Risk-Averse Second-Best Toll Pricing

Xuegang (Jeff) Ban, Department of Civil and Environmental Engineering,
Rensselaer Polytechnic Institute, banx@rpi.edu, Corresponding Author

Shu Lu, Department of Statistics and Operations Research,
University of North Carolina at Chapel Hill, shulu@email.unc.edu

Michael Ferris, Computer Sciences Department,
University of Wisconsin at Madison, ferris@cs.wisc.edu,

Henry Liu, Department of Civil Engineering,
University of Minnesota, Twin Cities, henryliu@umn.edu

Abstract

Existing second best toll pricing (SBTP) models determine an optimal toll for a given set of links in a transportation network by minimizing certain system objective, while the traffic flow pattern is assumed to follow user equilibrium (UE). In this paper, we show that such a toll design approach is risk-prone, which tries to optimize for the best-case scenario, if the UE problem has multiple solutions. Accordingly, we propose a risk-averse SBTP approach that aims to optimize for the worst-case scenario. By applying the robust optimization concept, we formulate the risk-averse SBTP as a "min-max" problem. We establish a general solution existence condition for the risk-averse model, whereas we discuss in detail that such a condition may not be always satisfied in reality. In case a solution does not exist, it is possible to replace the exact UE solution set by a set of approximate solutions. This replacement guarantees the solution existence of the risk-averse model. We then develop a scheme such that the solution set of an affine UE can be explicitly expressed. Using this explicit representation, an improved simplex method can be adopted to solve the risk-averse SBTP with affine UEs. We illustrate the risk-prone and risk-averse modeling approaches, the explicit expression of affine UEs, and the solution algorithm using a small example.

1 Introduction

The Second-Best Toll Pricing (SBTP) problem aims to determine an optimal toll for a given set of links in a transportation network so that traffic can be distributed more efficiently from the system point of view. SBTP can be further categorized as static SBTP or dynamic SBTP (DSBTP), depending on whether traffic dynamics are considered or not. In this paper, we focus on the static SBTP, which is referred to as SBTP hereafter in this paper.

The literature on SBTP is rich and still growing. Many researchers have modeled SBTP as a bilevel problem or an MPEC (mathematical program with equilibrium constraints) [1, 2, 3, 4]. The upper level is to optimize a certain objective function from the transportation system point of view and the lower level is a user equilibrium (UE) problem to account for the route choice behavior of individual motorists. Most existing SBTP methods try to find "optimal" tolls (i.e., the so-called "upper level" decision variables) for the selected set of links so that the upper level objective can be minimized or maximized. Meanwhile, by solving the bilevel model, the lower level UE solution associated with the optimal tolls can also be obtained. Hence the bilevel model implicitly assumes that by applying the obtained toll pricing scheme, the resulting UE flow pattern is exactly what is predicted by the model, and thus the desired system objective can be achieved. Here we make distinction between the "predicted" UE solution and the "realized" UE solution. The former refers to the UE solution obtained by solving the bilevel SBTP model, whereas the latter refers to the UE flow pattern that one obtains after imposing the optimal toll (e.g., via solving the bilevel model). If UE has a unique solution, the predicted and realized solutions are exactly the same. However, when UE has multiple solutions, each solution may correspond to different motorist behaviors. Further, once the toll is imposed, there is no other way in the context of toll pricing that one can "enforce" how drivers make their choice decisions. Therefore, it is possible that the "realized" UE flow pattern may deviate from what is predicted. If this is the case, the desired "optimal" objective may not be achieved and the designed toll scheme may not be effective.

The non-uniqueness of UE solutions therefore represents uncertainty in the SBTP design, which is not fully recognized in the SBTP literature. In this paper, we first show that existing SBTP modeling methods are "risk prone" if the UE solution is not unique. This is because the toll from these methods is designed in such a way that one "hopes" the realized UE solution is exactly the predicted UE solution (i.e. the "best case"). However, since one has no control which UE solution will be realized given a toll setting, a more appropriate toll pricing scheme should be able to account for this uncertainty.

To address the non-uniqueness of UE solutions, we adopt the robust optimization concept [5] in this paper. From the robust design perspective, the implemented toll should be optimal for the "worst case" scenario while the UE solution varies. Here the "worst case" for a given toll refers to the largest upper level objective value as UE solution varies under the toll. This corresponds to a "risk-averse" design approach, which can be further formulated as a "min-max" problem. Here "risk" refers to whether the toll designer's objective can be achieved after the toll is imposed, or more specifically, whether the objective is "better off" or "worse off". Clearly, for risk-prone SBTP, one aims to optimize for the "best-case" scenarios; as UE solution varies, the designer's objective will worse off. On the other hand, since the risk-averse approach aims to optimize for the "worst case", the design objective will always be better off as UE solution changes. As we result, we capture in this paper the risk-taking behaviors of the toll designer instead of individual motorists.

It has been recognized for long that SBTP is a special case of Stackelberg games that consider a two-player game: one of the players is the leader and the other is the follower. The upper level is to find an optimal strategy for the leader, and the lower level considers equilibrium responses of the follower. In terms of SBTP, the toll designer is the leader and individual drivers are the followers. The possible non-uniqueness of the followers' responses (i.e. the non-uniqueness of UE solution in our case) has been recognized in the game theory literature, leading to the weak Stackelberg games [6, 7, 8]. In particular, the weak Stackelberg game takes an *inf-sup* formulation, similar to the *min-max* model of the risk-averse SBTP design approach proposed in this paper. The lower level solution set for a weak Stackelberg game is defined by a lower level nonlinear programming problem (NLP) in [7, 8]. In our risk-averse SBTP approach, however, the solution set is defined by a variational inequality (VI). Moreover, previous studies on weak Stackelberg games have not recognized their connections to risk-taking behaviors of the leader (i.e. the toll designer in SBTP).

A general solution existence condition exists for the proposed *min-max* model of risk-averse SBTP. In case this condition does not hold, we can replace the original lower level solution set by a set of approximate solutions to the lower level problem. By doing so, we prove that the upper level problem has at least one solution under mild conditions. Moreover, the optimal objective values of such problems converge to that of the original problem as the approximation error goes to zero. These findings extend the theoretical results in [7, 8] on the solution existence of weak Stackelberg games (whose lower level problems are NLPs) to risk-averse SBTP (whose lower level problems are VIs).

The risk-prone toll pricing scheme has been extensively studied in the literature and many solution methods have been developed [1, 3]. The solution algorithm for weak stackelberg game however were not fully explored. Although some theoretical results are given for solution existence conditions, no solution algorithm is given [6, 7, 8]. To solve the proposed risk-averse SBTP model, we show that it is necessary to explicitly explore the characterization of the lower level UE solution set. In this paper, we start with affine UEs, in which the link travel time is a linear function of link flow. We show, through a link-node based nonlinear complementarity UE model, that the solution set of an affine UE can be represented as a polyhedral set. Based on this finding, we find that the objective value of the *min-max* formulation can be easily evaluated. This motivates us to adopt the fortified-descent simplex method developed in [9] for non-constrained optimization to solve the risk-averse SBTP. We then test the algorithm using a small example.

This paper is organized as follows. Section 2.1 introduces the VI-based UE model and shows that if the link travel time is monotone, the UE has a nonempty, compact and convex solution set. This is true for both tolled and un-tolled cases. Section 2.2 discusses the existing SBTP model that can be formulated as an MPEC. We find that such modeling approach is risk-prone when UE solution is not unique. The risk-averse model is developed in Section 3, which is formulated as a *min-max* problem. We prove the existence of the optimal objective value for this problem, and discuss conditions under which an optimal solution exists. We also show in this section the connection of the risk-averse SBTP with the weak stackelberg game. We then extend some of the theoretical results in studying weak stackelberg game to risk-averse SBTP. An illustrative example is provided in Section 4. We show that at least for this small example, the solution existence conditions of the risk-averse approach hold, and furthermore the risk-averse approach is superior to the risk-prone approach. In Section 5, we introduce a link-node based nonlinear complementarity model for UE. We show that the solution set of the UE, provided any of its solution is known, can be expressed as a polyhedral set. An iterative solution algorithm based on the fortified-descent simplex method [9] is proposed in Section 6. The solution algorithm is then tested on the small example. We conclude the paper in Section 7.

2 Risk-Prone SBTP

Most existing SBTP models aim to achieve an optimal toll by minimizing certain design objective, subject to the requirement that the resulting traffic flow pattern must follow UE. We show in this section that such approach is "risk prone" if UE solution is not unique. We start with some well-known results of VI-based UE models.

2.1 VI-Based UE Model

Assume that a traffic network can be represented as a directed graph $G(N, A)$ where N is the set of nodes and A is the set of links. In this paper, we use index i or j to denote a node, and a to denote a link. Let x_a be the total traffic flow on link a, and let $x = (x_a)_{a \in A}$ be the vector of link flows. Further $t_a(x)$ is the travel time of link a, which is a function of the total link flow vector x, and $t = (t_a)_{\forall a \in A}$. The traffic user equilibrium (UE) model, denoted as *UE(0)*, can be formulated as follows [10, 11]:

$$\textit{UE(0)} \qquad t(x)^T(x' - x) \geq 0, \forall x' \in K. \tag{1}$$

Here K denotes the feasible set of link flows, which is nonempty, compact and convex, and is in fact a polyhedral set in most applications. We use $UE(0)$ to indicate the model in which no toll is imposed. Note that here we focus on UEs with fixed demands only, so (1) applies. It is well-known that if t is strictly monotone on K, $UE(0)$ has a unique solution [12]. If t is only monotone (or pseudo-monotone [13]), the solutions of $UE(0)$ may not be unique. Let the solution set of $UE(0)$ be denoted by $S(0)$. An application of Facchinei and Pang [13, Corollary 2.2.5 and Theorem 2.3.5] shows that $S(0)$ is a nonempty, compact and convex set. We state this as the following lemma without proof.

Lemma 1 *If t is continuous and monotone, and K is nonempty, compact and convex, then $S(0)$ is nonempty, compact and convex.* □

Assume that tolls, denoted as y, are imposed on the network. Here, y is a vector whose dimension is the number of links, with its elements y_a denoting the toll imposed on link a. In many situations, tolls are only imposed on a subset, say P, of the link set A. Then for each link a that does not belong to P, we fix y_a as 0. Even for links that belong to P, the toll y_a usually has to lie in a reasonable range. Consequently, the toll y has to satisfy a bound constraint defined as $K_y = \{y | y_l \leq y \leq y_u\}$, where y_l and y_u are the lower and upper bounds of y respectively. If we introduce θ as the "value of time" parameter, we may use

$$c(x, y) = t(x) + y/\theta \tag{2}$$

to denote the generalized link travel time, which is a combination of the link travel time and toll. We then consider the following UE problem under the influence of y, denoted by $UE(y)$:

$$\textit{UE(y)} \qquad (t(x) + y/\theta)^T(x' - x) \geq 0, \forall x' \in K. \tag{3}$$

Let $S(y)$ denote the solution set of $UE(y)$. We have the following Lemma, which is a straightforward extension of Lemma 1.

Lemma 2 *If t is continuous and monotone, and K is nonempty, compact and convex, then $S(y)$ is nonempty, compact and convex for any given toll y.* □

2.2 Risk Prone SBTP Model

Assume that $f(x, y)$ is the objective function to determine an optimal toll, which may be the total system travel time or similar objectives the toll designer may have. Most existing SBTP models aim to find the optimal toll by solving the following MPEC, denoted as $MPECSBTP$:

$$MPECSBTP \qquad \min_{y,x} \quad f(y, x) \tag{4}$$

$$\text{s.t.} \quad y \in K_y \tag{5}$$

$$x \quad \text{solves UE}(y). \tag{6}$$

Since we use $S(y)$ to denote the solution set of $UE(y)$, we may rewrite the constraint that x solves $UE(y)$ as $x \in S(y)$. Under the hypothesis that t is monotone (not necessarily strictly monotone) with respect to x, $UE(y)$ may have multiple solutions. Hence, $S(y)$ is a set-valued map of the toll vector y; see Appendix A of this paper for the definition of set-valued maps. If we let

$$G = \{(y, x) \mid x \in S(y), \quad y \in K_y\}$$

be the graph of the set-valued map S, we can rewrite the $MPECSBTP$ model into the following single level problem:

$$RPSBTP \qquad \min_{y,x} \quad f(y, x) \tag{7}$$

$$\text{s.t.} \quad (y, x) \in G, \tag{8}$$

where we use the label $RPSBTP$ to stand for "risk-prone second-best toll pricing" for reasons we will see later in this subsection.

The following theorem provides mild conditions under which the $RPSBTP$ has at least a solution.

Theorem 1 *If t is continuous and monotone with respect to x, $f(y, x)$ is continuous with respect to (y, x), and K is nonempty, compact and convex, then the following two statements hold.*

(a) G is compact; and

(b) RPSBTP has at least one solution.

Proof. See Appendix B. □

We note that in the literature, similar conditions have been established for the solution existence of SBTP by many researchers (e.g. [2]). Theorem 1 provides an alternative way for such a proof. The above theorem guarantees that the $MPECSBTP$, or $RPSBTP$, has a solution. However, we need to ask the question whether its solution satisfies the initial objective after imposing a toll.

To answer this question, let (y^*, x^*) be the solution of $RPSBTP$. We call y^* the optimal toll scheme, and x^* the *predicted* UE flow under y^*. Suppose that we impose the optimal toll scheme y^* on the network and denote $\bar{x}$ the resulting UE solution called *realized* flow. If the solution of $UE(y^*)$ is unique, we must have $x^* = \bar{x}$, i.e. the network user equilibrium flow will be x^* under y^*. In this case, the optimal objective value $f(y^*, x^*)$ is achieved.

However, if t is monotone (not necessarily strictly monotone), there may be multiple UE solutions given the toll scheme y^*. This is illustrated in Figure 1. In the figure, y^* is the obtained optimal toll by $RPSBTP$ and the rectangle represents the UE solution set under y^* (i.e. $S(y^*)$). The objective function is assumed to be a convex curve for given y^*. Since $S(y^*)$ is not a singleton, it is possible that if y^* is imposed, the realized UE pattern could be $\bar{x}$ instead of x^*. The objective value $f(y^*, \bar{x})$, however, may be much larger than $f(y^*, x^*)$, implying that the obtained optimal toll y^* may not be the most desirable.

The above discussion shows that traditional SBTP approaches are risk-prone when the UE solution is not unique since it hopes for the best scenario to happen, that is the predicted UE solution is realized when a toll is imposed. The fact that $RPSBTP$ optimizes for the "best case" can be made clear if we rewrite (7) - (8) as $\min_{y \in K_y} \; \min_{x \in S(y)} \; f(y, x)$. Clearly, the inner minimization problem $\eta(y) \equiv \min_{x \in S(y)} \; f(y, x)$ represents the "best case" scenario, i.e. the smallest upper level objective value, for a given toll y. The $RPSBTP$ design approach then aims to find a toll scheme that minimizes $\eta(y)$ over K_y (i.e. the "best case").

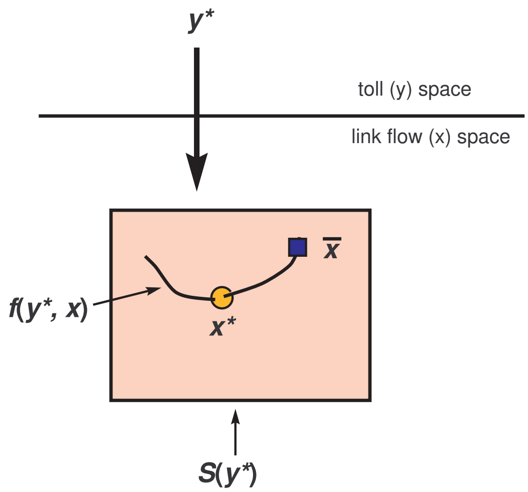

Figure 1: Illustration of Risk-Prone SBTP

3 Risk Averse SBTP

3.1 Model Formulation and the Boundedness

As aforementioned, if the UE solution is not unique, it will be uncertain which UE flow pattern (solution) will be realized once a toll is imposed. The UE solution set therefore represents uncertainty in SBTP design. One way to account for this uncertainty is to adopt the robust optimization concept to design tolls so that they optimal for the worst case scenario. This represents a "risk-averse" approach for toll pricing, which can be expressed by a *min-max* problem (denoted by $RASBTP$) as follows:

$$RASBTP \qquad \min_{y \in K_y} \quad \max_{x \in S(y)} \quad f(y, x) \tag{9}$$

We can see that $RASBTP$ aims to minimize, as y varies within K_y, the largest objective value over all x's in $S(y)$. Assume y^* the computed optimal toll by the risk-averse approach and $x^* \in S(y^*)$ its associated UE solution. We will have $f(y^*, x^*) \leq \max_{\forall x \in S(y)} f(y, x), \forall y \in K_y$. In other words, the risk-averse approach generates a solution that is optimal for the worst-case scenario. Further, at y^*, we have $f(y^*, x^*) \geq f(y^*, x), \forall x \in S(y^*)$. This means that at the optimal toll y^*, $f(y^*, x^*)$ represents the "worst" possible objective value. If x varies in $S(y^*)$, the objective value will never increase (i.e. always "better off").

We note that there are other ways to account for the uncertainty due to the UE solution non-uniqueness besides optimizing for the worst case. For example, one may want to achieve the optimal expected value of the objective as the UE solution changes within $S(y^*)$. This however requires knowledge of the distribution of the possible realization of the UE solution. In [14], we assume that such realization follows a uniform distribution and further show that such a design approach is actually *risk-neutral* by optimizing for the expected objective value.

The following theorem shows that the optimal objective value of $RASBTP$ exists; in other words, the function

$$\Phi(y) \equiv \max_{x \in S(y)} \quad f(y, x) \tag{10}$$

has a greatest lower bound as y varies within K_y.

Theorem 2 *If t is continuous and monotone with respect to x, $f(y, x)$ is continous with respect to (y, x), and K is nonempty, compact and convex, then*

$$\min_{y \in K_y} \Phi(y) > -\infty. \tag{11}$$

Proof. Note that

$$\min_{y \in K_y} \max_{x \in S(y)} f(y,x) \geq \min_{y \in K_y} \min_{x \in S(y)} f(y,x) \tag{12}$$

where the right hand side is exactly the optimal objective value of the *RPSBTP* problem studied in last section, which proven to be finite by Theorem 1. This completes the proof. □

3.2 Solution Existence Conditions for RASBTP

The next question to ask is whether the optimal objective value of the *RASBTP* is attained by some $y \in K_y$. A general condition for this is provided in the following theorem. It uses a property called lower semicontinuity; see its definition in Appendix A. The theorem is a standard result, see, e.g., Theorem 1.9 in [15]; we include its proof here for the sake of completeness.

Theorem 3 *Suppose that the set K_y is compact, and the function $\Phi(y)$ defined in* (10) *is lower semicontinuous at each point y in K_y, then the RASBTP has at least one solution.*

Proof. Let $\{y_n\}$ be a sequence in K_y such that the function values of (10) at y_n converge to the optimal objective value of the *RASBTP*. Such a sequence exists due to Theorem 2. By the compactness of K_y, the sequence $\{y_n\}$ has a limit point $y_0 \in K_y$. By the lower semicontinuity assumption, y_0 is an optimal solution for the *RASBTP*. □

The question on solution existence now reduces to whether the function $\Phi(y)$ in (10) is lower semicontinuous. The answer is not affirmative. See Appendix C for a simple example where this fails to hold. As a result, the *RASBTP* in that example does not attain its optimal objective value in K_y. That example reflects the key feature that prevents a general *RASBTP* from attaining its optimal value: the lack of global continuity of $S(y)$ as a set-valued map of y. We know that the set of feasible link flows K is usually a polyhedral convex set. Consequently, the map S is a polyhedral set-valued map of y, see, e.g., [16, Proposition 2.4]. By a polyhedral set-valued map, we mean a set-valued map whose graph is the union of finitely many polyhedral convex sets. It follows that S has the *outer Lipschitz continuity* property (also known as *upper Lipschitz continuity* or *calmness*, see [17]) that, there exists a positive scalar λ such that each $y \in K_y$ has a neighborhood V in K_y with

$$S(y') \subset S(y) + \lambda \|y' - y\| B$$

for each $y' \in V$, where B is the unit ball; see [17] for a proof of this property for polyhedral set-valued maps. The outer Lipschitz continuity prevents $S(y')$ from expanding at a rate faster than linear as y' goes away from y. However, there are no limits on the contraction rate, as is seen from the example above. Such discontinuity of $S(y)$ causes the lack of lower semicontinuity of the function (10).

This raises another question: can we obtain the continuity of S by strengthening the assumption on the travel time function? The monotonicity assumption requires the Jacobian matrix for an affine travel time function be positive semidefinite. It is known that, if the Jacobian matrix is positive definite, then the solution set $S(y)$ is a singleton, and is Lipschitz continuous with respect to y. Hence, if we would strengthen our assumption by requiring the Jacobian matrix be positive definite, we could guarantee the solution existence of the *RASBTP*. However, in this case $S(y)$ is also a singleton for each y, implying that *RPSBTP* and *RASBTP* are the same.

Is there a condition that lies between positive definiteness and positive semidefiniteness, under which $S(y)$ will be a nontrivial set depending Lipschitz continuously on y? For linear complementarity problems (LCP), if the matrix is positive semidefinite, then the solution set is Lipschitz continuous if and only if the matrix is a P matrix [18]., which then implies that the solution set for the LCP is a singleton. Here a P matrix is a matrix with all principal minors being positive [13]. For variational inequalities, there is not a similar result to the extent we know. Clearly, further investigations are needed to find out whether such condition exists. [1]

Given the above analysis, one would ask how to deal with the case in which the *RASBTP* does not attain its optimal objective value. We first notice that for the purpose of robust design, it is enough to find a toll scheme y so that the function value $\Phi(y)$ is close to the infimum. Our present methodology adopts this approach. Another

[1] The discussion so far shows that one cannot guarantee that the *RASBTP* attains its optimal objective value in general situations. Due to this reason, it might be mathematically appropriate to replace the "min" notation in the *RASBTP* formulation by the "inf" notation, to emphasize the difference between these two notations. Hereafter in this paper, however, we will still use the "min" notation.

possible approach is to enlarge (slightly) the set $S(y)$ by considering approximate solutions of the lower level UE problem. The next subsection discusses the latter approach in more detail. Also notice that due to the complicated nature of this problem, there is no guarantee that the function $\Phi(y) = \max_{x \in S(y)} f(y,x)$ be convex. In fact, Φ is not necessary convex even for the simplest case in which $S(y)$ is a singleton for each y. Hence the uniqueness of $RASBTP$ cannot be established.

3.3 Regularized Lower Level UE

It has been recognized for long that SBTP is a special case of Stackelberg games that consider a two-player game: one of the players is the leader and the other is the follower. The upper level is to find an optimal strategy for the leader, and it takes the same formulation as (9), if we would, for the moment, let K_y represent the set of strategies of the leader, f be the cost function for the leader, and $S(y)$ be the set of optimal responses for the follower when the leader's strategy is y. In terms of SBTP, the toll designer is the leader and individual drivers are the follower.

In the literature, strong and weak Stackelberg games have also been proposed and studied respectively [6, 7, 8] to the case that $S(y)$ is a singleton or not. In particular, the weak Stackelberg game takes a *inf-sup* formulation, conceptually the same to the *min-max* model of $RASBTP$. In this sense, the $RPSBTP$ and $RASBTP$ models proposed in this paper are specific applications of general strong and weak Stackelberg games in toll design. However, the lower level solution set $S(y)$ for the weak Stackelberg game is defined by a lower level optimization problem in [7, 8] whose objective function is the cost function for the follower. This is different from the risk-averse SBTP, in which $S(y)$ is defined by a VI.

In [7, 8], the authors pointed out that the upper level problem of weak Stackelberg game (which is similar to $RASBTP$ in this paper) may have no solutions even for nice cost functions. Consequently, they proposed to replace the set $S(y)$ by the set of approximate solutions to the lower level problem. By doing so, the upper level problem has solution existence under mild conditions. Moreover, the optimal objective values of such problems converge to that of the original problem as the approximation error goes to zero. We show next how we extend these results to $RASBTP$ whose lower level is a VI instead of an NLP.

We first define a *regularized lower level equilibrium (RLLE)* problem

$$S^\epsilon(y) = \{x \in K : (t(x) + y/\theta)^T (x' - y) > -\epsilon, \forall x' \in K\}, \tag{13}$$

for some $\epsilon > 0$. We then replace the set $S(y)$ by $S^\epsilon(y)$ in (9) to obtain the following problem

$$\varphi(\epsilon) := \min_{y \in K_y} \max_{x \in S^\epsilon(y)} f(y,x). \tag{14}$$

The following result states that S^ϵ is lower semicontinuous as a set-valued map of y.

Lemma 3 *Let ϵ be fixed, and suppose that the set K is compact and the function t is continuous. Then the set-valued map $S^\epsilon(y)$ is lower semicontinuous at each $\bar{y} \in K_y$.*

Proof. Let $\bar{x} \in S^\epsilon(\bar{y})$. The compactness of K implies the existence of $\eta > 0$ such that

$$(t(\bar{x}) + \bar{y}/\theta)^T (x' - \bar{y}) \geq \eta - \epsilon, \forall x' \in K.$$

Since K is bounded, there exist neighborhoods V of $\bar{x}$ and W of $\bar{y}$ such that

$$(t(x) + y/\theta)^T (x' - y) > -\epsilon$$

holds for each $x' \in K$, $x \in V$ and $y \in W$. Hence, we have $x \in S^\epsilon(y)$ for each $x \in V$ and $y \in W$. This completes the proof. □

Consequently, we have

Theorem 4 *Let ϵ be fixed, and suppose that sets K and K_y are compact, and functions t and f are continuous. The problem* (14) *has at least one solution.*

Proof. Lemma 3 implies that the function

$$\max_{x \in S^{\epsilon}(y)} f(y, x)$$

is lower semicontinuous with respect to y; see, e.g., [7, Theorem 6.1(i)]. The conclusion then follows from an argument similar to the proof of Theorem 3. □

The next result provides a relation between set-valued maps S^{ϵ} and S.

Lemma 4 *Suppose that the function t is continuous. Let $y \in K_y$, $\epsilon_n \to 0^+$, $y_n \to y$, and $x_n \in S^{\epsilon_n}(y_n)$. Let x be a limit point of $\{x_n\}$. Then $x \in S(y)$.*

Proof. The fact $x_n \in S^{\epsilon_n}(y_n)$ implies that

$$(t(x_n) + y_n/\theta)^T (x' - y_n) > -\epsilon_n$$

for each $x' \in K$. Taking limits on both sides, we have

$$(t(x) + y/\theta)^T (x' - y) \geq 0$$

for each $x' \in K$. Thus, $x \in S(y)$. □

Here $\epsilon_n \to 0^+$ means ϵ_n goes to zero from above. The sequence $\{x_n\}$ always has a limit point because it lies in the compact set K. Next we use the latter lemma to prove that $v(\epsilon)$ as defined in (14) converges to the optimal objective value of the $RASBTP$ as ϵ converges to 0.

Theorem 5 *Suppose that sets K and K_y are compact, and functions t and f are continuous. Then*

$$\lim_{\epsilon \to 0^+} \varphi(\epsilon) = \min_{y \in K_y} \max_{x \in S(y)} f(x, y).$$

Proof. For each $\epsilon > 0$ and each $y \in K_y$, we have

$$S(y) \subset S^{\epsilon}(y),$$

so

$$\min_{y \in K_y} \max_{x \in S(y)} f(x, y) \leq \min_{y \in K_y} \max_{x \in S^{\epsilon}(y)} f(x, y) = \varphi(\epsilon).$$

This proves

$$\liminf_{\epsilon \to 0^+} \varphi(\epsilon) \geq \min_{y \in K_y} \max_{x \in S(y)} f(x, y).$$

It remains to prove (see Appendix A for the definition of lim sup)

$$\limsup_{\epsilon \to 0^+} \varphi(\epsilon) \leq \min_{y \in K_y} \max_{x \in S(y)} f(x, y), \tag{15}$$

and to prove this we need to prove

$$\limsup_{\epsilon \to 0^+} \varphi(\epsilon) \leq \max_{x \in S(y)} f(x, y) \tag{16}$$

for each $y \in K_y$. Choose an arbitrary y from K_y, and let $\epsilon_n \to 0^+$ be a sequence such that

$$\lim_{n \to \infty} \varphi(\epsilon_n) = \limsup_{\epsilon \to 0^+} \varphi(\epsilon).$$

Choose x_n in $S^{\epsilon_n}(y)$ such that

$$f(x_n, y) \geq \max_{x \in S^{\epsilon_n}(y)} f(x, y) - 1/n \quad \geq \quad \varphi(\epsilon_n) - 1/n.$$

Let x be a limit point of $\{x_n\}$. Passing to a subsequence if necessary, take limits on both sides of the inequality above. We have

$$f(x, y) \geq \lim_{n \to \infty} \varphi(\epsilon_n) = \limsup_{\epsilon \to 0^+} \varphi(\epsilon)$$

where the latter equality holds by the choice of ϵ_n. We have $x \in S(y)$ by Lemma 4, so it follows

$$\limsup_{\epsilon \to 0^+} \varphi(\epsilon) \leq \max_{x \in S(y)} f(x, y).$$

This proves (16), and (15) follows. □

In summary, by replacing $S(y)$ by $S^{\epsilon}(y)$ in the $RASBTP$, we obtain a class of problems in the form of (14). The solution existence of such problems is easy to guarantee, according to Theorem 4. Moreover, as ϵ converges to 0, the optimal objective value of (14) converges to that of $RASBTP$. This not only provides a possible scheme to solve the $RASBTP$, but also shows that its optimal objective value is stable with respect to the approximation in computing the lower level equilibrium problem $RLLE$. Notice that solving the $RLLE$ problem (13) itself is not trivial. This is because 1) no existing solution technique can be directly applied for solving (13) as it is neither a VI nor NLP, and 2) we need to obtain an expression of the solution set $S^{\epsilon}(y)$ rather than a single solution.

4 An Illustrative Example

To better illustrate the risk-prone and risk-averse approaches, we provide a small example in this section.

Figure 2(a) depicts a hypothetical network with one origin-destination (OD) pair (from node r to node s) and three routes. A toll booth is located at the very beginning of route 2 and 3. The distance between node r and i is very small so that the travel time can be ignored (assume that toll is automatically collected and therefore the delay at the toll booth can be ignored as well). Further assume that the total demand $d = 10$ and the route (also link) flow are x_1, x_2, and x_3. The travel times of the links are assumed to have the following form:

$$\begin{aligned} t_1 &= 2x_1 + x_2 + x_3 \\ t_2 &= 2x_2 + 2x_3 \\ t_3 &= 2x_2 + 2x_3. \end{aligned}$$

In other words, link interactions do exist among the three links. For simplicity, we assume that the "value of time" $\theta = 1$. Then the link generalized travel times, with toll imposed, are:

$$\begin{aligned} c_1 &= t_1 \\ c_2 &= t_2 + y \\ c_3 &= t_3 + y. \end{aligned}$$

Here y is the toll and $y \in K_y = \{y | 0 \leq y \leq 15\}$. Denote $c = (c_1, c_2, c_3)^T$ and $x = (x_1, x_2, x_3)^T$. It is easy to observe that c (or t) is monotone, but not strictly monotone, with respect to x. To see this, we note that the Jacobian matrix of c over x is

$$J = \partial c / \partial x = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{bmatrix}.$$

Clearly, J is not symmetric and we have

$$(J + J^T)/2 = \begin{bmatrix} 2 & 0.5 & 0.5 \\ 0.5 & 2 & 2 \\ 0.5 & 2 & 2 \end{bmatrix},$$

which is symmetric and positive semidefinite, but not positive definite. Therefore, c is monotone with respect to x, but not strictly monotone.

We first look at the solution set of $UE(y)$, i.e., $S(y)$ for any given $y \in K_y$. Since we have $c_2 = c_3$, there are three cases that we need to consider: i) only route 1 carries flow, ii) only routes 2 and 3 carry flow, and iii) all three routes carry flow. For case i), we have $x_1 = 10, x_2 = x_3 = 0$. This leads to $c_1 = 20 > c_2 = c_3 = y \leq 15$. Therefore, case i) is impossible. For case ii), we have $x_1 = 0, x_2 + x_3 = 10$. This leads to $c_1 = 10 < c_2 = c_3 = 20 + y \geq 20$, which is also impossible. Therefore, all three routes must carry flow and we have $c_1 = c_2 = c_3$. This gives us

$$S(y) = \{x = (x_1, x_2, x_3)^T \geq 0 | x_1 = (10 + y)/3, \quad x_2 + x_3 = (20 - y)/3\}. \tag{17}$$

Clearly, for any given $y \in K_y$, $S(y)$ is a straight line (i.e., a nonempty polyhedral set) in the three dimension space $x_1 - x_2 - x_3$ as shown in Figure 2(b).

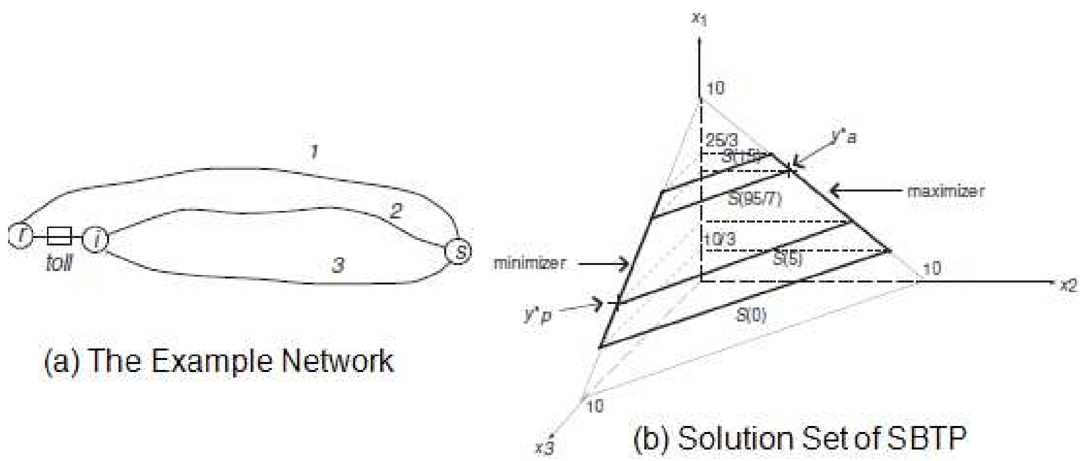

Figure 2: An Illustrative Small Example

To determine the "optimal" toll, we first assume that the objective function for the upper level as follows:

$$f(y, x) = t_1 x_1 + 3 t_2 x_2 + t_3 x_3. \tag{18}$$

In the above definition, we assign different weights to different links (routes). In particular, the weight of link 2 is set as 3. This may be appropriate if route 2 goes through an area which is more adversely impacted by traffic (in terms of vehicle-miles-traveled) than other areas.

Given the above, the risk-prone approach (i.e., the current SBTP practice) is to solve *RPSBTP*. First, since $S(y)$ can be explicitly expressed in equation (17) for a given y, the upper level objective function $f(y, x)$ can be rewritten as:

$$f(y, x) = \frac{y^2 - 10y + 400}{3} + \frac{4(20 - y)}{3} x_2. \tag{19}$$

Obviously, for a fixed $y \in K_y$, $f(y, x)$ is minimized when $x_2 = 0$ (notice $(20 - y) > 0$ always holds). Actually, $x_1 = (10 + y)/3, x_2 = 0, x_3 = (20 - y)/3$ is the unique and global minimizer of $f(y, x)$ when y is given since $f(y, x)$ is a linear function of x_2 for fixed y. This minimizer is the intersecting point of $S(y)$ and the $x_1 - x_3$ plane. Therefore, as y varies from 0 to 15, the trajectory of minimizers of $f(y, x)$ is the line on the $x_1 - x_3$ plane, as shown in Figure 2(b).

To find the solution to $RPSBTP$, therefore, we need to solve the following problem:

$$\min_{y \in K_y} \quad \eta(y) = \frac{y^2 - 10y + 400}{3}. \tag{20}$$

This is a convex quadratic programming problem and we have $y_p^* = 5$ as the (global) optimal solution. Here the subscript "p" denotes "risk-prone". The predicted UE solution is $x_{p,1}^* = 5, x_{p,2}^* = 0, x_{p,3}^* = 5$ and the associated objective value is $\eta(y_p^*) = 125$.

Most existing SBTP design methods will stop here with the above solution, which simply states that a toll $y = 5$ should be implemented. However, as the UE solution at the computed "best" toll $y_p^* = 5$ is not unique, the realized UE solution can be any point in $S(5)$, which is the line as shown in Figure 2(b). If the realized UE solution is on the $x_1 - x_2$ plane ($x_1 = 5, x_2 = 5, x_3 = 0$), the objective value will be much higher as $\bar{\eta}(y_p^*) = 225$. This illustrates that the risk-prone approach is not reliable when the UE solution is not unique.

For the "risk-averse" approach, we first find the maximizer of $f(y,x)$ for a given y, i.e. the expression of $\Phi(y)$. This is equivalent to maximize $f(y,x)$ in equation (19) over set $S(y)$. Clearly, this is achieved when x_2 is maximized at $x_2 = (20-y)/3$ since again $f(y,x)$ is a linear function of x_2. Thus the unique and global maximizer of $f(y,x)$ for a given y is $x_1 = (10+y)/3, x_2 = (20-y)/3, x_3 = 0$, which is at the $x_1 - x_2$ plane.

Next, substitute $x_2 = (20-y)/3$ to equation (19), we obtain the objective function $\Phi(y)$ for for the risk averse case as:

$$\Phi(y) = \frac{7y^2 - 190y + 2800}{9}. \tag{21}$$

Clearly $\Phi(y)$ in this case is continuous with respect to y and thus $RASBTP$ has at least one solution according to Theorem 6. To find the optimal solution, one needs to minimize (21) over K_y. This can be easily solved with a unique and global solution $y_a^* = 95/7 = 13.57$. Here the subscript "a" denotes "risk-averse". The predicted UE solution is $x_{a,1}^* = 7.86, x_{a,2}^* = 2.14, x_{a,3}^* = 0$ and the associated objective value is $z_a^* = 167.86$. Note that this value is less than that of the worst case scenario by the risk-prone toll scheme $y_p^* = 5$ (i.e., $\bar{\eta}(y_p^*)$).

Distinct from the risk-prone approach, the upper level objective value will decrease as the realized UE solution varies under the risk-averse optimal toll $y_a^* = 13.57$. In particular, if the UE solution on the $x_1 - x_3$ plane (i.e., $x_1 = 7.86, x_2 = 0, x_3 = 2.14$) is realized at y_a^*, the objective value is $\bar{\Phi}(y_a^*) = 149.49$. Since this UE solution is at the $x_1 - x_3$ plane, 149.49 is the lowest possible objective value (the best case) when $y = 13.57$.

To further compare the performance of the two toll pricing approaches, we compute the average value (the average of the best and worst scenarios) and variation (the difference of the best and worst scenarios) of the upper level objective value for a given toll. The risk-prone approach has a larger average value: 175 than the risk-averse approach: 158.67. The variation for the risk-prone approach is $|\bar{\eta}(y_p^*) - \eta(y_p^*)| = 100$, which is higher than that for the risk-averse approach: $|\bar{\Phi}(y_a^*) - \Phi(y_a^*)| = 18.37$. Clearly, $y = 13.57$ generates a set of solutions whose average objective value is less than that by $y = 5$, and with a smaller variation. Therefore, at least for this small example, we can conclude that the risk-averse design approach is superior to the risk-prone approach (which is currently the most popularly used approach for SBTP).

5 Solution Set Representation of Affine UEs

The above analysis shows that in order to solve the risk-averse SBTP model, it is necessary to explore the explicit representation of the solution set of UE, i.e. $S(y)$ for a given toll vector y. Due to the difficulty of characterizing the solution set of a general UE, we concentrate on affine UEs in this paper, in which link travel time is a linear function of link flow. We first introduce a link-node complementarity model for UEs in Section 5.1 and in Section 5.2 an explicit representation of the solution set of an affine UE is presented.

5.1 A Link-Node Nonlinear Complementarity Formulation for UEs

Assume that Q is the set of destination nodes in a network and $q \in Q$ is a given destination. Denote v_a^q the flow of link a with respect to destination q, d_i^q the traffic demand from node i to q, and π_i^q the minimum travel time from i to q. We then have $x_a = \sum_{q \in Q} v_a^q$. We also set $d_q^q = 0$ and $\pi_q^q = 0, \forall q \in Q$. Denote vectors $v^q = (v_a^q)_{\forall a \in A}, \pi^q = (\pi_i^q)_{\forall i \in N, i \neq q}, d^q = (d_i^q)_{\forall i \in N, i \neq q}$ as destination-specific variables. Define vectors $v = (v^q)_{\forall q \in Q}$, $\pi = (\pi^q)_{\forall q \in Q}$, and $u = (\pi^T \ v^T)^T$. We also denote $\Lambda \in \mathbb{R}^{|N|} \times \mathbb{R}^{|A|}$ the link-node incidence matrix, i.e.,

$$\Lambda_{i,a} = \begin{cases} 1, \text{if node } i \text{ is the tail (starting) node of link } a, \\ -1, \text{if node } i \text{ is the head (ending) node of link } a, \\ 0, \text{otherwise} \end{cases}$$

Further Λ_q is Λ with the row corresponding to destination q removed, which guarantees that Λ_q has full row rank. Given the above notation, UE can be formulated as a link-node nonlinear complementarity model as follows [19]:

$$NCPUE(0) \qquad 0 \leq (\Lambda_q v^q - d^q) \perp \pi^q \geq 0, \forall q \in Q, \tag{22}$$

$$0 \leq (-\Lambda_q^T \pi^q + t(\textstyle\sum_{q \in Q} v^q)) \perp v^q \geq 0, \forall q \in Q. \tag{23}$$

Here "⊥" reads as "perpendicular", i.e., $x \perp y \leftrightarrow x^T y = 0$. The above model is denoted as *NCPUE(0)*, where "0" represents the fact that no toll is imposed and NCPUE stands for Nonlinear Complementarity Problem formulation (NCP) for User Equilibrium. The function in (23), i.e. $\Lambda_q v^q - d^q$, represents the flow conservation at nodes of the network for a specific destination q, while the function defined in (23), i.e. $-\Lambda_q^T \pi^q + t(\sum_{q \in Q} v^q)$, is for the route choice condition at nodes of the network. Detailed discussions of the model can be found in [19]. If toll y is imposed, the UE problem can be modeled as (see Section 2.1):

$$NCPUE(y) \qquad 0 \leq (\Lambda_q v^q - d^q) \perp \pi^q \geq 0, \forall q \in Q, \tag{24}$$

$$0 \leq (-\Lambda_q^T \pi^q + t(\textstyle\sum_{q \in Q} v^q) + y/\theta) \perp v^q \geq 0, \forall q \in Q. \tag{25}$$

As show in [19], the following lemma holds for *NCPUE(y)*.

Lemma 5 *The following statements hold for NCPUE(y):*

(a) If travel time t is continuous with respect to x, then NCPUE(y) has at least one solution;

(b) If t is further strictly monotone with respect to x, then NCPUE(y) has a unique solution in terms of total link flow x.

Proof. See proofs for Theorems 2 and 3 in [19]. □.

Lemma 5 implies that if t is only monotone with respect to x, the optimal total link flow may not be unique for a given toll vector y. In other words, the upper level objective function, defined on total link flows and toll variables, may not have a unique value. In this case, the risk-prone and risk-averse approaches may produce quite different toll pricing schemes.

5.2 Characterization of the Solution Set of an Affine UE

For an affine UE, the link travel time t is a linear function of total link flow x, i.e., we can define t as:

$$t(x) = \alpha x + \beta. \tag{26}$$

Here $\beta \in \mathbb{R}^{|A|}$ is a vector of link free flow travel times and $\alpha \in \mathbb{R}^{|A|} \times \mathbb{R}^{|A|}$ is a matrix for link interactions among different links. In other words, its entry $\alpha_{a,b}$ represents the contribution of traffic flow of link b to the travel time of link a. Therefore, we would expect all the elements of matrix α are non-negative. In particular, its diagonal entries should be all positive since as flow increases on a link, its travel time should always increase monotonically. Further, if α is a symmetric matrix, there will be no link interactions or link interactions are symmetric. Otherwise, link interactions will be asymmetric [20].

Since $x = \sum_{q \in Q} v^q$, (26) can also be expressed as:

$$t(\sum_{q \in Q} v^q) = \alpha(\sum_{q \in Q} v^q) + \beta, \tag{27}$$

From these two definitions, we have

$$\partial t / \partial x = \partial t / \partial v^q = \alpha, \forall q \in Q. \tag{28}$$

Substituting (27) into *NCPUE(y)* and noticing (28), we have the following standard form for an affine UE:

$$0 \leq [Mu + p] \perp u \geq 0. \tag{29}$$

Here M is a matrix and p is a vector, defined as

$$M = \begin{bmatrix} 0 & \cdots & 0 & \Lambda_1 & 0 & 0 \\ \vdots & \ddots & \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \Lambda_{|Q|} \\ -\Lambda_1^T & 0 & 0 & \alpha & \cdots & \alpha \\ 0 & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & 0 & -\Lambda_{|Q|}^T & \alpha & \cdots & \alpha \end{bmatrix}, \tag{30}$$

$$p = \begin{pmatrix} \vdots \\ -d^q \\ \vdots \\ \beta + y/\theta \\ \vdots \\ \beta + y/\theta \end{pmatrix}. \tag{31}$$

Clearly, M is positive semidefinite if α is so [19]. In this case, the solution set of *NCPUE(y)* can be explicitly characterized using any known solution. This is formally stated in the following theorem.

Theorem 6 *Assume that t is an affine function of x as defined in (27) and α is positive semidefinite. Further assume that $\bar{u} = (\bar{\pi}^T \ \bar{v}^T)^T$ is a known solution to NCPUE(y), i.e., $\bar{u} \in S(y)$. Then the solution set $S(y)$ can be represented as follows:*

$$S(y) = \Big\{ x = \sum_{q \in Q} v^q \mid \exists (\pi^T \ v^T)^T \geq 0 \tag{32}$$

$$\Lambda_q v^q - d^q = 0, \forall q \in Q, \tag{33}$$

$$-\Lambda_q^T \pi^q + \alpha(\sum_{q \in Q} v^q) + \beta + y/\theta \geq 0, \forall q \in Q, \tag{34}$$

$$(\alpha + \alpha^T)(\sum_{q \in Q} v^q - \sum_{q \in Q} \bar{v}^q) = 0, \tag{35}$$

$$-\sum_{q \in Q} (d^q)^T (\pi^q - \bar{\pi}^q) + (\beta + y/\theta) \sum_{q \in Q} (v^q - \bar{v}^q) = 0 \Big\}. \tag{36}$$

Proof. First, M is positive semidefinite since α is positive semidefinite. Also, (29) is an NCP (nonlinear complementarity problem) defined on the non-negative orthant. Therefore, according to [13] (Lemma 2.4.12), $S(y)$ can be represented as:

$$S(y) = \Big\{ u \geq 0 \quad | \quad Mu + p \geq 0, \tag{37}$$

$$(M^T + M)(u - \bar{u}) = 0, \tag{38}$$

$$p^T (u - \bar{u}) = 0 \Big\}. \tag{39}$$

Substituting (30) and (31) into (37) - (39), we can obtain (32) - (36) for $S(y)$. □

The solution set representation $S(y)$ in (32) - (36) merits further discussions. First, for a given toll vector y, $S(y)$ is a nonempty polyhedral set since it contains at least $\bar{u}$. Second, matrix α represents the link interactions for calculating link travel times. If α is a diagonal matrix (i.e., no link interaction exists), since all its entries are positive, α is positive definite. This implies, based on Theorem 6, that *NCPUE(y)* has a unique solution in terms of total link flow. In this case, risk-prone and risk-averse approaches will produce the same solution since the upper level objective function is defined on total link flows. However, if α is not a diagonal matrix (i.e., link interaction does exist), multiple solutions may exist since α may not be positive-definite. Actually, for the affine case, we can see from (35) that if $\alpha + \alpha^T$ is non-singular, we will have $x = \sum_{q \in Q} v^q - \sum_{q \in Q} \bar{v}^q = \bar{x}$. This is a relaxed condition for the uniqueness of total link flow compared with Theorem 6 for the general UE case.

6 Solution Approach and Numerical Results

There are various methods in the literature for solving the risk-prone toll pricing model *MPECSBTP* or *RPSBTP*. One can refer to [1] or [3] for more details, or refer to [21] for solution algorithms for general MPECs. In this paper, we focus on solving the risk-averse model *RASBTP*.

6.1 Solution Approach for RASBTP

First, by the definition of $\Phi(y)$ in (10), *RASBTP* can be rewritten as

$$\min_{y \in K_y} \quad \Phi(y). \tag{40}$$

In most cases, $\Phi(y)$ does not have a close-form expression since it involves solving the maximization problem in (10). Therefore, computing the derivatives of $\Phi(y)$ is usually difficult. However, for a given y, evaluating the value of $\Phi(y)$ is relatively straightforward. This can be done in two steps. In the first step, one needs to solve $UE(y)$ in Section 2. In this paper, we focus on the NCP based UE model with toll (24) - (25), which can be sovled by the decomposition scheme developed in [22]. The solution, denoted as $(\bar{\pi}, \bar{v})$ can be used to construct the solution set $S(y)$ as shown in (32) - (36). In the second step, the maximization problem in (10) can be solved using standard NLP algorithms.

The above analysis motivates us to adopt certain direct search method to solve (40), which does not require to evaluate derivatives of $\Phi(y)$. The simplex method [23, 9] is adopted in this paper. Assume that the toll vector y is in an n-dimension space, i.e. $y \in R^n$. A simplex in R^n is the convex hull of $n+1$ points, denoted as $y^0, y^1, \ldots, y^n$. In particular, if we denote y_{good} and y_{bad} the "good" and "bad" vertices of the simplex, that is, they satisfy

$$\Phi(y_{good}) = \min\nolimits_{i=0,1,\ldots,n} \; \Phi(y^i), \tag{41}$$

$$\Phi(y_{bad}) = \max\nolimits_{i=0,1,\ldots,n} \; \Phi(y^i). \tag{42}$$

Denote $\hat{y}$ the centroid of the simplex formed by the vertices other than y_{bad}, i.e.

$$\hat{y} = \frac{1}{n}(-y_{bad} + \sum_{i=0}^{n} y^i). \tag{43}$$

The method starts with an initial simplex and replaces at each iteration y_{bad} via one of the three steps: reflection, expansion, and contraction. For this purpose, we further define three points. The reflection point y_{ref} lies on the line passing through y_{bad} and $\hat{y}$, and is symmetric to y_{bad} with respect to $\hat{y}$:

$$y_{ref} = 2\hat{y} - y_{bad}. \tag{44}$$

The expansion point y_{exp} is on the line passing y_{ref} and $\hat{y}$, and is symmetric to $\hat{y}$ with respect to y_{ref}:

$$y_{exp} = 2y_{ref} - \hat{y}. \tag{45}$$

Lastly, the contraction point y_{con} is the middle point of $\hat{y}$ and y_{bad} or $\hat{y}$ and y_{ref} depending on the objective values $\Phi(y_{bad})$ and $\Phi(y_{ref})$. More specifically,

$$y_{con} \quad = \quad \begin{cases} \frac{1}{2}(y_{bad} + \hat{y}), \text{ if } \Phi(y_{bad}) \leq \Phi(y_{ref}), \\ \frac{1}{2}(y_{ref} + \hat{y}), \text{otherwise.} \end{cases} \tag{46}$$

We can then define three replacement rules as follows:

(a) If the reflection point has the minimum objective, i.e. $\Phi(y_{good}) > \Phi(y_{ref})$, then use y_{exp} to replace y_{bad} if $\Phi(y_{exp}) < \Phi(y_{ref})$. Otherwise, use y_{ref} to replace y_{bad}. This is called the (attempt) expansion step;

(b) If the reflection point has an intermediate objective, i.e. $max(\Phi(y^i)|y^i \neq y_{bad}) > \Phi(y_{ref}) \geq \Phi(y_{good})$, use y_{ref} to replace y_{bad}. This is called the reflection step;

(c) If the reflection point has the maximum objective, i.e. $\Phi(y_{ref}) \geq max(\Phi(y^i)|y^i \neq y_{bad})$, use y_{con} to replace y_{bad}. This is called the contraction step.

After the replacement, a new simplex is generated and the simplex method starts the next iteration with this new simplex. Most simplex methods in the literature generate a single point in the reflection, expansion, and contraction steps. A general approach is proposed in [9], which can generate a set of points in the reflection/expansion/contraction steps. This method, called the fortified-descent simplex method, also improves the traditional simplex methods by accepting a trial simplex only if certain fortified descent criteria (stronger than the strict descent criteria) are satisfied. The fortified-descent criteria basically guarantee that the improvement of the new vertex at each iteration is larger than some threshold. Because of its improved performance, we adopt the fortified-descent simplex method in this paper to solve (40). All the simplex methods proposed so far (including the fortified-descent method in [9]) are for nonconstrained optimization problems. For our particular *RASBTP* problem, however, the toll vector must lie in a box constraint K_y. To address this issue, we introduce a penalized objective function by integrating the box constraint:

$$\Psi(y) \equiv \Phi(y) + C[max(0, y_l - y) + max(0, y - y_u)], \tag{47}$$

where C is a big positive number (10^6 is used in this paper). Clearly, the second term of the right hand side of (47) is zero if $y_l \leq y \leq y_u$. On the other hand, $\Psi(y)$ will admit a large value if y is outside the range of $[y_l, y_u]$. The detailed descriptions of the algorithm can be found in Appendix D, which is a simplified version of that proposed in [9].

The fortified-descent simplex method adopted in this paper only works well for problems with small dimensions (in terms of the toll vector y or the size of set P in Section 2.1). Since it is arguably true that the dimension of the toll vector y should be small in practice (e.g. in the San Francisco Bay Area, there are only 8 tolled bridges), the simplex method is expected to be able to solve the *RASBTP* model proposed in this paper on networks with reasonable size.

6.2 Numerical Example

We show in this section how the fortified-descent simplex method can be used to solve the example problem in Section 4. For this purpose, we implemented the algorithm in Matlab, except the evaluation of the objective $\Psi(y)$ for a given toll y. The latter was done in GAMS [24], including solving the NCP-based link-node UE model (24) - (25), constructing the solution set $S(y)$ via (32) - (36), and solving the maximization problem (10). We notice that although one can analytically derive $f(y, x)$ as a linear function of x_2 as shown in Section 4, when solving the maximization problem (10) directly, however, it has to be treated as an NLP (i.e. $f(y, x)$ is a quadratic function of x). Therefore, the NLP solver CONOPT in GAMS was used for solving (10).

For this particular example, since the toll y is a scalar, we have $n = 1$. Therefore, any simplex will contain only two vertices (scalars). As a result, the reflection/expansion/contraction point sets will all be singleton. Further, one of the two vertices will be y_{good} and the other one is y_{bad}. The centroid point will coincide with y_{good}.

We choose $y^0 = 0, y^1 = 4$ as the initial simplex which are within the constraint $K_y = \{y|1 \leq y \leq 15\}$. Table 1 illustrates how the new simplicies are generated for the first 8 iterations of the algorithm. In this table, each iteration contains two rows: one is for the toll variable and the other one is for the corresponding objective value. The five columns named "bad", "good", "ref", "exp", "con" are for the bad, good, reflection, expansion, and contraction points respectively. The "bad" and "good" vertices constitute the simplex at the *start* of an iteration, which are shown in italic texts. The "bad" vertex will be replaced by a new one which must be one of the reflection/expasion/contraction points. The selected new vertex for each iteration is highlighted in bold text in the table. Notice that when the fortified-descent criteria are satisfied, the generation of the contraction point will be skipped. In this case, the "con" column is filled with "/". Furthermore, "inf" in the table indicates that the given toll vector y is outside K_y so that its corresponding objective value is too large according to (47).

Iteration #		*bad*	*good*	ref	exp	con
1	y	0.0000	4.0000	8.0000	**12.0000**	/
	obj	311.1111	239.1111	288.0000	**169.7778**	/
2	y	4.0000	12.0000	20.0000	28.0000	**8.0000**
	obj	239.1111	169.7778	inf	inf	**192.0000**
3	y	8.0000	12.0000	16.0000	20.0000	**10.0000**
	obj	192.0000	169.7778	inf	inf	**177.7778**
4	y	10.0000	12.0000	**14.0000**	16.0000	/
	obj	177.7778	169.7778	**168.0000**	inf	/
5	y	12.0000	14.0000	16.0000	18.0000	**13.0000**
	obj	169.7778	168.0000	inf	inf	**168.1111**
6	y	13.0000	14.0000	15.0000	16.0000	**13.5000**
	obj	168.1111	168.0000	169.4444	inf	**167.8611**
7	y	14.0000	13.5000	13.0000	12.5000	**13.7500**
	obj	168.0000	167.8611	168.1111	168.7500	**167.8819**
8	y	13.7500	13.5000	13.2500	13.0000	**13.6250**
	obj	167.8819	167.8611	167.9375	168.1111	**167.8594**

Table 1: Fortified-Descent Simplex Method

We can see from the table that in the first iteration, $y = 0$ is the "bad" vertex and $y = 4$ is the "good" vertex. The reflection and expansion points are 8 and 12 respectively. Evaluations of the objective values at these points reveal that the expansion point $y = 12$ has the minimum objective (169.7778). The expansion point is thus selected to replace the "bad" vertex. The second iteration starts with the simplex $y^0 = 4, y^1 = 12$. Both the reflection and expansion points (20 and 28 respectively) are outside K_y. Based on (47), their objectives are so large that the fortified-descent criteria are violated. The contraction step is thus called, which generates a point $y = 8$ to replace the bad vertex $y = 4$. In the third iteration, a contraction step is also executed with 10 replacing 8. The reference point (14) is taken in the forth iteration, with the resulting simplex is $[12, 14]$. As the optimal solution is 13.57, the algorithm will always perform the contraction step starting from the fifth iteration. The contracting step will simply generate a point in between the two vertices and replace the "bad" point with the new point. The newly generated point may be the "bad" or "good" point in the next iteration depending on its objective value. This process repeats itself and after 20 iterations, the obtained solution is 13.5714, exactly the same as the optimal solution (or the relative error is less than 10^{-6}).

In Figure 3(a) and 3(b), we show respectively the change of the simplices and objective values among iterations. The thin vertical line for each iteration in Figure 3(a) represents the actual simplex at the end of the iteration (i.e. the new simplex). We can see that the simplices gradually shrink for later iterations (although not strictly monotonically). At the end of the 20th iteration, the difference of the two vertices of the simplex is 6×10^{-5}. Similarly, Figure 3(b) depicts that the objective values of the two vertices of the simplex become closer as the iteration number increases. After the 20th iteration, the difference of the two objectives is less than 10^{-6}. Both figures illustrate the convergence of the algorithm for solving the $RASBTP$ of the illustrative example.

7 Conclusion

We studied the SBTP problem under the situation where the solution of the lower level UE is not unique. For this purpose, we proposed to capture the risk-taking behavior of the toll designer, where "risk" is defined as whether the objective of the toll designer can be obtained or not. We showed that existing SBTP approaches, formulated as bilevel or MPEC problems, are risk-prone in the sense that they are optimal for the best case scenario. As UE solution varies under a given toll, the design objective will always worse off. To achieve more robust tolling, we proposed a risk-averse SBTP approach by optimizing for the worst-case scenario. As opposed to the risk-prone approach, the objective value will always better off as UE solution varies for the risk-averse approach. The illustrative example provided in this paper showed that in some cases, the risk-averse SBTP solution is superior to the risk-prone solution.

We provided a general solution existence condition for the *min-max* formulation of the risk-averse approach. It

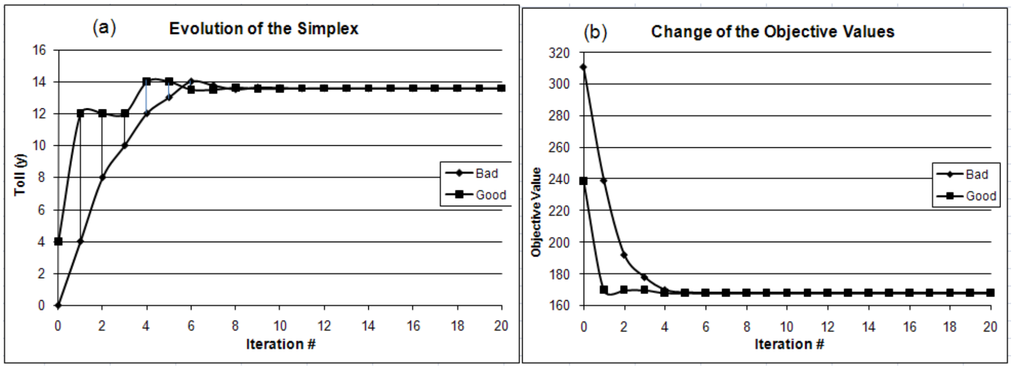

Figure 3: Performance of the Algorithm on the Test Problem

turned out that in order for this condition to hold, one requires some condition stronger than monotonicity but weaker than strict monotonicity. In case this condition does not hold, we replaced the original lower level solution set by a set of approximate solutions to the lower level problem. By extending the results in [7, 8] for weak Stackelberg games (whose lower level problems are NLPs) to risk-averse SBTP (whose lower level problems are VIs), we proved that such a replacement is effective and the upper level problem has at least one solution under mild conditions. Moreover, the optimal objective values of such problems converge to that of the original problem as the approximation error goes to zero.

To solve the risk-averse model, we first noticed that the solution set of the lower level UE needs to be explicitly expressed. We studied affine UE in this paper. By adopting the link-node nonlinear complmentarity formulation for UE, we showed that the solution set can be explicitly represented as a polyhedron if the UE is monotone. Using this explicit solution set representation, we observed that the function evaluation of the inner *max* problem can be easily conducted. This observation motivated us to use the fortified-descent simplex method to solve the risk-averse model. In the numerical example, we presented in detail the procedure and performance of applying this method to risk-averse SBTP.

The present paper shows that the uncertainty caused by non-uniqueness of UE solutions adds more complexity to model SBTP. By introducing the concept of a toll designer's risk-taking, we provided alternative ways (i.e. the risk-averse approach) to address this uncertainty compared with existing SBTP approaches (i.e. risk-prone). There are several issues in this line however that need further investigations, some of which are summarized below:

(a) Besides optimizing for the best-case and worst-case scenarios, a toll designer may want to minimize the "expected" objective value as UE solution varies. This leads to "risk-neutral" SBTP. The authors are investigating the modeling issues of risk-neutral SBTP, which can be formulated as a stochastic program. Some initial results can be found in [14].

(b) The solution existence conditions of risk-averse SBTP merits further investigations. First, what are the exact conditions in-between monotonicity and strict-monotonicity that can guarantee the solution existence of the risk-averse model? Obviously, we need more studies, especially for SBTP with general UE, to answer this question. Second, although applying the approximate solution set of the lower UE can easily guarantee the solution existence of risk-averse SBTP, such a scheme imposes more complexity in the solution process. In particular, the $RLLE$ problem (13) itself requires careful investigation since it is neither a VI or an NLP. How to efficiently solve $RLLE$ and evaluate $S^{\epsilon}(y)$ remains an open question.

(c) The solution algorithm for the risk-averse model requires an explicit expression of the solution set of the lower level UE. Although such an expression can be readily constructed for affine UEs, extending the results to general UEs requires further research. In addition, the fortified-descent simplex method needs to be evaluated on large-scale problems to test its solution performance.

(d) The UE "solution" in this article refers to link flows instead of path flow. It is well-known that the path flow solution is generally not unique even when the UE problem is strictly monotone (i.e. the link travel time function is strictly monotone with total link flows). The upper level of SBTP however still attains the same objective value if the objective function only involves link flow variables and the UE is strictly monotone (see Section 2.1. The proposed risk-averse model however may be applied to cases where the upper level objective function has to be expressed by path flows directly. This includes for example cases with nonadditive path travel times [25], which is worth further investigations.

References

[1] S. Lawphongpanich and D. Hearn. An MPEC approach to second-best toll pricing. *Mathematical programming B*, 101:33–55, 2004.

[2] S. Lawphongpanich, D.W. Hearn, and M.J. Smith. *Mathematical And Computational Models for Congestion Charging*. Springer, 2006.

[3] H. Yang and H.J. Huang. *Mathematical and Economic Theory of Road Pricing*. Elsevier, 2005.

[4] H.M Zhang and Y.E. Ge. Modeling variable demand equilibrium under second-best road pricing. *Transportation Research B*, 38:733–749, 2004.

[5] A. Ben-Tal and A. Nemirovski. Robust optimization - methodology and applications. *Mathematical Programming, Series B*, 92:324–343, 2002.

[6] G. Leitmann. On generalized Stackelberg strategies. *Journal of Optimization Theory and Applications*, (4):637–643, 1978.

[7] M.B. Lignola and J. Morgan. Topological existence and stability for Stackelberg problems. *Journal of Optimization Theory and Applications*, 84(1):145–169, 1995.

[8] P. Loridan and J. Morgan. Weak via strong Stackelberg problem: New results. *Journal of Global Optimization*, 8:263–287, 1996.

[9] P. Tseng. Fortified-descent simplicial search method: a general approach. *SIAM Journal of Optimization*, 10(1):269–288, 1999.

[10] M.J. Smith. The existence, uniqueness and stability of traffic equilibria. *Transportation Research, Part B*, 13:295–304, 1979.

[11] S. Dafermos. Traffic equilibrium and variational inequalities. *Transportation Science*, 14:42–54, 1980.

[12] A. Nagurney. *Network Economics: A Variational Inequality Approach (2nd Edition)*. Kluwer Academic Publishers, 1998.

[13] F. Facchinei and J.S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems: Vol. I, II.* Springer, 2003.

[14] X. Ban and M.C. Ferris. Risk-neutral second-best toll pricing. *In Preparation*, 2008.

[15] R. Tyrrell Rockafellar and Roger J–B Wets. *Variational Analysis*. Number 317 in Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin, 1998.

[16] Stephen M. Robinson. Solution continuity in monotone affine variational inequalities. *SIAM Journal on Optimization*, 18:1046–1060, 2007.

[17] Stephen M. Robinson. Some continuity properties of polyhedral multifunctions. *Mathematical Programming Studies*, 14:206–214, 1981.

[18] M. Gowda. On the continuity of the solution map in linear complementarity problems. *SIAM Journal on Optimization*, 2:619–634, 1992.

[19] X. Ban, M.C. Ferris, and H.X. Liu. An MPCC formulation and numerical studies for continuous network design with asymmetric user equilibria. *Submitted for publication*, 2007.

[20] Y. Sheffi. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods.* Prentice-Hall, Inc., 1985.

[21] Z.Q Luo, J.S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints.* Cambridge University Press, 1996.

[22] X. Ban, H.X. Liu, and M.C. Ferris. A link-node based complementarity model and its solution algorithm for asymmetric user equilibria. In *Proceedings of the 85th Annual Meeting of Transportation Research Board (CD-ROM)*, 2006.

[23] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, 1995.

[24] A. Brooke, D. Kendrick, A. Meeraus, and R. Raman. Gams, a user's guide. Technical report, GAMS Development Corporation, 1998.

[25] R.P. Agdeppa, N. Yamashita, and M. Fukushima. The traffic equilibrium problem with nonadditive costs and its monotone mixed complementarity problem formulation. *Transportation Research, Part B*, 41(8):862–874, 2007.

Appendices

A Definition and Properties of A Set-Valued Map

This appendix provides definitions and properties of set-valued maps, and some limit definitions for real-valued functions. The results shown here can also be found in [13] or [15].

Definition A.1 Denote map Φ is a set-valued map from $\mathbb{R}^n$ to the power of $\mathbb{R}^n$. Then for any $x \in \Phi$, $\Phi(x)$ is a subset of $\mathbb{R}^n$ (possibly empty). The domain of Φ, denoted domΦ, the range of Φ, denoted as ranΦ, and the graph of Φ, denoted as gphΦ, are defined respectively as:

$$\begin{aligned} \text{dom}\Phi &\equiv \{x \in \mathbb{R}^n : \Phi(x) \neq \emptyset\} \\ \text{ran}\Phi &\equiv \bigcup_{x \in dom\Phi} \Phi(x) \\ \text{ghp}\Phi &\equiv \{(x, y) \in \mathbb{R}^{2n} : y \in \Phi(x)\} \end{aligned}$$

Definition A.2 A set-valued map $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is said to be

(a) *closed* at at point $\bar{x}$ if

$$\left. \begin{array}{c} x^k \longrightarrow \bar{x} \\ y^k \in \Phi(x^k)\ \forall k \\ y^k \longrightarrow \bar{y} \end{array} \right\} \implies \bar{y} \in \Phi(\bar{x}) \ ;$$

(b) *closed* on a set S if Φ is closed at every point of S.

(c) *upper semicontinuous* at a point $\bar{x}$ if for every open set υ containing $\Phi(\bar{x})$, there exists an open neighborhood $\mathcal{N}$ of $\bar{x}$ such that, for each $x \in \mathcal{N}$, υ contains $\Phi(x)$.

(c) *lower semicontinuous* at a point $\bar{x}$ if for every open set υ meeting $\Phi(\bar{x})$, there exists an open neighborhood $\mathcal{N}$ of $\bar{x}$ such that, for each $x \in \mathcal{N}$, υ meets $\Phi(x)$.

Theorem A.1 The following statements are true for a set-valued map Φ.

(a) Suppose $\Phi(\bar{x})$ is a closed set. If Φ is upper semicontinuous at $\bar{x}$, then Φ is closed at $\bar{x}$;

(b) Φ is closed if and only if its graph is a closed set.

Definition A.3 Let g be a real-valued function from $\mathbb{R}^n$ to $\mathbb{R}$. The lower limit of g at a point $y_0 \in \mathbb{R}^n$ is defined by

$$\liminf_{y \to y_0} g(y) := \sup_{\epsilon > 0} \inf_{y \in B(y_0, \epsilon)} g(y),$$

and the upper limit of g at a point $y_0 \in \mathbb{R}^n$ is defined by

$$\limsup_{y \to y_0} g(y) := \inf_{\epsilon > 0} \sup_{y \in B(y_0, \epsilon)} g(y),$$

where $B(y_0, \epsilon)$ denotes the open ball around y_0 with radius ϵ. g is lower semicontinuous at y_0 if

$$g(y_0) \leq \liminf_{y \to y_0} g(y).$$

B Proof of Theorem 1

Proof. To prove (a), we rewrite the problem $UE(y)$ as a general equation $0 \in t(x) + y/\theta + N_K(x)$, where $N_K(x)$ is the set-valued map that denotes the normal cone of K at $x \in K$, defined by

$$N_K(x) = \{z \mid z^T(x' - x) \leq 0 \text{ for each } x' \in K\}.$$

Thus, x belongs to $S(y)$ if and only if it satisfies the last general equation. Consequently, we can rewrite G as

$$G = \{(y, x) \mid y_l \leq y \leq y_u, 0 \in t(x) + y + N_K(x)\}.$$

The set G above is bounded: y is bounded by its upper and lower bounds, and x is bounded because it has to belong to the compact set K.

If we let gph N_K denote the graph of the operator N_K, that is,

$$\text{gph } N_K = \{(x, t) \mid t \in N_K(x)\},$$

then we have

$$G = \{(y, x) \mid y_l \leq y \leq y_u, (x, -t(x) - y) \in \text{gph } N_K\}.$$

By the definition of the operator N_K and the definition of outer semicontinous property given in Appendix A, it is easy to check that N_K is outer semicontinuous. Consequently, gph N_K is a closed set (see Theorem A.1(a) in Appendix A). By assumption, $t(x)$ is a continuous map of x, so is the map $(x, -t(x) - y)$ with respect to (y, x). Therefore, the set

$$\{(y, x) \mid (x, -t(x) - y) \in \text{gph } N_K\}$$

is closed. The set G is the intersection of the latter set and another closed set

$$\{(y, x) \mid y_l \leq y \leq y_u\},$$

so G is closed. We already noted that G is bounded, so G is compact.

(b) Since G is compact and $f(y, x)$ is continuous with respect to (y, x) by the assumption, the problem *RPSBTP* has at least one solution. □

C Example in Which the RASBTP Has No Solutions

In the following example, the function $\Phi(y)$ as defined in (10) is not lower semicontinuous at certain points. As a result the *RASBTP* does not attain its optimal objective value in K_y.

Consider a small network in which two links connect a common origin-destination pair, with each link also being a route. Let the demand from the origin to the destination be d. Suppose that the link travel time does not depend on the link flow x, so $t(x)$ is a constant function, which is monotone but not strictly monotone. We consider tolls y_1 and y_2 on link 1 and link 2. The link generalized travel times with toll imposed are:

$$c_1 = 3 + y_1$$
$$c_2 = 3 + y_2.$$

Let $y = (y_1, y_2)$ take values in $K_y = \{(y_1, y_2) \mid 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1\}$, and define the function f to be

$$f(x, y) = c_1 x_1 + 2c_2 x_2.$$

There are the following three cases to consider.

1. When $y_2 < y_1$, we have $c_2 < c_1$, so $S(y)$ contains a single point $(0, d)$, and $\Phi(y) = \max_{x \in S(y)} f(x, y) = 2d(3+y_2)$.
2. When $y_2 > y_1$, we have $c_1 < c_2$, so $S(y)$ contains a single point $(d, 0)$, and $\Phi(y) = \max_{x \in S(y)} f(x, y) = d(3+y_1)$.
3. When $y_2 = y_1$, we have $c_1 = c_2$, so $S(y) = \{(x_1, x_2) \mid 0 \leq x_1 \leq d, 0 \leq x_2 \leq d, x_1 + x_2 = d\}$. In this case, it is not hard to verity

$$\Phi(y) = \max_{x \in S(y)} \quad f(y, x) = \max_{x \in S(y)} \quad c_1 x_1 + 2c_2 x_2 = 2d(3 + y_2).$$

When (y_1, y_2) lies on the line $y_1 = y_2$, the set $S(y)$ is a line segment with length d; when (y_1, y_2) leaves the line $y_1 = y_2$, the set $S(y)$ immediately shrinks to a singleton. Further, the function (10) is not lower semicontinuous at each point on the line $y_1 = y_2$: the lower limit of (10) at (y_1, y_1) is $d(3 + y_1)$, but the function value there is $2d(3 + y_1)$. In particular, the function (10) is not lower semicontinuous at $(0, 0)$, which is the only limit point for any sequence in K_y whose function value converge to the optimal value $3d$. Consequently, the *RASBTP* does not attain its optimal objective value in K_y.

D Fortified Descent Simplex Method for Risk-Averse SBTP

Algorithm FSMRASBTP

Step 0 Initialization. Choose $n+1$ points from K_y, denoted as $Y^0 = \{y^0, y^1, \dots, y^n\}$. Define two functions $\alpha(t) = 10^{-5} min\{0.5t^2, t\}, \beta(t) = 10^6 t^2$. Set $\theta_r = 0.1, \nu = 10^{-5}, \gamma_q = 0.5$. Set $k = 0$ and $Y = Y^k$.

Step 1 Construct the Set of Reflection Points. Let $\Delta = diam(Y), m = min\{n, l(F(Y), F(Y^k))\}$. Partition Y into two sets Y_{good} and Y_{bad} so that $|Y_{good}| = m, |Y_{bad}| = n+1-m$, and $\Psi_{max}(Y_{good}) \leq \Psi_{min}(Y_{bad})$. Compute the centroid point of the set Y_{good} and its associated objective value:

$$\hat{y} = \frac{1}{m} \sum_{y \in Y_{good}} y, \tag{48}$$

$$\hat{\Psi} = \frac{1}{m} \sum_{y \in Y_{good}} \Psi(y). \tag{49}$$

Compute the reflection set Y_r as follows:

$$Y_r = 2\hat{y} - Y_{bad}. \tag{50}$$

If $Y = Y^k$, set $y^k = \hat{y}, \bar{m} = m$.

Step 2 Check the Fortified Descent Criteria. If the following two conditions hold (*fortified descent criteria*):

$$\Psi_{min}(Y_r) \leq \Psi_{max}(Y_{good}) - \alpha(\Delta), \tag{51}$$

$$\Psi_{min}(Y_r) \leq \Psi_{max}(Y_{good}) - \theta_r(\Psi_{max}(Y_{bad}) - \hat{\Psi}) + \beta(\Delta), \tag{52}$$

go to Step 3; otherwise, go to Step 4.

Step 3 Expansion. Set $Y_e = 3\hat{y} - 2Y_{bad}$. If $\Psi_{min}(Y_e) \leq \Psi_{min}(Y_r)$, set $Y^{k+1} = Y_{good} \cup Y_e$ (*accept expansion*); else, set $Y^{k+1} = Y_{good} \cup Y_r$ (*accept reflection*). Set $\Delta_k = \Delta, \bar{m}_k = m, k = k+1$ and go to Step 0.

Step 4 Contraction. Define

$$Y_c = \begin{cases} 1.5\hat{y} - 0.5Y_{bad}, \text{ if } \Psi_{min}(Y_r) < \Psi_{min}(Y_{bad}), \\ 0.5\hat{y} + 0.5Y_{bad}, \text{otherwise} \end{cases} \tag{53}$$

We now look at the set $Y_{good} \cup Y_c$. If the following two conditions hold:

$$F_i(Y_{good} \cup Y_c) \leq F_i(Y^k), \forall i = 1, \dots, m+1, \tag{54}$$

$$\sum_{i=1}^{m+1} F_i(Y_{good} \cup Y_c) \leq \sum_{i=1}^{m+1} F_i(Y^k) - \alpha(\Delta), \tag{55}$$

then set $Y^{k+1} = Y_{good} \cup Y_c$ (*accept contraction*), $\Delta_k = \Delta, \bar{m}_k = m+1, k = k+1$, and go to Step 0. Otherwise, go to Step 5.

Step 5 Shrink the Simplex. Denote $y_{best} = argmin_{y \in Y} \Psi(y)$. Set $Y' = y_{best} + \gamma_q(Y - y_{best})$. If $\Psi_{min}(Y') \leq \Psi_{min}(Y^k) - \alpha(\Delta)$, set $Y^{k+1} = Y', \Delta_k = \Delta, \bar{m}_k = 1, k = k+1$, and go to Step 0 (accept the shrunken simplex). Otherwise, set $Y = Y'$ and go to Step 1 (*accept a nonimproving shrink*).

In Step 1, $|Y|$ denotes the cardinality of a set Y and the *diameter* of set Y is defined as [9]:

$$diam(Y) = max_{y \in Y, y' \in Y} \|y - y'\|,$$

where $\| \cdot \|$ denotes the 2-norm. The *consistency index* function l is defined for any two n-vectors c and d as [9]:

$$l(c, d) = \max\ i \in 0, 1, \ldots, n \text{ such that } c_j \leq d_j \text{ for } 1 \leq j \leq i.$$

Lastly, for a set Y with p vectors, the function $F(Y)$ is defined as a permutation of the objective values $\{\Psi(y)|y \in Y\}$ in an increasing order [9], i.e. ,

$$F(Y) \quad = \quad \begin{bmatrix} F_1(Y) \\ \ldots \\ F_p(Y) \end{bmatrix}, \quad \text{where } F_1(Y) \leq \ldots \leq F_p(Y),$$

and $F_i(Y)$ denotes the ith smallest element of $\{\Psi(y)|y \in Y\}$.

The set of reflection points Y_r is generated in Step 1. In Step 2, (51) and (52) are the fortified-descent criteria, which accept the expansion or reflection of the simplex only when the improvement of the reflection points is larger than some threshold represented by $\alpha(\Delta)$. In Step 4, the contraction step is accepted only when the two fortified criteria (54) and (55) are satisfied. They require that (a) the contraction points set must bring at least one point that improves Y_{good} and (b) the improvement (in terms of objective values) by bringing this point to Y_{good} must be larger than $\alpha(\Delta)$. In Step 5, if the algorithm returns too many times to Step 1 because of accepting a nonimproving shrink, the algorithm should stop with the best solution found.