DYNAMIC SYSTEM OPTIMAL TRAFFIC ASSIGNMENT ON CONGESTED MULTIDESTINATION NETWORKS

BYUNG-WOOK WIE
School of Travel Industry Management
University of Hawaii
2560 Campus Road
Honolulu, HI 96822
U.S.A.

1. INTRODUCTION

The dynamic system optimal traffic assignment model is formulated as a continuous time optimal control problem for a network with multiple origin-destination pairs. This paper resolves three important theoretical questions inherited from the previous models. First, more precise economic interpretation of the optimality conditions is provided. Second, it is established that the optimality conditions are both necessary and sufficient. Third, singular controls are tested for optimality using the generalized Legendre-Clebsch condition. Under the steady-state assumptions, the model is also proven to be a proper, dynamic extension of a static system optimal traffic assignment model.

2. DYNAMICS AND CONSTRAINTS

Let us consider a transportation network represented by a directed graph $G(N,A)$, where N is the set of nodes and A is the set of arcs. The cardinalities of N and A are assumed to be v and r, respectively. We also assume that N includes the three subsets of origin, destination and intersection nodes, which are not mutually exclusive because a node can be an origin, destination, and intersection at once. In this paper, the index "k" will denote an origin node or an intersection node, and the index "n" will denote a destination node. The cardinality of the set of destination nodes is assumed to be q. In addition, the set of paths connecting origin "k" and destination "n" will be denoted by P_{kn}.

The transportation network will be regarded as a controllable dynamic system that evolves over time and space. We will consider a fixed time period $[0,T]$. Let $x_a(t)$ denote the traffic volume on arc a at time t, that is, the number of vehicles traveling on arc a at time t. We assume that vehicles are uniformly distributed over the length of each arc at every instant in time. At any time t, the network is in some state which is described by r real numbers, $x_1(t)$, $x_2(t)$, ..., $x_r(t)$. Since the network has multiple destinations, we need to identify vehicles traveling on each arc by destination. Let $x_a^n(t)$ denote the traffic volume on arc a with destination n at time t. It follows that

$$(1) \qquad x_a(t) = \sum_{n\in N} x_a^n(t) \qquad \forall\ a\in A \quad \forall\ t\in[0,T].$$

In order to depict a physical phenomenon of congestion on each arc, the exit functions $g_a[x_a(t)]$ are assumed to be nonnegative, increasing, and continuously differentiable for all $x_a(t)\geq 0$ and $t\in[0,T]$

with additional restriction that $g_a(0) = 0$ for all $a \epsilon A$. The exit functions $g_a[x_a(t)]$ are further assumed to be linear in $x_a(t)$ in order to keep the property of separability for a multidestination network analysis as follows:

$$(2) \qquad g_a[x_a(t)] = \xi_a x_a(t) \qquad \forall\ a \epsilon A \quad \forall\ t \epsilon [0,T]$$

where ξ_a is a positive, time-invariant constant. Note that the exit functions were assumed to be strictly concave in a single destination case (see e.g. Wie, 1988b). We need to identify vehicles exiting from each arc at each instant by destination. Let $g_a^n[x_a^n(t),x_a(t)]$ denote the number of vehicles exiting from arc a with destination, called the exit flow rate. It follows

$$(3) \qquad g_a[x_a(t)] = \sum_{n \epsilon N} g_a^n[x_a^n(t),x_a(t)] \qquad \forall\ a \epsilon A \quad \forall\ t \epsilon [0,T].$$

In order to consider a first-in-first-out queue discipline, we assume that vehicles with different destinations are completely mixed on each arc at each instant in time and further assume that

$$(4) \qquad g_a^n[x_a^n(t),x_a(t)] = [x_a^n(t)/x_a(t)]\ g_a[x_a(t)]$$
$$= \xi_a x_a^n(t) \qquad \forall\ a \epsilon A \quad \forall\ n \epsilon N \quad \forall\ t \epsilon [0,T].$$

Equation (4) tells us that the exit functions $g_a^n[x_a^n(t),x_a(t)]$ are separable, nonnegative, increasing, and affine.

The dynamic evolution of the state of each arc is described by the first-order linear differential equations:

$$(5) \qquad dx_a(t)/dt \equiv \dot{x}_a(t) = u_a(t) - g_a[x_a(t)] \qquad \forall\ a \epsilon A \quad \forall\ t \epsilon [0,T]$$

where $u_a(t)$ denotes the inflow rate on arc a at time t, that is, the number of vehicles entering arc a at time t. It follows that

$$(6) \qquad u_a(t) = \sum_{n \epsilon N} u_a^n(t) \qquad \forall\ a \epsilon A \quad \forall\ t \epsilon [0,T]$$

where $u_a^n(t)$ is the inflow rate on arc a at time t with destination n. The dynamic evolution of the state of each arc for each destination can also be described by the first-order linear differential equations:

$$(7) \qquad dx_a^n(t)/dt \equiv \dot{x}_a^n(t) = u_a^n(t) - g_a^n[x_a^n(t),x_a(t)]$$
$$= u_a^n(t) - \xi_a x_a^n(t)$$
$$\forall\ a \epsilon A \quad \forall\ n \epsilon N \quad \forall\ t \epsilon [0,T].$$

Throughout the paper, $x_a^n(t)$ will be called the state variable and $u_a^n(t)$ the control variable. Equation (7), called the state equation, tells us that the instantaneous rate of change of the state variable with respect to time is a function of the exit flow rate and the inflow rate on each associated arc. Equation (7) can be interpreted as an infinity of scalar equality constraints indexed by time t and destination n. In addition, we assume that the number of vehicle traveling on arc a with destination n is a known nonnegative constant at the initial time t=0:

$$(8) \qquad x_a^n(0) = x_a^{n,o} \geq 0 \qquad \forall\ a\epsilon A \quad \forall\ n\epsilon N.$$

From equations (7) and (8), the value of the state variable can be determined as follows:

$$(9) \qquad x_a^n(t) = x_a^{n,o} + \int_o^t u_a^n(\tau)\ d\tau - \int_o^t \xi_a x_a^n(\tau)\ d\tau$$

$$\forall\ a\epsilon A \quad \forall\ n\epsilon N \quad \forall\ t\epsilon[0,T].$$

The flow conservation constraints are stated as follows:

$$(10) \qquad h_k^n[x(t),u(t)] \equiv S_{kn}(t) + \sum_{a\epsilon B(k)} \xi_a x_a^n(t) - \sum_{a\epsilon A(k)} u_a^n(t) = 0$$

$$\forall\ k\epsilon N \quad \forall\ n\epsilon N \quad \forall\ t\epsilon[0,T],\ k\neq n$$

where $S_{kn}(t)$ is the rate of flow generated at node k and destined to node n at time t; A(k) is the set of arcs whose tail node is k; and B(k) is the set of arcs whose head node is k. We assume that $S_{kn}(t)$ is a known nonnegative and continuous function of time. The scalar function $h_k^n[x(t),u(t)]$ can be regarded as the instantaneous mixed state-control equality constraint.

We ensure that the control variables are constrained as

$$(11) \qquad 0 \leq u_a^n(t) \leq S_{kn}(t) + \sum_{a\epsilon B(k)} \xi_a x_a^n(t)$$

$$\forall\ a\epsilon A \quad \forall\ n\epsilon N \quad \forall\ t\epsilon[0,T],\ k\neq n.$$

Equation (11) tells us that the value of the control variable $u_a^n(t)$ cannot exceed the upper bound that changes over time as a function of $S_{kn}(t)$ and $\sum_{a\epsilon B(k)}\xi_a x_a(t)$. Note that the previous models did not consider the upper bounds of the control variables. The control variables are assumed to be "piecewise continuous" for the period [0,T]. In this paper, the term "piecewise continuous" is used in the sense that a function is continuous except at a finite number of points and the one-sided limits always exist at the points of discontinuity. We also ensure that the state variables are nonnegative. However, we do not subsequently consider nonnegativity of the state variables in an explicit manner, because the assumption that $g_a^n[0,x_a(t)] = 0$ ensures that the state variables are always nonnegative.

We shall define the state and control vector functions to effect some economy of notation as follows:

$$(12) \qquad x(t) = (\ldots\ldots, x_a^1(t), \ldots, x_a^q(t), \ldots\ldots) \qquad \forall\ t\epsilon[0,T],$$

$$(13) \qquad u(t) = (\ldots\ldots, u_a^1(t), \ldots, u_a^q(t), \ldots\ldots) \qquad \forall\ t \in [0,T].$$

3. MODEL FORMULATION

The dynamic system optimal traffic assignment model is formulated as a continuous time optimal control problem:

$$(14) \qquad \text{Minimize } J = \sum_{a \in A} \int_0^T C_a[x_a(t)]\, dt$$

subject to

$$\dot{x}_a^n(t) = u_a^n(t) - \xi_a x_a^n(t) \qquad \forall\ a \in A \quad \forall\ n \in N \quad \forall\ t \in [0,T]$$

$$S_{kn}(t) + \sum_{a \in B(k)} \xi_a x_a^n(t) - \sum_{a \in A(k)} u_a^n(t) = 0$$

$$\forall\ k \in N \quad \forall\ n \in N \quad \forall\ t \in [0,T],\ k \neq n$$

$$x_a^n(0) = x_a^{n,o} \geq 0 \qquad \forall\ a \in A \quad \forall\ n \in N$$

$$0 \leq u_a^n(t) \leq S_{kn}(t) + \sum_{a \in B(k)} \xi_a x_a^n(t)$$

$$\forall\ a \in A \quad \forall\ n \in N \quad \forall\ t \in [0,T],\ k \neq n.$$

Let $C_a[x_a(t)]$ denote the instantaneous transportation cost rate on arc a where $x_a(t)$ is the number of vehicles traveling on arc a at time t. We assume that the functions $C_a[x_a(t)]$ are nonnegative, increasing, differentiable, and strictly convex for all $x_a(t) \geq 0$ and $t \in [0,T]$. Thus, the term $\sum_{a \in A} C_a[x_a(t)]$ is considered to be the instantaneous total transportation cost rate (e.g. measured in dollars per minute). Therefore, the performance index J has an economic interpretation as the total transportation cost spent during the fixed time interval [0,T] in the network (e.g. measured in dollars). It should be noted that J is additive and separable.

The optimal control problem (14) is to find optimal trajectories of the state and control variables that minimize J and satisfy all the constraints during the fixed time interval [0,T]. In other words, an optimal solution of the dynamic system optimal traffic assignment model can be referred to as a time-varying traffic flow pattern that minimizes the total transportation network cost. For example, each optimal control trajectory will represent the temporal distribution over the fixed time interval [0,T] of the number of vehicles entering arc a with destination n at each instant.

4. NECESSARY CONDITIONS

The necessary conditions for an optimal solution of the control problem (14) can be derived by the Pontryagin minimum principle (Pontryagin et al., 1962). We first construct the Hamiltonian:

(15) $$H[x(t),u(t),\lambda(t)] = \sum_{a\in A} C_a[x_a(t)] + \sum_{a\in A}\sum_{n\in N} \lambda_a^n(t)[u_a^n(t) - \xi_a x_a^n(t)]$$

where $\lambda_a^n(t)$ is the costate variable associated with the state equation (7) and $\lambda(t)$ is described by $(\ldots\ldots, \lambda_a^1(t), \ldots, \lambda_a^q(t), \ldots\ldots)$ for all $t\in[0,T]$. We then construct the Lagrangian, which is the augmented Hamiltonian:

(16) $$L[x(t),u(t),\lambda(t),\mu(t)] = H[x(t),u(t),\lambda(t)] + \sum_{k\in N}\sum_{n\in N} \mu_k^n(t)(S_{kn}(t) + \sum_{a\in B(k)} \xi_a x_a^n(t) - \sum_{a\in A(k)} u_a^n(t))$$

where $\mu_k^n(t)$ is the Lagrange multiplier associated with the flow conservation constraint (10) and $\mu(t)$ is described by $(\ldots\ldots, \mu_k^1(t), \ldots, \mu_k^q(t), \ldots\ldots)$.

The Pontryagin minimum principle requires that the differential equations governing the evolution of the costate variables must be defined as follows:

(17) $$\dot{\lambda}_a^n(t) = -\,\partial L[x(t),u(t),\lambda(t),\mu(t)]/\partial x_a^n(t) = -\,C_a'[x_a(t)] + \xi_a[\lambda_a^n(t) - \mu_k^n(t)]$$

$$\forall\ a\in B(k)\quad \forall\ k\in N\quad \forall\ n\in N\quad \forall\ t\in[0,T],\ k\neq n$$

where

$$\frac{\partial C_a[x_a(t)]}{\partial x_a^n(t)} = \frac{dC_a[x_a(t)]}{dx_a(t)} \cdot \frac{\partial x_a(t)}{\partial x_a^n(t)} = C_a'[x_a(t)].$$

Equation (17) will be called the costate equation. Let $\phi_a^n[x_a^n(T)]$ denote the salvage value function when the terminal state is $x_a^n(T)$. Since we impose no constraint on the values of the state variables at the terminal time T, the values of the salvage value functions must be equal to zero. Therefore, the terminal boundary conditions on the costate variables, which are also called the transversality conditions, are given by

(19) $$\lambda_a^n(T) = \partial\phi_a^n[x_a^n(T)]/\partial x_a^n(T) = 0 \qquad \forall\ a\in A\quad \forall\ n\in N.$$

The Pontryagin minimum principle allows us to convert a continuous time optimal control problem into a series of constrained static optimization problems at each instant in time. It follows that

(20) Minimize $L[x(t),u(t),\lambda(t),\mu(t)]$

subject to

$$0 \leq u_a{}^n(t) \leq S_{kn}(t) + \sum_{a \epsilon B(k)} \xi_a x_a{}^n(t)$$

$$\forall \ a \epsilon A(k) \quad \forall \ k \epsilon N \quad \forall \ n \epsilon N \quad \forall \ t \epsilon [0,T], \ k \neq n.$$

Thus, the Pontryagin minimum principle can be regarded as the extension of the method of Lagrange multipliers to dynamic optimization problem. Notice that we cannot minimize the Lagrangian with respect to the state variables because they are not subject to choice; however, we must choose the time-trajectories of the optimal control variables so that the resultant values of the state variables could minimize the total transportation cost over the period [0,T].

Given the values of $x(t)$, $\lambda(t)$ and $\mu(t)$, the necessary conditions for $u^*(t)$ to be optimal in the static optimization problem (20) are:

(21) $$L[x^*(t),u^*(t),\lambda(t),\mu(t)] \leq L[x^*(t),u(t),\lambda(t),\mu(t)]$$

$$\forall \ u(t) \epsilon U[x(t)] \quad \forall \ t \epsilon [0,T]$$

where $U[x(t)]$ denotes the set of admissible controls satisfying (11). Note that $x^*(t)$ and $u^*(t)$ are the optimal values of the state and control variables. Since the Lagrangian (16) is linear in the control variables, minimization of the Lagrangian with respect to $u(t)$ requires that only three cases can occur:

(22) $$u_a{}^{n*}(t) = \begin{cases} 0 & \text{if } \lambda_a{}^n(t) > \mu_k{}^n(t) \\ \text{singular} & \text{if } \lambda_a{}^n(t) = \mu_k{}^n(t) \\ S_{kn}(t) + \sum_{a \epsilon B(k)} \xi_a x_a{}^n(t) & \text{if } \lambda_a{}^n(t) < \mu_k{}^n(t) \end{cases}$$

$$\forall \ a \epsilon A(k) \quad \forall \ k \epsilon N \quad \forall \ n \epsilon N \quad \forall \ t \epsilon [0,T], \ k \neq n.$$

The possible optimal solution thus involves a combination of "bang-bang" control and "singular" control; in other words, the optimal control lies either on the boundary of the admissible control set or within the boundary. At each instant in time, bang-bang controls can be considered all-or-nothing traffic assignment, whereas singular controls can be considered stochastic multipath traffic assignment. Note that each optimal control is determined by the sign of the partial derivative of the Lagrangian with respect to the control variable evaluated along the optimal trajectories, that is, $\partial L^*/\partial u_a{}^n(t) = \lambda_a{}^n(t) - \mu_k{}^n(t)$. In general, the function $[\lambda_a{}^n(t) - \mu_k{}^n(t)]$ is known as the switching function, since the optimal control switches between bang-bang control and singular control when $[\lambda_a{}^n(t) - \mu_k{}^n(t)]$ changes its sign. This statement can also be expressed as a variational inequality:

(23) $$\{\partial L[x^*(t),u^*(t),\lambda(t),\mu(t)]/\partial u(t)\} \cdot [u(t) - u^*(t)] \geq 0$$

$$\forall \ u(t) \epsilon U[x(t)] \quad \forall \ t \epsilon [0,T].$$

In particular, for the case when $\lambda_a^n(t) = \mu_k^n(t)$, the Lagrangian cannot be minimized with respect to a choice of $u_a^n(t)$ and the Performance index J is said to be insensitive to $u_a^n(t)$. Generally, if minimization of the Lagrangian leads to non-unique determination of control variables, the optimal solution is referred to as the singular control. Note that singular controls will be discussed in the subsequent section.

The necessary conditions for optimality of the control problem (14) are summarized as follows:

(24) $\dot{x}_a^n(t) = u_a^n(t) - \xi_a x_a^n(t) \qquad \forall\ a \epsilon A \quad \forall\ n \epsilon N \quad \forall\ t \epsilon [0,T]$

(25) $x_a^n(0) = x_a^{n,o} \geq 0 \qquad \forall\ a \epsilon A \quad \forall\ n \epsilon N$

(26) $\dot{\lambda}_a^n(t) = -\ C_a'[x_a(t)] + \xi_a[\lambda_a^n(t) - \mu_k^n(t)]$

$\forall\ a \epsilon B(k) \quad \forall\ k \epsilon N \quad \forall\ n \epsilon N \quad \forall\ t \epsilon [0,T],\ k \neq n$

(27) $\lambda_a^n(T) = 0 \qquad \forall\ a \epsilon A \quad \forall\ n \epsilon N$

(28) $$u_a^{n*}(t) = \begin{cases} 0 & \text{if } \lambda_a^n(t) > \mu_k^n(t) \\ \text{singular} & \text{if } \lambda_a^n(t) = \mu_k^n(t) \\ S_{kn}(t) + \sum\limits_{a \epsilon B(k)} \xi_a x_a^n(t) & \text{if } \lambda_a^n(t) < \mu_k^n(t) \end{cases}$$

$\forall\ a \epsilon A(k) \quad \forall\ k \epsilon N \quad \forall\ n \epsilon N \quad \forall\ t \epsilon [0,T],\ k \neq n.$

Note that the differential equations for the state variables and the differential equations for the costate variables plus all boundary conditions are called the cononical equations, which may be solved as a two-point boundary-value problem. The control problem (14) will have an optimal solution in the form of bang-singular-bang controls.

5. SUFFICIENCY

The Pontryagin's necessary conditions for optimality (24)-(28) can propose a certain number of admissible solutions, while assuring that there will be no other candidate that can solve the optimal control problem (14). However, the Pontryagin minimum principle cannot tell us whether an admissible solution is optimal or not. In this section, we shall establish that the Pontryagin's necessary conditions are also sufficient for the dynamic system optimal traffic assignment model (14). To this end we present the Seierstad-Sydsaeter sufficiency theorem (Seierstad and Sydsaeter, 1977, p.376-378) and generalize it for our control problem as follows:

Seierstad-Sydsaeter Sufficiency Theorem

Let $[x^*(t),u^*(t)]$ be an admissible state-control pair. Assume that there exist vector functions $\lambda(t)$ and $\mu(t)$ where $\lambda(t)$ is continuous and $d\lambda(t)/dt$ and $\mu(t)$ are piecewise continuous. Then $[x^*(t),u^*(t)]$ is an optimal state-control pair if the Pontryagin's necessary conditions for optimality (24)-(28) are satisfied for all $t \epsilon [0,T]$ as well as the following conditions:

(i) The derived Hamiltonian $\tilde{H}[x(t),\lambda(t)]$ is a convex function of $x(t)$ on $A_u(t)$ if $A_u(t)$ is convex, where

$$\tilde{H}[x(t),\lambda(t)] = \min_{u(t)\epsilon A_x(t)} H[x(t),u(t),\lambda(t)] \quad \forall\ t\epsilon[0,T] \quad \forall\ x\epsilon A_u(t),$$

$A_x(t) = \{u(t)\epsilon U[x(t)]: h[x(t),u(t)] = 0\}$, and

$A_u(t) = \{x(t): h[x(t),u(t)] = 0 \text{ for some } u(t)\epsilon U[x(t)]\}$.

(ii) The linear independence constraint qualification is satisfied for $[x^*(t),u^*(t)]$ for all $t\epsilon[0,T]$, provided

$$\text{rank } [\partial h_k{}^n(x,u)/\partial u_a{}^n(t)]_{\substack{k\epsilon N \\ a\epsilon A}} = \text{the number of indices in } I(t)$$

where $I(t) = \{i: h_i{}^n[x^*(t),u^*(t)] = 0\}$.

We are now ready to state and prove the following theorem:

Theorem 1. The optimality conditions (24)-(28) are both necessary and sufficient to characterize an optimal solution of the dynamic system optimal traffic assignment model (14).

PROOF. To establish sufficiency we use the above Seierstad-Sydsaeter sufficiency theorem. It tells us that conditions (i) and (ii) must be satisfied for the necessary conditions (24)-(28) to be sufficient.

First, the derived Hamiltonian can be written as

$$\tilde{H}[x(t),\lambda(t)] = \sum_{a\epsilon A} C_a[x_a(t)] + \sum_{a\epsilon A}\sum_{n\epsilon N} \lambda_a{}^n(t)[\bar{u}_a{}^n(t) - \xi_a x_a{}^n(t)] \tag{29}$$

where $\bar{u}_a{}^n(t)\epsilon A_x(t)$ for all $a\epsilon A$, $n\epsilon N$ and $t\epsilon[0,T]$. We can see that the derived Hamiltonian is strictly convex for all $x_a{}^n(t)$ because $C_a[x_a(t)]$ was assumed to be a convex function of $x_a(t)$ for all $a\epsilon A$ and $\xi_a x_a{}^n(t)$ is linear. Furthermore, the set $A_u(t)$ is convex because both the state equations (7) and the flow conservation constraints (10) are linear in the state and control variables. Hence, condition (i) is satisfied.

Second, the v x r Jacobian matrix J^n can be written for each destination $n\epsilon N$ and time $t\epsilon[0,T]$:

$$J^n = [\partial h_k{}^n[x(t),u(t)]/\partial u_a{}^n(t)]_{\substack{k=1,\ldots\ldots,v \\ a=1,\ldots\ldots,r}} \tag{30}$$

Since any arc belongs to only one set of the form A(k) which is the set of arcs whose tail node is k, each column in the Jacobian matrix J^n has at most one non-zero element. It means that every Jacobian matrix J^n has v linearly independent rows; consequently J^n must be of rank v. In the contrary case, the constraint qualification is not satisfied, since at least one of the rows in the Jacobian matrix consists of only zeros and thus the rows in the matrix are linearly dependent. Hence, condition (ii) is satisfied. Q.E.D.

6. SINGULAR CONTROLS

The singular controls are encountered in our control problem (14) because the Lagrangian (16) is linear in the control variables. Hence, the optimality conditions (24)-(28) derived by the Pontryagin minimum principle yield no useful information to determine the optimal control variable $u_a^n(t)$ when the value of the associated switching function $[\lambda_a^n(t) - \mu_k^n(t)]$ is equal to zero for a finite time interval. This circumstance is referred to as a singular control. Nonetheless, it may be possible to derive an explicit expression for the singular control in terms of the associated state variables, costate variables, and Lagrange multipliers. If $\partial L/\partial u_a^n(t) = \lambda_a^n(t) - \mu_k^n(t) = 0$ for some $k \epsilon N$, $n \epsilon N$, and $a \epsilon A(k)$ during a finite time interval $[t_1, t_2] \subseteq [0,T]$, it follows that

(31) $$\frac{d}{dt}\left[\frac{\partial L}{\partial u_a^n(t)}\right] = \dot{\lambda}_a^n(t) - \dot{\mu}_k^n(t) = 0$$

(32) $$\frac{d^2}{dt^2}\left[\frac{\partial L}{\partial u_a^n(t)}\right] = \ddot{\lambda}_a^n(t) - \ddot{\mu}_k^n(t) = 0.$$

Let us consider an arc whose tail node is k and head node is s, that is, $a=(k,s) \epsilon A$. Using the costate equation (26) we may rewrite equations (31) and (32) as follows:

(33) $$- C_a'[x_a(t)] + \xi_a[\lambda_a^n(t) - \mu_s^n(t)] - \dot{\mu}_k^n(t) = 0$$

(34) $$- C_a''[x_a(t)]\dot{x}_a^n(t) + \xi_a[\dot{\lambda}_a^n(t) - \dot{\mu}_s^n(t)] - \ddot{\mu}_k^n(t) = 0$$

Substituting the state equation (24) and relationship (31) into the equation (34) yields an explicit expression for the singular control:

(35) $$u_a^n(t) = \frac{\xi_a[\dot{\mu}_k^n(t) - \dot{\mu}_s^n(t)] - \ddot{\mu}_k^n(t) + \xi_a x_a^n(t) C_a''[x_a(t)]}{C_a''[x_a(t)]}$$

$$\forall\ a=(k,s) \epsilon A\ \forall\ k \epsilon N\ \forall\ s \epsilon N\ \forall\ n \epsilon N\ \forall\ t \epsilon [t_1,t_2] \subseteq [0,T],\ k \neq s \neq n.$$

An important question now arises as to whether the singular control given by (35) is optimal or not. Recently, several necessary conditions have been derived to replace the Pontryagin's minimum principle so that the singular controls could be tested for optimality. In this paper, we shall use the generalized Legendre-Clebsch condition (see Bryson and Ho, 1975, pp.246-270) to test the optimality of singular controls as follows:

$$(36)\qquad \frac{\partial}{\partial u_a{}^n(t)}\left[\frac{d^2}{dt^2}\left[\frac{\partial L}{\partial u_a{}^n(t)}\right]\right] = -\,C_a''[x_a(t)] \le 0.$$

Since $C_a[x_a(t)]$ was assumed to be strictly convex for all $x_a(t)$, the generalized convexity condition (36) is satisfied. Hence, we can state that the singular control (35) is optimal. However, (36) is not a sufficient condition for optimality of singular controls.

7. ECONOMIC INTERPRETATION OF OPTIMALITY CONDITIONS

In this section, we shall explore the economic interpretation of optimality conditions (24)-(28). To this end we shall derive economic interpretations of the Lagrangian, costate variables and Lagrange multipliers, respectively. We first multiply the Lagrangian (16) by dt:

$$(37)\qquad Ldt = \sum_{a\epsilon A} C_a[x_a(t)]dt + \sum_{a\epsilon A}\sum_{n\epsilon N} \lambda_a{}^n(t)\dot{x}_a{}^n(t)dt$$

$$+ \sum_{k\epsilon N}\sum_{n\epsilon N} \mu_k{}^n(t)\{S_{kn}(t) + \sum_{a\epsilon B(k)} \xi_a x_a{}^n(t) - \sum_{a\epsilon A(k)} u_a{}^n(t)\}dt.$$

Provided that all flow conservation constraints are binding, (37) can be rewritten as

$$(38)\qquad Ldt = \sum_{a\epsilon A} C_a[x_a(t)]dt + \sum_{a\epsilon A}\sum_{n\epsilon N} \lambda_a{}^n(t)dx_a{}^n(t).$$

The first term represents the instantaneous total transportation cost spent during the time period [t,t+dt]. We may interpret the first term as the direct contribution to the performance index J in dollars from time t to t+dt. The second term represents the cost incurred by the incremental vehicle unit with destination n on arc a during [t,t+dt] and it may be interpreted as the indirect contribution to J in dollars. Hence, Ldt can be interpreted as the total contribution to J from time t to t+dt if the network is in state x(t) and control u(t) is applied.

We now can understand why the Lagrangian (16) must be minimized with respect to the control variables at each time t, as shown in (20). If only the first term $\sum_{a\epsilon A} C_a[x_a(t)]dt$ were minimized at each time t, the performance index J would not be minimized because the values of the optimal control variables in the vicinity of any time t influence the values of the corresponding state variables at all points of time after time t. It also implies that the indirect influence of the control variables in changing the instantaneous total transportation cost rate at each time t must be taken into account. Consequently, the Lagrangian can be interpreted as a surrogate instantaneous transportation cost rate to be minimized at each time t.

Let us define $J^o[x(t),t]$ as the optimal value function that has the minimum value of the performance index J of the control problem (14) beginning at initial state x(t) and initial time t:

$$(39)\qquad J^o[x(t),t] = \min_{u(t)\in A_x(t)} \sum_{a\in A} \int_t^T C_a[x_a(\tau)]\, d\tau.$$

It is well known (see Bryson and Ho, 1975, pp.134) that

$$(40)\qquad \lambda_a^n(t) = \frac{\partial J^o[x(t),t]}{\partial x_a^n(t)}$$

along the optimal state trajectory. Hence, $\lambda_a^n(t)$ can be interpreted as the marginal contribution of an additional infinitesimal vehicle unit with destination n on arc a at time t to the total transportation cost spent during the period [t,T] in the network. Since $J^o[x(t),t]$ has the dimension of a total cost in dollars, $\lambda_a^n(t)$ has the dimension of a marginal cost or a shadow price in dollars per vehicle. Therefore, the costate variables have an interpretation as the dynamic equivalent of the Lagrange multipliers encountered in a nonlinear programming problem. The costate variable $\lambda_a^n(t)$ is sometimes called the influence function because a small increment in $x_a^n(t)$ results in a small increment in the value of $J^o[x(t),t]$ at the rate $\lambda_a^n(t)$.

We shall further establish the economic interpretation of the costate variables. Multiplying the costate equation (28) by dt yields

$$(41)\qquad d\lambda_a^n(t) = \{- C_a'[x_a(t)] + \xi_a[\lambda_a^n(t) - \mu_k^n(t)]\}\, dt.$$

Integrate (41) from time t to T as follows:

$$(42)\qquad \int_t^T d\lambda_a^n(\tau) = - \int_t^T [\partial L[x(\tau),u(\tau),\lambda(\tau),\mu(\tau)]/\partial x_a^n(\tau)]\, d\tau.$$

Because $\lambda_a^n(T) = 0$, we may rewrite (42) as

$$(43)\qquad \lambda_a^n(t) = \int_t^T [\partial L[x(\tau),u(\tau),\lambda(\tau),\mu(\tau)]/\partial x_a^n(\tau)]\, d\tau.$$

Therefore, the marginal cost of a vehicle on arc a with destination n at time t is the integral of the surrogate instantaneous marginal transportation cost rate, $\partial L/\partial x_a^n(t)$, from time t to T. This suggests that $\lambda_a^n(t)$ includes the cost component of the additional cost burden that a small increment in $x_a^n(t)$ inflicts on each one of the vehicles that will traverse arc a in all future instants after the time t.

We shall conjecture the economic interpretation of the Lagrange multipliers associated with the flow conservation constraints. On the analogy of an economic interpretation of the Lagrange multiplier derived in Benavie (1972, pp.256-259), we know that

$$(44) \qquad \mu_k^n(t) = \frac{\partial J^o[x(t),t]}{\partial[S_{kn}(t) + \sum_{a \epsilon B(k)} \xi_a x_a^n(t)]}$$

Therefore, the Lagrange multiplier $\mu_k^n(t)$ can be interpreted as the marginal cost (or the shadow price) of a small relaxation of the flow conservation constraint $h_k^n[x(t),u(t)]$. In particular, along the optimal state trajectory, it represents the sensitivity of the optimal value of $J^o[x(t),t]$ to a small increment in the number of vehicles at node k with destination n at time t. Assuming that vehicles terminated at each destination have no impact on the optimum value of $J^o[x(t),t]$, we also know that

$$(45) \qquad \mu_n^n(t) = \frac{\partial J^o[x(t),t]}{\sum_{a \epsilon B(n)} \xi_a x_a^n(t)]} = 0.$$

We are now ready to interpret the Kuhn-Tucker optimality conditions (28). The relationship $\lambda_a^n(t) > \mu_k^n(t)$ implies that if the marginal cost of an additional vehicle on arc a with destination n is greater than the marginal cost of an additional vehicle at node k with destination n, then no vehicle will enter arc a at time t to travel to destination n from node k. Conversely, $\lambda_a^n(t) < \mu_k^n(t)$ means that all of vehicles generated at node k and vehicles exiting from upstream arcs connected to node k will enter arc a at time t in order to travel to destination n. This interpretation implies that holding a vehicle at node k is more expensive than sending it to any downstream arc at time t from node k. In particular, when $\lambda_a^n(t) = \mu_k^n(t)$, the number of vehicles entering arc a from node k with destination n at time t will be determined as the singular control (35).

Finally, the optimality conditions (28) are analogous to the flow of electricity in that only three cases can occur: [1] no electric current flow from a node with lower voltage to a node with higher voltage; [2] singular flow between nodes with the same voltage; and [3] maximum flow from a node with higher voltage to a node with lower voltage.

8. DYNAMIC GENERALIZATION OF WARDROP'S SECOND PRINCIPLE

Let us assume that all network users cooperate in minimizing the total transportation cost, or the rational behavior of all network users is completely controllable by a central traffic authority. Then we may state a dynamic generalization of Wardrop's second principle:

If the instantaneous marginal total costs for all paths that are being used are identical and equal to the instantaneous minimum marginal total cost at any instant in time for each origin-destination pair, then the corresponding flow pattern is said to be system optimized.

We now wish to ascertain whether a solution of the dynamic system optimal traffic assignment model (14), described by time trajectories of the state and control variables, corresponds to a dynamic generalization of Wardrop's second principle at each instant in time. To this end, let us consider a path p connecting an origin node $k \epsilon N$ and a destination node $n \epsilon N$, which may be expressed in the generic form as

(46) $$P \equiv [\ k=k_o,\ a_1,\ k_1,\ \ldots\ldots,\ k_{m-1},\ a_m,\ k_m=n\ ].$$

Let A(p) denote the set of arcs comprising a path p. For each $a \epsilon A(p)$, the costate equation (26) may be rewritten as

(47) $$\{C_a'[x_a(t)] + \dot{\lambda}_a^n(t)\}/\xi_a = \lambda_a^n(t) - \mu_k^n(t).$$

The left-side term of (47) has the units of instantaneous marginal cost because both $C_a'[x_a(t)]$ and $d\lambda_a^n(t)/dt$ have the units of instantaneous marginal cost rate, while ξ_a has the units of incremental exit flow rate per incremental vehicle unit. Obviously, the right-side term of (47) has the units of instantaneous marginal cost. Let $\Phi_p(t)$ denote the instantaneous marginal total cost on path p at time t:

(48) $$\Phi_p(t) \equiv \sum_{a \epsilon A(p)} \{C_a'[x_a(t)] + \dot{\lambda}_a^n(t)\}/\xi_a \qquad \forall\ p \epsilon P_{kn} \quad \forall\ t \epsilon [0,T].$$

It may be stated that $\Phi_p(t)$ consists of both static and dynamic terms. The term $C_a'[x_a(t)]/\xi_a$ is static in the sense that it represents the instantaneous marginal cost on arc a where $x_a(t)$ is the number of vehicles traveling on arc a at time t. The term $[d\lambda_a^n(t)/dt]/\xi_a$ is dynamic in the sense that it represents the time rate of change of the marginal cost of a vehicle unit on arc a which is weighted by ξ_a.

Let $\bar{u}_a^n(t)$ denote the maximum rate of flow entering arc a with destination n at time t, which is the sum of $S_{kn}(t)$ and $\sum_{a \epsilon B(k)} \xi_a x_a^n(t)$. We are now in a position to state and prove the following theorem:

> Theorem 2. If $0 < u_a^n(t) < \tilde{u}_a^n(t)$ for all $a \epsilon A(p)$ at some time $t \epsilon [0,T]$, then $\Phi_p(t) = \inf\ \{\Phi_b(t):\ \forall\ b \epsilon P_{kn}\}$ for a solution of the dynamic system optimal traffic assignment model (14).

PROOF. Let $p \epsilon P_{kn}$ have the generic form given in (46). Using (47), we may write (48) as follows:

(49) $$\Phi_p(t) = \sum_{i=1}^{m} [\lambda_{a_i}^n(t) - \mu_{k_i}^n(t)].$$

From (28) we know that $0 < u_a^n(t) < \tilde{u}_a^n(t)$ for all $a \epsilon A(p)$ if and only if

(50) $$\lambda_{a_i}^n(t) = \mu_{k_{i-1}}^n(t) \qquad \text{for } i = 1,\ \ldots.,\ m.$$

It follows at once from (49) and (50) that for $p \epsilon P_{kn}$

$$\Phi_p(t) = \sum_{i=1}^{m} [\mu_{k_{i-1}}^{n}(t) - \mu_{k_i}^{n}(t)] = \mu_{k_o}^{n}(t) - \mu_{k_m}^{n}(t) = \mu_k^{n}(t). \tag{51}$$

Hence, the theorem follows immediately. Q.E.D.

9. CONCLUSION

In this paper, we explored only theoretical aspects of the dynamic system optimal traffic assignment model. The future research must include [1] the development of an efficient solution algorithm for a large-scale network, and [2] its application to dynamic route guidance and traffic control system. Recently, the dynamic user optimal traffic assignment model was developed in Wie (1988a, 1988b) and Friesz, Luque, Tobin and Wie (1989). This model seems to provide a more realistic network analysis because it assumes that every network user simultaneously minimizes his/her measure of unit path travel cost while recognizing both the time-varying nature of travel demands and the presence of congestion externalities.

REFERENCE

Bryson, A.E. and Y-C Ho (1975), Applied optimal control, revised version, John Wiley & Sons, New York

Friesz, T.L., F.J. Luque, R.L. Tobin and B.W. Wie (1989) "*Dynamic network traffic assignment considered as a continuous time optimal control problem*," Operations Research, Volume 37, 6

Merchant, D.K. (1974) A study of dynamic traffic assignment and control, Ph.D. Dissertation, Cornell University

Merchant, D.K. and G.L. Nemhauser (1978a) "*A model and an algorithm for the dynamic traffic assignment problems*," Transportation Science, 12, 183-199

Merchant, D.K. and G.L. Nemhauser (1978a) "*Optimality conditions for a dynamic traffic assignment model*," Transportation Science, 12, 200-207

Pontryagin, L.S., V.A. Boltyanski, R.V. Gankrelidge, and E.F. Mishenko (1962) The Mathematical Theory of Optimal Process, Wiley, New York

Wardrops, J.G. (1952) "*Some theoretical aspects of road traffic research*," Proceedings, Institute of Civel Engineers, II, 1, 325-378

Wie, B.W. (1988a) "*An application of optimal control theory to dynamic user equilibrium traffic assignment*," Transporation Research Record, in press

Wie, B.W. (1988b) Dynamic Models of Network Traffic Assignment; A Control Theoretic Approach, Ph.D. Dissertation, Department of City and Regional Planning, University of Pennsylvania