INT. J. CONTROL, 1984, VOL. 40, NO. 4, 813–829

# Hierarchical optimal control of urban traffic networks

E. S. PARK†, J. H. LIM†, I. H. SUH† and Z. BIEN†

This paper deals with the problem of optimally controlling traffic flows in urban traffic networks. For this, a non-linear discrete-time model of an urban traffic network is first suggested in order to handle the phenomenon of traffic flows such as oversaturatedness and undersaturatedness. Then an optimal control problem is formulated and a hierarchical optimization technique is applied, which is based on a prediction-type two-level method of Hirvonen and Hakkala.

## 1. Introduction

Since Webster (1959) reported an efficient model of fixed-cycle traffic signals on the oversaturated urban traffic networks, many researchers have investigated the problem of effectively controlling traffic signals at traffic stops. Their major concerns have been to establish well-defined models of road traffic networks and their computational methods for control. In particular, Gazis (1964) applied the maximum principle to a road traffic problem. However, the high dimensionality of traffic networks causes serious computational difficulties in obtaining an optimal control law, which led Burhardt and Kulikowski (1970) and Tamura (1975) to apply decomposition and hierarchical optimization techniques for an effective control algorithm. Recently, Sarachik and Özgüner (1982) proposed a decentralized dynamic routing strategy for clearing congested traffic networks with deterministic inputs.

Singh and Tamura (1974) first handled the control problem of an oversaturated urban traffic network by suggesting a deterministic discrete-time dynamic model and applying a hierarchical computational method. In their problem formulation, however, a certain inequality constraint was unrealistically incorporated, and as a consequence, an optimal control law for a given problem may not be correctly obtained even though it exists. To remedy the shortcomings of Singh and Tamura's model, Lim *et al.* (1981) proposed a model of an oversaturated urban traffic network by employing inequality constraints which are slightly different from that of Singh and Tamura.

In many real traffic situations, the phenomenon of oversaturatedness and undersaturatedness must be handled concurrently in one framework, but the models considered in either Singh and Tamura (1974) or Lim *et al.* (1981) are restricted to the dynamics of the oversaturated urban traffic networks only. In this paper we suggest a model which describes the dynamics of urban traffic networks involving the oversaturatedness as well as the undersaturatedness. With the model containing non-linear terms of exponential type, an optimal control problem is then formulated and an hierarchical optimization technique based on Hirvonen and Hakkala (1979) is applied. The proposed model and hierarchical optimization method are tested by means of several examples.

Received 11 January 1984.
† Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, P.O. Box 150, Chongyangni, Seoul 131, Korea.

## 2. Model and control of urban traffic networks

Consider the problem of controlling signals of a traffic network of the type shown in Fig. 1. Let $i=1$ and $i=2$ indicate the horizontal traffic flow direction and the vertical direction, respectively. For each discrete-time $k$, let $q_i(k)$ denote the arrival rate of vehicles in the lane of direction $i$ and $\bar{g}_i(k)$ the averaged departure rate over the duration of cycle time $C$. If we let $l$ be the loss time due to the amber phase, $G_i$ the duration of the green phase in the direction $i$ and $\bar{G}_e$ the duration of effective green, then the relation

$$C=G_1+G_2+l=\bar{G}_e+l \tag{1}$$

holds.

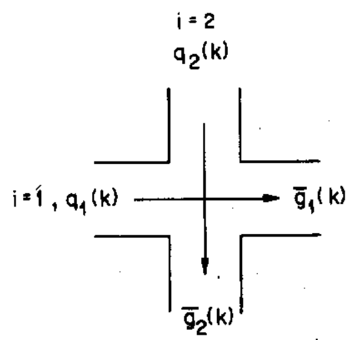


Figure 1. One-way no-turn intersection.

Define the state variable $x_i(k)$ to be the number of vehicles which are waiting at time $k$ to pass the intersection in the lane of direction $i$. Then, for the given cycle time $C$, it is easily shown that the evolution process of the states can be described by (Singh and Tamura 1974)

$$x_i(k+1)=x_i(k)+q_i(k)-\bar{g}_i(k), \quad i=1,2, \quad k=0,1,\ldots,k_f-1 \tag{2}$$

Here $k_f$ is the final time of the control duration.

Now consider two phenomena of the traffic networks such as undersaturated and oversaturated flow. In the case of the oversaturated traffic flow, the state queue $x_i(k)$ in the direction $i$ exists even at the end of the green interval started at the time $k$. However, in the case of the undersaturated flow, the sum of the state queue $x_i(k)$ and input queue $q_i(k)$ in the direction $i$ can be eliminated during the green phase of the direction $i$. To describe these two phenomena of the traffic flows more specifically, define the control variable $u_i(k)$ to be the percentage of ' green ' over $C$ in the direction $i$. It is assumed that the averaged departure rate $\bar{g}_i(k)$ can be given by the relationship

$$\bar{g}_i(k)=S_i(k,x_i(k),q_i(k))u_i(k) \tag{3}$$

where $S_i(k,x_i(k),q_i(k))$ is the output flow rate. It is further assumed that $S_i(k,x_i(k),q_i(k))$ is modelled by the relation

$$S_i(k,x_i(k),q_i(k))=S_{it}\left[1-\exp\left\{-K_i\frac{(x_i(k)+q_i(k))}{X_{ic}}\right\}\right] \tag{4}$$

where $S_{if}$ is the saturation flow rate in the direction $i$. That is, $S_{if}$ is the maximum number of vehicles which can pass through the intersection per cycle in the direction $i$ if the signal in this direction is all the available green during one cycle time. $K_i$ and $X_{ic}$ are positive real constants to be determined by the road capacity. Then, we find that eqns. (2) to (4) would describe the dynamics of the urban traffic flows such as undersaturatedness and oversaturatedness in the following sense. As shown in Fig. 2, if the sum of the state queue and the input queue in the direction $i$ at the instant $k$ exceeds some threshold value $X_{ic}$, then $S_i(k, x_i(k), q_i(k))$ will be at most $S_{if}$. This implies that the road in the direction $i$ becomes oversaturated at the instant $k$. However, if the sum of the state queue and the input queue in the direction $i$ at the instant $k$ does not exceed $X_{ic}$, the output flow rate, $S_i(k, x_i(k), q_i(k))$, would be less than $S_{if}$ and thus depends on the exponential function of the sum of the state queue and the input queue. This implies that the road in the direction $i$ is undersaturated at the instant $k$. Therefore, $X_{ic}$ and $K_i$ should be determined so as to accommodate these situations.

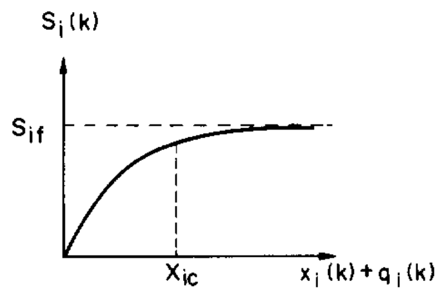


Figure 2. A model for the output flow rate.

*Remark* 1

The experimental traffic data in Lincoln Tunnel quoted by Greenberg (1959) as well as the Edie–Underwood's model (Edie 1961) justifies our choice of the output flow rate $S_i(k, x_i(k), q_i(k))$ in eqn. (4) as an exponential function of the state queue and the input queue. Similar modelling may be found in Lemieux (1978), in which an exponential type of output flow transfer curve is utilized to control the switched telephone traffic flows.

Now, from eqns. (1), (3) and (4), it follows that

$$\begin{aligned} x_i(k+1) &= x_i(k) + q_i(k) - \bar{g}_i(k) \\ &= x_i(k) + q_i(k) - S_i(k, x_i(k), q_i(k))u_i(k) \end{aligned} \tag{5}$$

for $i = 1, 2$, $k = 0, 1, \ldots, k_f - 1$, and

$$u_1(k) + u_2(k) = \frac{(G_1 + G_2)}{C} = 1 - \frac{l}{C} = G_e \tag{6}$$

where $G_e$ is the normalized effective green, i.e. $G_e = \bar{G}_e/C$. Following Singh and Tamura (1974), we require that the state and control variables be subject to the inequality constraints

$$0 \leqslant x_i(k) \leqslant x_{i,\max}, \quad k = 0, 1, \ldots, k_f, \quad i = 1, 2 \tag{7}$$

and

$$u_{i,\min} \leqslant u_i(k) \leqslant u_{i,\max}, \quad k=0, 1, \ldots, k_f-1, \quad i=1, 2 \tag{8}$$

For the interconnected road as shown in Fig. 3, the output, i.e. the averaged departure rate $\bar{g}_1(k)$ in the lane of direction $i=1$, becomes the delayed input of the lane of $i=3$. The model in this case can be obtained by incorporating a delay element (Lim *et al.* 1981). To be specific, let $d$ denote the delay time during which the output $\bar{g}_1(k)$ becomes the input. Further let

$$d=(\theta+a)c \tag{9}$$

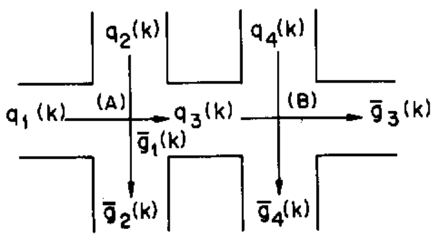


Figure 3. A model for the interconnected road.

where $\theta$ is an integer and $0 \leqslant a < 1$. Then

$$q_3(k)=(1-a)S_1(k-\theta, x_1(k-\theta), q_1(k-\theta))u_1(k-\theta) \\ +aS_1(k-\theta+1, x_1(k-\theta+1), q_1(k-\theta+1))u_1(k-\theta+1) \tag{10}$$

Assuming for simplicity, that $d=C$ ($\theta=1$, $a=0$), the traffic flow process of the interconnected road in Fig. 3 can be described in state variables as

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \\ x_4(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \\ x_4(k) \end{bmatrix}$$

$$+ \begin{bmatrix} -S_1(k, x_1(k), q_1(k)) & 0 & 0 & 0 \\ 0 & -S_2(k, x_2(k), q_2(k)) & 0 & 0 \\ 0 & 0 & -S_3(k, x_3(k), q_3(k)) & 0 \\ 0 & 0 & 0 & -S_4(k, x_4(k), q_4(k)) \end{bmatrix} \begin{bmatrix} u_1(k) \\ u_2(k) \\ u_3(k) \\ u_4(k) \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ S_1(k-1, x_1(k-1), q(k-1)) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u_1(k-1) \\ u_2(k-1) \\ u_3(k-1) \\ u_4(k-1) \end{bmatrix} + \begin{bmatrix} q_1(k) \\ q_2(k) \\ 0 \\ q_4(k) \end{bmatrix} \tag{11}$$

Thus, in general, the traffic flow process of a complex interconnected network can be described by a non-linear difference equation with delays in the controls of the form

$$x(k+1)=E_n x(k)+\sum_{j=0}^{\theta} B_j(k)u(k-j)+c(k), \quad k=0, 1, \ldots, k_f-1 \tag{12}$$

where $x(k)$ is the $n\times 1$ state vector, $E_n$ is an $n\times n$ identity matrix, $\theta$ is the largest integer unit of time delay in the overall networks, $B_j(k)$ $(j=0, 1, \ldots, \theta)$ are $n\times n$ matrices which include non-linear components $S_i(k-j, x_i(k-j), q_i(k-j))$, $u_i(k)$ is the $n\times 1$ control vector and $c(k)$ is the $n\times 1$ vector of inputs denoting the incoming flow rate from the outside of the network.

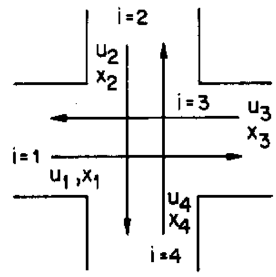


Figure 4. Two-way intersection road.

In eqn. (12), the dimension of the control vector is made equal to that of the state vector for ease of computation, although its dimension is usually reducible. For example, in the case of a two-way intersection road as shown in Fig. 4, the dimension of the control vector $[u_1, u_2, u_3, u_4]^T$ is made equal to that of state vector $[x_1, x_2, x_3, x_4]^T$ though $u_1=u_3$ and $u_2=u_4$. Thus, in the formulation of eqn. (12), some components $u_i(k)$ and $u_j(k)$ of the control vector $u(k)$ may be identical

$$u_i(k)=u_j(k), \quad \text{and for } k=0, 1, \ldots, k_f-1, \text{ for some } i \text{ and } j \tag{13}$$

For future development, let us define $I$ to be the set defined by

$$I=\{(i, j)\,|\,u_i(k)=u_j(k), \quad \text{for } k=0, 1, \ldots, k_f-1\}$$

and let $s$ be the number of elements in $I$.

Also observe that, as shown in eqn. (6), the sum of some control components at each intersection is equal to the normalized duration of the effective green. To be specific, let $t$ be the number of traffic junctions for the green traffic network. At an arbitrary $j$th traffic junction, let $M_j$ be the set of the indices $m_j$ such that the controls $u_{m_j}(k)$ satisfy

$$\sum_{m_j\in M_j} u_{m_j}(k)=G_e, \quad j=1, 2, \ldots, t, \quad k=0, 1, 2, \ldots, k_f-1 \tag{14}$$

In (14), the values of $u_{m_j}(k)$ are the controls associated with the $j$th traffic junction and are different from each other.

Let the cost functional be

$$J = \frac{1}{2} \sum_{k=0}^{k_f - 1} \{ \|x(k+1)\|_Q^2 + \|u(k)\|_R^2 \} \tag{15}$$

where $\|\cdot\|$ denotes the euclidean norm, $Q$ and $R$ are $n \times n$ positive semidefinite and definite diagonal matrices, respectively. Then the optimal control problem for traffic networks can be formulated as follows.

### 2.1. *An optimal control problem (p)*

Find an (optimal) control sequence $u(0), u(1), \ldots, u(k_f - 1)$ for the system described by eqn. (12) in such a way that the cost functional $J$ in eqn. (15) is minimized subject to the inequality constraints in eqns. (7), (8), (12) and (14).

*Remark* 2

Singh and Tamura (1974) developed a simple linear discrete-time model to describe the dynamic behaviour of an oversaturated urban road traffic network. However, in their model, a contradictory situation may occur in the sense that an optimal control scheme for a given problem is not correctly obtained even though it exists. An example was given by Lim *et al.* (1981). Because they utilized a constant saturation flow rate $S_{it}$ as the output flow rate $S_i(k, x_i(k), u_i(k))$ even for an undersaturated intersection, the averaged departure rate was calculated as being greater than the real one, and thus it might lead to a non-optimal control strategy. Lim *et al.* proposed a new model of an oversaturated urban traffic network to remedy the shortcomings of Singh and Tamura's model. To be more specific, they had used a constant output flow rate as in Singh and Tamura's model, but they employed an inequality constraint on the state variable which is different from that of Singh and Tamura. In the model of Lim *et al.*, the existence of the optimal solution is guaranteed only when the sum of the input queue exceeds the sum of averaged departure rates in each direction during the optimization interval. Thus, the model of Lim *et al.* is applicable only to the oversaturated traffic networks as illustrated in the Appendix.

## 3. A hierarchical optimization method

In this section, a hierarchical optimization method is suggested to solve the problem (p) in § 2. This method is based upon a prediction-type two-level method of Hirvonen and Hakkala (1979), which was used for an optimal control problem of non-linear systems. We use the idea of decomposing the problem with respect to time to handle the time delay (Tamura 1975). Furthermore, as given by Lim *et al.* (1981), for ease of computation the dimension of the control variables is made to be equal to that of the state variables as in problem (p). By doing this in addition to time decomposition, the subproblems can be further decomposed into simple quadratic problems which have only two variables $x_i(k)$, $u_i(k)$ and linear constraints in eqns. (7) and (8). In this way, lower-level subproblems are made easy to handle.

To employ a prediction-type two-level optimization method of Hirvonen and Hakkala (1979), we introduce additional variables $x^*(k)$ and $u^*(k)$ to predict the non-linear term $B_j(k)$ in eqn. (12). Then the problem (p) can be modified as the following problem (p*).

### 3.1. *Optimal control problem* $(p^*)$

Find an optimal control sequence $u(0), u(1), \ldots, u(k_t-1)$, while minimizing the cost function $J$ in eqn. (15) subject to the following constraints

(i)
$$x(k+1)=E_n x(k)+\sum_{j=0}^{\theta} B_j^*(k)u(k-j)+c(k) \tag{16}$$

where the entries of $B_j^*(k)$ may include $S_i^*(k-j)$ given by

$$S_i^*(k-j)=S_{it}\left[1-\exp\left\{-K_i\frac{x_i^*(k-j)+q_i^*(k-j)}{X_{ic}}\right\}\right] \tag{17}$$

(ii)
$$x_i(k)=x_i^*(k) \tag{18}$$

(iii)
$$u_i(k)=u_i^*(k) \tag{19}$$

and eqns. (7), (8), (13) and (14).

Here, note that the problem (p*) reduces to the problem (p) if $x_i^*(k)$ and $u_i^*(k)$ could be found such that the constraints in eqns. (18) and (19) are satisfied.

For the modified problem (p*) let us write the lagrangian $L$ as

$$\begin{aligned} L=\sum_{k=0}^{k_t-1}\Bigg[&\tfrac{1}{2}\|x(k+1)\|_Q^2+\tfrac{1}{2}\|u(k)\|_R^2+p^{\mathrm{T}}(k)\Bigg\{-x(k+1)+E_n x(k)\\ &+\sum_{j=0}^{\theta} B_j^*(k)u(k)+c(k)\Bigg\}+\sum_{(i,j)\in I}\beta_{ij}(k)\{u_i(k)-u_j(k)\}\\ &+\sum_{j=1}^{t} r_j(k)\Bigg\{\sum_{m\in M_j}u_m(k)-G_e\Bigg\}+\lambda_1^{\mathrm{T}}(k)\{x(k)-x^*(k)\}\\ &+\lambda_2^{\mathrm{T}}(k)\{u(k)-u^*(k)\}\Bigg]\end{aligned} \tag{20}$$

where $n\times 1$ vectors $p$, $\lambda_1$ and $\lambda_2$ and scalars $\beta_{ij}$ and $r_j$ are multipliers. Define $\beta$ and $r$ as

$\beta \triangleq s\times K_t$ dimensional vector whose elements are $\beta_{ij}(k)$

for $(i,j)\in I$ and $k=0,1,\ldots,k_t-1$,

$r\triangleq t\times K_t$ dimensional vector whose elements are $r_i(k)$

for $j=1,2,\ldots,t$ and $k=0,1,\ldots,k_t-1$. Also, let us define the dual function $\phi$ : $X\times U\times X^*\times R^{s\times k_t}\times R^{t\times k_t}\times X^*\times U^*\supset D(\phi)\to R$ for the modified problem

$$\phi(x^*,u^*,p,\beta,r,\lambda_1,\lambda_2)=\min_{(x,u)\in X\times U} L(x,u,x^*,u^*,p,\beta,r,\lambda_1,\lambda_2) \tag{21}$$

subject to eqns. (7) and (8) with

$$D(\phi)=\{(x^*,u^*,p,\beta,r,\lambda_1,\lambda_2)\in X\times U\times X^*\times R^{s\times k_t}\times R^{t\times k_t}\times X^*\times U^* \mid \text{the minimum in eqn. (20) exists}\} \tag{22}$$

where $x^*\in X$, $u^*\in U$, $p$, $\lambda_1\in X^*$, $\lambda_2\in U^*$, $\beta\in R^{s\times k_t}$ and $r\in R^{t\times k_t}$, and $X^*$ and $U^*$ are dual spaces of the euclidean space $X$ and $U$, respectively. For each $(x^*, u^*, p, \beta, r, \lambda_1, \lambda_2)\in D(\phi)$ define the set $\mu(x^*, u^*, p, \beta, r, \lambda_1, \lambda_2)\subset X\times U$ as follows

$$\mu(x^*, u^*, p, \beta, r, \lambda_1, \lambda_2) = \{(x, u)\in X\times U \,|\, \text{the minimum in eqn. (21) is attained as } (x, u)\} \quad (23)$$

Furthermore, let

$$A=\{(x, u)\in X\times U \,|\, (x, u) \text{ satisfy eqns. (7), (8), (13) and (14)}\} \quad (24)$$

Then it follows from the result of Hirvonen and Hakkala (1979) that, for $(x^{*0}, u^{*0}, p^0, \beta^0, \lambda_1{}^0, \lambda_2{}^0)\in D(\phi)$, $(x^0, u^0)\in\mu(x^{*0}, u^{*0}, p^0, \beta^0, r^0, \lambda_1{}^0, \lambda_2{}^0)$ and $A\subset D(\phi(\cdot, \cdot, p^0, \beta^0, r^0, \lambda_1{}^0, \lambda_2{}^0))$, if $x(k)=x^*(k)$, $u(k)=u^*(k)$ and if for all $(x^*(k), u^*(k))\in A$, $\phi(x^{*0}, u^{*0}, p^0, \beta^0, r^0, \lambda_1{}^0, \lambda_2{}^0)\leqslant\phi(x^*, u^*, p^0, \beta^0, r^0, \lambda_1{}^0, \lambda_2{}^0)$, then the problem (p) has a solution given by $(x^0, u^0)$ and furthermore $\phi$ has a saddle point at $(x^{*0}, u^{*0}, p^0, \beta^0, r^0, \lambda_1{}^0, \lambda_2{}^0)$. Thus the strategy for finding the optimal point of the dual function, which yields the solution $(x^0, u^0)$ to the problem (p), is to find a saddle point of the dual function (Hirvonen and Hakkala 1979).

To find the minimum in eqn. (21) more effectively, it may be necessary to use the decomposition method. Note that, as given by Lim *et al.* (1981), the lagrangian $L$ in eqn. (20) is additively separable with respect to $x_i(k)$ and $u_i(k)$ for given $x_i^*(k)$ and $u_i^*(k)$, and also note that the lagrangian can be decomposed with respect to time $k$ as

$$L=\sum_{k=0}^{k_t} L_k(x, u, x^*, u^*, p, \beta, r, \lambda_1, \lambda_2) \quad (25)$$

where, for $k=0$

$$\begin{aligned} L_0 &= \tfrac{1}{2}\|u(0)\|_R^2+p^{\mathrm{T}}(0)\{E_n x(0)+B_0{}^*(0)u(0)+c(0)\} \\ &\quad + \sum_{(i,j)\in I} \beta_{ij}(0)\{u_i(0)-u_j(0)\} + \sum_{j=1}^{t} r_j(0)\left\{\sum_{m\in M_j} u_m(0)-G_e\right\} \\ &\quad + \lambda_1{}^{\mathrm{T}}(0)(x(0)-x^*(0))+\lambda_2{}^{\mathrm{T}}(0)(u(0)-u^*(0)) \end{aligned} \quad (26)$$

for $k=1, 2, \ldots, k_t-1$

$$\begin{aligned} L_k &= \tfrac{1}{2}\|x(k)\|_Q^2+\tfrac{1}{2}\|u(k)\|_R^2-p^{\mathrm{T}}(k-1)x(k)+p^{\mathrm{T}}(k)x(k) \\ &\quad + \lambda_1{}^{\mathrm{T}}(k)x(k)+\sum_{j=0}^{\theta} p^{\mathrm{T}}(k+j)B_j{}^*(k)u(k)+\sum_{(i,j)\in I}\beta_{ij}(k)\{u_i(k)-u_j(k)\} \\ &\quad + \sum_{j=1}^{t} r_j(k)\sum_{m_j\in M_j} u_{m_j}(k)+\lambda_2{}^{\mathrm{T}}(k)u(k)+p^{\mathrm{T}}(k)c(k)-\sum_{j=1}^{t} r_j(k)G_e \\ &\quad - \lambda_1{}^{\mathrm{T}}(k)x^*(k)-\lambda_2{}^{\mathrm{T}}(k)u^*(k) \end{aligned} \quad (27)$$

and for $k=k_t$

$$K_{k_t}=\tfrac{1}{2}\|x(k_t)\|_Q^2-p^{\mathrm{T}}(k_t-1)x(k_t) \quad (28)$$

Recalling that, for ease of computation, the dimension of the control variables is made equal to that of the state variables and $Q$ and $R$ are diagonal matrices, it is easily shown that $L_k$ can be decomposed into the form of

$$L_k = \sum_{i=1}^{N} L_{ik}(x_i(k), u_i(k), x_i^*(k), u_i^*(k), p(k), \beta(k), r(k), \lambda_1(k), \lambda_2(k)) \tag{29}$$

Consequently, $L$ can be decomposed as

$$L = \sum_{k=0}^{k_t} \sum_{i=1}^{N} L_{ik}(x_i(k), u_i(k), x_i^*(k), u_i^*(k), p, \beta, r, \lambda_1, \lambda_2) \tag{30}$$

This leads to a prediction-type two-level algorithm in which for given $x^*(k)$, $u^*(k)$, $p$, $\beta$, $r$, $\lambda_1$ and $\lambda_2$, $(k_t+1)\times N$ independent subproblems are minimized with respect to $x_i(k)$ and $u_i(k)$ with linear constraints eqns. (7) and (8), and, at the second-level, predictions are made of $x^*(k)$, $u^*(k)$, $p$, $\beta$, $r$, $\lambda_1$ and $\lambda_2$ in order to improve them.

A coordination strategy for updating the values of $x^*(k)$, $u^*(k)$, $p$, $\beta$, $r$, $\lambda_1$ and $\lambda_2$ is proposed based on the gradients of the dual function. Specifically, a two-level algorithm can be given as follows (see Fig. 5).

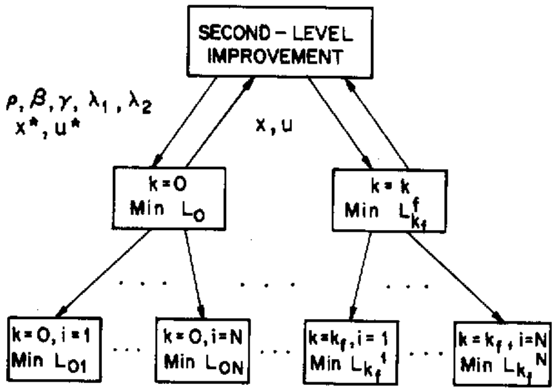


Figure 5. Two-level computation algorithm.

*Level* 1

For $i=1, \ldots, N$ and $k=0, 1, \ldots, k_t-1$, find $x_i(k)$ and $u_i(k)$ such that $L_{ik}$ in eqn. (30) is minimized with inequality constraints in eqns. (7) and (8).

*Level* 2

Update $x^*$, $u^*$, $\lambda_1$ and $\lambda_2$ by the reinjection method and also update $p$, $\beta$ and $r$ according to the multipliers method of Hestenes (1969), i.e.

$$\nabla_{\lambda_1}\phi = 0 \quad \text{gives } x^{*l+1}(k) = x^l(k) \tag{31}$$

$$\nabla_{\lambda_2}\phi = 0 \quad \text{gives } u^{*l+1}(k) = u^l(k) \tag{32}$$

$$\nabla_{x^*}\phi = 0 \quad \text{gives } \lambda_1^{l+1}(k) = \sum_{j=0}^{\theta} \left.\frac{\partial[B_j^*(k)u(k)]}{\partial x^*}\right|_{\substack{x^*=x^l \\ x^*=u^l}} p^l(k) \tag{33}$$

$$\nabla_{u^*}\phi=0 \quad \text{gives} \quad \lambda_2^{l+1}(k)=\sum_{j=0}^{\theta}\frac{\partial[B_j^*(k)u(k)]}{\partial u^*}\bigg|_{\substack{x^*=x^l\\ x^*=u^l}} p^l(k) \tag{34}$$

$$\begin{aligned} p^{l+1}(k) &= p^l(k)+\alpha_1\nabla_p\phi \\ &= p^l(k)+\alpha_1\{-x^l(k+1)+E_n x^l(k)+\sum_{j=0}^{\theta} B_j^l(k)u^l(k)+c(k)\} \end{aligned} \tag{35}$$

$$\beta_{ij}^{l+1}(k)=\beta_{ij}^l(k)+\alpha_2(u_i^l(k)-u_j^l(k)), \qquad \text{for } (i, j)\in I \tag{36}$$

and

$$r_j^{l+1}(k)=r_j^l(k)+\alpha_3\left(\sum_{m\in M_j} u_m(k)-G_e\right), \quad \text{for } j=1, 2, \ldots, t \tag{37}$$

where $l$ is the iteration index and the step sizes, $\alpha_i>0$, $i=1, 2, 3$, are chosen to maximize $\phi$ in each iteration.

*Remark* 3

As given by Hirvonen and Hakkala (1979), the two-level optimization strategy for finding a saddle point of the dual function $\phi$, which yields an optimal solution to the original problem, can be designed with the aid of the gradients of the dual function. In this case the solution is a stationary point of the lagrangian of the modified problem.

Often it would be difficult to choose suitable initial second-level variables to start. Some physical interpretation and computational experience of the problem can be useful. In this type of problem, the computer time and the number of second-level iterations may not be an effective measure for comparison, because they are quite dependent on the initial choice of second-level variables.

## 4. Simulation results

In this section, two examples are illustrated to show the validities of the results in § 2.

*Example* 1

Consider a simple one-way no-turn intersection as shown in Fig. 1. The state equations can be written

$$x_1(k+1)=x_1(k)-S_1(k, x_1(k), q_1(k))u_1(k)+q_1(k) \tag{38}$$

$$x_2(k+1)=x_2(k)-S_2(k, x_2(k), q_2(k))u_2(k)+q_2(k) \tag{39}$$

where $S_i(k, x_i(k), q_i(k))$, $i=1, 2$, are given by

$$S_i(k, x_i(k), q_i(k))=S_{it}\left[1-\exp\left\{-K_i\frac{(x_i(k)+q_i(k))}{X_{ic}}\right\}\right] \tag{40}$$

Here, it is assumed that state and control constraints are given by

$$0{\cdot}2\leqslant u_i(k)\leqslant 0{\cdot}8, \qquad i=1, 2 \tag{41}$$

$$0\leqslant x_i(k)\leqslant x_{i,\max}, \quad i=1, 2 \tag{42}$$

and

$$u_1(k)+u_2(k)=1 \tag{43}$$

Further, the cycle time $C$ and loss time $l$ are assumed to be 60 s and zero, respectively.

The optimal control problem is as follows. Minimize $J$ given by eqn. (15) under the constraints in eqns. (38)–(43). The weighting matrices $Q$ and $R$ are chosen to be

$$Q=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \quad R=\begin{bmatrix}100 & 0\\ 0 & 100\end{bmatrix} \tag{44}$$

Now, we consider the three cases of traffic phenomena which might occur frequently in an urban traffic area.

*Case* 1

The sum of the input queue lies between the minimum of the sum of the output flow in each direction and its maximum. It can be easily shown that the minimum of the sum of the output flows and the maximum are

$$\min\,[S_1(k, x_1(k), q_1(k)), S_2(k, x_2(k), q_2(k))]$$

and

$$\max\,[S_1(k, x_1(k), q_1(k)), S_2(k, x_2(k), q_2(k))]$$

respectively.

*Case* 2

The sum of the input queues exceeds the maximum of the sum of the output flow in each direction.

*Case* 3

The input queues in each direction vary with the elapse of time so that the oversaturatedness and undersaturatedness of the traffic flow can exist concurrently during the optimization interval.

Each of the above three cases is discussed below.

*Case* 1

In this case, we regard two traffic flows as the oversaturated and undersaturated flow, respectively. The initial conditions and parameter values are shown in Table 1. The dual cost $\phi$ and primal cost $J$ were

$$J=0{\cdot}82\times 10^3 \quad \text{and} \quad \phi=0{\cdot}82\times 10^3$$

The optimal state trajectories and control sequences are given in Fig. 6 and Table 2, respectively.

| $i$ | $S_{it}$ | $x_{i,\max}$ | $x_i(0)$ | $q_i(k)$ $k=0,1,2,3$ | $x_{ic}$ | $K_{ic}$ |
|---|---|---|---|---|---|---|
| 1 | 45 | 70 | 24 | 25 | 40 | 2·5 |
| 2 | 40 | 60 | 12 | 10 | 36 | 2·5 |

Table 1. The initial conditions and parameter values of Case 1 in Example 1.

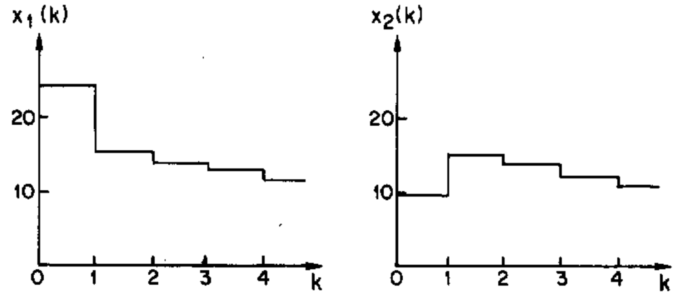


Figure 6. The optimal state trajectories for the oversaturated intersection of Case 1 in Example 1.

| $k$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $u_1(k)$ | 0·799 | 0·669 | 0·638 | 0·643 |
| $u_2(k)$ | 0·201 | 0·331 | 0·362 | 0·357 |

Table 2. The optimal control sequences of Case 1 in Example 1.

*Case* 2

The initial conditions and parameter values are given in Table 3, and the optimal state trajectories and control sequences are shown in Fig. 7 and Table 4, respectively. At the optimum, the dual cost $\phi$ and primal cost $J$ were

$$J=0{\cdot}235\times 10^4 \quad \text{and} \quad \phi=0{\cdot}235\times 10^4$$

*Case* 3

We regard the input queues as time-varying ones as shown in Table 5. The dual cost $\phi$ and primal cost $J$ were

$$J=0{\cdot}138\times 10^4 \quad \text{and} \quad \phi=0{\cdot}138\times 10^4$$

The optimal state trajectories are given in Fig. 8 and the optimal control sequences are given in Table 6.

*Example* 2

Consider a model for the interconnected intersection in Fig. 3. Let $x_i(k)$ and $u_i(k)$ be the state and the control variable, respectively, corresponding to the intersection $i$. The time delay between junction (A) and (B) is one cycle time.

| $i$ | $S_{it}$ | $x_{i,\max}$ | $x_i(0)$ | $q_i(k)$ $k=0,1,2,3$ | $X_{ic}$ | $K_i$ |
|---|---|---|---|---|---|---|
| 1 | 50 | 70 | 22 | 25 | 50 | 2·5 |
| 2 | 40 | 60 | 20 | 20 | 40 | 2·5 |

Table 3. The initial conditions and parameter values of Case 2 in Example 1.

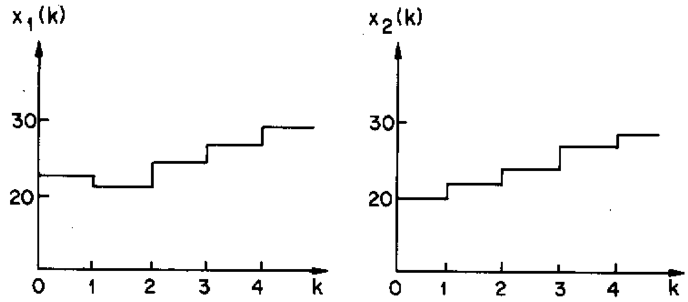


Figure 7. The optimal state trajectories for the oversaturated intersection of Case 2 in Example 1.

| $k$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $u_1(k)$ | 0·579 | 0·518 | 0·524 | 0·519 |
| $u_2(k)$ | 0·421 | 0·482 | 0·476 | 0·481 |

Table 4. The optimal control sequences of Case 2 in Example 1.

Then state equations for the traffic network are described by eqn. (11). In eqn. (11), $S_i(k, x_i(k), q_i(k))$ is given by

$$S_i(k, x_i(k), q_i(k)) = S_{it}\left[1-\exp\left\{-K_i\frac{(x_i(k)+q_i(k))}{X_{ic}}\right\}\right], \quad i=1,2,3,4 \tag{45}$$

Here, it is assumed that state and control constraints are given by

$$0{\cdot}2 \leqslant u_i(k) \leqslant 0{\cdot}8 \tag{46}$$

$$0 \leqslant x_i(k) \leqslant x_{i,\max}, \quad i=1,2,3,4 \tag{47}$$

$$u_1(k)+u_2(k)=1 \tag{48}$$

and

$$u_3(k)+u_4(k)=1 \tag{49}$$

Further, the cycle time $C$ and loss time $l$ are assumed to be 60 s and zero, respectively. Here, the initial conditions and parameter values are given in Table 7.

| $i$ | $S_{it}$ | $x_{i,\max}$ | $x_i(0)$ | $q_1(0)$ | $q_2(1)$ | $q_3(2)$ | $q_4(3)$ | $X_{ic}$ | $K_i$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 50 | 70 | 22 | 25 | 25 | 15 | 15 | 45 | 2·5 |
| 2 | 40 | 60 | 20 | 20 | 20 | 12 | 12 | 34 | 2·5 |

Table 5. The initial conditions and parameter values of Case 3 in Example 1.

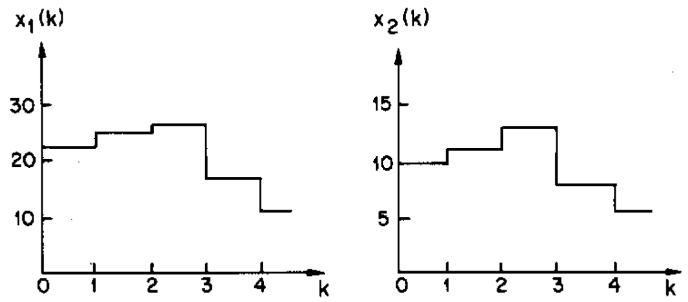


Figure 8. The optimal state trajectories for the oversaturated and/or undersaturated intersection of Case 3 in Example 1.

| $k$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $u_1(k)$ | 0·569 | 0·506 | 0·503 | 0·490 |
| $u_2(k)$ | 0·431 | 0·494 | 0·497 | 0·510 |

Table 6. The optimal control sequences of Case 3 in Example 1.

The optimal control problem is the following. Minimize $J$ given by eqn. (15) under the constraints in eqns. (11) and (45)–(49). In eqn. (15), the weighting matrices $Q$ and $R$ are chosen to be

$$Q=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix} \quad \text{and} \quad R=\begin{bmatrix}100&0&0&0\\0&100&0&0\\0&0&100&0\\0&0&0&100\end{bmatrix} \tag{50}$$

The time interval $k_f$ is taken as $k_f=4$.

The optimal control sequences were calculated for this problem using the prediction-type two-level optimization method in § 3. The dual cost $\phi$ and primal cost $J$ at the optimum were

$$J=0{\cdot}478\times10^4 \quad \text{and} \quad \phi=0{\cdot}478\times10^4$$

The optimal state trajectories are shown in Fig. 9, and the optimal control sequences in Table 8.

| $i$ | $S_{it}$ | $x_{i,\max}$ | $x_i(0)$ | $q_i(k)$ $k=0, 1, 2, 3$ | $x_{io}$ | $K_i$ |
|---|---|---|---|---|---|---|
| 1 | 50 | 70 | 22 | 25 | 49 | 2·5 |
| 2 | 40 | 60 | 20 | 20 | 38 | 2·5 |
| 3 | 50 | 70 | 22 | internal flow $q_3(1)=24$ | 46 | 2·5 |
| 4 | 40 | 60 | 20 | 20 | 36 | 2·5 |

Table 7. The initial conditions and parameters in Example 2.

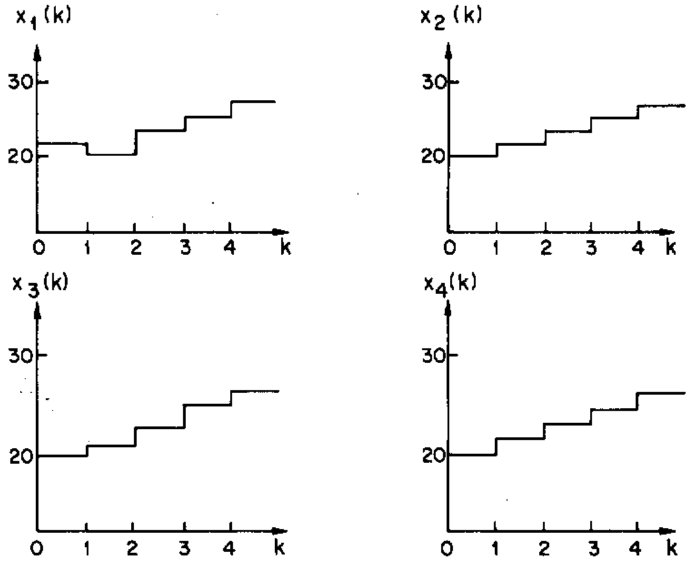


Figure 9. The optimal state trajectories for the interconnected intersection in Example 2.

| $k$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $u_1(k)$ | 0·579 | 0·520 | 0·525 | 0·522 |
| $u_2(k)$ | 0·421 | 0·480 | 0·475 | 0·478 |
| $u_3(k)$ | 0·516 | 0·584 | 0·456 | 0·482 |
| $u_4(k)$ | 0·484 | 0·416 | 0·544 | 0·518 |

Table 8. The optimal control sequences in Example 2.

*Remark* 4

In Fig. 9, the queue lengths $x_i(k)$ $(i=1, 2, 3, 4)$ increase even though the optimal control strategies were applied. From this, we know that the minimization of the quadratic cost functional $J$ in eqn. (15) does not guarantee full utilization of the output flow rate because the cost functional $J$ will tend to penalize the excessive increase of the queue length in one direction. And, also, these phenomena occur when the sum of the input queues is greater than that of the output flows for a given intersection.

## 5. Conclusions

The control problem of traffic networks was investigated by suggesting a new non-linear discrete-time model for traffic networks to describe the traffic phenomena such as oversaturated and undersaturated traffic flows. Then the prediction-type two-level hierarchical optimization method of Hirvonen and Hakkala together with a decomposition technique were utilized to solve the problem. In our model the output flow rate was modelled as an exponential function of the sum of the state queue and input queue.

## Appendix

*Brief discussion of the model of Lim* et al.

It will be shown that the model of Lim *et al.* is applicable only to the oversaturated traffic networks by investigating the existence of a solution for their optimal control problem. For simplicity, consider a one-way no-turn intersection network in Fig. 1. Then the optimal control problem of Lim *et al.* (1981) can be formulated as follows. Find the control sequences $u(0)$, $u(1)$ and $u(2)$ while minimizing the cost function $J$ given by

$$J=\tfrac{1}{2}\sum_{i=1}^{2}\sum_{k=0}^{2}\{x_i(k+1)+u_i(k)\} \tag{A 1}$$

subject to

$$x_i(k+1)=x_i(k)-S_{it}u_i(k)+q_i(k) \tag{A 2}$$

$$S_{it}u_i(k)\leqslant x_i(k)\leqslant x_{i,\max} \tag{A 3}$$

$$0\leqslant u_i(k)\leqslant 1 \tag{A 4}$$

$$u_1(k)+u_2(k)=1 \tag{A 5}$$

Now, for the above problem, we will consider the case of undersaturated traffic behaviour.

| $i$ | $S_{it}$ | $x_{i,\max}$ | $x_i(0)$ | $q_i(k)$ $k=0, 1, 2$ |
|---|---|---|---|---|
| 1 | 100 | 100 | 50 | 30 |
| 2 | 80 | 80 | 40 | 30 |

Table 9. The initial conditions and parameter values in Appendix.

Consider the initial conditions and parameter values in Table 9. Then, from eqns. (A 3) and (A 5), we can obtain that at $k=0$

$$u_1(0)\leqslant\frac{x_1(0)}{S_{1t}}=0{\cdot}5 \tag{A 6}$$

and

$$u_2(0)=1-u_1(0)\leqslant\frac{x_2(0)}{S_{2t}}=0{\cdot}5 \tag{A 7}$$

Thus from eqns. (A 6) and (A 7), $u_1(0)$ should be 0·5. With this value of $u_1(0)$, obtain $x_1(1)$ and $x_2(1)$ from eqn. (A 5). Then

$$x_1(1)=x_2(1)=30 \tag{A 8}$$

Now, from eqn. (A 4), we can obtain the inequality given by

$$0{\cdot}626 \leqslant u_1(0) \leqslant 0{\cdot}3 \tag{A 9}$$

From eqn. (A 9), we know that there is no control sequence $u_1(0)$, $u_1(1)$ and $u_1(2)$ which satisfies the constraints in eqns. (A 4)–(A 7).

We have shown that the model of Lim *et al.* cannot be applied to the undersaturated traffic networks.

## References


Burhardt, K. K., and Kulikowski, R., 1970, *Bull. Acad. pol. Sci. Sér. tech.*, **18**, 573.
Edie, L., 1961, *Ops Res.*, **12**, 66.
Gazis, D. C., 1964, *Ops Res.*, **12**, 815.
Greenberg, H., 1959, *Ops Res.*, **7**, 79.
Hestenes, M. R., 1969, *J. optim. Theory Applic.*, **4**, 303.
Hirvonen, J., and Hakkala, L., 1979, *Int. J. Systems Sci.*, **10**, 1311.
Lemieux, C., 1978, *Int. Symp. on Computer Network Protocols*, *I.F.I.P.*, New York, p. 217.
Lim, J. H., Hwang, S. H., Suh, I. H., and Bien, Z., 1981, *Int. J. Control*, **33**, 727.
Sarachik, P. E., and Özgüner, Ü., 1982, *I.E.E.E. Trans. autom. Control*, **27**, 1233.
Singh, M. G., and Tamura, H., 1974, *Int. J. Control*, **20**, 913.
Tamura, H., 1975, *Automatica*, **11**, 593.
Webster, F. V., 1959, *Road Research Technical Paper*, No. 39 (H.M.S.O.).