

# An algorithm for local continuous optimization of traffic signals

Jiang Qian Ying <sup>a,\*</sup>, Huapu Lu <sup>b,1</sup>, Jing Shi <sup>b,1</sup>

<sup>a</sup> Faculty of Regional Studies, Gifu University, Yanagido 1-1, Yanagido, Gifu 501, Japan

<sup>b</sup> Institution of Transportation Engineering, Tsinghua University, China

Received 2 November 2004; accepted 10 October 2005

Available online 3 May 2006

## Abstract

In this paper, an algorithm for sensitivity analysis for equilibrium traffic network flows with link interferences is proposed. Based on this sensitivity analysis algorithm, a general algorithm is provided for solving the optimal design and management problems for traffic networks. In particular, this algorithm is applied to the optimal traffic signal setting problem. A numerical example is given to demonstrate the effectiveness of our algorithm.

© 2006 Elsevier B.V. All rights reserved.

**Keywords:** Signal control; Link interference; Asymmetric link performance function; Sensitivity analysis; Continuous optimization

## 1. Introduction

Recently, signal setting that optimizes some network wide performance index has been studied based on the equilibrium modeling of traffic networks. See, e.g., Yang and Yagar (1995), Chiou (1999) and Wong et al. (2001a,b). The use of equilibrium modeling, which accounts for the driver's time sensitive routing behavior, is necessary when the influence of signal changes on flow patterns on the whole network is considered. Sensitivity analysis, which is by definition the computation of partial derivatives of link flows and link travel times with respect to network design variables, e.g., traffic

signal parameters, is used for finding the direction for adjusting the signal control variables. Recent works on signal optimization are based on the Wardropian equilibrium (cf. Wardrop, 1952) modeling of the network traffic flows. Wardropian equilibrium is a state of traffic network in which all the drivers choose minimum time paths. Unfortunately, the corresponding sensitivity analysis algorithms which are modified from the algorithm provided originally by Tobin and Friesz (1988), have to enumerate paths explicitly and are computationally inefficient for application in realistic networks. In the literature, some works use numerical approach for estimating relevant partial derivatives. See Cipriani and Fusco (2004) for details on this approach.

Since conventional exact sensitivity analysis and optimization algorithms have to enumerate paths on the traffic network thus need a computational time that is exponential in the number of links in the network, design of algorithms of polynomial

\* Corresponding author. Tel./fax: +81 58 293 3306.

E-mail addresses: [ying@cc.gifu-u.ac.jp](mailto:ying@cc.gifu-u.ac.jp) (J.Q. Ying), [luhp@mail.tsinghua.edu.cn](mailto:luhp@mail.tsinghua.edu.cn) (H. Lu), [jingshi@mail.tsinghua.edu.cn](mailto:jingshi@mail.tsinghua.edu.cn) (J. Shi).

<sup>1</sup> Tel.: +86 10 6277 2615; fax: +86 10 6277 1132.

computational time is the key step toward the development of efficient computational methods for solving traffic network optimization problems.

In this paper, an algorithm is said to be *efficient* if it has a computational time complexity that is a polynomial in the number of links of the traffic network. The purpose of this paper is just to develop efficient algorithms for solving sensitivity analysis and signal optimization problems for traffic networks.

In Ying and Miyagi (2001), efficient algorithms for sensitivity analysis of stochastic user equilibrium network flows have been developed that do not need to enumerate paths. Stochastic user equilibrium is a state of traffic network where each driver chooses a path which has minimum perceived travel time (see the next section). However, the algorithms present in Ying and Miyagi (2001) apply only to networks that do not have link interferences. Because at intersections there usually exist strong interferences between link flows, a general equilibrium model that can treat these interferences has to be developed for solving the optimal signal setting problem. A sensitivity analysis algorithm for such a general equilibrium model is proposed in this paper. Based on this algorithm, efficient computational methods are proposed for solving the traffic signal optimization problem. By using realistic time delay formulas that capture the interference properties of link flows at intersections, the efficiency of the proposed sensitivity algorithm for optimal signal setting is demonstrated.

In the following section the stochastic user equilibrium (SUE) model for a general network with link interferences is formulated. A sensitivity analysis method for computing the derivatives of equilibrium link flows and link travel times with respect to signal control variables is presented in Section 3. A framework of applying this algorithm for optimal signal setting is proposed in Section 4, which is illustrated by a numerical example in Section 5. Some problems regarding practical applications are discussed in Section 6.

## 2. SUE for network with interfering links

Stochastic user equilibrium (SUE) is a state of traffic network in which each driver chooses a path which has a *perceived* minimum travel time. In general, there are random errors in the driver's perception of travel time. Details on network equilibrium analysis can be found in, e.g., Sheffi (1985). In the

following a self contained SUE formulation is outlined.

### 2.1. Stochastic path choice behavior and traffic assignment

A traffic network can be represented by a directed graph which consists of a set of links  $A = \{a, b, \dots\}$ , and a set of nodes  $N = \{i, j, \dots\}$ . The pair of the origin and the destination of a trip is called an OD pair. Let  $W$  denote the set of OD pairs. Let  $q_{rs}$  be the volume of travel demand for OD pair  $rs \in W$ . ( $q_{rs}$  will be referred to as OD demand later). Let  $R_{rs}$  be the set of paths connecting  $r$  and  $s$ ,  $rs \in W$ .

Let  $t = (t_a, a \in A)$  denote the vector of travel times  $t_a$  incurred in the network links  $a \in A$ . Let  $C_k^{rs} = \sum_{a \in A} t_a \delta_{k,a}^{rs}$  be the travel time of path  $k$  from origin  $r$  to destination  $s$ , where  $\delta_{k,a}^{rs} = 1$  if  $a$  is a link in the path  $k$ , and  $\delta_{k,a}^{rs} = 0$  otherwise.

Assume that there is some *reference* set of path travel times  $(C_k^{rs}; rs \in W, k \in R_{rs})$  based on which drivers perceive path travel times. As will be described later, this reference vector depends on the average link flows in the network.

Assume that for a driver from  $r$  to  $s$ , the perceived travel time of path  $k$  is  $C_k^{rs} + \frac{1}{\theta} \xi_k$ , where  $\xi_k$  is a random variable with the following Gumbel distribution function (Sheffi, 1985)  $\Pr(\xi_k \leq \omega) = \exp(-\exp(-\omega + E))$ ,  $E = 0.5708$   $\theta > 0$  is a parameter characterizing the scale of the effect of perception error. This perception randomness is assumed to be the pure source of randomness in the traffic network.

Assume that the random variables  $\xi_k$ ,  $k \in R_{rs}$ , are independently and identically distributed. Then it can be derived that the probability that a path  $k \in R_{rs}$  is chosen is given by

$$P_k^{rs} = \Pr \left[ C_k^{rs} + \frac{1}{\theta} \leq \max_{p \in R_{rs}, p \neq k} \left\{ C_p^{rs} + \frac{1}{\theta} \right\} \right] = \frac{\exp(-\theta C_k^{rs}(t))}{\sum_p \exp(-\theta C_p^{rs}(t))}. \quad (1)$$

The probability that a traveler from  $r$  to  $s$ , that is passes link  $a$  is given by

$$P_a^{rs} = \sum_{k \in R_{rs}} \frac{\exp(-\theta C_k^{rs}(t)) \delta_{k,a}^{rs}}{\sum_{p \in R_{rs}} \exp(-\theta C_p^{rs}(t))} = \frac{\partial S_{rs}(t)}{\partial t_a}, \quad (2)$$

where

$$S_{rs} = -\frac{1}{\theta} \ln \left[ \sum_{p \in R_{rs}} \exp(-\theta C_p^{rs}(t)) \right] \quad (3)$$

is the expected minimum travel time from  $r$  to  $s$

$$S_{rs} = E \left[ \min_{k \in R_{rs}} \left\{ C_k^{rs} + \frac{1}{\theta} \zeta_k \right\} \right].$$

Although this term is not essential for the problem to be dealt with in this paper, it will be used as a convenient notation.

In (1), if  $C_k^{rs} > C_l^{rs}$ , for some path  $l$ , then  $P_k^{rs} \rightarrow 0$ , for  $\theta \rightarrow \infty$ . This means that in the extreme situation of  $\theta \rightarrow \infty$ , only the paths with minimum travel time will be chosen. This kind of routing behavior yields the so called Wardropian equilibrium of traffic network (see Wardrop, 1952 or Sheffi, 1985).

Given the link travel time vector  $\mathbf{t} = (t_a, a \in A)$ , if  $P_a^{rs}$  are known, then the link flows are immediately computed as

$$x_a = \sum_{rs \in W} q_{rs} P_a^{rs}. \quad (4)$$

This process of computing link flows is called traffic assignment.

Although  $P_a^{rs}$  are expressed as explicit functions of path travel times, they can be computed in a *link-based* manner in the sense that the path travel times  $C_k^{rs}$  do not need to be explicitly compared with and that the computation only involves link variables and is finished in polynomial time. A widely used link-based algorithm for traffic assignment was provided by Dial in 1971 (see Dial, 1971, or Sheffi, 1985). As is stated previously, such a link-based algorithm is efficient.

## 2.2. Stochastic user equilibrium

Given the link travel time vector  $\mathbf{t} = (t_a; a \in A)$ , the vector  $\mathbf{x} = (x_a; a \in A)$  of average link flows can be obtained as a result of driver's stochastic path choosing behavior described above

$$x_a = \sum_{rs \in W} q_{rs} P_a^{rs}(\mathbf{t}), \quad a \in A. \quad (4')$$

On the other hand, this deterministic link flow vector decides a deterministic pattern of travel times on links in the network. In general, link travel time is a function of this flow vector and some control vector  $\lambda$ , that is,

$$t_a = t_a(\mathbf{x}, \lambda), \quad a \in A,$$

or in a compact form

$$\mathbf{t} = \mathbf{t}(\mathbf{x}, \lambda). \quad (5)$$

Assume that this set of deterministic travel times is the reference travel time set based on which drivers estimate the path travel times. Then drivers routing behaviors are in an equilibrium state when the average link flows computed by (4') coincide with the link flows which are the arguments in link travel time function (5).

The stochastic user equilibrium on the traffic network is then characterized by the following nonlinear equations:

$$\begin{aligned} F_a(\mathbf{x}, \lambda) &= x_a - \sum_{rs \in W} q_{rs} \frac{\sum_{k \in R_{rs}} \exp(-\theta C_k^{rs}(\mathbf{t})) \delta_{k,a}^{rs}}{\sum_{p \in R_{rs}} \exp(-\theta C_p^{rs}(\mathbf{t}))} \\ &= x_a - \sum_{rs \in W} q_{rs} \frac{\partial S_{rs}(\mathbf{t})}{\partial t_a} = 0, \quad a \in A, \end{aligned} \quad (6)$$

where  $\lambda$  is contained in the terms  $\partial S_{rs}(\mathbf{t}) / \partial t_a$ . Eq. (6) are rewritten in a vector form

$$\mathbf{F}(\mathbf{x}, \lambda) = 0. \quad (7)$$

In the equilibrium model, the network structure and the OD data are assumed to be given. The parameter  $\theta$  is usually determined by experience. Link flow vector is the state variable in the equilibrium model.  $\lambda$  is a vector of control variables that can be adjusted in order to obtain desirable equilibrium state.

In this paper, we will particularly consider signal parameters as control variables. In general, the signal parameters may include cycle time, green splits and off-sets between signals. Suppose that there are a total of  $I$  signal parameters  $\lambda_1, \dots, \lambda_I$ , the control vector is expressed as  $\lambda = (\lambda_1, \dots, \lambda_I)$ .

For given signal control variables, any standard computational method developed for solving general non-linear equations can be used to find the equilibrium flow solutions. There have also been developed various computational methods for solving for the traffic network equilibria (see, e.g., Patriksson, 1994 for a review). The following method of successive averages (MSA, see Sheffi, 1985, or Patriksson, 1994) will be used later in this paper for finding equilibrium flow solutions in a numerical example for signal optimization.

## 2.3. Method of successive approximation

*Step 1.* Set an initial vector of link flows  $\mathbf{x}^1$ . Set  $n = 1$ .

*Step 2.* Compute  $\mathbf{t}^n = \mathbf{t}(\mathbf{x}^n, \lambda)$ .

*Step 3.* Based on  $\mathbf{t}^n$ , compute the link flow vector  $\mathbf{y}^n$  by Dial's assignment algorithm.

**Step 4.** Set  $\mathbf{x}^{n+1} = \mathbf{x}^n + (\mathbf{y}^n - \mathbf{x}^n)/n$ . If  $|\mathbf{y}^n - \mathbf{x}^n| <$  some threshold, then stop; Else, set  $n = n + 1$ , go to Step 2.

Interested readers may find details on properties of the MSA for solving the general network equilibrium equations in [Daganzo \(1983\)](#).

### 3. Sensitivity analysis

The basic procedure of sensitivity analysis is the computation of the derivatives of link flows  $\mathbf{x}$  with respect to  $\lambda$ . For the case that travel time on a link is determined solely by the flow on the link, for fixed network parameter  $\lambda$ , i.e.,  $t_a = t_a(x_a, \lambda)$ ,  $a \in A$ , an efficient sensitivity algorithm has been proposed by [Ying and Miyagi \(2001\)](#). In the following this algorithm is extended to the general case that travel time on a link depends not only on the flow on the link, but also on the flows on other links. This is the situation that there are interferences between link flows, which happens mainly at intersections.

In the following, a matrix with components  $m_{gh}$ , for indices  $g$  and  $h$  belonging to sets  $G$  and  $H$ , respectively, will be expressed as  $(m_{gh}; g \in G, h \in H)$ . From the SUE equation  $\mathbf{F}(\mathbf{x}, \lambda) = 0$  we have

$$\frac{\partial \mathbf{x}}{\partial \lambda} = -(\nabla_{\mathbf{x}} \mathbf{F})^{-1} \frac{\partial \mathbf{F}}{\partial \lambda}, \quad (8)$$

where

$$\nabla_{\mathbf{x}} \mathbf{F} = \left( \frac{\partial F_a}{\partial x_b}; a \in A, b \in A \right) \quad (9)$$

and

$$\frac{\partial F_a}{\partial x_b} = \delta_{a,b} - \sum_{rs} q_{rs} \frac{\partial(\partial S_{rs}/\partial t_a)}{\partial x_b}. \quad (10)$$

$\delta_{a,b} = 1$  if  $a = b$  and  $\delta_{a,b} = 0$  otherwise. Let  $A_b = \{c, \dots\}$  be the set of links interfered by link  $b$ . Then we have

$$\frac{\partial(\partial S_{rs}/\partial t_a)}{\partial x_b} = \sum_{c \in A_b} \frac{\partial^2 S_{rs}}{\partial t_a \partial t_c} \frac{\partial t_c}{\partial x_b}. \quad (11)$$

In [Ying and Miyagi \(2001\)](#), it has been shown that  $\partial^2 S_{rs}/\partial t_a \partial t_c$  can be computed in a link-based manner by using Dial's algorithm. The link performance function  $t_c(x_c, x_b, \dots; \lambda)$  is determined by the physical design of the links (including links at intersections) and can usually be computed in a few steps. This fact will be seen from the numerical example given in Section 5 where typical practical function

forms are provided. It then follows that  $\nabla_{\mathbf{x}} \mathbf{F} = (\partial F_a/\partial x_b; a \in A, b \in A)$  can be computed efficiently, with a polynomial complexity. The term  $\partial \mathbf{F}/\partial \lambda = (\partial F_a/\partial \lambda_k; a \in A, 1 \leq k \leq I)$  can be computed in the following way. Let  $(t_c)_{\lambda_k}$  be the derivative of  $t_c$  with respect to  $\lambda_k$  when all the link flows are fixed, and  $(\mathbf{t})_{\lambda} = ((t_a)_{\lambda_k}; a \in A, 1 \leq k \leq I)$ . Then we have

$$\frac{\partial F_a}{\partial \lambda_k} = - \sum_{rs \in W} q_{rs} \sum_{c \in A_k} \frac{\partial^2 S_{rs}}{\partial t_a \partial t_c} (t_c)_{\lambda_k}, \quad (12)$$

where  $A_k$  is the set of links whose travel times depend on  $\lambda_k$ .

As the inverse matrix can be computed in polynomial time in the size of the matrix, which is the number of links of the network. Therefore  $\partial \mathbf{x}/\partial \lambda = -(\nabla_{\mathbf{x}} \mathbf{F})^{-1} \partial \mathbf{F}/\partial \lambda$  can be computed in polynomial time in the number of links of the network.

Subsequently, we can obtain

$$\frac{\partial \mathbf{t}}{\partial \lambda} = \frac{\partial \mathbf{t}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \lambda} + (\mathbf{t})_{\lambda}, \quad (13)$$

where

$$\frac{\partial \mathbf{t}}{\partial \mathbf{x}} = \left( \frac{\partial t_a}{\partial x_b}; a \in A, b \in A \right).$$

For a network without interfering links,  $\partial \mathbf{t}/\partial \mathbf{x} = \text{diag}(\partial t_a/\partial x_a, a \in A)$  is a diagonal matrix. In general, if link interferences exist, the matrix will have non-diagonal elements. However, as is the case for a real world network, the range of links affected by some link flow is very limited, therefore the matrix is in general sparse in the sense that most of its elements are zero.

### 4. Optimal signal setting

In practice, the condition of a traffic network is usually evaluated by levels of services in roads and intersections. Level of service is usually measured by speed and delay time (cf. [Transportation Research Board, 2000](#)), which are, ultimately, reflected as travel times on network links. As a single index for evaluating traffic network condition, the following total travel time spent by all the vehicles in the network is widely used ([Sheffi, 1985](#)).

$$f(\lambda) = \sum_{rs \in W} \sum_{k \in R_{rs}} q_{rs} C_k^{rs} = \sum_{a \in A} x_a t_a. \quad (14)$$

In this section, we consider the problem of setting the signal variables so that this total time function is minimized. A more precise formulation may be one that imposes upper constraints on individual

link travel times. Such a formulation can be obtained by extending our basic formulation without individual link time constraints.

The gradient of  $f$  is given by

$$\frac{\partial f}{\partial \lambda} = \left( \frac{\partial f}{\partial \lambda_k}; 1 \leq k \leq I \right), \quad (15)$$

where

$$\frac{\partial f}{\partial \lambda_k} = \sum_a x_a \frac{\partial t_a}{\partial \lambda_k} + \sum_a t_a \frac{\partial x_a}{\partial \lambda_k}. \quad (16)$$

The gradient can be efficiently computed because so are the derivatives  $\partial t/\partial \lambda$  and  $\partial x/\partial \lambda$  by using the sensitivity analysis algorithm provided in Section 3.

There are usually some linear inequality constraints imposed on the control vector. For example, if  $\lambda_k$  is the green split at an intersection, a usual constraint is  $0 \leq \lambda_k \leq 0.9$ , which will be used in the following. Using the gradient data, one can use some standard mathematical programming techniques (cf., e.g., Luenberger, 1996 or Calamai and Moré, 1987) to find a local optimal solution with inequality constraints. In next section we will use the following projected gradient method for finding optimal signal green splits.

- Step 1. Set initial values for feasible signal green splits.
- Step 2. Compute the equilibrium link flows and times by the MSA algorithm.
- Step 3. Compute the gradient by the sensitivity analysis method.
- Step 4. Project the gradient on the cube  $\prod_{i=1}^I [0, 0.9]$ .
- Step 5. Do line search along the projected gradient to find a minimum.
- Step 6. Stop if certain convergence condition is satisfied; otherwise repeat from Step 2.

## 5. A numerical example

### 5.1. Delay functions at intersections

Two-phase signal-controlled intersections will be considered in this example, because such intersections have the most sophisticated characteristics of interruptions between traffic flows. In Fig. 1 is shown the composition of a North-South phase and a West-East phase in a traffic signal. Let  $C$  be the cycle length of the signal,  $G_{NS}$  and  $G_{WE}$  be the green times for the North-South and the West-East phases, respectively. Let  $g_{NS} = G_{NS}/C$ ,  $g_{WE} = G_{WE}/C$ , be the green splits for the phases. A certain time

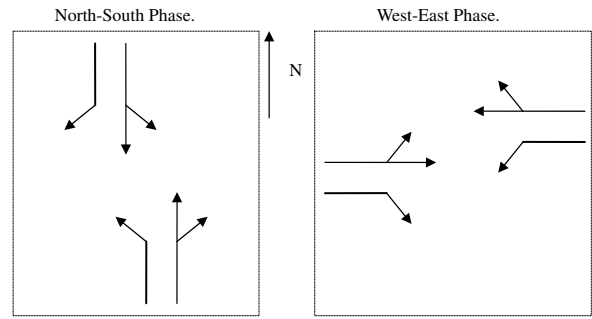


Fig. 1. The traffic signal phases.

loss occurs during switching over between the phases, its ratio to the cycle is assumed to be 0.1 in the following. Thus we have  $g_{NS} + g_{WE} = 0.9$ . Suppose that at an intersection, as shown in Fig. 2, the through and the right turning traffic flows share a common lane, while the left turning traffic flow occupies a single lane. In the following the maximal flow rate that can pass a lane within unit green time is called the *saturation flow rate* of the lane in the traffic engineering literature (cf. Transportation Research B, 2000).

Given the saturation flow rate, if the approach traffic in some lane at an intersection obeys a Poisson distribution, the average delay can be calculated by the following formula (p. 353, Sheffi, 1985).

$$d(x) = \frac{C(1-g)^2}{2(1-\rho)} + \frac{\rho/g}{2S(g-\rho)}, \quad (17)$$

$S$ : saturation flow rate of the approach lane;  $\rho = x/S$ : normalized flow;  $x$ : approach flow;  $g = G/C$ : green time split;  $G$ : green time for the approach lane;  $C$ : cycle length.

The saturation flow rate for the through and the right turning traffic flows are determined by the geometric conditions of the corresponding lane and are

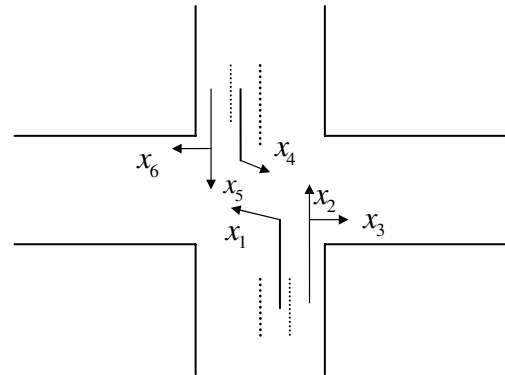


Fig. 2. A two-phase signal controlled intersection.

independent of signal parameters. A standard value for such a lane is 1800 car/hour, which is assumed to be the saturation flow rate of the lane for the through and the right turning traffic flows here.

The saturation flow rate for the left turning lane, however, depends on the opposing through flow in the following manner. As shown in Fig. 2, in the green time for the North-South phase, the flow  $x_1$  that turns left is interrupted by the opposing through flow  $x_5$ . During the green time, left turning traffic can pass by the gaps occurring in the opposing through traffic. Based on some statistical distribution assumption for traffic arrivals, the rate of left

turning flow that can pass during the green time can be estimated as (Miller, 1968)

$$1800\alpha(x_5)(S_5g - x_5)/(S_5 - x_5), \quad (18)$$

where  $S_5$  is the saturation flow rate of the opposing through traffic lane,  $\alpha$  is a function in  $x_5$ . From Japan's data (p. 234, Iida, 1992), this function can be estimated as

$$\alpha(x_5) = \exp(-0.432x_5/1000). \quad (19)$$

The maximal flow rate that can pass the lane within unit green time is then equal to

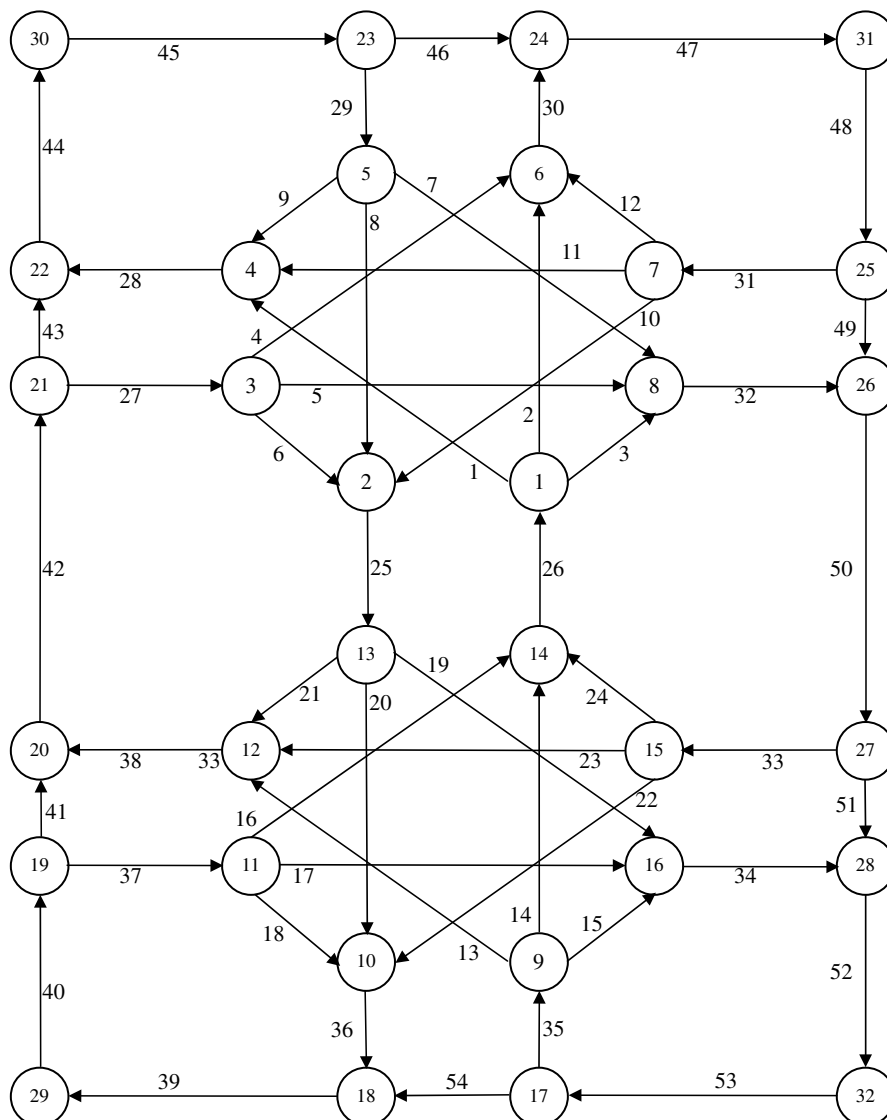


Fig. 3. The traffic network with two intersections.



$$\begin{aligned} & [1800\alpha(x_5)(S_5g - x_5)/(S_5 - x_5)]/g \\ & = 1800\alpha(x_5)(S_5 - x_5/g)/(S_5 - x_5). \end{aligned}$$

In addition, the number of cars that can pass the intersection during the switching over between phases is estimated as  $120/C$  ( $C$  in minute, cf. p. 234, Iida, 1992). Therefore, the saturation flow rate of the left turning lane is

$$S_1 = 1800\alpha(x_5) \frac{S_5 - x_5/g}{S_5 - x_5} + \frac{120}{C} (\text{car/hour}). \quad (20)$$

In summary, the delay for the left turning flow can be written as

$$\begin{aligned} t_1 = d_1(x_1, x_5) &= \frac{C(1-g)^2}{2(1-x_1/S_1(x_5))} \\ &+ \frac{x_1/(S_1(x_5)g)}{2S_1(x_5)(g-x_1/S_1(x_5))}. \end{aligned} \quad (21)$$

The through and the right turning traffic flows share a same lane and have the same delay that can be expressed as

$$\begin{aligned} t_2 = t_3 = d_2(x_2, x_3) &= \frac{C(1-g)^2}{2(1-\rho)} + \frac{\rho/g}{2S_2(g-\rho)}, \\ \rho &= \frac{x_2 + x_3}{S_2}. \end{aligned} \quad (22)$$

### 5.2. Network structure

The example network, as shown in Fig. 3, contains two intersections. There are a total of 54 links and 32 nodes in the network. Links 1 through 12 represent the links at the first intersection and links 13 through 24 represent the links at the second. The remaining 30 links are common links which are assumed to have a BPR performance function (developed by the US Bureau of Public Road, see Sheffi, 1985) of the following form:

$$t_a = C0_a \left[ 1 + b_a \left( \frac{x_a}{\text{Cap}_a} \right)^4 \right], \quad (23)$$

where  $C0_a$  is the free-flow travel time on link  $a$ ,  $\text{Cap}_a$  the capacity of link  $a$ . For the common links, these parameters are set as

$$\begin{aligned} C0_{25} = C0_{26} &= 2 \text{ minute}, \quad C0_a = 0.5 \text{ minute}, \quad a = 27, \dots, 54; \\ \text{Cap}_a &= 2000 \text{ car/hour}, \quad a = 25, \dots, 54, \quad \text{and} \\ b_a &= 0.15, \quad a = 25, \dots, 54. \end{aligned}$$

In reality, the operation of signal at one intersection affects the statistical property of flows going to the

downstream intersection. In this example we assume that this effect can be neglected so that the off set between the two signals is irrelevant to the performance of the traffic network. We also fix the signal cycle lengths as  $C = 1$  minute. Thus the only parameters to be adjusted are the green splits. Let  $\lambda_1 = g_{\text{NS},1}$  and  $\lambda_2 = g_{\text{NS},2}$  be the green splits for the North South phases of the two intersections, respectively. As stated previously,  $g_{\text{NS}} + g_{\text{WE}} = 0.9$ . It follows that  $g_{\text{WE}} = 0.9 - g_{\text{NS}}$ , and the only free variables in the signal optimization problem are  $\lambda_1$  and  $\lambda_2$  with constraints  $0 \leq \lambda_k \leq 0.9$ ,  $k = 1, 2$ .

### 5.3. Computational results

The projected gradient method described at the end of Section 4 is used to find optimal green splits. Initial values for  $\lambda_k$ ,  $k = 1, 2$ , are set to be 0.45, which is the middle value in the feasible range  $[0, 0.9]$ . Four sets, as shown in Table 1, of OD travel demands are investigated.

In the model,  $\theta$  is set to be 0.2. It was found that by 3–5 iteration steps, the objective function had no further significant improvement, and the optimization procedure was stopped. The computed optimal splits, total travel times for both the initial and the optimal signal settings, and travel time reduction ratios, are shown in Table 2, for the four sets of OD demands, respectively.

For the first set of travel demands, the West-East directional demand for OD pair  $29 \rightarrow 32$  tends to shorten the North-South directional green splits. The second set is generated by cutting down this demand ( $29 \rightarrow 32$ ) by half and adding a 400 car/hour demand for the OD pair  $30 \rightarrow 29$ . For this demand set, the North-South directional green split for the first signal (0.588) becomes greater than the West-East directional split ( $0.9 - 0.588 = 0.312$ ).

Demands of the third and the fourth sets are just half of those in the first and the second, respectively.

Table 1  
Four sets of travel demands

OD ( $r \rightarrow s$ )	Travel demand set (car/hour)			
	First	Second	Third	Fourth
29 $\rightarrow$ 31	400	400	300	300
29 $\rightarrow$ 32	800	400	600	300
30 $\rightarrow$ 29	0	400	0	300
30 $\rightarrow$ 32	400	400	300	300
31 $\rightarrow$ 29	400	400	300	300
32 $\rightarrow$ 30	400	400	300	300

Table 2  
Computational results

Demand set	$(\lambda_1, \lambda_2)$	Initial time (car minute)	Minimal time (car minute)	Reduction (%)
First	(0.444, 0.318)	7503.7	7182.4	4.3
Second	(0.588, 0.315)	8070.2	7830.5	3.0
Third	(0.451, 0.265)	5302.0	5152.2	2.8
Fourth	(0.601, 0.294)	5861.0	5719.2	2.4

The results also suggest that by applying the optimizing algorithm, larger reduction of travel time can be achieved for more congested network with greater travel demands.

## 6. Conclusions

In this paper an efficient algorithm for optimal traffic signal setting based on equilibrium sensitivity analysis was proposed. The algorithm was illustrated with a numerical example which captures the sophisticated time delay characteristics at signal controlled intersections.

Our method depends on two computational techniques, one is the traffic network equilibrium solution method, the other is the sensitivity analysis method. In reality, stochastic user equilibrium solution methods have already been widely applied in the engineering fields of transportation planning and management and have also been implemented in softwares (see e.g., the INRO web site <http://www.inro.ca/en/index.php> for description of a representative transportation network analysis software EMME2). This approves the fact that the first technique for solving SUE equations is actually efficient enough for implementation for realistic traffic network. As was seen in Section 3, in addition to the computational complexity for solving the equilibrium equations, our sensitivity analysis method has to compute the inverse of a matrix of the size equal to the number of network links. This technique is therefore also efficient in the sense that it can be carried out in polynomial time. The contribution of this technique is significant because conventional techniques depend on path-enumerating algorithms which have exponential time complexity.

From the practical point of view, in a realistic urban road network, there are usually thousands of links, inversion of matrix of such a size is not hard with high performance computer. Therefore

it is proper to conclude that our method has the potential for application in realistic networks. On the other hand, there are many alternatives for the projected gradient method adopted in this paper as an optimization procedure. Incorporation of second order derivative information may also be useful for the optimization procedure. Exploration of these techniques for application in realistic traffic network signal optimization problems would be interesting and important research topics for future.

## Acknowledgements

We are grateful to the referees whose comments have greatly improved the paper. We would like to thank W. Shen, Q. Liu of Tsinghua University, and G. Mu, J.G. Bao and B. Yang of Gifu University, for their assistance in preparing the paper. We would like to thank Dr. R. Ando of Toyota Transportation Research Institute for valuable discussion on delay functions at intersections.

## References

- Calamai, P.H., Moré, J.J., 1987. Projected gradients methods for linearly constrained problems. *Mathematical Programming* 39, 93–116.
- Chiou, S-W., 1999. Optimization of area traffic control for equilibrium network flows. *Transportation Science* 33, 279–289.
- Cipriani, E., Fusco, G., 2004. Combined signal setting design and traffic assignment problem. *European Journal of Operational Research* 155, 569–583.
- Daganzo, C.F., 1983. Stochastic network equilibrium with multiple vehicle types and asymmetric, indefinite link cost Jacobians. *Transportation Science* 17, 282–300.
- Dial, R.B., 1971. A probabilistic mutipath traffic assignment algorithm which obviates path enumeration. *Transportation Research* 5 (2), 83–111.
- Iida, Y. (Ed.), 1992. *Traffic Engineering*. Kokumin Kagaku Sha, Tokyo (in Japanese).
- Luenberger, D.G., 1996. *Linear and Nonlinear Programming*, second ed. Addison-Wesley, Reading, Mass.
- Miller, A.J., 1968. The capacity of signalized intersections in Australia. Australian Road Research Board, Bulletin No. 3.
- Patriksson, M., 1994. *The Traffic Assignment Problem, Models and Methods*. VSP, Utrecht, The Netherlands.
- Sheffi, Y., 1985. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Englewood Cliffs, NJ.
- Tobin, R.L., Friesz, T.L., 1988. Sensitivity analysis for equilibrium network flows. *Transportation Science* 22, 242–250.
- Transportation Research Board, 2000. *Highway Capacity Manual*. National Research Council, Washington, DC.
- Wardrop, J.G., 1952. Some Theoretic Aspects of Road Traffic Research, *Proceedings, Institute of Civil Engineers*, II (1), pp. 325–378.



- Wong, S.C., Yang, C., Lo, H.K., 2001a. A path-based traffic assignment algorithm based on the TRANSYT traffic model. *Transportation Research* 35B, 163–181.
- Wong, S.C., Yang, C., Tong, C.O., Wong, C.K., 2001b. Group-based optimization of signal timings for traffic equilibrium network. *Journal of the Eastern Asia Society for Transportation Studies* 4 (4), 133–148.
- Yang, H., Yagar, S., 1995. Traffic assignment and signal control in saturated road networks. *Transportation Research* 19B, 125–139.
- Ying, J.Q., Miyagi, T., 2001. Sensitivity analysis for stochastic user equilibrium network flows—a dual approach. *Transportation Science* 35, 124–133.