Finding the Most Reliable Path With and Without Link Travel Time Correlation:
A Lagrangian Substitution Based Approach

Tao Xing
Department of Civil and Environmental Engineering
University of Utah, Salt Lake City, UT 84112-0561
Email: tao.xing@utah.edu

Xuesong Zhou
Department of Civil and Environmental Engineering
University of Utah, Salt Lake City, UT 84112-0561
Email: zhou@eng.utah.edu
Tel: +1-801-585-6590
Fax: +1-801-585-5477
(Corresponding Author)

Abstract
Path travel time reliability is an essential measure of the quality of service for transportation systems and an important attribute in travelers’ route and departure time scheduling. This paper investigates a fundamental problem of finding the most reliable path under different spatial correlation assumptions, where the path travel time variability is represented by its standard deviation. To handle the nonlinear and nonadditive cost functions introduced by the quadratic forms of the standard deviation term, a Lagrangian substitution approach is adopted to estimate the lower bound of the most reliable path solution through solving a sequence of standard shortest path problems. A subgradient algorithm is used to iteratively improve the solution quality by reducing the optimality gap. To characterize the link travel time correlation structure associated with the end-to-end trip time reliability measure, this research develops a sampling-based method to dynamically construct a proxy objective function in terms of travel time observations from multiple days. The proposed algorithms are evaluated under a large-scale Bay Area, California network with real-world measurements.
1. Introduction

1.1. Motivation

As traffic systems can be viewed as stochastic processes with non-deterministic demand and capacity distributions, travel time reliability has been widely recognized as an important element of a traveler’s route and departure time scheduling. In recent years, operating agencies have begun to shift their focus more toward monitoring and improving the reliability of transportation systems through probe-based data collection, integrated corridor management and advanced traveler information provision. With a growing trend of incorporating trip time variability into traffic network analysis models, finding reliable path alternatives motivates substantial algorithmic development efforts.

For many common route finding criteria, such as physical distance and travel time, the (generalized) path cost functions are linear and additive across different links, so the resulting optimization problem can be directly solved by the standard label correcting or label setting shortest path algorithms. In an early study by Sen et al. (2001), the path travel time reliability is modeled as a linear combination of travel time mean and variance, and the resulting 0-1 quadratic integer program is solved by as a sequence of parametric subproblems. However, most end-to-end trip reliability measures, as discussed below, lead to nonlinear and nonadditive cost functions, which considerably increase the complexity and impose challenges for the path search procedures. A wide range of definitions and formulations have been proposed to measure travel time reliability, including (1) 90th- or 95th-percentile travel time, buffer and planning time index, (2) on-time arrival probability, (3) travel time variation expressed in terms of standard deviation or coefficient of variation. The first two definitions (1 and 2) are built on the probability distribution function of travel time. For example, in a report by Cambridge Systematics (2005), the 95th-percentile travel time is defined as the travel time within which 95th-percentile trips are completed, and the buffer and planning time indexes can be further calculated from the 90th- or 95th-percentile travel time. The on-time arrival probability measure, on the other hand, considers the percentage of trips that are completed within a reasonable buffered travel time (e.g. average travel time plus 20% buffer). In a study by Fan et al. (2005a), a path finding algorithm was proposed to minimize the probability of arriving at the destination later than a specified arrival time. Recently, Nie and Wu (2009a) developed solution algorithms with first-order stochastic dominance rules for the routing problem with on-time arrival reliability.

The third type of models, which characterizes the travel time reliability measure in terms of standard deviation, has been calibrated in various empirical studies (e.g. Small, 1982; Noland et al., 1998; Noland and Polak, 2002), and the corresponding utility function is also incorporated in dynamic traffic assignment models (e.g. Zhou et al., 2008). It is important to recognize that, within a Kalman filtering framework, which is the building block of real-time traffic state estimation and prediction systems (e.g. Ashok and Ben-Akiva, 1993; Zhou and Mahmassani, 2007), the variance of travel time estimates, and accordingly its standard deviation, can be analytically derived and calculated through a recursive estimation error propagation formula. In comparison, the first two types of reliability measures must be assessed by relatively complicated numerical probabilistic methods.
To allow further extensions in real-time traffic prediction and route guidance systems, we consider the most reliable path problem with a linear disutility function of mean trip travel time and its standard deviation: \( \min \{ \text{mean} + \beta \sqrt{\text{var}} \} \). In particular, we want to address two fundamental challenges introduced by this special functional form. First, the standard deviation of path travel time is not a linear summation of the standard deviation of related link travel times. Second, the square root transformation associated with the standard deviation term is, in fact, a concave function, so it is difficult to directly apply many convex programming techniques in this application.

1.2. Literature Review

Several previous pioneering research efforts have been devoted to addressing computational issues caused by nonlinear, nonadditive or concave objective functions in the shortest path problem. The early work by Henig (1986) presented efficient approximate methods on the shortest path problem with two criteria, which are assumed to be quasiconcave or quasiconvex utility functions. Scott and Bernstein (1997) developed an iterative solution method for the shortest path problem where the value of time function is nonlinear and non-decreasing. In their algorithm, the search space is decomposed to a series of resource-constrained shortest path subproblems, which can be solved by the Lagrangian relaxation technique (Handler and Zang, 1980). Gabriel and Bernstein (2000) further proposed a path-finding heuristic algorithm for the nonadditive path problem using a linear approximation reformulation.

By dualizing hard constraints to the objective function (Fisher, 1981), the Lagrangian relaxation method is a well-known solution procedure for integer programming problems. To further introduce separability in Lagrangian reformulations, one important extension of Lagrangian relaxation is the variable splitting and Lagrangian decomposition approach proposed by Joernsten and Naesberg (1986) and independently by Guignard and Kim (1987). This approach aims to split original variable \( x \) into the pair \( (x, y) \), and then link the auxiliary variable \( y \) with \( x \) through a linking constraint \( Ax = y \), which will be further relaxed to the objective function. Larsson et al. (1994) adapted this problem restatement approach to decompose a minimum cost network flow problem with a concave objective function to a standard linear minimum cost network flow subproblem and an easy-to-solve concave minimization problem. Along the same line, Tsaggouris and Zaroliagis (2004) combined the Lagrangian relaxation and hull approach to solve the nonadditive shortest path problem with a nonlinear, convex and non-decreasing cost function. Readers interested in general Lagrangian relaxation methods are referred to the review paper written by Guignard (2003).

In stochastic routing problems, spatial and temporal dependences have been exclusively studied by a number of researchers, and we refer interested readers to the comprehensive studies by Miller-Hooks and Mahmassani (2000) and Nie and Wu (2009a) on the a priori time-varying least travel time problem. Considering spatial dependence in terms of congestion level and state transfer probability, Fan et al. (2005b) proposed a multistage adaptive feedback control process to address shortest path problem with correlated link costs. Recently, limited spatial and temporal dependences have been considered by Boyles and Waller (2007) for the nonlinear disutility
shortest path problem, and by Nie and Wu (2009b) for the reliable routing problem. Specifically, the above studies characterize the randomness of link travel time by using certain probability density functions abstracted from a historical database, and incorporate limited spatial correlation through a Markovian model that considers the transition probabilities of link states.

1.3. Proposed Approach

To meet the challenges of developing efficient path search algorithms, we develop a new lower bound method to quantify the solution quality and accordingly guide the iterative search process to find close-to-optimal solutions. This approach is built on a reformulation of the complicating objective function using Lagrangian relaxation. Under different assumptions of data availability, two problem formulations are considered in this study: finding the most reliable path with and without link correlation. Specifically, with the same measure of reliability through standard deviations, the first model in this paper adapts independent distributions of link travel time and considers no link correlations, while the second model incorporates the spatial dependencies via a set of travel time samples from historical records. Separate lower bound methods are derived for both models within a Lagrangian substitution framework.

It is important to recognize the significance of optimal solutions and lower bound estimates for both theoretical development and practical applications. Obviously, any feasible solution can serve as an upper bound on the optimal solution for a general minimization problem. Many heuristic algorithms (e.g. the K-shortest path algorithm or the A* algorithm) can be used to generate feasible solutions and potentially tight upper bounds on the most reliable path problem. However, heuristic methods either require an inordinate amount of running time and memory space to enumerate possible solutions or produce results with no guarantee on solution accuracy. For example, a solution from a heuristic route generation algorithm could be 1%, 5%, or 20% away from the optimum, and simply accepting the current best solution without evaluating its solution quality could lead to unsatisfactory level-of-service for end-users and significant inefficiency in system-wide traffic flow management.

This study aims to seamlessly incorporate and significantly enhance several modeling/algorithic components from several previous studies. Based on the variable splitting approach in the Lagrangian reformulation framework, we first replace the complex quadratic portion of the objective function with equivalent equality constraint(s) to remove the nonadditivity, and the auxiliary constraint(s) can be further dualized to a simplified objective function that leads to easy subproblems. In particular, one integer subproblem involving linear link cost functions can be efficiently solved by standard shortest path algorithms, while another subproblem containing the concave square root objective function with a single variable can be solved analytically by checking the boundary values in the feasible region. The similar bounding technique was used by Larsson et al. (1994).

Another significant contribution of this research is an efficient and practical incorporation of spatial network correlations. Different from traditional methods focusing on the probability density functions of link travel times, this paper proposes a sampling-based algorithm to consider the spatial dependencies among links. By directly utilizing readily available historical travel time
measurements from traffic monitoring systems, the proposed approach can systematically incorporate the inherent spatial correlation into the reliable route searching process.

The remainder of this paper is structured as follows. The next section provides the formal problem statement and briefly discusses two different models for the most reliable path problem: with and without link correlation. Focusing on each of the two different models, sections 3 and 4 present the theoretical derivations and algorithmic development in detail, followed by illustrative examples. Finally, section 5 evaluates the performance of proposed algorithms through numerical experiments on a large-scale network.

2. Problem Statement and Model Assumptions

The following notation is used to represent variables in the problem formulation:

- $\beta$ = reliability coefficient
- $N$ = set of nodes
- $A$ = set of links
- $p$ = path index
- $m$ = link index in a path,
  - $c_p$ = travel time of path $p$
  - $\overline{c}_p$ = mean travel time of path $p$
- $i, j$ = subscript for node index
  - $l = \text{subscript for the index of a link in a path, } l = 1, \ldots, m$
- $a_l$ = link in path $p$, with index $l$
- $a_{ij}$ = directed link from node $i$ to $j$
- $c_{ij}$ = travel time of link $a_{ij}$
- $\overline{c}_{ij}$ = mean travel time of link $a_{ij}$
- $f(c_{ij})$ = probability distribution function of $c_{ij}$
- $f(c_{ij})$ = probability distribution function of $c_{ij}$
- $\sigma^2_{ij}$ = variance of link travel time $c_{ij}$
- $x_{ij}$ = binary variable that indicates link $a_{ij}$ is included in path solution if $x_{ij} = 1$
- $X = \text{set of binary variables } \{x_{ij} | ij \in A\}$
- $D$ = set of travel time measurement samples
- $n$ = number of samples in set $D$
- $d$ = subscript for samples, $d=1,\ldots,n$.
  - $c_{p,d}$ = travel time of path $p$ in sample $d$
  - $c_{l,d}$ = travel time of link $l$ in sample $d$
  - $c_{ij,d}$ = travel time of link $(i, j)$ in sample $d$
2.1. Problem Statement

Let $G (N, A)$ represent a transportation network, where $N$ is the set of nodes and $A$ is the set of links. Each link can be denoted as either a directed link $a_{ij}$ from node $i$ to $j$, or an indexed link $a_j$ in a path $p$ with $m$ links. Accordingly, the travel time of each link is denoted as $c_{ij}$ or $c_j$. The travel time of each link $(i, j)$ can be described as a random variable with probability distribution function $f(c_{ij})$ which has a mean of $\overline{c_{ij}}$ and a variance of $\sigma_{ij}^2$. Generally, the mean and variance of link travel times evolve considerably depending on the time of day and underlying traffic congestion levels. For simplicity, this study focuses on the static shortest path problem (to be used in static traffic assignment) and considers link travel times as time-invariant parameters in the underlying study horizon (e.g. morning peak hours).

Consider binary variable $x_{ij} \in \{0,1\}$ that indicates the selection of link $a_{ij}$ for the optimal path solution, the least mean travel time problem for a pre-specified OD pair $(o, d)$ is described as

$$
\begin{align*}
z^* = \min \sum_{j \in A} \overline{c_{ij}} x_{ij} \\
\text{Subject to the following flow balance constraints:}
\end{align*}
$$

$$
\sum_{j \in A} x_{ij} - \sum_{j \in A} x_{ji} = 
\begin{cases}
1 & i = o \\
0 & i \in N - \{o, d\} \\
-1 & i = d
\end{cases}
$$

The above integer linear program can be solved using regular label correcting or label setting shortest path algorithms (Ahuja et al., 1993).

In this paper, we formulate the most reliable path problem (P) as the following non-linear integer programming problem by combining mean path travel time and its standard deviation in the objective function:

$$
\begin{align*}
z^* = \min \sum_{j \in A} \overline{c_{ij}} x_{ij} + \beta \sqrt{\text{var}(c_p)},
\end{align*}
$$

subject to

$$
\sum_{j \in A} x_{ij} - \sum_{j \in A} x_{ji} = b
$$

where $b = \begin{cases}
1 & i = o \\
0 & i \in N - \{o, d\} \\
-1 & i = d
\end{cases}$ represents the flow status for each node $i$ in the network, and $\beta$ is the reliability coefficient, which reflects the significance of travel time variability. It can be derived as the ratio of Value of Reliability (VOR) and Value of Time (VOT). The reliability coefficient could also vary across different travelers and different trip purposes (e.g. business trip vs. recreational trip), and a typical value can be 1.27, calibrated by Noland et al. (1998).

Given link travel time statistics, the mean and variance of path travel time can be derived as
Obviously, it is computationally intractable to obtain the multi-dimension probability distribution function \( f(c_1, c_2, \ldots, c_m) \) for each path \( p \), especially when spatial correlation exists. Along this line, two different approximation modeling approaches are proposed below to calculate the path travel time variance, with and without link travel time correlation assumptions.

### 2.2. Finding Most Reliable Path without Link Travel Time Correlation: Independent Distribution Based Model

To consider travel time variance in the most reliable path problem, one simplifying approach is to assume that there is no spatial correlation among travel times on different links. That is, by assuming the link travel time distributions are independent, we can reduce Eq. (5) to

\[
\var(c_p) = \sum_{i=1}^{m} \left( \int_{c_i=0}^{+\infty} (c_i - \bar{c}_i)^2 f(c_i)dc_i \right) = \sum_{i=1}^{m} \var(c_i) \tag{6}
\]

so that the variance of the path travel time is now expressed as the sum of independent link travel time variances, which are relatively easy to measure based on historical traffic databases with multi-day observations. Given independent link travel time distributions, the most reliable path problem (P) is then rewritten as

\[
z^*_1 = \min \sum_{g \in A} c_g x_g + \beta \sqrt{\sum_{g \in A} \sigma_g^2 x_g} \quad \text{s.t.} \quad \sum_{j \in i \in A} x_{ij} = b \tag{7}
\]

### 2.3. Finding Most Reliable Path with Link Travel Time Correlation: Sampling-based Model

In reality, travel times among different links could be highly correlated, e.g. due to the propagation of traffic congestion from a lane drop or merge bottleneck to its upstream links along a freeway or arterial corridor. In order to explicitly consider the link correlation in path travel time variable calculation, a sampling-based approximation method is adapted in this research to formulate the most reliable path problem.

According to the Monte Carlo method, a continuous stochastic distribution can be approximated as a discrete function by taking \( n \) samples from its population:
\[
\text{var}(c_p) \approx \frac{1}{n-1} \sum_{d=1}^{g} (c_{p,d} - \bar{c}_p)^2 \\
= \frac{1}{n-1} \sum_{d=1}^{g} \left( \sum_{i=1}^{n} c_{i,d} - \sum_{i=1}^{n} \bar{c}_i \right)^2
\]

That is to say, one can take \( n \) days’ samples from a multi-day historical traffic database and use them to directly calculate the sample variance and sample standard deviation of path travel time. By doing so, the inherent correlation among link travel times has been automatically represented by the sample set without explicitly requiring the variance-covariance matrix.

According to the sampling approach in Eq. (8), the most reliable path problem \( P \) that allows link travel time correlation is reformulated as

\[
z_2^* = \min \sum_{y \in A} \bar{c}_y x_y + \beta \left\{ \frac{1}{n-1} \sum_{d=1}^{g} \left( \sum_{y \in A} c_{y,d} x_y - \sum_{y \in A} \bar{c}_y x_y \right)^2 \right\}
\text{ s.t. } \sum_{j \in \text{in} y} x_j - \sum_{j \in \text{out} y} x_j = b
\]

where \( \bar{c}_y = \frac{1}{n} \sum_{d=1}^{g} c_{y,d} \) is the sample mean of link travel time.

Compared to the model assuming independent link travel time distributions, the proposed sampling-based method obviously requires a much larger number of measurements across different days to reduce sampling error and capture any possible spatial correlation. This approach fully recognizes link travel time correlation in calculating path travel time variance, but it also considerably complicates the path search process, especially when an acceptable level of approximation accuracy is needed. For real-world applications lacking sufficient link travel time measurements, the first model without link travel time correlation is still a viable option, but we also need to recognize that it might not find the most reliable path (i.e. optimal solution) in a network with possible spatial correlation.

For solving the two models proposed above, two separate lower bound algorithms are addressed with the similar Lagrangian relaxation-based approach in the next two sections. In particular, the complex quadratic parts of the two problems in Eq. (7) and (9) are first replaced with auxiliary variables and equivalent equality constraints. Then, based on Lagrangian substitution, the auxiliary constraints are further dualized into the simplified objective functions and lead to easy-to-solve and variable-independent subproblems. Specifically, the new set of subproblems contains one integer program with linear and additive cost functions, one single-variable concave function that is solvable by checking the boundary values in the feasible region, and a set of convex subproblems particularly for the sampling-based model. Finally, the subgradient method is applied on both algorithms to update the lower bound.

3. Algorithm for Finding Most Reliable Path without Link Travel Time Correlation

In this section, an algorithm is presented for finding the most reliable path without assuming spatial dependency of link travel times. The Lagrangian relaxation-based model reformulation...
and derivation are first described in detail, followed by the subgradient method implementation and illustrative example demonstration.

### 3.1. Model Reformulation Using Lagrangian Substitution Method

As shown in the optimization program (7), the standard deviation of path travel time is a nonlinear, concave and nonadditive function of link travel time variance. The nonadditivity violates Bellman's Principle of Optimality, which forms the basis for standard label setting or label correcting shortest path algorithms. The overarching goal of the following model reformulation is to approximate the complicating nonadditive objective function by a linear additive cost function that is suitable for the regular shortest path algorithm. To remove the nonadditivity on decision variable \( x \), we first introduce a nonnegative auxiliary variable \( y \) to the program (7) to convert the variance term to an equality constraint (11):

\[
\begin{align*}
\text{min} & \quad \sum_{ij \in A} c_{ij} x_{ij} + \beta \sqrt{y} \\
\text{s.t.} & \quad \sum_{ij \in A} \sigma_{ij}^2 x_{ij} = y \\
& \quad \sum_{j' \in A} x_{ij'} - \sum_{j \in A} x_{j'i} = b \\
& \quad 0 \leq y \leq y'
\end{align*}
\]

where \( y' \) is the path travel time variance of the least expected travel time path.

![Fig. 1. Feasible region of optimal path travel time variance](image)

**Lemma 1**: The feasible interval of optimal path travel time variance is \([0, y']\), i.e. between 0 and the variance of the least expected travel time path \( y' \).

Proof: Shown in Fig.1, for any optimal solution to the most reliable path problem, its mean travel time is no less than that of the least expected travel time path. Moreover, it cannot have a path travel time variance greater than that of the least expected travel time path. Otherwise, this path has worse mean travel time and travel time variance compared to the least expected travel time path, which means that it is dominated by the least expected travel time path and should not be
the optimal solution for optimization problem \( \min \left\{ \text{mean} + \beta \sqrt{\text{var}} \right\} \). Since the variance of path travel time is always greater than or equal to zero, the feasible region of \( y \) is an interval between 0 and the variance of the least expected travel time path \( y' \).

After the reformulation in Eqs. (10) and (11), the original optimization program is transferred to a constrained shortest path problem with a linear cost function \( \sum_{y \in A} c_{ij} x_{ij} \) and a univariate function \( U(y) = \beta \sqrt{y} \) which is concave and monotonically increasing.

In a Lagrangian relaxation modeling framework, we can further relax the equality constraint (11) to an inequality (with a larger feasible region for \( x \)):

\[
\sum_{y \in A} \sigma_{y}^{2} x_{y} \leq y
\]  
(12)

and then introduce a Lagrangian multiplier \( \mu \geq 0 \) to bring the definitional linear constraint back to the original objective function (10):

\[
\min \sum_{y \in A} c_{y} x_{y} + \beta \sqrt{y} + \mu \left( \sum_{y \in A} \sigma_{y}^{2} x_{y} - y \right)
\]

s.t. \( \sum_{j \in A} x_{y} - \sum_{j \in A} x_{ji} = b \)

\[0 \leq y \leq y'\]

By further re-grouping variables, we can obtain the following Lagrangian dual function of the original primal problem (7),

\[
L(\mu) = \min \left\{ \sum_{y \in A} \left( c_{y} + \mu \sigma_{y}^{2} \right) x_{y} + \beta \sqrt{y} - y : \sum_{j \in A} x_{y} - \sum_{j \in A} x_{ji} = b, 0 \leq y \leq y' \right\}
\]

(14)

The Lagrangian function (14) is called a dual problem, in contrast to the primal problem (7), which can be divided and solved by two independent sub-functions:

\[
L(\mu) = L_{x}(\mu) + L_{y}(\mu)
\]

(15)

The first sub-function \( L_{x}(\mu) \) involves a linear combination of primal variables \( x \) with a new link cost function \( c_{y} + \mu \sigma_{y}^{2} \), and the resulting additive shortest path problem can be solved efficiently using label correcting or label setting algorithms (Ahuja et al., 1993).

\[
L_{x}(\mu) = \min \left\{ \sum_{y \in A} \left( c_{y} + \mu \sigma_{y}^{2} \right) x_{y} : \sum_{j \in A} x_{y} - \sum_{j \in A} x_{ji} = b \right\}
\]

(16)

The second sub-function \( L_{y}(\mu) \) is a univariate concave minimization problem with respect to auxiliary variable \( y \), and it can be solved through Lemma 1 by selecting the boundary value \( y' \). The least expected travel time path can be obtained by solving the optimization problem with objective function (1) and constraint set (2).

\[
L_{y}(\mu) = \min \left\{ \beta \sqrt{y} - \mu y : 0 \leq y \leq y' \right\}
\]

(17)
The optimal value of the concave function is \( \min \{0, \beta \sqrt{y'} - \mu y'\} \) because the optimal value of a concave function is attained at one of the extreme points of the feasible region (a similar modeling technique was used in Larsson et al., 1994), as shown in Fig. 2. Note that, for a general multi-dimensional concave minimization problem, its optimal solution is found by enumerating a large number of extreme points, while our proposed solution algorithm takes advantage of the fact that the simple concave function in Eq. (17) only involves a single variable.

![Fig. 2. \( L_y(\mu) \) as a function of Lagrangian multiplier \( \mu \) and optimal value obtained at one of the extreme points.](image)

In the Lagrangian dual problem in Eq. (14), essentially, the original nonadditive and nonlinear integer program (7) is approximated (lower-bounded) by a shortest path problem with linear and additive cost functions and another concave problem. In this new problem, the Lagrangian multiplier \( \mu \) corresponds to the weight of the travel time variance in its reconstructed objective function, and its optimal value can be identified by some iterative search methods to be described below.

### 3.2. Subgradient Method

For each positive value of Lagrangian multiplier \( \mu \), the corresponding value of the Lagrangian function \( L(\mu) \) provides a lower bound to the optimal objective function value \( z^* \) of the primal problem (7). Let us denote \( L^* \) to be the maximum value of \( L(\mu) \) according to \( \mu \):

\[
L^* = \max_{\mu} L(\mu)
\]

Let us define \( \varepsilon \) as the gap or tolerance level between the lower bound \( L^* \) and the upper bound of the optimal solution, while the upper bound \( UB \) can be derived based on a feasible solution. Since the true optimal solution to the primal problem must have an objective value within the lower bound \( L^* \) and the upper bound, the approximation error of the current best solution (corresponding to the upper bound \( UB \)) is no larger than the gap \( \varepsilon \) with respect to the primal
optimal value $z^\ast$. To reduce approximation error ($UB - L$), the Lagrangian multiplier $\mu$ and the auxiliary variable $y$ can be determined iteratively through the following subgradient method.

The search direction of $\mu$ is typically calculated from the subgradient of $L(\mu)$:

$$\nabla L(\mu) = \sum_{ij \in A} \sigma_y^2 x_{ij} - y$$  \hspace{1cm} (19)

Let $k$ denote the iteration number in the subgradient method. Starting from any feasible initial choice of the Lagrangian multiplier, to update the Lagrangian multiplier $\mu^k$ at iteration $k$, the sub-functions (16) and (17) must be solved first, with solutions denoted as $x_{ij}^k$ and $y^k$, respectively. With solutions at iteration $k$, we calculate the value of the Lagrangian multiplier as follows:

$$\mu^{k+1} = \mu^k + \theta^k \left( \sum_{ij \in A} \sigma_y^2 x_{ij}^k - y^k \right)$$  \hspace{1cm} (20)

A heuristic algorithm can be used to update the step size $\theta^k$.

$$\theta^k = \frac{\lambda^k \left[ UB - L(\mu^k) \right]}{\left| \sum_{ij \in A} \sigma_y^2 x_{ij}^k - y^k \right|^2}$$  \hspace{1cm} (21)

In this expression, $UB$ is computed as the current best objective function value $z^\ast$ in the primal problem and can be updated iteratively to speed up the optimization process. $\lambda^k$ is a scalar chosen between 0 and 2 to adjust the step-size of the process and make sure no negative value appears in the cost function.

As illustrated in Fig. 3, the overall algorithm for solving the most reliable path problem without link travel time correlation is described below.

**Algorithm 1:**

**Step 1: Initialization**

Choose an initial Lagrangian multiplier $\mu > 0$;

Initialize iteration number $k=0$;

Solve the least expected travel time path problem, set its objective function value as $UB$, and set the variance of the least travel time path as $y'$.

**Step 2: Solve decomposed dual problems**

Solve $L_x(\mu) = \min \left( \sum_{ij \in A} (c_{ij} + \mu \sigma_y^2) x_{ij} \right)$ using a standard shortest path algorithm;

Solve $L_y(\mu) = \min \left\{ 0, \beta \sqrt{y'} - \mu y' \right\}$;
Calculate primal, dual and gap values.

**Step 3: Update Lagrangian multiplier**

Calculate Lagrangian multiplier \( \mu \) with Eqs. (20-21)

**Step 4: Termination condition test**

If \( k > K_{\text{max}} \) or the gap is smaller than the predefined toleration gap \( \varepsilon \), terminate the algorithm, otherwise go back to Step 2. Here, \( K_{\text{max}} \) is the predefined maximum number of iterations.

---

**3.3. Relative Gap and Optimal Solution**

The duality gap is defined as the difference between the primal optimal value \( z^* \) and the dual optimal value \( L^* \), and it offers an important metric for evaluating the performance of solution methods. If the duality gap is equal to zero, then the primal solution is an optimal solution for sure.

In practice, the real primal and dual optimal values cannot be achieved directly. Therefore, at each iteration of the above searching process, we update the tight upper bound of the primal problem with the shortest path solution of the linear integer subproblem in Eq. (16). Upon the termination of the algorithm, the minimal value of the primal problem serves as the tightest upper bound (UB), while the maximal value of the dual problem serves as the best lower bound (LB). To evaluate the solution quality, we can define a relative optimality measure as

---

Fig. 3. Subgradient algorithm for solving Lagrangian dual problems.
\[ \varepsilon' = \frac{UB - LB}{UB}, \quad (22) \]

while \( \frac{UB - LB}{UB} \geq \frac{Z^* - L^*}{Z^*} \), meaning that this relative gap is no less than the real gap between the optimal primal and dual values. That is, if there exists any optimal solution other than the best solution found in the algorithm, its improvement on the route travel time reliability is no larger than the relative gap. Thus, with a reasonably small relative gap, we provide a satisfied solution quality guarantee on the suggested reliable routes. It is important to notice that, due to the nature of the approximation from the Lagrangian relaxation-based lower bound estimation method, there could still be a positive gap even if the optimal solution of the primal problem has been achieved.

3.4. Illustrative Example

Consider a single origin-destination pair with three parallel links/paths with the following path travel time data:

**Table 1**

<table>
<thead>
<tr>
<th>Path</th>
<th>Mean Travel Time</th>
<th>Travel Time Variance</th>
<th>Travel Time Standard Deviation</th>
<th>Value of Objective Function (( \beta = 1 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>B</td>
<td>29</td>
<td>49</td>
<td>7</td>
<td>36</td>
</tr>
<tr>
<td>C(opt)</td>
<td>31</td>
<td>4</td>
<td>2</td>
<td>33</td>
</tr>
</tbody>
</table>

By simply comparing the values of objective function, it is obvious that path C is the most reliable path for objective function: \( \text{mean} + \beta \sqrt{\text{var}} \). Fig. 4 shows the relationship between the Lagrangian multiplier and the value of objective function.
Fig. 4. Solution results for independent distribution-based algorithm.

When the Lagrangian multiplier is equal to 0.142, the best lower bound is found at 31.57, while the tightest upper bound is found at 33 for the primal problem. Path C, identified as the best solution by the proposed algorithm (with results shown in Fig. 4), has a relative gap of 4.33% to the best lower bound (31.57). As mentioned previously, although path C is in fact the optimal solution in this example, there still exists a positive duality gap between primal and dual solutions, as the Lagrangian relaxation-based lower bound estimation method is, essentially, an approximation.

4. Algorithm for Finding Most Reliable Path with Link Travel Time Correlation

In last section, we presented the model reformulation and solution algorithm for the most reliable path problem without travel time correlation. In order to take the link travel time correlation into account, a Monte Carlo-based approximation method is used to propose the sampling-based model in Eq. (9). Along this line, given the same transportation network $G (N, A)$, we construct a sample set $D$ with $n$ travel time measurements from the same time at the same day-of-week. The sample domain is denoted with the subscript $d$ for variables. Similar to the independent distribution-based model, because of the non-linear and nonadditive characteristics of the objective function, a Lagrangian substitution method is adapted here to solve the sampling-based model.

4.1. Lagrangian Substitution
In order to approximate the minimization problem Error! Reference source not found. with a linear optimization problem, we implement a two-step Lagrangian relaxation approach with two sets of auxiliary variables:

\[
\sum_{y \in A} c_{y,d}x_{y} - \sum_{y \in A} \bar{c}_{y}x_{y} = w_{d} \quad \forall d \in D
\] (1)

\[
\frac{1}{n-1} \sum_{d=1}^{n} w_{d}^{2} = y
\] (2)

After relaxing the equality constraints in Eq. (1) and (2) with inequality constraints, the minimization problem can be reformulated as

\[
z^{*} = \min \sum_{y \in A} \bar{c}_{y}x_{y} + \beta \sqrt{y}
\] (3)

s.t.

\[
\sum_{y \in A} c_{y,d}x_{y} - \sum_{y \in A} \bar{c}_{y}x_{y} \leq w_{d} \quad \forall d \in D
\] (4)

\[
\frac{1}{n-1} \sum_{d=1}^{n} w_{d}^{2} \leq y
\] (5)

\[
\sum_{y \in A} x_{y} - \sum_{y \in A} x_{y} = b
\]

\[
0 \leq y \leq y'
\]

In this reformulation, \( n+1 \) auxiliary variables are introduced. The variable \( w_{d} \) for each sample \( d \) represents the difference between the mean path travel time and the path travel time on sample \( d \). The variable \( y \) corresponds to the average path travel time deviation between samples and the sample mean, which defines the path travel time variance.

To further remove constraints in Eq. (4) and (5), a set of positive Lagrangian multipliers, denoted as \( \mu_{d} \) and \( \nu \) sequentially, is introduced in order to move the explicit inequality constraints into the objective function (3):

\[
L(\mu_{1},...,\mu_{n},\nu)
\]

\[
= \min \left\{ \sum_{y \in A} \bar{c}_{y}x_{y} + \beta \sqrt{y} + \sum_{d=1}^{n} \mu_{d} \left( \sum_{y \in A} c_{y,d}x_{y} - \sum_{y \in A} \bar{c}_{y}x_{y} - w_{d} \right) + \nu \left( \frac{1}{n-1} \sum_{d=1}^{n} w_{d}^{2} - y \right) \right\}
\] (6)

By regrouping the variables, we will have a clearer view on the components of the dual problem:
\[ L(\mu_1, \ldots, \mu_n, \nu) \]
\[
= \min \left\{ \sum_{j \in A} \left[ \left( 1 - \sum_{d=1}^{n} \mu_d \right) c_{ij} + \sum_{d=1}^{n} \mu_d c_{ij,d} \right] x_{ij} : \sum_{j \in A} x_{ij} = b, 0 \leq y \leq y' \right\} \quad (7)
\]

The dual function (7) has a linear objective function corresponding to the primal variable \( X \). For each link, the cost function is a combination of weighted sample travel times on different days. By adjusting the Lagrangian multipliers \( \mu_d \) and \( \nu \), we may iteratively construct an optimal linear cost function that will maximize the dual function and, more specifically, best approximate the non-linear objective function in the primal problem.

4.2. Dual Function Decomposition

We decompose the dual function (29) into a set of sub-functions:

\[
L_s(\mu_1, \ldots, \mu_n, \nu) = \min \left\{ \sum_{j \in A} \left[ \left( 1 - \sum_{d=1}^{n} \mu_d \right) c_{ij} + \sum_{d=1}^{n} \mu_d c_{ij,d} \right] x_{ij} : \sum_{j \in A} x_{ij} = b \right\} \quad (8)
\]

\[
L_{w_d}(\mu_1, \ldots, \mu_n, \nu) = \min \left\{ \frac{1}{n-1} \nu w_d^2 - \mu_d w_d \right\} \quad \forall d \in D \quad (9)
\]

\[
L_y(\mu_1, \ldots, \mu_n, \nu) = \min \left\{ \beta \sqrt{y - \nu y'} : 0 \leq y \leq y' \right\} \quad (10)
\]

The first sub-function (8) can be easily solved using shortest path algorithms. Notice that in this sub-shortest path problem, the cost function for each link is a weighted combination of mean travel time and travel time of each day. In other words, the set of Lagrangian multiplier \( \mu \) indicates the weights of the travel time variance for each sample. By adjusting the set of multipliers, the influences of different samples on the overall path travel time reliability measure are evaluated and assessed systematically at each iteration.

The second sub-function set (9) contains one convex minimization problem for each auxiliary variable \( w_d \) and can be solved using the first-order gradient:

\[
\frac{\partial L_{w_d}}{\partial w_d} = \frac{2}{n-1} \nu w_d - \mu_d = 0
\]

\[
w_d = \frac{\mu_d (n-1)}{2\nu} \quad (11)
\]

The third sub-function (10) is a concave minimization problem for variable \( y \). Since \( y \) represents the variance of the path travel time, the feasible region is between zero and the variance of the path with least travel time, and the minimization point locates at one of the extreme points of the feasible region, i.e. \( L_y(\mu_1, \ldots, \mu_n, \nu) = \min \left\{ 0, \beta \sqrt{y' - \nu y'} \right\} \).

4.3. Subgradient Method
The subgradient method is also used in the second algorithm considering spatial correlation. The search direction for each Lagrangian multiplier is found using the following equations.

\[
\nabla L(\mu_1, \ldots, \mu_n, \nu) = \left( \sum_{ij \in A} (c_{ij} - \bar{c}_y) x_{ij} - w_i, \ldots, \sum_{ij \in A} (c_{ij} - \bar{c}_y) x_{ij} - w_n, \frac{1}{n-1} \sum_{d=1}^{n} w_d^2 - y \right)
\]

(34)

\[
\mu_d^{k+1} = \mu_d^k + \theta_{\mu_d}^k \left( \sum_{ij \in A} (c_{ij} - \bar{c}_y) x_{ij}^k - w_d^k \right) \quad \forall d \in D
\]

(35)

\[
\nu^{k+1} = \nu^k + \theta_{\nu}^k \left( \frac{1}{n-1} \sum_{d=1}^{n} (w_d^k)^2 - y^k \right)
\]

(36)

The step size of each iteration \(k\) is calculated using heuristic algorithms:

\[
\theta_{\mu_d}^k = \lambda_{\mu_d}^k \left[ L_{UB}(\mu_1, \ldots, \mu_n, \nu) - L(\mu_1^k, \ldots, \mu_n^k, \nu) \right] \quad \forall d \in D
\]

(37)

\[
\theta_{\nu}^k = \lambda_{\nu}^k \left[ L_{UB}(\mu_1, \ldots, \mu_n, \nu) - L(\mu_1^k, \ldots, \mu_n^k, \nu) \right] \quad \frac{1}{n-1} \sum_{d=1}^{n} (w_d^k)^2 - y^k
\]

(38)

Similar to Algorithm 1, an iterative algorithm for solving the most reliable path problem with link correlations are written as the following.

**Algorithm 2:**

**Step 1: Initialization**

Choose initial values for the set of positive Lagrangian multipliers, \(\mu_d\) and \(\nu\);

Initialize iteration number \(k=0\);

Solve the least expected travel time path problem, set its objective function value as \(UB\), set variance of the least travel time path as \(y'\).

**Step 2: Solve decomposed dual problems**

Solve the first sub-function (30) using a standard shortest path algorithm;

Solve the second sub-function set (31) using Eq. (33);

Solve the third sub-function (32) with \(L_j(\mu_1, \ldots, \mu_n, \nu) = \min \left\{ 0, \beta \sqrt{y' - \nu y'} \right\}\)

Calculate primal, dual and gap values.

**Step 3: Update Lagrangian multipliers**
Calculate Lagrangian multipliers with Eqs. (34-38)

**Step 4: Termination condition test**

If \( k > K_{\text{max}} \) or the gaps are smaller than the predefined toleration gap, terminate the algorithm, otherwise go back to Step 2.

**4.4. Illustrative Example**

Consider an illustrative network with two paths from a single origin to destination pair, where two paths each contain two links and travel time measurements from four days, as shown in Fig. 5.

![Illustrative network and daily measurements for sampling-based model.](image)

The path travel times are calculated in Table 2 below, and path A was found to be the optimal path for the most reliable path model with link travel time correlation.

**Table 2**

<table>
<thead>
<tr>
<th>Path</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Mean Travel Time</th>
<th>Travel Time Variance</th>
<th>Value of Objective Function (( \beta = 1 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (opt)</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3.25</td>
<td>0.25</td>
<td>3.75</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1.33</td>
<td>4.15</td>
</tr>
</tbody>
</table>

Fig. 6 shows the change of the dual value, Eq. (29), depending on different combinations of \( \mu \). To visualize the 5-dimentional result in 3-dimentional figures, we set \( \mu_1 = \mu_2 = \mu_3 = \mu_4 \). The maximum dual value (3.05) is achieved at \( \mu = (0.1, 0.1, 0.1, 0, 0) \) and \( v = 0.1 \). The best primal value found in the search process is 3.75 for path A. A relative gap of 16\% exists due to the approximation characteristics of the Lagrangian relaxation approach.
5. Numerical Experiments

5.1. Test Network Overview

In this section, numerical experiments are conducted on a large-scale real-world transportation network for the Bay Area, California, which is comprised of 53,124 nodes and 93,900 links. Specifically, 8,511 links (9.1% of links) of the entire network are freeways with a total length of 1,774.8 miles (i.e., 15.8% of the total mileage); while 85,389 links (90.9%) are arterial roads with a total length of 9,431.8 miles (84.2%). The algorithm is implemented in C# on the Windows Vista platform and evaluated on a personal computer with an Intel Core Duo 1.8GHz CPU and 2 GB memory.

For the first model without considering link travel time correlation, the mean travel time and travel time variance for each link are calculated from available historical travel time records from the NAVTEQ traffic database, and the data used in this study (mainly from freeway segments) cover about 4.1% of the total mileage in the Bay Area (Fig. 7). Note that the underlying network includes a large number of major and minor arterial streets which do not have temporally continuous traffic observations for us to calculate variability statistics. For the second sampling-based model that recognizes spatial correlation, 73 days of travel time measurements between November 2009 and February 2010 are collected for the time interval of 9:00 AM to 9:15 AM of

Fig. 6. Contour of Lagrangian function (z axis) with varying Lagrangian multipliers (x/y axes) for sampling-based model.
each sample day. For simplicity, this research does not remove traffic data from weekend days and holidays.

Fig. 7. Measurement coverage of travel time for Bay Area, California. Links with measurements are highlighted. A sub-corridor on Bayshore Freeway from Mountain View to San Jose, California is enlarged.

5.2. Spatial Correlation Analysis

This study extracts and examines the data from a segment of the Bayshore Freeway between Mountain View and San Jose, California. On this 11-mile freeway corridor, 6 miles of links have travel time measurements, as enlarged in Fig. 7. The spatial link correlation is visualized in Fig. 8, where the planar xy coordinates represent link sequence numbers along the corridor and the vertical z axis shows the value of correlation. As a simple verification test, the reliable routing algorithm for the model with link correlation is carried out from the origin to the destination of this sub-corridor, leading to a relative gap of 3.5%.
In our experiments, in order to correctly model the randomness of link travel time for links without measurements, random travel time values from a Normal distribution are generated in order to incorporate their variance. It is important to recognize that, if only single mean travel time values are used for links without measurements, then the calculated path travel time reliability measure would assume zero variability for those links, leading to potentially biased solutions. In this research, the travel time index (calculated as the travel time divided by the free flow travel time of each link) is calculated first to capture the variance of the travel time index for links with measurements. The corresponding average variance of the travel time index is then applied to links without measurements to generate random travel times for each sampling day. It should be also noted that the randomly generated travel times in the above procedure are inherently independent and uncorrelated across different links, so the calculated path travel time reliability without complete data coverage is likely to be lower than the actual end-to-end travel time reliability measure experienced by commuters.

5.3. Numerical Performance and Solution Quality Analysis

As short-distance OD pairs might be covered by no or inadequate raw observations, and they typically have very limited alternative routes to examine, this study particularly imposes the following rules to select OD pairs to be tested: (1) the average path travel time is larger than 45 minutes, and (2) the measurement coverage on the least expected travel time path is larger than 30% in distance. As a result, an OD-pair set $S$ containing $s=246$ random OD pairs is generated from the Bay Area network.
The performance of our approach is assessed using the average relative gap, which is calculated as the average value of all 246 OD pairs under a predefined maximum number of iterations $K_{\text{max}}$ and reliability coefficient $\beta$ in Eq. (3), e.g. $\bar{\epsilon} = \frac{\sum_{s \in S} E_{\text{rel},\beta,K_{\text{max}}}}{s}$. To consider the impact of the reliability coefficients on the experiment results, this study considers two sets of reliability coefficients: $\beta = 1.27$ as suggested in the travel time reliability study by Noland et al. (1998), and $\beta = 4$ as a comparatively large value for testing purposes. The travel time reliability coefficient represents how much travel time reliability (in terms of standard deviation) the trip-maker is willing to trade for unit time saving.

In this paper we adopts a utility function of travel time reliability from the study by Noland et al. (1998). They performed a state-preference survey that used a sample set of more than 700 commuters in the Los Angeles region to empirically estimate the user preferred value of travel time reliability. Specifically, among the 543 valid questionnaires collected in their survey, each respondent was asked 9 stated preference questions each with two commute routes under different distributions of travel times and departure times. A value of 1.27 was calibrated in their study for the travel time reliability coefficient $\beta$. Additionally, we also test $\beta = 4$ in our study, by considering that some road users, such as commercial fleet companies, have larger values of travel time reliability, compared to regular commuters.

As shown in Figs. 9 and 10, the average gap decreases along with the increase of the predefined maximum number of iterations $K_{\text{max}}$. To characterize the statistics distribution of the optimality gap measure, the standard deviation of relative gaps under variant $K_{\text{max}}$ and $\beta$ are calculated in Fig. 11. Overall, the relative gap for the reliable routing algorithm without considering link travel time correlation is dramatically lower than the gap for the algorithm with correlation consideration, e.g. 1.7% vs. 5.4% when $\beta = 1.27$. It should be remarked that this difference does not indicate that the first algorithm is superior to the second one, as these two algorithms use different primal and dual objective functions (with vs. without spatial correlation). A more systematic comparison scheme should be the following: first, extract the route solutions (in terms of node sequence) from these two different algorithms, and then evaluate their solution quality in terms of the same travel time utility function with spatial correlation. Conceptually, possible solution quality loss occurs for those models that ignore the underlying correlations. This problem can be viewed as a special case of “price of correlation” in the field of stochastic optimization, as investigated recently by Agrawal et al. (2010).

From Figs. 9 and 10, we find that 20 maximum iterations will be sufficient for the model without correlation to achieve a solution with relatively small gap value, and the decreasing trends for the model considering spatial dependencies begin to slow down after 10 iterations. As mentioned previously, a small duality gap can still exist despite a considerably large number of iterations, mainly caused by the inherent limitation of Lagrangian lower bound estimation techniques. As expected, a larger travel time reliability coefficient, representing higher weight on the standard derivation, could result in larger relative gaps for both models.
Fig. 9. Relative gap in model WITHOUT link travel time correlation. Beta is the travel time reliability coefficient.
Fig. 10. Relative gap in model WITH link travel time correlation. Beta is the travel time reliability coefficient.

Fig. 11. Standard deviation of relative gaps for 246 OD pairs under two models and reliability coefficients.

Clearly, the most computational consuming step of the proposed algorithms is the shortest path calculation in each iteration of the searching process. As a result, the average computational complexity of the algorithms is determined by the number of shortest path calculations. For the model without correlations with a maximum number of 20 iterations (as the termination condition), the average computing time for finding the most reliable path for a single OD pair on our test network is 1.2 seconds (including 20 shortest path calculations). For the sampling-based model, about 0.7 seconds is needed with 10 shortest path calculations (to gain significant optimality improvements), and 1.4 seconds for at most 20 shortest path calculations.

6. Conclusions and Future Research
To meet the emerging needs for modeling travel time reliability, especially in the area of spatial network analysis, this paper proposes two models for the most reliable path problem, each with a different spatial dependency assumption. The path travel time reliability measure in this research is expressed through the standard deviation of path travel time. The computational challenges introduced by this reliability functional form stem from the nonlinearity and nonadditivity of standard deviation, as well as the concave characteristics of the corresponding square root transformation.

To tackle the above modeling and computational challenges, this study proposed two new approximation methods for solving the reliable path searching problem. First, focusing on the nonadditive and concave characteristics of the original reliability representation, a Lagrangian substitution-based lower bound approach is introduced to quantify the quality of solutions found by an iterative search process. More specifically, with efficient evaluation of feasible solutions and their dual problem results, a tight lower bound is achieved and a close-to-optimal solution can be obtained with a guaranteed level-of-service. Second, to incorporate the spatial correlation among link travel times, this paper constructs a sampling-based solution algorithm. Instead of relying on the (independent or limited correlated) probability density functions of link travel times, a set of individual historical measurements are utilized to explicitly capture the inherent spatial correlation. Comprehensive experiment results on a large-scale network show that 10-20 iterations of standard shortest path algorithms for the reformulated models can offer a very small duality gap of about 2-6%.

Future research interests will cover four major extensions. (1) Expand the realm of application of these models from static travel times to time-varying travel times. (2) Jointly consider temporal and spatial correlation in finding the most reliable path. (3) Extend current reliable routing algorithms from single OD case to the one-origin to all-destination application. Such one-to-all most reliable path problem may serve an important role in the dynamic traffic assignment. (4) Apply distributed computing techniques such as cloud computing to improve the computational efficiency.
Acknowledgement

This paper is based on research work partially supported through the Safe Trip 21 project, sponsored by the Research and Innovative Technology Administration (RITA), US Department of Transportation. We would like to thank the California Partners for Advanced Transit and Highways (PATH) in the University of California, Berkeley and NAVTEQ Corporate for providing traffic measurement and map data of the Bay Area, CA. We also want to thank our colleague Mr. Jeffrey Taylor in the University of Utah for his careful review and comments. The work presented in this paper remains the sole responsibility of the authors.
Reference


Scott, K., Bernstein, D., 1997. Solving a best path problem when the value of time function is nonlinear. preprint 980976 of the *Transportation Research Board*.


