, where the numbers on the arcs correspond to the mean costs. a_{ij} represents the link (i, j), so that in the variance-covariance matrix, $Cov(12, 25) = 0.75$ and $Cov(34, 32) = 0.70$ for instance.

4.1.1 Complete Algorithm Review

Both sub-gradient and deep-cut ellipsoid updating strategy are used to give some insights of how the complete Lagrangian substitution algorithm works for finding RSP solutions. The problem setting is $\eta = 1.0, \epsilon = 0.5\%, \mu = 1.0$. Table 4.1 and 4.2 show how Lagrangian multipliers, lower bound and upper bound, and relative gap evolve to get the optimal solution.

The algorithm starts with initial Lagrangian multipliers all set to 1.0, and the estimated Z_{UB} is calculated from a standard shortest path (1-2-5-6), which results in the initial upper bound 15.30. The lower bound is evaluated from equation (3.5) by solving the standard shortest path problem $L_x(\mu)$. At iteration 2, solving the sub-problem $L_x(\mu)$ gives us a new shortest path (1-3-4-6) and the Z_{UB} evaluated at this solution lowered to 14.43. Continuing the algorithm, the shortest path solution of $L_x(\mu)$ keep unchanged, but with the adjustment of Lagrangian multipliers, lower bound keeps increasing until the gap at iteration 6 reaches 0.29% so that the predefined gap threshold is met. Deep-cut ellipsoid method also starts with all Lagrangian multipliers set to 1.0, same with the



	a_{12}	a_{13}	a_{24}	a_{25}	a_{32}	a_{34}	a_{46}	a_{54}	a_{56}
a_{12}	5	1.2	0.75	0.75	-0.325	0.75	0.5	0.7625	1.5
a_{13}	1.2	3	0.875	-0.5	0.55	-0.8	-1.125	0.95	0.7875
a_{24}	0.75	0.875	2	1	1	0.8	-0.3	0.3875	0.5125
a_{25}	0.75	-0.5	1	3	1.25	0.9	0.5	-0.325	1
a_{32}	-0.325	0.55	1	1.25	2	0.7	-0.5	0.375	0.275
a_{34}	0.75	-0.8	0.8	0.9	0.7	3.5	0.625	-0.3625	-0.1375
a_{46}	0.5	-1.125	-0.3	0.5	-0.5	0.625	2	0.8125	1
a_{54}	0.7625	0.95	0.3875	-0.325	0.375	-0.3625	0.8125	3.5	-0.6125
a_{56}	1.5	0.7875	0.5125	1	0.275	-0.1375	1	-0.6125	4

Figure 4.1: Illustrative network topology and link travel time variance-covariance matrix

sub-gradient method, it finds the new shortest path at the second iteration. The similar increasing lower bound and decreasing relative gap is observed, and the only difference is that deep-cut ellipsoid method takes fewer iterations than sub-gradient method. The convergence trajectories for these two examples are shown in Figure 4.2 and 4.3.

Table 4.1: Results of Lagrangian substitution algorithm using sub-gradient method for the illustrative network with $\eta = 1.0$

iteration	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	$L_x(\mu)$	Z_{LB}	Z_{UB}	relative gap(%)
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	10.42	10.42	15.30	31.90
2	-0.172	0.184	0.540	-0.367	0.123	0.573	0.927	0.004	-1.472	11.88	11.88	14.43	17.67
3	0.267	-0.418	0.652	0.239	0.140	0.324	0.782	-0.097	-1.850	13.75	13.75	14.43	4.67
4	0.268	-0.577	0.682	0.400	0.145	0.258	0.744	-0.123	-1.950	14.16	14.16	14.43	1.85
5	0.268	-0.666	0.679	0.488	0.147	0.222	0.723	-0.139	-2.005	14.32	14.32	14.43	0.73
6	0.268	-0.676	0.699	0.498	0.148	0.218	0.721	-0.140	-2.011	14.25	14.39	14.43	0.29

Table 4.2: Results of Lagrangian substitution algorithm using deep-cut ellipsoid method for the illustrative network with $\eta = 1.0$

iteration	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	$L_x(\mu)$	Z_{LB}	Z_{UB}	relative gap(%)
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	10.84	10.84	15.30	29.18
2	-0.267	-0.082	0.408	-0.773	-0.130	0.459	0.921	-0.286	-2.220	12.18	12.18	14.43	15.59
3	0.267	-0.690	0.544	-0.038	-0.110	0.156	0.746	-0.410	-2.678	14.34	14.34	14.43	0.62
4	0.268	-0.690	0.553	0.007	-0.108	0.137	0.735	-0.417	-2.718	14.41	14.41	14.43	0.12

4.1.2 Lagrangian Multiplier Updating Method Comparison: Sub-gradient and Deep-cut Ellipsoid

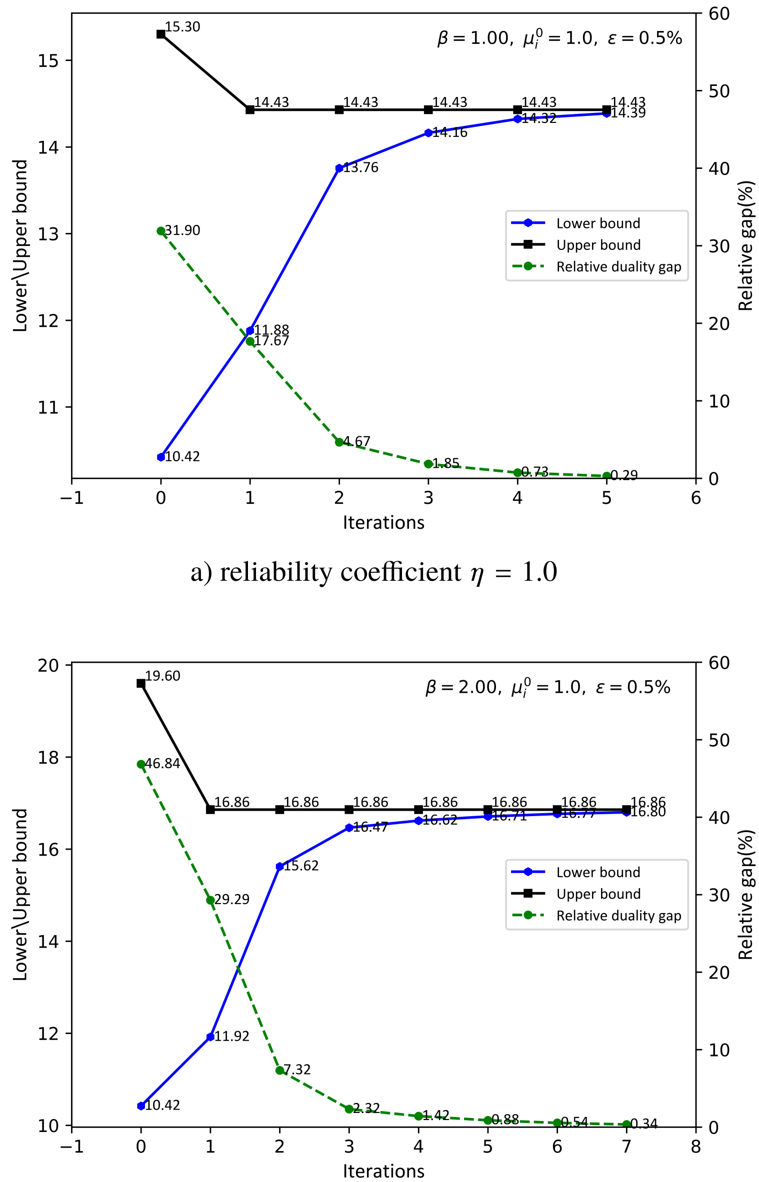
Table 4.3 compares the solutions obtained from the sub-gradient method and deep-cut ellipsoid method with different reliability coefficient settings. In the sequel, Lagrangian substitution based sub-gradient method deep-cut ellipsoid method will be referred to as LS-SG and LS-DE for convenience, respectively. The case where reliability coefficient is set to 0 corresponds to a standard shortest path problem. With reliability coefficient increasing to 1.0, shortest paths for both OD pairs deviate from the standard shortest paths. Continuing increasing value of the reliability coefficient, the reliable shortest paths keep unchanged with respect to the case where η equals 1.0. One can expect the same reliable shortest path for even higher values of reliability coefficient since when the standard deviation of path travel time dominates the average travel time, a higher weight of reliability coefficient will only foster the reliable path. Thus, to find a critical point of reliability coefficient where reliable shortest path first changes would be an interesting problem to explore. Here, the example shows that standard shortest path may be different from a reliable shortest path.

For the tests shown in Table 4.3, the relative duality gap threshold was set to 0.5%. Comparing sub-gradient and deep-cut ellipsoid Lagrangian multiplier updating methods, RSP solution and the optimal value of both methods are exactly same. Computational time, the number of iterations and relative duality gap are also comparable. It shows in the table LS-DE algorithm solves problem slightly faster than the LS-SG for the illustrative network.

Figure 4.1 and 4.2 respectively show the convergence trajectory for the two updating strategies with cases that reliability coefficient setting to 1.0 and 2.0 respectively. Sub-gradient method tends to find the optimal solution at the beginning of the algorithm, which finds both reliable shortest paths at the second iteration, whereas ellipsoid method needs more iterations for finding the optimal solution. Compared with the deep-cut ellipsoid method, the sub-gradient method has a better convergence trend due to the adjustable step size according to the relative difference between the upper bound and the lower bound.

Table 4.3: Solution summary for Lagrangian substitution algorithm with sub-gradient and deep-cut ellipsoid method of illustrative network

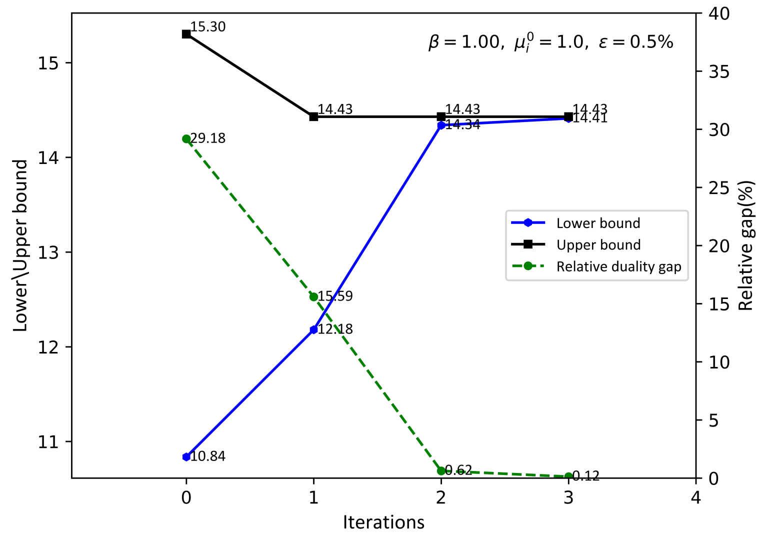
		RSP solution		Optimal value		Computational time (s)		# of iterations		Relative duality gap (%)	
Reliability coefficient η	OD pair	LS-SG	LS-DE	LS-SG	LS-DE	LS-SG	LS-DE	LS-SG	LS-DE	LS-SG	LS-DE
0	(1,6)	[1-2-5-6]	[1-2-5-6]	11.0	11.0	0.000	0.000	2	1	0.0	0.0
	(2,6)	[2-5-6]	[2-5-6]	6.0	6.0	0.000	0.000	2	1	0.0	0.0
1.0	(1,6)	[1-3-4-6]	[1-3-4-6]	14.43	14.43	0.016	0.000	6	4	0.29	0.03
	(2,6)	[2-4-6]	[2-4-6]	8.84	8.84	0.06	0.013	7	5	0.33	0.17
2.0	(1,6)	[1-3-4-6]	[1-3-4-6]	16.86	16.86	0.015	0.016	8	4	0.34	0.47
	(2,6)	[2-4-6]	[2-4-6]	10.69	10.69	0.06	0.015	7	4	0.49	0.37



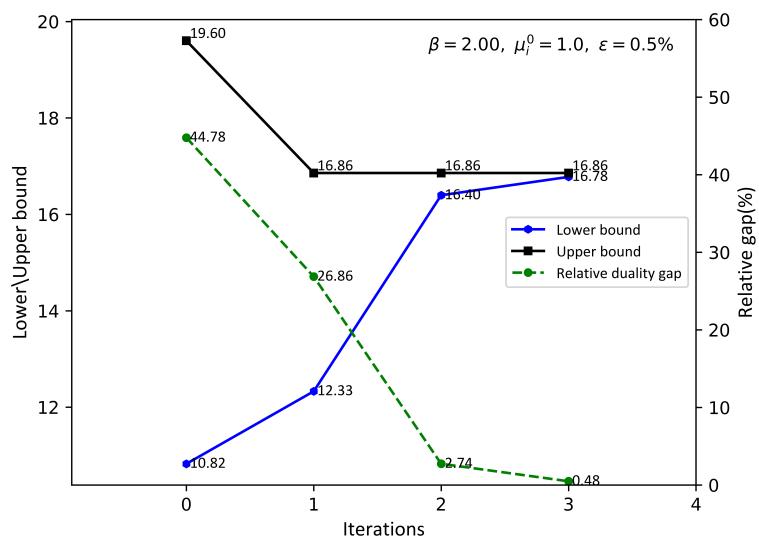
a) reliability coefficient $\eta = 1.0$

b) reliability coefficient $\eta = 2.0$

Figure 4.2: Convergence trajectory for the sub-gradient Lagrangian multiplier updating method



a) reliability coefficient $\eta = 1.0$



b) reliability coefficient $\eta = 2.0$

Figure 4.3: Convergence trajectory for the deep-cut ellipsoid Lagrangian multiplier updating method

4.1.3 Comparison with GUROBI

We also compared our solution algorithm with the commercial optimization solver GUROBI 7.5.0. GUROBI claims to solve mixed integer quadratically constrained problem (MIQCP). To solve this mean-standard deviation RSP problem, we slightly modified the problem in $\mathbb{P}$ to accommodate the solver as following, where t is an auxiliary continuous variable.

($\mathbb{P}3$)

$$\min \quad Z = c^T x + \eta t \tag{4.1}$$

$$\text{s.t.} \quad t^2 \geq x^T \Sigma x \tag{4.2}$$

$$\sum_{j:ij \in A} x_{ij} - \sum_{j:ji \in A} x_{ij} = \begin{cases} -1 & i = r \\ 0 & i \in \mathbb{N} - \{r, s\} \\ 1 & i = s \end{cases} \tag{4.3}$$

$$x_{ij} \in \{0, 1\} \tag{4.4}$$

GUROBI is readily available for solving the problem formulated above. Besides, as is mentioned in the literature review section, the RSP problem defined as mean-variance is another way to model the variability of path travel times. We referred to this mean-variance problem as $\mathbb{P}4$ in the sequel, whose objective function is $c^T x + \eta x^T \Sigma x$. It would also be an interesting problem to solve the mean-variance problem and compare the results. Here, we slightly modified our algorithm to solve the mean-variance problem. The second sub-problem $\bar{L_y}(\mu)$ now becomes $inf_y\{\eta y^T \Lambda y - \mu^T y\}$, and the convex optimization problem can be solved by any convex optimization solution algorithm. After obtaining the solutions for y, the same sub-gradient method for updating the Lagrangian multipliers can be applied and the only difference is that the sub-gradient, in this case, is $\nabla L(\mu) = V^T x - y$. Follow the LS-SG algorithm proposed for the mean-standard deviation problem, the mean-variance problem can be solved.

We obtained solutions from our proposed method for mean-variance and mean-standard deviation problems, and compared the results with the commercial optimization solver GUROBI for this

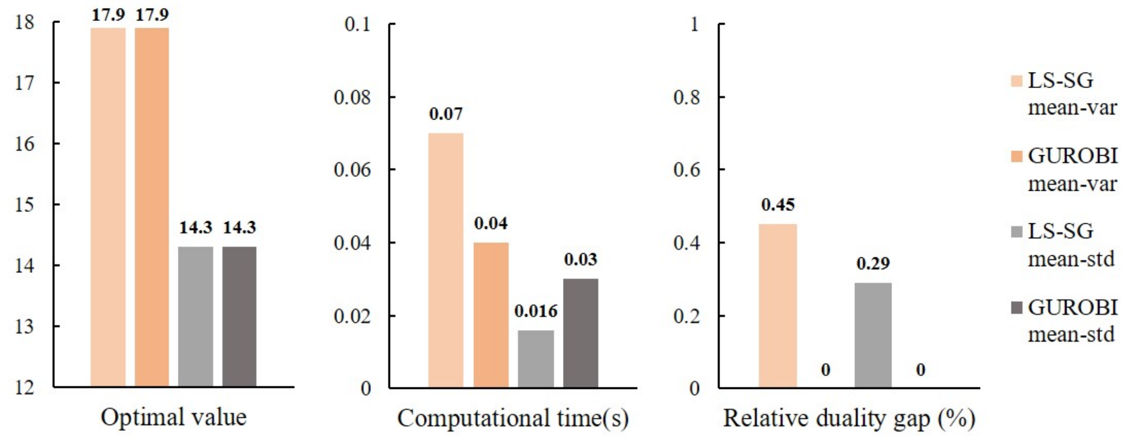


Figure 4.4: Comparison with GUROBI solver of mean-variance and mean-standard deviation problem

illustrative network example with $\eta = 1.0$ and $\epsilon = 0.5\%$. A detailed comparison is presented in Figure 4.4. To compare the mean-variance problem first, experiment results show that both our proposed algorithm and GUROBI found the optimal path (1-3-4-6) whose travel time is 17.9. There is a 0.45% relative duality gap by our method in contrast to 0 relative gap of GUROBI for the mean-variance problem (4.4). In terms of computational time, our proposed algorithm spends more time. As to the mean-standard deviation problem, again both the proposed algorithm and GUROBI found the optimal RSP (1-3-4-6) with total travel time being 14.30. Computational time for LS-SG is around half of the GUROBI's with a 0.29% relative duality gap. Later in the large-scale network experiments, however, GUROBI fails to solve an RSP problem within a reasonable amount of time. The algorithm GUROBI mainly adopts for solving mixed integer programming is branch-and-cut.

To sum up, the comparison with the commercial solver GUROBI, shows that the proposed algorithm is accurate and the performances are comparable to the branch-and-cut algorithm. More experiments will be tests on larger networks and sample sizes in the following section.

4.2 Real Network Computational Tests

Numerical tests on several large-scale real transportation networks were conducted to evaluate the performance of the proposed solution algorithm. The algorithm was coded in Python. All the tests were executed on a 2.60 GHz Core i7-6700HQ computer with the 64-bit version of the Windows 10 operating system and 8GB RAM. To show the necessity of considering link travel time correlations, travel times computed with and without link travel time correlations were first analyzed. In terms of the algorithm performance, we compared the proposed algorithm with the existing outer approximation method [8] and sampling based Lagrangian relaxation method [2], and also compared with the sub-gradient projection algorithm [23]. These analysis will be presented in the following subsections.

4.2.1 Data Preparation

Three real networks, Anaheim, Barcelona, and Chicago sketch, are used to conduct the tests. The details of the test networks are presented in Table 4.4. The average travel time data come from transportation test networks [39]. We also need standard deviation of link travel times to capture the variability of link travel times. In this regard, we first generate the standard deviation by the following equation and then generate the normally distributed random samples based on the average and standard deviation. Note that in this study, link travel time spatial correlation is not limited, which means that link travel time on link ij, for instance, is possibly correlated with all the other links in the network. Given link travel time samples, then the variance-covariance matrix can be determined. The standard deviation of link travel time was generated randomly in $(0, v)$ range. This is the same method adopted by Shahabi [8]. Parameter v controls the maximum value of the coefficient of variation. Different values of v will be tested on real networks.

$$\sigma_{ij} = Uniform(0, v)c_{ij} \tag{4.5}$$

Due to the numerical issues of the covariance matrix computation, following the above-mentioned method may not necessarily return a variance-covariance matrix that is PSD. An alternative method is to approximate the generated matrix to the nearest PSD matrix to obtain a valid variance-covariance matrix. We refer readers to the work by Higham [40], which is the method we adopted to get a PSD matrix. Table 4.5 shows the parameters and their values that are used in the real network experiments.

Table 4.4: Test network summary

Test network	Number of links (m)	Number of nodes (n)
Anaheim	914	416
Chicago sketch	2950	933
Barcelona	2522	1020

Table 4.5: Experiment parameters

Parameter	Selected testing values	Parameter description
ν	$\{0.15, 0.25, 0.5\}$	maximum value of coefficient of variation
η	$\{1.0, 2.0, 3.0\}$	reliability coefficient (risk aversion weights)

4.2.2 Comparison Between Correlated and Uncorrelated Link Travel Times

We generalized the mean-standard deviation defined RSP problem from the one without link travel time correlations (see equation (2.1), which can be solved to optimality [5]. However, the solution of the problem defined in equation (2.1) leaves out the interactions of the links on the shortest path that can potentially affect the travel time of the current shortest path. Therefore, we feel the necessity of carrying out the comparison of shortest travel times with and without link travel time correlations. We randomly selected some OD pairs from the Chicago sketch and Barcelona network, computed shortest path travel time under different cases and recorded the results in Table 4.6. Shortest path travel time with link correlations was computed using the LS-SG method. Shortest path travel time without link correlations was computed using the same algorithm, where the only difference is that the off-diagonal elements of the variance-covariance, Σ, were modified to zero leaving a diagonal PSD matrix that only contains the variances of link travel times. Using the solution obtained from the case without correlations, we computed the correlated travel time, which is presented in the table as “corrected travel time”.

Table 4.6: Comparison of shortest travel time for correlated and uncorrelated link travel time

Network	OD pair	Shortest travel time with correlations (sec)	Shortest travel time without correlations (sec)	Corrected travel time (sec)
Chicago Sketch	(426, 885)	5750.93	5804.34 (+)	5791.28 (+)
	(390, 778)	4905.68	4976.82 (+)	4974.48 (+)
	(526, 549)	2252.96	2248.26 (-)	2290.00 (+)
	(608, 501)	2622.61	2601.97 (-)	2622.61
	(403, 883)	4728.69	4734.64 (+)	4728.69
Barcelona	(286, 1007)	672.82	673.72 (+)	679.08 (+)
	(70, 24)	736.28	743.61 (+)	741.09 (+)
	(272, 759)	294.15	293.28 (-)	294.15
	(919,259)	533.92	538.38 (+)	533.92

In the table, the plus and minus sign following the travel times are indicators of whether the figure is greater or smaller than the shortest travel time with correlations. For example, in the Chicago sketch network the travel time of OD pair (426, 885) is 5750.93 seconds with correlations. If correlations are ignored, the shortest path travel time is 5804.34, which is greater than the one with correlations, and if we compute the travel time using the solution, the travel time is 5791.28, which is smaller than 5804.34 but still greater than 5750.93. Other records can be interpreted similarly. It has been observed that the shortest path travel time without correlations can be greater or smaller than the one with link travel correlations, but the corrected travel time is always no less than the shortest travel time without correlations. In other words, if someone follows the path that has minimum travel time without link travel times, he/she may end up experience longer actual travel time due to the exclusion of link travel time correlations. Therefore, it is necessary that link travel time correlations are considered when the goal is to find *reliable* paths.

The cases where corrected travel time equals travel time with correlations reveal the fact that some paths travel time variations are dominated by the link travel time variance per se, which is intuitive. The authors have also observed that around half of the paths' travel time without correlations coincide with the corrected travel time. Nonetheless, in either case, the inclusion of link travel time correlations is necessary.

4.2.3 Real Network Comparisons

The purpose of tests summarized in Table 4.7 is to compare computational efficiency of the proposed algorithm with both of the Lagrangian multiplier updating methods, varying reliability coefficients and variation coefficients. For all the tests reported in Table 4.7, 100 OD pairs from the tested network are randomly drawn, and both the Lagrangian substitution algorithm with sub-gradient method and deep-cut ellipsoid method for finding the RSP were conducted. The relative gap threshold was set to 1% and the maximum number of iterations K was set to 200. Table 4.7 recorded the average computational time, average duality gap, and average number of iterations. Note that all statistics related to computational time reported are ones only for the main algorithm time (step 2 to step 5), which does not include the variance-covariance decomposition time. This is true for all the following tests.

Table 4.7: Large scale network experimental results summary-Sub-gradient and Deep-cut ellipsoid method

			Average computational time (s)		Average relative gap (%)		Average number of iterations		Max computational time (s)		Min computational time (s)	
	η	v	LS-SG	LS-DE	LS-SG	LS-DE	LS-SG	LS-DE	LS-SG	LS-DE	LS-SG	LS-DE
Anaheim	1.0	0.15	0.02	0.10	0.64	0.45	2.4	2.2	0.08	0.29	0.01	0.04
		0.25	0.09	0.28	0.83	0.83	6.7	6.0	0.47	0.03	0.03	0.07
		0.50	0.15	0.46	0.83	0.81	11.2	9.2	0.69	1.19	0.04	0.08
	2.0	0.15	0.04	0.15	0.68	0.66	2.8	3.2	0.08	0.56	0.02	0.07
		0.25	0.12	0.46	0.84	0.82	9.6	9.3	0.51	1.51	0.02	0.13
		0.50	0.22	0.54	0.86	0.82	17.0	11.4	2.30	2.00	0.05	0.12
	3.0	0.15	0.05	0.18	0.65	0.59	3.1	3.5	0.10	0.51	0.02	0.09
		0.25	0.13	0.56	0.80	0.84	10.1	12.2	0.28	3.21	0.04	0.13
		0.50	0.21	0.69	0.88	0.84	15.9	14.6	0.58	3.07	0.25	0.09
Chicago sketch	1.0	0.15	0.45	0.82	0.85	0.65	8.3	3.2	6.80	4.16	0.10	0.50
		0.25	0.57	1.07	0.87	0.62	13.9	3.7	6.92	3.29	0.15	0.52
		0.50	0.90	1.39	0.89	0.65	15.2	4.7	4.25	3.15	0.22	0.53
	2.0	0.15	0.50	1.06	0.86	0.66	9.4	3.8	1.35	3.38	0.15	0.52
		0.25	0.74	1.81	0.87	0.71	12.8	6.9	3.40	10.34	0.12	0.53
		0.50	1.18	2.76	0.87	0.75	19.9	8.9	7.66	15.09	0.16	0.56
	3.0	0.15	0.66	1.56	0.91	0.65	12.3	5.6	1.90	7.99	0.21	0.51
		0.25	1.29	1.99	0.85	0.70	21.0	6.9	10.40	9.24	0.11	0.50
		0.50	1.37	3.08	0.92	0.72	25.8	10.7	4.12	25.61	0.20	0.54
Barcelona	1.0	0.15	0.29	1.43	0.87	0.85	3.8	4.1	0.81	9.94	0.14	0.68
		0.25	0.53	1.90	0.83	0.79	7.3	5.1	1.55	9.44	0.14	0.69
		0.50	1.39	4.36	0.86	0.82	18.1	12.4	11.11	28.52	0.15	0.69
	2.0	0.15	0.71	2.94	0.83	0.83	10.2	7.3	2.36	18.22	0.14	0.69
		0.25	0.82	3.67	0.88	0.78	10.8	10.2	2.62	27.06	0.15	0.69
		0.50	1.35	6.16	0.85	0.84	17.8	17.1	4.24	26.78	0.28	0.74
	3.0	0.15	0.94	3.13	0.88	0.83	11.5	8.2	5.36	12.92	0.15	0.69
		0.25	1.28	4.19	0.87	0.75	15.7	11.7	10.62	30.53	0.23	0.67
		0.50	**1.47**	**7.33**	0.87	0.85	21.2	18.3	10.77	22.83	0.24	1.05

Both Lagrangian substitution algorithm with sub-gradient updating method (LS-SG) and with deep-cut ellipsoid updating method (LS-DE), are sensitive to reliability coefficient and variance coefficient. For the same network, say Anaheim, with fixed $\eta = 1.0$, we observe an increase for computational time with increasing variance coefficient v. Similarly, with fixed $v = 0.15$, increasing computational time and iterations can be observed with increased reliability coefficient. As a result, for Anaheim network, the problem with largest η and v takes the longest computational time, and this is true for both methods (on average 0.21 seconds and 0.69 seconds for LS-SG and LS-DE, respectively). Comparison between networks shows that computational effort also relates to the network size, which is not surprising. OD pair drawn from larger network tends to have longer travel time and thus it takes longer for convergence. In this problem setting, LS-SG takes as long as 1.37 seconds for finding a reliable shortest path in Chicago sketch network with $\eta = 3.0, v = 0.50$. For LS-DE, this number increases to 3.08 seconds, more than twice as much as LS-SG. Similar patterns also exist among other cases that LS-DE usually takes longer than LS-SG for meeting the same duality gap threshold. The LS-SG solves all the problems within 1.47 seconds, and the LS-DE solves all the problems within 7.33 seconds, which both of the longest running time happen in Barcelona network with $\eta = 3.0, v = 0.50$. In terms of average number of iterations, numbers vary according to network sizes, reliability coefficients and variance coefficients. For LS-SG, it varies from 2.4 to 21.2 and for LS-DE it changes in the range of $(2.2, 18.3)$. Longer computational time for LS-DE is not surprising considering the Lagrangian multiplier updating method in deep-cut ellipsoid method. To update the multipliers, inverse matrix calculation and matrix multiplication are needed, which is the main reason that slows down the LS-DE method despite fewer iterations.

From Table 4.7, we observe that deep-cut ellipsoid method usually takes more computational time than the sub-gradient method to reach similar duality gaps. To test on the efficiency of these two updating methods and to show if there is any benefit in applying ellipsoid method, we designed the following experiments as well. The first two comparisons for the Anaheim (Chicago sketch) network, we fix the maximum computational time as 0.1 (0.5) seconds and 0.5 (1.0) seconds respectively, in order to compare which method gives the better solution within a fixed running time. The last test was designed in such a way that allows for long enough computational time (5 seconds and 10 seconds respectively), and then compare the solution of the two methods. Tests were performed on Anaheim and Chicago sketch networks with $\eta = 3.0, \nu = 0.25$. Figure 4.5 summarizes the results.

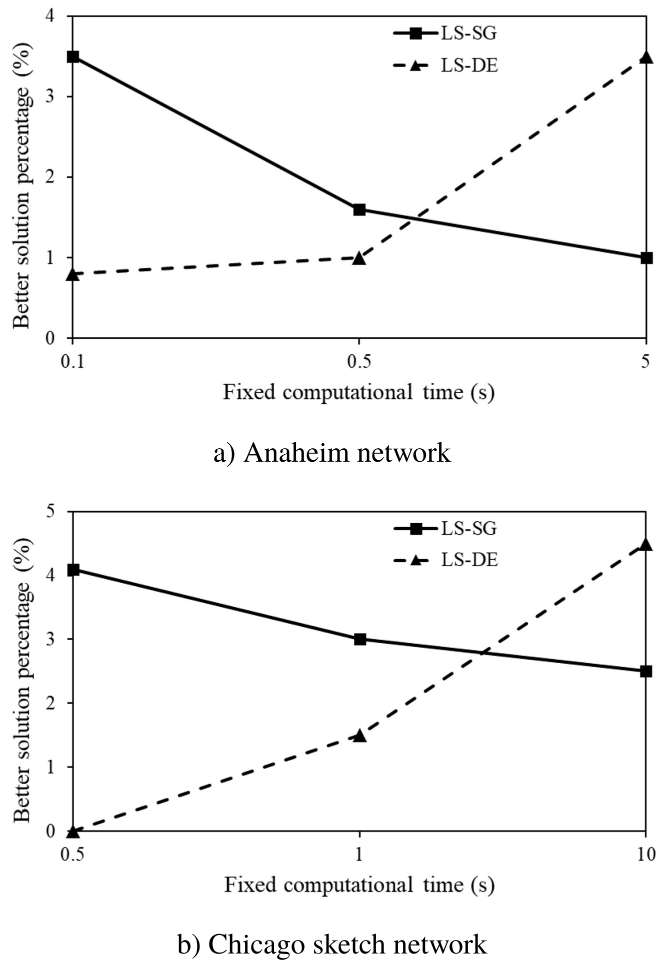


Figure 4.5: Fixed computational time solution comparison

Given a very short computational time, say, 0.1 seconds for Anaheim network and 0.5 seconds for Chicago sketch network, solutions obtained by LS-SG are better than those obtained by LS-DE for more instances. Increasing computational time for Anaheim to 0.5 seconds, LS-DE improves solutions gradually, which is reflected from the decreased percentage of better solutions by LS-SG, but LS-SG still outperforms LS-DE in obtaining more better solutions. Giving longer enough computational time, it is observed that solutions by LS-DE outperform those by LS-SG. Similar trends for the Chicago sketch network exist. These tests show that though LS-DE has a relatively slow convergence rate compared with LS-SG, it can find better solutions that are not found by LS-SG. Given the fact that the duality gaps of LS-SG solutions are already small, this indicates that the LS-SG method might converge to local optimal solutions. This partly explains why LS-DE algorithm tends to obtain smaller gaps than LS-SG. It is because LS-DE are more robust in finding better feasible solutions. Thus, if computational time is not a concern, deep-cut ellipsoid method shall be applied to update the Lagrangian multipliers since the solutions by LS-DE usually finds better solutions than LS-SG. On the other hand, considering the longer computational time LS-DE consumes, if one wants to obtain some reliable shortest path solution very quickly, LS-SG is the better choice.

In summary, tests in this section show that LS-SG converges faster than LS-DE, but LS-DE may give better solutions even though computational time can be significantly longer(2-4 times) than LS-SG's. One may choose between the two Lagrangian multiplier updating methods based on whether the speed of finding the solution is more crucial or not. Also, the multi-start strategy can be applied to the sub-gradient method to make solutions more reliable.

4.2.4 Comparison with OA, Sampling based Algorithm, and GUROBI

In this section, we compared our proposed algorithm with existing algorithms in the literature, that is the sampling-based algorithm and Outer Approximation, and the commercial optimization package, GUROBI.

A comparison between GUROBI and our proposed algorithm is of interest as is shown in the illustrative example. We transformed the problem to the form of $\mathbb{P}3$ and coded the problem in Python. However, preliminary tests on Anaheim and Barcelona with GUROBI shows that the solver was unable to handle the problem of size like these. It already took over 38,000 seconds for Anaheim and 171,000 seconds for Barcelona with $\eta = 1.0$ and $\nu = 0.15$ but still no solutions were found. Regarding of this, there is no necessity of carrying out the complete comparison with GUROBI any longer.

The following table shows the computational efficiency comparison between the proposed algorithm and the Outer Approximation method. Note that the statistics selected for Outer Approximation method are the ones reported in the paper [8] for corresponding variance coefficient with correlation factor equal to 0, which are the cases usually with less computational time. All the tests were based on $\eta = 1.0$.

Table 4.8: Computation comparison between the proposed algorithm (LS-SG and LS-DE) and OA

		Average computational time (s)			Average number of iterations			Average relative duality gap (%)		
	v	LS-SG	LS-DE	OA	LS-SG	LS-DE	OA	LS-SG	LS-DE	OA
Anaheim	0.1	0.02	0.08	5.64	2.2	2.3	2	0.83	0.68	0
	0.3	0.08	0.28	5.66	7.2	6.3	2	0.81	0.86	0
	0.5	0.15	0.46	5.77	11.2	9.2	2	0.83	0.81	0
Chicago sketch	0.1	0.13	0.70	58.74	6.3	3.6	2	0.84	0.76	0.04
	0.3	0.45	1.04	59.29	9.8	4.2	2	0.82	0.74	0.03
	0.5	0.90	1.39	60.11	15.2	4.7	2	0.89	0.65	0.05
Barcelona	0.1	0.21	1.20	155.54	4.6	5.3	2	0.84	0.83	0.07
	0.3	0.83	3.60	157.36	11.2	9.2	2	0.83	0.81	0.08
	0.5	1.39	4.36	160.19	18.1	12.4	2	0.86	0.82	0.07

According to Table 4.8, the OA algorithm outperforms in terms of the relative gap and average number of iterations. However, the OA needs more computational time compared with the proposed algorithm. The fact that computational times of the same network is rather stable shows that the OA algorithm is insensitive to the variance coefficient. The proposed algorithm's computational time increases with increasing network size and variance coefficient. Despite this, the total computational times for all the tests are still significantly less than those of OA. For larger networks like Barcelona, the RSP problem is solved within 1.4 seconds using LS-SG and less than 5 seconds using LS-DE while it takes around 160 seconds for the OA method. The average computational time improvement are 98.2% and 95.2%, respectively for methods LS-SG and LS-DE. The OA method finds a one-to-all reliable shortest paths tree, whereas, the proposed algorithm finds the reliable shortest path between a specific OD pair each time.

In the literature, there is also a one-to-one algorithm, namely sampling based Lagrangian relaxation [2]. Since the test network is different in that study, and there is no computational time available, we compared the relative gap. When $\eta = 1.27$, the best relative gap from the sampling-based algorithm is around 5.5%, and when η increases to 4, the gap increases to around 13.5%. In our case, the average relative gap is less than 1% with a very short computational time. Be noted that the referenced network is larger with 53,124 nodes and 93,900 links.

4.2.5 Comparison with the Sub-gradient Projection based Lagrangian Substitution Method

Using Lagrangian substitution method that decomposes the problem into two convex optimization problems, the most recent research work is believed to be conducted by Zhang et. al. [4], which shows a smaller gap and faster computational time compared to the sampling based Lagrangian substitution method. In this section, we compared our proposed algorithm with their proposed algorithm, namely the Lagrangian relaxation method with sub-gradient projection (LR-SP). The authors implemented their algorithm and all the tests were conducted in the same computation environment. To detect negative cycle problem, Bellman-Ford labelling algorithm was used to find shortest paths in LR-SP [41].

Table 4.9 compares the computational cost for the methods LS-SG and LR-SP. There are 6 sets of tests on the three networks, and in each set 50 OD pairs were randomly drawn to perform the experiment. The relative duality gap threshold was set to 0.1% and v was set to 0.50 for all cases. Comparing the computational time, LS-SG is 2-3 time faster than the LR-SP. Both of the algorithms are comprised of finding the shortest path and updating Lagrangian multipliers these two main steps. However, the main cause of the larger computational time for LR-SP is the negative cycle detection, adjusts of Lagrangian multipliers, and additional computations of the auxiliary variable y, which are not necessary in LS-SG. In terms of iteration number, LR-SG outperforms the proposed algorithm, which indicates that LR-SG algorithm reduces the objective value more drastically in each iteration than LS-SG does. The faster algorithm, LS-SG, takes more iterations to achieve similar duality gaps. To shed more light on the convergence of the algorithms, a line chart that shows how the relative duality gap reduces with number of iterations of LS-SG, LS-DE and LR-SP is presented (Figure 4.6). The test were performed on Barcelona network with $\eta = 3.0$ and $v = 0.50$. Maximum number of iterations was set to 50.

Table 4.9: Computation comparison between the proposed algorithm (LS-SG) and sub-gradient projection method (LR-SP)

Network	η	Average computational time (s)		Average relative gap (%)		Average number of iterations	
		LS-SG	LR-SP	LS-SG	LR-SP	LS-SG	LR-SP
Anaheim	1.0	0.21	0.63	0.091	0.076	10.0	4.6
	3.0	0.34	1.01	0.084	0.085	11.8	7.2
Chicago sketch	1.0	0.88	2.02	0.079	0.085	13.6	5.4
	3.0	1.34	3.14	0.077	0.088	16.2	8.4
Barcelona	1.0	1.58	2.97	0.086	0.082	14.8	4.8
	3.0	2.22	5.30	0.090	0.091	20.4	8.6

Duality gap of LR-SP drops faster than both LS-SG and LS-DE at the beginning of the algorithm, which explains why LR-SP takes fewer iterations in achieving 0.1% gap than LS-SG. However, it is noted that the final gap of LR-SP is greater than LS-SG and LS-DE. The LR-SP line crosses the LS-DE and LS-SG line at around 20 and 30 iterations, respectively. This shows that the proposed algorithms converge to a smaller gap than LR-SP. In this case, the relative duality gap at iteration 50 are 0.06%, 0.02% and 0.004% for LR-SP, LS-SG and LS-DE, respectively. Comparing LS-DE and LS-SP, it again reflects the fact that the deep-cut ellipsoid method has a better convergence rate and smaller gap compared with the sub-gradient projection Lagrangian updating method.

To sum up, the results reveal the following: LR-SG reduces the objective value faster initially than both of the proposed algorithm; LS-SG reduces computational time by around 58% as opposed to the LR-SG algorithm and improves the duality gap by around 67%, and LS-DE has an even better duality gap.

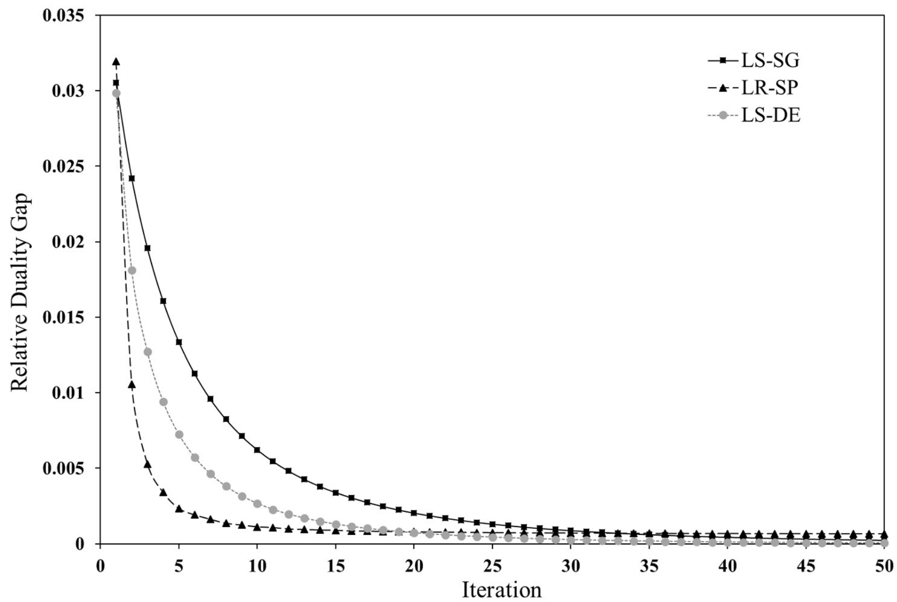


Figure 4.6: Convergence comparison of LR-SP, LS-SG and LS-DE

Chapter 5

Conclusion

In this thesis, we studied the mean-standard deviation reliable shortest path problem with link travel time correlations. Thanks to the matrix decomposition technique, it becomes possible to introduce continuous variables to establish the relationship between the covariance matrix and the original binary variables. Combined with the Lagrangian relaxation method, we decompose the non-linear and non-additive original problem into two sub-problems. One of which is a standard shortest path problem that can be solved efficiently with labeling algorithms and the other convex optimization problem is proved to show its optimal solution which further saves computational time. The complete algorithm was proposed with both sub-gradient and deep-cut ellipsoid method for updating Lagrangian multipliers. An illustrative example was provided to shed light on the mechanism of the proposed algorithm with two different Lagrangian multiplier updating strategies. Large-scale tests were also conducted to evaluate the computational performances. We compared the relative gap, running time and average iteration with the existing methods in the literature. From the large-scale test results, following conclusions are drawn:

- The proposed algorithm is more efficient in terms of computational time compared with OA with a speed improvement at least 98% for LS-SG, and it outperforms the sampling-based Lagrangian relaxation method in terms of the relative duality gap.
- The proposed algorithm converges to a smaller gap than the LR-SP algorithm in the literature. In addition, tests with the same input show that the proposed algorithm is computationally more efficient even though more iterations are usually needed.
- Large-scale networks experiments show that even though deep-cut ellipsoid usually takes

longer to find an optimal solution, it is promising for obtaining a better path compared with the sub-gradient method whose strength is finding a solution faster.

With the increasing research and practical interests on the variability of traffic conditions, the significant decrease in the computational time of the RSP problem provides promises for further comprehensive studies and real applications. The contributions of this study include the proposal of the Lagrangian substitution solution method based on eigendecomposition that has not been studied in the literature, the efficiency improvement from OA method and sampling based and sub-gradient projection based Lagrangian relaxation method, and the comparison between the widely adopted sub-gradient method and deep-cut ellipsoid method for Lagrangian multiplier updating and its implications.

Future research extensions include multiple directions. First, efficient matrix decomposition method can be a research subject because larger network size makes decomposition of the variance-covariance matrix computationally expensive. Second, closely related to the first direction, link travel time spatial correlation can be explored, limiting the correlation of link travel time may give some special structure of the variance-covariance matrix and provide simplicity in matrix decomposition. Third, as mentioned in the text, exploration on the critical value of η where the variability of path travel time dominates the average travel time can be an interesting problem for future study.

Bibliography

[1] Takashi Uchida and Yasunori Iida. Risk assignment: A new traffic assignment model considering the risk of travel time variation. *Transportation and Traffic Theory*, 1993.

[2] Tao Xing and Xuesong Zhou. Finding the most reliable path with and without link travel time correlation: A lagrangian substitution based approach. *Transportation Research Part B: Methodological*, 45(10):1660–1679, 2011.

[3] Weiliang Zeng, Tomio Miwa, Yukiko Wakita, and Takayuki Morikawa. Application of lagrangian relaxation approach to α-reliable path finding in stochastic networks with correlated link travel times. *Transportation Research Part C: Emerging Technologies*, 56:309–334, 2015.

[4] Yuli Zhang, Zuo-Jun Max Shen, and Shiji Song. Lagrangian relaxation for the reliable shortest path problem with correlated link travel times. *Transportation Research Part B: Methodological*, 104:501–521, 2017.

[5] Alireza Khani and Stephen D Boyles. An exact algorithm for the mean–standard deviation shortest path problem. *Transportation Research Part B: Methodological*, 81:252–266, 2015.

[6] Borzou Rostami, André Chassein, Michael Hopf, Davide Frey, Christoph Buchheim, Federico Malucelli, and Marc Goerigk. The quadratic shortest path problem: complexity, approximability, and solution methods. *European Journal of Operational Research*, 268(2):473–485, 2018.

[7] Hao Hu and Renata Sotirov. Special cases of the quadratic shortest path problem. *Journal of Combinatorial Optimization*, 35(3):754–777, 2018.

[8] Mehrdad Shahabi, Avinash Unnikrishnan, and Stephen D Boyles. An outer approximation algorithm for the robust shortest path problem. *Transportation Research Part E: Logistics and Transportation Review*, 58:52–66, 2013.

[9] Mehrdad Shahabi, Stephen D Boyles, et al. Robust optimization strategy for the shortest path problem under uncertain link travel cost distribution. *Computer-Aided Civil and Infrastructure Engineering*, 30(6):433–448, 2015.

[10] Ravi Seshadri and Karthik Srinivasan. Algorithm for determining most reliable travel time path on network with normally distributed and correlated link travel times. *Transportation Research Record: Journal of the Transportation Research Board*, (2196):83–92, 2010.

[11] Bi Yu Chen, William HK Lam, Agachai Sumalee, and Zhi-lin Li. Reliable shortest path finding in stochastic networks with spatial correlated link travel times. *International Journal of Geographical Information Science*, 26(2):365–386, 2012.

[12] Bi Yu Chen, William H. K. Lam, Agachai Sumalee, Qingquan Li, Hu Shao, and Zhixiang Fang. Finding Reliable Shortest Paths in Road Networks Under Uncertainty. *Networks and Spatial Economics*, 13(2):123–148, jun 2013.

[13] Bi Yu Chen, Qingquan Li, and William H.K. Lam. Finding the k reliable shortest paths under travel time uncertainty. *Transportation Research Part B: Methodological*, 94:189–203, dec 2016.

[14] Xing Wu and Yu Marco Nie. Modeling heterogeneous risk-taking behavior in route choice: A stochastic dominance approach. *Transportation research part A: policy and practice*, 45(9):896–915, 2011.

[15] Xing Wu. Study on mean-standard deviation shortest path problem in stochastic and time-dependent networks: A stochastic dominance based approach. *Transportation Research Part B: Methodological*, 80:275–290, 2015.

[16] YY Fan, RE Kalaba, and JE Moore. Arriving on time. *Journal of Optimization Theory and Applications*, 127(3):497–513, 2005.

[17] Yu Nie and Yueyue Fan. Arriving-on-time problem: discrete algorithm that ensures convergence. *Transportation Research Record: Journal of the Transportation Research Board*, (1964):193–200, 2006.

[18] Yu Marco Nie and Xing Wu. Shortest path problem considering on-time arrival probability. *Transportation Research Part B: Methodological*, 43(6):597–613, 2009.

[19] Hong K Lo, XW Luo, and Barbara WY Siu. Degradable transport network: travel time budget of travelers with heterogeneous risk aversion. *Transportation Research Part B: Methodological*, 40(9):792–806, 2006.

[20] David Watling. User equilibrium traffic network assignment with stochastic travel times and late arrival penalty. *European journal of operational research*, 175(3):1539–1556, 2006.

[21] Judith YT Wang, Matthias Ehrgott, and Anthony Chen. A bi-objective user equilibrium model of travel time reliability in a road network. *Transportation Research Part B: Methodological*, 66:4–15, 2014.

[22] Hu Shao, William Lam, Qiang Meng, and Mei Tam. Demand-driven traffic assignment problem based on travel time reliability. *Transportation Research Record: Journal of the Transportation Research Board*, (1985):220–230, 2006.

[23] Yudong Zhang, Yan Jun, Geng Wei, and Lenan Wu. Find multi-objective paths in stochastic networks via chaotic immune pso. *Expert Systems with Applications*, 37(3):1911–1919, 2010.

[24] Yuting Li and Linhan Guo. Multi-objective optimal path finding in stochastic time-dependent transportation networks considering timeliness reliability and travel expense. In *Prognostics and System Health Management Conference (PHM-Chengdu), 2016*, pages 1–6. IEEE, 2016.

[25] Anthony Chen and Zhaowang Ji. Path finding under uncertainty. *Journal of advanced transportation*, 39(1):19–37, 2005.

[26] Claude Lemaréchal. Lagrangian relaxation. In *Computational combinatorial optimization*, pages 112–156. Springer, 2001.

[27] Robert B Noland, Kenneth A Small, Pia Maria Koskenoja, and Xuehao Chu. Simulating travel reliability. *Regional science and urban economics*, 28(5):535–564, 1998.

[28] Carlos Carrion and David Levinson. Value of travel time reliability: A review of current evidence. *Transportation research part A: policy and practice*, 46(4):720–741, 2012.

[29] Marco Taboga. *Lectures on probability theory and mathematical statistics*. CreateSpace Independent Publishing Platform, 2012.

[30] Hervé Abdi. The eigen-decomposition: Eigenvalues and eigenvectors. *Encyclopedia of measurement and statistics*, pages 304–308, 2007.

[31] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[32] Lixing Yang and Xuesong Zhou. Constraint reformulation and a lagrangian relaxation-based solution algorithm for a least expected time path problem. *Transportation Research Part B: Methodological*, 59:22–44, 2014.

[33] Lixing Yang and Xuesong Zhou. Optimizing on-time arrival probability and percentile travel time for elementary path finding in time-dependent transportation networks: Linear mixed integer programming reformulations. *Transportation Research Part B: Methodological*, 96:68–91, 2017.

[34] Monique Guignard. Lagrangean relaxation. *Top*, 11(2):151–200, 2003.

[35] Marshall L Fisher. The lagrangian relaxation method for solving integer programming problems. *Management science*, 27(1):1–18, 1981.

[36] Robert G Bland, Donald Goldfarb, and Michael J Todd. The ellipsoid method: A survey. *Operations research*, 29(6):1039–1091, 1981.

[37] Jean-Louis Goffin. Variable metric relaxation methods, part ii: The ellipsoid method. *Mathematical Programming*, 30(2):147–162, 1984.

[38] Stephen Boyd and C Barratt. Ellipsoid method. *Notes for EE364B, Stanford University*, 2008, 2008.

[39] Hillel Bar-Gera. Transportation network test problems, 2002.

[40] Nicholas J Higham. Computing a nearest symmetric positive semidefinite matrix. *Linear algebra and its applications*, 103:103–118, 1988.

[41] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2009.