This is the Preprint Version.

This is the preprint version of the following article: Xu, Z., Chen, A., Li, G., Li, Z., & Liu, X. (2024). Elastic-demand bi-criteria traffic assignment under the continuously distributed value of time: A two-stage gradient projection algorithm with graphical interpretations. Transportation Research Part E: Logistics and Transportation Review, 183, 103425 which is available at https://doi.org/10.1016/j.tre.2024.103425.

A two-stage gradient projection algorithm for elastic demand continuous bi-criteria traffic assignment

Zhandong Xu[1], Anthony Chen*[2], Guoyuan Li[2], Zhengyang Li[2], and Xiaobo Liu[1]

[1]School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, China

[2]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

June 2, 2023

Abstract

This paper studies the elastic demand continuous bi-criteria traffic assignment (ED-CBiTA) problem, in which users' time-cost tradeoff heterogeneity is considered through the continuously distributed value of time. The origin and destination (O-D) demand is endogenously determined by the expected generalized travel time across the travel population. A variable demand formulation and an equivalent excess demand reformulation are presented for the ED-CBiTA problem. Based on two types of Gauss-Seidel decomposition schemes, we propose a novel two-stage gradient projection (TSGP) algorithm, which implicitly delivers visual interpretations to account for the interactions between the demand and supply functions. The first stage, called demand equilibration, aims to adjust O-D demand and all efficient path flows "vertically upwards or downwards" based on the network congestion level. The second stage, namely boundary equilibration, is to perform the boundary movements and adjust adjacent efficient flows "horizontally forwards or backward", to achieve exact positions along the Pareto frontier. Numerical results on a small network show TSGP's features and confirm that TSGP significantly outperforms two link-based benchmark algorithms. For instances of practical network size, TSGP consistently promises to obtain high-quality solutions with a rather smaller CPU time.

Keywords: transportation; elastic demand; bi-criteria traffic assignment; continuous multi-class; gradient projection

*Corresponding author, E-mail: anthony.chen@polyu.edu.hk

1 Introduction

Elastic demand traffic assignment aims to establish a spatial equilibrium to predict network flows, resulting from origin-destination (O-D) demands, route choice behaviors, and the interactions between the supply functions and the demand functions (Sheffi, 1985). It consistently models the trip generation and traffic assignment in strategic transport planning. The supply functions are determined by the route attributes, whereas the demand functions depend on perceived benefits derived from travel. It the literature, many have extensively studied a single criterion (time) for the elastic demand traffic assignment problem (EDTAP), ranging from deterministic (e.g., see Babonneau and Vial, 2008; Ryu et al., 2014, 2017) to stochastic (Yu et al., 2014; Cantarella et al., 2015). Under the user equilibrium (UE) principle (Wardrop, 1952), all used paths have equal travel times and no unused path has a lower time, and O–D demands determined by the minimum travel time also satisfy the demand functions (Beckmann et al., 1956).

However, travelers' route choices may depend not only on travel times alone but also on toll costs. The value of time (VOT), which combines time and toll cost in a linear manner to form generalized travel times, plays a central role to describe how individuals make trade-offs when facing these two criteria. Many empirical studies have reported that VOT varies significantly across individuals due to many factors, such as social-economic characteristics and trip purposes (e.g., see Cirillo and Axhausen, 2006; Schmid et al., 2021). Concerning continuously distributed VOT and demand elasticity among heterogeneous users leads to the so-called *elastic demand continuous bi-criteria traffic assignment (ED-CBiTA)* problem, which is the focus of this study.

1.1 Related studies

Initial efforts on time and toll bi-criteria routing choices on uncongested networks could date back to Dial (1979), who developed a parametric path finding algorithm to search for non-dominated paths along the Pareto frontier. To consider congestion effects on travel times, Leurent (1993) proposed a path-based mathematical programming (MP) formulation for ED-CBiTA. The associated equilibrium condition was established based on the monetary expense class, which corresponds to the travelers having different VOTs but choosing the same efficient path. The variable O-D demand relates to the level of service is determined by the *expected generalized travel time (EGTT)* of that O-D pair. Later, Leurent (1996) presented a more general variational inequality (VI) formulation for ED-CBiTA, based on which the sensitivity analysis on some perturbation parameters was conducted (Leurent, 1998). Despite ED-CBiTA's potential applications in transportation planning, this complex problem that involves a potentially infinite number of user classes under demand elasticity receives rare attention since then.

Existing studies mainly focus on mathematical models and solution algorithms for the fixed demand CBiTA problem. Based on the result of Leurent (1993) and Dafermos (1981), Marcotte and Zhu (1994) presented an infinite-dimensional VI formulation and developed the Frank-Wolfe (FW) algorithm to solve the problem. In a sequel, Marcotte et al. (1996) generalized the model

to consider asymmetric congestion effects and provided implementation details to FW (Marcotte and Zhu, 1997). Dial (1996, 1997) designed another link-based algorithm, called T2, that permits both criteria to be flow-dependent based on the restricted simplicial decomposition (RSD) algorithm of Lawphongpanich and Hearn (1984). An ingenious 'lazy' trip loader is developed to efficiently generate extreme points with an origin-based tree structure. Both FW and RSD bypass the need to store and manipulate efficient paths thanks to the additive route cost structure for the two criteria. However, they could fail to fetch path flow information from the link-based equilibrium pattern. Alternatively, Marcotte (1998) provided a reformulation for CBiTA with VOT boundaries based on ordered efficient paths, which enables to the transformation of an infinite-dimensional VI into a finite-dimensional problem (Marcotte and Patriksson, 2007). Calderone et al. (2017) revisited the formulation of Leurent (1993) and Marcotte (1998) to allow the random attribute (e.g., VOT) to be negative in a potential game. Recently, Xie et al. (2021) extended the boundary-based formulation of Marcotte (1998) to design a new path-based algorithm with a Newton-type flow equilibration procedure, namely single-boundary adjustment. It was found that a path-based algorithm enables to efficiently obtaining high-quality solutions for large networks. From the perspective of application, Leurent et al. (2012) applied CBiTA to model the urban housing market, where travel demand is disaggregated based on several land-use attributes. Another typical application that is built on CBiTA is the congestion pricing problem, which has been widely studied for many years (e.g., see Dial, 1999a,b; Wu and Huang, 2014). More recently, Xu et al. (2023) considered the non-additive cost structure in the form of time surplus to investigate the effects of non-linear VOTs.

Another category of previous studies that address user heterogeneity is the discrete multi-class approach, in which the entire feasible VOT range is divided into several predetermined intervals (Dafermos, 1981; Yang and Huang, 2004; Han and Yang, 2008). If demand elasticity is further considered, each class is differentiated with a class-specific VOT value and also a demand function for standard practices (Clark et al., 2009; Slavin et al., 2014). However, this approach could have several limitations. Firstly, each predetermined interval implicitly assumes homogeneity within such a user class, bringing potential estimation errors. Second, it is difficult to specify how many user classes is needed to obtain a satisfactory solution beforehand. Third, the discretization method also poses computational challenges because the number of O-D pairs potentially grows one-fold for one additional divided class for each O-D pair. Inevitably, one may encounter huge computational burdens to preset many user classes if involving millions of O-D pairs of a real-world problem.

1.2 Summary, contribution and structure

In summary, elastic demand traffic assignment has been extensively studied either in the context of the single criterion of travel time (i.e., EDTAP) or discrete multi-class. However, ED-CBiTA receives little attention since initial efforts made by Leurent (1993, 1996, 1998), and the algorithmic development is rather limited to date. A possible reason for this lack of progress is the complexity

of ED-CBiTA itself, since the problem not only involves an infinite number of user classes, but also requires specifying endogenously determined travel demands characterized by EGTT from a continuous distribution. To solve the ED-CBiTA problem, Leurent (1993) firstly suggested the method of successive average (MSA) and Leurent (1996) later provided a more efficient path-based algorithm, referred to as Procedure of Equation-by-Transfer (PET). However, to the best of our knowledge, the algorithm detail of PET still remains unavailable and was never refined to test real-world problems. This study draws its inspiration from the remarkable performance of the advanced path-based gradient projection (GP) algorithm for various equilibrium assignment problems (e.g., see Xu et al., 2020, 2022; Zhang et al., 2023).

In this paper, we revisit the work of Leurent (1993) and present an alternative path-based formulation by extending the result of Marcotte (1998) and Xie et al. (2021) to further incorporate demand elasticity. The proposed formulation uses O-D demand and VOT-based boundaries as decision variables, which allows for replacing conventional path-based conservation constraints with simple boxes. Moreover, an equivalent excess demand reformulation is presented under a modified network representation. Unlike EDTAP, it is worth emphasizing that ED-CBiTA cannot simply be transformed into a fixed demand CBiTA problem because of continuously distributed VOT. The above observations make solving the ED-CBiTA problem a considerable challenge.

The main contribution of this paper is the development of a novel path-based *two-stage gradient projection (TSGP)* algorithm for solving the ED-CBiTA problem. In the spirit of the Gauss–Seidel method, TSGP predicates on the idea of decomposing the original problem into two types of subproblems, i.e., the O-D excess demand subproblem for the first stage (namely demand equilibration) and the single-boundary subproblem for the second stage (namely boundary equilibration). Guided by the objective function, both subproblems can be solved very efficiently using the Newton-type GP method. As we shall see, both two equilibrium operations are in fact adjusting path flows in different *"directions"* to move toward equilibrium. Interestingly,

Stage 1: Solving the subproblem is to adjust O-D demand and the path flows *vertically upwards or downwards* among all efficient paths, thereby achieving demand equilibration based on the network congestion level;

Stage 2: Solving the subproblem is to move boundaries and adjust path flows *horizontally forwards or backwards* between adjacent efficient paths, thereby achieving exact positions along the Pareto frontier.

In short, TSGP implicitly delivers visual interpretations to account for the interactions between the demand and supply functions. We use a two-link network to demonstrate TSGP's features and ED-CBiTA's generality compared with the discrete multi-class approach. We found that discretization errors on the equilibrium flow pattern could still exist even when a large number of user classes are considered. Moreover, toll revenues can be underestimated without addressing user heterogeneity. For the convergence performance, computational experiments consistently confirm that TSGP significantly outperforms the two link-based benchmark algorithms. For in-

stances of practical network size, TSGP still promises to obtain high-quality solutions with a rather smaller CPU time.

The rest of this paper is structured as follows. Section 2 describes the assumption, formulation, and solution properties for the variable demand formulation. Section 3 presents an excess demand reformation and highlights the differences with the discrete multi-class problem. The TSGP algorithm is presented in Section 4, followed by the numerical results on small and large networks in Section 5. The concluding remarks and future research directions are specified in Section 6.

2 Formulation

2.1 Definitions and assumptions

Consider a network represented as a directed graph $G(N, A)$, where N and A denote the sets of nodes and links, respectively. Let RS denote the set of O-D pairs, K_{rs} the set of simple paths from origin r to destination s, f_k the flow on path $k \in K_{rs}$. The travel demand is represented as the number of trips for each O-D pair rs, denoted by q^{rs}. Each link $a \in A$ is associated with two attributes: travel time t_a and toll cost τ_a. The path travel time c_k and path toll τ_k are, respectively, the sum of the time and cost of each link belonging to path k, i.e.,

$$c_k = \sum_{a \in A} \delta_{a,k} t_a, \quad \forall k \in K_{rs}, rs \in RS, \tag{1a}$$

$$\tau_k = \sum_{a \in A} \delta_{a,k} \tau_a, \quad \forall k \in K_{rs}, rs \in RS, \tag{1b}$$

where $\delta_{a,k} = 1$ if path k passes link a, and 0 otherwise.

The generalized path travel time is defined as a linear combination of the two objectives through a user's value of time (VOT, \$/hour):

$$C_k(\alpha) = c_k + \frac{\tau_k}{\alpha} \equiv C_k(\beta) = c_k + \beta \tau_k, \tag{2}$$

where α is VOT, $\beta = 1/\alpha$ is denoted as the time equivalence of money (TEM) to simplify notation.

To consider heterogeneous users, the VOT α (or equivalently TEM β) is treated as a random variable that follows a probability density function (PDF) across individuals. Accordingly, let $h_{rs}(\cdot)$ and $H_{rs}(\cdot)$ denote the PDF and cumulative distribution function (CDF) of β for O-D pair rs, respectively. For practical purpose, we assume that $h_{rs}(\cdot) > 0$ lies in a closed positive interval $\Phi^{rs} = [\underline{\beta}, \overline{\beta}]$. Since each user with a particular β attempts to minimize his/her generalized travel time, the continuous bi-criteria traffic assignment problem involves a potentially infinite number of user classes. Nevertheless, it is possible to track users who have different TEMs but choose the same *efficient path*, which is defined as follows:

Definition 1 (Efficient path (Dial, 1979))**.** *A path $k \in K_{rs}$ is said to be an efficient path if it is chosen by at least one user with TEM $\beta \in \Phi^{rs}$, i.e., $C_k(\beta) \leq C_{k'}(\beta)$, $\forall k' \in K_{rs}$, and $k' \neq k$.*

To explain the above concept, Figure 1(a) shows 10 paths, each of which corresponds to a certain travel time c_k and toll cost τ_k. The black line represents the Pareto frontier, along with efficient paths 1 to 4, such that the generalized time defined in Eq. (2) is minimized within the support Φ^{rs}. Consider a user with β such that $\beta_2 < \beta < \beta_3$ as plotted Figure 1(a), we have $C_3(\beta) < C_k(\beta)$ for any other path $k \neq 3$. In other words, those dominated paths (paths 5 to 10) will not be taken by any user. The consecutive slope of the frontier that connects two adjacent efficient paths is the TEM threshold across which the efficient path choice changes (i.e., β_1 to β_3). More specifically, a user will choose path 1 if his/her $\beta \in [\underline{\beta}, \beta_1)$, path 2 if $\beta \in [\beta_1, \beta_2)$, path 3 if $\beta \in [\beta_2, \beta_3)$ and path 4 if $\beta \in [\beta_3, \overline{\beta}]$. The distribution of TEM is shown in Figure 1(b), which illustrates four path-choice groups. The integral of each shaded area, i.e., $\int_{\beta_{k-1}^{rs}}^{\beta_k^{rs}} h_{rs}(z)dz$, is the choice probability corresponding to the selected efficient path.

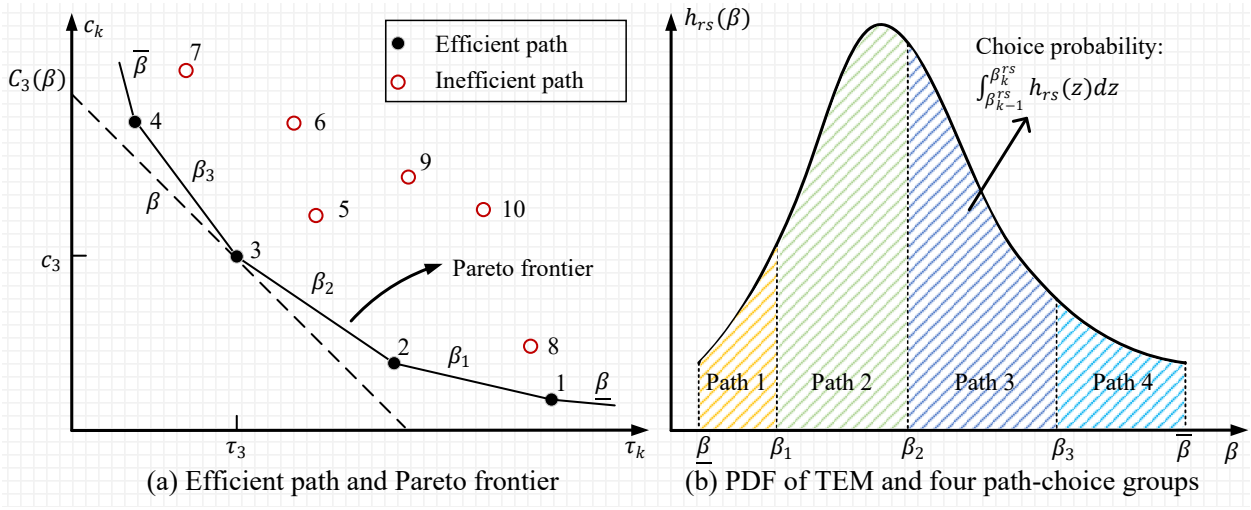


Figure 1: Illustration of efficient paths and path-choice groups.

Following Leurent (1993), Marcotte (1998) and Xie et al. (2021), we rank all paths in K_{rs} in a descending order of τ_k. A set of boundaries to generate TEM intervals is defined as

$$\Phi_k^{rs} = \begin{cases} [\beta_{k-1}^{rs}, \beta_k^{rs}), & \text{if } k = 1, \ldots, |K_{rs}| - 1 \\ (\beta_{k-1}^{rs}, \beta_k^{rs}], & \text{if } k = |K_{rs}| \end{cases} \quad \forall rs \in RS, \tag{3}$$

where $\beta_0^{rs} = \underline{\beta}$, $\beta_{|K_{rs}|}^{rs} = \overline{\beta}$, and $\beta_{k-1}^{rs} \leq \beta_k^{rs}$ for any $k = 1, \ldots, |K_{rs}|$.

After determining the TEM boundaries, given O-D demand q^{rs}, the flow on any path can be readily obtained:

$$f_k = q^{rs} \int_{\beta_{k-1}^{rs}}^{\beta_k^{rs}} h_{rs}(z)dz, \quad \forall k = 1, \ldots, |K_{rs}|, rs \in RS. \tag{4}$$

Notably, if both travel times and toll costs are flow-independent, one can first apply the parametric shortest path algorithm of Dial (1979) to identify those efficient paths for a given O-D pair, then assign a portion of flows to each efficient path according to its choice probability.

In reality, however, travel time is generally flow-dependent and the O-D travel demand would be influenced by the level of service in a congested network. To model more realistic travel characteristics, this paper is built on the following widely adopted assumptions (e.g., see Leurent, 1993; Babonneau and Vial, 2008).

Assumption 1. *The link toll cost τ_a, $\forall a \in A$ is a non-negative constant.*

Assumption 2. *The link travel time $t_a(\cdot)$, $\forall a \in A$ is a positive, continuously differentiable and strictly increasing function with respect to (w.r.t) its own flow x_a.*

Assumption 3. *The O-D travel demand q^{rs} is determined by the equilibrium expected generalized travel time (EGTT) T^*_{rs} for O-D pair rs among the population, i.e.,*

$$q^{rs} = D_{rs}(T^*_{rs}) = D_{rs}\left(\sum_{k \in K_{rs}} \int_{\beta^{rs*}_{k-1}}^{\beta^{rs*}_k} C_k(z) \cdot h_{rs}(z) dz\right), \quad \forall rs \in RS, \tag{5}$$

where $\boldsymbol{\beta}^* = \{\beta^{rs*}_k\}$ is the equilibrium boundaries corresponding to the efficient Pareto frontier, and $D_{rs}(\cdot)$ is a continuously differentiable, non-negative, upper bounded, and strictly decreasing demand function.

Assumption 4. *The inverse of demand function $D^{-1}_{rs}(\cdot)$ is integrable and separable.*

We proceed to define *elastic demand continuous bi-objective traffic assignment (ED-CBiTA)* UE condition as follows:

Definition 2 (ED-CBiTA UE)**.** *A path flow vector $\mathbf{f}^* = \{f^*_k\}$ is a ED-CBiTA UE flow if for any O-D pair $rs \in RS$ such that (i) the positive O-D travel demand $q^{rs*} = \sum_{k \in K_{rs}} f^*_k = D_{rs}(T^*_{rs})$, and (ii) for any path $k \in K_{rs}$ and $f^*_k > 0$, then $C_k(\mathbf{f}^*, \beta) \leq C_{k'}(\mathbf{f}^*, \beta)$, $\forall \beta \in \Phi^{rs*}_k \neq \emptyset$, $k' \in K_{rs}$.*

In Definition 2, Condition (i) states that the O–D travel demand must satisfy the demand functions according to EGTT at equilibrium, whereas Condition (ii) implies that any user whose TEM belongs to Φ^{rs*}_k would choose efficient path k to achieve the shortest generalized travel time per Definition 1. In a congested network under above assumptions, the TEM boundaries for efficient paths vary with the flows assigned to each path as travel times are flow-dependent, while the total travel demand depends on the network condition also relates to the boundaries, leading to complex interactions between the supply and demand functions. We now set up to present a mathematical program (MP) according to Definition 2.

2.2 Variable demand formulation

The ED-CBiTA problem simultaneously accounts for both trip generation and traffic assignment while involving an infinite number of user classes. To model ED-CBiTA, Leurent (1993) firstly proposed a convex MP that works in the space of path flows under Assumptions 1-4. However, a mapping had to be constructed between path flow $\mathbf{f}$ and TEM boundaries $\boldsymbol{\beta}$ in Leurent (1993).

Following Marcotte (1998) and Xie et al. (2021), we present an alternative MP that uses demand and TEM boundaries as decision variables, i.e.,

$$\min_{\mathbf{q},\boldsymbol{\beta}} \quad Z_1 = \underbrace{\sum_{a\in A}\int_0^{x_a} t_a(z)dz}_{\text{Beckman transformation}} + \underbrace{\sum_{rs\in RS}\sum_{k\in K_{rs}} q^{rs}\cdot\tau_k\int_{\beta_{k-1}^{rs}}^{\beta_k^{rs}} z\cdot h_{rs}(z)dz}_{\text{Continuous monetary expense}} - \underbrace{\sum_{rs\in RS}\int_0^{q^{rs}} D_{rs}^{-1}(z)dz}_{\text{Inverse demand integrals}} \tag{6a}$$

$$s.t. \quad \beta_k^{rs} - \beta_{k-1}^{rs} \geq 0, \quad \forall rs \in RS, k = 1,\ldots,|K_{rs}|, \tag{6b}$$

$$x_a = \sum_{rs\in RS}\sum_{k\in K_{rs}} \delta_{a,k} q^{rs}\int_{\beta_{k-1}^{rs}}^{\beta_k^{rs}} h_{rs}(z)dz, \quad \forall a \in A, \tag{6c}$$

$$q^{rs} \geq 0, \quad \forall rs \in RS, \tag{6d}$$

where $\mathbf{q} = \{q^{rs}\}$ is a vector of O-D travel demand, $\boldsymbol{\beta} = \{\beta_k^{rs}\}$ is a vector of boundaries, $D_{rs}^{-1}(\cdot)$ is the inverse of the demand function associated with O-D pair rs.

For objective function (6a), the Beckman transformation term and the Continuous monetary expense term relate to the path travel time and toll incurred by choosing a path, respectively. Together, as we shall see, they aim to minimize the generalized travel time for each user class. While the inverse demand integral term reflects the elasticity of O–D demand associated with network level of service. Constraint (6b) corresponds to the relationship between adjacent boundaries by Eq. (3). Constraint (6c) ensures that the link flow x_a is the sum of all path flows traversing link a. Non-negativity constraint on the O–D demand is imposed in Eq. (6d).

Although the solution variables (i.e., $\mathbf{q}$ and $\boldsymbol{\beta}$) of the above formulation are not commonly seen as path flows in many traffic assignment problems, path flows can be readily obtained by calculating definite integrals. Therefore, MP (6) still belongs to a path-based formulation but eliminates the necessity to map path flows to TEM boundaries and O-D travel demand. Indeed, the above formulation equips with the simple box and non-negative constraints, i.e., Eqs. (6b) and (6d), can remove traditional flow conservation constraints as in Leurent (1993). We proceed to formally prove the equivalence between the proposed formulation with that of Leurent.

Proposition 1. *MP (6) is equivalent to that of path-based formulation in Leurent (1993).*

Proof. See Appendix A. □

Next, we show an optimal solution must coincide with the ED-CBiTA UE condition.

Theorem 1. *A solution $(\mathbf{q},\boldsymbol{\beta})$ to MP (6), along with its corresponding path flows $\mathbf{f}$, satisfies the ED-CBiTA UE conditions specified in Definition 2.*

Proof. The Lagrangian function associated with MP (6) can be written as follows:

$$L(\mathbf{q},\boldsymbol{\beta},\boldsymbol{\mu}) = Z_1(\mathbf{q},\boldsymbol{\beta}) - \sum_{rs}\sum_{k}\mu_k^{rs}(\beta_k^{rs} - \beta_{k-1}^{rs}), \tag{7}$$

where μ_k^{rs} is the multiplier associated with Constraint (6b).

The Karush–Kuhn–Tucker (KKT) conditions for this program are:

$$\frac{\partial L}{\partial q^{rs}} \geq 0, \quad \forall rs \in RS, \tag{8a}$$

$$q^{rs}\frac{\partial L}{\partial q^{rs}} = 0, \quad \forall rs \in RS, \tag{8b}$$

$$\frac{\partial L}{\partial \beta_k^{rs}} = 0, \quad \forall rs \in RS, k = 1, \ldots, |K_{rs}|, \tag{8c}$$

$$\mu_k^{rs} \geq 0,\ \mu_k^{rs}(\beta_k^{rs} - \beta_{k-1}^{rs}) = 0, \quad \forall rs \in RS, k = 1, \ldots, |K_{rs}|, \tag{8d}$$

where

$$\begin{aligned}\frac{\partial L}{\partial q^{rs}} &= \sum_{k\in K_{rs}}\sum_{a\in A}\delta_{a,k}t_a\int_{\beta_{k-1}^{rs}}^{\beta_k^{rs}} h_{rs}(z)dz + \sum_{k\in K_{rs}}\tau_k\int_{\beta_{k-1}^{rs}}^{\beta_k^{rs}} z\cdot h_{rs}(z)dz - D_{rs}^{-1}(q^{rs}) \\ &= \sum_{k\in K_{rs}}\int_{\beta_{k-1}^{rs}}^{\beta_k^{rs}} [c_k + z\tau_k]\cdot h_{rs}(z)dz - D_{rs}^{-1}(q^{rs}),\end{aligned} \tag{9a}$$

$$\frac{\partial L}{\partial \beta_k^{rs}} = q^{rs}\cdot h_{rs}(\beta_k^{rs})\cdot[c_k + \beta_k^{rs}\tau_k - c_{k+1} - \beta_k^{rs}\tau_{k+1}] - (\mu_k^{rs} - \mu_{k+1}^{rs}). \tag{9b}$$

Consequently, Eqs. (8a) and (8b) can be rewritten as

$$T_{rs} - D_{rs}^{-1}(q^{rs}) \geq 0, \text{ and } q^{rs}\left(T_{rs} - D_{rs}^{-1}(q^{rs})\right) = 0, \tag{10}$$

which ensures a positive O–D demand satisfying Condition (i) in Definition 2.

In case of zero O-D demand (i.e., $q^{rs} = 0$), Eqs. (8c)-(8d) become trivial since they can be directly removed corresponding to this O-D pair. Otherwise, it has been shown that Eqs. (8c)-(8d) by fixing demand q^{rs} yields (Calderone et al., 2017; Xie et al., 2021):

$$C_k(\beta) \leq C_{k'}(\beta), \quad \forall \beta \in \Phi_k^{rs}, k' \neq k, k' \in K_{rs}, \tag{11}$$

which ensures the UE Condition (ii) in Definition 2. To see this, consider two adjacent used paths k and $k+l$ such that $\beta_{k-1}^{rs} < \beta_k^{rs} < \beta_{k+l}^{rs}$ (i.e., $\beta_k^{rs} = \beta_{k+1}^{rs} = \cdots = \beta_{k+l-1}^{rs}$ and $\mu_k^{rs} = \mu_{k+l}^{rs} = 0$) and $\tau_k \geq \tau_{k+l}$, summarizing Eq. (8c) from k to $k+l$ yields

$$c_k + \beta_k^{rs}\tau_k - c_{k+l} - \beta_k^{rs}\tau_{k+l} = 0. \tag{12}$$

Therefore, paths k and $k+l$ either share the same position with equal path times and tolls or produce the TEM boundary along the efficient frontier, which reads:

$$\beta_k^{rs} = \frac{c_{k+l} - c_k}{\tau_k - \tau_{k+l}}. \tag{13}$$

This completes the proof. □

We proceed to discuss the solution existence and uniqueness results concerning MP (6). Because the objective function Z_1 is continuous w.r.t the solution variables, and the feasible set is nonempty, bounded and compact, there must exists at least one optimal solution. As shown in Lemma 1, MP (6) is equivalent with that of path-based convex formulation in Leurent (1993), implying that Z_1 has a global minimum once the ED-CBiTA UE condition is satisfied. Below, we show the uniqueness w.r.t $\mathbf{q}$, $\mathbf{x}$ and $\boldsymbol{\beta}$ under certain conditions.

Theorem 2. *Under Assumptions 1-4, MP (6) has unique link flows* $\mathbf{x}^*$ *and O-D demands* $\mathbf{q}*$.

Proof. It is straightforward to show that MP (6) is strictly convex w.r.t $\mathbf{q}$ and $\mathbf{x}$, which guarantees the uniqueness of total link flows and O-D demands. □

Theorem 3. *Under Assumptions 1-4, if every path belonging to the same O-D pair rs has a unique toll, i.e.,* $\tau_k \neq \tau_{k'}, \forall k, k' \in K_{rs}$, *then MP (6) has unique TEM boudaries* $\{\beta_k^{rs*}, k = 1, \ldots, |K_{rs}|\}$ *corresponding to O-D pair rs.*

Proof. Given unique link flows $\mathbf{x}^*$ by Theorem 2, we first note that path travel times $\{c_k^*\}$ is also unique. For O-D pair rs, the efficient frontier with unique c_k and τ_k shown in Figure 1 is also unique, resulting in unique optimal boundaries $\{\beta_k^{rs*}\}$ corresponding to rs. □

The result of Theorem 3 further implies a unique path flow pattern $\mathbf{f}^*$ if a set of paths for any O-D pair is differentiated from each other in terms of the toll cost (see also Theorem 3 in Leurent, 1993). However, such a strong condition may be hard to satisfy in real-world networks.

3 Excess demand reformulation

This section reformulates MP (6) by replacing the O-D demand with excess demand, which is represented by a modified network and further compared with the discrete multi-class problem. Recall that the travel demand determined by the demand function is consistent with the network level of service via EGTT for each O–D pair, as illustrated in Figure 2. Given the O-D demand vector $\mathbf{q}$, EGTT is the integral of the shaded area associated with equilibrium boundaries across all efficient paths (see Figure 2(a)). Meanwhile, EGTT is equal to the inverse of the demand function (see Figure 2(b)). Let $\bar{q}^{rs}$ be the upper bound demand between O–D pair rs, $e^{rs} = \bar{q}^{rs} - q^{rs}$ is interpreted as the excess demand. Then, the inverse demand integrals term in Eq. (6a) can be rewritten as follows:

$$\begin{aligned} -\sum_{rs\in \mathrm{RS}} \int_0^{q^{rs}} D_{rs}^{-1}(z)dz &= -\sum_{rs\in \mathrm{RS}} \int_0^{\bar{q}^{rs}} D_{rs}^{-1}(z)dz + \sum_{rs\in \mathrm{RS}} \int_{q^{rs}}^{\bar{q}^{rs}} D_{rs}^{-1}(z)dz \\ &= -\sum_{rs\in \mathrm{RS}} \int_0^{\bar{q}^{rs}} D_{rs}^{-1}(z)dz + \sum_{rs\in RS} \int_0^{e^{rs}} W_{rs}(z)dz, \end{aligned} \tag{14}$$

where the first term on the right-hand side of Eq. (14) is a constant, and $W_{rs}(z) = D_{rs}^{-1}(\bar{q}^{rs} - z)$ is defined as the excess demand function for O-D pair rs.

Using the excess demand variable, MP (6) can be reformulated as follows:

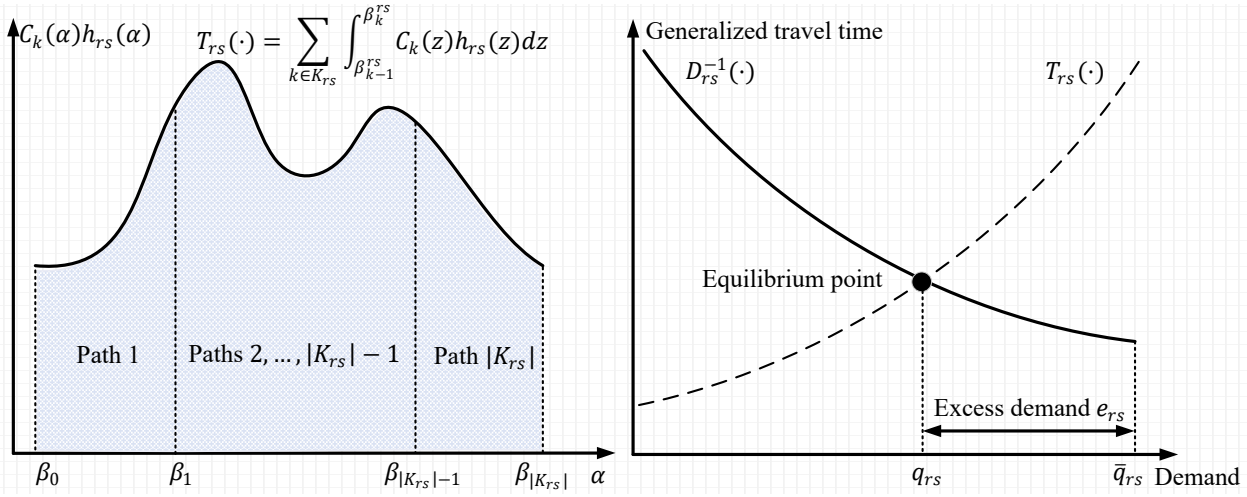
$$\min_{\mathbf{e},\boldsymbol{\beta}} Z_2 = \underbrace{\sum_{a\in A}\int_0^{x_a} t_a(z)dz}_{\text{Beckman transformation}} + \underbrace{\sum_{rs\in RS}\sum_{k\in K_{rs}} (\bar{q}^{rs}-e^{rs})\tau_k \int_{\beta_{k-1}^{rs}}^{\beta_k^{rs}} z\cdot h_{rs}(z)dz}_{\text{Continuous monetary expense}} + \underbrace{\sum_{rs\in RS}\int_0^{e^{rs}} W_{rs}(z)dz}_{\text{Excess demand integrals}} \quad (15a)$$

$$s.t. \quad \beta_k^{rs} - \beta_{k-1}^{rs} \geq 0, \quad \forall rs \in RS, k = 1,\dots,|K_{rs}|, \quad (15b)$$

$$x_a = \sum_{rs\in RS}\sum_{k\in K_{rs}} \delta_{a,k}(\bar{q}^{rs}-e^{rs}) \int_{\beta_{k-1}^{rs}}^{\beta_k^{rs}} h_{rs}(z)dz, \quad \forall a \in A, \quad (15c)$$

$$0 \leq e^{rs} \leq \bar{q}^{rs}, \quad \forall rs \in RS, \quad (15d)$$

where $\mathbf{e} = \{e^{rs}\}$ is a vector of the excess demand of O-D pairs.



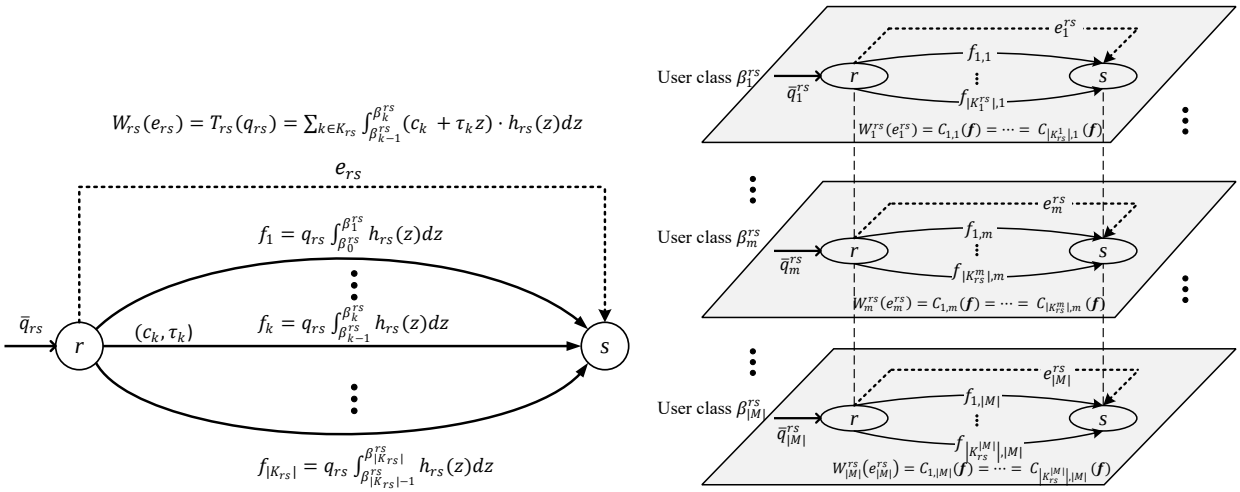
(a) Expected generalized travel time (b) Equilibrium point of elastic demand

Figure 2: Illustration of demand and supply equilibrium for ED-CBiTA.

The above excess demand reformulation brings a modification of network representation involving an infinite number of user classes, as shown in Figure 3(a). An excess demand link (dashed) directly connecting origin r and destination s is created to represent the excess demand flow e^{rs} with a link performance function $W_{rs}(\cdot)$. The equilibrium is achieved when $W_{rs} = T_{rs}$ and the latter is obtained with the optimal TEM boundaries along the efficient frontier.

Other than the continuous model, a more commonly used approach is the discrete multi-class model, in which the entire TEM (or VOT) range is divided into several predetermined intervals. However, it implicitly assumes homogeneity among each class m, which is usually represented by a class-specific TEM value β_m^{rs}. For completeness, the mathematical formulation of the discrete multi-class excess demand problem is presented in Appendix B. Let K_m^{rs} denote the set of used paths for class m between O-D pair rs, Figure 3(b) illustrates the excess-demand representation for the discrete multi-class problem. Each layer stores class-based flow information, and an excess demand link is added to each class associated with an excess demand function $W_m^{rs}(\cdot)$. Through

the discretization method, the number of O-D pairs potentially grows one-fold for one additional divided class for each O-D pair.



(a) Continuous multi-class problem (15)

(b) Discrete multi-class problem (36)

Figure 3: Excess demand network representation for O–D pair *rs*.

Nevertheless, the convenience offered under discretization is that MP (36) can be solved as a fixed demand multi-class problem using existing efficient solution algorithms, such as the modified path-based gradient projection (GP) algorithm suggested by Ryu et al. (2014). As for the ED-CBiTA Problem (15) illustrated in Figure 3(a), individuals are distinguished using a continuous distribution and considered as a whole. As a result, travel demand is endogenously determined by the TEM boundaries across all efficient paths among all users, which in turn are determined by the travel demand. There is still a lack of efficient algorithms that can be directly applied to solve this complex traffic assignment problem involving elastic demand, bi-criteria path cost structure and random TEMs, which motivates us to fill this gap.

4 TSGP algorithm

This section presents a two-stage gradient projection (TSGP) algorithm to solve the ED-CBiTA Problem (15), as depicted in Figure 4. In the spirit of the Gauss–Seidel method, TSGP is designed based on two decomposition schemes: (i) the first stage (called demand equilibration) decomposes the original Problem (15) into an O-D excess demand subproblem; and (ii) the second stage (called boundary equilibration) decomposes the original Problem (15) into a series of single-boundary adjustment problems. Both two types of subproblems can be solved very efficiently using GP so as to satisfy the KKT conditions (10)-(11) for each stage. As we shall see, path flows are accordingly adjusted in vertical and horizontal directions in the first and second stages, respectively.

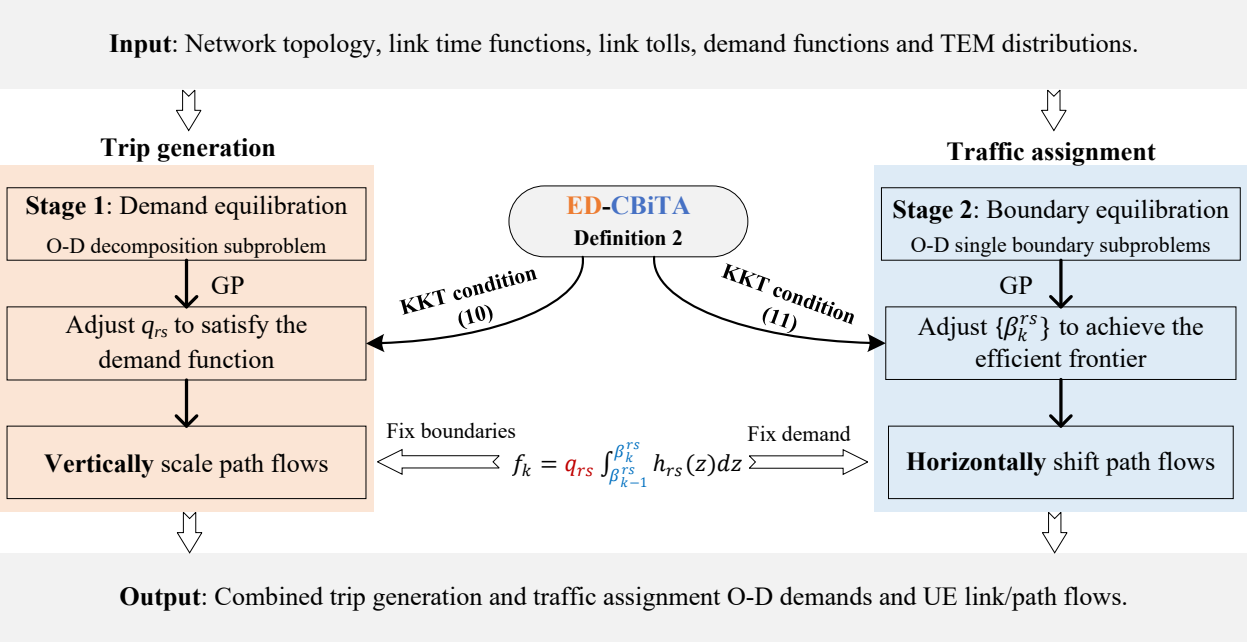


Figure 4: Two-stage gradient projection algorithm framework in solving the ED-CBiTA problem.

In the following, Sections 4.1 and 4.2 describe the GP algorithm for the first and second stage, respectively. Finally, the full solution procedure and the convergence proof are presented in Section 4.3.

4.1 First stage: Demand equilibration

This stage aims to simultaneously adjust the O-D demand and path flows based on the current network congestion level. The idea is to solve a single O-D based excess demand subproblem, while holding the excess demand associated with other O-D pairs and boundaries constant. Given an O-D pair (we drop the subscript rs for convenience), the one-dimensional excess demand subproblem with respect to e can be expressed as follows:

$$\min \quad Z_3(e) = \sum_{a\in A} \int_0^{x_a} t_a(z)dz + \sum_{k\in K_{rs}} (\bar{q} - e) \cdot \tau_k \int_{\tilde{\beta}_{k-1}}^{\tilde{\beta}_k} z \cdot h(z)dz + \int_0^e W(z)dz \tag{16a}$$

$$s.t. \quad 0 \le e \le \bar{q}, \tag{16b}$$

$$x_a = \tilde{x}_a + \sum_{k\in K_{rs}} \delta_{a,k}(\bar{q} - e) \int_{\tilde{\beta}_{k-1}}^{\tilde{\beta}_k} h(z)dz, \quad \forall a \in A, \tag{16c}$$

where $\tilde{x}_a$ is the constant link flow contributed by O-D pairs other than rs, $\tilde{\beta}_k$ is the constant boundary.

Given current solution $e(n)$ at iteration n, the first- and second-order derivatives of Z_3 with

respect to e are, respectively, determined by

$$\frac{\partial Z_3}{\partial e} = W[e(n)] - \sum_{k\in K_{rs}} \int_{\tilde{\beta}_{k-1}}^{\tilde{\beta}_k} [c_k + z\tau_k] \cdot h(z)dz = W[e(n)] - T(n), \tag{17}$$

$$\begin{aligned}\frac{\partial^2 Z_3}{(\partial e)^2} = S(n) &= \frac{\partial W[e(n)]}{\partial e} - \frac{\partial \sum_{k\in K_{rs}} \int_{\tilde{\beta}_{k-1}}^{\tilde{\beta}_k} [c_k + z\tau_k] \cdot h(z)dz}{\partial e} \\ &= \frac{\partial W[e(n)]}{\partial e} + \sum_{k\in K_{rs}} \left[\int_{\tilde{\beta}_{k-1}}^{\tilde{\beta}_k} h(z)dz \cdot \sum_{a\in A} \delta_{a,k} \frac{\partial t_a(x_a)}{\partial x_a} \left(\sum_{k'\in K_{rs}} \delta_{a,k'} \int_{\tilde{\beta}_{k'-1}}^{\tilde{\beta}_{k'}} h(z)dz \right) \right]. \end{aligned} \tag{18}$$

To solve the problem, we apply gradient projection (GP) to update excess demand solution:

$$e(n+1) = \begin{cases} \min\left(\bar{q}, e(n) - \mu \dfrac{W[e(n)] - T(n)}{S(n)}\right) & \text{if } T(n) > W[e(n)], \\ \max\left(0, e(n) - \mu \dfrac{W[e(n)] - T(n)}{S(n)}\right) & \text{if } T(n) < W[e(n)], \end{cases} \tag{19}$$

where the step size μ is determined by performing a line search.

Or equivalently, the O-D demand $q = \bar{q} - e$ can be updated as follows

$$q(n+1) = \begin{cases} \min\left(0, q(n) - \mu \dfrac{T(n) - D^{-1}[q(n)]}{S(n)}\right) & \text{if } T(n) > D^{-1}[q(n)], \\ \max\left(\bar{q}, q(n) - \mu \dfrac{T(n) - D^{-1}[q(n)]}{S(n)}\right) & \text{if } T(n) < D^{-1}[q(n)], \end{cases} \tag{20}$$

Remark 1. *Subproblem (16) is strictly convex because its Hessian is positive definite (i.e., $S(n) > 0$).*

Remark 2. *Many stepsize determination schemes can be adopted to help convergence, e.g., self-regulated averaging scheme (Liu et al., 2009), self-adaptive Armijo scheme (Chen et al., 2013). In our implementation, a unit stepsize is used because the second-order derivative information serves an automatic scaling (e.g., see Bertsekas et al., 1984; Jayakrishnan et al., 1994; Chen et al., 2002; Ryu et al., 2014).*

Further, flows on all paths corresponding to this O-D pair can be adjusted accordingly, i.e.,

$$f_k(n+1) = f_k(n) + [q(n+1) - q(n)] \cdot \int_{\tilde{\beta}_{k-1}}^{\tilde{\beta}_k} h(z)dz, \quad \forall k \in K_{rs}. \tag{21}$$

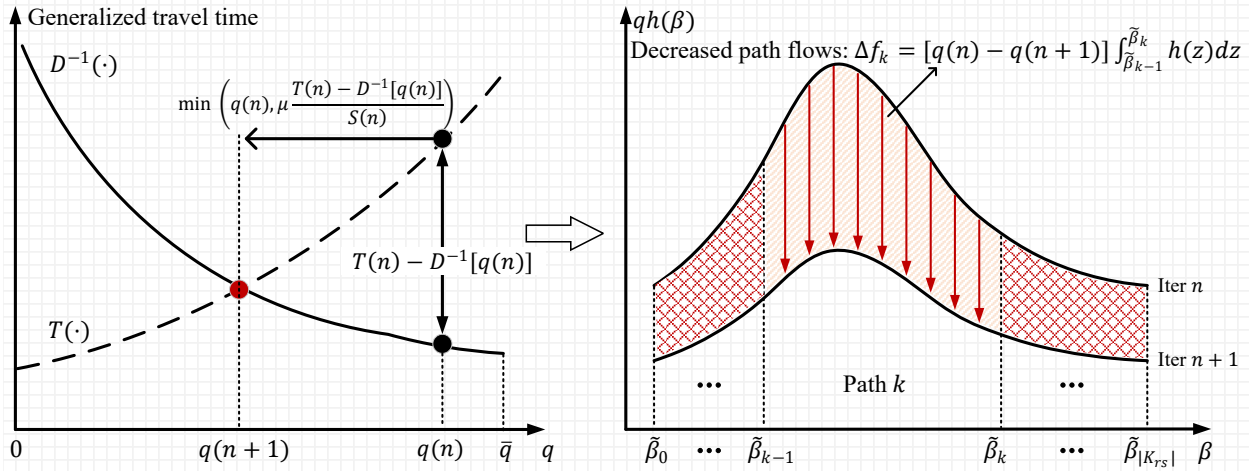
As shown in Figure 5, we provide a graphical illustration for the flow adjustments using Eqs. (20)-(21). The purpose of this illustration is to visually demonstrate the adjustments made to the flow distribution in response to varying congestion levels.

Specifically, if $T(n) > D^{-1}[q(n)]$ (see Figure 5(a)), i.e., the expected congestion level by EGTT is higher than the O-D cost at iteration n, then the demand of that O–D pair will be decreased. Accordingly, the flow on each path is "vertically" decreased due to the reduced demand. If $T(n) < D^{-1}[q(n)]$ (see Figure 5(b)), i.e., the expected congestion level (i.e., EGTT) is lower than the O-D cost at iteration n, then the O-D demand will be increased and the flow on each path is

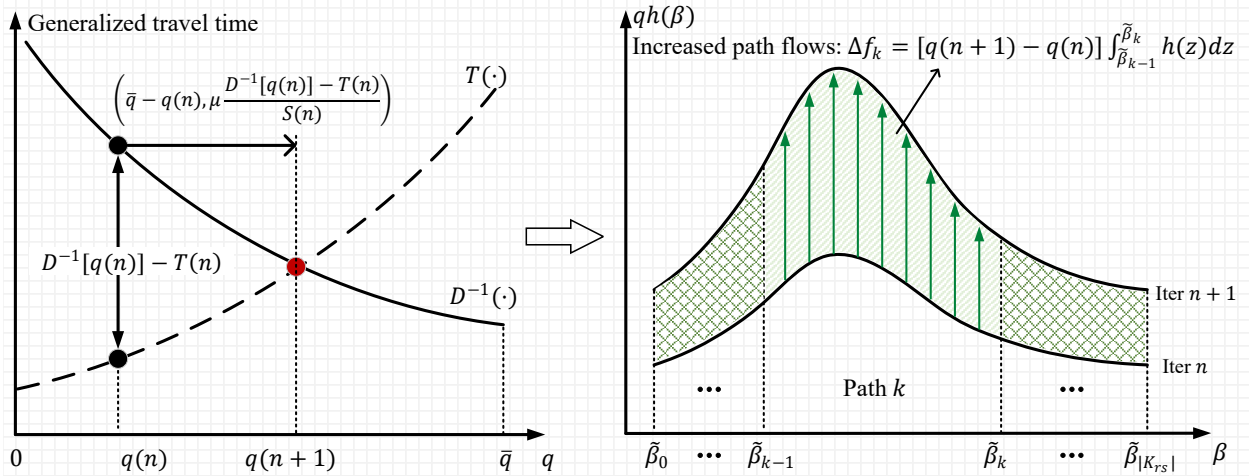
"vertically" enlarged. Consequently, both cases attempt to achieve demand equilibration based on the current congestion level. The detailed first stage GP algorithm for solving a single O-D pair subproblem problem is reported in Algorithm 1.

Algorithm 1 Gradient projection algorithm for a single O-D pair subproblem

1: **Input:** TEM boundaries Φ^{rs} and excess demand $e(n)$ for O-D pair rs.
2: Compute $W[e(n)] - T(n)$ and $S(n)$ according to Eqs. (17) and (18), respectively.
3: Update $q(n+1)$ according to Eq. (20).
4: **for** each path $k \in K_{rs}$ **do**
5: Compute the path flow change $\Delta = [q(n+1) - q(n)] \cdot \int_{\tilde{\beta}_{k-1}}^{\tilde{\beta}_k} h(z)dz$.
6: Update the path flow $f_k^{n+1} = f_k^n + \Delta$.
7: Update $x_a = x_a + \Delta$, $t_a(x_a)$ and $\frac{dt_a(x_a)}{dx_a}$ of each link a on path k.
8: **end for**
9: **Output:** Updated e, q, $\mathbf{f}$ and $\mathbf{x}$.



(a) Decrease demand and path flows if $T(n) > D^{-1}(q^n)$



(b) Increased demand and path flows if $T(n) < D^{-1}(q^n)$

Figure 5: Graphical illustration of O-D demand and path flow adjustments at iteration n.

4.2 Second stage: Boundary equilibration

To simultaneously adjust TEM boundaries and path flows, the original MP (15) is decomposed with respect to O-D based TEM boundaries, while holding other boundaries and O-D travel demand constant. Noteworthy this decomposition scheme was recently studied by Xie et al. (2021) for fixed demand CBiTA. In our study, we provide visual demonstrations illustrating the horizontal adjustments in path flow to align with the slope observed on the Pareto frontier.

Given an O-D pair (we drop the subscript rs for convenience), the one-dimensional single-boundary subproblem with respect to β_k can be expressed as follows:

$$\min Z_4(\beta_k) = \sum_{a\in A}\int_0^{x_a} t_a(z)dz + (\bar{q}-\tilde{e})\cdot\left(\tau_k\int_{\tilde{\beta}_{k-1}}^{\beta_k} z\cdot h(z)dz + \tau_{k+1}\int_{\beta_k}^{\tilde{\beta}_{k+1}} z\cdot h(z)dz\right) \tag{22a}$$

$$s.t. \quad \tilde{\beta}_{k-1} \le \beta_k \le \tilde{\beta}_{k+1}, \tag{22b}$$

$$x_a = \tilde{x}_a + (\bar{q}-\tilde{e})\left(\delta_{a,k}\int_{\tilde{\beta}_{k-1}}^{\beta_k} h(z)dz + \delta_{a,k+1}\int_{\beta_k}^{\tilde{\beta}_{k+1}} h(z)dz\right), \quad \forall a\in A, \tag{22c}$$

where $\tilde{x}_a$ is the constant link flow contributed by paths other than paths k and $k+1$, $\tilde{e}$ is treated as the constant excess demand at current iteration, $\tilde{\beta}_{k-1}$ and $\tilde{\beta}_{k+1}$ are treated as the constant lower and upper bound of β_k, respectively.

Given the current solution $\beta_k(n)$ at iteration n, the first- and second-order derivatives of Z_4 w.r.t β_k are, respectively, determined by

$$g_k = \frac{\partial Z_3}{\partial \beta_k} = (\bar{q}-\tilde{e})\cdot h(\beta_k^n)\cdot\left[C_k(\beta_k(n)) - C_{k+1}(\beta_k(n))\right], \tag{23}$$

$$G_k = \frac{\partial^2 Z_3}{(\partial\beta_k)^2} = \{(\bar{q}-\tilde{e})\cdot h[\beta_k(n)]\}^2\sum_{a\in A}(\delta_{a,k}-\delta_{a,k+1})^2\cdot\frac{dt_a(x_a)}{dx_a} + (\bar{q}-\tilde{e})\cdot h[\beta_k(n)]\cdot(\tau_k-\tau_{k+1})$$
$$+ (\bar{q}^{rs}-e_{rs}^n)\cdot h'(\beta_k^n)\cdot\left[C_k(\beta_k^n) - C_{k+1}(\beta_k^n)\right]. \tag{24}$$

Remark 3. *Subproblem (22) is strictly pseudoconvex because $G_k > 0$ if $g_k = 0$ (Cambini and Martein, 2008). Although the Hessian G_k could be possibly negative by the third term in Eq. (24), we can still prove it is positive at the stationary point. To see this, $g_k = 0$ implies that $C_k(\beta_k(n)) - C_{k+1}(\beta_k(n)) = 0$, leading to the third term of G_k be zero. In Eq. (24), because the first term is strictly positive and the second term is non-negative, we must have $G_k > 0$. Note that the property of strict pseudoconvexity ensures the subproblem is still uni-modal and has a strictly global minimum.*

To solve subproblem (22), GP is also adopted with a modified Hessian to obtain a new solution $\beta_k(n+1)$ as follows:

$$\beta_k(n+1) = \begin{cases} \min\left(\tilde{\beta}_{k+1}, \beta_k(n) - \mu\dfrac{g_k}{\tilde{G}_k}\right) & \text{if } g_k < 0, \\ \max\left(\tilde{\beta}_{k-1}, \beta_k(n) - \mu\dfrac{g_k}{\tilde{G}_k}\right) & \text{if } g_k > 0, \end{cases} \tag{25}$$

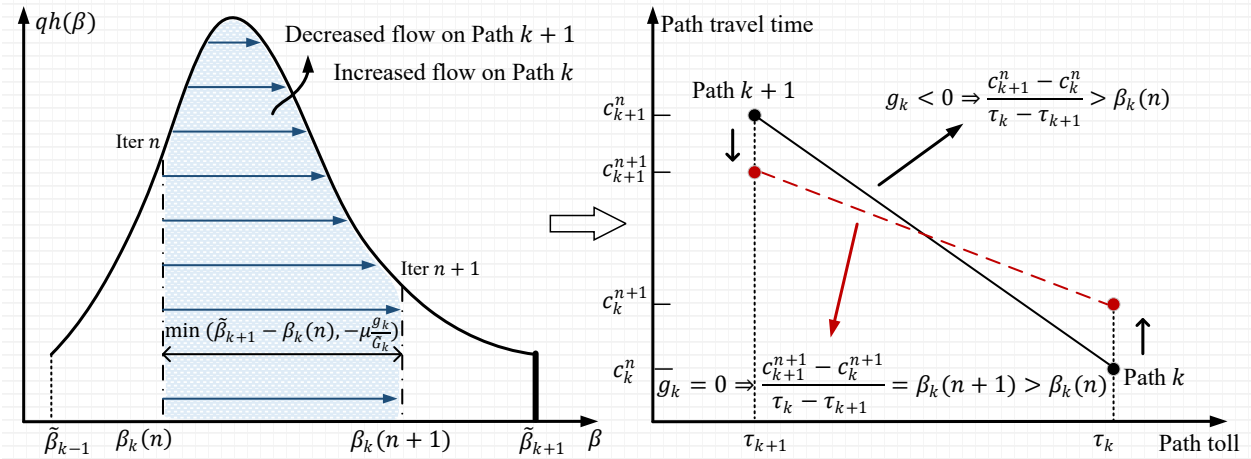
where

$$\tilde{G}_k = \{(\bar{q} - \tilde{e}) \cdot h[\beta_k(n)]\}^2 \sum_{a \in A} (\delta_{a,k} - \delta_{a,k+1})^2 \cdot \frac{dt_a(x_a)}{dx_a} + (\bar{q} - \tilde{e}) \cdot h[\beta_k(n)] \cdot (\tau_k - \tau_{k+1}). \tag{26}$$

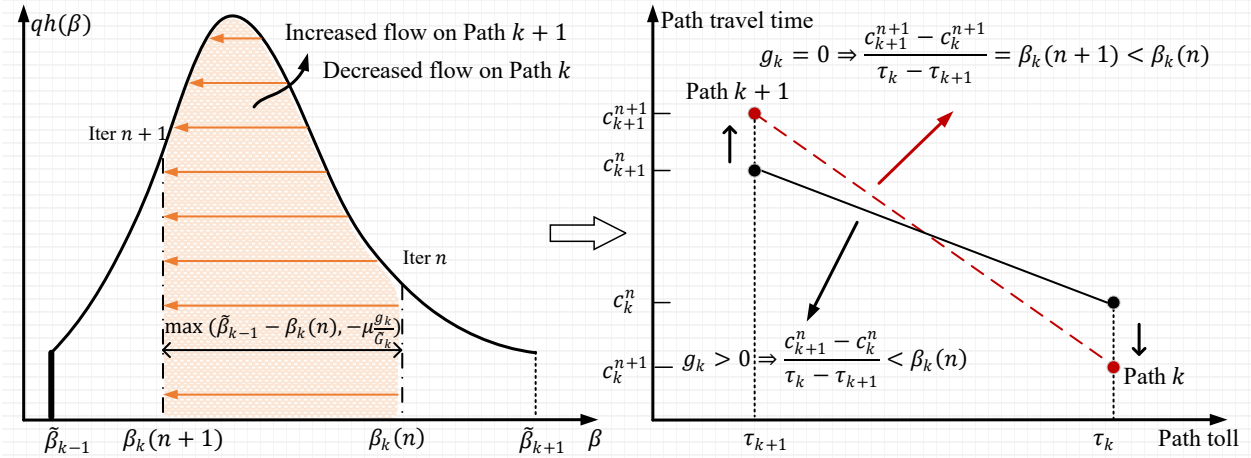
Note that $\tilde{G}_k$ is an approximation to G_k such that it only contains the first and second terms of the latter to ensure a positive scaling (see also in Xie et al., 2021). After updating the boundary, the flows on path k and $k+1$ can be adjusted accordingly, i.e.,

$$f_k(n+1) = f_k(n) + (\bar{q} - \tilde{e}) \cdot \int_{\beta_k(n)}^{\beta_k(n+1)} h(z)dz, \tag{27a}$$

$$f_{k+1}(n+1) = f_{k+1}(n) - (\bar{q} - \tilde{e}) \cdot \int_{\beta_k(n)}^{\beta_k(n+1)} h(z)dz. \tag{27b}$$



(a) Move forward if $g_k < 0$



(b) Move backward if $g_k > 0$

Figure 6: Graphical illustration of boundary and path flow adjustments at iteration n.

Figure 6 provides a graphical illustration for flow adjustments using Eqs. (25)-(27). That is, solving subproblem (22) is to horizontally move β_k forward or backward and thus adjust the flows between path k and $k+1$. In a more general case, let us consider paths k and $k+1$ have different tolls, i.e., $\tau_k > \tau_{k+1}$. When $g_k < 0$ (see Figure 6(a)), the boundary $\beta_k(n)$ would move forward by $-\mu g_k / \tilde{G}_k$ with a maximum of $\tilde{\beta}_{k+1} - \beta_k(n)$. Consequently, f_k is increased while f_{k+1} is decreased, causing c_k to become larger while c_{k+1} to become smaller. Once the subproblem (22) is optimally solved, the slope $\frac{c_{k+1}^{n+1} - c_k^{n+1}}{\tau_k - \tau_{k+1}}$ drops to equal to $\beta_k(n+1)$, which would coincidence with the slope of the Pareto frontier between two paths (see KKT condition (12)). On the other hand, as shown in Figure 6(b), $\beta_k(n)$ moves backward such that traffic flows are shifting from path k to $k+1$ when $g_k > 0$.

Algorithm 2 describes the procedure for boundary equilibration in detail. Note that we drop the unused path (i.e., k and $k+1$) as necessary to ensure that traffic flows can be effectively adjusted between adjacent paths (cf. Lines 10-19). That is, we drop path k when it has zero flows (i.e., $f_k = 0$) and it does not have the incentive to receive flows from path $k+1$ (i.e., $\Delta = 0$). To remove path $k+1$ as necessary (cf. Lines 14-19), we have to further check whether the objective function value can be decreased if path $k+1$ also has no incentive to receive flows from path $k+2$ (i.e., $g_{k+1} > 0$) in latter boundary adjustments.

Algorithm 2 Gradient projection algorithm for a single boundary subproblem

1: **Input:** TEM boundaries Φ^{rs} and excess demand $\tilde{e}$ for O-D pair rs.
2: Initialize $k = 1$ and a zero-flow path counter $z = 0$.
3: **while** $k \leq |K_{rs}| - 1$ **do**
4: Compute g_k and G_k according to Eqs. (23) and (24), respectively.
5: Update boundary $\beta_k(n+1)$ given $\beta_k(n)$ according to Eq. (25).
6: Compute the path flow change $\Delta = (\bar{q} - \tilde{e}) \cdot \int_{\beta_k(n)}^{\beta_k(n+1)} h(z)dz$.
7: Adjust the path flow $f_k = f_k + \Delta$, $f_{k+1} = f_{k+1} - \Delta$.
8: Update $x_a = x_a + \Delta$, $t_a(x_a)$ and $\frac{dt_a(x_a)}{dx_a}$ of each link a on path k.
9: Update $x_a = x_a - \Delta$, $t_a(x_a)$ and $\frac{dt_a(x_a)}{dx_a}$ of each link a on path $k+1$.
10: **if** $f_k = 0$ and $\Delta = 0$ **then** ▷ Drop path k as necessary.
11: Remove path k from K_{rs}.
12: Set $z = z + 1$.
13: **end if**
14: **if** $f_{k+1} = 0$ and $\Delta = 0$ **then** ▷ Drop path $k+1$ as necessary.
15: **if** $\nexists$ path $k+2 \in K_{rs}$ or ($\exists$ path $k+2 \in K_{rs}$ and $g_{k+1} > 0$) **then**
16: Remove path $k+1$ from K_{rs}.
17: Set $z = z + 1$.
18: **end if**
19: **end if**
20: Set $k = k + 1 - z$ and reset $z = 0$.
21: **end while**
22: **Output:** Updated boundaries Φ^{rs} and path flows $\mathbf{f}$.

4.3 Full solution procedure

For completeness, Algorithm 3 presents the full solution procedure of TSGP, which mainly consists of four steps. Several remarks corresponding to each step are in order as below.

- In Step 1 (cf. Lines 2-10), the algorithm starts to assign an initial demand to the network by calling the Dial's parametric shortest path tree algorithm (Dial, 1996, 1997), in which all efficient paths and associated TEM intervals with the same origin can be simultaneously obtained.

- In Step 2 (cf. Lines 12-20), the column generation generates new potential efficient paths and adds them into the working path set K_{rs} based on the current flow pattern.

- Step 3 (cf. Line 21) measures the overall convergence by the relative gap (RG) defined as follows:

$$RG = 1 - \frac{\sum_{rs \in RS} D_{rs}(\bar{T}_{rs}) \cdot \bar{T}_{rs}}{\sum_{rs \in RS} q^{rs} \cdot T_{rs}}, \tag{28}$$

  where $\bar{T}_{rs} = \sum_{k \in \bar{K}_{rs}} \int_{\bar{\beta}_{k-1}^{rs}}^{\bar{\beta}_k^{rs}} [c_k + z \cdot \tau_k] \cdot h(z) dz$ and T_{rs} are EGTT using path sets $\bar{K}_{rs}$ and working set K_{rs}, respectively. Notably, path set $\bar{K}_{rs}$ and TEM boundaries $\bar{\boldsymbol{\beta}} = \{\bar{\beta}_k^{rs}\}$ are generated by Step 2 (i.e., column generation).

- In Step 4 (cf. Lines 21-36), an adaptive inner loop is deigned to efficiently solve the restricted version of the original problem defined upon the working path set. The central idea intends to accelerate convergence by performing an adaptive equilibration for O-D pairs that are not fully equilibrated on two stages before naturally turning to column generation. One may witness this strategy has significantly accelerated various traffic assignment solvers (e.g., Chen and Jayakrishnan, 1998; Xu et al., 2020, 2022; Xie et al., 2021; Tan et al., 2022). Moreover, we drop the unused paths for each O-D pair to keep each TEM interval set Φ^{rs} compact (cf. Line 36).

  To implement the strategy that ensures each O-D pair can be solved with a similar precision, a convergence indicator to the first stage for O-D pair rs is defined as

$$\Delta_{rs}^{[1]} = \left| 1 - \frac{D_{rs}(T_{rs})}{\bar{q}^{rs} - e^{rs}} \right|, \tag{29}$$

  which is effectively a measure of the relative demand differences for O-D pair rs.

  As suggested by Xie et al. (2021), a convergence indicator to the second stage for O-D pair rs is defined as the maximum value of the first-order derivative, i.e.,

$$\Delta_{rs}^{[2]} == \max_{k=1,\ldots,|K_{rs}|-1} \{|g_k|\}. \tag{30}$$

Algorithm 3 Two-stage gradient projection (TSGP) algorithm for ED-CBiTA

1: **Input:** Network $G(N, A)$, the maximum inner loop iterator L, the convergence criteria ϵ, two control parameters Γ_1, Γ_2.
2: **Step 1. Initialization:** Update $\mathbf{t}$ based on zero-flow network. Set $\mathbf{e} = \mathbf{0}, \mathbf{q} = \bar{\mathbf{q}}$, $K_{rs} = \emptyset, \Phi^{rs} = \emptyset, \forall rs$.
3: **for** each origin r **do**
4: Call Dial's parametric shortest path tree algorithm for origin node r .
5: **for** each destination s **do**
6: Set $K_{rs} = \bar{K}_{rs}$, and $\Phi_k^{rs} = [\tilde{\beta}_{k-1}^{rs}, \tilde{\beta}_k^{rs})$ for all $k \in \bar{K}_{rs}$.
7: Assign flows to any path $k \in K_{rs}$ and links such that $\delta_{a,k} = 1$ according to Eq. (4).
8: Update $t_a(x_a)$ and $\frac{dt_a(x_a)}{dx_a}$ of each link a on path k.
9: **end for**
10: **end for**
11: **Step 2. Column generation:** Lines 12-20.
12: **for** each origin r **do**
13: Call Dial's parametric shortest path tree algorithm for origin node r .
14: **for** each destination s **do**
15: **for** each $k \in \bar{K}_{rs} \setminus K_{rs}$ **do**
16: Set $\Phi_k^{rs} = [\beta_n^{rs}, \beta_n^{rs})$ such that $n = \texttt{argmin}_{k'}\{\tau_{k'} \leq \tau_k | k' = 0, \ldots, |K_{rs}|\}$.
17: Add k into K_{rs} based on the ordering index n.
18: **end for**
19: **end for**
20: **end for**
21: **Step 3. Convergence test:** Computer RG according to Eq. (28). If $RG < \epsilon$ or the maximum run-time is achieved, then terminate the algorithm; otherwise go to Step 4.
22: **Step 4. Adaptive inner loop:** Lines 23-38.
23: Set a counter $l = 0$ and an indicator $\nu =$ TRUE for the next inner loop.
24: **while** $l < L$ and $\nu =$ TRUE **do**
25: Set $l = l + 1$ and $\nu =$ FALSE.
26: **for** each origin r **do**
27: **for** each destination s **do**
28: **if** $\Delta_{rs}^{[1]} < \Gamma_1 RG$ **then** ▷ *First stage*
29: Run Algorithm 1 for demand equilibration.
30: Set $\nu =$ TRUE if $\nu =$ FALSE.
31: **end if**
32: **if** $\Delta_{rs}^{[2]} < \Gamma_2 RG$ **then** ▷ *Second stage*
33: Run Algorithm 2 for boundary equilibration.
34: Set $\nu =$ TRUE if $\nu =$ FALSE.
35: **end if**
36: **end for**
37: **end for**
38: **end while**
39: Return to Step 2 (column generation, Line 11).
40: **Output:** $\mathbf{x}^*$, $\mathbf{q}^*$, $\mathbf{f}^*$ and Φ^{rs*}.

5 Numerical results

In this section, we use the proposed path-based TSGP algorithm and two link-based algorithms, MSA (Leurent, 1993) and a modified FW (Marcotte et al., 1996), to solve the ED-CBiTA problem. Note that FW conducts a line search at each iteration whereas MSA uses a diminishing stepsize. Implementation details of MSA and FW can be found in Appendix C and D, respectively. To see the results difference between the continuous problem (15) and the discrete problem (36) as shown in Figure 3, the latter is solved by the modified path-based GP algorithm of Ryu et al. (2014). All algorithms are implemented in the C++ programming language, and the implementations share codes wherever possible to ensure a fair comparison of the computational performance.

The remaining section is divided into two parts. The first part uses a two-link network to demonstrate the correctness of the formulation and the feature of the TSGP algorithm. The second part describes the evaluation of the algorithmic performance on large networks.

5.1 A two-link network

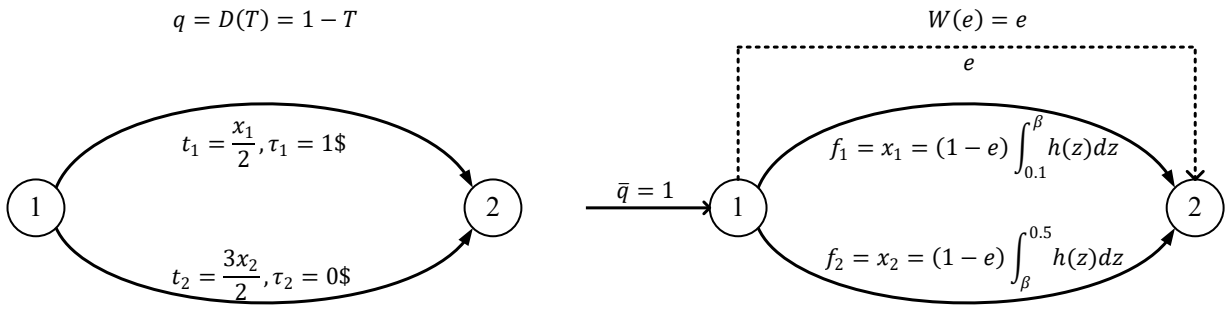
The numerical results of this two-link network are divided into four parts. Firstly, We present the excess demand formulation and check the correctness of the optimal solution. Secondly, we illustrate the features of different algorithms, including TSGP, MSA and FW. Next, we compare the flow pattern between the continuous and discrete multi-class problems. Finally, toll revenues using the discrete and continuous approach are further examined.

5.1.1 Analytical analysis on the formulation

The simple two-link network is illustrated in Figure 7(a), which has an upper bound demand of one unit. Note that the link index is also the path index since $\tau_1 > \tau_2$. Because there is only one O–D pair, the subscript rs is omitted for simplicity. We assume that TEM follows a uniform distribution $U(0.1, 0.5)$, thus the PDF $h(\beta) = \frac{1}{0.5-0.1} = 2.5$ and the CDF $H(\beta) = \frac{\beta-0.1}{0.5-0.1}$. Accordingly, the excess demand formulation shown in Figure 7(b) can be written as below:

$$\begin{aligned}
\min_{e,\beta} Z_5 = & \frac{x_1^2}{4} + \frac{3x_2^2}{4} + (1-e)\cdot\left(\frac{100\beta^2-1}{80}\right) + \frac{e^2}{2} \\
s.t. \quad & 0.1 \le \beta_1 \le 0.5, \\
& 0 \le e \le 1 \\
& x_1 = (1-e)\cdot(2.5\beta - 0.25), \\
& x_2 = (1-e)\cdot(1.25 - 2.5\beta).
\end{aligned}$$

The optimal solution to this excess demand formulation reads $\beta^* = 0.3, e^* = 0.4$, and accordingly, $x_1^* = x_2^* = 0.3$, $t_1^* = 0.15$, $t_2^* = 0.45$ and $Z_5^* = 0.23$. To see the correctness of the flow pattern that exactly satisfies the ED-CBiTA UE condition: (i) EGTT $T^* = \int_{0.1}^{0.3} 2.5 * (0.15 + z)dz + \int_{0.3}^{0.5} 2.5 *$



(a) Original network with elastic demand (b) Modified network representation with excess demand

Figure 7: A small network.

$0.45dz = 0.4 = W(e^*)$, which matches the excess demand function; and (ii) the slope between two efficient paths $= \frac{t_2^* - t_1^*}{\tau_1 - \tau_2} = \frac{0.45-0.15}{1-0} = 0.3 = \beta^*$, which is able to achieve exact positions along the Pareto frontier. Consequently, the optimal solution satisfies the ED-CBiTA UE Definition 2.

5.1.2 Illustrative comparison of different algorithms

Starting from an initial solution with $\beta = 0.45$ and $e = 0$, Figure 8 depicts three iterative trajectories using TSGP, MSA and FW. We can observe that TSGP can efficiently approach the optimal solution in almost two iterations. As for two link-based algorithms, FW begins to exhibit the well-known zigzagging effect when it approaches the optimal solution, whereas MSA uses a diminishing stepsize and needs a rather larger number of iterations to converge. In a nutshell, TSGP significantly outperforms MSA and FW in solving this two-link network.

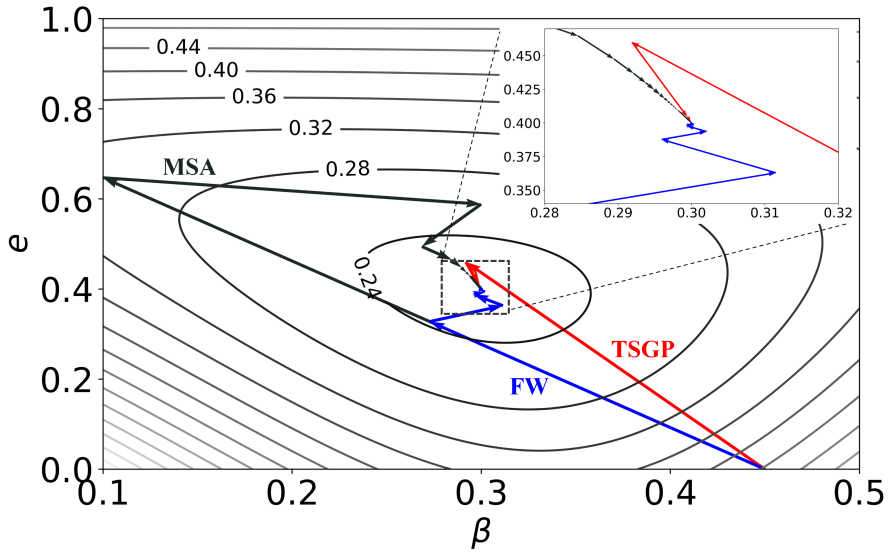


Figure 8: Iterative trajectories by three algorithm.

5.1.3 Flow pattern: Continuous versus discrete

In subsections 5.1.3 and 5.1.4, we use more general settings on the two-link network to compare the equilibrium results of the continuous and discrete multi-class model. The link tolls $\tau_1 = 5$ \$ and $\tau_2 = 0$. The link travel time function takes the form of the Bureau of Public Roads (BPR) function:

$$t_a = t_a^0 \left(1 + 0.15 \left(\frac{x_a}{\rho_a}\right)^4\right), \tag{32}$$

where t_a^0 denotes the free-flow travel time on link a, ρ_a is the capacity on link a. Herein, we set $t_1^0 = 10$ and $t_2^0 = 15$ min, $\rho_1 = 1000$ and $\rho_2 = 4000$ veh/h.

Following Ryu et al. (2014), the O–D demand function adopts the exponential functional form:

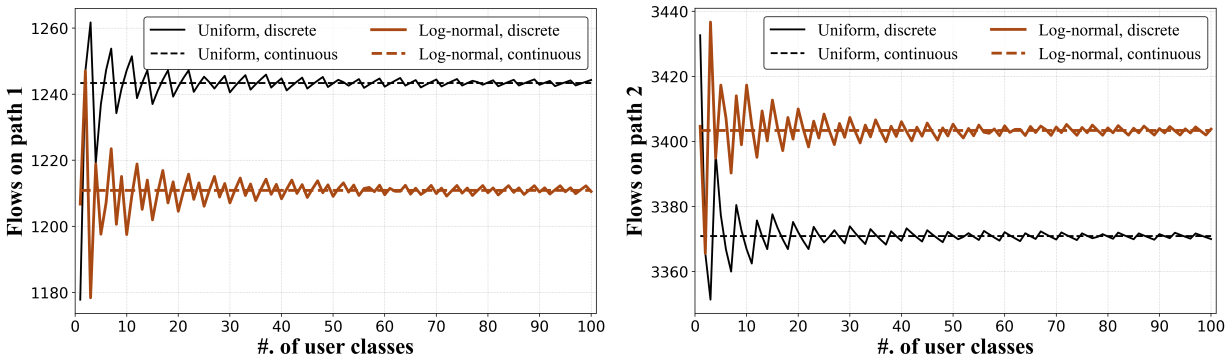
$$D_{rs}(\bar{T}_{rs}) = \bar{q}^{rs} \exp(-\gamma \bar{T}_{rs}), \tag{33}$$

where $\bar{T}_{rs}$ is the equilibrium EGTT (h), $\bar{q}^{rs}$ is the upper bound demand of O–D pair rs and γ is a positive scaling parameter. Herein, we set $\gamma = 0.3$ and $\bar{q}^{rs} = 5000$ veh/h.

We consider two types of continuous distributions of VOT, including uniform and log-normal distributions. The lower and upper bound of VOT are set as 6 and 30 \$/h respectively as suggested by Xie et al. (2021). Graphical illustrations of two distributions and the transformation between VOT (α) and TEM (β) are provided in Appendix E.

Next, we briefly describe how to set parameters and functions for a discrete multi-class problem with elastic demand. For a single-class problem ($|M| = 1$), the class-specific TEM value is set as the mean value of the continuous distribution, and the demand function is the same as the continuous problem. For a multi-class problem ($|M| > 1$), we evenly divided the TEM range $[1/30, 1/6]$ with a total of $|M|$ intervals, each taking the mean as the class-specific value. For the demand function, the upper bound demand of each class in Eq. (33) is scaled by its proportion such that the total O-D trips are comparable.

Figure 9 shows equilibrium flows on paths 1 and 2 for a different numbers of user classes ranging from 1 to 100. We can observe that the flow differences significantly fluctuate at the beginning (i.e., $|M| < 10$). After that, the fluctuation gradually reduces as the number of classes increases. When the continuous distribution is discretized with a large number of classes, e.g., more than $|M| = 50$, it would produce a closer solution to the continuous problem. Nevertheless, we can still observe the flow differences compared to the continuous solution.

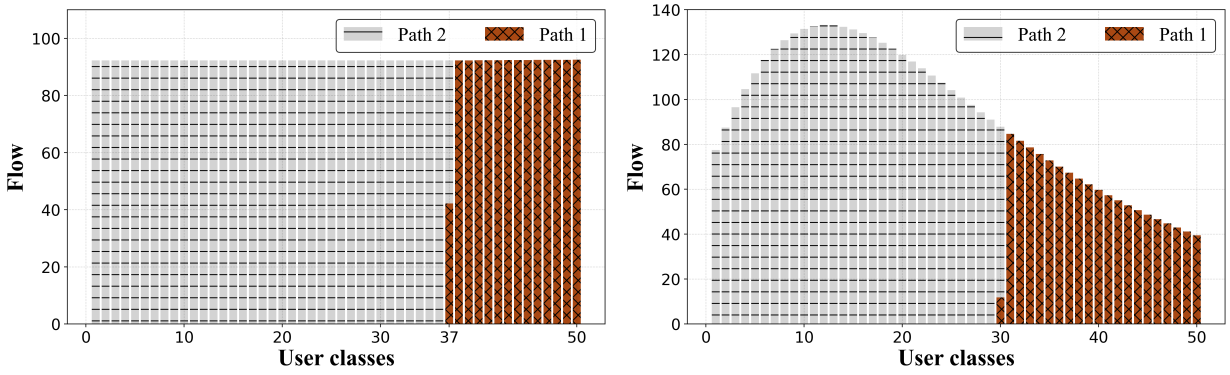


(a) Flows on path 1

(b) Flows on path 2

Figure 9: Equilibrium path flows for different number of user classes ranging from 1 to 100.

To see why the discrete problem fails to achieve the exact flow pattern with that of the continuous problem, Figure 10(a) and 10(b) plots the class-specific equilibrium path flows with $|M| = 50$ for uniform and log-normal distributions, respectively. All 50 groups are ordered in increasing order of VOT on x-axis. We found that the preferred path order by different VOT groups has been achieved through the discretization method. However, Group 37 for uniform distribution and Group 30 for log-normal distribution both carry positive flows on two paths, respectively. That is, those travelers grouped into the same class can choose different paths. The homogeneity among these two groups will bring discretization errors compared to the continuous problem.



(a) Uniform distribution

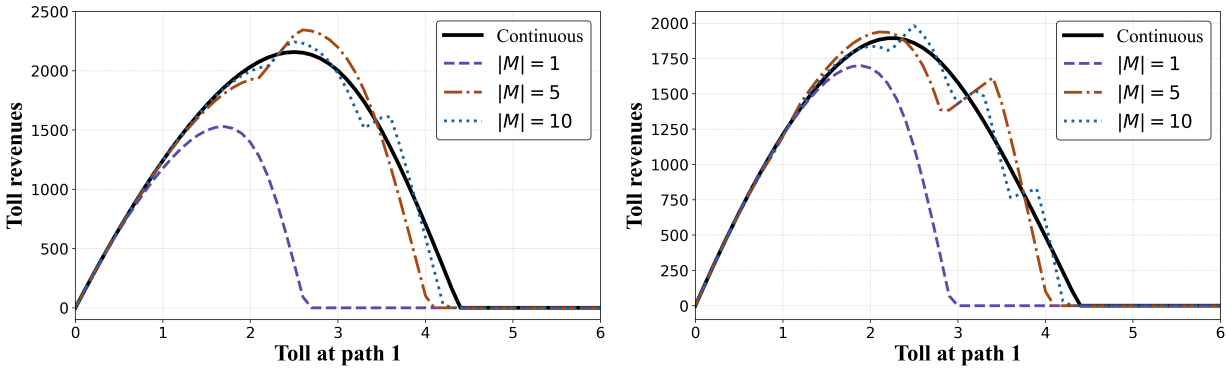
(b) Log-normal distribution

Figure 10: Equilibrium class-based path flows for discrete multi-class problem when $|M| = 50$.

5.1.4 Toll revenues: Continuous versus discrete

This subsection further compares the toll revenues of the continuous and discrete multi-class model ($|M| = 1, 5, 10$). We changes the toll on link/path 1 from 0 to 6 \$, while the path 2 is always a toll-free choice. Figure 11(a) and 11(b) show toll avenues using uniform and log-normal distributions, respectively. We can observe that single-class problem (i.e., $|M| = 1$) cannot receive

the maximum toll revenues as achieved by a continuous problem. Moreover, no travelers will choose path 1 when τ_1 is greater than about 3 \$. When user heterogeneity is considered to include more user classes (i.e., $|M| = 5$ and 10), we found that toll avenues for the discrete multi-class model would be competitive with the continuous problem.



(a) Uniform distribution of VOT

(b) Log-normal distribution of VOT

Figure 11: Toll avenues with different toll-settings on path 1.

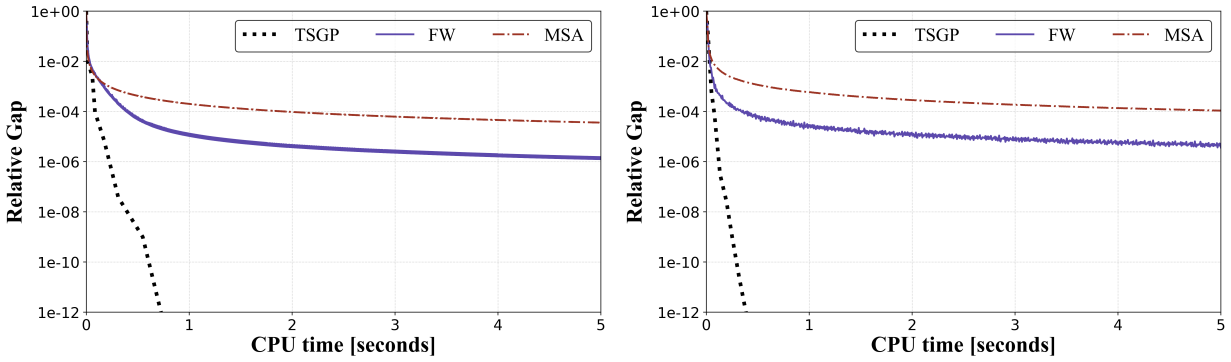
5.2 Large networks

In this section, we compare the computational performance of three algorithms (TSGP, MSA and FW) using large networks. Table 1 presents details of six large traffic networks, the information of which is available at https://github.com/bstabler/TransportationNetworks. Because the original instances (except Philadelphia) from the website do not contain tolls, the toll on each link is set as the marginal toll cost by solving the system optimum standard traffic assignment problem as suggested by Xie et al. (2021). VOT is assumed to follow a uniform distribution ranging from 6 to 30 \$/h. We use the BPR functions (see Eq. (32)) for link travel times and exponential demand functions (see Eq. (33)) for all O-D pairs. The upper bound demand is assumed to be the O-D demand from the original O-D matrix, and the parameter γ is set as 0.3 as suggested by Ryu et al. (2014). In our implementation, the maximum inner loops L and the control parameter Γ_1 are set by default to 500 and 0.1, respectively. Note that another control parameter Γ_2 takes the same setting as in Xie et al. (2021).

Table 1: Details of four six networks.

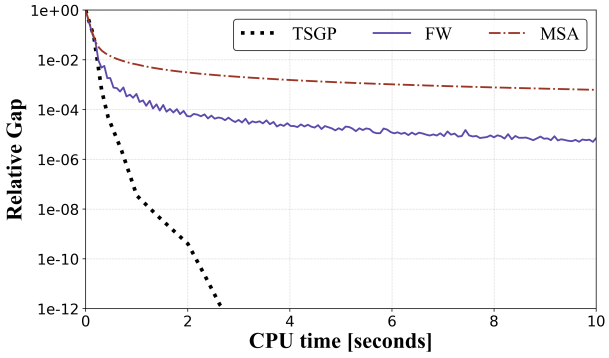
Networks	# Nodes	# Links	# Zones	# O-D pairs	# Trips	# Tolled links
Sioux Falls	24	76	24	528	360,600	68
Anaheim	416	914	38	1,406	104,694	43
Winnipeg	1,052	2,836	147	4,344	64,775	37
Barcelona	1,020	2,522	110	7,865	184,680	53
Chicago Sketch	933	2,950	387	93,512	1,260,907	862
Philadelphia	13,389	40,003	1,525	1,149,795	14,336,062	31

Note: The number of trips is the total demand from the original demand matrix.

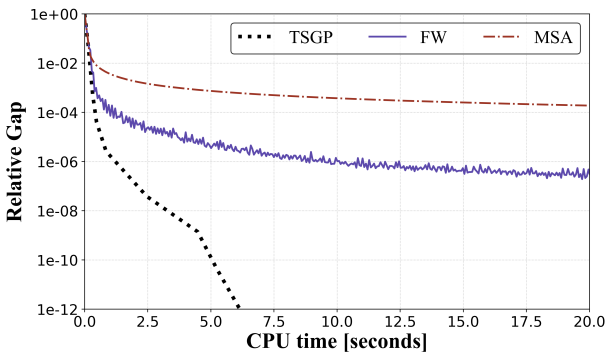


(a) Sioux Falls

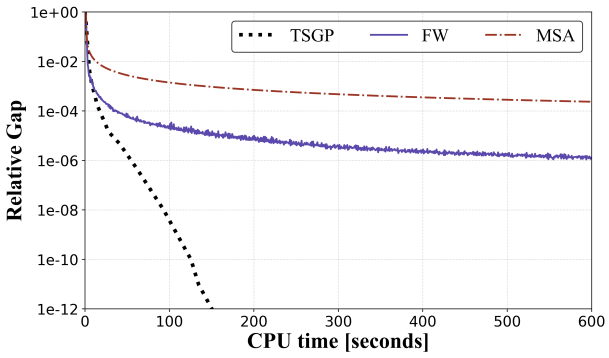
(b) Anaheim



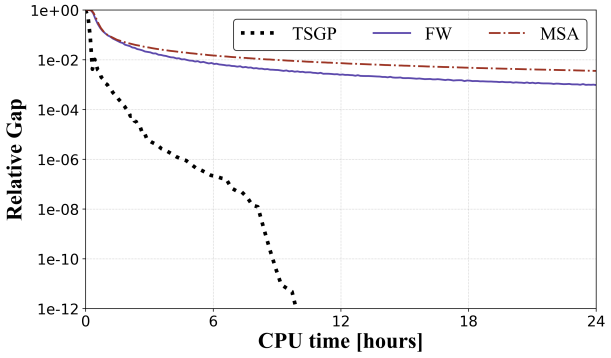
(c) Winnipeg



(d) Barcelona



(e) Chicago Sketch



(f) Philadelphia

Figure 12: Convergence performance for six large networks.

The convergence performances (CPU time versus relative gap) on six networks are compared in Figure 12. In our experiments, an algorithm is terminated if either it reaches a relative gap of 10^{-12} or the maximum running time. It can be seen that the superiority of TSGP becomes much more prominent in the larger examples. We can observe that TSGP is able to attain a relative gap of 10^{-12} for all networks, whereas the two link-based algorithms, FW and MSA, are unable to achieve highly precise equilibrium solutions. Besides Philadelphia, FW can only get to around 10^{-6}, and MSA for 10^{-4}. Their sluggish convergence behavior, which has been illustrated in a

two-link network in Section 5.1.2, is aligned with the algorithm's reputation as a standard solver for traffic assignment. For an especially computational challenge of Philadelphia, it can be seen that FW and MSA are struggling at around 10^{-3} after 24 hours. The remarkable performance of TSGP confirms its capability to deal with real-world instances involving millions of O-D pairs.

6 Conclusions

In this paper, we revisit the work of Leurent (1993) for the ED-CBiTA problem and present an alternative path-based formulation by extending the model of Marcotte (1998) and Xie et al. (2021) to further incorporate elastic demand. The proposed ED-CBiTA formulation is defined based on the O-D travel demand and TEM boundaries, which closely account for the trip generation and bi-criteria traffic assignment, respectively. We propose a highly efficient two-stage gradient projection (TSGP) algorithm. This novel algorithm predicated on the idea of reducing the original-complex problem into two types of subproblems, which can be solved very efficiently using the Newton-type GP method. More specifically,

- At the first stage, we apply GP to solve the O-D excess demand problem, which aims to achieve demand equilibration based on the current congestion level.
- At the second stage, we apply GP to solve the O-D single-boundary subproblem, which is to achieve exact positions of two adjacent efficient paths along the Pareto frontier.

Graphical illustrations are provided to show that both two equilibrium operations are in fact adjusting path flows in different "directions" to move toward equilibrium. That is, path flows are adjusted in "vertically" and "horizontally" directions at the first and second stage, respectively. We use small and larger networks to demonstrate TSGP's efficiency compared to two link-based benchmark algorithms, i.e., MSA and FW. Moreover, we compare and investigate the features between continuous and discrete multi-class models, and found that discretization errors could still exist even when a larger number of user classes is considered.

The directions pursued by this paper have opened many other extensions that would be interesting to explore in future studies. First, we can tailor the methodological tool developed in this paper to other variants and applications of ED-CBiTA. For example, one may be interested in finding first- or second-best tolls build upon TSGP. Second, one of the criteria (i.e., toll) of this study is assumed to be flow-independent. It is of particular interest to study a continuous multi-class problem with both attributes being flow-dependent. However, the path order no longer being deterministic in such a general problem, which would pose new challenges to both the model and the solution algorithm. Another possible direction for future research is to include more criteria (e.g., travel time reliability) concerned by travelers, leading to the continuous multi-criteria traffic assignment problem. The plan is to conduct experiments on large-scale networks to see how flow pattern changes and how much computational costs increase by considering a joint two-dimensional continuous distribution.

Acknowledgments

The work described in this paper was jointly supported by the Research Grants Council of the Hong Kong Special Administrative Region (PolyU 15222221), the Areas of Excellence Committee of the Hong Kong Polytechnic University (ZE0A), the Research Institute for Sustainable Urban Development at the Hong Kong Polytechnic University (1-BBWF), and the National Nature Science Foundation of China (Grant No. 72201220). Their support is gratefully acknowledged.

References

Babonneau, F., Vial, J.P., 2008. An efficient method to compute traffic assignment problems with elastic demands. Transportation Science 42, 249–260.

Beckmann, M., McGuire, C.B., Winsten, C.B., 1956. Studies in the Economics of Transportation. Technical Report.

Bertsekas, D., Gafni, E., Gallager, R., 1984. Second derivative algorithms for minimum delay distributed routing in networks. IEEE Transactions on Communications 32, 911–919.

Calderone, D., Dong, R., Sastry, S.S., 2017. External-cost continuous-type wardrop equilibria in routing games, in: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), IEEE. pp. 1–6.

Cambini, A., Martein, L., 2008. Generalized convexity and optimization: Theory and applications. volume 616. Springer Science & Business Media.

Cantarella, G.E., Cartenì, A., de Luca, S., 2015. Stochastic equilibrium assignment with variable demand: theoretical and implementation issues. European Journal of Operational Research 241, 330–347.

Chen, A., Jayakrishnan, R., 1998. A path-based gradient projection algorithm: Effects of equilibration with a restricted path set under two flow update policies, in: Proceedings of the Transportation Research Board Annual Meeting.

Chen, A., Lee, D.H., Jayakrishnan, R., 2002. Computational study of state-of-the-art path-based traffic assignment algorithms. Mathematics and computers in simulation 59, 509–518.

Chen, A., Xu, X., Ryu, S., Zhou, Z., 2013. A self-adaptive armijo stepsize strategy with application to traffic assignment models and algorithms. Transportmetrica A: Transport Science 9, 695–712.

Cirillo, C., Axhausen, K.W., 2006. Evidence on the distribution of values of travel time savings from a six-week diary. Transportation Research Part A: policy and practice 40, 444–457.

Clark, A., Sumalee, A., Shepherd, S., Connors, R., 2009. On the existence and uniqueness of first best tolls in networks with multiple user classes and elastic demand. International Journal of Control 5, 141–157.

Dafermos, S.C., 1981. A multicriteria route-mode choice traffic equilibrium model. Lefschetz Center for Dynamical Systems .

Dial, R.B., 1979. A model and algorithm for multicriteria route-mode choice. Transportation Research Part B: Methodological 13, 311–316.

Dial, R.B., 1996. Bicriterion traffic assignment: Basic theory and elementary algorithms. Transportation science 30, 93–111.

Dial, R.B., 1997. Bicriterion traffic assignment: Efficient algorithms plus examples. Transportation Research Part B: Methodological 31, 357–379.

Dial, R.B., 1999a. Network-optimized road pricing: Part I: A parable and a model. Operations Research 47, 54–64.

Dial, R.B., 1999b. Network-optimized road pricing: Part II: Algorithms and examples. Operations Research 47, 327–336.

Han, D., Yang, H., 2008. The multi-class, multi-criterion traffic equilibrium and the efficiency of congestion pricing. Transportation Research Part E: Logistics and Transportation Review 44, 753–773.

Jayakrishnan, R., Tsai, W., Prashker, J., Rajadhyaksha, S., 1994. A faster path-based algorithm for traffic assignment. Transportation Research Record 1443, 75–83.

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

Leurent, F., 1993. Cost versus time equilibrium over a network. European Journal of Operational Research 71, 205–221.

Leurent, F., 1996. The theory and practice of a dual criteria assignment model with a continuously distributed value-of-time, in: ISTTT, Pergamon. pp. 455–477.

Leurent, F., 1998. Sensitivity and error analysis of the dual criteria traffic assignment model. Transportation Research Part B: Methodological 32, 189–204.

Leurent, F., Coulombel, N., Poulhès, A., 2012. A disaggregate residential equilibrium assignment model. Procedia-Social and Behavioral Sciences 54, 758–771.

Liu, H.X., He, X., He, B., 2009. Method of successive weighted averages (mswa) and self-regulated averaging schemes for solving stochastic user equilibrium problem. Networks and Spatial Economics 9, 485–503.

Marcotte, P., 1998. Reformulations of a bicriterion equilibrium model, in: Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods. Springer, pp. 269–291.

Marcotte, P., Nguyen, S., Tanguay, K., 1996. Implementation of an efficient algorithm for the multiclass traffic assignment problem, in: Transportation and Traffic Theory. Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Pergamon, Oxford, Citeseer. pp. 217–236.

Marcotte, P., Patriksson, M., 2007. Chapter 10 Traffic equilibrium, in: Transportation. volume 14 of *Handbooks in Operations Research and Management Science*, pp. 623–713.

Marcotte, P., Zhu, D., 1994. An efficient algorithm for a bicriterion traffic assignment problem, in: Advanced methods in transportation analysis. Springer, pp. 63–73.

Marcotte, P., Zhu, D.L., 1997. Equilibria with infinitely many differentiated classes of customers. Complementarity and variational problems, state of art , 234–258.

Ryu, S., Chen, A., Choi, K., 2014. A modified gradient projection algorithm for solving the elastic demand traffic assignment problem. Computers & operations research 47, 61–71.

Ryu, S., Chen, A., Choi, K., 2017. Solving the combined modal split and traffic assignment problem with two types of transit impedance function. European Journal of Operational Research 257, 870–880.

Schmid, B., Molloy, J., Peer, S., Jokubauskaite, S., Aschauer, F., Hössinger, R., Gerike, R., Jara-Diaz, S.R., Axhausen, K.W., 2021. The value of travel time savings and the value of leisure in zurich: Estimation, decomposition and policy implications. Transportation Research Part A: Policy and Practice 150, 186–215.

Sheffi, Y., 1985. Urban transportation networks: Equilibrium analysis with mathematical programming methods. Prentice Hall, Englewood cliffs, NJ.

Slavin, H., Brandon, J., Lam, J., Sundaram, S., 2014. A brief note on toll road assignment methods.

Technical Report. Caliper Corporation.

Tan, H., Du, M., Chen, A., 2022. Accelerating the gradient projection algorithm for solving the non-additive traffic equilibrium problem with the barzilai-borwein step size. Computers & Operations Research 141, 105723.

Wardrop, J.G., 1952. Road paper. some theoretical aspects of road traffic research. Proceedings of the institution of civil engineers 1, 325–362.

Wu, W.X., Huang, H.J., 2014. Finding anonymous tolls to realize target flow pattern in networks with continuously distributed value of time. Transportation Research Part B: Methodological 65, 31–46.

Xie, J., Wang, Q., Nie, Y.M., 2021. An efficient algorithm for continuous bi-criteria traffic assignment. Preprint online URL: https://doi.org/10.13140/RG.2.2.24929.43364.

Xu, Z., Chen, A., Liu, X., 2023. Time and toll trade-off with heterogeneous users: A continuous time surplus maximization bi-objective user equilibrium model. Transportation Research Part B: Methodological 173, 31–58.

Xu, Z., Xie, J., Liu, X., Nie, Y.M., 2020. Hyperpath-based algorithms for the transit equilibrium assignment problem. Transportation Research Part E: Logistics and Transportation Review 143, 1–18.

Xu, Z., Xie, J., Liu, X., Nie, Y.M., 2022. Hyperbush algorithm for strategy-based equilibrium traffic assignment problems. Transportation science 56, 877–903.

Yang, H., Huang, H.J., 2004. The multi-class, multi-criteria traffic network equilibrium and systems optimum problem. Transportation Research Part B: Methodological 38, 1–15.

Yu, Q., Fang, D., Du, W., 2014. Solving the logit-based stochastic user equilibrium problem with elastic demand based on the extended traffic network model. European Journal of Operational Research 239, 112–118.

Zhang, H., Liu, Z., Wang, J., Wu, Y., 2023. A novel flow update policy in solving traffic assignment problems: Successive over relaxation iteration method. Transportation Research Part E: Logistics and Transportation Review 174, 103111.

Appendix A Proof for the formulation equivalence

The path-based convex formulation for ED-CBiTA in Leurent (1993) is formulated as

$$\min_{\mathbf{f}} \quad Z_f = \underbrace{\sum_{a\in A}\int_0^{x_a} t_a(z)dz}_{\text{Beckman transformation}} + \underbrace{\sum_{rs\in RS}\sum_{k\in K_{rs}} q^{rs}\tau_k\left[E_{rs}\left(\frac{Q_k^{rs}}{q^{rs}}\right) - E_{rs}\left(\frac{Q_{k-1}^{rs}}{q_{rs}}\right)\right]}_{\text{Monetary expense expectation}} - \underbrace{\sum_{rs\in RS}\int_0^{q^{rs}} D_{rs}^{-1}(z)dz}_{\text{Inverse demand integrals}} \tag{34a}$$

$$s.t. \quad f_k \geq 0, \quad \forall k \in K_{rs}, rs \in RS \tag{34b}$$

$$Q_k^{rs} = \sum_{j\leq k} f_j, \quad \forall rs \in RS,\ k = 1,\ldots,|K_{rs}|, \tag{34c}$$

$$q^{rs} = \sum_{k\in K_{rs}} f_k, \quad \forall rs \in RS, \tag{34d}$$

$$x_a = \sum_{rs\in RS}\sum_{k\in K_{rs}} \delta_{a,k} f_k, \quad \forall a \in A, \tag{34e}$$

where $E_{rs}(\cdot)$ is the primitive of the inverse CDF (i.e., $H_{rs}(\cdot)$), Q_k^{rs} is the cumulative flow on paths whose tolls are less than or equal to the toll on path k corresponding to O-D pair rs. Although we herein don't group flows for paths such that have the same toll as in Leurent (1993), it is easy to prove this operation won't affect the formulation.

Constraint (34c) stores the path flow information with a descending order of toll, which is consistent with the definition of TEM intervals as in Eq. (3). Both feasible sets in MPs (34) and (6) are the same, i.e., including non-negative and conservation constraints for path flows (implicitly the non-negatively of O-D demands in Eq. (34d)). Next, we show that the objective functions (34a) and (6a) are equivalent, i.e., $Z_f = Z_1$. Recall that the first Beckman-type term and the third inverse demand integrals term are the same for both formulations. While the second term in Eq. (34a) that relates to toll costs can be reconstructed as follows:

$$\sum_{rs\in RS}\sum_{k\in K_{rs}} q^{rs}\cdot\tau_k\cdot\left[E_{rs}\left(\frac{Q_k^{rs}}{q^{rs}}\right) - E_{rs}\left(\frac{Q_{k-1}^{rs}}{q^{rs}}\right)\right] \tag{35a}$$

$$= \sum_{rs\in RS}\sum_{k\in K_{rs}} q^{rs}\cdot\tau_k \int_{\frac{Q_{k-1}^{rs}}{q^{rs}}}^{\frac{Q_k^{rs}}{q^{rs}}} H_{rs}^{-1}(u)du = \sum_{rs\in RS}\sum_{k\in K_{rs}} q^{rs}\cdot\tau_k \int_{H_{rs}^{-1}(\frac{Q_{k-1}^{rs}}{q^{rs}})}^{H_{rs}^{-1}(\frac{Q_k^{rs}}{q^{rs}})} z\cdot d(H_{rs}(z)) \tag{35b}$$

$$= \sum_{rs\in RS}\sum_{k\in K_{rs}} q^{rs}\cdot\tau_k \int_{H_{rs}^{-1}(\frac{Q_{k-1}^{rs}}{q^{rs}})}^{H_{rs}^{-1}(\frac{Q_k^{rs}}{q^{rs}})} z\cdot h_{rs}(z)dz = \underbrace{\sum_{rs\in RS}\sum_{k\in K_{rs}} q^{rs}\cdot\tau_k \int_{\beta_{k-1}^{rs}}^{\beta_k^{rs}} z\cdot h_{rs}(z)dz}_{\text{Continuous monetary expense}}. \tag{35c}$$

Consequently, MPs (6) and (34) are equivalent, which completes the proof.

Appendix B Formulation of the discrete multi-class model

Mathematically, the discrete multi-class excess demand problem can be formulated as below:

$$\min_{\{f_{km},\, e_m^{rs}\}} \quad Z_{\text{discrete}} = \underbrace{\sum_{a\in A}\int_0^{x_a} t_a(z)dz}_{\text{Beckman transformation}} + \underbrace{\sum_{rs\in RS}\sum_{m\in M}\sum_{k\in K_{rs}} \beta_m^{rs}\tau_k f_{km}}_{\text{Monetary expense summation}} + \underbrace{\sum_{rs\in RS}\sum_{m\in M}\int_0^{e_m^{rs}} W_m^{rs}(z)dz}_{\text{Excess demand integrals}} \tag{36a}$$

$$s.t. \quad \bar{q}_m^{rs} = \sum_{m\in M}\sum_{k\in K_{rs}} f_{km} + e_m^{rs}, \quad \forall rs \in RS, m \in M, \tag{36b}$$

$$x_a = \sum_{rs\in RS}\sum_{m\in M}\sum_{k\in K_{rs}} \delta_{a,k} f_{km} \tag{36c}$$

$$f_{km} \geq 0, \quad \forall rs \in RS, m \in M, k \in K_{rs}, \tag{36d}$$

$$e_m^{rs} \geq 0, \quad \forall rs \in RS, m \in M, \tag{36e}$$

where each user class $m \in M$ is associated with a TEM β_m^{rs}. Flows are subscripted with user class m, i.e., f_{km} is the flow on path k for class $m \in M$, $W_m^{rs}(\cdot)$ is the excess demand function for class m, e_m^{rs} and $\bar{q}^{rs}$ denote the excess and upper bound demand of O-D pair rs for class m, respectively.

The above formulation directly operates in the space of class-specific path flows and excess demands. Notice that the unique equilibrium solution is ensured w.r.t. total link flows and O-D demands, while it does not generally hold for class-specific link and path flows.

Appendix C The MSA procedure

Algorithm 4 MSA

1: **Initialization:** Given the initial demand vector $\mathbf{q}_0 = \hat{\mathbf{q}} = \bar{\mathbf{q}}$, call Dial's T2 algorithm to obtain link flow vector $\mathbf{x}^0$, link moment vector $\mathbf{u}^0$, and the O-D based expected generalized travel time $\mathbf{T}_0$. Set the loop counter $n = 0$ and update link cost $\mathbf{t}$.

2: **Descent direction finding:** Set travel demand $\hat{\mathbf{q}}_{n+1} = D(\mathbf{T}_n)$. Perform Dial's T2 algorithm to obtain the auxiliary link flow vector $\bar{\mathbf{x}}^n$, auxiliary link moment vector $\bar{\mathbf{u}}^n$, the O-D based expected generalized travel time $\mathbf{T}_{n+1}$.

3: **Move:** Set $\mathbf{x}^{n+1} = \frac{1}{n+1}\bar{\mathbf{x}}^n + \frac{n}{n+1}\mathbf{x}^n$, $\mathbf{u}^{n+1} = \frac{1}{n+1}\bar{\mathbf{u}}^n + \frac{n}{n+1}\mathbf{u}^n$, $\mathbf{q}_{n+1} = \frac{1}{n+1}\hat{\mathbf{q}}_{n+1} + \frac{n}{n+1}\mathbf{q}_n$.

4: **Update:** Update link cost vector $\mathbf{t}$.

5: **Convergence test**: If the relative gap $RG = 1 - \frac{\sum_{rs\in RS} D_{rs}(T_{rs}^n)\cdot T_{rs}^n}{\sum_{a\in A}(x_a^{n+1}t_a^{n+1} + u_a^{n+1}m_a)}$ is less than the convergence criteria or the maximum running time is reached, stop; otherwise set $n = n+1$ and return to Step 2.

Appendix D The FW procedure

Algorithm 5 FW

1: **Initialization:** Given the initial demand vector $\mathbf{q}_0 = \hat{\mathbf{q}} = \bar{\mathbf{q}}$, call Dial's T2 algorithm to obtain link flow vector $\mathbf{x}^0$, link moment vector $\mathbf{u}_0$, and the O-D based expected generalized travel time $\mathbf{T}_0$. Set the loop counter $n = 0$ and update link cost $\mathbf{t}$.

2: **Descent direction finding:** Set travel demand $\hat{\mathbf{q}}_{n+1} = D(\mathbf{T}_n)$. Perform Dial's T2 algorithm to obtain the auxiliary link flow vector $\bar{\mathbf{x}}^n$, auxiliary link moment vector $\bar{\mathbf{u}}^n$, the O-D based expected generalized travel time $\mathbf{T}_{n+1}$.

3: **Line search:** Solving the following one-dimensional problem to find an optimal stepsize λ_n^*:

$$\underset{\lambda_n \in [0,1]}{\operatorname{argmin}} \left\{ \left[\int_0^{\lambda_n \bar{x}^n + (1-\lambda_n) x_a^n} t_a(z) dz + \tau_a \cdot (\lambda_n \bar{u}_a^n + (1-\lambda_n) u_a^n) \right] - \sum_{rs} \int_0^{\lambda_n \hat{q}_n^{rs} + (1-\lambda_n) q_n^{rs}} D_{rs}^{-1}(z) dz \right\}$$

4: **Move:** Set $\mathbf{x}^{n+1} = \lambda_n^* \bar{\mathbf{x}}^n + (1-\lambda_n^*)\mathbf{x}^n$, $\mathbf{u}^{n+1} = \lambda_n^* \bar{\mathbf{u}}^n + (1-\lambda_n^*)\mathbf{u}^n$, $\mathbf{q}_{n+1} = \lambda_n^* \hat{\mathbf{q}}_{n+1} + (1-\lambda_n^*)\mathbf{q}_n$.

5: **Update:** Update link cost vector $\mathbf{t}$.

6: **Convergence test**: If the relative gap $RG = 1 - \frac{\sum_{rs \in RS} D_{rs}(T_{rs}^n) \cdot T_{rs}^n}{\sum_{a \in A} (x_a^{n+1} t_a^{n+1} + u_a^{n+1} m_a)}$ is less than the convergence criteria or the maximum running time is reached, stop; otherwise set $n = n + 1$ and return to Step 2.

Appendix E Distributions of TEM from uniform and log-normal distributions of VOT

- Given a uniform distribution of VOT $\alpha \sim U(\underline{\alpha}, \overline{\alpha})$, it is easy to derive the distribution of TEM $\beta = 1/\alpha$ as follows:

$$h(\beta) = \frac{1}{\beta^2(\underline{\alpha} - \overline{\alpha})}, \tag{37a}$$

$$H(\beta) = \int_{\frac{1}{\beta}}^{\beta} h(z) dz = \frac{\overline{\alpha} - 1/\beta}{(\underline{\alpha} - \overline{\alpha})}, \tag{37b}$$

$$\int \beta \cdot h(\beta) = \frac{\log(\beta)}{\underline{\alpha} - \overline{\alpha}} + C, \tag{37c}$$

  where C is a constant.

- Given a log-normal distribution of VOT $\alpha \sim Logn(\mu, \sigma)$ such that $\alpha \in [\underline{\alpha}, \overline{\alpha}]$, then TEM $\beta \sim Logn(-\mu, \sigma)$ such that $\beta \in [1/\overline{\alpha}, 1/\underline{\alpha}]$. Let Δp denote the portion within the range of the truncated support, i.e.,

$$\Delta p = \frac{1}{2} \left[erf\left(\frac{\ln(1/\underline{\alpha}) + \mu}{\sigma\sqrt{2}} \right) - erf\left(\frac{\ln(1/\overline{\alpha}) + \mu}{\sigma\sqrt{2}} \right) \right], \tag{38}$$

  where the Gauss error function $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-z^2)\, dz$.

Further, the distributions of β can be analytically derived as:

$$h(\beta) = \frac{1}{\beta \cdot \Delta p \cdot \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln \beta + \mu)^2}{2\sigma^2}\right), \tag{39a}$$

$$H(\beta) = \frac{1}{2\Delta p}\left[erf\left(\frac{\ln \beta + \mu}{\sigma\sqrt{2}}\right) - erf\left(\frac{\ln(1/\overline{\alpha}) + \mu}{\sigma\sqrt{2}}\right)\right], \tag{39b}$$

$$\int \beta \cdot h(\beta) = \frac{1}{2\Delta p} \exp\left(-\mu + \frac{\sigma^2}{2}\right) \cdot \left[1 + erf\left(\frac{\log \beta + \mu - \sigma^2)}{2\sigma}\right)\right] + C. \tag{39c}$$