A Semi-Analytical Approach for Solving the Bottleneck Model with General User Heterogeneity

Yang Liu
Department of Civil and Environmental Engineering
Northwestern University, 2145 Sheridan Road, Evanston, IL 60208
Email: yangliu@u.northwestern.edu

Yu (Marco) Nie*
Department of Civil and Environmental Engineering
Northwestern University, 2145 Sheridan Road, Evanston, IL 60208
Email: y-nie@northwestern.edu
Phone: 1-847-467-0502

Jonathan Hall
Department of Economics
Northwestern University, 2001 Sheridan Road, Evanston, IL 60208
Email: jhall@northwestern.edu

October 22, 2013

Abstract
This paper proposes a novel semi-analytical approach for solving the dynamic user equilibrium (DUE) of a bottleneck model with general heterogenous users. The proposed approach makes use of the analytical solutions from the bottleneck analysis to create an equivalent assignment problem that admits closed-form commute cost functions. More specifically, the underlying assignment problem is a static and asymmetric traffic assignment problem, which may be formulated and solved as a variational inequality problem (VIP). This approach provides a new tool to understand the analytical properties of the bottleneck model with general heterogeneity, and the existence and the uniqueness of DUE solution is proved by examining the P-property of the Jacobian matrix. It is also a useful tool to conduct numerical analysis for congestion management policies. The numerical experiments show that a decomposition algorithm is able to quickly solve the equivalent VIP to high precision. Thus, the proposed approach may be used to perform numerical analysis for the bottleneck models that are analytically intractable. It is also extended to address simultaneous departure time and route choice in a single O-D network with multiple parallel routes.

*Corresponding author
Keywords: bottleneck model; general heterogeneity; dynamic user equilibrium; variational inequality problem


1 Introduction

The analysis of dynamic commute travel patterns, originated by Vickrey (1969a) and refined subsequently in Hendrickson and Kocur (1981); Smith (1984); Daganzo (1985); Newell (1987); Arnott et al. (1990, 1994); Yang and Huang (1997), has been extensively studied in the literature. The underlying assumption in these analyses is that travelers make trade-off between the anticipated costs of travel time and schedule delay (incurred when travelers cannot arrive at their destination at a desired time). Accordingly, the general pattern of commuters’ departure time choices is explained as a dynamic user equilibrium at which nobody can reduce his/her own commute cost by unilaterally shifting his/her route and departure time choice. Although it provides a stunning example in which the Pigouvian toll (Pigou, 1920) completely eliminates congestion at nobody's cost, Vickrey’s original bottleneck model ignored user heterogeneity in the valuation of travel time and schedule delay. Such heterogeneity has since attracted much attention because of its potential impacts on the equilibrium solutions as well as welfare effects of various demand management policies (Small, 1982; Cohen, 1987; Arnott et al., 1994; Lindsey, 2004; Small et al., 2005; van den Berg and Verhoef, 2011; Liu and Nie, 2011; Hall, 2013).

It is well known that analytical solutions for the bottleneck model with user heterogeneity exist in special cases (e.g., for a small number of groups, or assumptions are imposed on how the value of time and schedule cost parameters can be distributed). Vickrey (1973) studied the case that value of time is proportional to unit cost of schedule delay. Cohen (1987) considered two typical groups of commuters: low-income commuters who have low value of time but rigid work schedule and high-income commuters who value their time higher and have more flexible work schedule. Arnott et al. (1988, 1994) generalized Cohen’s analysis and also considered the other dimensions of user heterogeneity under certain assumptions. Hall (2013) allows agent preferences to be continuously distributed in three dimensions, but requires that the cost of being late be proportional to the cost of being early. If a general joint distribution of these user-specific parameters is considered, however, obtaining an equilibrium solution for bottleneck model seems analytically intractable.

Addressing the route choice of heterogeneous users in a parallel bottleneck model would significantly increase the complexity of the problem. Arnott et al. (1990) examined a system of parallel routes with homogeneous users and found that Pigouvian toll does not alter route choice. The heterogeneity setting of Cohen (1987) was adopted to study route choice with two parallel routes by Arnott et al. (1992); Liu and Nie (2011); and Huang (2000).

Vickrey’s model also inspired a large body of literature under the umbrella of Dynamic Traffic Assignment (DTA) which seeks to forecast dynamics of urban traffic in more realistic settings. Because the DTA models aim at representing realistic traffic phenomena (e.g. physical queue, traffic control measures), the commute cost is typically evaluated through simulation instead of close-form formulae. Accordingly, the equilibrium problem may only be solved approximately in most cases. The reader is referred to Peeta and Ziliaskopoulos (2001) for a comprehensive review of DTA research. Recently, Ramadurai et al. (2010) and Pang et al. (2012) tackled the Vickrey’s bottleneck problem with heterogeneous users using a DTA approach. Specifically, they proposed formulating and solving the problem as a general linear complementarity system.

The approach adopted in this paper differs from those in the classical bottleneck analysis.
and the DTA research. It makes use of the analytical solutions from the bottleneck analysis to create an equivalent assignment problem that admits closed-form commute cost functions. More specifically, the underlying assignment problem is a static and asymmetric traffic assignment problem, which may be formulated as a variational inequality problem. We call this approach "semi-analytical" because it blends analytical and numerical methods. This allows us to analyze the basic properties of the underlying problem, such as the existence and uniqueness of equilibrium, since we have a closed-form commute cost function (cf. Newell, 1987; Daganzo, 1985; Lindsey, 2004). We prove the existence and the uniqueness of DUE of our model by examining the P-property of VIP’s Jacobian matrix. Perhaps more important for practical purposes, this problem can be solved to high precision with simple assignment algorithms, which makes it a useful instrument to perform numerical analysis for congestion management policies using bottleneck model with user heterogeneity.

For the reminder, Section 2 introduces a single bottleneck model with a fixed number of heterogeneous commuters. In Section 3, a semi-analytical approach is developed to transforms the DUE problem of the bottleneck model into static traffic assignment problem, which is then formulated as a variational inequality problem. Section 4 proves the existence and the uniqueness of DUE solution. In Section 5, we extend the variational inequality formulation to solve DUE in a single origin-destination corridor network with multiple parallel routes. Section 6 reports numerical experiments using empirical results from California State Route 91. Section 7 concludes our findings.

## 2 Model Setting

Consider a fixed number of travelers who commute from home to work through a corridor during morning peak-hour. Without loss of generality, we assume that a bottleneck is located at the exit of the corridor, such as an off-ramp leading to downtown. When the demand (the departure rate) exceeds the capacity of the bottleneck, denoted as $s$, a queue forms and consequently commuters experience queuing delays. Therefore, the travel time along the corridor consists of two parts: 1) the fixed free-flow travel time $T$, i.e., the time needed to traverse the corridor when there is no congestion, and 2) queuing delay. When commuters arrive at their work place, their schedule delay will be the difference between the actual and desired arrival times. Each commuter chooses a departure time so as to minimize his/her commute cost, which consists of the costs of travel delay and schedule delay as in the classic bottleneck analysis Vickrey (1969a). To consider user heterogeneity, the population is assumed to consist of $n$ groups of commuters. All the groups are assumed to have an identical desired arrival time $t^*$.

To formulate the equilibrium problem, four one-to-one mappings are created between the set of group IDs $i$ and the set of departure orders $k$. Arnott et al. (1994) has shown that the departure order in the dynamic user equilibrium of a bottleneck is predetermined by the relative schedule delay and travel time costs, i.e., $\beta_i/a_i$ for early arrival and $\gamma_i/a_i$ for late arrival. Commuters
choose their departure time to make the trade-off between the travel delay and the schedule delay. The farther from the center of rush hour a group travels, the lower travel time and the higher schedule delay is. Therefore, these with relatively higher cost of schedule delay to travel time (\(\beta_i/\alpha_i\) and \(\gamma_i/\alpha_i\)) will prefer to travel closer to the peak. Arnott et al. (1994) also proved that each group’s early/late departure interval is a connected set, and there is no mixing of groups as long as the ratios \(\beta_i/\alpha_i\) and \(\gamma_i/\alpha_i\) are different across groups. Therefore, it is also assumed here that \(\beta_i/\alpha_i \neq \beta_j/\alpha_j, \gamma_i/\alpha_i \neq \gamma_j/\alpha_j, \forall i \neq j\). As shown in Table 1, group \(i\) is the \(F(i)\)-th departing from the beginning of congestion who arrives at the work place earlier than the desired arrival time \(t^*\); group \(E(k)\) is the \(k\)-th departing from the end of congestion who arrives at the work place earlier than the desired arrival time \(t^*\). Because it is assumed that the queue at bottleneck follows the First-In-First-Out (FIFO) principle, the groups arrive at work in the order they depart from home. Similarly, we have two mappings, \(G(i)\) and \(A(k)\), for the case of late arrival. The mapping \(F(i)\) and \(G(i)\) can be obtained by ranking the ratios of \(\beta_i/\alpha_i\) and \(\gamma_i/\alpha_i\), and mapping \(E(k)\) and \(A(k)\) are the inverses of mapping \(F(i)\) and \(G(i)\) respectively. The parameter assumptions are summarized here:

Assumption 1: \(\beta_i < \alpha_i, \forall i\)  
Assumption 2: all the groups are assumed to have an identical desired arrival time \(t^*_i = t^*, \forall i\)  
Assumption 3: \(\beta_i/\alpha_i \neq \beta_j/\alpha_j, \gamma_i/\alpha_i \neq \gamma_j/\alpha_j, \forall i \neq j\)

Let \(N_{ia}\) denote the number of commuters in group \(i\) who arrive later than their desired arrival time \(t^*\) and \(N_{ie}\) denote the number of commuters in group \(i\) who arrive earlier than their desired arrival time \(t^*\). The following demand constraints must be satisfied:

\[
N_{ie} + N_{ia} = N_i, \forall i  \tag{1}
\]

\[
N_{ie}, N_{ia} \geq 0  \tag{2}
\]

Without loss of generality, let us assume that each group has both arrival early and arrival late intervals (two intervals). If all the commuters in group \(i\) choose to arrive early (or late), then \(N_{ia} = 0\) (or \(N_{ie} = 0\)). Therefore, there is a total of \(2N\) departure intervals. We also select a commuter from each departure interval to represent the commute cost of the departure interval: the last commuter in each early-arrival interval and the first commuter in each late-arrival interval. The reason we select in this way is that the commute cost of the first and the last commuter of each departure interval can be formulated as an affine map of equilibrium flows (\(N_{ia}\) and \(N_{ie}\) in Table 2), which will be demonstrated in Section 3. Let \(x_k\) and \(y_k\) denote the travel time and the schedule delay of the last user in the \(k\)-th early-arrival group from the beginning of congestion. Similarly, let \(u_k\) and \(v_k\) denote the travel time and the schedule delay of the first user in the \(k\)-th late-arrival group from the end of congestion. The commute cost of the representative commuter in each departure interval, i.e., the sum of the costs of travel delay and schedule delay, can be formulated
to commute cost

commute cost for both the early-arrival and late-arrival intervals of group \(i\)

increase at a rate \(b\)

intuitively, the queuing time (the additional travel time caused by congestion) for group \(i\)

Because commuters in group \(i\)

the commute cost across commuters within the same departure interval will remain the same.

relationship which imposes the user equilibrium condition within each departure interval, i.e.

\(N\)

will build an affine map from equilibrium flow \(N\) to commute cost across commuters within the same group,

as:

\[
c_{ie}(N) = \alpha_i x_f(i) + \beta_i y_f(i)
\]

\[
c_{ia}(N) = \alpha_i u_g(i) + \gamma_i v_g(i)
\]

where \(c_{ie}\) represents the commute cost of the \(k\)-th early-arrival group from the beginning and \(c_{ia}\) represents the commute cost of the \(k\)-th late-arrival group from the end. The rest of this section will build an affine map from equilibrium flow \(N = (N_{E(1)e}, N_{E(2)e}, \ldots, N_{E(n)e}, N_{A(1)a}, N_{A(2)a}, \ldots, N_{A(n)a})\) to commute cost \(C = (c_{E(1)e}, c_{E(2)e}, \ldots, c_{E(n)e}, c_{A(1)a}, c_{A(2)a}, \ldots, c_{A(n)a})\), i.e., \(C(N) = J^T N, J \in \mathbb{R}^{2n \times 2n}\).

With the set of unknown variables defined in Table 2, we can write the queuing delay-flow relationship which imposes the user equilibrium condition within each departure interval, i.e. the commute cost across commuters within the same departure interval will remain the same. Because commuters in group \(i\) value their travel time by \(\alpha_i\) and value their unit of schedule delay by \(\beta_i\) and \(\gamma_i\) to keep the commute cost identical across the commuters within the same group, intuitively, the queuing time (the additional travel time caused by congestion) for group \(i\) will increase at a rate \(\beta_i/\alpha_i\) with respect to the departure rate within the early-arrival interval and decrease at a rate \(\gamma_i/\alpha_i\) with respect to the departure rate within late-arrival interval Vickrey (1969b). Therefore, we have the following relationship between queuing delay and equilibrium flows (\(N_{ie}\) and \(N_{ia}\)):

\[
x_1 = \frac{N_{E(1)e}}{s} \beta_E(1) \frac{1}{\alpha_E(1)}
\]

\[
x_k = x_{k-1} + \frac{N_{E(k)e}}{s} \beta_E(k) \frac{1}{\alpha_E(k)}, k = 2, \ldots, n
\]

\[
u_1 = \frac{N_{A(1)a}}{s} \gamma_1 \frac{1}{\alpha_1}
\]

\[
u_k = u_{k-1} + \frac{N_{A(k)a}}{s} \gamma_A(k) \frac{1}{\alpha_A(k)}, k = 2, \ldots, n
\]

Note that Equations (5)-(8) just partially impose the equilibrium condition. That is, the same commute cost for both the early-arrival and late-arrival intervals of group \(i\) can not be guaranteed at this point. Additional constraints will be added in the next section.

Moreover, the following schedule delay-flow relationship:

\[
y_k = \frac{N_{E(n)e} + N_{E(n-1)e} + \ldots + N_{E(k+1)e}}{s}, k = 1, \ldots, n - 1
\]

\[
y_n = 0
\]

\[
v_k = \frac{N_{A(n)a} + N_{A(n-1)a} + \ldots + N_{A(k+1)a}}{s}, k = 1, \ldots, n - 1
\]

\[
v_n = 0
\]

Table 2: Unknown variables in the equilibrium problem

<table>
<thead>
<tr>
<th></th>
<th>early arrival</th>
<th>late arrival</th>
</tr>
</thead>
<tbody>
<tr>
<td>equilibrium flow of group (i)</td>
<td>(N_{ie})</td>
<td>(N_{ia})</td>
</tr>
<tr>
<td>travel time of the k-th group</td>
<td>(x_k) (the last commuter)</td>
<td>(u_k) (the first commuter)</td>
</tr>
<tr>
<td>schedule delay of the k-th group</td>
<td>(y_k) (the last commuter)</td>
<td>(v_k) (the first commuter)</td>
</tr>
</tbody>
</table>
can be easily derived based on the fact that the schedule delay (i.e., the difference between the
actual arrival time and the desired arrival time) of a commuter is equivalent to the time needed
to serve the demand between this commuter and the commuter who arrives at \( t^* \). Substituting
Equations (5)-(12) into commute cost function (Equations (3)-(4)) gives an affine map from flows
to commute costs \((N \rightarrow C)\) which will be further explained in the next section.

3 Equivalent Assignment Formulation

Now, with an explicit formulation of the commute cost function \( C(N) \), we are ready to formulate
the dynamic user equilibrium problem of the bottleneck model with heterogeneous commuters.
A semi-analytical approach is proposed here that transforms the dynamic equilibrium problem
into a static, multi-class, asymmetric traffic assignment problem, which is subsequently formu-
lated and solved as a variational inequality problem (VIP).

According to the definition of the user equilibrium, the commute costs should be the same
across commuters within a group whether they arrive before or after the desired arrival time
\( t^* \). In Section 2, equilibrium condition is partially imposed by letting queuing delay increase
at at rate \( b_i / \alpha_i \) and decrease at rate \( g_i / \alpha_i \) (Equations (5)-(12)), which means that commuters
within each departure interval have the same commute cost. To determine the user equilibrium,
additional constraints are needed since each group has two departure intervals. Consequently, if
"the early-arrival" and "the late-arrival" are each treated as one of the parallel routes connecting
a single origin-destination pair, the dynamic user equilibrium problem is equivalent to assigning
demand of each group to their two "imaginary" routes. The above choice can be stated as a
mixed linear complementarity system, which corresponds to an equilibrium problem.

\[
(c_{ie}(N) - u_i)N_{ie} = 0, i = 1, ..., n \tag{13}
\]
\[
(c_{ia}(N) - u_i)N_{ia} = 0, i = 1, ..., n \tag{14}
\]
\[
N_{ie} + N_{ia} = N_i, i = 1, ..., n \tag{15}
\]
\[
N_{ie}, N_{ia} \geq 0, i = 1, ..., n \tag{16}
\]

Specifically, Figure 1 demonstrates that for each group \( i \), commuters have their own origin \( i \) and
destination \( i + 1 \) and choose between the two routes: the early arrival route and the late arrival
route. In Section 2, a closed-form commute cost function has been developed (Equations (3)-(4)).Thus, the dynamic user equilibrium can be determined by solving the following equivalent
static traffic assignment problem.

**Equivalent Static Traffic Assignment Problem:** Consider a network with a set of \( n + 1 \) nodes
and a set of \( 2n \) routes, as shown in Figure 1. There are \( n \) Origin-Destination (O-D) Pairs, and
commuters entering the network at node \( i \) always exits at node \( i + 1 \). The demand between origin
\( i \) and destination \( i + 1 \) is denoted by \( N_i \). There are two routes connecting a pair of nodes \( i \) and
\( i + 1 \). Therefore, commuters between origin \( i \) and destination \( i + 1 \) choose between two routes, so
that the flows on the two routes, denoted as \( N_{ie} \) and \( N_{ia} \), are at user equilibrium. The commute
cost functions \( (c_{ie} \text{ and } c_{ia}) \) are given in Equations (3)-(4).

The commute cost functions (Equations (3)-(4)) are not separable, i.e., the commute cost of a
route depends on not only the flow on its, but also the flows on other routes. To see this, note
that the route interactions, i.e., the marginal effect of one additional route flow on the commute cost of any other route, can be derived as:

\[
\frac{\partial C_{ie}}{\partial N_{ja}} = 0; \quad \frac{\partial C_{ia}}{\partial N_{je}} = 0
\]  \hspace{1cm} (17)

\[
\frac{\partial C_{ie}}{\partial N_{je}} = \begin{cases} \frac{a_i}{s} \frac{b_j}{x} & \text{if } F(j) < F(i) \\ \frac{b_j}{s} & \text{if } F(j) \geq F(i) \end{cases}
\]  \hspace{1cm} (18)

\[
\frac{\partial C_{ia}}{\partial N_{ja}} = \begin{cases} \frac{a_i}{s} \frac{g_j}{x} & \text{if } G(j) < G(i) \\ \frac{g_j}{s} & \text{if } G(j) \geq G(i) \end{cases}
\]  \hspace{1cm} (19)

This reveals the asymmetric link interactions between the early-arrival (or late-arrival) commute costs and early-arrival (or late-arrival) flows, i.e.:

\[
\frac{\partial C_{ie}}{\partial N_{je}} \neq \frac{\partial C_{je}}{\partial N_{ie}} \quad \text{and} \quad \frac{\partial C_{ia}}{\partial N_{ja}} \neq \frac{\partial C_{ja}}{\partial N_{ia}}, \quad \text{for } i \neq j
\]  \hspace{1cm} (20)

The Jacobian matrix is formed by arranging the derivatives of all route commute functions \( C \) in the order of departure \( C = (c_{E(1)e}, c_{E(2)e}, \ldots, c_{E(n)e}, c_{A(1)a}, c_{A(2)a}, \ldots, c_{A(n)a}) \) (from the first-departure group to the last-departure group for early arrival, then from the last-departure group to the first-departure group for late arrival), with respect to route flows \( N \).

\[
J = \nabla_{N} C
\]
With the Jacobian matrix, we can write the commute function as:

$$ C(N) = J^T N, J \in \mathbb{R}^{2n \times 2n} $$

(21)

When the Jacobian matrix of the commute cost functions is symmetric, i.e., the link interactions are symmetric, it is possible to find an equivalent convex minimization problem (Dafermos, 1971).

However, in our model the link interactions and Jacobian matrix are asymmetric, and therefore there is no known equivalent minimization program. That said, asymmetric traffic assignment problems can be formulated as a variational inequality (see Smith, 1979; Dafermos, 1980). We state the formulation as follows:

A route flow pattern $$ N^* = (N_E(1)e, N_E(2)e, \ldots, N_E(n)e, N_A(1)a, N_A(2)a, \ldots, N_A(n)a) $$ is a user equilibrium of the equivalent static traffic assignment problem if and only if it satisfies the variational inequalities problem $$ VI(K, C) $$:

$$ \langle C(N^*)^T, (N - N^*) \rangle \geq 0 $$

(22)

where $$ \langle \cdot, \cdot \rangle $$ denotes the inner product in $$ 2n $$-dimensional Euclidean space, $$ C(N) $$, denote an affine map from $$ N $$ to commute cost function $$ C = (c_{E(1)e}, c_{E(2)e}, \ldots, c_{E(n)e}, c_{A(1)a}, c_{A(2)a}, \ldots, c_{A(n)a}), N $$ denotes flow pattern that lies in a convex set $$ K $$ of vectors in $$ \mathbb{R}^{2n} $$ satisfying the following demand constraints:

$$ N_{ie} + N_{ia} = N_{i}, i = 1, 2, \ldots, n $$

$$ N_{ie}, N_{ia} \geq 0 $$

(23)

(24)

Based on the theory of variational inequality in literature (see Harker and Pang (1990) for an excellent survey of theory, algorithms and applications in VIP), we will further discuss the existence and the uniqueness of the dynamic user equilibrium in the next section.

4 Existence and Uniqueness

Lindsey (2004) has proved the existence and uniqueness of a more general model, implying that the equilibrium of our model exists uniquely. The variational inequality formulation offers a new

---

1See Beckmann et al. (1956) and Sheffi (1985) for examples of this for the static traffic assignment problem.
perspective to examine these basic properties. In particular, thanks to the transformation, we only need to examine the existence and the uniqueness of the equivalent static traffic assignment or the VIP formulated in Equations (22)-(24).

**Proposition 1** (Existence). Under Assumption 1-3, there exists a solution to the problem \( VI(K, C) \).

Proposition 1 can be proved by showing that \( K \) is a nonempty, compact and convex subset of \( \mathbb{R}^{2n} \) and \( C \) is a continuous mapping from \( \mathbb{R}^{2n} \) to itself (Eaves, 1971). \( K \) in our problem is a closed and bounded polyhedral, hence compact. Also, since \( C \) is an affine mapping, it is continuous. Therefore, we can conclude that a solution of \( VI(K, C) \) exists.

In order to prove the uniqueness of the equilibrium solution, let’s first make some necessary definitions. Fiedler (1962) defined a P-matrix:

**Definition 1.** A matrix \( A \) is a P-matrix if it fulfills one of the following conditions:

- all principal minors of \( A \) are positive;
- to every vector \( x \neq 0 \), there exists an index \( k \) such that \( x_k y_k > 0 \) where \( y = Ax \).

**Lemma 1.** The following two properties of a matrix \( A \) are equivalent:

- all leading principal minors are positive;
- there exists a lower triangular matrix \( L \) and a upper triangular matrix \( U \) both with positive diagonal elements such that \( A = LU \).

Proof: see Theorem 3.1 Fiedler (1962).

We consider a class of real square matrices \( A \) in the following form:

\[
A = \begin{bmatrix}
\alpha_1 \theta_1 & \alpha_2 \theta_1 & \cdots & \alpha_i \theta_1 & \cdots & \alpha_n \theta_1 \\
\alpha_1 \theta_1 & \alpha_2 \theta_2 & \cdots & \alpha_i \theta_2 & \cdots & \alpha_n \theta_2 \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\alpha_1 \theta_1 & \alpha_2 \theta_i & \cdots & \alpha_i \theta_i & \cdots & \alpha_n \theta_i \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\alpha_1 \theta_1 & \alpha_2 \theta_n & \cdots & \alpha_i \theta_n & \cdots & \alpha_n \theta_n
\end{bmatrix}
\]

(25)

where \( \alpha_i > 0, \theta_i > 0, \forall i \), and \( \theta_1 < \theta_2 < \cdots < \theta_i < \cdots < \theta_n \). We want first prove that any matrix in the form of \( A \) is a P-matrix, which will subsequently lead to that Jacobian matrix \( J = C' \) of the problem \( VI(K, C) \) is a P-matrix. The P property is essential for the proof of the uniqueness of equilibrium in the later part.

**Lemma 2.** A matrix \( A \) of form (25) is a P-matrix if \( \alpha_i > 0, \theta_i > 0, \forall i \), and \( \theta_1 < \theta_2 < \cdots < \theta_i < \cdots < \theta_n \).

Proof: Because

\[
A = LLI = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
1 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
1 & 1 & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
\alpha_1 \theta_1 & \alpha_2 \theta_1 & \cdots & \alpha_n \theta_1 \\
0 & \alpha_2 \theta_1 & \cdots & \alpha_n \theta_1 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \alpha_n \theta_1
\end{bmatrix}
\begin{bmatrix}
\alpha_1 \theta_1 & \alpha_2 \theta_1 & \cdots & \alpha_n \theta_1 \\
0 & \alpha_2 \theta_1 & \cdots & \alpha_n \theta_1 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \alpha_n \theta_1
\end{bmatrix}
\]
that is, there exists a lower triangular matrix $L$ and an upper triangular matrix $U$ both with positive diagonal elements such that $A = LU$. Per Lemma 1, this is equivalent to all leading principal minors of $A$ is are positive. It can be easily verified that any principal minor of $A$, i.e., $A(M)$ where $M$ is a subset of indices $\{1, 2, \ldots, n\}$, also belong to class $A$. As proved above, all leading principal minors of $A(M)$ are positive, including $A(M)$ itself. Thus, we conclude all the principal minors of $A$ are positive. According to Definition 1, $A$ is a P-matrix.

Lemma 3. Matrix $J = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$ is a P-matrix if both $A_1$ and $A_2$ are P-matrices.

Proof: According to definition 1, to every vector $x \neq 0$, there exists an index $k$ such that $x_k(A_1x)_k > 0$. Therefore, to every vector $x = \begin{pmatrix} x \\ w \end{pmatrix} \neq 0$, there exists an index $k$ such that:

$$x_k[\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} \begin{pmatrix} x \\ w \end{pmatrix}]_k = x_k[\begin{pmatrix} A_1x \\ A_2w \end{pmatrix}]_k = x_k(A_1x)_k > 0$$

We invoke Definition 1 again, and conclude that $J$ is a P-matrix. Proof is completed.

Now, let’s define a class of functions, P-function (Definition 2.5 in Moré and Rheinboldt (1973)), and we will prove that $C(N)$ (Equation (21)) is a P-function.

Definition 2. A mapping $F: Q \subset R^n \rightarrow R^n$ is a P-function if for any $x, y \in Q, x \neq y$, there is an index $k = k(x, y) \in N$ such that $(x_k - y_k)[f_k(x) - f_k(y)] \geq 0$ and $x_k \neq y_k$.

Proposition 2. The affine mapping $C(N) = f'N$ in the problem $VI(K, C)$ is a P-function when $C'$ is a P-matrix.

Proof: Moré and Rheinboldt (1973) (Theorem 5.2) proved that if $C$ is differentiable on a rectangle $Q$ and its Jacobian matrix $C'$ is a P-matrix for each $x$ in $Q$, then $C$ is a P-function on $Q$. Lemma 3 proved that the Jacobian matrix $J = C'$ of the problem $VI(K, C)$ is a P-matrix. According to the definition of P-function, that is, for any $x, y \in Q, x \neq y$, there is an index $k = k(x, y) \in N$ such that $(x_k - y_k)[f_k(x) - f_k(y)] \geq 0$ and $x_k \neq y_k$.

Let $Q$ be a large rectangle $Q = \prod_{i=1}^{2n} [0, \sigma], \sigma \gg N_i, i = 1, \ldots, n$, so that $K \subset Q$. It can be easily verified that $C$ is still a P-function on $K \subset Q$, because for any $x, y \in K \subset Q, x \neq y$, there is an index $k = k(x, y) \in N$ such that $(x_k - y_k)[c_k(x) - c_k(y)] \geq 0$ and $x_k \neq y_k$. Proof is completed.

Proposition 3 (Uniqueness). Under Assumption 1-3, the problem $VI(K, C)$ has a unique solution.

Proof: see Theorem 2.3 in (Moré, 1974) that there are at most one solution to $VI(R^n_+, C)$ if $C$ is P-function. Because $K \subset R^n_+$, there is at most one solution to $(K, C)$. It is proved in Proposition 1 that there exists a solution to the problem $VI(K, C)$. We can conclude that $VI(K, C)$ has a unique solution. Proof is completed.

Remark: It is worthwhile noting that the uniqueness of the VIP solution can be also guaranteed if:
(1) the Jacobian matrix of commute cost functions is positive definite; or

(b) \( C(N^+) \) is strictly monotone, i.e.,

\[
(C(N_1) - C(N_2))^T(N_1 - N_2) > 0, \forall N_1, N_2 \in \mathcal{K}, N_1 \neq N_2
\]

where the class of positive definite matrices also belong to P-matrices. The asymmetric Jacobian matrix \( J \) is positive definite if and only if matrix \( H = (J^T + J)/2 \) is positive definite because \( <x, Hx> = <x, Jx> \) always holds. However, we found the Jacobian matrix \( J \) may not be positive definite and the commute cost function \( C(N^+) \) may not be strictly monotone. Conditions (a) and (b) are both sufficient conditions for the uniqueness of solution. In other words, violating these conditions does not necessarily mean non-uniqueness.

5 Model with Route Choices

It is straightforward to extend the above semi-analytical approach to address the route choice in a single origin-destination network with a number of \( q \) parallel routes. For each group, each route will be treated as two “imaginary” routes corresponding to early and late arrivals, thus there are a total of \( 2nq \) routes for each group in the equivalent static traffic assignment problem. Let \( S = (s_1, s_2, \ldots, s_q) \) be the capacity; \( T = (T_1, T_2, \ldots, T_q) \) be the free flow travel time of a set of routes; \( N_{ie} \) (or \( N_{ia} \)) be the flow of an early-arrival (or late-arrival) group \( i \) on route \( j \); \( c_{ie} \) (or \( c_{ia} \)) be the commute cost of an early-arrival (or late-arrival) group \( i \) on route \( j \); \( C = (C^1, C^2, \ldots, C^q) \) be the commute cost function, where \( C^i = (c_{E(1)e}^i, c_{E(2)e}^i, \ldots, c_{E(n)e}^i, c_{A(1)a}^i, c_{A(2)a}^i, \ldots, c_{A(n)a}^i) \); \( N = (N^1, N^2, \ldots, N^q) \) denote the flow, where \( N^i = (N_{ie}^1, N_{ie}^2, \ldots, N_{ie}^n, N_{ia}^1, N_{ia}^2, \ldots, N_{ia}^n) \).

\( C \) can be shown as an affine mapping of \( N \) with a constant vector \( \omega \):

\[
C(N) = J'N + \omega, J \in \mathbb{R}^{2nq \times 2nq}, \omega \in \mathbb{R}^{2nq \times 2nq}
\]

where

\[
J = \begin{pmatrix}
J_1 & 0 & \ldots & 0 \\
0 & J_2 & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \ldots & 0 & J_q
\end{pmatrix}
\]

(26)

and \( J_j \) can be obtained by replacing \( s \) of the Jacobian matrix in Section 3 by \( s_j \), and the constant vector depends on the value of time vector \( a = (a_1, a_2, \ldots, a_n) \):

\[
\omega = (T_1 a, T_1 a, T_2 a, T_2 a, \ldots, T_q a, T_q a)
\]

Similarly, the variational inequality problem with route choice will be:

\[
C(N^+)^T(N - N^+) \geq 0
\]

(27)

where \( N \) denotes flow pattern that lies in a convex set \( \mathcal{K} \) of vectors in \( \mathbb{R}^{2nq} \) satisfying the following demand constraints:

\[
\sum_{j=1}^{q} N_{ie}^j + N_{ia}^j = N_i, i = 1, 2, \ldots, n
\]

(28)

\[
N_{ie}^j, N_{ia}^j \geq 0
\]

(29)
Existence of the user equilibrium with route choices can also be verified by showing that \( \mathcal{K} \) is a nonempty, compact and convex subset of \( \mathcal{R}^{2nq} \) and \( \mathbf{C} \) is a continuous mapping from \( \mathcal{R}^{2nq} \) to itself. The proof is omitted here. This goes beyond Lindsey (2004) which only proves existence and uniqueness for a single route.

**Proposition 4. Uniqueness** The problem VI(\( \mathcal{K}, \mathbf{C} \)) in Equations (27)-(29) has a unique solution.

Proof: We can invoke Lemma 3 and prove that Jacobian matrix in Equation (26) is a P-matrix. Similarly, as Proposition 3 proved, we can show \( \mathbf{C} \) is P-function and there is at most one solution to (\( \mathcal{K}, \mathbf{C} \)). The detailed proof is omitted here. The existence of solution is verified above. We can conclude that VI(\( \mathcal{K}, \mathbf{C} \)) has a unique solution.

\[
\mathbf{6 \ Numerical Experiments}
\]

We use a decomposition algorithm and follow the Gauss-Seidel principle (Pang, 1985) to quickly solve our problem with high precision. The intuitive idea underlying this method is to solve a sequential of \( n \) subproblems VI(\( \mathcal{K}, \bar{c}_i^{k+1} \)) where \( \bar{c}_i^{k+1} = (\bar{c}_{ie}^{k+1}, \bar{c}_{ia}^{k+1}) \), \( i = 1, \ldots, n \), and

\[
\bar{c}_{ie}^{k+1}(N_{ie}, N_{ia}) = c_{ie}(N_{1e}^{k+1}, \ldots, N_{(i-1)e}^{k+1}, \ldots, N_{ne}^{k+1}, N_{1a}^{k+1}, \ldots, N_{(i-1)a}^{k+1}, N_{ia}, N_{(i+1)a}^{k+1}, \ldots, N_{na}^{k})
\]

\[
\bar{c}_{ia}^{k+1}(N_{ie}, N_{ia}) = c_{ia}(N_{1e}^{k+1}, \ldots, N_{(i-1)e}^{k+1}, \ldots, N_{ne}^{k+1}, N_{1a}^{k+1}, \ldots, N_{(i-1)a}^{k+1}, N_{ia}, N_{(i+1)a}^{k+1}, \ldots, N_{na}^{k})
\]

\( c_{ie} \) and \( c_{ia} \) are formulated by Equations (3)-(4). Fixing the off-diagonal elements of commute cost function in each subproblem gives an equivalent minimization problem. The solution of each sequential subproblems is \( N^{k+1} = (N_{1e}^{k+1}, \ldots, N_{ne}^{k+1}, N_{1a}^{k+1}, \ldots, N_{na}^{k+1}) \). The procedure is repeated until a satisfactory solution is obtained where \( \| N^{k+1} - N^k \| < \sigma \).

\[
\mathbf{6.1 \ Joint Distribution from State Route 91}
\]

Following the same methodology as in Hall (2013) we can estimate the distribution of \( \beta/\alpha \). We will use the generalized method of moments (Hansen, 1982) to find the parameters of the distribution of \( \beta/\alpha \) that best fit the empirical travel time profile. Specifically, our moment conditions are that the model predicted travel times match the same average travel times.

To do so we make several assumptions. First, we assume that \( \gamma/\beta \) is constant for all agents. Second, we assume that the distribution of \( \beta/\alpha \) is uniform over \( [0, \delta] \) with a point mass at \( \delta \), where \( \delta \) is unknown and must be estimated along with the weight of the point mass.

In order to map the model to the data we also estimate the length of rush hour, free flow travel times, \( T \), and agents’ desired arrival time \( t^* \).

To do so we use data from the California Department of Transportation’s Performance Measurement System (PeMS) (California Department of Transportation, 1999), which contains road detector data from almost all of the highways in California. From this database we calculate
travel times for each arrival time for every non-holiday weekday in 2004 for trips from on California State Route 91 westbound from the middle of Corona to I-605. We do this for the start of every five minute interval between 4:00 a.m. and 10:00 a.m.

We find that $\beta/\alpha$ follows a uniform distribution between $[0,0.23]$ with a mass point at $\gamma/\alpha = 0.23$ with density 0.59 and the constant $\gamma/\beta = 0.40$. We also find that rush hour is 7.5 hours long, meaning total demand is 7.5 times capacity in the morning. We estimate that agents’ desired arrival time $t^*$ is 6:50 a.m. Based on Lam and Small (2001) study of SR-91 commuters we assume that agents’ value of time $\alpha$ is log-normally distributed with parameters ($\$21, \$110$).

We find that $\gamma/\beta < 1$ contrasts with our intuition that being late is more costly than being early as well as with previous research, however, it is largely a result of how I estimate the ratio of the cost of being late to early. I only observe $\beta/\alpha$ for drivers arriving before the peak of rush hour and $\gamma/\alpha$ for drivers arriving after the peak. The finding that the cost of being late is less than the cost of being early is best interpreted as being that those drivers who choose to arrive late pay less of a cost then those drivers who arrive early; there is nothing unreasonable about this. Furthermore, being late doesn’t necessarily mean literally arriving late to an appointment, but can be that you would prefer to go to the doctor at 9 a.m. but instead schedule the appointment for 11 a.m. to avoid traffic. You arrive exactly on-time to your 11 a.m. appointment, but still have schedule delay costs.

We assume that there is no correlation between $\beta/\alpha$ and $\alpha$. The cumulative density function is shown in Figure 2. Each dimension of distribution, $\alpha$ and $\beta/\alpha$, is discretized into 10 ranges and the weighted $\alpha$ and $\beta/\alpha$ of each range represent the value of time and relative cost of schedule delay to value of time of the commuters within each range. Therefore, the demand is discretized into 10 by 10, i.e., 100, groups. We also disicrete $\alpha$ dimension of the mass point at $\gamma/\alpha = 0.23$ into 10 ranges. Therefore, there are 110 classes. The demand $d$ of a class within each joint interval $\alpha \in [a_1,a_2]$ and $\beta/\alpha \in [b_1,b_2]$ can be calculated by:

$$d = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f(\alpha, \beta) d\alpha d\beta$$

where $f(\alpha, \beta)$ denotes the density function.

We solved the DUE of the 110 classes described above using the decomposition algorithm. Figure 3 shows how the predicted travel times solved by our approach fit the actual travel time from PeMS data. It provides the curve which demonstrates the trade-off between schedule delay and travel time at user equilibrium solution. The closer to the center of rush hour a commuter is, the higher the queuing delay she/he will suffer. Because of the constant ratio $\beta_i/\gamma_i \equiv 0.40$ here, our results are verified by comparing with the analytical solution in Arnott et al. (1994) and we found the decomposition algorithm can converge to the analytical solution with high precision.

### 6.2 Four-Class Example

A four-class case is considered here to compare the no-toll equilibrium and system optimum (system cost is minimized) under general user heterogeneity. The input parameters for the

---

2We define holidays as the ten United States Federal Holidays.

3This roughly represents the typical commute for those living in Corona and using State Route 91; which we calculated using data from Sullivan (1999).

4For examples see Small (1982); Hendrickson and Plank (1984); Parthasarathi et al. (2010).
Figure 2: Cumulative Density Function of Commuters on SR-91

Figure 3: Predicted Travel Times vs Actual Travel Times
Table 3: Four-Class Example: NTE and SO Flows

<table>
<thead>
<tr>
<th>group</th>
<th>No-toll Equilibrium</th>
<th>System Optimum</th>
<th>benefit from toll($)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>early</td>
<td>late</td>
<td>early</td>
</tr>
<tr>
<td>1</td>
<td>1500</td>
<td>0</td>
<td>1269</td>
</tr>
<tr>
<td>2</td>
<td>873</td>
<td>267</td>
<td>891</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>800</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>459</td>
<td>41</td>
<td>500</td>
</tr>
</tbody>
</table>

four classes are as follows: value of time $\alpha = [6.4\ 2.5\ 2.0\ 1.7]$, unit cost of schedule delay $\beta = [3.9\ 1.95\ 1.8\ 1.5]$, $\gamma = [15.21\ 4.5\ 3.5\ 5]$, and demand is $N = [1500\ 1140\ 800\ 500]$. The capacity of the bottleneck is 2400/hr.

The no-toll equilibrium (NTE) and system optimum (SO) flows are given in Table 3. At no-toll equilibrium, the departure order is determined by the relative ratio $\beta_i/\alpha_i$ for early arrival and $\gamma_i/\alpha_i$ for arrival late. As visualized in Figure 4 a), all the group 3 commuters arrive later than their desired arrival time $t^*$ and all the group 1 commuters arrive earlier than their desired arrival time $t^*$. This is because group 3 has a higher relative cost of early arrival to travel time (i.e., $\beta/\alpha$) than the other groups and group 1 has a higher relative cost of late arrival to travel time (i.e., $\gamma/\alpha$) than the other groups.

At system optimum, the commuters are forced by toll to depart in the order determined by the absolute value of schedule, i.e., unit cost of schedule delay $\beta_i$ for early arrival and $\gamma_i$ for late arrival. The congestion is eliminated by the time-depending toll, as shown in Figure 4 b). Group 3 is absent from early-arrival route and group 4 is absent from late-arrival route. The departure order in Figure 4 b) is very different from the departure order in Figure 4. By comparing the commute cost of each group before and after pricing, the benefit from toll is provided in the last column in Table 3. Only group 1, who have value their schedule delay highest among the four groups, benefit from SO toll, and the other three classes are worse off. The group 3, who can be considered as an intermediate class, have the highest increase in travel cost, and suffer the greatest loss in this example. While not reported due to space limit, we have also found examples that the class with either the highest or the lowest value of time becomes the critical class. This is consistent with the findings based on the static two-mode network model Liu and Nie (2012).

6.3 Multi-Route Example

We proceed to apply our approach on a single origin-destination network with multiple routes. The same two-route example as in Liu and Nie (2011) is solved here by the gradient projection algorithm. The convergence performance of the algorithm is reported in Figure 5, in which $X$ represents the CPU time in second and $Y$ represents the gap, i.e., the Euclidean distance between two consecutive solutions $d^2(N_k, N_{k+1})$. The convergence curve shows that the algorithm converged quickly with sharp drop in the first few seconds. And it converged to the analytical solution provided in Liu and Nie (2011), i.e., $N = [840, 907, 216, 233, 35, 605, 9, 155]$. 
Table 4: Parameters adopted in the multi-route example

<table>
<thead>
<tr>
<th>Category</th>
<th>Travel time</th>
<th>VOT parameters</th>
<th>Capacity</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit</td>
<td>(minutes)</td>
<td>$/hour</td>
<td>veh/hour</td>
<td>N/A</td>
</tr>
<tr>
<td>Parameter</td>
<td>$T_1$</td>
<td>$T_2$</td>
<td>$\alpha_1$</td>
<td>$\alpha_2$</td>
</tr>
<tr>
<td>Value</td>
<td>18min</td>
<td>30min</td>
<td>6.4</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Figure 4: No-toll Equilibrium of Four-Class Example

Figure 5: Convergence curves of the multi-route example
7 Conclusion

This paper proposes a new approach for solving the dynamic equilibrium problem of a bottleneck model. This semi-analytical approach transforms the dynamic equilibrium problem of a bottleneck with heterogeneous users into a static traffic assignment problem. We show that the commute cost function is not separable and the link interaction is asymmetric. Since no equivalent optimization formulation exists for the equivalent traffic assignment problem, it is formulated and solved as a variational inequality problem. The variational inequality formulation proposed here provides a new perspective to prove the existence and the uniqueness of dynamic user equilibrium of the bottleneck model with heterogenous users. Our numerical experiments suggest that:

- The decomposition algorithm can quickly solve the bottleneck model with reasonably large number of user groups. We tested the proposed approach using empirical results on California State Route 91 and found the predicted travel time can well fit the actual arrival time from PeMS data.

- The proposed semi-analytical approach can be used to perform policy analysis using the bottleneck models that are otherwise intractable. For example, the welfare analysis of the system optimum toll revealed that users with either high, low or intermediate value of time could suffer the greatest loss from toll, which is consistent with the findings based on a two-model network model (Liu and Nie, 2012).

- The proposed approach can address simultaneous departure time and route choice in a single O-D network with multiple parallel routes.

The semi-analytical approach proposed herein can be extended to design more realistic toll scheme strategies, such as step tolls (toll invariant within certain period), which is an on-going research project. Another interesting, and seemingly straightforward extension, would be to use our proposed framework to determine impact of elastic demand on welfare analysis. From the modeling point of view, it would only require adding a dummy route to accommodate the latent demand.

Acknowledgements

The work was partially supported by National Science Foundation under the award number CMMI-1256021.

References


Hall, J., 2013. Pareto improvements from lexus lanes: The case for value pricing on heavily congested highways. In: Kuhmo NECTAR Conference.


