Optimization of Area Traffic Control for Equilibrium Network Flows

SUH-WEN CHIOU

University of London Centre for Transport Studies, University College London, London, England

A bilevel programming approach is used to tackle an optimization problem for area traffic control and equilibrium flows. The signal timing plan is defined by common cycle time, and by starts and durations of greens. The system performance index is defined as the sum of a weighted linear combination of rate of delay and number of stops per unit time for all traffic streams, which is evaluated by the traffic model from TRANSYT. User equilibrium traffic assignment is formulated as a variational inequality problem. Approximate mathematical expressions for various components of the performance index and the average delay to a vehicle at the downstream junction in the TRANSYT model for both undersaturated and oversaturated links have been derived. For a locally optimal search, the gradient projection method is used in deciding whether feasible descent directions leading to a Karush–Kuhn–Tucker point, which is potentially a local optimum, can be identified. A global search heuristic is proposed in this paper by which successively better Karush–Kuhn–Tucker points can be found with reasonable computation effort. The mixed search procedure, including the locally optimal search and global search heuristic, is proposed. Encouraging results for Allsop and Charlesworth's network have confirmed that the approximately optimizing mixed search procedure in solving the area traffic control optimization problem can achieve substantially better results than does the nonoptimizing calculations of mutually consistent signal timings and link flows.

The combined problem for optimization of area traffic control and network flows has long been recognized as an important problem in the road traffic field. A number of solution methods to this combined problem have been discussed, and good results have been given in a medium sized trial road network. ALLSOP and CHARLESWORTH (1977) carried out a mutually consistent calculation for the area traffic control optimization problem and equilibrium network flows, in which the signal timings and link flows were treated alternately and obtained, respectively, by solving the area traffic control optimization problem for the current link flows and by solving user equilibrium traffic assignment for the resulting signal timings. These two subproblems can be solved alternately until an intuitively expected convergence is achieved. The resulting mutually consistent signal timings and network flows will, however, in general, be a nonoptimal solution, as has been illustrated by GERSHWIN and TAN (1979) and DICKSON (1981). In contrast, a bilevel formulation has been adopted as an appropriate tool in dealing with the combined problem. For example, HEYDECKER and KHOO (1990) formulated the combined problem as a bilevel program, in which an optimization of signal timings was regarded as the upper level program while the user equilibrium traffic assignment was regarded as the function of signal timings, and, therefore, it was dealt with as the lower level program. In the bilevel formulation to the optimization for area traffic control and equilibrium network flows, it is recognized that, although the link flows must be in equilibrium for the resulting signal timings, the signal timings will not, in general, be optimal for the resulting link flows if the latter are regarded as fixed. In a comparison with the results obtained from the mutually consistent calculation in a trial road network, it was reported that, as expected, the total system performance achieved by the bilevel formulation was an improvement on that given by mutually consistent calculation (Heydecker and Khoo, 1990).

Application of sensitivity analysis for equilibrium network flows was introduced by TOBIN and FRIESZ

Transportation Science
Vol. 33, No. 3, August 1999

0041-1655/99/3303-0279 $01.25
© 1999 Institute for Operations Research and the Management Sciences

Copyright © 1999. All rights reserved.

(1988). Extension of the application of sensitivity analysis of equilibrium network flows to a number of topics of interest within this aspect of transportation have been investigated in subsequent studies (YANG and YAGAR, 1995, YANG, 1997). The mathematical relationships between the network flows and signal timings are implicit and nonlinear, necessitating the bilevel formulation, which is nonconvex because the performance index is a nonconvex function of the signal timings and the network flows. As a practical way of solving this bilevel program, FRIESZ, et al. (1990) discussed a sensitivity analysis method, in which all the first partial derivatives of the objective function and network flows, with respect to the decision variables, are taken into account. By use of this method, the perturbations around a current solution are estimated from these derivatives and then gradient-based solution methods can be used to identify feasible directions at current solution in which a better local solution can be achieved, even though an explicit expression for the link flows, in terms of the decision variables, is not available. The sensitivity analysis method was used by Yang and Yagar (1995) in solving the bilevel traffic assignment and traffic control problem in saturated road networks with simple forms of signal control and a simple measure of performance.

In this paper, we formulate the combined problem for a very general form of coordinated fixed-time signal control with the performance index used in one of the world's most widely used practical procedures for optimization of signal timings for area traffic control. In the upper level problem, optimization for area traffic control is defined by the phase-based signal setting variables: the common cycle time, and the starts and durations of all green times. Also, the performance index is defined as the sum of a weighted linear combination of the estimated rate of delay and number of stops per unit time for all traffic streams, as evaluated by the traffic model in TRANSYT (VINCENT, MITCHELL and ROBERTSON, 1980). Approximate mathematical expressions for various components of the performance index and the average delay to a vehicle at the downstream junction in the TRANSYT model for both undersaturated and oversaturated links are considered. Undersaturated traffic streams with and without initial queues are included. As a step toward finding solutions to this bilevel problem, the partial derivatives, with respect to signal setting variables and link flows, are derived (CHIOU, 1997a) following previous work by WONG (1995). In the lower level problem, a general expression for a user equilibrium traffic assignment is formulated as a variational inequality problem, which has been presented by SMITH (1979) and also identified by DAFERMOS (1980).

Since the objective function in the bilevel problem is nonconvex, only locally optimal solutions can be achieved; in this paper, a locally optimal search is presented based on the gradient projection method (LUENBERGER 1989), which determines the descent direction at each iteration and along which the optimal step length can be determined by the bisection method. In this way, local solutions, a Karush–Kuhn–Tucker (KKT) point can be identified as each local solution. Furthermore, a global search heuristic of making equal and simultaneous changes in the starts of green for all signal groups at any one junction, corresponding to making changes in offsets at various junctions, changes on which there are no practical constraints, is used to carry the search widely across the feasible region. The mixed search procedure, which includes the locally optimal search in terms of gradient projection method, and global search heuristic, in terms of changes in offsets, is adopted here after carrying out a series of empirical studies (CHIOU, 1997b, 1998a,b). In the next section, the bilevel formulation for optimization of area traffic control and equilibrium network flows is given. In Section 2, solution method and mixed search procedure for the bilevel program at any given signal settings are discussed, including both the identification of descent direction and optimal choices of step length. In Section 3, Allsop and Charlesworth's example network (1977) is used for an illustrative numerical example of the proposed solution method, and comparison is made with mutually consistent calculations. Conclusions for this paper are given in Section 4.

1. BILEVEL FORMULATION

1.1 Notation

A directed network $\mathbf{G}(\mathbf{N}, \mathbf{L})$ includes N_{J} fixed-time signal controlled junctions, and N_{L} links.

ζ is the reciprocal of the common cycle time.
θ is the vector of starts of green for various links as proportions of cycle time.
ϕ is the vector of durations of green for various links as proportions of cycle time.

For any link a in $\mathbf{L}$,

D_a is the rate of delay on link a.
d_a is the average delay to a vehicle arriving on link a.
S_a is the number of stops per unit time on link a.

Regarding the user equilibrium traffic assignment problem:

Copyright © 1999. All rights reserved.

q is the vector of average link flows.
f is the vector of path flows between points of entry to and points of exit from the network.
$\bar{\mathbf{D}}$ is the matrix of numbers of vehicles traveling per unit time between each point of entry to the network and each point of exit from it.
δ, Δ are, respectively, the link-path and entry point/exit point-path incidence matrices.
c is the vector of link travel times, where the travel time on a link is defined as the sum of the undelayed travel time and average delay incurred by a vehicle at the downstream junction.

1.2 Problem Formulation

A bilevel problem for optimization of signal timings for equilibrium network flows can be described in the following way.

$$\underset{\psi}{\text{Minimize}}\ P = P_0(\psi, \mathbf{q}^*(\psi))$$

$$\text{subject to} \quad \psi \in S_0, \tag{1}$$

where S_0 is the feasible set of ψ. In Problem 1 the performance index P is expressed via function P_0 in terms of signal timings $\psi = (\zeta, \theta, \phi)$ and equilibrium flows $\mathbf{q}^*(\psi)$. The signal timings are determined in the upper level problem while the equilibrium network flows are determined in terms of ψ in the lower level problem in a way that defines the function $\mathbf{q}^*$.

By definitions $P_1(\psi) = P_0(\psi, \mathbf{q}^*(\psi))$, Problem 1 can be re-expressed as to

$$\underset{\psi}{\text{Minimize}}\ P = P_1(\psi)$$

$$\text{subject to} \quad \psi \in S_0. \tag{2}$$

Let W_{aD}, W_{aS} be, respectively, link-specific weighting factors for rate of delay and number of stops per unit time as estimated by the TRANSYT traffic model, and M_D, M_S be the corresponding monetary factors common to all links. The objective function in Problem 2 can be expressed as

$$\sum_{a \in L} D_a W_{aD} M_D + S_a W_{aS} M_S.$$

Mathematical expressions representing D_a and S_a have been reported in Chiou (1997a). The constraints defining feasible signal timings in Problem 2 are all linear in the components at ψ (ALLSOP, 1992).

Let **A** and **B** be, respectively, the coefficient matrix and constant vector for linear constraints in matrix form defining S_0. Problem 2 becomes

$$\underset{\psi}{\text{Minimize}}\ P = P_1(\psi)$$

$$\text{subject to} \quad \mathbf{A}\psi^{\mathrm{T}} \leqslant \mathbf{B}, \tag{3}$$

where superscript T denotes the matrix transposition. Determination of equilibrium network flows $\mathbf{q}^*(\psi)$ in Problem 1 can be formulated as a variational inequality problem as follows.

$$\text{To find values} \quad \mathbf{q}^*(\psi) \text{ of } \mathbf{q}(\psi)$$

$$\text{such that} \quad \mathbf{c}(\mathbf{q}^*(\psi))(\mathbf{q}(\psi) - \mathbf{q}^*(\psi))^{\mathrm{T}} \geqslant 0$$

$$\forall \mathbf{q}(\psi) \in \Omega = \{\mathbf{q}(\psi);\ \mathbf{q}(\psi)^{\mathrm{T}} = \delta \mathbf{f}(\psi)^{\mathrm{T}},\ \Delta \mathbf{f}(\psi)^{\mathrm{T}} = \hat{\mathbf{D}}^{\mathrm{T}},\ \mathbf{f}(\psi) \geqslant 0\}, \tag{4}$$

where the link travel time function $\mathbf{c}(\mathbf{q}^*(\psi))$ is evaluated on the basis of the average delay d_a, $\forall a \in \mathbf{L}$ corresponding to the rate of delay D_a, as reported in Chiou (1997a).

2. SOLUTION METHOD

BECAUSE THE FUNCTION P_0 is nonconvex in terms of signal timing variables, only locally optimal solutions to Problem 2 can be found by gradient-based methods. Following previous work by Wong (1995) and newly derived results by Chiou (1997a), the partial derivatives of P_0 have been obtained, and those of the equilibrium flows, with respect to the signal timings, follow from the sensitivity analysis of Tobin and Friesz (1988). In this paper, the gradient projection method (Luenberger 1989) is used to decide a descent direction from the current signal timings. The search process in this direction is usually confined to a short interval, and, therefore, a first-order linear approximation to the corresponding equilibrium flow is adopted. In this way, a good local solution can be identified within the neighborhood of current signal timings.

To extend the search to many parts of the feasible region, advantage is taken of the fact that movement from the current solution in directions corresponding to changes in offsets at the various junctions is unconstrained. The search intervals in these directions are correspondingly long, and optimal choices of step length require a global search, which is carried out by evaluating the objective function at equally spaced points within the search interval. For each such evaluation, the equilibrium link flows need to be recalculated by making a fresh assignment because of the large changes in signal timings. The mixed search procedure proposed by Chiou (1997b, 1998a,b) has been refined in this paper, in

Copyright © 1999. All rights reserved.

which the first partial derivatives reported by Wong (1995) and Chiou (1997a) for the function P_0 and for equilibrium link flows $\mathbf{q}^*$ are continuous; therefore, the following relevant theorems are given immediately without proofs.

THEOREM 1 (The gradients of $P_1(\psi)$ at signal timings ψ^0). *In Problem 2, for any given point ψ^0 that satisfies $\mathbf{A}(\psi^0)^{\mathrm{T}} \leqslant \mathbf{B}$ and $\mathbf{q}^0 = \mathbf{q}^*(\psi^0)$ the first partial derivatives of $P_1(\psi)$ with respect to $\psi = (\zeta, \theta, \phi)$ can be expressed as follows:*

$$\frac{\partial P_1(\psi^0)}{\partial \zeta} = \frac{\partial P_0(\psi^0, \mathbf{q}^0)}{\partial \zeta} + \frac{\partial P_0(\psi^0, \mathbf{q}^0)}{\partial \mathbf{q}^*} \frac{\partial \mathbf{q}^*(\psi^0)}{\partial \zeta}, \tag{5}$$

$$\frac{\partial P_1(\psi^0)}{\partial \theta} = \frac{\partial P_0(\psi^0, \mathbf{q}^0)}{\partial \theta} + \frac{\partial P_0(\psi^0, \mathbf{q}^0)}{\partial \mathbf{q}^*} \frac{\partial \mathbf{q}^*(\psi^0)}{\partial \theta}, \tag{6}$$

$$\frac{\partial P_1(\psi^0)}{\partial \phi} = \frac{\partial P_0(\psi^0, \mathbf{q}^0)}{\partial \phi} + \frac{\partial P_0(\psi^0, \mathbf{q}^0)}{\partial \mathbf{q}^*} \frac{\partial \mathbf{q}^*(\psi^0)}{\partial \phi}. \tag{7}$$

COROLLARY 2. (Linear approximation of $\mathbf{q}^*(\psi)$ at ψ^0). *Let ψ^0 and $\psi^0 + \alpha\mathbf{h}$ be two distinct points where α is a small multiplier along direction $\mathbf{h}$ such that the straight line segment joining ψ^0 and $\psi^0 + \alpha\mathbf{h}$ lies in the domain set, and let $\mathbf{q}^0 = \mathbf{q}^*(\psi^0)$. Then, the linear approximation of $\mathbf{q}^*$ can be expressed below as a Taylor polynomial of degree 1 at ψ^0*

$$\mathbf{q}^*(\psi^0 + \alpha\mathbf{h})^{\mathrm{T}} = (\mathbf{q}^0)^{\mathrm{T}} + \alpha \nabla \mathbf{q}^*(\psi^0)\mathbf{h}^{\mathrm{T}}.$$

COROLLARY 3. (Linear approximation of $P_1(\psi)$ at ψ^0). *Let ψ^0 and $\psi^0 + \alpha\mathbf{h}$ be two distinct points where α is a small multiplier along direction $\mathbf{h}$ such that the straight line segment joining ψ^0 and $\psi^0 + \alpha\mathbf{h}$ lies in the domain set. Then, the linear approximation for $P_1(\psi)$ can be expressed below as a Taylor polynomial of degree 1 at ψ^0*

$$P_1(\psi^0 + \alpha\mathbf{h}) = P_1(\psi^0) + \alpha \nabla P_1(\psi^0)\mathbf{h}^{\mathrm{T}}. \tag{8}$$

2.1 Gradient Projection Method

The following theorem is related to a gradient projection method (Luenberger 1989), and are adapted from BAZARAA, SHERALI, and SHETTY (1993, 448–454).

THEOREM 4. (Finding a descent direction for $P_1(\psi^0)$ in Eq. 8). *In Problem 3, for any feasible point ψ^0 that satisfies $\mathbf{A}(\psi^0)^{\mathrm{T}} \leqslant \mathbf{B}$ such that $\mathbf{A}_b(\psi^0)^{\mathrm{T}} = \mathbf{B}_b$ and $\mathbf{A}_{nb}(\psi^0)^{\mathrm{T}} < \mathbf{B}_{nb}$ where $\mathbf{A}^{\mathrm{T}} = [\mathbf{A}_b^{\mathrm{T}}, \mathbf{A}_{nb}^{\mathrm{T}}]$, $\mathbf{B}^{\mathrm{T}} = [\mathbf{B}_b^{\mathrm{T}}, \mathbf{B}_{nb}^{\mathrm{T}}]$, if $\mathbf{M} = \mathbf{A}_b$ has full rank, and if $\mathbf{g}$ is of the form $\mathbf{g} = \mathbf{I} - \mathbf{M}^{\mathrm{T}}(\mathbf{M}\mathbf{M}^{\mathrm{T}})^{-1}\mathbf{M}$ such that $\mathbf{g}\nabla_{\psi} P_1(\psi^0)^{\mathrm{T}} \neq \mathbf{0}$, then $\mathbf{h}$ is an improving feasible direction for $P_1(\psi^0)$ in Eq. 8 and is given by*

$$\mathbf{h}^{\mathrm{T}} = -\mathbf{g}\nabla P_1(\psi^0)^{\mathrm{T}}, \tag{9}$$

where $\mathbf{g}$ is the projection matrix for $\nabla P_1(\psi^0)$. If $\mathbf{g}\nabla P_1(\psi^0)^{\mathrm{T}} = \mathbf{0}$, a new projection matrix $\hat{\mathbf{g}}$ can be generated as

$$\hat{\mathbf{g}} = \mathbf{I} - \hat{\mathbf{M}}^{\mathrm{T}}(\hat{\mathbf{M}}\hat{\mathbf{M}}^{\mathrm{T}})^{-1}\hat{\mathbf{M}},$$

where $\hat{\mathbf{M}} = \bar{\mathbf{A}}_b$ and $\bar{\mathbf{A}}_b$ is obtained from $\mathbf{A}_b$ by deleting the rows of $\mathbf{A}_b$ corresponding to negative Lagrange multipliers at ψ^0.

COROLLARY 5. (Stopping condition). *Let ψ_{opt} locally solve the problem $P_1(\psi)$ in Eq. 3, and the constraints in Eq. 3 satisfy the Linear Independence Constraint Qualification* (see Bazaraa, Sherali and Shetty, 1993, 192–193) *then ψ_{opt} is a KKT point for Problem 3. Following the gradient projection method that is given in Theorem 4 in determining the feasible descent direction for Problem 3 at current point ψ_{opt}, this search procedure will terminate only at a KKT point, which is regarded as a plausible heuristic for solving this nonlinear and nonconvex problem.*

2.2 Determination of the Maximum Step Length

Given a feasible point ψ^0 and an improving feasible direction $\mathbf{h}$ for $P_1(\psi^0)$ in Eq. 8, a one-dimensional search problem can be expressed as

$$\begin{aligned} &\underset{\alpha \geqslant 0}{\text{Minimize}}\ Z(\alpha) = P_1(\psi^0 + \alpha\mathbf{h}) \\ &\text{subject to} \quad \mathbf{A}(\psi^0 + \alpha\mathbf{h})^{\mathrm{T}} \leqslant \mathbf{B}. \end{aligned} \tag{10}$$

Let $\mathbf{C} = \mathbf{B}_{nb} - \mathbf{A}_{nb}(\psi^0)^{\mathrm{T}}$, $\mathbf{D} = \mathbf{A}_{nb}\mathbf{h}^{\mathrm{T}}$ and the ith rows of $\mathbf{C}$, $\mathbf{D}$ be $C_i = B_i - \Sigma_j A_{ij}\psi_j^0$, $D_i = \Sigma_j A_{ij}h_j$, where B_i and A_{ij} are, respectively, the ith, ijth element of $\mathbf{B}_{nb}$ and $\mathbf{A}_{nb}$. Then Eq. 10 can be expressed as follows, and the good step length α can be solved by the bisection method,

$$\begin{aligned} &\underset{\alpha}{\text{Minimize}}\ Z(\alpha) = P_1(\psi^0 + \alpha\mathbf{h}) \\ &\text{subject to} \quad 0 \leqslant \alpha \leqslant \alpha_{max}, \end{aligned} \tag{11}$$

where

$$\alpha_{\max} = \begin{cases} \infty & \text{if} \quad \mathbf{D} \leqslant 0 \\ \text{Min}\{(C_i/D_i)\colon D_i > 0\} & \text{otherwise.} \end{cases}$$

2.3 Determination of a Good Step Length

In Problem 11, for each α_0 in $[0, \alpha_{\max}]$, let $\psi_0 = \psi^0 + \alpha_0\mathbf{h}$, the total derivative $Z'(\alpha_0)$ can be expressed as

$$Z'(\alpha_0) = \nabla P_1(\psi_0)\mathbf{h}^{\mathrm{T}}. \tag{12}$$

The bisection method for $Z(\alpha)$ along the feasible direction $\mathbf{h}$ at current signal settings ψ^0 starting

Copyright © 1999. All rights reserved.

with $Z'(0) < 0$ can be carried out in the following steps.

Step 0. Set the search interval $[0, \alpha_{\max}]$ as $[\alpha^l, \alpha^r]$ where $\alpha^l = 0$ and $\alpha^r = \alpha_{\max}$, set index $k = 0$, and set $N_{\max}$ as the maximum iterations for performing the bisection method, and let $\alpha^{(k)} = (\alpha^l + \alpha^r)/2$.

Step 1. If $k = N_{\max}$, then take $\alpha^{(k)}$ as the good step length, i.e., $\alpha_{\text{good}} = \alpha^{(k)}$ along the feasible direction $\mathbf{h}$ at current signal settings ψ^0, the new signal settings is $\psi_{\text{new}} = \psi^0 + \alpha_{\text{good}}\mathbf{h}$ and the corresponding equilibrium flows are $\mathbf{q}_{\text{new}} = \mathbf{q}^*(\psi_{\text{new}})$ via the reassignment process.

Otherwise, perform evaluations of $Z(\alpha^{(k)})$ and of $Z'(\alpha^{(k)})$ by Eq. 11, where $\alpha^{(k)} = (\alpha^l + \alpha^r)/2$. Various possibilities then need to be considered.

1. If $Z'(\alpha^{(k)}) > 0$, it is not necessary to search for the good step length over the right half interval, and a reduced search interval can be decided by letting $\alpha^r = \alpha^{(k)}$; let $k = k + 1$ and return to Step 1.
2. If $Z'(\alpha^{(k)}) < 0$, we evaluate the values of $Z(\alpha)$ when $\alpha = \alpha^l$ and $\alpha = \alpha^{(k)}$, where the signal settings is $\psi = \psi^0 + \alpha\mathbf{h}$ and the corresponding equilibrium flows are in the linear approximation form as given by Corollary 2, and two cases are discussed.
   (i). If $Z(\alpha^{(k)}) \geqslant Z(\alpha^l)$, then, since, for all sufficiently small $\varepsilon > 0$ in the neighborhood of α^l, we have $Z(\varepsilon) < Z(\alpha^l)$, it follows that a reduced search interval can be decided by letting $\alpha^r = \alpha^{(k)}$; let $k = k + 1$ and return to Step 1.
   (ii). If $Z(\alpha^{(k)}) < Z(\alpha^l)$, then, in the right half interval, we may find a good step length that can keep decreasing the value of Z and a new search interval $[\alpha^{(k)}, \alpha^r]$ is decided by letting $\alpha^l = \alpha^{(k)}$; let $k = k + 1$ and return to Step 1.
3. If $Z'(\alpha^{(k)}) = 0$ and $Z(\alpha^{(k)}) \geqslant Z(\alpha^l)$, we follow sub-step 2.i; if $Z(\alpha^{(k)}) < Z(\alpha^l)$, we choose the left half interval as our new search interval because we are looking for the local optimum in the neighborhood of current signal settings; let $k = k + 1$ and return to Step 1.

2.4 Optimal Choice of Step Length for Changes of Offsets

Extensions of locally optimal search via gradient projection method to many parts of the feasible region can be carried out as a global search heuristic as follows. Let ζ^0, θ^0, ϕ^0 be, respectively, current signal timings for common cycle time, starts, and durations of greens; let $\psi^0 = [\zeta^0, \theta^0 + \Theta, \phi^0]$, where changes at various junctions in a vector form is $\Theta = [\Theta_1, \ldots, \Theta_m, \ldots, \Theta_{N_J}]$, and, for each element in Θ_m, is the vector $\theta_m[1, 1, \ldots, 1]$ of the order of the number of signal groups at junction m. Then, each element θ_{jm} for particular signal group j at junction m can be expressed as

$$\theta_{jm} = \theta_{jm}^0 + \theta_m, \quad 1 \leqslant m \leqslant N_J.$$

The optimal choice of changes in offsets in the steepest descent direction $\mathbf{h} = [0, \Theta, \mathbf{0}]$ is to

$$\underset{0 \leqslant \alpha \leqslant \alpha_{\max}}{\text{Minimize}}\ Z_1(\alpha) = P_1(\psi^0 + \alpha\mathbf{h})$$

$$\text{where} \quad \alpha_{\max} = \underset{1 \leqslant m \leqslant N_J}{\text{Max}} \{|\theta_m|^{-1}\}, \tag{13}$$

and $|\theta_m|^{-1}$ is the whole range of global search for offsets at junction m. In Problem 13, a global search for optimal choice of changes in offsets is carried out by evaluating the objective function Z_1 at equally spaced points within search interval until the least is found and can be described as the following steps.

Step 0. Set the maximal number of iterations as $N_{\max}$.

Step 1. Decide the evaluated interval γ, $\gamma = (\alpha_{\max}/N_{\max})$ and $\alpha_{\max} = \gamma N_{\max}$.

Step 2. Evaluate the objective function $Z_1(\alpha)$ at current signal timings ψ^0 when $\alpha = k\gamma$, where $k = 1, 2, \ldots, N_{\max}$, and choose the minimum of $Z_1(\alpha)$ as the $Z_1(\alpha_{\text{opt}})$, where α_{opt} is the approximately optimal step length for $Z_1(\alpha)$ along the direction $\mathbf{h}$ at ψ^0.

Step 3. A new point can be decided below as the feasible point for the next iteration. $\psi_{\text{new}} = \psi^0 + \alpha_{\text{opt}}\mathbf{h}$ and the corresponding equilibrium flow, $\mathbf{q}_{\text{new}} = q^*(\psi_{\text{new}})$, which can be decided by reassignment process.

2.5 Mixed Search Procedure

Following the mixed search procedure in earlier work (Chiou, 1997b; 1998a,b), in this paper a practical way to search for a better solution at current signal timings using three kinds of types is as follows.

Type I. Apply the locally optimal search for signal setting variables with respect to ζ, θ, ϕ in the neighborhood of current signal timings, such that the value of the performance index is the minimal one and the number of the search dimension is the total number of signal-setting variables. The descent direction at each iteration is determined by means of the gradient projection method, along which a good step length is decided by the bisection method.

Copyright © 1999. All rights reserved.

Type II. Apply the locally optimal search similarly as in Type I, but, for signal settings variables with respect to θ, ϕ for given common cycle time, by specifying ζ as ζ^0, so that the number of search dimensions is the number of signal-setting variables θ and ϕ. The descent direction at each iteration is determined by means of the gradient projection method, along which a good step length is decided by the bisection method.

Type III. Apply the global search heuristic for signal setting variables with respect to suitable changes in θ only for given common cycle time and the durations of green times for all signal groups by specifying ζ, ϕ as ζ^0, ϕ^0, respectively, and changing the starts of greens for all signal groups at each junction by the same amount specific to that junction, so that the number of search dimensions is reduced to the number of signal-controlled junctions, but the step length in the chosen direction is unconstrained. The search direction for the unconstrained optimization Problem 13 at each iteration is determined by the steepest descent direction, along which a good step length is decided by the global search throughout the whole length lying within the feasible region.

The purpose of the mixed search procedure for the bilevel problem given in Eq. 3 is to find a good local optimum by adopting the TRANSYT-like hill-climbing technique. A step of Type I allows each element of the signal-setting variables to vary within the feasible constraint set; therefore, a good local solution can be obtained. A step of Type II specifies the common cycle time as a fixed value while still allowing other signal-setting variables to vary within the restricted feasible region by which less computation efforts will be incurred but a good local solution for that cycle time can be located. Furthermore, a step of Type III specifies the common cycle time and the durations of green times for all signal groups at each junction as fixed values, and makes equal and simultaneous changes in the starts of green times for all signal groups at any one junction, which corresponds to making changes in the offset variables; therefore, the search in this restricted set of directions will be extended over the whole feasible region. In summary, the rationale for this mixed search procedure is to apply the three types of search steps, in turn, to find the near-optimal in many parts of the feasible region and locate them within the corresponding neighborhoods.

Two variants of the mixed search procedure, mix a and mix b, obtained by applying various sequences of the three types of step are stated as follows. The mix a starts from any initial solution that is not a local optimum and follows the sequence of search steps of type I–II, III, and I–II, and so on until a good local optimum for the bilevel formulation is found and located. The mix b starts with the search step of type III first and, therefore, by searching many parts of the feasible region so that a good initial solution can be found and used as a new starting point for the subsequent search process, carried out by making steps of type I and II of the mixed search procedure; the sequence of search steps is type III, I–II, and type III, I–II, and so on. A stopping criterion for performing the mixed search will be met if a predetermined threshold is satisfied between successive iterations.

3. NUMERICAL EXAMPLE

FOR ILLUSTRATION, the network of Allsop and Charlesworth (1977) is used. The layout of the network for use in signal settings and allocations for signal groups at each junction are given in Figs. 1 and 2. Travel demands for each pair of origin and destination are those used by Allsop and Charlesworth (1977), also given in Table I. This numerical test includes 22 origin–destination pairs, 23 links, 40 feasible routes, and 18 signal groups at 6 signal-controlled junctions. Because of the nonconvexity of the bilevel Problem 3, only a local optimum is expected to be obtained in this numerical test. For reasons of understanding the effectiveness and robustness of the proposed solution method in this test, four distinct sets of initial signal timings are specified. The first and third sets of initial signal timings are specified with the start times for signal groups equally distributed over the cycle at each junction, whereas the second and fourth sets of initial signal timings are specified to favor the roads via junctions 5 and 6 for north–south and south–north traffic flows.

3.1 Computation Results

Computation results for Allsop and Charlesworth's road network are summarized in Table II for the two variants of the mixed search procedure, i.e., mix a and mix b, and the mutually consistent calculations, starting from the four sets of initial signal settings. First, the improvement rates for mix a are 88.26, 96.27, 91.43, and 94.01%, and the final values of performance index are 34.20, 34.27, 42.72, and 39.60 veh·h/h with an average of 37.70 veh·h/h and standard deviation of 3.63 veh·h/h. Similarly, the improvement rates for mix b are 88.42, 96.33, 92.14, and 94.42%, and the final values are 33.74, 33.78, 39.19, and 36.88 veh·h/h with an average of 35.90 veh·h/h and standard deviation of 2.29 veh·h/h. The

Copyright © 1999. All rights reserved.

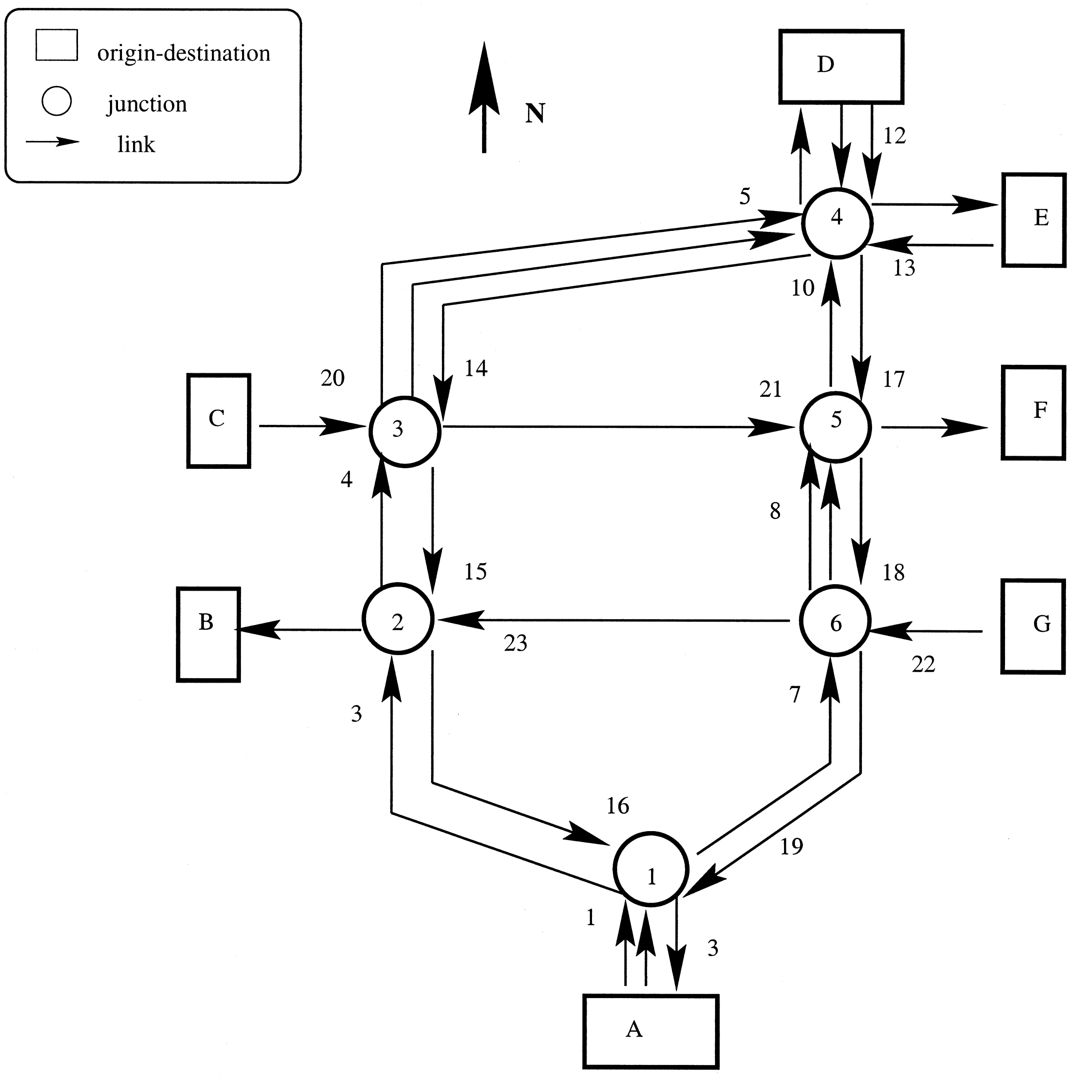

Fig. 1. Layout for Allsop and Charlesworth's example network.

differences for the two variants, mix a and mix b, on the four sets of signal timings are 1.35, 1.43, 8.26, and 6.87%. Furthermore, improvement rates for the mutually consistent calculations are 57.06, 49.11, 82.13, and 84.09%, which are substantially worse than those for the mixed search procedure; and the corresponding final values are 125.07, 467.96, 89.13, and 105.17 veh·h/h with an average of 196.83 veh·h/h and standard deviation of 157.05 veh·h/h. Because the mutually consistent calculations are nonoptimal, the values of the performance index fluctuate between iterations, as can be seen in Table II.

Second, for the two variants of mixed search procedure, mix a and mix b, at the first set of signal timings, for example, mix a starts locally optimal search by using type I and proceeds throughout iterations 1–4 until the contribution to system performance with respect to common cycle time becomes negative, then continues type II in iterations 5–8 until the KKT points are identified, and, at iteration 9, applies the global search heuristic via type III to other parts of the feasible region, then uses the locally optimal search type II again to improve current system performance. Similarly, mix b starts the global search heuristic via type III firstly, but takes fewer iterations than does mix a by carrying out the locally optimal search via type I, and again applies the global search heuristic to other parts of the feasible region, then uses the locally optimal search to identify better KKT points, which will continue until the system improvement rate is negligible. As we mentioned, the differences in the final values for the system performance index between the two variants of mixed search procedure are very small; however, it is worth noting that applying the global

Copyright © 1999. All rights reserved.

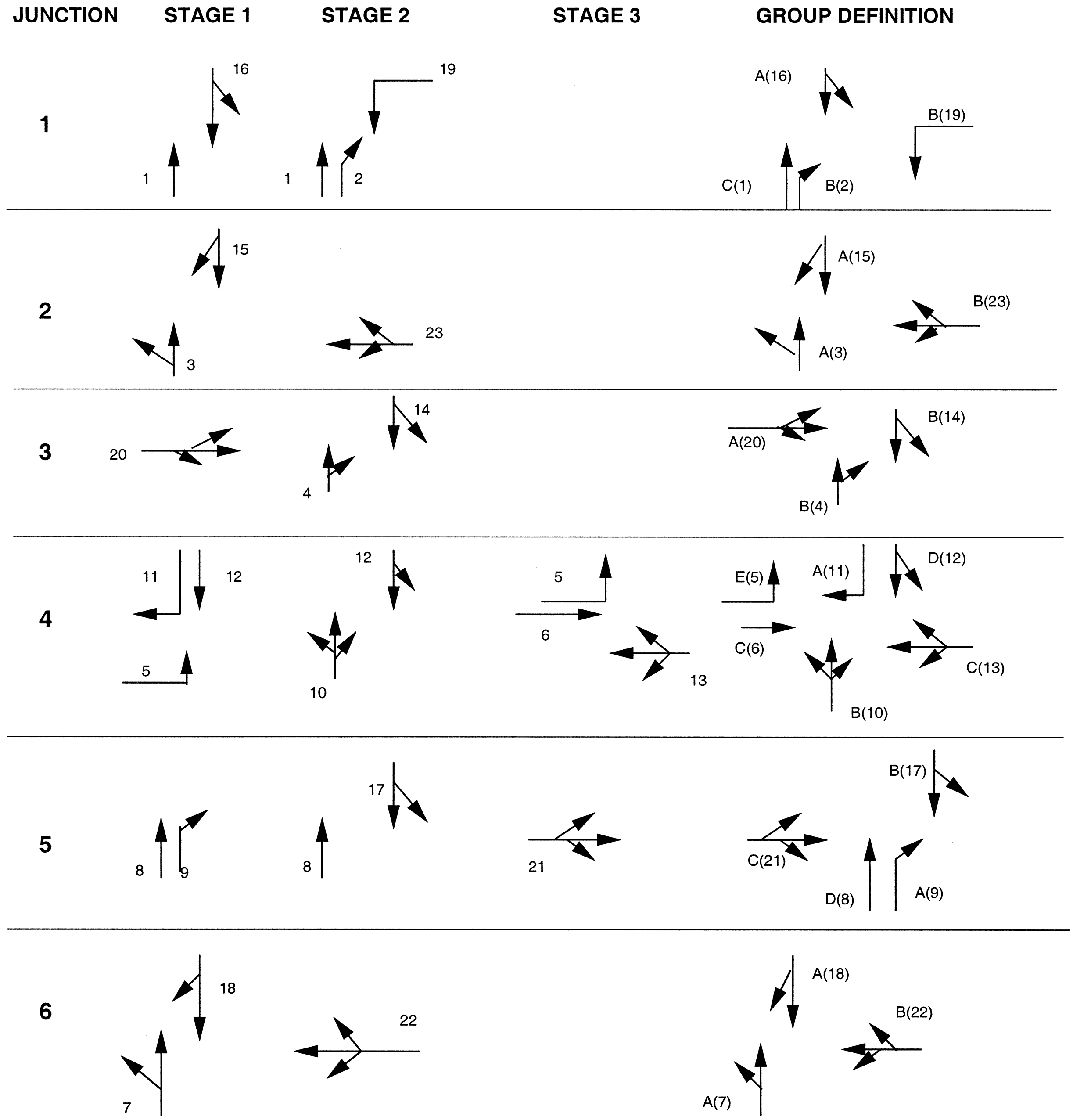

Fig. 2. Signal groups allocation for Allsop and Charlesworth's example network.

TABLE I

Travel Demands for Allsop and Charlesworth's Network[1]

Origin/Destination	A	B	D	E	F	Origin Totals
A	—	250	700	30	200	1180
C	40	20	200	130	900	1290
D	400	250	—	50	100	800
E	300	130	30	—	20	480
G	550	450	170	60	20	1250
Destination Totals	1290	1100	1100	270	1240	5000

[1]Each pair trip is measured in vehicles per hour.

Copyright © 1999. All rights reserved.

TABLE II

Comparison of Mixed Search Procedure and Mutually Consistent Calculations

Iteration	Mix a		Mix b		Mutually Consistent Calculations	
	PI[1]	Improving Rate[2]	PI[1]	Improving Rate[2]	PI[1]	Improving Rate[2]
1st Signal Settings						
1	291.25	64.45	291.25	0.74	291.25	70.46
2	103.53	32.73	289.10	86.83	86.04	1.86
3	69.64	4.24	38.07	2.21	84.44	15.67
4	66.69	8.25	37.23	4.83	71.21	−24.88
5	61.19	13.12	35.43	0.37	88.93	−1.34
6	53.16	7.19	35.30	4.42	90.12	−2.44
7	49.34	4.58	33.74		92.32	−14.84
8	47.08	3.78			106.02	−18.10
9	45.30	6.05			125.21	−0.02
10	42.56	7.73			125.23	0.13
11	39.27	3.21			125.07	
12	38.01	5.10				
13	36.07	5.18				
14	34.20					
2nd Signal Settings						
1	919.56	23.64	919.56	3.14	919.56	49.89
2	702.16	43.08	890.70	80.78	460.81	−17.23
3	399.64	25.32	171.19	73.38	540.23	20.22
4	298.47	53.25	45.57	22.14	431.02	−11.80
5	139.53	19.34	35.48	4.79	481.88	2.89
6	112.54	37.51	33.78		467.96	
7	70.33	10.55				
8	62.91	31.74				
9	42.94	12.58				
10	37.54	7.38				
11	34.77	1.44				
12	34.27					
3rd Signal Settings						
1	498.76	65.32	498.76	1.91	498.76	80.00
2	172.97	61.69	489.23	88.63	99.78	11.57
3	66.26	6.43	55.63	27.79	88.24	0.32
4	62.00	11.06	40.17	2.44	87.96	−1.24
5	55.14	10.26	39.19		89.05	−0.09
6	49.48	6.81			89.13	
7	46.11	7.03				
8	42.87	0.35				
9	42.72					
4th Signal Settings						
1	661.04	68.53	661.04	3.14	661.04	67.31
2	208.03	52.19	640.26	89.10	216.12	37.98
3	99.46	24.25	69.80	46.17	134.04	14.26
4	75.34	25.82	37.57	1.84	114.93	9.18
5	55.89	28.16	36.88		104.38	−16.28
6	40.15	1.37			121.37	0.33
7	39.60				120.97	13.06
8					105.17	

[1]PI = performance index (in vehicles).
[2]Improving rate is shown in percentage.

search heuristic first then carrying out locally optimal search like mix b takes a fewer number of iterations in this numerical test network as compared to those obtained by mix a.

Third, additional mutually consistent calculations are carried out by starting with the equilibrium flows given by mix b, and the corresponding results for the four sets of signal timings are given in Table III. The values of performance index still fluctuate from iteration to iteration and the final values converge to a much higher one than those given by the mix b, which again illustrates the nonoptimal characteristics of the solutions obtained by the mutually consistent calculations.

The computation time for the mixed search procedure implemented on PC 486SX 25/33 Zenith machine, was that each iteration for this numerical example was performed in less than 30 seconds CPU

Copyright © 1999. All rights reserved.

TABLE III
Rerun Mutually Consistent Calculations Starting with Flows Given by Mix b

Iteration	1st Signal Settings		2nd Signal Settings		3rd Signal Settings		4th Signal Settings	
	PI[1]	IR[2]	PI[1]	IR[2]	PI[1]	IR[2]	PI[1]	IR[2]
1	33.74	−152.43	33.78	−271.55	39.19	−139.30	36.88	−173.24
2	85.17	−5.58	125.51	3.72	93.78	−1.10	100.77	−0.10
3	89.92	−2.62	120.84	0.06	94.81	1.28	100.87	1.14
4	92.28	−0.09	120.77	0.46	93.60	−0.53	99.72	−1.97
5	92.36	1.50	120.21	0.60	94.10	−0.50	101.68	1.55
6	90.97	1.67	119.49		94.57	0.04	100.10	0.52
7	89.45	1.29			94.53		99.58	
8	88.30	1.64						
9	86.85	3.42						
10	83.88							

[1]PI = performance index (in vehicles).
[2]IR = improving rate (in percentage).

time in the C++ integrated development environment. Total computation time for complete run of the mixed search procedure did not exceed 10 minutes of CPU time on that machine.

4. CONCLUSIONS AND DISCUSSIONS

IN THIS PAPER, a bilevel program for optimization of area traffic control and network flows has been formulated and solved. A practical way to search for a better solution to this nonconvex problem, a mixed search procedure, has been modified in this paper. Allsop and Charlesworth's (1977) example network has been used as an illustrative example for showing the effectiveness of the mixed search procedure, in which four distinct sets of signal timings were specified as the initial values for the numerical test. The robustness of the mixed search procedure in solving the nonconvex bilevel program has been shown by the fact that the resulting values were within negligible difference. Furthermore, in this paper, two variants of the mixed search procedure, mix a and mix b, have been proposed and conducted on a series of empirical studies. Applying the global search heuristic via type III first, i.e., mix b, to search for a better locally optimum, can effectively reduce the number of iterations that need to be taken afterward.

Consider further extensions of this mixed search procedure solving the area traffic control optimization problem; investigations on current issues of dynamic traffic assignment models need to be undertaken and also the corresponding mathematical expressions in the objective function need to be derived. Further work will continue on this area traffic control optimization problem.

ACKNOWLEDGEMENTS

THE AUTHOR WOULD LIKE to express her gratitude to Professor Allsop and Dr Heydecker for their invaluable discussions on this work. Thanks also go to three anonymous referees for their helpful comments on the earlier version of this paper. The work reported here was sponsored by the studentship of the Taiwan government and formed part of a Ph.D. thesis submitted to the University of London for the award of the Ph.D. degree.

REFERENCES

R. E. ALLSOP AND J. A. CHARLESWORTH, "Traffic in a Signal-Controlled Road Network: An Example of Different Signal Timings Inducing Different Routeings," *Traffic Eng. Control* **18,** 262–264 (1977).

R. E. ALLSOP, "Evolving Application of Mathematical Optimisation in Design and Operation of Individual Signal-Controlled Road Junctions," in *Mathematics in Transport Planning and Control*, J. D. Griffths (ed.), Oxford University Press. Oxford, U.K. 1–25, 1992.

M. S. BAZARAA, H. D. SHERALI, AND C. M. SHETTY, *Nonlinear Programming Theory and Algorithms.* John Wiley and Sons, New York, 1993.

S.-W. CHIOU, "Derivatives in an Area Traffic Control Problem subject to User Equilibrium Traffic Assignment," *Universities Transport Study Group 29th Annual Conference*, Bournemouth University, Poole, U.K. January 1997. unpublished (1997a).

S.-W. CHIOU, "Optimisation of Area Traffic Control subject to User Equilibrium Traffic Assignment," *25th European Transport Forum, Proc. Seminar F, Transportation Planning Methods.* Vol. II. PTRC, London, 53–64, 1997b.

S.-W. CHIOU, "Optimisation of Traffic Signal Timings for User Equilibrium Flow," in *Proceedings of Third TRB International Symposium on Highway Capacity*, Vol. I, R. Rysgaard (ed.), Danish Road Directorate, Copenhagen. 321–326, 1998a.

Copyright © 1999. All rights reserved.

S.-W. CHIOU, "Bi-level Formulation for Equilibrium Traffic Flow and Signal Settings," in *Mathematics in Transportation Planning and Control.* J. D. Griffths (ed.), Elsevier Science, Amsterdam, 59–68, 1998b.

S. C. DAFERMOS, "Traffic Equilibrium and Variational Inequalities," *Transp. Sci.* **14,** 42–54 (1980).

T. J. DICKSON, "A Note on Traffic Assignment and Signal Timings in a Signal-Controlled Road Network," *Transp. Res.* **15B,** 267–271 (1981).

T. L. FRIESZ, R. L. TOBIN, H.-J. CHO, AND N. J. MEHTA, "Sensitivity Analysis Based Heuristic Algorithms for Mathematical Programs with Variational Inequality Constraints," *Math. Program.* **48,** 265–284 (1990).

S. B. GERSHWIN, AND H. N. TAN, "Hybrid Optimisation: Optimal Static Traffic Control Constrained by Drivers' Route Choice Behaviour," Massachusetts Institute of Technology, laboratory for Information and Decision System Report LIDS-p-870, 1979.

B. G. HEYDECKER AND T. K. KHOO, "The Equilibrium Network Design Problem," *Proc. AIRO'90 Conference on Models and Methods for Decision Support,* Sorrento, Italy, 587–602, 1990.

D. G. LUENBERGER, *Linear and Nonlinear Programming.* Addison-Wesley, Massachusetts, 330–337, 1989.

M. J. SMITH, "The Existence, Uniqueness and Stability of Traffic Equilibria," *Transp. Res.,* **13B,** 295–304 (1979).

R. L. TOBIN AND T. L. FRIESZ, "Sensitivity Analysis for Equilibrium Network Flow," *Transp. Sci.* **22,** 242–250 (1988).

R. A. VINCENT, A. I. MITCHELL, AND D. I. ROBERTSON, *User Guide to TRANSYT Version 8, TRRL Report,* LR888, Transport and Road Research Laboratory, Crowthorne, U.K., 1980.

S. C. WONG, "Derivatives of the Performance Index for the Traffic Model from TRANSYT," *Transp. Res.* **29B,** 303–327 (1995).

H. YANG AND S. YAGAR, "Traffic Assignment and Signal Control in Saturated Road Networks," *Transp. Res.* **29A,** 125–139 (1995).

H. YANG, "Sensitivity Analysis for the Elastic-Demand Network Equilibrium Problem with Applications," *Transp. Res.* **31B,** 55–70 (1997).

(Received: July 1997; revisions received: August 1998; accepted: December 1998)

Copyright © 1999. All rights reserved.