A Bilevel Optimization Model for the Network Signal Timing Design Problem with Traveler Trip-chain Route Choice Behavior Consideration

Chung-Yung WANG [a], Shou-Ren HU [b], Chih-Peng CHU [c], Ya-Tin JHUANG [d]

[a,d]Department of Logistics Management, National Defense University, Taipei City,
  112, Taiwan;
[a] E-mail: stellar@ms35.url.com.tw
[b] Department of Transportation and Communication Management Science, National Cheng
  Kung University, Tainan City, 701, Taiwan; E-mail: shouren@mail.ncku.edu.tw
[c] Department of Business Administration, National Dong Hwa University, Hualien City, 974
  Taiwan; E-mail: chpchu@mail.ndhu.edu.tw

Abstract: This study proposed a bilevel optimization model for the network signal timing design problem by considering the link flows reflected by the trip-chain route choice behaviors of road users. The bilevel programming model is formulated based on the interactions between signal timing control and trip-chain behavior, and a solution algorithm by combining variational inequality sensitivity analysis, the generalized inverse matrix method, and a gradient projection approach revised for trip-chain user equilibrium is developed for the transportation design problem. The performance of the developed model framework was verified through numerical analysis under different test scenarios.

Keywords: signal timing design, bilevel programming model, trip-chain user equilibrium.

1. INTRODUCTION

In road network analysis, the design of traffic signal timing plans is highly dependent on the accurate observations of road users' route choice behaviors and/or decisions. Signal timing settings can have effects on distributed traffic volume, which influences link travel costs, in turn affecting the routes that road users choose. The aggregation of user behaviors is reflected in the traffic volume of various road segments and intersections, and the traffic volume at intersections is the foundation of signal timing design. Fisk (1984) established a leader-follower relationship between signal timing designs and the route choices behavior of users, with the former being the leader of road network systems and the latter being the follower. Followers (the road users) are influenced by signal timings in their selection of their routes. In order to develop a desirable signal timing plan that will potentially result in the optimal transportation performance in a road network, it is necessary to develop a suitable model framework for the behaviors of followers. Therefore Stackelberg equilibrium exists between signal control and user behavior. Many previous studies explored the optimization of signal timings using bilevel programming, with the upper level considering signal timing optimization and the lower level considering user equilibrium route choices behavior. However, these researches only considered the route selections between trip origins and destinations, neglecting the trip-chain behavior that can arise from incidental social or economic activities. Consequently, the bilevel programming models proposed in these studies did not concern the situation of trip-chain behavior on signal timing design. This is a significant research gap by incorporating trip-chain needs in the travel behavior of urban road users into the signal timing design problem.

In order to take the trip chain behavior of road users into account, we integrated the Variational Inequality sensitivity analysis approaches proposed by Tobin (1986) and Tobin and Friesz (1988) and the generalized inverse matrix method to derive the objective descent direction of the bilevel programming model and developed a solution algorithm. For the lower level of the model, we constructed a trip-chain user equilibrium traffic assignment model and developed a route-based algorithm to solve it.

The remainder of this paper is organized as follows. Section 2 analyzes the research problem and relevant literature, while Section 3 describes the modeling process for the bilevel signal timing optimization model that considers trip-chain route choice decisions. Section 4 explains how to combine the sensitivity analysis method and the generalized inverse matrix method to solve the bilevel programming model. A method for solving the lower-level trip-chain-based user equilibrium traffic assignment model is also presented. Section 5 presents numerical analysis on the verification of the proposed method using a test network, and Section 6 summarizes this reserch by conclusions and suggestions.

2. PROBLEM STATEMENT AND LITERATURE REVIEW

A number of researchers have considered the route selection behavior of road users when optimizing signal timing designs using bilevel programming, including Yang and Yagar (1995), Chen and Hsueh (1997), Chen and Chou (2001), Cipriani and Fusco (2004), Chiou (2008), Smith (2011), Ukkusuri *et al*. (2013), and Chiou (2014). However, their models only considered the route choices between origin-destination (O-D) pairs, and neglected trip-chain behavior. In fact, there are also some needs to complete multiple activities into a single trip, which results in trip-chain behavior. Hägerstrand (1970) proposed an activity-based model to consider trip-chain behavior. Many empirical studies, such as those conducted by Hensher and Reyes (2000), McGuckin *et al*. (2005), Morency and Valiquette (2010), Currie and Delbosc (2011), and Zhao *et al*. (2012), showed the trip-chain behavior should be considered. Therefore, we consider the trip-chain based route choices of road users to identify the actual transportation demand within a network before an appropriate traffic signing timing design scheme can be established.

With regard to the trip-chain behavior of road users, Maruyama and Harata (2005) indicated that if the individual segments of a road user's trip-chain needs are considered separately, the results would be unable to reflect the relationships among them, thereby reducing the accuracy of predictions for subsequent transportation demand. To overcome this problem, Maruyama and Harata (2006) established a trip-chain based model under the assumption of static network equilibrium to conduct cordon-based congestion pricing. Maruyama and Sumalee (2007) indicated that a direct relationship exists between congestion pricing and the trip-chain behavior of road users and therefore used the trip-chain based network equilibrium model proposed by Maruyama and Harata (2006) to compare the validity and fairness of cordon-based and area-based pricing schemes. Higuchi *et al*. (2011) developed a novel combined transport mode and trip-chain based route choice network equilibrium model with two stages based on variational inequality problems and adopted the relaxation method to solve it. However, all of the above studies fixed the order of activities, thereby eliminating order as a factor influencing route choices. Trip-chain descriptions in the models were also overly simplified, merely indicating that if route n was on the trip-chain route between O-D pair (r,s), then trip-chain variable $\eta_n^{rs}=1$. However, an urban road network is not directly consisted of routes, but rather routes comprising respective links. Expressing trip-chain behavior in this manner renders it from analyzing the relationships between

trip-chain needs and link flows in the road network. Furthermore, these studies used link-based Frank-Wolfe algorithm approaches, which exhibit poor computational efficiency. Wang and Chen (2013) developed a trip-chain user equilibrium traffic assignment model that considered the effect of the orders in which road users participated in activities on their route choice decisions and used route-based gradient projection (GP) method to develop an algorithm. The results showed that conventional traffic assignment models which did not consider trip-chain behavior were merely a special case of trip-chain-based traffic assignment models. Thus, the model framework they developed was a more generalized trip-chain-based traffic assignment model.

To solve the bilevel programming models for optimal signal timing plans, Allsop (1974) and Gartner (1974) used iterative optimization-assignment (IOA) algorithms to update signal timing designs based on fixed network flow and then solve network equilibrium flows based on fixed signal timing designs. They first fixed the decision variables in the upper-level problem before solving the lower-level problem and then fixed the decision variables in the lower-level problem to solve the upper-level problem. This process was then repeated until the upper-level problem converged. However, the IOA algorithm obtains Nash solutions rather than the true solutions for the bilevel optimization models. Marcotte (1983) expanded the optimization problems hidden within network design problems to form a bilevel programming problem, indicating that bilevel programming problems are actually a type of non-cooperative Stackelberg game (NSG). Fisk (1984) also stated that a Stackelberg equilibrium exists between the signal control and road user's route choice behavior problem.

In the Stackelberg equilibrium solutions of bilevel programming models, the decision variables of the upper-level and lower-level models have an implicit functional relationship, and for this reason, the partial derivatives of the decision variables cannot be calculated directly. To overcome this problem, Tobin (1986) proposed a variational inequalities sensitivity analysis method and verified that when a variational inequality has a unique solution and the variables and functions satisfy the condition of strict complementary slackness, parameter perturbation can be used to obtain the optimal solution in the domain of the parameter $\varepsilon = 0$. Tobin and Friesz (1988) also adopted a variational inequalities sensitivity analysis approach that used the trace perturbations in the decision variables near the equilibrium solution to effectively estimate the derivative function near the equilibrium solution. The function provides descent direction in the search for the optimal solution and then returns the Stackelberg solution, which is the optimal solution for the bilevel programming model. While investigating bilevel programming models for intersection signal designs, Wong and Yang (1997) used the sensitivity analysis approach proposed by Tobin (1986) to determine the relationships among signal timing, link flow, and user behavior, but they overlooked the fact that route solutions may not be unique and thus derived a singular solution.

To address this issue, Cho (1991) used the generalized inverse matrix method to convert routes into unique links before performing the sensitivity analysis. This approach eliminated the possibility of degenerate solutions occurring in the sensitivity analysis. Wang (1999) combined the variational inequalities sensitivity analysis approach with the generalized inverse matrix method and successfully solved a time-dependent bilevel programming model for signal timing optimization with link capacity constraints. Chen and Chou (2001) similarly used the variational inequalities sensitivity analysis method and the generalized inverse matrix method to solve a time-dependent bilevel programming model for signal timing optimization. The research problem considered by this study can be summarized as follows:

    1) Bilevel programming models can be used to optimize traffic signal timings. The upper-level models are signal timing optimization problems with minimizing total

travel costs in the system as the objective, and the travelers route choice behaviors can serve as constraints in the lower-level models.

2) In previous researches with bilevel programming models for signal timing optimization, lower-level models did not consider trip-chain behavior and thus could not comprehensively reflect the actual transportation demands in a road network.

3) The solution of a bilevel programming model for signal timing optimization is a Stackelberg equilibrium solution. The decision variables of the upper-level and lower-level models have an implicit functional relationship and do not form a closed function. For this reason, the partial derivatives of the decision variables cannot be calculated directly. Using the variational inequalities sensitivity analysis method proposed by Tobin (1986) and the generalized inverse matrix method presented by Cho (1991) can provide the derivative of the implicit function.

3. MODEL FORMULATION

3.1 Notation list

a link number

c_a travel costs of link a

c_{a_0} free travel costs of link a

c_a' travel cost derivative of link a

c_p^{rs*} travel costs of using route p between OD pair (r,s) under user equilibrium principle

$c_{\hat{p}}^{rs*}$ travel costs of using trip-chain route $\hat{p}$ between OD pair (r,s) under user equilibrium principle

$\overline{C}^m$ signal cycle time at intersection m

CAP_a^m capacity and saturation flow rate of road link a connected to intersection m

$d_{\hat{p}}^{rs}$ descent direction

$\mathbf{d}$ vector of descent direction

g_a^{Im} effective green time in the Ith phase at intersection m which connected to link a

$\mathbf{g}$ vector of effective green time

$\underline{g}_a^{Im}$ the minimum effective green time.

h_p^{rs*} traffic flow on route p between OD pair (r,s) under user equilibrium principle

$h_{\hat{p}}^{rs*}$ traffic flow on trip-chain route $\hat{p}$ between OD pair (r,s) under user equilibrium principle

l_a^{Im} loss time in the Ith phase at intersection m connected to link a

n_i activity nodes passed by the trip-chain routes between OD pair (r,s)

N^{rs} set of activity nodes passed by the trip-chain routes between OD pair (r,s)

p path variable

$\hat{p}$ route passing activity nodes on trip-chain

$\hat{p}^*$ shortest route passing activity nodes on trip-chain

$P_{\hat{p}}^{rs}$ set of trip-chain routes

$\overline{q}^{rs}$ trip demand between OD pair (r,s)

r origin variable

s destination variable

S_a^m saturation flow rate of link a connected to intersection m

x_a traffic flow on link a

$\mathbf{z}$ objective function

$\alpha_{\hat{p}}^{rs}$ step size

ε perturbation parameter

φ convergence criterion

$\overline{\delta}_{a\hat{p}}^{rs}$ The link/path indicator variable, it is a zero-one indicator variable that equals 1 when trip-chain route $\hat{p}$ passes link a and 0 when it does not.

$\overline{\gamma}_{n_i p}^{rs}$ The activity node/path indicator variable, it is a zero-one indicator variable. When route p between trip-chain OD pair (r,s) passes the nth activity node on the trip-chain between the trip-chain OD pair, then $\overline{\gamma}_{n_i p}^{rs}=1$; otherwise, $\overline{\gamma}_{n_i p}^{rs}=0$.

υ step adjustment parameter

3.2 The model

Before formulating the trip-chain based optimization network signal timing design model, we define the trip-chain and its user equilibrium principle. In this study, all the locations of the trip-chain activities between the trip ends should be passed. We assumed that the travelers have perfect information such that they could make correct decisions regarding trip-chain route choice. The trip-chain based user equilibrium principle can be defined as no traveler can improve his trip-chain route travel time by unilaterally changing trip-chain route. That is, the route choice must pass all of the particular intermediary activity locations between their origin and destination and incur the minimal travel costs.

    The objective of this model is to minimize the total travel time. We assumed that the traffic signals at all of the intersections were two-phase signals with fixed cycle times and minimum green light time. The loss of green light time in each phase and the saturation flow rates at each intersection are known. The established model is as follows:

$$\min \sum_a c_a\left(g_a^{Im}\right) x_a\left(g_a^{Im}\right) \tag{1}$$

subject to the following constraints:

    Cycle conservation constraint:

$$\sum_I \sum_a \left(g_a^{Im} + l_a^{Im}\right) = \overline{C}^m \quad \forall I,a \tag{2}$$

    Definitional constraint:

$$CAP_a^m = S_a^m \frac{g_a^{Im}}{\overline{C}^m} \quad \forall I,m,a \tag{3}$$

    Boundary constraint:

$$g_a^{Im} \geq \underline{g}_a^{Im} \quad \forall I,a,m \tag{4}$$

    The trip-chain user equilibrium constraint:

$$\sum_a c_a^*\left(x_a, g_a^{Im}\right)\left[x_a\left(g_a^{Im}\right) - x_a^*\left(g_a^{Im}\right)\right] \geq 0 \quad \forall x_a \in \mathbf{\Omega(g)} \tag{5}$$

Eq. (1) is the target function for the signal timing optimization of a single road network with green light time as the decision variable. Eq. (2) is the constraint for cycle conservation in the signal timings, where the cycle time of the signal at intersection m is $\overline{C}^m$ and equals the total sum of effective green time g_a^{Im} and loss time l_a^{Im}, the effective green time and loss time in the Ith phase at intersection m which connected to link a, respectively. Eq. (3) defines the relationship among CAP_a^m, S_a^m, g_a^{Im}, and $\overline{C}^m$, which denote the capacity and saturation flow rate of link a connected to intersection m, the effective green time at intersection m connected to link a, and signal cycle time at intersection m. Eq. (4) is a boundary constraint stipulating that the effective green light time must be greater than or equal to the minimum effective green time $\underline{g}_a^{Im}$. Eq. (5) constrains the trip-chain route selection behavior of road users; it comprises a trip-chain based user equilibrium route selection model in the form of a variational inequality and encompasses the following constraints for flow conservation, non-negative route flow, definition, and trip-chain definition:

Flow conservation constraint:
$$\sum_{\hat{p}} h_{\hat{p}}^{rs} = \overline{q}^{rs}, \forall r, s, \hat{p} \tag{6}$$

Non-negativity constraints on route flows:
$$h_{\hat{p}}^{rs} \geq 0, \forall r, s, \hat{p} \tag{7}$$

$$x_a \geq 0 \quad \forall a \tag{8}$$

Definitional constraints:
$$x_a = \sum_r \sum_s \sum_{\hat{p}} h_{\hat{p}}^{rs} \overline{\delta}_{a\hat{p}}^{rs} \geq 0, \forall a \tag{9}$$

$$\overline{\delta}_{a\hat{p}}^{rs} = \{0,1\}, \forall r, s, a, \hat{p} \tag{10}$$

Trip-chain definitional constraints:
$$h_{\hat{p}}^{rs} = h_p^{rs} \prod_{n_i \in N^{rs}} \overline{\gamma}_{n_i p}^{rs}, \forall r, s, p \in (r,s), \hat{p} \in (r,s) \tag{11}$$

$$\overline{\gamma}_{n_i p}^{rs} = \{0,1\}, \forall r, s, n_i \in N^{rs}, p \tag{12}$$

Eq. (6) is the constraint for the conservation of trip-chain flow, meaning that the total route flows of the trip-chains between any given O-D pair (r,s) must equal $\overline{q}^{rs}$, the trip-chain demand for said O-D pair. Eq. (7) is the non-negativity constraint on trip-chain route flows, and Eq. (8) is the non-negativity constraint on link flows. Eq. (9) is presents the definitional constraint, which indicates the relationship between the traffic flow on each link in the network and the flow on the routes in the trip-chain. Eq. (10) defines $\overline{\delta}_{a\hat{p}}^{rs}$ as a known indicator variable that equals 1 if $\hat{p}$, a trip-chain route that passes the intermediate activity points between O-D pair (r,s), passes through link a and 0 otherwise. Eq. (11) defines the relationship between $h_{\hat{p}}^{rs}$ and h_p^{rs}, the route flows of a trip-chain route that passes the intermediate activity points between O-D pair (r,s) and a common route; $\overline{\gamma}_{n_i p}^{rs}$ is an

activity point/route adjacency matrix that is a known $\{0,1\}$ indicator variable, and N^{rs} is a set of activity points (nodes) that the trip-chain route must pass between O-D pair (r, s) ($N^{rs} = \{n_1, n_2, ..., n_m\}$). If route p passes the ith activity point, n_i, then $\bar{\gamma}_{n_i p}^{rs} = 1$; otherwise, $\bar{\gamma}_{n_i p}^{rs} = 0$. If route p passes all of the activity points in the trip-chain between O-D pair (r, s), then the product of all the $\bar{\gamma}_{n_i p}^{rs}$ values for this route must equal 1; in other words, $\prod_{n_i \in N^{rs}} \bar{\gamma}_{n_i p}^{rs} = 1$. If route p does not pass one or more of the activity points in the trip-chain, then the product of all the $\bar{\gamma}_{n_i p}^{rs}$ values for this route must equal 0; in other words, $\prod_{n_i \in N^{rs}} \bar{\gamma}_{n_i p}^{rs} = 0$.

Thus, the constraining conditions of Eq. (11) can restrict the flows of the trip-chain routes between O-D pair (r, s) and ensure that all of the activity points in the trip-chain are passed. Furthermore, $\prod_{n_i \in N^{rs}} \bar{\gamma}_{n_i p}^{rs} = 1$ does not restrict the order in which the activity points are passed.

Finally, Eq. (12) is a constraint for the trip-chain definition, in which $\bar{\gamma}_{n_i p}^{rs}$ is an activity point/route adjacency matrix that is a known $\{0,1\}$ indicator variable.

As can be seen, the model is a bilevel programming model. The upper-level model optimizes the road network system, and the lower level model is a trip-chain based user equilibrium traffic assignment model encompassing the constraints of the upper-level model. With regard to trip-chain based user equilibrium conditions, Wang and Chen (2013) stated that when users must pass certain intermediate activity points, the route that they will select is the one that passes all of the intermediate activity points between their origin and destination with minimal travel costs. In view of this, we placed the trip-chain based user equilibrium traffic assignment model in the lower level of our bilevel model for signal timings to better reflect the relationship between signal timings design and road user behavior.

4. SOLUTION ALGORITHM

4.1 Sensitivity analysis and generalized inverse matrix method

The objective function of the bilevel programming model for signal timing optimization in this study pursues minimum network costs and is the sum of the products of link cost function **c(g)** and link flow **x**. Link costs are functions of green light time **g**, and differences in green light time are reflected on link capacity, which in turn influence travel costs and the route choice behavior of road users. Thus, link flow **x** varies with green light time **g** and an implicit functional relationship exists between the two, without a closed function to express it. Consequently, the derivative of link flow **x** with regard to green light time **g** cannot be calculated directly, which makes the signal timing optimization model of this study difficult to solve. Generally, the solution of a bilevel programming model is also referred to as a Stackelberg equilibrium solution. To address the problem in which partial derivatives cannot be obtained directly from the implicit functional relationship between the decision variables of the upper and lower levels, Tobin (1986) and Tobin and Friesz (1988) proposed sensitivity analysis methods to derive the descent direction of the objective function in the upper level and obtain the local optimum solution of the model. We also adopted this approach to solve the model in this study.

According to the conclusions made by Tobin (1986) and Tobin and Friesz (1988), an implicit functional relationship exists between link flow **x** and effective green light time **g**; an sensitivity analysis approach can be used to derive the partial derivative of the implicit function and the descent direction to search for the optimal solution of **g**, the decision variable of the upper-level model, as shown in the equation below.

$$\nabla_\varepsilon \mathbf{x}(\mathbf{0}) = \mathbf{J}_{\mathbf{x}}^{-1}(\mathbf{0})\left[-\mathbf{J}_\varepsilon(\mathbf{0})\right] \tag{13}$$

However, the sensitivity analysis method requires a route solution, which in general may not be unique for any O-D pair, and this creates some difficulty in obtaining the solution. To enhance the generality of the sensitivity analysis method in user equilibrium problems, we adopted the generalized inverse matrix approach proposed by Cho (1991) to conduct the sensitivity analysis. The generalized inverse matrix approach for sensitivity analysis in user equilibrium problems utilizes the fact that link solutions are unique. This uniqueness is used to convert the relationships among the link variables into an O-D pair/route adjacency matrix and a linear independent link/route adjacency matrix that express the route solution under the restrictions of flow conservation and link/route definitions.

We conducted a numerical simulation to verify whether the proposed approach can accurately obtain the derivative of the implicit function in trip-chain based network equilibrium problems. Figure 1 displays the test network comprising six nodes and twelve links. Nodes 1 and 2 are the origins, and Nodes 5 and 6 are the destinations. Node 3 is an intermediate activity point that must be passed between Origin 2 and Destination 5. We adopted an FHWA cost function for the link costs, shown in Eq. (14). The free-flow travel time for each link was set as 1, and Tables 1 and 2 exhibit the transportation demands for each O-D pair in the trip-chains and the data regarding the road links in the network and the signal timings of the corresponding intersections. Tables 3 and 4 present the flow data of the various routes and road links when the network reaches the conditions of trip-chain user equilibrium.
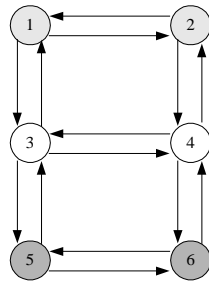

Fig. 1 Test Network 1

$$c_a(x_a) = c_{a_0}\left[1 + 0.15\left(\frac{x_a}{Cap_a}\right)^4\right], \forall a \tag{14}$$

Table 1. Transportation demand between trip-chain O-D pairs in Test Network 1

No.	O-D pair	Activity node	Travel demend (pcu/hr)
1	1-6	--	30
2	2-5	3	50

Table 2. Link and intersection signal timing data

Cycle time	Initial green time	Minimum green time	Saturation flow rate	Lost time
60	27	7	50	3

Table 3. Trip-chain route flows when Test Network 1 reaches trip-chain user equilibrium

O-D pair	Activity node	No. of path	Trip-chain path	flow	Travel time
1-6		1	1→3→4→6	14.55	5.27
		2	1→2→4→6	15.45	
2-5	3	3	2→4→3→5	24.64	7.51
		4	2→1→3→4→6→5	3.31	
		5	2→1→3→5	22.04	

Table 4 Link flows when Test Network 1 reaches trip-chain user equilibrium

No.	Link	Flow	No.	Link	Flow
1	1→2	15.45	8	4→2	0.00
2	1→3	39.91	9	4→3	24.64
3	2→1	25.36	10	4→6	33.31
4	2→4	40.09	11	5→3	0.00
5	3→1	0.00	12	5→6	0.00
6	3→4	17.86	13	6→4	0.00
7	3→5	46.69	14	6→5	3.31

Next, based on the solution of trip-chain user equilibrium, we put a small perturbation on the green time. And compare the actual solution with the estimated solution that is calculated by variational inequality sensitivity analysis and generalized inverse approach. The results are summarized in Table 5. When we put a small perturbation to green time, the estimated solution is almost same as the actual solution. If the perturbation was too big, the difference between estimated solution and actual solution would be increased. The numerical test demonstrates that the implicit difference could be calculated by variational inequality sensitivity analysis and generalized inverse approach. Then the exactly descent search direction could be found. Therefore, with the optimal step sizes, the Stackelberg solution of bilevel programming model would be easily solved.

Table 5 Comparison of actual equilibrium solutions and first-order approximate solutions

No.	Link	$\varepsilon = 0$	$\varepsilon = 0.1$		$\varepsilon = 0.5$	
			Actual sol.	Estimated sol.	Actual sol.	Estimated sol.
1	1→2	15.45	15.43	15.43	15.38	15.38
2	1→3	39.91	39.91	39.91	39.92	39.92
3	2→1	25.36	25.34	25.34	25.29	25.29
4	2→4	40.09	40.09	40.09	40.08	40.08
6	3→4	17.87	17.79	17.79	17.46	17.47
7	3→5	46.69	46.78	46.78	47.16	47.15
9	4→3	24.64	24.66	24.66	24.71	24.71
10	4→6	33.31	33.22	33.22	32.84	32.85
14	6→5	3.31	3.22	3.22	2.84	2.85

* ε is a small perturbation parameter

4.2 Solution algorithm

Numerical simulations demonstrate that the sensitivity analysis approach and generalized inverse matrix methods can produce the partial derivative of the implicit function and can be applied to the bilevel programming model of this study to provide direction in the search for the objective function of the upper-level model. Below, we describe the steps of the solution algorithm in this study.

Step 1: Set the initial green light times $\mathbf{g}^n$ for each intersection and let $n = 1$.

Step 2: Calculate the capacities of each link using the formula below.

$$CAP_a^{Im^n} = S_a^{Im} \frac{g_a^{Im^n}}{\overline{C}^m} \quad \forall I, m, a \tag{15}$$

Step 3: Employ the solution algorithm developed by Wang and Chen (2013) based on gradient projection to solve the trip-chain based user equilibrium traffic assignment model below:

$$\min \mathbf{z} = \sum_a \int_0^{x_a} c_a(\omega, \mathbf{g}^n) d\omega \tag{16}$$

S.t.

Flow conservation constraint

$$\sum_{\hat{p}} h_{\hat{p}}^{rs} = \overline{q}^{rs} \quad \forall r, s \tag{17}$$

Non-negativity constraint on route flows

$$h_{\hat{p}}^{rs} \geq 0 \quad \forall r, s \tag{18}$$

Definitional constraint

$$x_a = \sum_{rs} \sum_{\hat{p}} h_{\hat{p}}^{rs} \overline{\delta}_{a\hat{p}}^{rs} \quad \forall a \tag{19}$$

$$c_{\hat{p}}^{rs} = \sum_a c_a \overline{\delta}_{a\hat{p}}^{rs} \quad \forall r, s, \hat{p} \tag{20}$$

$$\overline{\delta}_{a\hat{p}}^{rs} = \{0,1\} \quad \forall r, s, a, \hat{p} \tag{21}$$

The trip-chain definitional constraint

$$h_{\hat{p}}^{rs} = h_p^{rs} \prod_{n_i \in N^{rs}} \overline{\gamma}_{n_i p}^{rs}, \forall r, s, p \in (r,s), \hat{p} \in (r,s) \tag{22}$$

$$\overline{\gamma}_{n_i p}^{rs} = \{0,1\}, \forall r, s, n_i \in N^{rs}, p \tag{23}$$

Step 4: Use the aforementioned sensitivity analysis and generalized inverse matrix methods to derive $\nabla_\varepsilon \mathbf{x}(\mathbf{0})$.

Step 5: Calculate the descent direction for green light time:

$$\mathbf{d}^n = -\nabla_\varepsilon(\mathbf{xc}(\mathbf{x}, \mathbf{0})) \tag{24}$$

Using the FHWA cost function, for example, the descent direction is

$$\mathbf{d}^n = -\nabla_\varepsilon \left(\mathbf{xc}(\mathbf{x},\mathbf{0}) \right) = -\left(\begin{array}{c} \mathbf{c}_0 \nabla_\varepsilon \mathbf{x}(\mathbf{0}) + 0.75 \mathbf{c}_0 \mathbf{x}^4 \nabla_\varepsilon x(\mathbf{0}) \left(\dfrac{\mathbf{Sg}^n}{\overline{\mathbf{C}}} \right)^{-4} \\[4mm] -0.6\mathbf{c}_0 \mathbf{x}^5 \left(\mathbf{g}^n \right)^{-5} \left(\dfrac{\mathbf{S}}{\overline{\mathbf{C}}} \right)^{-4} \end{array} \right)$$ (25)

Step 6: Update the green time to $\mathbf{g}^{n+1} = \mathbf{g}^n + \dfrac{1}{n+1} \mathbf{d}^n$.

Step 7: Adjust the green times at each intersection until the constraints below are satisfied.

$$\underline{g}_a^{Im} \le g_a^{Im^{n+1}} \quad \forall I, m$$ (26)

$$\sum_m \left(g_a^{Im^{n+1}} + l^{Im} \right) = \overline{C}^I \quad \forall I$$ (27)

Step 8: Let $n = n+1$ for the convergence test: stop if $\mathbf{g}^n \approx \mathbf{g}^{n+1}$; otherwise, return to Step 2.

In the algorithm steps above, the procedure for the gradient projection method to solve the trip-chain based user equilibrium traffic assignment model in Step 3 is as follows:

Step 0: Algorithm initialization

Step 0.1: Let $n=0$, set free-flow travel time $\{c_{a_0}\}$ as the starting solution for travel time of the links in the network, and calculate the shortest route that passes all of the activity points between O-D pair (r,s).

Step 0.2: Based on the starting solution, generate a set of trip-chain routes and define the flow $h_{\hat{p}}^{rs} = \overline{q}^{rs}, \forall r,s$ for the trip-chain that passes the activity nodes between O-D pair (r,s) as $\{h_{\hat{p}}^{rs}\}^{(n+1)}$.

Step 1: Calculations for the master problem

Step 1.1: Let $n=n+1$, calculate the link flows based on $\{h_{\hat{p}}^{rs}\}^{(n)}$, and update travel time $\{c_a^{(n)}(x)\}$ for each link in the network.

Step 1.2: Calculate the shortest route that passes all the activity points between O-D pair (r,s), $\hat{p}^{rs*}$, and list it as the first in the set of feasible trip-chain routes, $\{\hat{p}^{rs}\}^{(n)}$.

Step 2: Calculations for restricted master problem

Step 2.1: Use Eqs. (28)~(30) to update trip-chain route flow $\{h_{\hat{p}}^{rs}\}^{(n+1)}$ and link flow $x_a^{(n+1)}$.

$$h_{\hat{p}}^{rs(n+1)} = \max \left\{ 0, \left(h_{\hat{p}}^{rs(n)} + \alpha_{\hat{p}}^{rs(n)} d_{\hat{p}}^{rs(n)} \right) \right\} \quad \forall r \in R, s \in S, \hat{p} \ne \hat{p}^*$$ (28)

$$h_{\hat{p}^*}^{rs(n+1)} = \overline{q}^{rs} - \sum_{\hat{p} \ne \hat{p}^*} h_{\hat{p}}^{rs(n+1)}, \quad \forall r \in R, s \in S$$ (29)

$$x_a^{(n+1)} = \sum_{rs} \sum_{\hat{p}} h_{\hat{p}}^{rs(n+1)} \overline{\delta}_{a\hat{p}}^{rs}, \quad \forall a$$ (30)

Step 2.2: Convergence test: convergence is achieved if the percentage difference between two rounds in link flow is less than a certain convergence criterion φ, as

shown in Eq. (31); otherwise, return to Step 1.

$$\max_{a} \left| \frac{x_a^{(n+1)} - x_a^{(n)}}{x_a^{(n)}} \right| \leq \varphi \tag{31}$$

Let the shortest route that passes all of the activity points between O-D pair (r,s) in the route set be $\hat{p}^*$, the trip-chain link flow of which can be expressed as below:

$$h_{\hat{p}^*}^{rs} = \overline{q}^{rs} - \sum_{\hat{p} \neq \hat{p}^*} h_{\hat{p}}^{rs} \qquad \forall r,s \tag{32}$$

The descent direction $\mathbf{d}^{(n)}$ in Step 2.1 can be obtained from the first-order derivative of the trip-chain route variable $h_{\hat{p}}^{rs}$ in objective equation Eq. (16), as shown below:

$$\frac{\partial z(\mathbf{x})}{\partial h_{\hat{p}}^{rs}} = c_{\hat{p}}^{rs} - c_{\hat{p}^*}^{rs}, \quad \forall r \in R, s \in S, \hat{p} \neq \hat{p}^* \tag{33}$$

The step size $\alpha^{(n)}$ in Step 2.1 can be obtained from the reciprocal of the second-order partial derivative of the trip-chain route variable $h_{\hat{p}}^{rs}$ in objective equation Eq. (16), as shown below:

$$\frac{\partial^2 z(\mathbf{x})}{\partial (h_{\hat{p}}^{rs})^2} = \sum_a c_a' \overline{\delta}_{a\hat{p}}^{rs} + \sum_a c_a' \overline{\delta}_{a\hat{p}^*}^{rs} - \sum_{a \in \hat{p} \cap \hat{p}^*} 2c_a' \tag{34}$$

Thus,

$$\alpha_{\hat{p}}^{rs} = \frac{\upsilon}{\sum_a c_a' \overline{\delta}_{a\hat{p}}^{rs} + \sum_a c_a' \overline{\delta}_{a\hat{p}^*}^{rs} - \sum_{a \in \hat{p} \cap \hat{p}^*} 2c_a'}, \quad \forall r \in R, s \in S, \hat{p} \neq \hat{p}^* \tag{35}$$

where parameter υ is a constant between 0 and 1 that indicates a faster solution speed closer to 1 and higher solution precision closer to 0.

5. NUMERICAL ANALYSIS

The algorithm presented above can find a local optimum solution for the bilevel programming model for signal timing optimization. Below, we used the test network in Fig. 2 to conduct numerical analysis on the verification of the proposed model framework. In Fig. 3, Nodes 1, 5, 9, and 13 are the origins and destinations; Node 3 is the intermediate activity point that must be passed between O-D pair Nodes 1 and 13, and Node 11 is the intermediate activity point that must be passed between O-D pair Nodes 5 and 9. We adopted the FHWA cost function in Eq. (13) to calculate link costs. Table 6 displays the transportation demands and the intermediate nodes that must be passed between the various O-D pairs. The intersections have two-phase signals, and the free-flow travel time for all the links was set as 1. Table 7 presents the cycle time, minimum green time, saturation flow rate, and time loss.
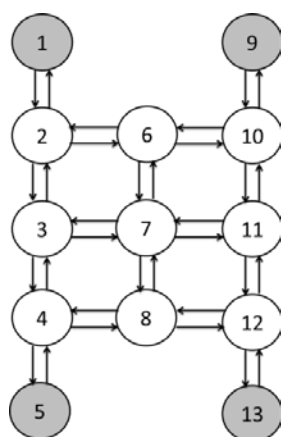
Figure. 2 Test Network 2

Table 6. Transportation demand between trip-chain O-D pairs in Test Network 2

O-D Pair	Activity Node	Travel Demand
1-13	3	40
5-9	11	50
9-5	--	40
13-1	--	50

Table 7. Intersection data for Test Network 2

Cycle Time	Initial Green Time	Minimum Green Time	Saturation Flow Rate	Lost Time
60*	27	5	60	3

*Unit: second.

We analyzed the road network and O-D pair data above using the sensitivity analysis and generalized inverse matrix approach developed in Section 4.2. The computing environment is at the PC under Pentium 4 3.4Gb Hz platform, and the solution algorithms were implementing by applying Borland C++ Version 5.02 language.

The results are shown in Fig. 3 and Tables 8 and 9. Figure 3 displays the convergence of the objection function values during the solving process of the Stackelberg solution to the upper-level model. Table 8 presents the trip-chain route decisions and user equilibrium in the lower-level model when the Stackelberg solution of the upper-level model has been obtained. Table 9 shows the travel time, flow, capacity, and green time for the road links connecting each intersection when the Stackelberg solution of the upper-level model has been obtained.
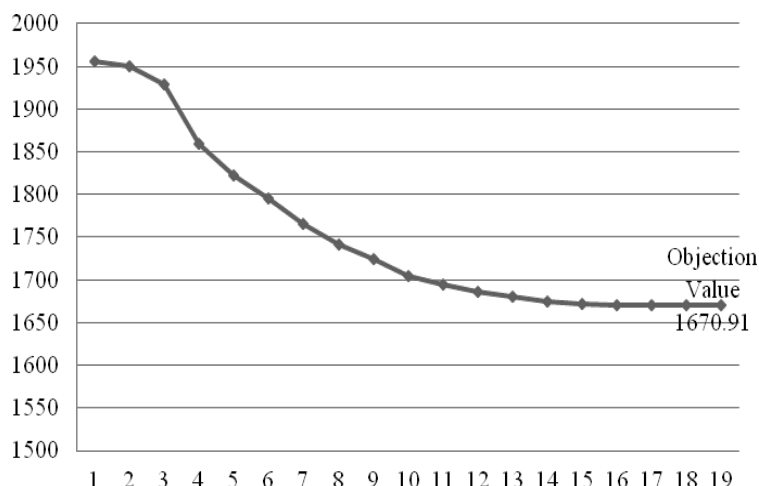
Figure 3. Convergence of objection values for Stackerlberg solutions in Test Network 2

Table 8. Link flow and travel time in Test Network 2

Trip-chain Path	Flow	Travel Cost	Obj. Fun. Values
1→2→3→2→6→10→11→12→13	5.97		
1→2→3→7→8→12→13	11.14	9.58	
1→2→3→7→11→12→13	14.34		
1→2→3→4→8→12→13	8.55		
5→4→3→7→11→10→9	11.61		
5→4→3→2→6→10→11→10→9	9.82	10.46	
5→4→8→12→11→10→9	17.06		
5→4→8→7→11→10→9	11.51		1670.91
9→10→11→7→8→4→5	10.77		
9→10→11→7→3→4→5	0.01	7.56	
9→10→6→7→3→4→5	21.13		
9→10→6→7→8→4→5	8.09		
13→12→8→7→3→2→1	8.09		
13→12→8→4→3→2→1	10.69		
13→12→8→7→6→2→1	12.26	9.24	
13→12→11→7→3→2→1	0.38		
13→12→11→7→6→2→1	15.57		

Table 9. Stackelberg solutions of Test Network 2

No.	Link	Free Flow Travel Time	Travel Time	Flow	Capacity	Green Time Ratio
1	1→2	1	1.37	40.00	31.95	0.53
2	2→1	1	1.9	50.00	31.95	0.53
3	2→3	1	1.37	40.00	31.95	0.53
4	2→6	1	1.04	15.79	22.05	0.37
5	3→2	1	1.22	34.96	31.95	0.53
6	3→4	1	1.11	29.69	31.95	0.53
7	3→7	1	2.2	37.09	22.05	0.37
8	4→3	1	1.15	32.12	31.95	0.53
9	4→5	1	1.2	40.00	37.35	0.62
10	4→8	1	2.2	37.12	22.05	0.37
11	5→4	1	1.9	50.00	31.95	0.53
12	6→2	1	1.57	30.83	22.05	0.37

13	6→7	1	1.11	29.22	31.95	0.53
14	6→10	1	1.12	15.79	16.65	0.28
15	7→3	1	1.49	29.62	22.05	0.37
16	7→6	1	1.13	30.83	31.95	0.53
17	7→8	1	1.12	29.99	31.95	0.53
18	7→11	1	2.25	37.46	22.05	0.37
19	8→4	1	1.48	29.55	22.05	0.37
20	8→7	1	1.15	31.87	31.95	0.53
21	8→12	1	2.16	36.75	22.05	0.37
22	9→10	1	1.2	40.00	37.35	0.62
23	10→6	1	1.46	29.22	22.05	0.37
24	10→9	1	1.48	50.00	37.35	0.62
25	10→11	1	1.07	26.57	31.95	0.53
26	11→7	1	1.5	29.73	22.05	0.37
27	11→10	1	1.48	50.00	37.35	0.62
28	11→12	1	1.02	20.31	31.95	0.53
29	12→8	1	1.59	31.05	22.05	0.37
30	12→11	1	1.24	36.01	31.95	0.53
31	12→13	1	1.37	40.00	31.95	0.53
32	13→12	1	1.9	50.00	31.95	0.53

The results of the numerical analysis on Test Network 2 indicate the following:

1. The results in Fig. 4 show that in the solution algorithm, the objective function value of each round decreases until convergence. This indicates that the combination of the variational inequalities sensitivity analysis approach and the generalized inverse matrix method can effectively obtain the derivative of the implicit function of the decision variables in the upper and lower level models in the signal timing optimization model with trip-chain route choices behavior in this study. Furthermore, the proposed approach can produce the descent direction for the objective function in the upper-level model, which facilitates the search for the Stackerlberg solution of the model.

2. The results in Table 8 reveal that the used trip-chain paths between each O-D pair have the same travel time and satisfy the trip-chain user equilibrium conditions when the Stackerlberg solution of the upper-level model has been obtained.

3. This study optimizes the design of signal timings at intersections while minimizing total costs in the road network, as shown in Table 9. This is achieved through solution of the signal timing optimization model with trip-chain route selection behavior. The proposed model is therefore a pioneer component worthy of adding to the transportation network science literature.

6.CONCLUSIONS AND SUGGESTIONS

Previous studies have pointed out that the factors influencing route choice behavior include incidental activities during travel as well as the locations of origins and destinations. We therefore developed a trip-chain based bilevel programming model for signal timing optimization that makes more reasonable assumptions of user route choice behavior. The lower-level of the model developed in this study could be used to calculate the trip demands with and without activities nodes between OD pairs based on trips as well as based on trip-chains. Therefore, this increases the generalizability of the model as well as the flexibility of problem analysis.

To solve the proposed bilevel optimization model, we adopted an approach combining sensitivity analysis for variational inequalities, the generalized inverse matrix method, and a gradient projection method developed for trip-chain user equilibrium behaviors. The numerical tests on Test Network 1 demonstrated that the results derived using the combined method approximate the actual trip-chain user equilibrium flow solutions. This approach can therefore effectively estimate the changes in link flow and provide the descent direction for the objective function in the upper-level model. The numerical analysis results based on Test Network 2 showed that the solution algorithm developed in this study can produce an appropriate descent direction at the end of each round during the solution of the bilevel programming model as well as a converging target value. The final optimized signal timing results can fulfill the constraints on trip-chain route choice behaviors.

To further increase the alignment of signal timing designs with actual road network traffic demands, we suggest the following directions for future research:

1) The model in this study was established in light of a closed road network system with normal two-phase signal settings. For the sake of convenience in the model verification, the developed model and adopted parameters did not take the changes in cycle time. This requires improvement in the future.
2) The inclusion of time dependence and dynamic network designs in the trip-chain-based bilevel programming model for signal timing optimization would increase the accuracy of prediction of network demand so that the transportation demand can be satisfied in real time.
3) This study did not consider multiple vehicles types trip-chain route selection behaviors or the influence of link capacity on signal timing designs. Giving consideration to multiple vehicles types, such as urban bus and scooter that are common in Taiwan, and capacity limits would enable the model to more accurately reflect actual trip-chain route selection behaviors.
4) Numerical analysis shows that the first-order partial derivative of link flow with regard to the perturbation parameter of green time derived using the sensitivity analysis method for variational inequalities produces results that approximate the actual equilibrium flows. This approach can therefore effectively estimate changes in link flows. Nevertheless, the error increases with the perturbation parameter, and thus the influence of the extent of perturbation on the solution process warrants further investigation.

ACKNOWLEDGEMENTS

REFERENCES

Allsop, R. (1974) Some Possibilities for Using Traffic Control to Influence Trip Distribution and Route Choice. in Buckley, D. J. (eds), *Proceedings of the Sixth International Symposium on Transportation and Traffic Theory*, 345–374, Elsevier, Sydney.

Chen, H. K. and Hsueh, C. F. (1997) Combining signal timing plan and dynamic traffic assignment. *Presented at the 76th Annual Meeting of the Transportation Research Board*, Washington, D.C., U.S.A.

Chen, H. K. and Chou, C. Y. (2001) A dynamic signal timing control problem.

Transportation Planning Journal, 30(4), 823-848.

Cho, H. J. (1991) Generalized inverse approach to sensitivity analysis of equilibrium network flow. *Transportation Planning Journal*, 20(1), 1-14.

Cipriani, E. and Fusco, G. (2004) Combined signal setting design and traffic assignment problem. *European Journal of Operational Research*, 155(3), 569-583.

Currie, G. and Delbosc, A. (2011) Exploring the Trip Chaining Behaviour of Public Transport Users in Melbourne," Transport Policy, Vol. 18, pp. 204-210.

Fisk, C. S. (1984) Game theory and transportation system modeling. *Transportation Research Part B*, 14(3), 301-313.

Gartner, N.H. (1974) Area traffic control and network equilibrium methods. In: Floran, M.(eds), *Traffic Equilibrium Methods*. 274-297, Springer-Verlag, Berlin.

Hägerstrand, T. (1970) What about people in regional science. *Papers of The Regional Science Association*, 24(1), 7-21.

Higuchi, T., Shimamoto, H., Uno, N., and Shiomi, Y. (2011) A trip-chain based combined mode and route choice network equilibrium model considering common lines problem in transit assignment model. *Procedia Social and Behavior Sciences*, 20, 354-363.

Hensher, D. and Reyes, A. (2000) Trip chaining as a barrier to the propensity to use public transport. *Transportation*, 27, 341-361.

McGuckin, N., Zmud, J., and Nakamoto, Y. (2005) Trip chaining trends in the united states: understanding travel behavior for policy making. *Transportation Research Record*, 1917, 199-204.

Marcotte, P. (1983) Network optimization with continuous control parameters. *Transportation Science,* 17(2), 181-197.

Maruyama, T. and Harata, N. (2005) Incorporating trip chaining behavior in network equilibrium analysis. *Transportation Research Record*, 1921, 11-18.

Maruyama, T. and Harata, N. (2006) Difference between area-based and cordon-based congestion pricing: investigation by trip-chain-based network equilibrium model with non-additive path costs. *Transportation Research Record*, 1964, 1-8.

Maruyama, T. and Sumalee, A. (2007) Efficiency and equity comparison of cordon- and area-based road pricing schemes using a trip-chain equilibrium model. *Transportation Research Part A*, 41, 655-671.

Morency, C. and Valiquette, F. (2010) Trip chaining and its impact on travel behaviour. *Activity-Based Analysis and Modeling*, 12th World Conference on Transport Research, Lisbon, Portugal.

Smith, M. (2011) Dynamics of route choice and signal control in capacitated networks. *Journal of Choice Model*, 4(3), 30-51.

Chiou, S. W. (2008) A hybrid approach for optimal design of signalized road network. *Applied Mathematical Modelling*, 32, 195-207.

Chiou, S. W. (2014) Optimal signal-setting for road network with maximum capacity. *Information Sciences*, 273, 287-303.

Tobin R. L. (1986) Sensitivity analysis for variational inequalities. *Journal of Optimization Theory and Applications*, 48(1), 191-204.

Tobin, R. L. and Friesz, T. (1988) Sensitivity analysis for equilibrium network flow. *Transportation Science*, 22(4), 242-250.

Ukkusuri1, S., Doan, K., and Abdul Aziz, H. M. (2013) A bi-level formulation for the combined dynamic equilibrium based traffic signal control. *Procedia - Social and Behavioral Sciences,* 80(7),729-752.

Wong, S.C. and Yang, H. (1997) Reserve Capacity of a Signal-Controlled Road Network. *Transportation Research Part B*, 31(5), 397-402.

Wang, C. Y. and Chen, S. W. (2014) A Trip-Chain Based User Equilibrium Traffic Assignment Model. *Proceeding of 2013 International Conference of the Chinese Institute of Transportation*, 191-212. ILan, Taiwan.

Wang, C. Y. (1999) *Dynamic travel choice models with link capacity side constraints*. Ph.D Dissertation, Department of Civil Engineering, National Central University, Taiwan.

Yang, H. and Yagar S. (1995) Traffic assignment and signal control in saturated road networks. *Transportation Research Part A,* 29(2),125-139.

Zhao, Z., Chua, G., and Zhao, J. (2012) Evolution of trip chaining patterns in London from 1991 to 2010. *Innovations in Improving the Sensitivity of Models*, 4th Conference on Innovations in Travel Modeling, Tampa, FL.