Multi-class percentile user equilibrium with flow-dependent stochasticity

Yu (Marco) Nie*

Department of Civil and Environmental Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, United States

ABSTRACT

Travelers often reserve a buffer time for trips sensitive to arrival time in order to hedge against the uncertainties in a transportation system. To model the effects of such behavior, travelers are assumed to choose routes to minimize the percentile travel time, i.e. the travel time budget that ensures their preferred probability of on-time arrival; in doing so, they drive the system to a percentile user equilibrium (UE), which can be viewed as an extension of the classic Wardrop equilibrium. The stochasticity in the supply of transportation are incorporated by modeling the service flow rate of each road segment as a random variable. Such stochasticity is flow-dependent in the sense that the probability density functions of these random variables, from which the distribution of link travel time are constructed, are specified endogenously with flow-dependent parameters. The percentile route travel time, obtained by directly convolving the link travel time distributions in this paper, is not available in closed form in general and has to be numerically evaluated. To reveal their structural properties, percentile UE solutions are examined in special cases and verified with numerical results. For the general multi-class percentile UE traffic assignment problem, a variational inequality formulation is given and solved using a route-based algorithm. The algorithm makes use of the diagonal elements in the Jacobian of percentile route travel time, which is approximated through recursive convolution. Preliminary numerical experiments indicate that the algorithm is able to achieve highly precise equilibrium solutions.

1. Introduction

Urban transportation systems are affected by uncertainties of various sorts, which can be broadly classified as those affecting the supply of transportation (e.g. weather, accidents, natural and man-made disasters) and those associated with the demand for transportation (e.g. travel and activity behavior, special events). Taken individually or in combination, these factors could adversely affect the quality of transportation services. Travel behavior researchers have established that unanticipated long delays on highways typically produce much worse frustration among motorists than "predictable" ones. To hedge against travel time fluctuations, travelers usually budget a sizable buffer time. In 1982, a 20-min free-flow trip in the US requires on average an extra 12-min buffer time if on-time arrival is important (FHWA, 2005). A similar trip would require 60% more buffer time in 2003. An important question that often arises in the context of transportation planning and network design is: how travelers' "buffer-reserving" behavior to cope with network uncertainties affects the overall performance of the transportation system. This paper provides a basis to address this question by formulating and solving a traffic assignment model that incorporates such behavior.

Traffic assignment is widely used to generate facility-level predictions by mapping travel demands onto the transportation network. The classic model, which postulates that travelers always choose the shortest route (Wardrop, 1952), produces a steady network flow pattern known as user equilibrium (UE) as the solution to the traffic assignment problem (Beckmann et al., 1956; Sheffi, 1985). Despite its remarkable success in practice, the classic traffic assignment model ignores the...
uncertainties inherited in both supply and demand sides of the transportation, which has been frequently cited as one of its major limitations. Many models attempt to incorporate uncertainties in traffic assignment. This paper is focused on the impacts of supply uncertainties on travelers’ route choices, that is, the only source of uncertainties considered herein is from the supply of transportation. More specifically, we assume that the service flow rate (SFR)\(^1\) of each road segment is random, and subject to the influence of exogenous (e.g. weather) and endogenous (e.g. traffic breakdown) random factors. Fluctuations in travel demands as well as individuals’ imperfect information and irrational choice behavior are excluded to simplify the analysis and maintains the tractability of the resulting model.

In this paper, the travel time on each link of the transportation network is treated as an independent random variable, whose distribution is endogenously determined from the distribution of SFR and a functional relationship between SFR, traffic flow and travel time. Accordingly, the route travel time is also a random variable, and its distribution is formulated by convolving the distributions of its member links. Travelers are assumed to have perfect knowledge of the route travel time distributions, and choose best routes to fulfill their desired on-time arrival probability, i.e., the probability of arriving at the destination on-time or earlier. Clearly, the “best” route in this context is the route that requires least percentile travel time (or time budget) corresponding to the desired on-time arrival probability. Consequently, the Wardrop equilibrium can be extended to the percentile user equilibrium, where all travelers with the same on-time probability should experience minimum and equal percentile travel time between the same origin-destination (O–D) pair.

Although the percentile user equilibrium is not a new concept, several features distinguish this study from those found in the literature.

- First of all, the proposed model enables flow-dependent stochasticity, which means that the distribution of SFR is endogenously determined based on link flow. In contrast, previous studies usually require such distributions to be specified exogenously, independent of traffic flow. As explained in detail in Section 6, the distribution of random road capacities is likely to be flow-dependent because the two important contributing factors of stochasticity, accident rates and traffic breakdown, are known to be affected by traffic flow.
- Second, the percentile route travel time is modeled as an implicit function of the route flow, which is evaluated by adding distributions directly through convolution. We postulate that the implicit percentile route travel time function is not monotone in general and provide supporting numerical evidence. A sufficient but restrictive condition that ensures monotonicity of the function is provided. Most previous studies depend on the central limit theorem (CLT) to evaluate this function. For one thing, CLT requires imposing restrictions on the type of distributions that can be used.\(^2\) Moreover, since the solution provided by CLT is exact only when the number of distributions \(n\) approaches infinity, its approximations may be unsatisfactorily large when \(n\) is small.
- Thirdly, analytical solutions are examined in special cases in order to provide insights to understanding the properties of the general percentile UE solutions. These solutions are then verified through numerical experiments.
- Finally, a route-based algorithm is proposed for the percentile UE problem. The algorithm makes use of the partial Jacobian of the percentile route travel time, which is approximated by convolving the distributions of the link travel time derivatives. While not implemented and tested in this paper, the proposed algorithm can be combined with a column generation scheme to obviate path enumeration. One possibility is using the reliable a priori shortest path algorithm (Nie and Wu, 2009) to find all non-dominated paths, which collectively provide minimum percentile travel time for any on-time arrival probability.

The remaining of this article is organized as follows. Section 2 briefly reviews the related studies. Section 3 discusses the adopted stochastic performance model and its analytical properties. Section 4 characterizes the percentile UE and presents an equivalent formulation based on variational inequality (VI). The properties of the percentile route travel time function and the evaluation of its Jacobian are also presented in Section 4. Section 5 reveals the structure of the percentile UE solutions in special cases. Section 6 discusses flow-dependent stochasticity. The solution algorithm and numerical results are presented in Sections 7 and 8, respectively. Section 9 concludes the study with remarks on future research.

2. Literature review

2.1. Classic stochastic user equilibrium models

It is well known that travelers do not always choose the best route, owing to random disturbances such as measurement and perception errors. Probabilistic choice theory most often used in transportation postulates that (1) travelers always want to choose the best route and (2) each route has a probability of being the best. Accordingly, a traffic assignment problem can be formulated to achieve stochastic user equilibrium (SUE) (Daganzo and Sheffi, 1977; Fisk, 1980), where the proportion of

\(^1\) We use service flow rate instead of capacity because random disruptions, such as accidents and traffic breakdown, affect service flow rate but not capacity. Notes in Acknowledgement explains why SFR might be more precise term than “capacity” when used in the circumstances such as concerned in this paper.

\(^2\) In the most restrictive form, CLT requires that all distributions are independently and identically distributed (IID). When the distributions are not identically distributed, other conditions, such the Lyapunov condition or the Lindeberg condition, are needed to ensure the validity of the theorem (Ash and Dolans-Dade, 2000).

Please cite this article in press as: Nie, Y. Multi-class percentile user equilibrium with flow-dependent stochasticity. Transportation Research Part B (2011), doi:10.1016/j.trb.2011.06.001
flows assigned to each route equals the choice probability. Note that in the classic SUE models, the disturbances in travel time are unobservable and their distributions has to be selected exogenously. The choice of this exogenous distribution dictates the nature of the underlying choice model.

The widely used logit model (Dial, 1971; Fisk, 1980; Bell, 1995; Maher, 1998) offers a closed form choice probability by assuming the disturbances in the travel time of each alternative route are independently and identically distributed (IID) Gumbel variables (Ben-Akiva and Lerman, 1985). However, this IID assumption is problematic, perhaps more so in route choice than in other circumstances, because the travel time disturbances are likely to be highly correlated (due to overlapping links) and have very different variances (due to the differences in topological, geographic and traffic characteristics) (Sheffi, 1985). Another random utility model, the probit model (Daganzo and Sheffi, 1977), assumes that travel time disturbances follow a multivariate normal distribution. The probit model allows complete flexibility in the variance-covariance structure of the disturbances, which makes it more suitable for route choice (Daganzo and Sheffi, 1977; Sheffi and Powell, 1982; Daganzo, 1983; Maher, 1997). However, the probit model does not have a closed form choice probability and must rely on numerical evaluation, either through multi-dimensional integration or Monte Carlo simulation (Sheffi, 1985). The expensive computational costs associated with these numerical procedures have limited large-scale applications of the probit model.

2.2. Reliability-based user equilibrium models

Hall (1983) noted that travelers tend to reserve a safety margin to hedge against variations of travel times. Such a behavioral assumption leads to a reliability-based extension of the Wardrop equilibrium, at which all used routes between an O-D pair have minimum and identical time budget, defined as the sum of mean travel time and a safety margin. An early traffic assignment model of this kind is Uchida and Iida (1993), who assume that link travel times are normally distributed with flow-dependent mean and variance. Since the route travel time is also normally distributed (as the sum of normal variables), its safety margin can be calculated from the inverse cumulative probability function (CDF) for a given reliability requirement. Instead of directly making assumptions about the travel time distribution, Lo and Tung (2003) and Lo et al. (2006) developed a stochastic performance model which derives the distribution of the travel time on a link from that of its capacity through a travel time function. The central limit theorem is then employed to obtain the distribution of route travel times. This stochastic performance model places the source of the uncertainties in road capacities, whose distribution is exogenously specified. Shao et al. (2006) showed that the distribution of link travel time may be derived from those of the O-D demands, recognizing that the travel time can be expressed as a closed-form function of all O-D demands if path choice proportions are given. Later, Lam et al. (2008) extended the above demand-driven stochastic performance model to combine the supply-side uncertainties such as those associated with extreme weather conditions. However, the interdependence between the demand and supply effects is ignored. Chen and Zhou (2009) argued, along the line of Mirchandani and Soroush (1987), that travelers’ perception errors should be integrated with the distributions of random road travel times to obtain the perceived travel time distributions as the basis for decision-making. Moment analysis is employed to add the distribution of the random travel time to that of the perception error; the latter is conditional on the realization of random travel time.

Other “reliability indexes” have been employed in the literature to model route choice when route travel times are random with known distributions. Examples include mean-excess travel time (Chen and Zhou, 2010), which is defined as the conditional expectation of the travel time larger than the minimum percentile travel time, an the expected utility theory (Yin et al., 2004; Watling, 2006). Moreover, travelers may not always choose the route with the best reliability index due to imperfect information and irrational behavior. To capture the probabilistic choice behavior, Shao et al. (2006) and Siu and Lo (2006) proposed adding a Gumbel random term, interpreted as the perception error, into the associated reliability index. Accordingly, the logit model is employed to compute the choice probability of each route, using the associated reliability indexes as the systematic components of the route utility.

A behavioral assumption very different from what has been discussed so far concerns pessimistic travelers who always prepare for the worst (Bell, 2000). Using this assumption, Bell and Cassir (2002) proposed a robust optimization formulation to maximize the minimum expected travel time over all possible link failure scenarios. It was also shown (Bell and Cassir, 2002) that the resulting stochastic assignment problem can be formulated as a Nash–Cournot game between travelers and an imaginary “demon” who seeks to maximize the damage to the network. Szeto et al. (2006, 2007) extended Bell’s work (Bell and Cassir, 2002) to consider elastic demand and to allow multiple demons. Ordóñez and Stier-Moses (2010) assumed that travelers will select the “robust shortest path”, which minimizes the worst-case travel time. They compared the equilibrium solution resulting from the robust route choice behavior with the percentile UE and concluded that their robust formulation provides reasonable approximations.

3. Stochastic performance model

Travel time variability is a result of many interacting random factors, including those endogenous to traffic stream and those that can be treated as exogenous. We intend to capture both effects by modeling the service flow rate (SFR) of the road...
as a random variable whose distribution may depend on congestion levels. Travel time on link \( a \), denoted as \( \tilde{t}_a \), is assumed to be strictly increasing with link volume \( x_a \) and strictly decreasing with service flow rate (SFR) \( \tilde{c}_a \). \( \tilde{t}_a \) also depends a vector of other parameters \( \beta_a = [\beta_a^1, \ldots, \beta_a^n] \), written as \( \tilde{t}_a = g(x_a, \tilde{c}_a, \beta_a) \). Note that a variable headed with \( \tilde{\cdot} \) is a random variable. We shall assume that the distribution of the SFR \( \tilde{c}_a \) can be calibrated from historical data, and that the stochasticity of \( \tilde{t}_a \) follows from that of \( \tilde{c}_a \), through the link performance function \( g(\cdot) \). Later in Section 6, a flow-dependent function is proposed to model the cumulative distribution function (CDF) of the random SFR. Suffices it now to say that \( \tilde{c}_a \) is a continuous random variable and its CDF is a continuous function

\[ F_{\tilde{c}}(y) = Pr(\tilde{c}_a \leq y) = \int_0^y \nu_{\tilde{c}}(z)dz \]  

(1)

where \( \nu_{\tilde{c}}(\cdot) \) is the probability density function (PDF) of \( \tilde{c}_a \). One can write the random SFR as a function of the random travel time, i.e., \( \tilde{c}_a = g^{-1}(x_a, \tilde{t}_a, \beta_a) \). Therefore, once \( F_{\tilde{c}}(y) \) is given, the CDF for \( \tilde{t}_a \) can be obtained using the following relationship:

\[ F_{\tilde{t}}(y) = Pr(\tilde{t}_a \leq y) = Pr(g(x_a, \tilde{c}_a, \beta_a) \leq y) = Pr(g^{-1}(x_a, y, \beta_a) \leq \tilde{c}_a) = 1 - F_{\tilde{c}}(g^{-1}(x_a, y, \beta_a)) \]  

(2)

The third equality holds because travel time is a decreasing function of SFR by assumption. Accordingly, the probability density function of \( \tilde{t}_a \) can be represented by

\[ \nu_{\tilde{t}}(y) = \frac{\partial F_{\tilde{t}}(y)}{\partial y} = -\nu_{\tilde{c}}(g^{-1}(x_a, y, \beta_a)) \frac{\partial g^{-1}(x_a, y, \beta_a)}{\partial y} \]  

(3)

**Definition 1** (\( \alpha \)-percentile). \( \alpha \)-percentile link travel time is the time required to complete traversing the link with a probability \( \alpha \), i.e., \( t_{a}^{\alpha} = F_{\tilde{t}}^{-1}(\alpha) \). \( \alpha \)-percentile link SFR \( c_{a}^{\alpha} \) is such that \( c_{a}^{\alpha} = F_{\tilde{c}}^{-1}(1 - \alpha) \).

The percentile travel time is a deterministic variable that travelers can use to ensure arriving on-time with a desired probability. As shown in [Wu and Nie (in press)](https://doi.org/10.1016/j.trb.2011.06.001), the percentile travel time is equivalent to the travel time budget such as discussed in [Lo et al. (2006)](https://doi.org/10.1016/j.trb.2011.06.001) when the travel time is normally distributed.

**Proposition 1.** Given a CDF \( F_{\tilde{c}}(y) \) and a link performance function \( g(\cdot) \), the \( \alpha \)-percentile link travel time \( t_{a}^{\alpha} = g(x_a, c_{a}^{\alpha}, \beta_a) \).

**Proof.** Note that \( t_{a}^{\alpha} = F_{\tilde{t}}^{-1}(\alpha) \). Hence,

\[ \alpha = F_{\tilde{t}}(t_{a}^{\alpha}) = 1 - F_{\tilde{c}}(g^{-1}(x_a, t_{a}^{\alpha}, \beta_a)) \]

The second equality is due to the relationship (2). Thus, we have

\[ g^{-1}(x_a, t_{a}^{\alpha}, \beta_a) = F_{\tilde{c}}^{-1}(1 - \alpha) \rightarrow t_{a}^{\alpha} = g\left(x_a, F_{\tilde{c}}^{-1}(1 - \alpha), \beta_a\right) \]

Noting \( c_{a}^{\alpha} = F_{\tilde{c}}^{-1}(1 - \alpha) \) completes the proof. \( \square \)

**Proposition 1** suggests that the \( \alpha \)-percentile link travel time can be computed using a deterministic link performance function with a nominal SFR that properly reflects reliability requirements. It follows that if all routes in the network consist of no more than one link, the percentile UE assignment problem can be converted to a deterministic counterpart.

The derivative of \( \tilde{t}_a \) with respect to \( x_a \), denoted as

\[ \tilde{t}_a = \frac{dg(x_a, \tilde{c}_a, \beta_a)}{dx_a} = g'(x_a, \tilde{c}_a, \beta_a) \equiv h(x_a, \tilde{c}_a, \beta_a) \]  

(4)

has similar properties as \( \tilde{t}_a \). Specifically, the CDF and PDF of \( \tilde{t}_a \) are

\[ F_{\tilde{t}}(y) = 1 - F_{\tilde{t}}(h^{-1}(x_a, y, \beta_a)), \quad \nu_{\tilde{t}}(y) = -\nu_{\tilde{t}}(h^{-1}(x_a, y, \beta_a)) \frac{\partial h^{-1}(x_a, y, \beta_a)}{\partial y} \]  

(5)

Furthermore, the \( \alpha \)-percentile link travel time derivative can be expressed as

\[ t_{a}^{\alpha} = h(x_a, c_{a}^{\alpha}, \beta_a) \]  

(6)

Recalling **Proposition 1**, it is clear that the derivative of \( \alpha \)-percentile link travel time \( \left( \frac{dt_{a}^{\alpha}}{dx_a} \right) \) is the corresponding \( \alpha \)-percentile derivative of link travel time \( \left( \frac{dt_{a}}{dx_a} \right) \), i.e.,

\[ t_{a}^{\alpha} = h(x_a, c_{a}^{\alpha}, \beta_a) = \frac{dg(x_a, c_{a}^{\alpha}, \beta_a)}{dx_a} = \frac{dt_{a}}{dx_a} \]  

(7)
4. Percentile user equilibrium: Formulation and properties

Consider a network consisting of a set of nodes $N$, a set of links $A$, and a set of OD pairs $W \subset N^2$. Each OD pair $w \in W$ is connected by a set of routes $K_w$. Let $K = \bigcup_{w \in W} K_w$ denote the set of all routes, and let $m = |A|$, $o = |W|$ and $r = |K|$ denote the cardinalities of $A$, $W$ and $K$ respectively. Let the matrix $(A = [A_{ab}]) \in \mathbb{R}^{m \times r}$ denote the OD-route incidence matrix in which $A_{ak} = 1$ if route $k \in K_w$ and $A_{ak} = 0$ otherwise, and the matrix $(A = [A_{ik}]) \in \mathbb{R}^{r \times m}$ denote the link-route incidence matrix; here $A_{ik} = 1$ if link $k$ is on route $i$ and $A_{ik} = 0$ otherwise. Further, travelers with the same on-time arrival reliability $a$ are grouped into class $a \in P$, where $s = |P|$. We use column vectors $(q = [q^i]) \in \mathbb{R}^{r \times s}$, $(f = [f^k]) \in \mathbb{R}^{s \times r}$, $(\xi = [\xi^i]) \in \mathbb{R}^{s \times s}$ and $(\chi = [\chi^k]) \in \mathbb{R}^{s \times r}$ to denote the travel demand, the route flow, the percentile route travel time and the link flow respectively, where $(q^i = [q^i]) \in \mathbb{R}^s$, $(f^k = [f^k]) \in \mathbb{R}^r$, $(\xi^i = [\xi^i]) \in \mathbb{R}^s$ and $(\chi^k = [\chi^k]) \in \mathbb{R}^r$ are corresponding column vectors for class $a$. Finally, let $f^{wa}_{k}$ and $q^{wa}_{k}$ denote the flow and the associated percentile route travel time for class $a$ travelers from OD pair $w$ on route $k$. If all travelers choose routes based on the percentile route travel time according to their own $a$, the user equilibrium (UE) conditions imply that any used route has the identical and minimum percentile route travel time, i.e.,

$$f^{wa}_{k} > 0 \Rightarrow \xi^{wa}_{k} = \pi^{wa}_{k}, \quad \forall k, w, a; \quad f \in \Omega$$

where $\Omega = \{ f | f^a = x, A^f q = q^a, f^k > 0, \forall k, a \}$. To distinguish it from the Wardrop’s classic definition (Wardrop, 1952), the route flow pattern that satisfies (8) is called a multi-class percentile UE pattern.

We proceed to show how $\xi^{wa}_{k}$ may be evaluated. Let $\xi^{wa}_{k}$ be the random travel time on route $k \in K_w$. Clearly, the CDF of $\xi^{wa}_{k}$ can be constructed from the CDFs of path $k$’s member links, i.e.

$$\xi^{wa}_{k} = \sum_{a} A_{a} \xi^{wa}_{a} = \sum_{a} A_{a} g(x_a, \xi_a, \rho_a)$$

Assuming that all $\xi_a$ are independently distributed, $F_{\xi_a}(y)$ can be evaluated recursively by convolution. Let $k$ be a subpath of $k$, that is, $k = a \cup k$ where $a$ is the first link on path $k$. We have

$$F_{\xi_k}^{wa}(y) = \int_0^y \xi_a \left( F_{\xi_a}^{wa}(y) \right) \, dz$$

and $\xi^{wa}_{k} = F_{\xi_k}^{wa}(z)$. Thus, although $\xi^{wa}_{k}$ is not available in closed form in general, it can be evaluated numerically using (10).

The derivative of the random vector $\xi(f)$ with respect to the route flow $f$ can be evaluated using

$$\psi_{ki} = \frac{\partial \xi^w_{ki}}{\partial f^r_{ki}} = \sum_{a} A_{a} A_{a} \xi^a_{ki}$$

Observing the analog at the link level (cf. Eq. (7)). If this postulation is correct, it would open the door to efficient algorithms to the percentile UE problem, since the Jacobian of $\xi$ can be evaluated numerically using the recursive convolution. However, whether or not the relationship (12) is valid remains an open question, although our numerical experiments suggest that it does hold approximately.

**Proposition 2 (Monotonicity).** The mapping from route flow to the percentile route travel time, $\xi(f) : \mathbb{R}^{r \times s} \rightarrow \mathbb{R}^{s \times s}$ is monotone if (1) for each link $a$, $g(x_a, \xi_a, \rho_a)$ is a continuous and strictly increasing function of link volume $x_a$ and (2) all routes are separable (i.e., paired routes that share one or more links do not exist).

**Proof.** $\xi(f)$ is monotone if its Jacobian is positive semi-definite (PSD). The Jacobian of $\xi(f)$ has the following structure:

$$\begin{bmatrix}
\left( \frac{\partial \xi^1}{\partial f^r} \right)_{r,r} & \left( \frac{\partial \xi^1}{\partial f^r} \right)_{r,1} & \cdots & \left( \frac{\partial \xi^1}{\partial f^r} \right)_{r,s} \\
\left( \frac{\partial \xi^2}{\partial f^r} \right)_{r,r} & \left( \frac{\partial \xi^2}{\partial f^r} \right)_{r,1} & \cdots & \left( \frac{\partial \xi^2}{\partial f^r} \right)_{r,s} \\
\vdots & \vdots & \ddots & \vdots \\
\left( \frac{\partial \xi^s}{\partial f^r} \right)_{r,r} & \left( \frac{\partial \xi^s}{\partial f^r} \right)_{r,1} & \cdots & \left( \frac{\partial \xi^s}{\partial f^r} \right)_{r,s}
\end{bmatrix}$$

where $\frac{\partial \xi^i}{\partial f^j}$ is the Jacobian of class $i$’s percentile route travel time with respect to the route flow of class $j$. First, note that

$$\frac{\partial \xi^i}{\partial f^j} = \frac{\partial \xi^i}{\partial f^j} \cdots = \frac{\partial \xi^i}{\partial f^j} \quad \forall i$$

Please cite this article in press as: Nie, Y. Multi-class percentile user equilibrium with flow-dependent stochasticity. Transportation Research Part B (2011), doi:10.1016/j.trb.2011.06.001
because the same amount of route flow change of any class will have the same impact on the same $\alpha$-percentile route travel time. Second, since routes are separable, each $\frac{dt}{dx}$ is a diagonal matrix. Thus, in order to show the Jacobian is PSD, we only need to show that each $\frac{dt}{dx}$ is a PSD diagonal matrix. This, in turn, requires demonstrating all diagonal elements of $\frac{dt}{dx}$ are non-negative. Thanks to the separability, any positive increase on $f_{k}^{\text{fix}}$ will definitely increase $x_{a}, \forall a \in k$. According to Proposition 1, $t_{a}^{*} = g(x, c_{a}^{*}, \beta)$. Hence, $t_{a}^{*}$ will increase with $x_{a}$ for any $x_{a}$, because the function $g(\cdot)$ is strictly increasing with $x_{a}$. This implies that the CDF of $t_{a}$ will shift to the right. Consequently, $\int_{0}^{y} \nu_{\alpha}(z)dz$ and hence $F_{\alpha}(y)$ decrease for every $y$ because of the recursive relationship (10), which means the CDF of $t_{a}^{\text{fix}}$ will also shift to the right. Thus, for any given $x$, $\xi_{k}^{\text{fix}}$ always increases for any positive increase on $f_{k}^{\text{fix}}$.

The condition (2) in Proposition 2 is a rather strong condition that is unlikely to hold in real networks. When overlapping links exist among routes, however, proving (or disproving) monotonicity is much more challenging. Since it cannot be proven, we postulate that monotonicity does not hold in general. While the numerical experiments conducted in Section 8 seem to support this postulation, we caution that more analyses are needed to confirm the lack of monotonicity and to fully understand its implications.

The problem of finding percentile UE route flows (8) may be formulated as a variational inequality (VI) problem as follows:

$$\langle \xi(f^{*}), f - f^{*} \rangle \geq 0, \quad \forall f \in \Omega \quad (13)$$

where $\xi(f)$ is the percentile route travel time corresponding to a route flow pattern $f$. The equivalence between the VI formulation and the optimal conditions that have a complementarity structure such as (8) is well understood and hence ignored here for brevity.

5. Analytical solutions

To reveal the structure and properties of percentile UE solutions, this section analyzes a simple two-route problem. For the purpose of demonstration, the BPR-type function is employed as the link performance function, i.e.,

$$\bar{t} = g(x, \bar{c}, \beta) = v \left( 1 + \eta \left( \frac{x}{\bar{c}} \right)^{\gamma} \right), \quad x > 0 \quad (14)$$

where the set of parameters $\beta = \{v, \eta, \gamma\}$, $v$ is the free flow travel time on link $a$ and $\eta$ and $\gamma$ are non-negative scalars. To simplify the notation, the subscript $a$ is suppressed from (14). We shall ignore $a$ hereafter as long as it does not cause confusions. It is worth noting that the link travel time is a random variable in our model only when $x > 0$. If $x = 0$, the travel time always equal $v$ regardless of the realization of $\bar{c}$. Accordingly, $\bar{c}$ can be expressed as the function of $\bar{t}$, i.e.,

$$\bar{c} = g^{-1}(x, \bar{t}, \beta) = \left( \frac{\eta}{\bar{t}/v - 1} \right)^{1/\gamma} = \delta x (\bar{t} - v)^{-1/\gamma} \quad (15)$$

where $\delta = (\eta v)^{1/\gamma}$. To further simplify the analysis, we assume $\bar{c}$ follows a uniform distribution, i.e., $\bar{c} \sim U(kC_{u}, C_{u})$, where $k \in (0, 1)$, and $C_{u}$ and $kC_{u}$ denote the highest and lowest possible service flow rates (SFR), respectively. Thus, the CDF of $\bar{c}$ is

$$F_{\bar{c}}(y) = \begin{cases} 0 & y < kC_{u} \\ \frac{kC_{u}}{C_{u}(1-k)} & kC_{u} \leq y \leq C_{u} \\ 1 & y > C_{u} \end{cases} \quad (16)$$

Using the result in (2), the CDF of $\bar{t}$ can be derived as

$$F_{\bar{t}}(y) = 1 - F_{\bar{c}}(g^{-1}(x, y, \beta)) = \begin{cases} 0 & y < T_{l} \\ 1 - \frac{\delta x (y - v)^{-1/\gamma} - kC_{u}}{C_{u}(1-k)} & T_{l} \leq y \leq T_{u}, \quad x > 0 \\ 1 & y > T_{u} \end{cases} \quad (17)$$

where

$$T_{l} = v \left( 1 + \eta \left( \frac{x}{C_{u}} \right)^{\gamma} \right), \quad T_{u} = v \left( 1 + \eta \left( \frac{x}{kC_{u}} \right)^{\gamma} \right)$$

Similarly, the PDF of $\bar{t}$ can be obtained as

$$f_{\bar{t}}(y) = \frac{dF_{\bar{t}}(y)}{dy} = \begin{cases} 0 & y < T_{l} \\ \frac{\delta x}{x(1-k)} (y - v)^{-1/\gamma - 1} & T_{l} \leq y \leq T_{u}, \quad x > 0 \\ 0 & y > T_{u} \end{cases} \quad (18)$$
According to Proposition 1, the \( \alpha \)-percentile link travel time is

\[
t^\alpha = v \left( 1 + \eta \left( \frac{x}{c} \right)^\gamma \right), \quad c^\alpha = C_u(2k - \alpha + 1)
\]  

(19)

Also, the derivative of the \( \alpha \)-percentile link travel time is

\[
t^\alpha(x) = \frac{dt^\alpha(x)}{dx} = \frac{\eta \gamma}{(c^\alpha)^{\gamma-1}} x^{\gamma-1}, \quad c^\alpha = C_u(2k - \alpha + 1)
\]  

(20)

With the above stochastic performance model, we now examine the percentile UE solutions on a network with two parallel routes, each consisting of only one link. The two routes connect the same O–D pair, which has a total travel demand of \( q \).

For simplicity, let us assume that the two routes have the same maximum SFR \( C_u \), and the SFRs on routes 1 and 2 follow uniform distributions \( U[k_1C_u, C_u] \), and \( U[k_2C_u, C_u] \), respectively. Without loss of generality, assume \( 1 > k_2 > k_1 > 0 \); that is, route 2 is a more reliable option in that its SFR has a smaller variance and a higher mean. The free flow travel time on route 1 and 2 are denoted as \( m_1 \) and \( m_2 \), respectively. The other two parameters in the BPR function, \( g \) and \( c \), are identical on both routes. The desired on-time arrival probability is assumed to be uniformly distributed in the population between 0 and 1, that is, the CDF of \( \alpha \) can be written as

\[
F^\alpha(w) = w, \quad w \in [0, 1]
\]

which implies that the number of travelers whose \( \alpha \leq w \) can be simply written as \( q w \).

5.1. \( v_2 > v_1 \)

In this case, the more reliable route also has a higher free flow travel time. Fig. 1 illustrates two possible percentile UE solutions. A corner solution as shown in Fig. 1a occurs when the maximum possible travel time on route 1 is always lower than the free flow travel time on route 2, i.e.,

\[
T^\alpha_1(q) = v_1 \left( 1 + \eta \left( \frac{q}{kC_u} \right)^\gamma \right) \leq q \leq kC_u \left( \frac{v_2 - v_1}{\eta} \right)^{1/\gamma} \equiv q^*
\]  

(21)

When \( q > q^* \), both routes will be used; those who have the higher requirement for reliability will switch to route 2 earlier. At equilibrium, the flow distributions will be such that the percentile travel times on two routes are identical for a particular \( x_0 \). In this case, all travelers with an \( \alpha > x_0 \) would switch to the more reliable route (route 2) whereas all traveler with an \( \alpha < x_0 \) would remain on the less reliable route (route 1). As visualized in Fig. 1b, \( x_0 \) corresponds to the intersection of the CDFs of the two route travel times. It is worth noting that the optimal choice for users with all possible on-time arrival probabilities constitutes a so-called Pareto frontier, which contains the lower segment of route 1’s CDF and the upper segment of route 2’s CDF. Generally, percentile UE solutions must be on such frontiers. \( x_0 \) can be solved from the following equations

\[
T^\alpha_1 = v_1 \left( 1 + \eta \left( \frac{x_1}{c_1^\alpha} \right)^\gamma \right) = v_2 \left( 1 + \eta \left( \frac{x_2}{c_2^\alpha} \right)^\gamma \right) = T^\alpha_2 \equiv x_1 + x_2 = q
\]  

(22)

Fig. 1. Different percentile UE solution patterns for the two-route network \( (v_1 < v_2) \).
where $c_0^a$ is given in Eq. (19). Eq. (24) above requires that all users with an on-time arrival probability less than $\alpha_0$ stay on link 1. To show this is always the case, let's first define $p(\alpha_0) = t_1^\alpha - t_2^\alpha$.

**Lemma 1.** If $\alpha_0$ is a solution to Eqs. (22)–(24), the function $p(\alpha_0)$ is always a strictly increasing function of $\alpha_0$ when $v_1 < v_2$.

**Proof.** See Appendix A. ⊓⊔

**Proposition 3.** Let $x$ be a solution to Eqs. (22)–(24) and $T^a_\alpha(x_\alpha)$ and $T^a_{\alpha_2}(x_\alpha)$ be the maximum and minimum realization of travel time on route $a$, $a = 1, 2$ at equilibrium, respectively. If $v_1 < v_2$, then $T^a_1(x_1) > T^a_2(x_2)$: $T^a_1(x_1) < T^a_2(x_2)$.

**Proof.** See Appendix A. ⊓⊔

Note that Proposition 3 implies that $p(0) < 0$ and $p(1) > 0$. It follows from Lemma 1 that there is one and only one $\alpha_0 \in [0, 1]$ such that $p(\alpha_0) = 0$, which leads to Eq. (22). Therefore, the pattern depicted in Fig. 1b is the only possible solution when $q > q^\alpha$.

5.2. $v_1 > v_2$

In this case, the more reliable route is also faster under the free flow traffic condition. Thus, travelers will first use route 2 as long as

$$T^a_2(q) = v_2 + \left(1 + \eta \left(\frac{q}{KCu}\right)^\gamma\right) \leq v_1 \Rightarrow q \leq kC_u \left(\frac{v_1 - v_2}{\eta}\right)^{1/\gamma} \equiv q^*$$

When $q > q^*$, both routes will be used. One can solve an equation system similar to (22)–(24) to obtain $\alpha_0$, as well as $x_\alpha$, $a = 1, 2$. However, $p(\alpha_0)$ is not necessarily an increasing function in this case. Moreover, the sign of $p(0)$ and $p(1)$ cannot be determined independent of the parameters. Consequently, other than the corner solution there are multiple possible solution patterns, which include, but may not be limited to, the three cases as shown in Fig. 2. Note that in each case, the Pareto frontier consists of different segments of the two CDFs. Determining which pattern emerges at the percentile UE a priori is difficult. However, a trial-and-error approach may be sufficiently effective in simple cases. That is, one can construct equations according to one of the possible patterns, and check whether or not the solution obtained from the equations agree with the assumption. For example, for the case shown in Fig. 2c, the percentile travel time on two routes will be identical for two different values of $\alpha$, namely $\alpha_1$ and $\alpha_2$ with $0 < \alpha_1 < \alpha_2 < 1$. This indicates that travelers with $\alpha < \alpha_1$ or $\alpha > \alpha_2$ (i.e., those with high or low reliability requirement) will stay on route 2; and those with medium reliability requirement ($\alpha_1 \leq \alpha \leq \alpha_2$) will switch to route 1. To solve for $\alpha_1$ and $\alpha_2$, let $x_1^a$ and $x_2^a$ be the number of travelers with $\alpha > \alpha_2$ and $\alpha < \alpha_1$, respectively. Thus, $x_1^a + x_2^a = x_2$. The percentile equilibrium solution can be solved from the following equation system.

![Figure 2](image-url) Fig. 2. Different percentile UE solution patterns for the two-route network ($v_1 > v_2$).
The cases (a) and (c) in Fig. 2 reveal that a more reliable route may not always be used by those who have higher reliability requirement at percentile UE. Essentially, if a reliable route is “slow”, it would only attract travelers with high reliability requirement; if a reliable route is also “fast”, it may be attractive to all travelers.

Remark

More complicated solution patterns may arise when more than two values of \( x \) can equilibrate the percentile travel time on both routes. It is suspected that such cases may be rare but the current analysis does not exclude them. These cases should be considered only if the primary patterns shown in Fig. 2 do not yield a correct percentile UE solution.

6. Flow-dependent stochasticity

The distribution of the link SFR may be flow-dependent for a couple of reasons. First, many studies show that the relationship between hourly accident rates and volume to capacity (v/c) ratio follows a U-shaped pattern, and that serious accidents tend to decrease while the v/c ratio increases (Ceder and Livneh, 1982; Zhou and Sisiopiku, 1997; Lord et al., 2005). Because accidents may sharply reduce SFR, the distribution of SFR is likely to be affected by accident rates, and thereby dependent on traffic flow levels. Another flow-dependent impact on SFR distribution has to do with a phenomenon known as traffic breakdown. Empirical evidence suggests that abrupt increase of travel times is often observed in accord with the onset of congestion, or traffic breakdown. Traffic breakdown does not necessarily occur at the same location, nor always at the same service flow rate (Persaud et al., 1998; Kerner, 2000; Evans et al., 2001). Generally speaking, however, traffic breakdown is more likely to occur at higher levels of traffic flow. That is, flow levels affect traffic breakdown probability, which in turn affects SFR distributions.

To capture the flow dependence of SFR distribution, the CDF of the SFR is modeled in this paper using a rational function as follows:

\[
F_q(w) = \frac{(1 + \rho)(w - kC_u)}{\rho(w - kC_u) + (1 - k)C_u} \tag{30}
\]

Note that this function is concave when \( \rho > 0 \) and convex when \(-1 \leq \rho \leq 0\). We shall further assume that \( \rho \) is flow dependent, specifically,

\[
\rho(x) = \frac{\theta x}{C_u} - \sigma \tag{31}
\]

where \( \theta \geq 0, \sigma \in [-\infty, 1) \) are external parameters that may be calibrated to determine the level of flow-dependence of the SFR distributions. When \( \theta = 0, \sigma = 0 \), the distribution follows a uniform distribution from \( kC_u \) to \( C_u \) and is not dependent on link flows. When \( \sigma > 0/kC_u \) (i.e., \( \rho(x) < 0 \)), the distribution is skewed to the higher end, implying that the random SFR is more likely to take a value closer to the upper bound \( C_u \); when \( \sigma < 0/kC_u \) (i.e., \( \rho(x) > 0 \)), the opposite is true: the resulting random SFR is more likely to take a value closer to the lower bound \( kC_u \). In summary, the intuition behind Eq. (31) is that the reliability of SFR is likely to decrease as the flow level increases. To see this, note that when \( \theta > 0 \), the concavity of the CDF modeled by the rational function always increases with \( x \), which renders the resulting distribution to skew to the lower bound. Fig. 3 demonstrates how the parameter \( \theta \) interacts with the level of traffic flow \( x \) in the CDF based on the above rational function.

Using the BPR function (14), the percentile link travel time associated with the CDF (30) can be written as

\[
t^q = v \left( 1 + \eta \left( \frac{x}{C_v(x)} \right)^\frac{1}{\beta} \right) \quad \text{and} \quad c^q(x) = \frac{C_u(\alpha k + 1 - x - k \alpha \sigma) + \alpha k \theta x}{1 - \alpha \sigma + \alpha \theta x/C_u} \tag{32}\]

It is easy to verify that the above formula collapses to Eq. (19) when \( \theta = 0, \sigma = 0 \). Also, the function \( c^q(x) \) is a monotone decreasing function for \( x \geq 0 \). To see this, note that the derivative of \( c^q(x) \) is

\[
\frac{dc^q(x)}{dx} = \frac{\alpha \theta (k - 1)(1 - x)}{(1 - \alpha \sigma + \alpha \theta x/C_u)^2} \leq 0 \tag{33}
\]

The inequality holds because \( k \leq 1, \theta \geq 0 \) by assumption. Hence, the percentile link travel time is an increasing function of \( x \). The percentile derivative of link travel time can be derived using Eq. (7)
shown in Fig. 2c, it can still be solved from Eqs. (26)–(29) in the case of flow-dependent stochasticity, provided that network problem discussed before as an example. If the percentile UE solution is known to correspond to a Pareto frontier as (Bertsekas, 1976; Jayakrishnan et al., 1994). The GP algorithm is described in the following.

presented in Section 4, inspired by the application of the same algorithm in the standard traffic assignment problem (Bertsekas, 1976; Jayakrishnan et al., 1994). The GP algorithm is described in the following.

7. Solution algorithm

In this section a gradient projection (GP) algorithm is proposed to solve the VI formulation for the percentile UE problem presented in Section 4, inspired by the application of the same algorithm in the standard traffic assignment problem (Bertsekas, 1976; Jayakrishnan et al., 1994). The GP algorithm is described in the following.

Gradient projection algorithm

- **Step 0**: Initialization. Assign all O–D flows to shortest paths to obtain an initial flow pattern, set \( \lambda = 1.0 \), \( \epsilon_1 = \infty \).
- **Step 1**: Flow and distribution update. Update link flows, the PDF of all link travel times using (3), and the PDF of all link travel time derivatives using (5).
- **Step 2**: Equilibration. Set convergence measure \( \epsilon = 0.0 \). For each O–D pair \( w \) and each user class \( x \):
  - **Step 2.1.** For each path \( k \), evaluate the percentile route travel time \( x_k^w \) using the recursive relationship (10). Find \( \tilde{k} = \arg\min_k (x_k^w) \).
  - **Step 2.2.** For each path \( k \neq \tilde{k} \), set \( G_k^w = x_k^w - x_{\tilde{k}}^w \), and \( H_k = \frac{\partial x_k^w}{\partial \rho} \).
  - **Step 2.3.** Update path flows, set
    \[
    f_k^w = f_k^w - \rho G_k / H_k, \quad \forall k \neq \tilde{k}, \quad f_{\tilde{k}}^w = q_{\tilde{k}}^w - \sum_{k \neq \tilde{k}} f_k^w
    \]
    \[
    (35)
    \]
  - **Step 2.4.** Set
    \[
    \epsilon = \epsilon + \left( \sum_k f_k^w / H_k \right) / \left( q_{\tilde{k}}^w / H_k \right)
    \]
    \[
    (36)
    \]
- **Step 3**: Convergence test. If \( \epsilon < \epsilon_0 \) (\( \epsilon_0 \) is a predetermined convergence criterion), terminate the procedure; otherwise, if \( \epsilon > \epsilon_1 \), set \( \lambda = 0.8\lambda \), update \( \epsilon_1 = \epsilon \), go to Step 1.
Remark 1. In Step 1, the PDFs of link travel times and derivatives have to be computed and stored in a discrete form. In this study, we use the percentile link travel time function \( \bar{t}^\alpha \) to construct a discrete CDF and then convert CDF to PDF.

Remark 2. In Step 2.3 is a step size, which is initialized as a constant 1.0, and then reduced by a constant scalar whenever the convergence measure increases from one iteration to the next. The constant scalar is set to 0.8 in our implementation, based on the previous computational experience with the GP algorithm.

Remark 3. To approximate \( \frac{\partial \bar{t}^\alpha_k}{\partial f} \) in Step 2.2, note that

\[
\frac{d(\bar{t}^\alpha_k - \bar{t}^\alpha_{\bar{k}})}{df_k} = \sum_a (A_{ak} - A_{\bar{a}k})(A_{\bar{a}k} - A_{ak})f_a 
\]

(37)

Thus, we can evaluate the CDF of route travel time derivative over a dummy route using

\[
F_{\bar{y}l}(y) = \int_0^y \nu_{\bar{a}l}(z)F_{\bar{y}l}(y - z)dz 
\]

(38)

where dummy route \( l \) consists of all links that belong to either route \( k \), or route \( \bar{k} \), but not both. Then

\[ H_k = F_{\bar{y}l}^{-1}(x) = \frac{\partial \bar{t}^\alpha_k}{\partial f} \]

\( H_k \) above is proposed as an approximation to the diagonal element of the percentile route travel time’s Jacobian (note that the percentile route travel time is defined on a reduced route flow space, i.e., for all routes \( k \neq \bar{k} \)). However, the quality of this approximation largely depends on the validity of Eq. (12), which has yet to be proven. Until that equation is fully verified, the above algorithm should be considered as a heuristic procedure.

Remark 4. The algorithm presented above does not have a route generation module. In practice, since routes may not be enumerated, they have to be augmented iteratively, with routes that have the smallest current percentile travel times. The smallest percentile travel time routes may be found by generating all non-dominated (admissible) routes under the first-order stochastic dominance rule (Nie and Wu, 2009; Wu and Nie, 2009). While the problem is generally NP hard, pseudo-polynomial heuristics can be developed to obtain high quality approximated solutions quickly, even on fairly large networks (Nie et al., in press). Combining the reliable routing algorithms with the above assignment procedure is left to a more computation-oriented study.

8. Numerical results

Numerical experiments are conducted in this section in order to verify the analytical results and examine the convergence performance of the GP algorithm. The GP algorithm was coded and run in MATLAB7 on a Dell Optiplex 960 desktop (with a 2.83 GHz Intel Q9550 CPU and 8G RAM).

8.1. Jacobian of the percentile route travel time

We first examine the Jacobian of the percentile route travel time \( \bar{t} \). As explained before, when a route contains more than one link, the derivative of its percentile travel time does not have a closed form. Two numerical methods are thus implemented and compared here. The first approach is based on brute-force perturbation, that is, perturbing the route flow \( f_w = f_{w'} \) by \( \delta f \) and approximate the derivatives by

\[
\frac{\partial \bar{t}^{w^\prime}}{\partial f} = \frac{\bar{t}^{w^\prime} - \bar{t}^{w}}{\delta f}, \quad \forall w, w', l
\]

where \( \bar{t}^{w^\prime} \) is the percentile route travel time after perturbation on path \( k \) for O–D pair \( w \) and class \( a \). This method is neither efficient (since the percentile travel time has to be evaluated \( r \) times for each route, where \( r \) is the number of routes) nor numerically stable. The second approach evaluates the Jacobian using Eqs. (11) and (12), which only requires computing the percentile derivative of travel time once for each route. The problem associated with this approach, as mentioned before, is that the relationship (12) remains to be validated. In what follows, \( J_e \) and \( J_h \) refer to the approximated Jacobians produced by the first and second methods, respectively.

In the example shown in Fig. 4, the performance function takes the form \( (14) \), where \( \nu \) on each link is given in Fig. 4, \( \eta = 1 \), \( \gamma = 4 \), and \( C \sim U(c_c,k_c) \). The Jacobian is evaluated at \( f = [0.75, 0.75, 0.75, 0.75] \) for different values of \( \alpha \) \((0.5, 0.9)\), \( \delta f \) \((0.1, 0.05, 0.01, 0.005, 0.001)\) and discretization schemes \((M = 1000, 5000, 10000, 50000, 100000)\), where \( M \) is the number of discrete support points. Ideally, the perturbation method can produce better approximations for smaller values of \( \delta f \). However, since the evaluation of percentile route travel times is subject to discretization errors in computing convolution, the estimation quality of the perturbation method is also affected by the discretization parameters such as \( M \). Intuitively, for smaller \( \delta f, M \) should be increased to achieve the same estimation quality.
When \( \alpha = 0.9, \delta f = 0.001, M = 100000 \),

\[
J_E = \begin{pmatrix}
3.42 & 1.72 & 1.72 & 0.00 \\
1.38 & 82.39 & 0.00 & 81.01 \\
1.90 & 0.00 & 2.27 & 0.38 \\
0.00 & 81.01 & 0.33 & 81.34
\end{pmatrix}, \quad J_A = \begin{pmatrix}
3.24 & 1.87 & 1.87 & 0.00 \\
1.87 & 82.20 & 0.00 & 80.93 \\
1.87 & 0.00 & 2.10 & 0.63 \\
0.00 & 80.93 & 0.63 & 81.16
\end{pmatrix}
\] (39)

Although the two matrices are similar to each other, noticeable differences do exist. While it is difficult to determine which one is actually closer to the “true” Jacobian (since we do not have a validated closed form), \( J_E \) is evidently problematic because it fails to meet the symmetric condition, likely due to numerical errors. This comparison indicates that the heuristic method based on Eqs. (11) and (12) is a competitive alternative to the brute-force approach: it provides comparable if not better estimation of the Jacobian with much less computational efforts.

Fig. 5 reports how the relative discrepancy between \( J_E \) and \( J_A \) changes with \( \delta f \) and \( M \) for two different \( \alpha \). The relative discrepancy is measured using

\[
\phi(J_E, J_A) = \frac{|J_E - J_A|_{\infty}}{|J_A|_{\infty}}
\]

where \( |A|_{\infty} \) denotes the maximum row sum of matrix \( A \). In the case of \( \alpha = 0.9 \), \( J_E \) evidently converges to \( J_A \) as \( \delta f \) changes from 0.1 to 0.001. More importantly, in all three cases, the discrepancy becomes stabilized after \( M \) reaches certain threshold, indicating that the impacts of \( M \) have been properly controlled. The case of \( \alpha = 0.5 \) is less conclusive, probably because the sharp changes of route travel time CDF in the neighborhood of \( \alpha = 0.5 \) demands larger \( M \) to achieve the same level of precision.

Nevertheless, the general trend is still similar, that is, the two matrices tend to converge to each other as \( M \) increases and \( \delta f \) decreases. Since \( J_E \) approaches to the true Jacobian as \( \delta f \to 0 \) (by definition), the above results suggest that \( J_A \) is likely to be a good approximation to the true Jacobian.

Finally, the minimum eigenvalue of \( J_E \) and \( J_A \) in Eq. (39) are \(-0.0068\) and \(-0.476\), respectively, which suggests that the percentile route travel time in this example may not be a monotone function of the route flow.

8.2. Two-route network

This section verifies the analyses for the two-route network problem discussed in Section 5 with numerical results. In this experiment, the demand \( q \) is set to 4, the maximum SFR \( C_u \) is set to 2, and \( \eta \) and \( \gamma \) in the BPR function are set to 0.15 and 4,
For brevity we only test the more complicated case when $k_2 > k_1$ and $v_2 < v_1$ (i.e., route 2 has a more reliable SFR and is faster at the free flow condition). Three sets of values are selected for parameters $v_a$ and $k_w$, $a = 1, 2$, each leading to a different equilibrium pattern (cf. Table 1). Eq. (30) is used to model the distribution of SFR. Two cases are considered regarding flow-dependent stochasticity: the case with flow-dependent stochasticity, in which $\theta$ and $\sigma$ are set to 3 and 0.8 respectively for route 1 (the less reliable route) and 0 for route 2; the case without flow-dependent stochasticity, in which case $\theta$ and $\sigma$ are 0 for both routes (the distribution is reduced to a uniform distribution).

Table 1 reports the flow distributions at percentile UE for the three parameter sets. Plots in Fig. 6 compare the CDFs of the travel times on both routes at percentile UE. In Set I, the travelers with higher reliability requirement always use the more reliable route (route 2). This result is reversed when the parameters in Set II are used: the travelers with higher reliability requirement always use the less reliable route (route 1). In Set III, the more reliable route is occupied by travelers with either high or low $\alpha$. It is clear that the percentile travel time are identical wherever the two CDFs intersect each other, and travelers with $\alpha$ corresponding to theses intersections are “indifferent” to the two routes in that their percentile travel times are identical.

Fig. 6 shows that the equilibrium flow patterns are different when route 1’s SFR is subject to flow-dependent stochasticity. In particular, the CDF of the travel time on route 1 becomes “less concave” with flow-dependent stochasticity. In Set II, it is actually changed from a concave function to a convex function. Such a change is expected because considering flow-dependent stochasticity tends to increase the concavity of the SFR’s CDF for the selected values of $\sigma$ and $\theta$ (cf. Fig. 3). When concavity of the SFR’s CDF increases, the probability concentration on the low end of the feasible SFR range becomes greater. Therefore, the route becomes less attractive because it is more likely to yield high travel time within the feasible range. This explains why flows on route 1 decrease in all three parameter sets when its SFR distribution is made flow-dependent.

8.3. General network

We now test the proposed model algorithm on a general problem, adopted from Lo et al. (2006). For the convenience of the reader, the network topology and other inputs of the problem are reproduced in Fig. 7. To be consistent with Lo et al. (2006), the BPR function (14) is again adopted in this study, with $\eta$ and $\gamma$ set to 1 and 4, respectively. Three classes of users are considered in this example, with on-time arrival probability $\alpha = 0.504, 0.805, 0.957$, respectively. The share of each class in the population is $0.495, 0.38, 0.125$, respectively, which was obtained from a survey conducted by Lo et al. (2006). The parameters of link properties are such selected that the faster routes (such as those using links 2, 4, 5, 6, 7) are less reliable than the slower ones (such as those using links 1, 3, 8, 9). Further, the CDF of the link SFRs takes the form of (30). Recall that when $\theta = 0, \sigma = 0$, the CDF (30) is reduced to a linear form corresponding to the uniform distribution, which is consistent with the assumption made in Lo et al. (2006). This is treated as benchmark and compared against two other cases with flow-dependent distributions on links 2, 4 and 5. The first case sets $\theta = 0.2, \sigma = -3$, and hence produces concave CDFs for SFR on these links ($\rho(x) > 0$); in the second case, $\theta = 0.1, \sigma = 0.8$, which results in convex CDFs for SFR unless $x$ is very large. These two cases are referred to as Concave- and Convex-case respectively. Note that a concave CDF is less reliable because the random SFR that follow such a CDF is more likely to take a value closer to the lower bound. In the following tests, the number of discrete points $M = 200$ and the convergence criterion $\varepsilon = 1e^{-8}$.

Table 2 reports the percentile UE solution for the benchmark problem. Noticeable differences are observed between the above solution and the solution given in Lo et al. (2006) (Table 5 on Page 803). For example, the majority of the Class-2 users between O-D pair 1–6 take route 5 according to the above solution, whereas the solution from Lo et al. (2006) does not assign any Class-2 users to this route. A close look reveals that the two solutions actually produce quite different total link flows. For example, the flow on link 2 is 12.159 in our solution but only 0.5 in the solution given by Lo et al. (2006). Such discrepancies are likely due to the fact that the model of Lo et al. (2006) relies on the central limit theorem to approximate

---

5 Note that the increase in concavity of the SFR’s CDF generally corresponds to the increase in convexity of the travel time’s CDF, which can be shown mathematically using Eq. (3).

6 These on-time arrival probabilities correspond to the punctual arrival factor $\lambda = 0.1, 0.86$, and 1.72, respectively (cf. Page 803 Lo et al., 2006).
the random route travel time distributions, which could generate large errors in such a small problem. As shown in Fig. 8, the route travel time distributions are clearly not good approximations of normal distributions.

Table 3 reports the percentile UE solution for the Concave-case, i.e., when $h = 0.2$, $r = -3$ on links 2, 4, and 5. Compared to the benchmark, the Class-1 users shift from route 2 to routes 1 and 3, from route 6 to route 7, and from route 9 to 8. The overall usage of route 1 increases, because the alternatives become less desirable due to the introduced concavity in the CDF of the SFR on links 2, 4 and 5. Also, most travelers are worse off, in that they have to budget more time to achieve the same on-time arrival probability, with the exception of the Class-3 travelers between O–D pairs 1–6 and 5–2.

Table 4 reports the percentile UE solution for the Convex-case, i.e., when $h = 0.1$, $r = 0.8$ on links 2, 4, 5. In this case, the usage of route 3 is significantly increased by the Class-1 users diverting from other routes. This is expected since the links that constitute route 3 now become more reliable (in the sense that their SFRs are more likely to realize higher values within its feasible range). Also expected are the reduced percentile travel time for those who travel to node 2 and hence benefit from

![Fig. 6. CDFs of the route travel times at percentile UE.](image-url)
the improvement of reliability on route 3 (which leads to node 2) through the introduced flow-dependent stochasticity. Interestingly, for those who travel to node 6, the required time budget is slightly increased, except for the Class 3 users between O–D pair 1–6. This is likely due to that the induced flow changes render the routes leading to node 6 less reliable.

Fig. 8 compares the equilibrium travel time distributions on the first three routes in the three cases. Note that in the Concave-case, the CDF of the route travel time became "less concave" compared to the benchmark - the travel times on these routes are almost uniformly distributed. On the contrary, the concavity of the route travel time CDFs increases in the Convex-case (i.e. the CDF of the SFR is convex). As explained in Footnote 5, the increase in concavity of the SFR's CDF corresponds to the increase in convexity of the travel time's CDF because of the relationship (3).
Finally, the convergence performance of the GP algorithm in solving all the three cases is reported in Fig. 9. The left panel of Fig. 9 indicates that the step size did not decrease endlessly, but rather stabilized at a reasonably high level (larger than 0.2 in all cases). Furthermore, the right panel shows that the algorithm converged to the percentile UE quickly, especially after the step size is stabilized. The CPU time per iteration is approximately 0.06–0.07 s in the three tests with $M = 200$. 

**Table 4**
Percentile UE solution for the Convex-case ($\theta = 0.1$, $\sigma = 0.8$).

<table>
<thead>
<tr>
<th>O–D</th>
<th>Route</th>
<th>Class-1 ($\alpha = 0.504$)</th>
<th>Class-2 ($\alpha = 0.805$)</th>
<th>Class-3 ($\alpha = 0.957$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Flow</td>
<td>Percentile time</td>
<td>Flow</td>
<td>Percentile time</td>
</tr>
<tr>
<td>1–2</td>
<td>1</td>
<td>8.354</td>
<td>11.4</td>
<td>184.486</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>307.085</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6.496</td>
<td>0</td>
<td>322.676</td>
</tr>
<tr>
<td>1–6</td>
<td>4</td>
<td>14.85</td>
<td>7.241</td>
<td>456.699</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0</td>
<td>4.159</td>
<td>456.699</td>
</tr>
<tr>
<td>5–2</td>
<td>6</td>
<td>1.828</td>
<td>11.4</td>
<td>304.335</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>13.022</td>
<td>0</td>
<td>322.331</td>
</tr>
<tr>
<td>5–6</td>
<td>8</td>
<td>10.063</td>
<td>11.4</td>
<td>250.908</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>4.787</td>
<td>0</td>
<td>454.061</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>456.366</td>
</tr>
</tbody>
</table>

**Fig. 8.** Travel time distributions on routes 1, 2 and 3 at percentile UE.
9. Conclusions

This paper studies a variant of the reliability-based traffic assignment model, in which travelers seek to minimize the percentile travel time, or the travel time budget that ensures their own desired on-time arrival probability. The stochasticity in the supply of transportation is incorporated by modeling the service flow rate (SFR) of each road segment as a random variable. The probability density functions of SFR are specified exogenously as rational functions, of which the CDF of uniform distribution is a special case. The parameters in these functions are assumed to be flow-dependent, and their relationship is set such that the reliability of SFR decreases when the flow level increases. While not attempted in this study, these parameters may be estimated from traffic data. The link travel time distribution is derived from the associated SFR distribution through a link performance function. By directly convolving these link travel time distributions, the distribution of route travel time is obtained. Note that this approach is more flexible than relying on the central limit theorem to add random distributions, because it applies to any distribution and problem size.

The percentile route time function derived from convolution does not have a closed form. Although we showed that this function is monotone with respect to route flow in a restrictive case, we postulated that monotonicity does not hold in general. While our experiments do support this postulation, more analyses are needed for further proof or disproof. Numerical schemes are proposed to evaluate the percentile route travel time and to approximate its Jacobian through recursive convolution. Although direct convolution is a computationally challenging procedure in large networks, several strategies may be exploited to improve its efficiency. For example, the structure of the partially overlapping routes can be used to combine convolutions carried out over them. If the routes are consolidated as an acyclic subnetwork, the convolution may be executed following a topological order over the subnetwork. Finally, efficient convolution algorithms such as those based on Fast Fourier Transform (FFT) can also be employed.

The solution properties of the percentile UE problem are analyzed using a two-route network, assuming a continuous distribution for on-time arrival probability. It is revealed that the percentile UE solutions can be found from a Pareto frontier of the CDFs of travel times on all used routes. Also, the route with more reliable SFR may not always be chosen by those who have high reliability requirements, particularly when such a route has low free flow travel time and hence tend to attract more users. These results are verified by numerical experiments. To solve the general percentile UE problem, we first formulate it as a VI problem and then propose a route-based gradient projection algorithm. Similar to most route-based traffic assignment algorithms, the proposed algorithm organize assignment by O–D pair in the spirit of Gauss–Seidel decomposition, by shifting flows to the minimum percentile travel time route from the other routes. An important feature of our algorithm is the use of the diagonal elements in the Jacobian of percentile route travel to scale the flow shift, which mimics the Quasi-Newton method for the classic traffic assignment problem. In the preliminary experiments, the algorithm demonstrated satisfactory convergence performance and was able to achieve highly precise equilibrium solutions in all test scenarios. The next step is to combine the assignment algorithm with a column generation scheme, which iteratively solves the shortest path problem considering on-time arrival probability, and to conduct more comprehensive computational studies. Finally, it is unclear whether or not the proposed method for evaluating the derivative of percentile travel time is applicable when the link performance function is non-separable. If the method cannot be directly extended to handle the non-separable case, assignment algorithms that obviate derivative information, such as those based on projection, may be considered. We leave this also to the future studies.

Acknowledgements

Professor Hani Mahmassani at Northwestern University had provided many valuable comments. In particular, he convinced me that the service flow rate is a better term than “capacity” for the purpose of this study. At his suggestion, the cur-
rent manuscript explicitly distinguishes the service flow rate, on which the calculation of travel time is supposed to be based, from capacity, which is typically defined as the maximum hourly volume that has a reasonable expectation of occurrence. Comments of an anonymous referee have helped to improve the analytical results presented in Section 5. The research was supported by National Science Foundation under the Award Number CMMI-0928577.

Appendix A

Proof of Lemma 1.

Note that \( p(x_0) = t^{x_0}_1 - t^{x_0}_2 \), and

\[
\frac{dt^{x_0}_1}{dx_0} = v_1 \eta \left( \frac{x_1}{C_1} \right)^\gamma (1 - \alpha + \alpha k_1)^{-1} (1 - k_1)
\]

Because \( k_1 < k_2 \), we have \( 1 - \alpha + \alpha k_1 < 1 - \alpha + \alpha k_2 \rightarrow (1 - \alpha + \alpha k_1)^{-1} > (1 - \alpha + \alpha k_2)^{-1} \) and \( 1 - k_1 > 1 - k_2 \). Moreover, \( v_1 < v_2 \) and Eq. (22) imply

\[
v_1 \eta \left( \frac{x_1}{C_1} \right)^\gamma > v_1 \eta \left( \frac{x_2}{C_2} \right)^\gamma
\]

Therefore,

\[
\frac{dp(x_0)}{dx_0} = \frac{dt^{x_0}_1}{dx_0} - \frac{dt^{x_0}_2}{dx_0} > 0.
\]

This proves that \( p(\cdot) \) is a strictly increasing function of \( x \). \( \square \)

Proof of proposition 3.

Note that

\[
T^*_1(x_1) > T^*_2(x_2)
\]

\[
\begin{align*}
&\iff v_1 \left( 1 + \left( \frac{\alpha_0 k_1 - \alpha_0 + 1}{k_1} \right) \right) \eta \left( \frac{x_1}{C_1} \right)^\gamma > v_2 \left( 1 + \left( \frac{\alpha_0 k_2 - \alpha_0 + 1}{k_2} \right) \right) \eta \left( \frac{x_2}{C_2} \right)^\gamma \\
&\iff \left( \frac{\alpha_0 k_1 - \alpha_0 + 1}{k_1} \right) \eta \left( \frac{x_1}{C_1} \right)^\gamma > \frac{\alpha_0 k_2 - \alpha_0 + 1}{k_2} \eta \left( \frac{x_2}{C_2} \right)^\gamma \\
&\iff \left( \frac{\alpha_0 k_1 - \alpha_0 + 1}{k_1} \right) \eta \left( \frac{x_1}{C_1} \right)^\gamma > \left( \frac{\alpha_0 k_2 - \alpha_0 + 1}{k_2} \right) \eta \left( \frac{x_2}{C_2} \right)^\gamma
\end{align*}
\]

Therefore,

\[
\frac{dt^{x_0}_1}{dx_0} = \frac{dt^{x_0}_2}{dx_0} > 0.
\]

because \( v_1 < v_2 \) and (22). Finally, since \( k_1 < k_2 \rightarrow (\alpha_0 k_1 - \alpha_0 + 1)/k_1 > (\alpha_0 k_2 - \alpha_0 + 1)/k_2 > 1 \), the inequality (40) always holds. Similarly, \( T^*_1(x_1) < T^*_2(x_2) \) requires

\[
\begin{align*}
&v_1 \left( 1 + \left( \frac{\alpha_0 k_1 - \alpha_0 + 1}{k_1} \right) \right) \eta \left( \frac{x_1}{C_1} \right)^\gamma < v_2 \left( 1 + \left( \frac{\alpha_0 k_2 - \alpha_0 + 1}{k_2} \right) \right) \eta \left( \frac{x_2}{C_2} \right)^\gamma \\
&\iff \left( \frac{\alpha_0 k_1 - \alpha_0 + 1}{k_1} \right) \eta \left( \frac{x_1}{C_1} \right)^\gamma < \left( \frac{\alpha_0 k_2 - \alpha_0 + 1}{k_2} \right) \eta \left( \frac{x_2}{C_2} \right)^\gamma \\
&\iff \left( \frac{\alpha_0 k_1 - \alpha_0 + 1}{k_1} \right) \eta \left( \frac{x_1}{C_1} \right)^\gamma < \left( \frac{\alpha_0 k_2 - \alpha_0 + 1}{k_2} \right) \eta \left( \frac{x_2}{C_2} \right)^\gamma
\end{align*}
\]

This inequality also always holds because \( k_1 < k_2 \rightarrow \alpha_0 k_1 - \alpha_0 + 1 < \alpha_0 k_2 - \alpha_0 + 1 \rightarrow (\alpha_0 k_1 - \alpha_0 + 1)^\gamma - 1 < (\alpha_0 k_2 - \alpha_0 + 1)^\gamma - 1 < 0 \). \( \square \)

References


Please cite this article in press as: Nie, Y. Multi-class percentile user equilibrium with flow-dependent stochasticity. Transportation Research Part B (2011), doi:10.1016/j.trb.2011.06.001


Y. Nie / Transportation Research Part B xxx (2011) xxx–xxx

Please cite this article in press as: Nie, Y. Multi-class percentile user equilibrium with flow-dependent stochasticity. Transportation Research Part B (2011), doi:10.1016/j.trb.2011.06.001