Optimisation of Area Traffic Control for Equilibrium Network Flows

Suh-Wen Chiou

A thesis submitted to the University of London for the Degree of Doctor of Philosophy

Centre for Transport Studies
University College London

June 1998

Abstract

A bi-level programming approach is used to tackle an optimisation problem for area traffic control with equilibrium flows. The signal timing plan is defined by common cycle time, and by starts and durations of green. 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 (Vincent, Mitchell and Robertson 1980). 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 are considered and the corresponding derivatives with respect to signal setting variables and link flows are derived (Chiou, 1997b). The gradient projection method is used in deciding feasible descent directions. Optimal choices of step length along feasible directions for changes in offsets are particularly considered. A mixed search procedure is proposed in this thesis, in which better local optima can be found by solving the unconstrained optimisation problem with respect to the offset variables; therefore a global search for particular directions can be carried out in many parts of the feasible region. Two test road networks have been chosen to illustrate the effectiveness of the proposed method and comparisons for the values of the performance index with other conventional methods have been made. Encouraging results on the Allsop and Charlesworth's road network (1977) have shown the robustness and efficiency of the proposed mixed search procedure as values of the performance index were improved in comparison with the non-optimising calculation of mutually consistent TRANSYT-optimal signal timings and equilibrium assignment.

Acknowledgements

I would like to express my truly deep appreciation to Professor Richard E. Allsop for his continuing supervision and invaluable discussions on the area traffic control optimisation with equilibrium network flows throughout the period of this work. Without his meditative suggestions, there would have been little opportunity for completion of this work. In addition, I would also like to thank Dr Benjamin G. Heydecker for his stimulating discussions on the topic of equilibrium traffic assignment and for his never-ending patience in advising me to carry out every aspect of the complicated numerical computations.

Thanks go to the Centre for Transport Studies for every aspect of resources which has been provided during the period of this work.

Thanks also go to the Ministry of Education in Taiwan, Republic of China for providing the four years scholarships.

Furthermore, I would also like to express my gratitude to my dear brothers and sisters in church of London for their consistent pray and encouragement throughout this duration.

Finally, but most importantly, I would like to thank my dear parents and brothers for their understanding and moral support for the past years.

Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background

Following the advent of applications of microprocessors in traffic signal control, such applications have achieved practical effectiveness during the last two decades. Firstly, as far as the scope of traffic control is concerned, it can be classified into the following two ways: isolated junction control and area traffic control. Isolated junction control assumes that the traffic flow arrival patterns are of Poisson-type distribution whilst in area traffic control, the traffic flow arrival distribution follows platooning patterns when the short spacing between adjacent links is considered. Secondly, as far as the way of operating traffic signals is concerned, two methods have been developed: the stage-based control and the phase-based control. Combining the scope in control and the method of control in a road network, we can have the stage-based isolated junction control, the stage-based area traffic control, the phase-based isolated junction control and the phase-based area traffic control. Ways of optimising the isolated junction using the stage-based method which can be expressed as a delay-minimising convex programming have been developed by Allsop (1971). Furthermore, ways of optimising the isolated junction using the phase-based method by adopting similar criteria for the stage-based method have been proposed by Gallivan and Heydecker (1988). On the other hand, ways of optimising area traffic control using the stage-based method, in which performance is expressed as weighted linear combinations of vehicular mean delay and number of stops per unit time have been developed as a practical traffic study tool for urban networks by Robertson (1969). Moreover, by applying the traffic model from TRANSYT (Vincent, Mitchell and Robertson, 1980) but taking account of a more flexible form of control in terms of the phase-based control variables, Wong (1995) proposed a phase-based optimisation approach for the area traffic control problem.

However, it may be appropriate not just to regard the setting of signal timings as simply an effective tool in alleviating congested traffic but it may also play a more flexible role to influence the traffic movements in a preferable way by which the total travel cost incurred by traffic throughout the whole network may be reduced. As

Allsop (1974) noted, ' *when all or part of the network is subject to traffic control, the relationships between travel cost and traffic flow on some or all of the links in the network depend on the control parameters, and these can therefore be used to influence the number of journeys made through the network and the routes taken*' .

As a result, the area traffic control problem can be combined with the equilibrium assignment in such a way that the equilibrium flows are treated as functions of signal setting variables. Considering the fact that the equilibrium flows can be treated as the functions of traffic signal settings, a bi-level mathematical programming approach is adopted. In dealing with the hierarchical relationships between area traffic control problem and equilibrium traffic assignment, the upper level problem is the area traffic control problem while the lower level problem represents the equilibrium assignment which is a constraint of the area traffic control problem.

By combining the two problems into a bi-level problem, there are some mathematical difficulties that result. Firstly, in the upper level problem, the objective function for the area traffic control optimisation with equilibrium network flows needs to be specified as an explicit mathematical expression, in which the chosen indicators for the system performance can fully represent the practical travel cost incurred by traffic where the objective function is to minimise the total travel cost. Secondly, in the lower level problem, the travel time function of traffic flow and signal setting variables for each link needs to be decided, which corresponds to the average travel cost incurred by the traffic at the end of the link controlled by the signal settings. Furthermore, the relationship between the equilibrium flows and signal setting variables forms a non-linear constraint on the choice of signal setting variables, which needs to be taken into account as the following solution methods being developed for the bi-level problem.

Thirdly, the ways of searching for a locally optimisation for this bi-level problem need to be decided in accordance with the following aspects.

1. For given signal settings, to find a descent direction along which the values of the objective function are consistently decreasing, where the chosen indicators for the system performance are to minimise the total travel cost for all users in the signal-controlled road network.

2. Along the descent direction for given signal settings, the optimal value of the objective function for the bi-level problem can be obtained by any convenient one-dimensional search programming technique.

3. A globally optimal search for the minimal value of the objective function for the bi-level problem can be achieved in this way if the bi-level problem is a convex programme with respect to signal setting variables; otherwise where the bi-level problem is a non-convex programme with respect to the signal setting variables, only a local optimum can be obtained.

4. In the latter case, the feasible search directions can be found in which unconstrained steps can be made to carry the search for better local optima to many parts of the feasible region.

1.2 Objectives

Therefore, the objectives of this thesis in dealing with the area traffic control optimisation with equilibrium network flows represented as a bi-level problem can be identified in the following way.

1. To formulate a bi-level mathematical problem which specifies the area traffic control problem as the upper level problem and user equilibrium traffic assignment as the lower level problem.

2. To define the objective function in the upper level problem in terms of convenient mathematical expressions with respect to signal setting variables and to identify the link travel time function for user equilibrium traffic assignment in the lower level problem.

3. To obtain the derivatives for the indicators of the system performance for the bi-level problem and carry out a sensitivity analysis, in which the consequential changes on the equilibrium flows caused by changes of the signal setting variables are estimated.

4. To develop solution methods for this bi-level problem by using these derivatives

for the indicators of the system performance with respect to the signal setting variables, in which the descent direction needs to be obtained so that the values of the objective function are consistently decreasing.

5. To identify the optimal step length along each descent direction so that the locally optimal value of the objective function can be obtained.

6. To develop good search strategies for the better local optima for the bi-level problem of area traffic control optimisation and equilibrium flows so that a global search can be carried out in many parts of the feasible region and the poor local solution can be avoided.

7. To illustrate the effectiveness of the solution method to the bi-level problem of the area traffic control optimisation for equilibrium network flows on test road networks, and to demonstrate the effectiveness of the solution method by comparing the values of the chosen indicators for the system performance with those obtained by other solution methods.

1.3 Structure of the Thesis

The remainder of this thesis can be organised in the following way.

In Chapter 2, a detailed review of work on the combined problem of the area traffic control optimisation and user equilibrium traffic assignment is given. In relation to the bi-level formulation of the area traffic control optimisation with equilibrium network flows, formulations of both the upper level and lower level problems are given in Chapter 3. In Chapter 4, a sensitivity analysis for the bi-level problem with respect to signal setting variables is given, in which detailed expressions for each term in the chosen indicators for the system performance of the bi-level problem are discussed. In Chapter 5, solution methods to this bi-level problem of area traffic control optimisation for network equilibrium flows are given and the corresponding search strategies for the bi-level problem are discussed as well. In Chapter 6, numerical calculations are illustrated on the two example road networks for the proposed solution method and other conventional methods to the bi-level problem of the area traffic control optimisation and equilibrium network flows. The corresponding results are

discussed as well. Conclusion for this thesis and future recommendations are given in Chapter 7. Further mathematical material is provided in three appendices.

Chapter 2 Literature Review

2.0 Introduction

The optimisation problem of area traffic control has been studied over the past few decades. The TRANSYT model proposed by Robertson (1969) has been widely recognized as one of the most useful tools in studying the optimisation of area traffic control. Furthermore, due to recent advanced technologies on the development of microprocessors applied to the operation of signal control, new approaches for studying signal control in urban road networks have been greatly encouraged. For example, a phase-based optimisation approach to the individual signal-controlled junction, which is designed directly in relation to the practical operation of signal groups, has been shown to be successful in achieving better system performance in comparison with previous ways. On the other hand, modelling in relation to users' behaviour in terms of choice of routes has also been widely researched over the past few decades. However, as it has been noted (Allsop 1974; Gartner 1976) a full optimisation process needs to be applied where the two individual problems are both relevant: the area traffic control optimisation with user equilibrium traffic assignment. The way of dealing with such a combined optimisation problem can be actually regarded as an equilibrium network design problem (Marcotte 1983; Heydecker 1986).

In the following sections, a brief review of the development of techniques in dealing with the optimisation problem of area traffic control is given in Section 2.1. An equivalent optimisation of formulation for user equilibrium traffic assignment and the corresponding solution method are reviewed in Section 2.2. The combination of the optimisation problem for area traffic control and user equilibrium assignment is regarded as one equilibrium network design problem for which the problem formulation and the corresponding solution method are reviewed as in Section 2.3. Following the topics discussed in Section 2.3, applications of sensitivity analysis to the equilibrium network flows and area traffic control optimisation problem are reviewed in Section 2.4. Section 2.5 forms a brief conclusion for this Chapter.

2.1 Optimisation of Area Traffic Control

2.1.1 The TRANSYT model

TRANSYT (TRAffic Network StudY Tool) is a stage-based optimisation programme including a numerical model of traffic movement in which platooning of vehicular movements between adjacent junctions is a central feature; the TRANSYT model was invented by Robertson (1969). The performance index in TRANSYT is the sum for all signal-controlled traffic streams of a weighted linear combination of estimated rate of delay and number of stops per unit time which are each evaluated by the numerical model of traffic movement and a simple analytical expression. In TRANSYT, there are two main modules: the traffic model and the signal optimiser; the signal optimiser is the optimisation procedure adjusting signal control variables by repeated evaluation of the traffic model so that the performance index is approximately minimised. The aim of the TRANSYT programme is to find good signal settings for a road network under coordinated control by adjusting the signal settings repeatedly in such a way that only changes in settings that reduce the performance index are adopted. A signal-controlled traffic stream is the basic element for the calculation of signal settings, which either can represent traffic in one or more lanes leading to a junction and used by vehicles in taking a particular movement or can represent for one or more lanes leading to a junction and used by vehicles in taking different movements in such a way that each of the movements shares a lane with one of the other movements. Each traffic stream is in turn represented by its own link in a node-link network with one node for a signal-controlled road junction in relation to the definition of the following performance index.

Let L be the set of links, W be overall cost per average vehicle-hour of delay, K be overall cost per vehicle-stop, and for each link a in L : let D_a be the delay rate in vehicles, S_a be the number of vehicle-stops per hour, and w_a , k_a be respectively link-specific weighting factors for delay rate and number of stops per unit time.
Then the performance index used in TRANSYT is given as

$$\sum_{a \in L} (W\, w_a\, D_a + K\, k_a\, S_a) \tag{2.1}$$

In the performance index, both the delay rate and number of stops per unit time are

classified into two components according to the assumption of periodic traffic movement on a cycle by cycle basis and additional random variations resulting from variations in vehicle movements from cycle to cycle: the uniform component of delay rate and of number of stops per unit time, which results from the assumption of identical vehicle arrivals cycle after cycle, and the random plus oversaturation component of delay rate and of number of stops, which results from the random variations in vehicle arrivals and assumption of steadily increasing oversaturation queues when the mean link arrival flow exceeds its capacity.

In calculations of the uniform component for any one link, the common cycle time for the whole road network can be divided into a number of equal intervals called 'steps'. Robertson employed the following three types of flow pattern as functions of time in the cycle in describing the movement of traffic on a link in terms of cyclic flow profiles:

(i). the IN profile is the pattern of traffic at the downstream end of the link that would arrive if the traffic were not stopped by the signals.

(ii). the GO profile is the pattern of traffic that would leave the stop line if there was enough traffic to saturate the green.

(iii). the OUT profile is the output pattern of traffic that leaves the link, and is derived from the IN and GO profiles.

For the purpose of deriving the IN profile at the downstream link,which results from the OUT profiles from relevant upstream links, the following recursive expression is used.

$$\bar{q}^{k+t} = f F q^{k} + (1 - F) \bar{q}^{k+t-1} \tag{2.2}$$

where $\bar{q}^{k}$ is the flow in the step k of the IN profile, q^{k} is the flow in step k of the OUT profile from a relevant upstream link, f is the proportion of the OUT flow from that link which enters the link being considered, and t is 0.8 times the mean undelayed travel time[1] (measured in steps) over the distance for which dispersion is being calculated, and

$F = \dfrac{1}{1 + 0.35\, t}$ is a smoothing factor.

[1]Undelayed travel time is the typical travel time of a vehicle which travels along the link under prevailing conditions but without being delayed by the downstream signal.

In addition, as for calculations for the random plus oversaturation of delay rate, D^{r+o} and of the number of stops per unit time, S^{r+o}, simple analytical equations are used (Vincent, Mitchell and Robertson 1980, pp7-8).

$$D^{r+o} = \frac{T}{4}\left[\left((q - \mu)^2 + \frac{4q}{T}\right)^{0.5} + (q - \mu)\right]$$

$$S^{r+o} = 2q\left(1 - \exp\left(\frac{-D^{r+o}}{2\bar{c}q}\right)\right) \tag{2.3}$$

where $\bar{c}$ is the common cycle time of the network, q is the average arrival flow, μ is the absolute capacity at the downstream stop line as estimated by the GO pattern, and T is the period over which the estimate is made.

As for the optimisation procedure in searching for the approximately optimal signal settings, a 'hill-climbing' search process is used. As far as the hill-climbing process is concerned, two kinds of signal setting variables are taken into account: the offset, which affects the coordinations between junctions, and the start times of individual stages at each junction, thus taking a stage-based approach. The hill-climbing search process can be stated in the following way. First, TRANSYT calculates the value of performance index for an initial set of signal settings for the road network in which all constraints are satisfied for considerations of safety. Next, one of the signal control variables is changed by a predetermined number of steps and the corresponding value of the performance index is evaluated. If the evaluated value of performance index decreases, which shows that the system performance is improved as the signal setting variable changes in the direction by the predetermined number of steps, then the signal setting variable is changed again in the same direction by the same number of steps until a minimal value of performance index is achieved. On the other hand, if the evaluated value of the performance index does not decrease, which means that the system performance is not improved as the signal setting variable changes in that direction, then the same variable is changed again in the opposite direction instead by the same number of steps until a minimal value of performance index is achieved. This process continues for all signal setting variables in the road network in turn and a sequence of numbers of steps by which the different variables are changed can be determined in advance.

In the TRANSYT programme, the traffic movements are represented by a numerical simulation model and no explicit mathematical expression is used. Since the performance index is a non-convex function, only a local minimum can be obtained during the hill-climbing search process. In order to reduce the possibility of missing out a good local solution, the TRANSYT programme includes a mixture of both small and large numbers of steps in searching by changing the signal setting variables. For example, TRANSYT uses various numbers of steps in searching by changing the offset variable since there is no limitation over the search region of the offset variable. In addition to changing offsets by various numbers of steps, the TRANSYT programme considers a finer search in changes one of step for the start time of each stage at each junction. As a result, the hill-climbing search process in the signal optimiser of TRANSYT employs a flexible way of adjusting each kind of signal setting variable and by this means an approximation to the global minimum may be found.

In relation to another signal setting variable at the network level, the common cycle time, which is specified as taking a common value for all junctions of the road network (or a submultiple of the common value for some particular junctions), the TRANSYT programme has yet to take account of it as a decision variable in the optimiser. However, the TRANSYT programme evaluates a wide variety of common cycle times on the basis of green times calculated for individual junctions and arbitrary offsets. For example, TRANSYT 8 (Vincent *et al.*, 1980) provides the information for a wide choice of cycle times, including the corresponding highest degree of saturation and green-time distribution at each junction, the values of performance index for all relevant links, and also options for double cycling at some particular junctions.

2.1.2 Phase-based Optimisation

Another approach to an area traffic control optimisation problem is the phase-based optimisation approach which takes full advantages of higher flexibility provided by recent development on microprocessor controller technology. Optimisation techniques for calculating signal settings using the phase-based optimisation approach deal with the signal control variables in relation to sets or groups of signals: those which are switched simultaneously. Mutually compatible groups can receive green concurrently; on the other hand, for mutually incompatible groups there are minimum intergreen times between them.

In this approach the orders of occurrence of stages and the structures of the interstages can be decided automatically during the optimisation process. It has been reported (Silcock and Sang 1989; Silcock 1992,1997) that considerable advantages can be achieved in the system performance by means of the phase-based approach in comparison with the system performance achieved by means of the stage-based approach when applied to individual signal-controlled junctions. In the phase-based approach the signal control variables for a road network: a sequence of stages, durations of stages, structures of interstages, offsets and even common cycle times, can be decided within the optimisation process. Detailed examples and discussions for the advantages provided by this phase-based optimisation approach at individual junctions are given by Heydecker and Dudgeon (1987), Gallivan and Heydecker (1988) and Allsop (1992).

Furthermore, optimisation techniques using the phase-based approach to the area traffic control problem have been investigated over recent years. Firstly, Heydecker (1996) proposed a complementary approach to the area traffic control optimisation problem, in which the results obtained for individual signal-controlled junctions using the phase-based approach are combined with the optimisation of the road network where the network level decision variables such as the offsets and common cycle time are to be taken into account. Such a complementary approach to the area traffic control optimisation problem can be decomposed into two separate levels when carried out in practice: firstly, the individual junction level, in which the structures of interstage and the order of stage occurrences for each individual junction can be optimised according to various choices of criteria; secondly, the network level, in which a full optimisation can be carried out for the system performance of the road network with respect to the durations of stages and the offset variables while the common cycle time, interstage structures and stage sequences are supposed to be fixed at this network level.

At the individual junction level, as mentioned above, the interstage structures and stage sequences for each individual junction can be optimised for various criteria: the minimum cycle time, the maximum capacity and the minimum rate of delay. Choosing different criteria in deciding the interstage structures and stage sequences at the individual junction level and in optimising at network level will give rise to some discrepancies between them. In general, the interstage structures and stage sequences optimised by the criteria of minimum cycle time and maximum capacity will achieve equally appropriate effects at the

individual junction and network levels. However, because only the critical movement of each signal group is taken into account, the choice of the criterion of the minimum cycle time or the maximum capacity will give rise to some degree of indeterminacy in signal settings. On the other hand, the interstage structures and stage sequences optimised by the criterion of minimum delay rate will not normally achieve equally appropriate effects at the individual junction and network levels, which is due to the nature of the assumptions about vehicle arrivals, but the corresponding signal settings are fully determined by the optimisation process. After the individual junction decision variables, i.e. the interstage structures and stage sequences, have been optimised, the common cycle time for a road network can be decided in the following way. First, the TRANSYT programme is run with the interstage structures and stage sequences provided by optimising at the individual junction level, and the common cycle time is chosen as the one with the best value of system performance within the possible range of cycle times, which can be identified by the cycle time selection results in the TRANSYT output. The interstage structures and stage sequences are then optimised again with the chosen common cycle time for each individual junction. If the resulting interstage structures and stage sequences are mutually consistent with the common cycle time, then the optimisation of the individual junction level is complete; otherwise, more iterations will be needed for the determination of the common cycle time, interstage structures and stage sequences.

At the network level, as it is mentioned above, a full optimisation is then carried out for the remaining decision variables, i.e. the durations of stages and offsets, while the common cycle time and the individual junction decision variables remain fixed. If the resulting performance of each junction appears satisfactory, the optimisation at the network level is then regarded as complete; otherwise, further adjustments of signal setting variables are made at the individual junction level and the whole optimisation process is performed again.

Secondly, Wong (1996) proposed a group-based (i.e. phase-based) optimisation approach to an area traffic control problem. The group-based optimisation approach formulates a set of nonlinear programmes in terms of the phase-based control variables explicitly, the common cycle time and the starts and durations of green, as a mathematical optimisation programme subject to the constraints by taking account of safety reasons. The objective function of the mathematical programme is defined as a weighted linear combination of the

delay rate and number of stops per unit time as in TRANSYT. The mathematical programme can be solved by an integer programming method which is mainly influenced by the derivatives of the value of the performance index with respect to the phase-based control variables. Wong (1995) showed that these derivatives of the performance index as estimated by approximate expressions were in a good agreement with results of numerical differentiation in the case of a test road network. In relation to the use of the phase-based optimisation approach to formulating an area traffic control problem, the following formulations are introduced.

Let $\bar{c}$ be the common cycle time, and N_J be the number of signal-controlled junctions in the road network. For each signal-controlled junction m, $1 \leq m \leq N_J$, let N_{pm} be the number of signal groups and θ_{jm}, ϕ_{jm} be respectively the start and duration of green for signal group j, $1 \leq j \leq N_{pm}$; furthermore let $\bar{\psi} = [\, \psi_m,\ 1 \leq m \leq N_J \,]$ and $\psi_m = [\, \theta_{jm},\ \phi_{jm};\ 1 \leq j \leq N_{pm} \,]$. As far as the constraints for the signal control variables are concerned, we let $A = [\, A_m,\ 1 \leq m \leq N_J \,]$ be a diagonal supermatrix with matrices A_m, $1 \leq m \leq N_J$ which are the coefficient matrix with respect to the signal control variables, and a_0, b_0 be respectively coefficient and constant vectors for the common cycle time $\bar{c}$, and $a = [\, a_m,\ 1 \leq m \leq N_J \,]$, $b = [\, b_m,\ 1 \leq m \leq N_J \,]$ be respectively the coefficient and constant vectors for signal control variables.

Then the area traffic control problem with *unspecified common cycle time* can be formulated as to

$$\underset{\psi}{Minimise}\ \ P_0(\psi)$$

$$subject\ \ to\ \ \begin{bmatrix} a_0 & 0 \\ a & A \end{bmatrix} \psi^T \leq \begin{bmatrix} b_0 \\ b \end{bmatrix} \tag{2.4}$$

where $\psi = [\, \bar{c},\ \bar{\psi} \,]$ and the superscript T is the matrix transpose operator.

Similarly, the area traffic control problem with *specified common cycle time* can be formulated as to

$$\begin{aligned} &\underset{\bar{\psi}}{Minimise} \quad P_1(\bar{\psi}) \\ &subject \quad to \\ &\qquad A\ \bar{\psi}^T \leq \bar{b} \end{aligned} \tag{2.5}$$

where $\bar{b} = b - \bar{c}\ a$

The objective functions $P_0(\psi)$ and $P_1(\bar{\psi})$ in (2.4)-(2.5) take the following form as shown in expression (2.1)

$$\sum_{a \in L} W\ w_a\ D_a + K\ k_a\ S_a$$

where $D_a = D_a^U + D_a^{r+o}$ is the delay rate on link a, the sum of the uniform and random plus oversaturation components of delay rate on the link, and $S_a = S_a^U + S_a^{r+o}$ is the number of stops per unit time on link a, the sum of the uniform and random plus oversaturation components of number of stops per unit time on the link.

In relation to the solution methods to the optimisation problems in (2.4)-(2.5), Wong employed an integer programming method to solve the optimisation problems, which can be divided into two components: the integer linear sub-problem which aims to find a feasible integer direction on the basis of the values of derivatives of performance index with respect to the phase-based signal control variables, and the sub-problem which aims to find the maximum of move size along the feasible direction.

For any one set of initial signal settings, $\psi^0 = [\ \bar{c}^0\ ,\ \bar{\psi}^0\ ]$, the integer linear sub-problem for the area traffic control optimisation problem (2.5) with *specified common cycle time* is described as follows.

<Integer linear sub-problem>

Find integer direction d in the $\bar{\psi}$ space to

$$\underset{d}{Minimise} \quad \nabla P_1(\bar{\psi}^0)\, d^T$$

$$subject\ to \quad A\, d^T \le \bar{b}' \qquad (2.6)$$

$$-1 \le d \le 1$$

where $\nabla P_1(\bar{\psi}^0)$ is the gradient evaluated at $\bar{\psi}^0$ of P_1 , which has been reported in Wong (1995), and $\bar{b}' = \bar{b} - A\,(\bar{\psi}^0)^T$.

<Sub-problem>

Find integer α to

$$\underset{\alpha}{Minimise} \quad P_1(\bar{\psi}^0 + \alpha\, d)$$

$$subject\ to \quad 0 \le \alpha \le \alpha_{max} \qquad (2.7)$$

where $\alpha_{max} = \underset{i}{Minimise} \left\{ \alpha_i = \dfrac{\bar{b}_i'}{\sum_{j=1}^{M} A_{ij}\, d_j} ,\ i \in \left\{ i\ ;\ \sum_{j=1}^{M} A_{ij}\, d_j > 0 \right\} \right\}$

$$and \quad M = 2\sum_{m=1}^{N_J} N_{pm}$$

For any one set of initial signal settings, $\psi^0 = [\ \bar{c}^0\ ,\ \bar{\psi}^0\ ]$, and given changes $\Delta\, \bar{c}$ the integer linear sub-problem for the area traffic control optimisation problem (2.4) with *unspecified common cycle time* is described as follows.

<Integer linear sub-problem>

Find direction d in the $\bar{\psi}$ space to

$$\underset{d}{Minimise} \quad \bar{\nabla} P_0(\psi^0)\, d^T$$

$$subject\ to \quad A\, d^T \le b' \qquad (2.8)$$

$$-1 \le d \le 1$$

where $\bar{\nabla} P_0(\psi^0)$ is $\nabla P_0(\psi^0)$, which is the gradient evaluated at ψ^0 of P_0 and has

been reported in Wong (1995), with the first element $\frac{\partial P_0}{\partial \bar{c}}$ omitted, and

$$b' = b - A\,(\bar{\psi}^0)^T - a\,(\bar{c}^0 + \Delta\,\bar{c})$$

<Sub-problem>

Find integer α to

$$\begin{aligned} &\underset{\alpha}{Minimise} && P_0(\psi^0 + \alpha\,\bar{d}) \\ &subject\ \ to && 0 \le \alpha \le \alpha_{max} \\ &where && \bar{d} = [\,\Delta\,\bar{c}\,,\,d\,] \end{aligned} \tag{2.9}$$

$$\text{let } \alpha_0 = \begin{cases} \dfrac{\bar{c}_{max} - \bar{c}^0}{\Delta\,\bar{c}}\,, & if\ \Delta\,\bar{c} > 0 \\[2ex] \dfrac{\bar{c}_{min} - \bar{c}^0}{\Delta\,\bar{c}}\,, & if\ \Delta\,\bar{c} < 0 \\[2ex] \infty & if\ \Delta\,\bar{c} = 0 \end{cases}$$

where $\bar{c}_{min}$, $\bar{c}_{max}$ are respectively the minimum and maximum cycle time.

$$\alpha_{max} = Minimise \left\{ \alpha_0\,,\ \underset{i}{Minimise} \left\{ \alpha_i = \frac{b_i - a_i\,\bar{c}^0 - \sum_{j=1}^{M} A_{ij}\,\bar{\psi}_j^0}{a_i\,\Delta\,\bar{c} + \sum_{j=1}^{M} A_{ij}\,d_j} \right. \right.$$

$$\left. \left. i \in \left\{ i\ ;\ a_i\,\Delta\,\bar{c} + \sum_{j=1}^{M} A_{ij}\,\bar{d}_j > 0 \right\} \right\} \right\}$$

$$where \qquad M = 2 \sum_{m=1}^{N_J} N_{pm}$$

As for the characteristics of the solutions to the problems (2.4)-(2.5), only local optima can be found due to the non-convexity of the objective function. Wong presented an optimisation process for implementing the phase-based signal settings by combining the

hill-climbing search process in TRANSYT with the integer programming method. The optimisation process consists of two classes of optimisation steps: a network-wide optimisation step, changes all signal control variables simultaneously over the road network aiming to approach the neighbourhood of a good local optimum, and a junction-based optimisation step, which changes the signal control variables junction by junction aiming to fine-tune the timings at each junction. This optimisation process for the area traffic control problem for any given set of initial signal settings can be summarized in the following steps.

Steps 1&2: Use a combination of large and small step sizes for the adjustments of the offset variables, which is the same procedure as implemented in the hill-climbing search process as in TRANSYT, such that a good initial solution can be obtained.

Step 3: Adjust the common cycle time and green times by changing the signal control variables simultaneously by means of the integer programming method, which is implemented by the network-wide optimisation step.

Steps 4&5: Repeat Steps 1&2 to find another better local optimum.

Step 6: Reduce the move size into one unit interval for the adjustment of offset variables to locate the neighbourhood of the local minimum.

Step 7: Reallocate green times by implementing the integer programming method on the junction-based step and changing the green times junction by junction.

Step 8: Repeat Step 6 for final checking in the search for the near-global optimum.

2.2 User Equilibrium Traffic Assignment

A user equilibrium traffic assignment problem is to find the traffic flows and the corresponding travel times for all links in a given road network by assigning the given origin-destination trip rates for the period to be analyzed. The resulting assignment represents conditions in which each road user chooses the shortest route with the minimum travel time from his origin to destination and no one can improve his travel time by unilaterally changing routes for alternatives between any pair of specified origin and

destination. As it has been enunciated by Wardrop (1952), a user equilibrium traffic assignment can be described as follows.

" The journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route."

For a given number of origin-destination trip rates over a period of analysis, the resulting equilibrium traffic flows and the corresponding travel times can be calculated by solving an equivalent mathematical programme. Beckmann, McGuire and Winsten (1956) first proposed an equivalent user equilibrium mathematical programme, in which the solutions satisfy the conditions required by the user equilibrium traffic assignment. Furthermore, a unique solution to the user equilibrium programme can be obtained when some strict mathematical conditions are satisfied by the link travel time function. In relation to the equivalent user equilibrium mathematical programme, the mathematical formulation will be given first, and the corresponding characteristics of the solutions will be discussed later.

2.2.1 Equivalent mathematical programme to user equilibrium assignment

2.2.1.1 Beckmann transformation

The formulation of the user equilibrium programme according to Beckmann transformation is as follows.

Let $\boldsymbol{L}$ be the set of links, $\boldsymbol{W}$ be the set of origin-destination pairs, and $\boldsymbol{P}_w$ be the set of paths between each origin-destination pair w , $\forall\, w \in \boldsymbol{W}$.

Let $\delta = [\, \delta_{ap} \,]$ be the link/path incidence matrix where $\delta_{ap} = 1$ if link a is in path p , and $\delta_{ap} = 0$ otherwise, $\forall\, a \in \boldsymbol{L}$, $\forall\, p \in \boldsymbol{P}_w$, $\forall\, w \in \boldsymbol{W}$, and $\Delta = [\, \Delta_{wp} \,]$ be the origin-destination/path incidence matrix where $\Delta_{wp} = 1$ if path p connects origin-destination pair w , and $\Delta_{wp} = 0$ otherwise, $\forall\, p \in \boldsymbol{P}_w$, $\forall\, w \in \boldsymbol{W}$.

Let $\hat{\boldsymbol{D}} = [\, \hat{D}_w \,]$ be the row vector of travel demands between each origin-destination pair

w in $\boldsymbol{W}$, $\boldsymbol{f} = [\, f_p \,]$ be the row vector of all path flows, and $\boldsymbol{q} = [\, q_a \,]$ be the row vector of all link flows.

Let $\boldsymbol{c}(\boldsymbol{q}) = [\, c_a(q_a) \,]$ be the row vector of link travel times, in which $c_a(q_a)$ denotes the link travel time depends only on the flow on that link, i.e., separable relationship between the link travel time and corresponding flow is assumed, and $\boldsymbol{C}(\boldsymbol{q}) = [\, C_p(\boldsymbol{q}) \,]$ be the row vector of path travel times.

Then the equivalent programme is to

$$\underset{\boldsymbol{q}}{Minimise} \quad Z(\boldsymbol{q}) = \sum_{a \in \boldsymbol{L}} \int_0^{q_a} c_a(w)\, dw$$

$$\text{subject to} \qquad \sum_{p \in \boldsymbol{P}_w} f_p = \hat{D}_w \,, \quad \forall\, w \in \boldsymbol{W} \tag{2.10}$$

$$\sum_{w \in \boldsymbol{W}} \sum_{p \in \boldsymbol{P}_w} f_p\, \delta_{ap} = q_a \,, \quad \forall\, a \in \boldsymbol{L}$$

$$f_p \geq 0 \,, \quad \forall\, p \in \boldsymbol{P}_w \,, \; \forall\, w \in \boldsymbol{W}$$

Furthermore, the dependence of link flows $\boldsymbol{q}$ on path flows $\boldsymbol{f}$ and the dependence of path travel times $\boldsymbol{C}(\boldsymbol{q})$ on link travel times $\boldsymbol{c}(\boldsymbol{q})$ can be expressed in terms of the incidence matrix as follows.

$$\boldsymbol{q}^T = \delta\, \boldsymbol{f}^T \tag{2.11}$$

$$\boldsymbol{C}(\boldsymbol{q})^T = \delta^T\, \boldsymbol{c}(\boldsymbol{q})^T \tag{2.12}$$

2.2.1.2 Equivalent conditions for user equilibrium programme

As we have mentioned earlier, the resulting traffic flows for the user equilibrium traffic assignment are such that the travel times on all used paths connecting any given origin-destination pair will be equal; furthermore the travel times on all used paths will also be less than or equal to the travel times on any of the unused paths. Therefore, at this point no road user can experience a lower travel time by unilaterally changing routes and hence

the network is in user equilibrium. Following this equilibrium conditions, the first-order necessary conditions for the mathematical programme (2.10) have been shown equivalent to the equilibrium conditions for the user equilibrium traffic assignment (Sheffi, 1985, pp 63-66). The first-order necessary conditions for (2.10) can be given as follows.

$$\left.\begin{aligned} f_p\,(C_p - u_w) &= 0 \\ C_p - u_w &\geq 0 \\ f_p &\geq 0 \end{aligned}\right\} \quad \forall\, p \in \boldsymbol{P}_w\,,\ \forall\, w \in \boldsymbol{W} \tag{2.13}$$

$$\sum_{p \in \boldsymbol{P}_w} f_p = \hat{D}_w \quad , \quad \forall\, w \in \boldsymbol{W} \tag{2.14}$$

where u_w is the minimum path travel time for specified origin-destination pair w, $\forall\, w \in \boldsymbol{W}$.

Expressions (2.13)-(2.14) show that for all used paths carrying traffic flows the corresponding travel times equal the minimum travel time while for any unused path carrying no flow, the corresponding travel time is not less than the minimum travel time. Therefore, that can be expressed in the following statements.

$$\begin{aligned} C_p &= u_w\,, \quad \textit{if} \quad f_p > 0 \\ C_p &\geq u_w\,, \quad \textit{if} \quad f_p = 0 \end{aligned} \quad , \quad \forall\, p \in \boldsymbol{P}_w\,,\ \forall\, w \in \boldsymbol{W} \tag{2.15}$$

2.2.1.3 Uniqueness condition for user equilibrium programme

The uniqueness condition for the user equilibrium programme with respect to link flows shows only one solution that minimises the objection function (2.10) and the solution also satisfies the equivalent conditions in expressions (2.13)-(2.14) if and only if the strict convexity condition of the user equilibrium programme with respect to link flows is satisfied.

In (2.10), the objective function is the sum of the integrals of the link travel time functions with respect to link flows. The separability relationship is assumed for the dependence of the link travel time function on the corresponding link flow; that is, any one link travel time function depends only on the flow on the link and is independent of the flow on other

links. Furthermore, for the reasons of traffic congestion on the road network, it assumes that the link travel time will increase strictly as the corresponding flow increases. In relation to mathematical expressions, both the assumptions of separability and strict increase of link travel time due to traffic congestion can be given below.

$$\frac{d\, c_a(q_a)}{d\, q_b} = 0 \,, \qquad b \neq a \,, \forall\, a \,, b \in \boldsymbol{L}$$
$$\frac{d\, c_a(q_a)}{d\, q_a} > 0 \,, \qquad \forall\, a \in \boldsymbol{L} \tag{2.16}$$

Following the assumption in terms of expression (2.16), it can be seen that the Hessian matrix of the equivalent programme is a diagonal matrix with strictly positive diagonal elements and hence positive definite; therefore the solution to the user equilibrium programme (2.10) (i.e. the solution that satisfies the equivalent conditions (2.13)-(2.14)) is unique.

Following the equivalent user equilibrium programme, a general formulation for the user equilibrium traffic assignment can be formulated in terms of a variational inequality problem by Smith (1979).

2.2.1.4 Variational inequality problem formulation

A general formulation for a user equilibrium traffic assignment problem can be expressed in terms of a variational inequality problem as shown by Smith (1979) and identified by Dafermos (1980).

To find values $q^\star$ of q such that

$$c(q^\star)\,(q - q^\star)^T \geq 0 \tag{2.17}$$

$$\forall\; q \in \Omega = \left\{ q \,;\, q^T = \delta\, f^T \,,\, \Delta f^T = \hat{\boldsymbol{D}}^T \,,\, f \geq \boldsymbol{0} \right\}$$

2.2.2 Solution method to user equilibrium traffic assignment

Following the formulation of equivalent user equilibrium programme in (2.10), a convex combination method based on the Frank-Wolfe algorithm has been widely used to solve the equivalent user equilibrium programme. In relation to the convex combination method, it can be carried out in the following two ways. First, a linear programme is formulated on

the basis of current travel times in deciding a feasible direction along which the value of objective function of (2.10) subject to constraint set can be reduced. Second, a sub-programme is formulated, in which the optimal move size along the feasible direction can be decided and hence the value of the objective function is minimised. In the linear programme, the first derivatives of the objective function with respect to path flows are used to determine the minimum path travel time for specified current traffic flows. This step is no more than a shortest path problem, in which the all-or-nothing traffic assignment is carried out in search for auxiliary flows on the basis of current travel times. In the sub-programme, a one-dimensional optimisation problem is carried out for which an optimal move size can be decided such that the value of objective function in terms of a linear combination of current flows and auxiliary flows is minimised. The linear programme and the corresponding sub-programme for the convex combination method can be carried out alternately in a series of iterative processes until a required criterion is satisfied.

<Linear programme>

Let $q_0^{(n)}$ be current link flows at iteration n and $c^{(n)}$, $C^{(n)}$ be respectively current link and path travel times at iteration n.

Find the shortest path at iteration n between each origin-destination pair w in $\boldsymbol{W}$ based on current traffic flows to

$$\underset{f}{\textit{Minimise}} \quad \sum_{w \in \boldsymbol{W}} \sum_{p \in \boldsymbol{P}_w} C_p^{(n)} f_p^{(n)}$$

$$\textit{subject to} \quad \sum_{p \in \boldsymbol{P}_w} f_p^{(n)} = \hat{D}_w \,, \quad \forall\, w \in \boldsymbol{W} \tag{2.18}$$

$$f_p^{(n)} \geq 0 \quad , \forall\, p \in \boldsymbol{P}_w \,, \forall\, w \in \boldsymbol{W}$$

The linear programme (2.18) is to find the shortest paths with the smallest travel times at iteration n, i.e. $\boldsymbol{u}^{(n)} = \left[u_w^{(n)} \right]$ for which the travel demands of all origin-destination pairs, $\hat{\boldsymbol{D}} = \left[\hat{D}_w \right]$, then can be assigned, which is carried out on the basis of current

flows $q_0^{(n)}$ and the corresponding current travel times $c^{(n)}$ and $C^{(n)}$. In other words, the resulting traffic assignment is no more than a problem of finding the shortest path for given specified travel time and thus it can be performed by the technique of all-or-nothing traffic assignment. The all-or-nothing traffic assignment for programme (2.18) can be described as follows.

For each origin-destination pair w in $\boldsymbol{W}$ and given trip rate over a period of analysis, $\hat{D}_w$, there exists a path h connecting origin-destination pair w at iteration n, such that

$$\left\{ \begin{array}{ll} & f_h^{(n)} = \hat{D}_w \,, \quad if \quad C_h^{(n)} \leq C_p^{(n)} \,, \quad \forall\, p \in \boldsymbol{P}_w \\ and & f_p^{(n)} = 0 \,, \; \forall\, p \neq h \end{array} \right.$$

The corresponding auxiliary link flow $q_1^{(n)}$ can be derived in terms of the link-path incidence matrix.

$$(q_1^{(n)})^T = \delta\ (f^{(n)})^T \tag{2.19}$$

<Sub-programme>

Following the auxiliary flows $q_1^{(n)}$ obtained by expression (2.19), a one-dimensional line search to find the optimal move length $\alpha^{(n)}$ along the descent direction,

$d^{(n)} = q_1^{(n)} - q_0^{(n)}$ which minimises (2.10) is to

$$\begin{array}{ll} \underset{\alpha^{(n)}}{Minimise} & Z(c_0^{(n)} + \alpha^{(n)}\ (q_1^{(n)} - q_0^{(n)})) \\ subject\ to & 0 \leq \alpha^{(n)} \leq 1 \end{array} \tag{2.20}$$

This represents a one-dimensional optimisation problem in searching for the optimal move size along the feasible direction; it is also a one-dimensional step size problem for a linear combination of auxiliary and current flows such that the value of objective function for mixed traffic flows can be minimised.

One of solution methods to a one-dimensional optimisation problem is a Bolzano search-bisection method. When this is applied to (2.20), the optimal move size can be found if the following condition is met.

$$\left.\begin{aligned} &\frac{d\ Z(\alpha^{(n)})}{d\ \alpha^{(n)}} = 0 \\ &c^{(n)}(v^{(n)})\ (q_1^{(n)} - q_0^{(n)})^T = 0 \\ &c^{(n)}(v^{(n)})\ (q_1^{(n)})^T = c^{(n)}(v^{(n)})\ (q_0^{(n)})^T \end{aligned}\right\} \quad 0 \le \alpha^{(n)} \le 1 \qquad (2.21)$$

where $v^{(n)} = q_0^{(n)} + \alpha^{(n)}\ (q_1^{(n)} - q_0^{(n)})$

The (2.21) corresponds to the *'cost equilibration'* as mentioned by Van Vliet (1987). When the sub-programme is solved by using a one-dimensional search process, new traffic flows can be obtained by the linear combination of auxiliary and current flows by means of the optimal move size and then the corresponding travel times can be updated accordingly. In relation to the stopping criterion for the convex combination method when carried out in a sequence of iterative processes by solving alternately the linear programme and sub-programme, it is suggested (Sheffi, 1985, pp119) to stop when the difference in the total origin-destination travel times between successive iterations is not greater than a predetermined threshold value, i.e.,

$$\sum_{w \in W} \frac{|\ u_w^{(n)} - u_w^{(n-1)}|}{u_w^{(n)}} \le \varepsilon \qquad (2.22)$$

where $u_w^{(n)}$ is the minimum travel time between a specified origin-destination pair w in W at iteration n, which can be found by (2.18), and ε, $\varepsilon \ge 0$, is a predetermined value of threshold.

In summary, as referred to Sheffi (1985, pp119-120) the procedure for carrying out the convex combination method to solve an equivalent user equilibrium programme can be stated in the following steps.

Step 0: Initialization.

0.1 Set indicator $n = 0$.

0.2 Perform all-or-nothing traffic assignment based on current travel times,

$c^{(n)}(q_0^{(n)} = \mathbf{0})$ by solving the linear programme (2.18) and yield the corresponding link

flows $q_0^{(n)}$.

0.3 Set indicator $n = 1$.

Step 1: Update.

Set $c^{(n)} = c^{(n)}(q_0^{(n)})$.

Step 2: Direction finding.

Perform all-or-nothing traffic assignment based on $c^{(n)}$ by solving the linear programme

(2.18) and yield the auxiliary link flows $q_1^{(n)}$ by means of (2.19).

Step 3: Line search.

Find $\alpha^{(n)}$ along the descent direction $d^{(n)} = q_1^{(n)} - q_0^{(n)}$ by solving (2.20), which also

can be carried out by means of (2.21).

Step 4: Move.

Set $q_0^{(n+1)} = q_0^{(n)} + \alpha^{(n)} (q_1^{(n)} - q_0^{(n)})$

Step 5: Convergence test.

If condition (2.22) is met then the convex combination procedure is complete and

$q_0^{(n+1)}$ is the equilibrium flow; otherwise set $n = n + 1$ and return to Step 1.

2.3 Equilibrium Network Design Problem

2.3.1 Introduction

In Section 2.2, a user equilibrium traffic assignment has been formulated as an equivalent mathematical programme given by fixed travel demands during a specified time of analysis for a general road network represented in terms of a number of nodes and links. In a general road network, when some or all nodes are subject to signal control, the travel time function for each link needs to be expressed in a form in which the average delay incurred by traffic at the downstream junction can be taken into account. Furthermore, it has been widely recognized (Allsop 1974, Gartner 1976) that the resulting equilibrium flows and the corresponding travel times in a road traffic network are strongly influenced by the operation of signals. Therefore, a revised equivalent user equilibrium programme can be expressed in the following way when the signal setting variables are taken into account in the link travel time functions. Recall the Beckmann transformation (2.10) for an equivalent user equilibrium programme, in which (2.10) can be revised as follows when signal settings are taken into account in the equilibrium flows and the corresponding travel time functions.

For any given signal settings ψ , a revised equivalent mathematical programme to the equilibrium traffic assignment when signal settings are considered is to

$$\underset{q}{Minimise} \quad Z_1(\boldsymbol{q}\ ,\ \boldsymbol{\psi}) = \sum_{a \in L} \int_0^{q_a(\boldsymbol{\psi})} c_a(\boldsymbol{\psi}\ ,\ w)\ dw$$

$$\textit{subject to} \qquad \sum_{p \in P_w} f_p(\boldsymbol{\psi}) = \hat{D}_w\ ,\quad \forall\ w \in \boldsymbol{W} \tag{2.23}$$

$$\sum_{w \in W} \sum_{p \in P_w} f_p(\boldsymbol{\psi})\ \delta_{ap} = q_a(\boldsymbol{\psi})\ ,\quad \forall\ a \in \boldsymbol{L}$$

$$f_p(\boldsymbol{\psi}) \geq 0\ ,\quad \forall\ p \in \boldsymbol{P}_w\ ,\ \forall\ w \in \boldsymbol{W}$$

Recall the generalized formulation given by fixed travel demands, $\hat{\boldsymbol{D}}$, during a specified time of analysis, for which the user equilibrium traffic assignment is in terms of a variational inequality problem, (2.17) can be revised as follows when signal settings are taken into account in the equilibrium flows and the corresponding travel time functions.

To find values $\boldsymbol{q}^{\star}(\psi)$ of $\boldsymbol{q}(\psi)$ such that

$$c(q^{\star}, \psi)\ (q(\psi) - q^{\star}(\psi))^T \geq 0 \tag{2.24}$$

$$\forall\ q(\psi) \in \Omega(\psi) = \{\ q(\psi)\ ;\ q(\psi)^T = \delta f(\psi)^T,\ \Delta f(\psi)^T = \hat{D}^T,\ f(\psi) \geq 0\ \}$$

As it is noted in expressions (2.23)-(2.24), given by the fixed travel demands, $\hat{D}$, of all pairs of origin-destination during a specified time of analysis, the resulting equilibrium flows and the corresponding travel times depend on the signal setting variables, ψ . Therefore, for a road traffic network subject to signal control operation, the sequential effects on the resulting equilibrium flows and the corresponding travel times will be caused if any element of the signal setting variables varies. In Section 2.1, the optimisation problem of area traffic control has been reviewed, in which we supposed the average traffic flow patterns remain fixed while the signal settings are optimised. This assumption can be relaxed now and a full optimisation process can be performed by which the problem of the area traffic control optimisation and the problem of user equilibrium traffic assignment are brought together as a combined optimisation problem. In addition, this combined optimisation problem can be regarded as one special case of the equilibrium network design problem.

As far as an equilibrium network design problem is concerned, a best system performance can be achieved by optimising a chosen objective function with respect to the corresponding decision variables, in which users' behaviour of choice of routes is supposed to follow Wardrop's first principle; that is, a road user will choose his route between a specified origin-destination pair with the minimal travel time which is in turn dependent on the choice of system decision variables. Therefore, in the optimisation process for the equilibrium network design problem, not only the decision variables themselves need to be considered but also the consequential effects on the equilibrium flows caused by the corresponding decision variables need to be taken into account. In relation to the formulation of the equilibrium network design problem for area traffic control, two classes below are distinguished according to the characteristics of the corresponding solutions: a mutually consistent formulation and an optimal formulation. Furthermore, the solution methods for the formulations stated above are accordingly separated into the mutually consistent calculation and mathematical optimisation methods. In the following subsections,

the ways of formulating the equilibrium network design problem for the area traffic control will be introduced first and the corresponding solution methods will be discussed accordingly.

2.3.2 Formulations

The mathematical optimisation formulation is to achieve the optimal value of the objective function with respect to decision variables of interest, in such a way that the dependence of equilibrium traffic flows on signal settings is taken into account. The optimisation formulation applied to the area traffic control problem while user equilibrium flows are assumed is described in the following way.

Let ψ be the signal settings, $q^{\star}(\psi)$ be the corresponding equilibrium flow which can be obtained by solving (2.23) or (2.24), and S_0 be the feasible region for signal settings.

Find the signal settings ψ within the feasible region S_0 to

$$\begin{aligned} &\underset{\psi}{Minimise} \quad Z_0(\psi\ ,\ q^{\star}(\psi)) \\ &subject\ to \quad \psi \in S_0 \end{aligned} \tag{2.25}$$

Furthermore, the mutually consistent formulation for the optimised area traffic control problem and user equilibrium traffic assignment can be expressed as follows.

To find $\psi^{\star}$ such that

$$\begin{aligned} Z_2(\psi^{\star}\ ,\ q^{\star}(\psi^{\star})) = &\underset{\psi \in S_0}{Minimise} \quad Z_2(\psi\ ,\ q^{\star}(\psi^{\star})) \\ &subject\ to \quad \psi^{\star} \in S_0 \end{aligned} \tag{2.26}$$

In (2.26), the objective of the formulation for mutually consistent signal settings and the corresponding equilibrium flows is to find a particular signal setting $\psi^{\star}$ which minimises the value of the objective function, $Z_2(\psi^{\star}\ ,\ q^{\star})$, which is in terms of chosen indicators of performance index for the resulting equilibrium traffic flows under Wardrop's first principle. Allsop and Charlesworth (1977) carried out a mutually consistent calculation

for the area traffic control optimisation problem and equilibrium network flows, in which the signal settings and link flows were treated alternately and obtained respectively by solving the area traffic control optimisation problem for the assumed link flows and by solving user equilibrium traffic assignment for the resulting signal settings until an intuitively expected convergence is achieved. The resulting mutually consistent signal settings and link flows will, however, in general be a non-optimal solution as has been discussed by Gershwin and Tan (1979) and Dickson (1981).

On the other hand, in relation to the ways of the mathematical formulation for (2.25), the bi-level formulation has been adopted as an appropriate tool in dealing with this combined problem. For example, Heydecker and Khoo (1990) formulated this combined problem as a bi-level programme, in which the optimisation of signal timings was regarded as the upper level programme whilst the user equilibrium traffic assignment was regarded as the function of signal timings and therefore it was dealt with as the lower level programme. In the bi-level formulation to the optimisation 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 comparison with the results obtained from the mutually consistent formulation (2.26) in a trial road network, it was reported that as expected the total system performance achieved by the bi-level formulation was an improvement on that given by mutually consistent formulation (Heydecker and Khoo 1990).

The approaches of the bi-level formulation (2.25) and mutually consistent formulation (2.26) for the equilibrium network design problem can be illustrated (Allsop 1997) by the two-dimensional contours in Figure 2.1. In Figure 2.1, the two-dimensional contours represent the signal settings ψ and the corresponding link flows q that give particular values of performance index. For example, the inner contour represents a smaller value of the performance index than does the outer contour. The objective is to find the minimum value of the performance index represented by the two-dimensional contours by adjusting the signal settings ψ and the corresponding equilibrium flows q . The curve $q(\psi)$ represents the assumed relationships of the equilibrium flows and the signal settings. Suppose the point A is the solution given by the bi-level formulation (2.25) with the

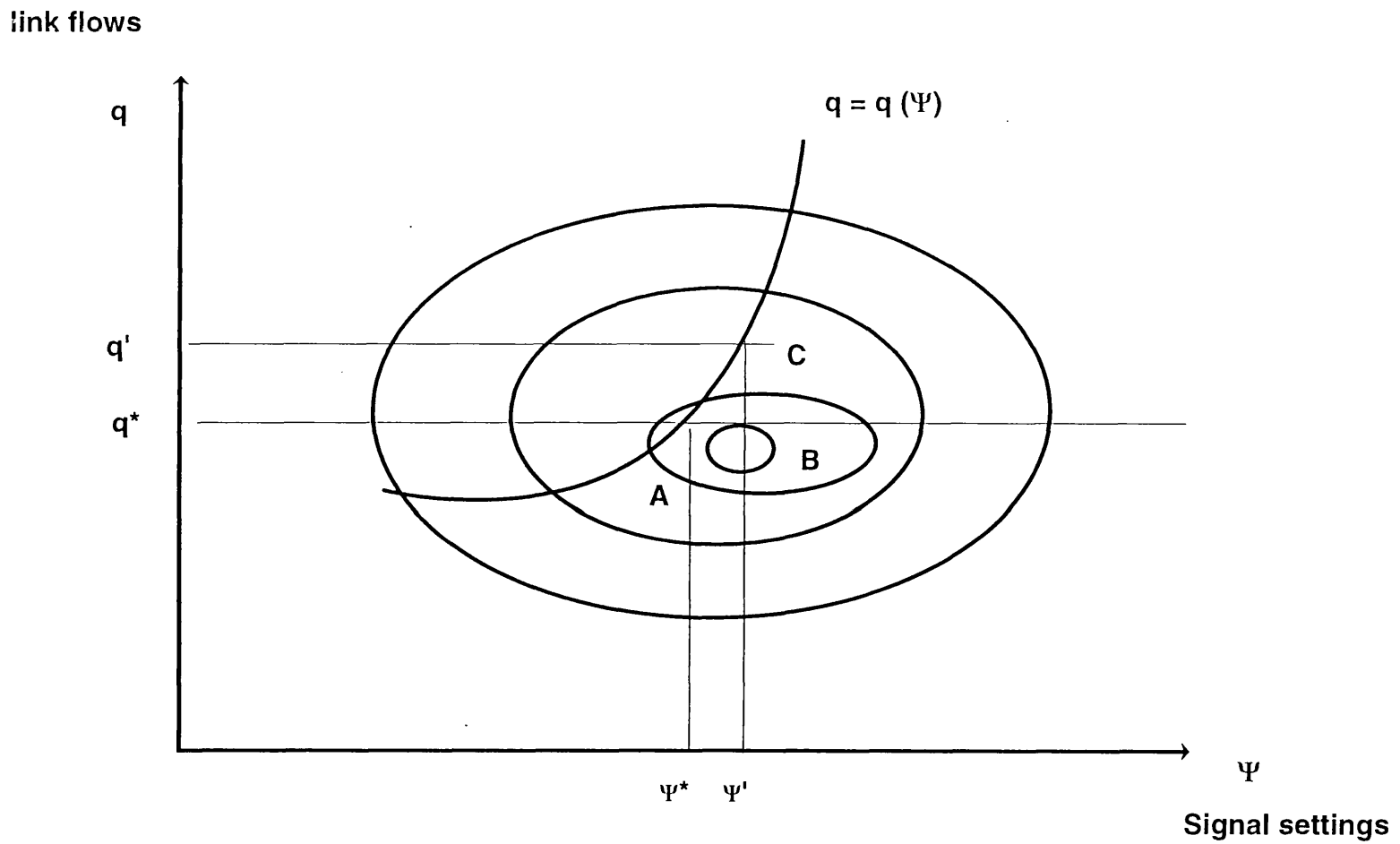


Figure 2.1 Illustration for the solutions of bi-level formulation and mutually consistent calculations

optimal signal settings $\psi^\star$ and the corresponding equilibrium flows $q^\star = q(\psi^\star)$.

Taking the optimal equilibrium flows $q^\star = q(\psi^\star)$ as the fixed value for the next step in searching for the mutually consistent solution, we may find that the optimal signal settings move from $\psi^\star$ to ψ' and the solution point moves from point A to B . Given the mutually consistent signal settings ψ' as the fixed value, the corresponding equilibrium flows can be solved by the user equilibrium traffic assignment (2.23) or (2.24) and represented by $q' = q(\psi')$ corresponding to the point C with a rather higher value of the performance index contour. Since optimising signal timings for this equilibrium flows will make little difference to ψ' we see that ψ' , q' is an approximate solution to the mutually consistent formulation (2.26), which is quite far from the optimal solution, i.e. $\psi^\star$, $q^\star$ given by (2.25).

2.3.3 Solution methods

In relation to the solution methods to the equilibrium network design problem for area traffic control, there are two classes of solution method according to the corresponding formulations: the mathematical optimisation method and the mutually consistent calculation. As it is noted in (2.26), Allsop and Charlesworth (1977) and Charlesworth (1977) reported a mutually consistent calculation for the signal settings in a test road network and the corresponding equilibrium traffic flows. The mutually consistent calculation of signal settings and the corresponding equilibrium traffic flows involved two classes of calculation: one is calculation of the optimal signal settings for an area traffic control problem with fixed traffic flows, and the other is calculation of user equilibrium traffic assignment problem with fixed signal settings.

A mutually consistent calculation to this equilibrium network design problem for area traffic control optimisation can be found as the limit of a sequence of alternate calculations of the following two problems: one is the area traffic control optimisation given by fixed traffic flows and the other is the user equilibrium traffic assignment where the link travel times are composed of the average delays decided by given signal settings. Furthermore,

this process can be carried out in a series of iterations, where one iteration of the mutually consistent calculation can be stated as follows. First, for the area traffic control optimisation problem when the traffic flows remain fixed, the optimal signal settings which minimised the chosen indicators of the system performance index can be obtained by means of the techniques discussed in Section 2.1, and thus the corresponding average traffic delay occurred at the downstream junction of each link can be specified as a component of the travel time function for further use in the user equilibrium traffic assignment. Second, following the convex combination method as discussed in Section 2.2 to solve the user equilibrium traffic assignment problem (2.23) or (2.24) with specified signal settings and corresponding incurred average traffic delay, the resulting equilibrium traffic flows can be obtained and therefore replace the fixed ones in the area traffic control optimisation problem for next iteration. Mathematical expressions for the alternate calculation for the above two problems can be given as follows.

$$\begin{aligned} &\underset{\psi \in S_0}{\textit{Minimise}} \quad Z_2(\psi\ ,\ \boldsymbol{q}) \\ &\textit{subject to} \quad \boldsymbol{q} = \bar{\boldsymbol{q}} \end{aligned} \tag{2.27}$$

and

$$\begin{aligned} &\underset{q \in \Omega(\psi)}{\textit{Minimise}} \quad Z_1(\psi\ ,\ \boldsymbol{q}) \\ &\textit{subject to} \quad \psi = \bar{\psi} \end{aligned} \tag{2.23}$$

where $\bar{\boldsymbol{q}}$, $\bar{\psi}$ are respectively the fixed values of equilibrium flows and signal settings, and $\Omega(\psi)$ is the set of user equilibrium flows when the signal settings ψ is taken into account.

The stopping criterion for alternate calculation of mutually consistent signal settings and the corresponding equilibrium flows in (2.27) and (2.23) will be met when the difference in the values of signal settings or of equilibrium flows between successive iterations is not greater than a predetermined threshold value. As it was mentioned in Section 2.3.2, the non-optimal formulation of the mutually consistent calculation can not achieve an optimal solution to the area traffic control optimisation problem. Another mathematical optimisation method is discussed below and in which the dependence of the equilibrium flows on signal settings will be dealt with in various ways.

To find the optimal value of the performance index for the equilibrium network design problem (2.25), a variety of mathematical optimisation methods have been reported for dealing with the dependence of equilibrium flows on signal settings. As far as these mathematical optimisation methods are concerned, two distinct classes have been noted according to the way of dealing with the dependence of equilibrium flows: firstly, a direct search approach by means of Hooke and Jeeves' method, which solved an unconstrained optimisation problem with respect to the decision variables where the equilibrium flows were regarded as one of the decision variables; secondly, a bi-level programming approach, which solved a constrained optimisation problem where a two level formulation was considered and the user equilibrium traffic assignment was solved by the lower level problem. In the following paragraphs, we will have a look at the direct search approach to the equilibrium network design problem (2.25) by means of Hooke and Jeeves' method first; then the bi-level programming approach will be investigated for the application to the equilibrium network design problem.

Dealing with the dependence of equilibrium flows on decision variables can be solved by means of a direct search approach where Hooke and Jeeves' method (1961) is used. Abdulaal and LeBlanc (1979) reported the formulation and solution method by means of Hooke and Jeeves' method for an equilibrium network design problem with continuous variables. Since no explicit mathematical expression is available for the relationships of the equilibrium flows to the decision variables, in relation to the application of Hooke and Jeeves' method to an equilibrium network design problem, Abdulaal and LeBlanc solved an unconstrained optimisation problem in which the dependence of equilibrium flows on decision variables was dealt with as one of the decision variables and solved directly by means of the convex combination method. Following the Hooke and Jeeves' method, the equilibrium flows need to be calculated each time when a decision variable changes and the value of the objective function will be changed accordingly. As for the application of Hooke and Jeeves' method to the equilibrium network design problem for area traffic control optimisation in (2.25), the following steps as referred to Abdulaal and LeBlanc (1979, pp32) are given below.

Step 1. Initialization:

1.1 For signal settings $\psi = [\ \psi_1, \psi_2, \ldots, \psi_N\ ]$, *where* N is the number of decision

variables for (2.25), choose an initial solution point $\psi^{(0)}$ satisfied the constraint set S_0 in (2.25), and let $\psi = \psi^{(0)}$.

1.2 Set an initial step size α , $\alpha > 0$ for the *exploratory search* in Step 2 and the acceleration factor β for the *pattern search* in Step 3; set indicator $\eta = 1$

1.3 Set indices $j = 1$, $k = 0$

<u>*Step 2. Exploratory search:*</u>

2.0 If $j = N + 1$, go to Step 3

2.1 Let e_j , $j = 1, \ldots, N$ be a vector with 1 in the j th component of the decision variable vector and 0 elsewhere, and set $\hat{\psi} = \psi + \alpha \eta e_j$.

Evaluate the objective function Z_0 at $\hat{\psi}$ along its positive coordinate direction with predetermined step move size α .

2.2 If $Z_0(\hat{\psi}) < Z_0(\psi)$ put $\psi = \hat{\psi}$, $j = j + 1$, $\eta = 1$ and go to Step 2.0; otherwise go to Step 2.3.

2.3 If $\eta = 1$, put $\eta = -1$ and go to Step 2.1, otherwise if $\eta = -1$ put $j = j + 1$, $\eta = 1$ and go to Step 2.0.

<u>*Step 3. Pattern search:*</u> when the evaluation of the objective function with respect to each component of the signal settings has been conducted in the *exploratory search*, we need to compare Z_0 at $\hat{\psi}$ *and* $\psi^{(k)}$

3.1 If $Z_0(\hat{\psi}) < Z_0(\psi^{(k)})$, put $\psi^{(k+1)} = \hat{\psi}$ and $\psi = \psi^{(k)} + \beta(\psi^{(k+1)} - \psi^{(k)})$ and $j = 1$ then go to Step 2, otherwise go to Step 3.2.

3.2 If the step size α is sufficiently small within the predetermined threshold, then the process is complete with the optimal signal settings $\psi^{(k)}$; otherwise put $\alpha = 0.5\ \alpha$, $j = 1$ *and* $\psi = \psi^{(k)}$ and return to Step 2.

It has been shown (Suwansirikul, Friesz and Tobin 1987; Marcotte and Marquis 1992) in using the direct search approach by means of Hooke and Jeeves' method to deal with the dependence of equilibrium flows on decision variables when solving an equilibrium network design problem did, as was to be expected, achieve better system performance in comparison with the result obtained by means of the mutually consistent calculation for a small test road network. The direct search approach by means of Hooke and Jeeves' method to solve an equilibrium network design problem, however, has its disadvantages. As we observe in the computing process of Hooke and Jeeves' method, the equilibrium flows need to be recalculated as each element of the decision variables changes in the steps of exploratory and pattern search; furthermore, a series of reevaluations of the value of the objective function will be made accordingly. Thus a huge computation load will be incurred as the number of the decision variables increases when a medium-sized test road network is taken into account.

The other alternative to solve an equilibrium network design problem is a bi-level programming approach. In the upper level problem, this optimises the system performance with respect to the decision variables of interest, while in the lower level problem, it solves a user equilibrium traffic assignment problem. The dependence of equilibrium flows on the decision variables of the upper level problem is regarded as a constraint for the upper level problem and is solved by the lower level problem. Therefore, the bi-level programme can be regarded as a constrained optimisation problem and solved by the techniques from optimisation theory. However, it has been pointed out (Heydecker and Khoo 1990), that although the dependence of equilibrium flows on the signal settings can be regarded as a lower level problem and solved by user equilibrium traffic assignment, no explicit mathematical expression for this dependence is available; furthermore, due to the non-linearity of the relationships between the equilibrium flows and signal settings, such a constrained optimisation problem leads to one unfortunate fact - the lower level problem forms a non-convex constraint set to the bi-level programme.

In relation to the way of dealing with the dependence of the equilibrium flows on the traffic signal settings for the equilibrium network design problem (2.25), there are two distinct classes proposed for the bi-level programme. First, the techniques of using the results in sensitivity analysis for network equilibrium flows to calculate the perturbations of the value of the objective function for an equilibrium network design problem with respect to the decision variables of interest have been employed and explored over the past years (Friesz, Tobin, Cho and Mehta, 1990; Suh and Kim 1992; Yang and Yagar 1995). Furthermore, in relation to the calculation of the perturbations of equilibrium flows caused by changes in travel time functions with respect to the decision variables of interest, work has been reported by Tobin and Friesz (1988) in which the perturbations of equilibrium flow can be expressed in terms of derivatives of the travel time functions. Second, the techniques of using a linear regression to fit the relationship for the dependence of the equilibrium flows on decision variables of interest has been proposed by Heydecker and Khoo (1990). In the following paragraphs, two bi-level programming approaches are investigated for the equilibrium network design problem (2.25) in terms of the ways of dealing with the dependence of the equilibrium flows on the decision variables of interest: a linear constraint approximation, which approximates the dependence of equilibrium flows on decision variables in the mathematically convenient form of a linear expression (Heydecker and Khoo 1990), and a sensitivity analysis based solution method, which employs the derivatives of the equilibrium flows calculated by the methods of Tobin and Friesz (Yang and Yagar 1995).

Heydecker and Khoo proposed a linear constraint approximation to the equilibrium flows with respect to signal setting variables and solved the bi-level problem (2.25) as a constrained optimisation problem subject to a linear constraint set as follows.

Let $\boldsymbol{d}$ be a feasible direction in the space ψ and γ be a parameter within the chosen range $[\gamma_L, \gamma_U]$ for given signal settings ψ^0, the feasible signal settings $\hat{\psi} \in S_0$ therefore can be represented by $\hat{\psi} = \psi^0 + \gamma \boldsymbol{d}$ and the approximate equilibrium flows $\hat{q}$ can be fitted by $\hat{q} = \alpha + \gamma \beta$, where α and β are the vectors of regression parameters fitted by least squares to the values calculated from (2.23).

The approximate problem for (2.25) is to

$$\begin{aligned} &\underset{\gamma}{Minimise} \quad Z_0(\hat{\psi}\ ,\ \hat{q}) \\ &subject\ to \quad \gamma_L \le \gamma \le \gamma_U \end{aligned} \tag{2.28}$$

In relation to the linear constraint approximation to equilibrium flows when a bi-level problem is being solved, one iteration k of the linear constraint method, for example, can be expressed in the following steps.

<u>*Step 1.*</u> For given signal settings ψ^0 , one feasible direction $\boldsymbol{d}$ is identified for each component of ψ^0 by making an increase in that particular component but making appropriate decreases in other components relating to the same junction so that the constraints on the components of vector ψ^0 relating to each junction are satisfied.

<u>*Step 2.*</u> Improved signal settings are sought in one of these feasible directions, $\boldsymbol{d}$, that is parameterised by γ in a given range which is determined by the extent of the feasible region in direction $\boldsymbol{d}$. This is derived by constructing a line segment $\psi^0 + \gamma\ \boldsymbol{d}$ such that it extends to the boundary of the feasible region, i.e., $\psi^0 + \gamma\ \boldsymbol{d} \in S_0$

<u>*Step 3.*</u> Select values of γ from the specified interval, $[\ \gamma_L\ ,\ \gamma_U\ ]$, and solve the users equilibrium assignment problem (2.23) at each of the corresponding points,

$\hat{\psi} = \psi^0 + \gamma\ \boldsymbol{d}$ then regress the resulting equilibrium link flows $\boldsymbol{q} = \boldsymbol{q}^{\star}(\hat{\psi})$ on γ to fit the linear approximation as follows.

$$\hat{\boldsymbol{q}} = \alpha + \gamma\ \beta \tag{2.29}$$

<u>*Step 4.*</u> Solve (2.28) by golden section method for optimal value $\gamma^{\star}$ of γ :

Then put $\psi^0 = \psi^0 + \gamma^{\star}\ \boldsymbol{d}$

Step 5. If there is another search direction needed to be considered, replace d by the next such direction and return to *Step 2*, otherwise the iteration k is complete.

Using the linear constraint approximation method to solve the bi-level problem (2.28) can be carried out in a number of iterations by which the resulting equilibrium flows are regressed as the signal setting variable changes in a simple linear form (2.29). As it is shown in (2.29), the dependence of equilibrium flows on signal settings can be dealt with as a linear constraint for (2.25) in the neighbourhood of specified values of the signal setting variables. Therefore, the equilibrium network design problem (2.25) for area traffic control optimisation can be solved by a number of iterations of the one dimensional optimisation problem shown in (2.28). However, the dependence of equilibrium flows on signal settings is by no means as simple as a linear form like (2.29), and although using a linear approximation to the constraint of the equilibrium flows may reduce the computation effort as compared to that incurred by carrying out a direct search by means of Hooke and Jeeves' method; no mathematical evidence has been found so far that convergence of the solution obtained by carrying out the linear approximation constraint method might be achieved for a general road network.

On the other hand, another possible approach to solve the equilibrium network design problem (2.25) is a sensitivity analysis based solution method. For an equilibrium network design problem the sensitivity analysis based method is used to estimate the perturbations of the value of the objective function caused by changes of the signal setting variables. Because of the dependence of equilibrium flows on the signal setting variables, the perturbations of equilibrium flows caused by changes in signal settings are particularly significant as we estimate the perturbations of the value of the objective function. Tobin and Friesz (1988) derived a rigorous mathematical sensitivity analysis method for equilibrium network flows, which will be investigated in the following subsection. Furthermore, following the result proposed by Tobin and Friesz, Yang and Yagar (1994,1995) and Yang (1997) reported the applications of the sensitivity analysis for equilibrium network flows in a bi-level programme for an equilibrium network design problem, in which they investigated solving the bi-level problem by means of the derivatives of equilibrium flows and of travel time functions.

Yang and Yagar (1995), for example, used derivatives of equilibrium flows and of the corresponding travel times to solve a bi-level programme for the equilibrium network design problem for a signal control optimisation. In this bi-level programme for a signal control optimisation, at the upper level, Yang and Yagar simplified the control variables to comprise intervals between starts of stages which are link-specific variables, and assumed in a rudimentary example that each signal group controls only one signal-controlled traffic stream; in addition, the common cycle time and offsets were taken fixed throughout the network which could be undersaturated or oversaturated. As the performance index of the signal-controlled network, the sum of running time, signal delay and queuing delay spent in the network by all vehicles over a given time period was defined as the performance index for the signal-controlled network. In the lower level problem, on the basis of Beckmann transformation, Yang and Yagar dealt with the user equilibrium traffic assignment as a function of green splits but took into account the link flow capacity constraints by regarding the queuing delay for each link as equal to the Lagrange multiplier from the corresponding link flow capacity inequality constraint. Also, Yang and Yagar proposed a sensitivity analysis based solution method to the bi-level programme, the critical point of which lies in the way of obtaining the first partial derivatives for the equilibrium flows and queuing delays with respect to the green splits. Applying the results from Tobin and Friesz for the sensitivity analysis for equilibrium network flow to this case, Yang and Yagar obtained the total derivatives of the objective function with respect to the green splits. By using these derivatives for a given feasible solution, the upper level problem was therefore approximated as one linear programming problem and the simplex method was used to solve it. The bi-level programme for the signal control optimisation formulated by Yang and Yagar (1995, pp 131-134) can be given as follows.

Let N be the set of signal-controlled junctions in the network, L be the set of links and L_m be the subset of L related to junction m, $\forall\, m \in N$ to which the corresponding links are leading. For each link a in L, let λ_a be the green split for the link, and for each junction m, $m \in N$ to which the links in a subset L_m are leading, we have $\sum_{a \in L_m} \lambda_a = 1.0$, $\forall\, m \in N$ because it is assumed that each signal

group has only one traffic stream, and let $\lambda_a^{\min}$, $\lambda_a^{\max}$ be respectively the minimum and maximum green splits; furthermore let $c_a(q_a , \lambda_a)$, d_a be respectively the travel time function of the average flow q_a and green split λ_a and the average delay to a vehicle travelling along the link, and let s_a , $Q_a^{\max}$ be respectively the saturation flow and an indicator of the maximum storage capacity on that link.
Then the whole problem can then be stated as to

$$
\begin{aligned}
&\underset{\boldsymbol{\lambda}}{\textit{Minimise}} \quad Z(\boldsymbol{\lambda} , \boldsymbol{q}(\boldsymbol{\lambda}) , \boldsymbol{d}(\boldsymbol{\lambda})) = \sum_{a \in L} q_a(\boldsymbol{\lambda}) \, (c_a(q_a , \lambda_a) + d_a(\boldsymbol{\lambda})) \\
&\textit{subject to} \qquad d_a(\boldsymbol{\lambda}) \, q_a(\boldsymbol{\lambda}) \le Q_a^{\max} , \qquad \forall \, a \in \boldsymbol{L} \\
&\qquad \sum_{a \in L_m} \lambda_a = 1.0 , \qquad \forall \, m \in \boldsymbol{N} \\
&\qquad \lambda_a^{\min} \le \lambda_a \le \lambda_a^{\max} , \qquad \forall \, a \in \boldsymbol{L}
\end{aligned}
\tag{2.30}
$$

where $\boldsymbol{q}(\lambda)$ and $\boldsymbol{d}(\lambda)$ are obtained by solving user equilibrium traffic assignment problem with specified signal settings as similarly given by (2.23) as follows.

$$
\begin{aligned}
&\underset{\boldsymbol{q}(\boldsymbol{\lambda})}{\textit{Minimise}} \quad \sum_{a \in L} \int_0^{q_a(\boldsymbol{\lambda})} c_a(\boldsymbol{\lambda} , w) \, dw \\
&\textit{subject to} \quad \sum_{p \in P_w} f_p(\boldsymbol{\lambda}) = \hat{D}_w , \quad \forall \, w \in \boldsymbol{W} \\
&\qquad \sum_{w \in W} \sum_{p \in P_w} f_p(\boldsymbol{\lambda}) \, \delta_{ap} = q_a(\boldsymbol{\lambda}) , \quad \forall \, a \in \boldsymbol{L} \\
&\qquad f_p(\boldsymbol{\lambda}) \ge 0 , \quad \forall \, p \in \boldsymbol{P}_w , \forall \, w \in \boldsymbol{W} \\
&\qquad \lambda_a \, s_a \ge q_a , \quad \forall \, a \in \boldsymbol{L}
\end{aligned}
\tag{2.31}
$$

In (2.31) it assumes that queue forms on the saturated link and the corresponding delay, d_a , $a \in \boldsymbol{L}$, can be decided accordingly in the following conditions:

$$\left.\begin{array}{ll} d_a = 0 & \text{if } q_a < \lambda_a s_a \\ d_a \geq 0 & \text{if } q_a = \lambda_a s_a \end{array}\right\} \quad \forall\, a \in \boldsymbol{L}$$

The first-order necessary conditions of (2.31) are

$$\left.\begin{array}{r} f_p \,(\sum_{a \in L} \delta_{ap}\,(c_a(q_a\,,\lambda_a) + d_a(\boldsymbol{\lambda})) - u_w) = 0 \\ \sum_{a \in L} \delta_{ap}\,(c_a(q_a\,,\lambda_a) + d_a(\boldsymbol{\lambda})) - u_w \geq 0 \\ f_p \geq 0 \end{array}\right\} \;, \forall\, p \in \boldsymbol{P}_w \,, \forall\, w \in \boldsymbol{W}$$

$$\sum_{p \in P_w} f_p = \hat{D}_w \quad , \forall\, w \in \boldsymbol{W}$$

and for queueing delay on saturated link, d_a , $a \in \boldsymbol{L}$,

$$\begin{aligned} (\lambda_a s_a - q_a)\, d_a &= 0 \\ d_a &\geq 0 \\ \lambda_a s_a - q_a &\geq 0 \end{aligned}$$

Therefore the equilibrium conditions for (2.31) are

$$\left.\begin{array}{ll} \sum_{a \in L} \delta_{ap}\,(c_a(q_a\,,\lambda_a) + d_a(\boldsymbol{\lambda})) = u_w & \text{if } f_p > 0 \\ \sum_{a \in L} \delta_{ap}\,(c_a(q_a\,,\lambda_a) + d_a(\boldsymbol{\lambda})) \geq u_w & \text{if } f_p = 0 \end{array}\right\} \quad \forall\, p \in \boldsymbol{P}_w \,,\, w \in \boldsymbol{W}$$

and for queueing delay on saturated link a , $a \in \boldsymbol{L}$

$$\begin{aligned} d_a &= 0 \quad \text{if } q_a < \lambda_a s_a \\ d_a &\geq 0 \quad \text{if } q_a = \lambda_a s_a \end{aligned}$$

A linearized approximation to the expression (2.30) in the neighbourhood of $\lambda = \lambda^\star$ is as follows.

$$\underset{\boldsymbol{\lambda}}{\textit{Minimise}} \;\; \hat{Z}(\boldsymbol{\lambda}) = \sum_{i \in L} (\lambda_i - \lambda_i^\star) \sum_{a \in L} [(\frac{\partial q_a(\boldsymbol{\lambda}^\star)}{\partial \lambda_i}(c_a(q_a^\star,\lambda_a^\star) + q_a(\boldsymbol{\lambda}^\star)\frac{\partial c_a(q_a^\star,\lambda_a^\star)}{\partial q_a}) +$$

$$q_a(\boldsymbol{\lambda}^*)\frac{\partial c_a(q_a^*,\lambda_a^*)}{\partial \lambda_i} + d_a(\boldsymbol{\lambda}^*)\frac{\partial q_a(\boldsymbol{\lambda}^*)}{\partial \lambda_i} + q_a(\boldsymbol{\lambda}^*)\frac{\partial d_a(\boldsymbol{\lambda}^*)}{\partial \lambda_i}] \tag{2.32}$$

subject to

$$\sum_{i \in L} \lambda_i(d_a(\boldsymbol{\lambda}^*)\frac{\partial q_a(\boldsymbol{\lambda}^*)}{\partial \lambda_i} + q_a(\boldsymbol{\lambda}^*)\frac{\partial d_a(\boldsymbol{\lambda}^*)}{\partial \lambda_i}) \le \hat{Q}_a^{\max}, \quad \forall\, a \in \boldsymbol{L}$$

$$\sum_{a \in L_m} \lambda_a = 1.0, \quad \forall\, m \in \boldsymbol{N}$$

$$\lambda_a^{\min} \le \lambda_a \le \lambda_a^{\max}, \quad \forall\, a \in \boldsymbol{L}$$

where $\hat{Q}_a^{\max} = Q_a^{\max} - d_a(\boldsymbol{\lambda}^*)q_a(\boldsymbol{\lambda}^*) + \sum_{i \in L} \lambda_i^*(d_a(\boldsymbol{\lambda}^*)\frac{\partial q_a^*}{\partial \lambda_i} + q_a(\boldsymbol{\lambda}^*)\frac{\partial d_a^*}{\partial \lambda_i})$

The solution method using the sensitivity analysis based approach for (2.30)-(2.31) can be summarized in the following steps.

<u>*Step 1*</u>. Determine a set of initial signal splits.

<u>*Step 2*</u>. Solve the lower level problem (2.31) corresponding to a queuing network equilibrium problem which takes into account the signal setting variables denoted as green splits.

<u>*Step 3*</u>. Calculate the derivatives for the equilibrium flows and queuing delays with respect to green splits at the basis of sensitivity analysis for equilibrium network flow, which results from Tobin and Friesz.

<u>*Step 4*</u>. Linearize the approximation for the upper level problem and use the derivatives obtained from *Step 3,* then solve the linear programming problem (2.32) in order to obtain an auxiliary solution.

<u>*Step 5*</u>. Update the green splits by linearly combining the auxiliary variable and the previous green split.

<u>*Step 6*</u>. Convergence test: if the difference in the resulting green splits, or equilibrium flow between successive iterations is within a predetermined threshold then the solution is achieved; otherwise return to *Step 2*.

As far as the sensitivity analysis based method is concerned, in (2.30)-(2.31), for example, it deals with the dependence of equilibrium flows on signal setting variables such as green splits by means of the derivative expressions with the rigorous mathematical background of the theory of the sensitivity analysis. By comparison with the linear constraint approximation method, the sensitivity analysis based method widens the choices of the solution methods to the bi-level programme when the derivatives of the value of the objective function are available. For example, as these derivatives of the objective function in (2.25) are available, optimisation techniques by means of these derivatives can be applied to solving this bi-level optimisation problem. In the following Chapters, we will concentrate on the equilibrium network design problem for the area traffic control optimisation, in which the bi-level programming technique is employed and the method of solutions by using sensitivity analysis to obtain the required derivatives of the objective function will be particularly important.

2.4 Sensitivity Analysis

In relation to the solution method by means of the sensitivity analysis to the bi-level programme (2.25), the perturbations of equilibrium flows caused by the changes of signal setting variables can be estimated by derivatives with respect to the signal setting variables. Use of sensitivity analysis in mathematical programming provides an approach to obtaining these derivatives from other known derivatives. Techniques of using the sensitivity analysis in dealing with the nonlinear programming problem has been presented by Fiacco (1976, 1983). Furthermore, for the purpose by applying sensitivity analysis to a generalized variational inequality problem, Tobin (1986) proposed a parallel development for which the perturbations of the variational inequality function and the feasible region by changes in the decision variables are available. When the equilibrium network design problem for an area traffic control optimisation is considered, an extension of the techniques used in the sensitivity analysis for variational inequalities to the specific equilibrium network flow problems has been studied by Tobin and Friesz (1988). From the results obtained by Tobin and Friesz, as we have mentioned before, the perturbations of equilibrium flow with respect to signal setting variables can be calculated. In the following subsections, a

brief mathematical review of the sensitivity analysis for variational inequalities will be introduced first. Then the application of sensitivity analysis for variational inequalities to the equilibrium network flow problem will be investigated. Furthermore, in relation to an area traffic control optimisation problem with respect to phase-based signal control variables for given traffic flows, the derivatives of the performance index have been reported by Wong (1995). These resulting derivatives will be quoted in the last part of this section.

2.4.1 Sensitivity analysis for variational inequalities

The theorems without proof related to the derivations of sensitivity analysis for equilibrium network flows in terms of a variational inequalities are adapted from Tobin (1986) and referred to Tobin and Friesz (1988, pp 243-244) as follows.

Let $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be continuous, $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be differentiable, and

$d : \mathbb{R}^n \rightarrow \mathbb{R}^p$ be linear affine.

Define $K = \{ x \in \mathbb{R}^n \mid g(x) \geq 0 \ , \ d(x) = 0 \}$ (2.33)

if there exists $x^\star \in K$ such that $F(x^\star)^T (x - x^\star) \geq 0$, *for all* $x \in K$ (2.34)

then, inequality (2.34) is a variational inequality problem and $x^\star$ is a solution.

<Theorem 2.1> Necessary conditions for solution:

If the vector $x^\star \in K$ *is a solution for the variational inequality* (2.34) *and the gradients* $\nabla g_i(x^\star)$, i *such that* $g_i(x^\star) = 0$, *and* $\nabla d_i(x^\star)$, $i = 1 ,..., p$ *are linearly independent, then there exists* $\pi \in \mathbb{R}^m$, $u \in \mathbb{R}^p$ *such that*

$$F(x^\star) - \nabla g(x^\star)^T \pi - \nabla d(x^\star)^T u = 0 \tag{2.35}$$

$$\pi^T g(x^\star) = 0 \tag{2.36}$$

$$\pi \geq 0 \tag{2.37}$$

■

<Theorem 2.2> Sufficient conditions for solution:

If $g_i(x)$ *for* $i=1,\ldots\ldots,m$ *are concave,* $x^\star \in K$, $\pi \in \mathbb{R}^m$ *and* $u \in \mathbb{R}^p$ *satisfy the expressions* (2.35)-(2.37), *then* $x^\star$ *is a solution to the variational inequality* (2.34). ■

<Theorem 2.3> Sufficient conditions for a locally unique solution:

If the conditions of Theorem 2.2 hold, and, in addition F *is differentiable and*

$$y^T \nabla F(x^\star)\, y > 0 \tag{2.38}$$

For all $y \neq 0$ *such that*

$$\nabla g_i(x^\star)\, y \geq 0 \;,\; \textit{for all } i \textit{ such that } g_i(x^\star) = 0 \tag{2.39}$$

$$\nabla g_i(x^\star)\, y = 0 \;,\; \textit{for all } i \textit{ such that } \pi_i > 0 \tag{2.40}$$

$$\nabla d_i(x^\star)\, y = 0 \;,\; \textit{for } i = 1, \ldots, p \tag{2.41}$$

then $x^\star$ *is a locally unique solution to the variational inequality* (2.34). ■

Let ε be a vector of perturbation parameters of dimension q , and let

$F : \mathbb{R}^n \times \mathbb{R}^q \to \mathbb{R}^n$ be once continuously differentiable in (x, ε) , and let

$g : \mathbb{R}^n \times \mathbb{R}^q \to \mathbb{R}^m$ be concave in x and twice continuously differentiable in

(x, ε) , and let $d : \mathbb{R}^n \times \mathbb{R}^q \to \mathbb{R}^p$ be linear affine in x and once continuously

differentiable in ε , and $K(\varepsilon) = \{ x \mid g(x, \varepsilon) \geq 0 \,,\, d(x, \varepsilon) = 0 \}$

Consider the following perturbed variational inequality, denoted as $VI(\varepsilon)$:

Find $x^\star \in K(\varepsilon)$ such that $F(x^\star, \varepsilon)^T (x - x^\star) \geq 0$, $\forall x \in K(\varepsilon)$ (2.42)

<Theorem 2.4> Implicit function theorem:

Let the conditions of Theorem 2.3 be satisfied for **VI** $(\mathbf{0})$ *with* $F(x^\star)$, $g(x^\star)$,

$d(x^\star)$, π , u *replaced by* $F(x^\star, \mathbf{0})$, $g(x^\star, \mathbf{0})$, $d(x^\star, \mathbf{0})$, $\pi^\star$, $u^\star$ *respectively;*

with the gradients $\nabla g_i(x^\star, \mathbf{0})$, i *such that* $g_i(x^\star, \mathbf{0}) = 0$ *and* $\nabla d_i(x^\star, \mathbf{0})$,

$i = 1,..., p$ *linearly independent; and, in addition, let the strict complementary slackness condition*

$\pi_i^\star > 0$ *when* $g_i(x^\star, \mathbf{0}) = 0$

be satisfied. Then, $\pi^\star$, $\mu^\star$ *are unique; and, in a neighbourhood of* $\varepsilon = \mathbf{0}$, *there exists a unique once continuously differentiable function*

$[\, x\,(\varepsilon)^T , \pi\,(\varepsilon)^T , u\,(\varepsilon)^T \,]^T$ *where* $x(\varepsilon)$ *is a locally unique solution to* **VI** (ε)

and $\pi\,(\varepsilon)$, $u\,(\varepsilon)$ *are unique associated multipliers satisfying the conditions of Theorem 2.3 for a locally unique solution of* **VI** (ε) *and with*

$[\, x\,(\mathbf{0})^T , \pi\,(\mathbf{0})^T , u\,(\mathbf{0})^T \,] = [\, (x^\star)^T , (\pi^\star)^T , (u^\star)^T \,]$.

In addition, in a neighbourhood of $\varepsilon = \mathbf{0}$ *the set of binding inequalities is unchanged, strict complementary slackness holds, and the binding constraint gradients are independent at* $x\,(\varepsilon)$. ■

For $\varepsilon = \mathbf{0}$ and $(\, x , \pi , u \,) = (\, x^\star , \pi^\star , u^\star \,)$, by *Theorem 2.1*

$$F(x , \mathbf{0}) - \sum_{i=1}^{m} \pi_i \nabla g_i(x , \mathbf{0})^T - \sum_{i=1}^{p} u_i \nabla d_i(x , \mathbf{0})^T = \mathbf{0} \tag{2.43}$$

$$\pi_i\, g_i(x , \mathbf{0}) = 0 , \qquad \textit{for } i = 1 ,..., m \tag{2.44}$$

$$d_i(x , \mathbf{0}) = 0 , \qquad \textit{for } i = 1 ,....., p \tag{2.45}$$

Let the Jacobian matrix of the system (2.43)-(2.45) with respect to

$y = (x, \pi, u)$ be denoted as $J_y^\star$ and with respect to ε as $J_\varepsilon^\star$.

<Corollary 2.5> Derivatives of the solution vector of ***VI*** **(0)** *with respect to* ε *:*

Under the assumptions of Theorem 2.4, the inverse of $J_y^\star$ *exists and the partial*

derivatives of $(x^\star, \pi^\star, u^\star)$ *with respect to* ε *are given by*

$$\nabla_\varepsilon y = \begin{bmatrix} \nabla_\varepsilon x^\star \\ \nabla_\varepsilon \pi^\star \\ \nabla_\varepsilon u^\star \end{bmatrix} = \left[J_y^\star \right]^{-1} \left[-J_\varepsilon^\star \right] \tag{2.46}$$

■

2.4.2 Sensitivity analysis for equilibrium network flow

The difficulties are clear in directly applying the results of *Theorem 2.4* in Section 2.4.1 to the sensitivity analysis of the equilibrium network flow due to the lack of the uniqueness of path flows when fixed origin-destination demands are assigned to the feasible paths on which the travel cost are minimum. Tobin and Friesz (1988) proposed a restricted equilibrium problem which regarded the positive path flows for each fixed origin-destination demand as the extreme points of the feasible region and therefore for a selected set of positive path flows the corresponding unique link flows can be identified. Simply by confining attention to the selected set of positive path flows, a restricted equilibrium problem was suggested by Tobin and Friesz for which the requirement of uniqueness by *Theorem 2.4* can be satisfied. Furthermore, Tobin and Friesz have shown that the derivatives which were obtained from the restricted equilibrium problem were equivalent to those from the original unrestricted equilibrium problem. In relation to the results from Tobin and Friesz (1988, pp 245-248), it can be briefly described as follows.

Recall the user equilibrium traffic assignment in variational inequality formulation in (2.17)

To find values $q^\star$ of q such that

$$c(q^\star)\ (q - q^\star)^T \geq 0 \tag{2.47}$$

$$\forall\ q \in \Omega = \left\{ q\ ;\ q^T = \delta f^T\ ,\ \Delta f^T = \hat{D}^T\ ,\ f \geq \mathbf{0} \right\}$$

Also, a set of the corresponding equilibrium path flows for $q^\star$ in (2.47) can be given

by $$K^\star = \left\{ f \mid \delta f^T = (q^\star)^T\ ,\ \Delta f^T = \hat{D}^T\ ,\ f \geq \mathbf{0} \right\} \tag{2.47a}$$

The equivalent variational inequality formulation for (2.47) in terms of path flows can be given as follows.

To find values $f^\star$ of f such that

$$C(f^\star)\ (f - f^\star)^T \geq 0 \tag{2.47b}$$

$$\forall\ f \in \Omega' = \left\{ f\ ;\ \Delta f^T = \hat{D}^T\ ,\ f \geq \mathbf{0} \right\}$$

Furthermore, let ε be a vector of perturbation parameters and suppose that $c(q\ ,\ \varepsilon)$ is once continuously differentiable in $(\ q\ ,\ \varepsilon\ )$, and $\hat{D}(\varepsilon)$ is once continuously differentiable in ε then the perturbed variational inequality for (2.47) can be given as follows.

To find values $q^\star$ of q such that

$$c(\ q^\star\ ,\ \varepsilon\ )\ (q - q^\star)^T \geq 0 \tag{2.48}$$

$$\forall\ q \in \Omega(\varepsilon) = \left\{ q\ ;\ q = \delta f^T\ ,\ \Delta f^T = \hat{D}\ (\varepsilon)^T\ ,\ f \geq \mathbf{0} \right\}$$

and a set of the corresponding path flows for $q^\star$ in (2.48) can be given by

$$K^\star\ (\varepsilon) = \left\{ f \mid \delta f^T = (q^\star)^T\ ,\ \Delta f^T = \hat{D}\ (\varepsilon)^T\ ,\ f \geq \mathbf{0} \right\} \tag{2.48a}$$

The equivalent variational inequality formulation for (2.48) in terms of path flows can be given as follows.

To find values $f^\star$ of f such that

$$C(f^\star, \varepsilon)(f - f^\star)^T \geq 0 \tag{2.48b}$$

$$\forall\ f \in \Omega'(\varepsilon) = \left\{ f\ ;\ \Delta f^T = \hat{D}(\varepsilon)^T,\ f \geq \mathbf{0} \right\}$$

Because $f^\star$ is a solution to the perturbed variational inequality (2.48b) at $\varepsilon = \mathbf{0}$, according to *Theorem 2.1* there exists a solution to the system

$$\begin{aligned} C(f^\star, \mathbf{0})^T - \pi^T - \Delta^T u^T &= \mathbf{0} \\ \pi\,(f^\star)^T &= 0 \\ \Delta\,(f^\star)^T - \hat{D}(\mathbf{0})^T &= \mathbf{0} \\ \pi^T &\gtrsim \mathbf{0} \end{aligned} \tag{2.49}$$

A restricted problem for (2.49) can be focused on those links with positive flow; and for each path p in P_w, $\forall\, w \in W$,

$$\left.\begin{aligned} C_p(f^\star, \mathbf{0}) &= u_w \\ \pi_p &= 0 \end{aligned}\right\} \quad \forall\, p \in P_w,\ \forall\, w \in W \tag{2.49a}$$

In addition, the restricted problem of (2.49) can be focused only on those path flow variables f_p, $\forall\, p \in P_w$, $\forall\, w \in W$ which are positive in $f^\star$.

Let $f^{0\star}$ be the path flows which are positive in $f^\star$ and P^0 be the set of the corresponding paths

$C^0(f^\star, \mathbf{0})$ be the corresponding path flow travel times

δ^0 and Δ^0 be respectively the link/path and origin-destination path incidence matrices.

Therefore the restricted problem for (2.49) can be given by

$$\begin{aligned} C^0 (f^\star , \mathbf{0})^T - (\Delta^0)^T u^T &= \mathbf{0} \\ \Delta^0 (f^{0\star})^T - \hat{D}(\mathbf{0})^T &= \mathbf{0} \end{aligned} \tag{2.50}$$

Since the conditions required by *Theorem 2.4* are satisfied by the restricted problem (2.50) (for details see Tobin and Friesz (1988, pp 246-247) the Jacobian matrices of (2.50) with respect to (f^0 , u) and ε at $\varepsilon = \mathbf{0}$ can be given by

$$J_{f^0 , \mu} = \begin{bmatrix} \nabla C^0(f^\star , \mathbf{0}) & -(\Delta^0)^T \\ \Delta^0 & \mathbf{0} \end{bmatrix} \tag{2.51a}$$

$$J_\varepsilon = \begin{bmatrix} \nabla_\varepsilon C^0 (f^\star , \mathbf{0})^T \\ -\nabla_\varepsilon \hat{D}(\mathbf{0})^T \end{bmatrix} \tag{2.51b}$$

Apply the results from *Corollary 2.5*, the derivatives of (f^0 , u) with respect to ε at $\varepsilon = \mathbf{0}$ is given by

$$\begin{bmatrix} (\nabla_\varepsilon f^0)^T \\ (\nabla_\varepsilon u)^T \end{bmatrix} = \left[J_{f^0 , u} \right]^{-1} \left[-J_\varepsilon \right] = \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix} \begin{bmatrix} -\nabla_\varepsilon C (f^\star , \mathbf{0})^T \\ \nabla_\varepsilon \hat{D}(\mathbf{0})^T \end{bmatrix} \tag{2.52}$$

where

$$B_1 = \nabla C^0(f^\star, \mathbf{0})^{-1} \left[I - (\Delta^0)^T \left[\Delta^0 \nabla C^0(f^\star, \mathbf{0})^{-1} (\Delta^0)^T \right]^{-1} \Delta^0 \nabla C^0(f^\star, \mathbf{0})^{-1} \right]$$

$$B_2 = \nabla C^0(f^\star, \mathbf{0})^{-1} (\Delta^0)^T \left[\Delta^0 \nabla C^0(f^\star, \mathbf{0})^{-1} (\Delta^0)^T \right]^{-1}$$

$$B_3 = -\left[\Delta^0 \nabla C^0(f^\star, \mathbf{0})^{-1} (\Delta^0)^T \right]^{-1} \Delta^0 \nabla C^0(f^\star, \mathbf{0})^{-1}$$

$$B_4 = \left[\Delta^0 \nabla C^0(f^\star, \mathbf{0})^{-1} (\Delta^0)^T \right]^{-1}$$

Since

$$(\nabla_\varepsilon q^0)^T = \delta^0 (\nabla_\varepsilon f^0)^T$$

$$\nabla_\varepsilon C^0(f^\star, 0)^T = (\delta^0)^T \nabla_\varepsilon c(q^\star, 0)^T \tag{2.53}$$

$$\nabla_f C^0(f^\star, 0) = (\delta^0)^T \nabla_q c(q^\star, 0)\, \delta^0$$

and by (2.52)

$$(\nabla_\varepsilon f^0)^T = -B_1 \nabla_\varepsilon C^0(f^\star, 0)^T + B_2 \nabla_\varepsilon \hat{D}(0)^T \tag{2.52a}$$

Substituting (2.53) into (2.52a), the derivatives of link flows with respect to ε at $\varepsilon = 0$ for the restricted problem (2.50) can be given by

$$(\nabla_\varepsilon q^0)^T = -\delta^0 B_{11} (\delta^0)^T \nabla_\varepsilon c(q^\star, 0)^T + \delta^0 B_{12} \nabla_\varepsilon \hat{D}(0)^T \tag{2.53a}$$

where

$$B_{11} = [(\delta^0)^T \nabla_q c(q^\star, 0)\, \delta^0]^{-1} [I - (\Delta^0)^T [\Delta^0 [(\delta^0)^T \nabla_q c(q^\star, 0)\, \delta^0]^{-1} (\Delta^0)^T]^{-1} \cdot \Delta^0 [(\delta^0)^T \nabla_q c(q^\star, 0)\, \delta^0]^{-1}]$$

$$B_{12} = [(\delta^0)^T \nabla_q c(q^\star, 0)\, \delta^0]^{-1} (\Delta^0)^T [\Delta^0 [(\delta^0)^T \nabla_q c(q^\star, 0)\, \delta^0]^{-1} (\Delta^0)^T]^{-1}$$

and I is the identity matrix.

In relation to the derivatives of link flows with respect to the perturbed parameters in (2.53a) which are calculated independent of the choice of the positive flows $f^\star$ in the restricted problem (2.50), and furthermore the results shown in (2.53) for the restricted problem (2.50) are equivalent to those for the original problem (2.49), the following theorem has been shown by Tobin and Friesz (1988, Theorem 6, pp 248).

<u>*Theorem 2.6*</u>

The values of $\nabla_\varepsilon q^0$ in (2.53) are independent of the choice of extreme point $f^\star$ and therefore, are the derivatives of link flows with respect to ε, $\nabla_\varepsilon q$, for the original (unrestricted) equilibrium problem. ■

2.4.3 Sensitivity analysis for area traffic control optimisation

In this subsection, results from application of sensitivity analysis by means of expressions of derivatives of the value of performance index with respect to signal settings for an area traffic control problem are quoted. In relation to the optimisation problem for area traffic control as reviewed in Section 2.1, Wong (1995) proposed expressions for the derivatives for the indicators of the performance index with respect to phase-based signal control variables for an area traffic control optimisation. Recall the objective function (2.1) in terms of the sum of a weighted linear combination of rate of delay and number of stops per unit time

$$\sum_{a \in L} W\, w_a\, D_a + K\, k_a\, S_a \tag{2.1}$$

Let $\hat{\theta}_{jm}$, $\hat{\phi}_{jm}$ be respectively the starts and duration of greens for specified signal-controlled junction m and the corresponding signal group j in the network, and

A_{jm} is the set of links a which is controlled by signal group j at junction m,

B_{jm} $=\{b$: *link b is the immediately downstream link for some link* $a \in A_{jm}\}$

and C_{jm} $=\{$ *c: link c is a further downstream link for some link* $a \in A_{jm}\}$ and

$L_{jm} = A_{jm} \cup B_{jm} \cup C_{jm}$.

The corresponding derivatives are

$$\sum_{a \in L_{jm}} W\, w_a \frac{\partial D_a}{\partial x} + K\, k_a \frac{\partial S_a}{\partial x}\ ; \quad \forall\, x \in \left\{ \hat{\theta}_{jm}\ ,\ \hat{\phi}_{jm} \right\} \tag{2.48a}$$

Since $D_a = D_a^U + D_a^{r+o}$ and $S_a = S_a^U + S_a^{r+o}$, (2.48a) can be written

$$\sum_{a \in L_{jm}} W\, w_a \left(\frac{\partial D_a^U}{\partial x} + \frac{\partial D_a^{r+o}}{\partial x}\right) + K\, k_a \left(\frac{\partial S_a^U}{\partial x} + \frac{\partial S_a^{r+o}}{\partial x}\right)\ ; \tag{2.48b}$$

$$\forall\, x \in \left\{ \hat{\theta}_{jm}\ ,\ \hat{\phi}_{jm} \right\}$$

Detailed expressions for the term of (2.48b) quoted in the following tables, i.e. Tables 2.1-2.7, are those derived by Wong for the case in which the queues for conditions of the uniform component on undersaturated links start to form at the start of the effective red.

2.5 Conclusion

In this Chapter, the optimisation problem of area traffic control and user equilibrium traffic assignment has been dealt with as a combined problem, i.e. the equilibrium network design problem. Various formulations and the corresponding solution methods for this equilibrium network design problem for area traffic control optimisation have been investigated respectively. The bi-level programming technique applied to the formulation of such problem has been considered as an appropriate approach in dealing with the dependence of equilibrium flow on signal timings. Furthermore, reviews of sensitivity analysis and the corresponding applications for the bi-level programme have been conducted. As we have mentioned before, with the background of rigorous mathematical sensitivity analysis for equilibrium network flows, the bi-level programme for the area traffic control optimisation problem regarded as a constrained optimisation problem can be solved by the techniques using derivatives in mathematical programming for optimisation problems. In the following chapters, the bi-level formulation for the optimisation of area traffic control with equilibrium network flows will be proposed to which the solution method in terms of the derivatives for system performance values with respect to the decision variables of interest will be discussed accordingly.

Table 2.1 Approximate expressions for the derivatives of the uniform components for upstream links; i.e. $\forall\, a \in A_{jm}$

$\dfrac{\partial D_a^U}{\partial \hat{\theta}_{m,n}}$	$\dfrac{\partial D_a^U}{\partial \hat{\phi}_{m,n}}$	$\dfrac{\partial S_a^U}{\partial \hat{\theta}_{m,n}}$	$\dfrac{\partial S_a^U}{\partial \hat{\phi}_{m,n}}$
$\zeta\,(s_a(z_{a2} - \bar{c}) - I_a(\phi_a)(z_{a2} - \phi_a))$	$-\zeta\, I_a(\phi_a)(z_{a2} - \phi_a)$	$\dfrac{\zeta\, s_a(I_a(z_{a2}) - I_a(\phi_a))}{s_a - I_a(z_{a2})}$	$\dfrac{-\zeta\, s_a\, I_a(\phi_a)}{s_a - I_a(z_{a2})}$

note: $\bar{c}$ is common cycle time; $\zeta = \dfrac{1}{\bar{c}}$.

θ_a , ϕ_a are respectively the starts and duration of effective greens for link a controlled by signal controlled group j at junction m in the network, i.e. $\dfrac{\partial}{\partial \theta_a} = \dfrac{\partial}{\partial \hat{\theta}_{jm}}$, $\dfrac{\partial}{\partial \phi_a} = \dfrac{\partial}{\partial \hat{\phi}_{jm}}$

$I_a(\bullet)$, q_a , s_a are respectively the arrival rate as represented by the IN profile, average flow and saturation flow for link a and z_{a1} , z_{a2} are respectively the starting and ending time in one cycle for queue formed on link a.

Table 2.2 Approximate expressions for the derivatives of the random plus oversaturation components for upstream links; i.e. $\forall\, a \in A_{jm}$

$\dfrac{\partial D_a^{r+o}}{\partial \hat{\theta}_{jm}}$	$\dfrac{\partial D_a^{r+o}}{\partial \hat{\phi}_{jm}}$	$\dfrac{\partial S_a^{r+o}}{\partial \hat{\theta}_{jm}}$	$\dfrac{\partial S_a^{r+o}}{\partial \hat{\phi}_{jm}}$
0	$\frac{1}{2}[(A_a \dfrac{\partial A_a}{\partial \hat{\phi}_{jm}} + \frac{1}{2}\dfrac{\partial B_a}{\partial \hat{\phi}_{jm}})(A_a^2 + B_a)^{-\frac{1}{2}} - \dfrac{\partial A_a}{\partial \hat{\phi}_{jm}}]$	0	$\zeta \exp(\dfrac{-\zeta\, D_a^{r+o}}{2q_a})\, \dfrac{\partial D_a^{r+o}}{\partial \phi_{m,n}}$

note: $\dfrac{\partial A_a}{\partial \hat{\phi}_{jm}} = \dfrac{\zeta\, s_a\, T\, [(\mu_a\, T)^2 - 2C(2 + x_a)\, \mu_a\, T - 4C\, L_a^0]}{2\,(\mu_a\, T - 2C)^2}$, $\dfrac{\partial B_a}{\partial \hat{\phi}_{jm}} = \dfrac{-2\zeta\, s_a\, T\, C\, (2\, L_a^0 + x_a\, \mu_a\, T)^2}{(\mu_a\, T - 2C)^2}$

$D_a^{r+o} = \frac{1}{2}\,[(A_a^2 + B_a)^{\frac{1}{2}} - A_a]$; $A_a = \dfrac{(1 - x_a)\,(\mu_a\, T)^2 + 2(2Cx_a - L_a^0)\, \mu_a\, T + 8C\, L_a^0}{2(\mu_a\, T - 2C)}$, $B_a = \dfrac{2C(2L_a^0 + x_a\, \mu_a\, T)^2}{\mu_a\, T - 2C}$

where μ_a , x_a , L_a^0 are respectively the capacity, degree of saturation and initial queue for link a over a period of analysis T, and C is a given constant.

Table 2.3 Approximate expressions for the derivatives of the uniform components for downstream links; i.e. $\forall\, b \in \boldsymbol{B}_{jm}$

$\dfrac{\partial D_b^U}{\partial x}$	$\dfrac{\partial S_b^U}{\partial x}$
$\zeta \displaystyle\int_{z_{b1}}^{z_{b2}} \frac{\partial I_b(t)}{\partial x} (z_{b2} - t)\, dt$	$\dfrac{\zeta\, s_b}{s_b - I_b(z_{b2})} \displaystyle\int_{z_{b1}}^{z_{b2}} \frac{\partial I_b(t)}{\partial x}\, dt$

$x \in \{ \hat{\theta}_{jm} , \hat{\phi}_{jm} \}$ and $I_b(\bullet)$, s_b , z_{b1} , z_{b2} are defined as in Table 2.1.

Table 2.4. The changes in the IN pattern on downstream links; i.e. $\forall\, b \in \boldsymbol{B}_{jm}$

$\dfrac{\partial I_b}{\partial \hat{\theta}_{jm}}$	$\dfrac{\partial I_b}{\partial \hat{\phi}_{jm}}$
$f_{ab}(I_a(\phi_a)R(y_1,t) - s_a R(y_2,t) + (s_a - I_a(\phi_a))R(y_3,t))$	$f_{ab} I_a(\phi_a)(R(y_1,t) - R(y_3,t))$

note: *for* $p=1,2,3$, $R(y_p, t) = \dfrac{\hat{F}\, e^{-\hat{F}(t - y_p)}}{1 - e^{\frac{-\hat{F}}{\zeta}}}$, *for* $y_p \le t \le y_p + \bar{c}$ *such that* $R(y_p, t + k\bar{c}) = R(y_p, t)$,

k is integer, f_{ab} is the proportion of traffic on upstream link a into downstream link b, and

$\hat{F} = \zeta\, F\, N_s$ is a modified smoothing factor used as a counterpart for F defined in TRANSYT,

N_s is the number of time steps into which the cycle time is divided for the evaluation of performance index. The values of y_p , for $p=1,2,3$ are given as

$y_1=(\theta_a + \phi_a + \tau_b - \theta_b)$ *mod* $\bar{c}$, $y_2=(\theta_a + \bar{c} + \tau_b - \theta_b)$ *mod* $\bar{c}$, $y_3=(\theta_a + z_{a2} + \tau_b - \theta_b)$ *mod* $\bar{c}$

where τ_b is 0.8 times the undelayed travel time on link b; where θ_b is defined as in Table 2.1.

Table 2.5 Approximate expressions for the derivatives of the random plus oversaturation components for downstream links;i.e. $\forall\, b \in \boldsymbol{B}_{jm}$

$\frac{\partial D_b^{r+o}}{\partial \hat{\theta}_{jm}}$	$\frac{\partial D_b^{r+o}}{\partial \hat{\phi}_{jm}}$	$\frac{\partial S_b^{r+o}}{\partial \hat{\theta}_{jm}}$	$\frac{\partial S_b^{r+o}}{\partial \hat{\phi}_{jm}}$
0	0 , *if upstream link a undersaturated* $\zeta f_{ab} s_a \frac{\partial D_b^{r+o}}{\partial q_b}$, *otherwise*	0	$\zeta \exp(\frac{-\zeta D_b^{r+o}}{2q_b}) \frac{\partial D_b^{r+o}}{\partial \hat{\phi}_{jm}}$

note: $\frac{\partial D_b^{r+o}}{\partial q_b} = \frac{1}{2}[(A_b \frac{\partial A_b}{\partial q_b} + \frac{1}{2}\frac{\partial B_b}{\partial q_b})(A_b^2 + B_b)^{-\frac{1}{2}} - \frac{\partial A_b}{\partial q_b}]$, $\frac{\partial A_b}{\partial q_b} = \frac{-\mu_b T^2 + 4CT}{2(\mu_b T - 2C)}$, $\frac{\partial B_b}{\partial q_b} = \frac{4CT(2L_b^0 + x_b \mu_b T)}{\mu_b T - 2C}$

A_b , B_b , μ_b , L_b^0 , x_b are defined as in Table 2.2.

Table 2.6. Approximate expressions for the derivatives of the uniform components for further downstream links; i.e. $\forall\, c \in \boldsymbol{C}_{jm}$

$\frac{\partial D_c^U}{\partial \hat{\theta}_{jm}}$	$\frac{\partial D_c^U}{\partial \hat{\phi}_{jm}}$	$\frac{\partial S_c^U}{\partial \hat{\theta}_{jm}}$	$\frac{\partial S_c^U}{\partial \hat{\phi}_{jm}}$
0	$\frac{1}{2}\zeta^2 f_{ac} s_a (z_{c2} - z_{c1})^2$	0	$\frac{\zeta^2 f_{ac} s_d s_c (z_{c2} - z_{c1})}{s_c - I_c(z_{c2})}$

note: where $I_c(\bullet)$, s_c , z_{c1} , z_{c2} are defined as in Table 2.1 and f_{ac} is defined as in Table 2.4.

Table 2.7 Approximate expressions for the derivatives of the random plus oversaturation components for further downstream links; i.e. $\forall\, c \in \boldsymbol{C}_{jm}$

$\frac{\partial D_c^{r+o}}{\partial \hat{\theta}_{jm}}$	$\frac{\partial D_c^{r+o}}{\partial \hat{\phi}_{jm}}$	$\frac{\partial S_c^{r+o}}{\partial \hat{\theta}_{jm}}$	$\frac{\partial S_c^{r+o}}{\partial \hat{\phi}_{jm}}$
0	0	0	0

Chapter 3 Problem Formulation

3.0 Introduction

In Chapter 3, the bi-level programming is used to formulate the optimisation problem of area traffic control for equilibrium network flows, in which the upper level problem deals with the optimisation of area traffic control with respect to the common cycle time, and the starts and durations of green times, whilst the lower level problem deals with the equilibrium network flows which corresponds to solving the user equilibrium traffic assignment. In the following sections, firstly, a general description of the bi-level formulation for equilibrium network flows is given in Section 3.1. Fundamental terminology and notation for the optimisation of area traffic control will be then given in Sections 3.2 and 3.3 respectively. Secondly, in relation to the formulation of the upper level problem, in Section 3.4, the mathematical expressions of the components for the system performance which are in terms of a weighted linear combination of mean delay and number of stops per unit time adopted from TRANSYT will be discussed. Thirdly, in Section 3.5, the formulation for user equilibrium traffic assignment in the lower level problem will be discussed, in which a separable link travel time function is determined by the sum of the undelayed travel time, i.e. travel time in prevailing traffic conditions if not delayed by downstream signals, and the average junction delay incurred by traffic at the downstream end of the link. Conclusions for this Chapter will be given in Section 3.6.

3.1 The Bi-level Problem

A bi-level formulation for equilibrium network flows problem can be described in the following way.

$$\begin{aligned} &\underset{\psi}{Minimise} \quad Z = Z_0(\psi\ ,\ q^{\star}(\psi)) \\ &subject\ \ to \\ &\qquad\qquad \psi \in S_0 \end{aligned} \tag{3.0}$$

In problem (3.0) the total travel cost Z can be minimised via function Z_0 with

respect to the decision variable ψ , which can be determined in the upper level problem whilst the equilibrium network flows q are determined in the lower level problem via function $q^{\star}$ which is in terms of the decision variable ψ , and S_0 is the feasible set of decision variable ψ .

3.2 Terminology

1. *Traffic streams*: Traffic streams are the basic segments which the approaching traffic on a particular road is considered to form when it enters a road junction. Then it can be identified either as two or more separate streams when disjoint sets of one or more lanes leading to the junction are exclusively used by all vehicles taking particular directions or as a single stream when no such disjoint sets of lanes can be identified.

2. *Approaches*: Approaches are sets of lanes used to enter a junction by vehicles in the corresponding streams.

3. *Signal groups and phases*: Sets of traffic signals that always show identical aspects and the sets of traffic streams that are controlled by them.

4. *Compatibility*: For the purpose of safety, the two streams are compatible if they can safely have right of way simultaneously and the two signal groups are compatible if and only if every stream controlled by one group is compatible with every stream controlled by the other group.

5. *Cycle time*: The shortest duration over which the displays of all signal groups at a junction are repeated identically.

6. *Stage*: The period during which a particular set of mutually compatible signal groups display green.

7. *Interstage*: The interval between any two consecutive stages during which some of the signal groups that displayed green in the preceding stage cease to do so and some

of those that displayed red in the previous stage begin to show green. The exact timing of various changes of signal aspects during it is called the interstage structure.

8. *Compatibility graph*: A graph composed of connected vertices which denote the mutually compatible signal groups for a junction.

9. *Maximal clique*: A maximal subset of vertices in the compatibility graph which are fully connected by arcs to one another.

10. *Sequence:* A sequence is formed by a series of the maximal cliques such that
(a). every signal group appears in at least one maximal clique.
(b). the cliques in which any particular signal group appears are consecutive ones in the sequence.

11. *Saturation flow*: The maximum average discharging rate at which vehicles pass the stop line from an approach as long as there is traffic queuing on the approach.

12. *Effective green time*: The period during which the traffic in a stream is regarded as passing the signals at the saturation flow rate if there is traffic queuing on the approach or otherwise passing the signal as it arrives.

13. *Effective red time*: The period equals to the cycle time minus the effective green time during which no vehicles are regarded in passing the signals.

14. *Green start lag*: The time from the start of actual green to that of effective green.

15. *Green end lag*: The time from the end of actual green to that of effective green.

16. *Link*: Representing a basic segment of traffic moving between two signal-controlled junctions and for which the performance index is measured.

17. *Upstream links*: The links leading towards a particular junction and feeding vehicles into each traffic stream approaching the junction.

18. *Downstream links*: The links leading away from a particular junction.

19. *Rate of delay:* The average delay incurred per unit time by traffic on a link.

20. *Number of stops per unit time*: The rate at which vehicles are obliged to stop on a link.

21. *Performance index*: A weighted linear combination of the rate of delay and the number of stops per unit time over all links.

22. *Local clock*: A basis on which signal timings are measured at a particular junction within a fixed-time controlled network.

23. *Master clock*: A common time clock on the basis of which the signal timings at all junctions in a fixed-time controlled network are measured.

24. *Offset at a junction*: The time difference between the master clock origin and the next local clock origin of the junction concerned.

25. *Time span:* A period forming the scope of this study and composed of a series of consecutive time slices during each of which fixed signal settings are to operate.

3.3 Notation

Consider a signal-controlled road network for which a given timing plan is implemented in each of the consecutive time slices forming the whole time span. The following notation can be stated first.

3.3.1. General

Let $\boldsymbol{T}$ be the time span comprising a series of time slices denoted by

$$\boldsymbol{T}^i = \left[t^{\,i}\ ,\ t^{i+1}\right] \text{ ; we have } \boldsymbol{T} = \bigcup_i \boldsymbol{T}^i = \bigcup_i \left[t^{\,i}\ ,\ t^{i+1}\right] .$$

Let $\boldsymbol{G}(\ \boldsymbol{N}\ ,\ \boldsymbol{L}\ ,\ \hat{\boldsymbol{R}}\ ,\ \hat{\boldsymbol{S}}\ )$ be a directed road network, where $\boldsymbol{N}$ is a set of N_J nodes

each of which represents one of N_J fixed-time signal controlled junctions, that is,

$\boldsymbol{N} = \{\ 1,\ 2,\ .....,\ N_J\}$, $\boldsymbol{L}$ is a set of N_L links, that is, $\boldsymbol{L} = \{\ 1,\ 2,\ .....,\ N_L\}$

where each traffic stream approaching any junction is represented by its own link,

$\hat{R}$ is a set of origin nodes and $\hat{S}$ is a set of destination nodes.

For each signal controlled junction m in N, let L_m be the corresponding set of the links leading to that junction.

3.3.2. Signal timings

The following notation is used and is supposed to be fixed within any one given time slice. For each signal controlled junction m in N : let N_{pm}, N_{sm} be respectively the numbers of signal groups and stages at junction m, and let

$P_m = \{ 1, 2, ..., N_{pm} \}$ be the set of numbers of signal groups at junction m,

$S_m = \{ 1, 2, ..., N_{sm} \}$ be the set of numbers of stages at junction m,

$\hat{A}_m = \{ \hat{a}_{ijm}, i \in S_m, j \in P_m ; \hat{a}_{ijm} = 1 \ or \ 0 \}$ be the stage-signal group incidence matrix at junction m, where $\hat{a}_{ijm} = 1$ if signal group j has green in the stage i at junction m, and $\hat{a}_{ijm} = 0$ otherwise; and

$\hat{B}_m = \{ \hat{b}_{ajm}, a \in L_m, j \in P_m ; \hat{b}_{ajm} = 1 \ or \ 0 \}$ be the link-signal group incidence matrix at junction m, where $\hat{b}_{ajm} = 1$ if the traffic stream a is controlled by the signal group j at junction m, and $\hat{b}_{ajm} = 0$ otherwise; and

g_{jm} be the specified minimum green for signal group j at junction m, $\forall j \in P_m$,

h_{im} be the specified minimum green for stage i at junction m, $\forall i \in S_m$,

u_m be a common multiplier applied to the average arrival flow for each link a in L_m leading to the junction m,

ω_{im} be the time at which the stage i begins at junction m expressed as a proportion of common cycle time relative to an arbitrary time origin to the network as a whole, $\forall\, i \in \boldsymbol{S}_m$,

$\hat{c}_{jlm}$ be the clearance time between the end of green for signal group j and the start of green for incompatible signal group l at junction m, for $j \neq l$, $\forall\, j$, $l \in \boldsymbol{P}_m$, and

$\Omega_m(\, j\ ,\ l\,)$ be the successor function which is defined by Heydecker (1992), a collection of numbers 0 and 1 for each pair of incompatible signal groups at junction m; where $\Omega_m(\, j\ ,\ l\,) = 0$, if the start of green for the signal group j precedes that of l, and $\Omega_m(\, j\ ,\ l\,) = 1$ otherwise, $\forall\, j$, $l \in \boldsymbol{P}_m$.

For the whole road network $\boldsymbol{G}$, let

ζ be the reciprocal of the common cycle time $\bar{c}$,

$\theta = [\ \theta_{jm}\ ;\ \ \forall\, j \in \boldsymbol{P}_m\ ,\ \forall\, m \in \boldsymbol{N}\ ]$ be the vector of starts of green, where element θ_{jm} is start of next green for signal group j at junction m as a proportion of common cycle time relative to an arbitrary time origin to the network as a whole, and

$\phi = [\ \phi_{jm}\ ;\ \ \forall\, j \in \boldsymbol{P}_m\ ,\ \forall\, m \in \boldsymbol{N}\ ]$ be the vector of durations of green, where element ϕ_{jm} is the duration of green as a proportion of common cycle time for signal group j at junction m.

For each link a in $\boldsymbol{L}$, let

e be the extra effective green for all links which equals green end lag e_2 minus green start lag e_1 and it can be denoted as $e = e_2 - e_1$,

$\bar{\theta}_a$ be start of next effective green for link a expressed as a proportion of common cycle time relative to an arbitrary time origin to the network as a whole, the relationship between $\bar{\theta}_a$ and θ_{jm} is $\bar{\theta}_a = \sum_{jm} \hat{b}_{ajm}\theta_{jm} + e_1\zeta \quad \forall \; a \in \boldsymbol{L}$,

Λ_a be the duration of effective green for link a, which is expressed as a proportion of common cycle time, the relationship between Λ_a and ϕ_{jm} is

$$\Lambda_a = \sum_{jm} \hat{b}_{ajm}\, \phi_{jm} + e\,\zeta \;, \quad \forall\, a \in \boldsymbol{L} \;, \text{ and}$$

p_a be the maximum degree of saturation for link a.

3.3.3. Flows

For each link a in $\boldsymbol{L}$, the traffic conditions in each timeslice $\boldsymbol{T}^i$, $t \in \boldsymbol{T}^i$ are described by the following indicators. Let

$I_a(t)$ be the arrival rate of traffic at the downstream end of link a at time t,

$\bar{I}_a(t)$ be the cumulative amount of traffic arriving at the downstream end of link a over the time interval $[\, t^i \,,\, t \,]$,

$\boldsymbol{q} = [\; q_a \;;\; \forall\, a \in \boldsymbol{L} \;]$ be the vector of the average flow q_a on link a,

s_a be the saturation flow on link a,

μ_a be capacity for link a, which can be expressed by definition as $\mu_a = \Lambda_a s_a$, and

x_a be the degree of saturation on link a, which can be expressed by definition as

$$x_a = \frac{q_a}{\mu_a} \;.$$

3.3.4. Indicators of traffic conditions

Four indicators of traffic conditions are used, each of which is a function of

t , $t \in T^i$. For each link a in L , let

$D_a(t)$ be the rate of delay on link a over the interval $[\, t^i \,,\, t \,]$,

$d_a(t)$ be the average delay to a vehicle arriving on link a during the interval $[t^i, t]$,

$L_a(t)$ be the number of queuing vehicles on link a at time t, and

$S_a(t)$ be the number of stops per unit time on link a over the interval $[\, t^i \,,\, t \,]$.

3.3.5. User equilibrium traffic assignment

For the road network $G(N \,,\, L \,,\, \hat{R} \,,\, \hat{S}\,)$, let

$W = \{\, w = (\, r \,,\, s \,) \;;\; \forall\, r \in \hat{R} \,,\; \forall\, s \in \hat{S} \,\}$ be the set of origin-destination pairs.

P_w be the set of paths between the origin-destination pair w in W ,

$\hat{D} = [\, \hat{D}_w \;;\; \forall\, w \in W \,]$ be the vector of travel demands, where element $\hat{D}_w$ is the travel demand between origin-destination pair w,

$f = [\, f_p \;;\; \forall\, p \in P_w \,,\; \forall\, w \in W \,]$ be the vector of path flows, where element f_p is traffic flow on path p ,

$c^0 = [\, c_a^0 \;;\; \forall\, a \in L \,]$ be the vector of undelayed link travel times, where element c_a^0 is the undelayed travel time for link a, and

$c(\, q \,,\, \psi\,) = [\, c_a(\, q_a \,,\, \psi) \,]$ be the vector of link travel times, where element $c_a(q_a \,,\, \psi)$ is travel time on link a as a function of flow on the link itself and the

signal setting variables, $\psi = (\zeta, \theta, \phi)$, which is the sum of the undelayed travel time under prevailing traffic condition, i.e. c_a^0 and the average delay to a vehicle at the end of the link leading to the downstream signal-controlled junction over the timeslice $[t^i, t]$, i.e. $d_a(t)$, therefore by definition for each link a in $\boldsymbol{L}$, if $c_a(q_a, \psi)$ is averaged over the interval $[t^i, t]$ then

$$c_a(q_a, \psi) = c_a^0 + d_a(t), \ \forall a \in \boldsymbol{L} .$$

Let $\delta = [\delta_{ap}; \forall a \in \boldsymbol{L}, \forall p \in \boldsymbol{P}_w, \forall w \in \boldsymbol{W}]$ be the link/path incidence matrix where $\delta_{ap} = 1$ if link a is in path p, and $\delta_{ap} = 0$ otherwise,

$\Delta = [\Delta_{wp}; \forall p \in \boldsymbol{P}_w, \forall w \in \boldsymbol{W}]$ be the origin-destination/path incidence matrix where $\Delta_{wp} = 1$ if path p connects origin-destination pair w, and $\Delta_{wp} = 0$ otherwise, and

$\boldsymbol{C} = [C_p; \forall p \in \boldsymbol{P}_w, \forall w \in \boldsymbol{W}]$ be the vector of path travel times, where element C_p is the travel time on path p, and the relationship between C_p and $c_a(q_a, \psi)$ is $C_p = \sum_{a \in L} \delta_{ap} c_a(q_a, \psi)$ and in vector form $\boldsymbol{C}^T = \delta^T \boldsymbol{c}(q, \psi)^T$ where the superscript T is the matrix transpose operator.

3.4 The Upper Level Problem

The optimisation of area traffic control at the upper level for bi-level problem (3.0) can be expressed as follows. Let P be a performance index which is a function

P_0 of ζ , θ , ϕ , q and is to be minimised with respect to $\psi = (\zeta , \theta , \phi)$ for given q subject to $\psi \in S_0$. Then the upper level problem can be rewritten as to

$$\begin{aligned} &\underset{\psi}{Minimise} \quad P = P_0(\psi , q) \\ &subject \quad to \\ &\qquad \psi \in S_0 \end{aligned} \tag{3.1}$$

3.4.1. Objective function

The function P_0 for the performance index P in problem (3.1) is taken to be a linear combination of the rate of delay and the number of stops per unit time for each traffic stream, averaged over the interval $[t^i , t]$. This is adopted from TRANSYT and then can be expressed as follows.

$$P_0 = \sum_{a \in L} [D_a(t) W_{aD} M_D + S_a(t) W_{aS} M_S] \tag{3.2}$$

where W_{aD} , W_{aS} are respectively link-specific weighting factors for rate of delay and number of stops per unit time and M_D , M_S are corresponding monetary factors common to all links.

3.4.2. Constraints

In relation to the feasible region S_0 for the signal setting variables in problem (3.1), i.e. the common cycle time, and the starts and durations for all signal groups, the corresponding constraints are expressed in the following way.

1. The constraints on timings for phases.

For the whole road network G :

a. The cycle time constraint.

$$\zeta_{min} \le \zeta \le \zeta_{max} \tag{3.3}$$

For each signal controlled junction m in $\boldsymbol{N}$:

b. The phase green time constraints.

$$g_{jm}\,\zeta \le \phi_{jm} \le 1\ ,\quad \forall\, j \in \boldsymbol{P}_m \tag{3.4}$$

c. The capacity constraints.

$$u_m\, q_a \le p_a\, s_a\, \Lambda_a\ ,\quad \forall\, a \in \boldsymbol{L}_m \tag{3.5}$$

d. The clearance time constraints.

$$\theta_{jm} + \phi_{jm} + \hat{c}_{jlm}\,\zeta \le \theta_{lm} + \Omega_m(j\ ,\ l)\ ;\ j \ne l\ ,\quad \forall\, j\ ,\ l \in \boldsymbol{P}_m \tag{3.6}$$

2. The constraints on durations of stages.

For each signal controlled junction m in $\boldsymbol{N}$, the beginning θ_{1m} of green for the first signal group at junction m is

$$0 \le \theta_{1m} < 1 \tag{3.7}$$

For each signal group $j \in \boldsymbol{P}_m \setminus \{\ 1\ \}$, it follows that

$$\begin{aligned} &0 \le \theta_{jm} < \theta_{1m} \quad \textit{and} \quad \Omega_m(1\ ,\ j) = 1 \\ &\theta_{1m} \le \theta_{jm} < 1 \quad \textit{and} \quad \Omega_m(1\ ,\ j) = 0 \end{aligned} \tag{3.8}$$

Let stage 1 at junction m be the stage in which green for group 1 begins, and let

$$\begin{aligned} &\omega_{1m} = \theta_{1m} + 1 \\ &\theta_{1m} < \omega_{im} < \theta_{1m} + 1\ ;\ \forall\, i \in \boldsymbol{S}_m \setminus \{\ 1\ \} \end{aligned} \tag{3.9}$$

Therefore the following four sets of inequalities are equivalent to the requirement that the duration of stage $i \ge h_{im}$, $\forall\, i \in \boldsymbol{S}_m$.

1. For any signal group j which does not have green in stage 1, but does have green in stage $i \in \boldsymbol{S}_m \setminus \{\ 1\ \}$ (i.e. for any $i \in \boldsymbol{S}_m \setminus \{\ 1\ \}$, and $j \in \boldsymbol{P}_m$ such that $\hat{a}_{1jm} = 0$ and $\hat{a}_{ijm} = 1$) according to expressions (3.7)-(3.9),

$$\theta_{jm} + \Omega_m(1, j) \leq \omega_{im} \leq \omega_{im} + h_{im}\zeta \leq \theta_{jm} + \phi_{jm} + \Omega_m(1, j) \leq \omega_{1m} \tag{3.10}$$

2. For any signal group j which does have green in the stage 1 (i.e. for any $j \in \boldsymbol{P}_m$ such that $\hat{a}_{1jm} = 1$) according to expressions (3.7)-(3.9),

$$\theta_{jm} + \Omega_m(1, j) \leq \omega_{1m} \leq \omega_{1m} + h_{1m}\zeta \leq \theta_{jm} + \phi_{jm} + \Omega_m(1, j) \tag{3.11}$$

3. For any signal group j which has green in the stage 1 and in the $k - 1$ subsequent consecutive stages as well (i.e. for any $k \in \boldsymbol{S}_m \setminus \{ 1 \}$ and $j \in \boldsymbol{P}_m$ such that $\prod_{i=1}^{k} \hat{a}_{ijm} = 1$) according to expressions (3.7)-(3.9),

$$\theta_{jm} + \Omega_m(1, j) - 1 \leq \omega_{1m} - 1 \leq \omega_{km} + h_{km}\zeta \leq \theta_{jm} + \phi_{jm} + \Omega_m(1, j) - 1 \tag{3.12}$$

4. For any signal group j which has green in the stage 1 and in all the preceding consecutive stages beginning at stage k as well (i.e. for any $k \in \boldsymbol{S}_m \setminus \{ 1 \}$ and $j \in \boldsymbol{P}_m$ such that $\hat{a}_{1jm} \prod_{i=k}^{N_{sm}} \hat{a}_{ijm} = 1$) according to expressions (3.7)-(3.9),

$$\theta_{jm} + \Omega_m(1, j) \leq \omega_{km} \leq \omega_{1m} + h_{1m}\zeta \leq \theta_{jm} + \phi_{jm} + \Omega_m(1, j) \tag{3.13}$$

3.4.3. Calculations for indicators of traffic conditions

Since there is no available closed mathematical form for the TRANSYT traffic model for the estimation of the indices of traffic conditions for each link, three cases which represent the relevant patterns of queue formations in TRANSYT are discussed in order to derive approximate mathematical expressions. For each of these, the indicator is made up of the uniform component and the random plus oversaturation component.

To calculate the uniform component of the indicator, a link-based formulation is used to calculate the indicator values during a particular timeslice when a specific timing plan for the whole time period is given. In the following subsections, the sheared delay formulae for traffic indicators are used (Allsop 1993 following Kimber and Hollis 1979), in each of which initial queue length is included. The uniform component which is evaluated by the traffic model in TRANSYT is approximated as convenient mathematical expressions. For specification of these mathematical expressions during some specific timeslice, all links are classified into the following three groups by the degree of saturation and initial queue length at the start of this timeslice: the oversaturated links, and the undersaturated links with and without initial queues at the start of this timeslice.

3.4.3.1. Random plus oversaturation component

As for the random plus oversaturation component of the indicators, sheared delay formulae derived on the basis of the steady state and the deterministic solutions are used and summarized in Table 3.1. Detailed discussions are as follows.

1. Steady state solutions

Given the timeslice T^i , for each link a in L with $x_a < 1$, let

L_a^e , D_a^e , d_a^e be respectively the steady state expressions for the number of queuing vehicles, the rate of delay and average delay to a vehicle on link a when $x_a < 1$ and C be a given constant approximately equal to 0.5 ~ 0.6 for a signal-controlled traffic stream. We then have the following expressions for the indicators.

$$L_a^e = \frac{C\, x_a^2}{1 - x_a} + x_a \tag{3.14}$$

$$D_a^e = \frac{C\, x_a^2}{1 - x_a} + x_a \tag{3.15}$$

$$d_a^e = \frac{1}{\mu_a} \left(\frac{C\, x_a}{1 - x_a} + 1\right) \tag{3.16}$$

Table 3.1. Expressions of random plus oversaturation component for each link $a \in \boldsymbol{L}$

$L_a^{r+o}(t)$	$D_a^{r+o}(t)$	$S_a^{r+o}(t)$	$d_a^{r+o}(t)$
$\frac{\sqrt{A_1^2 + B_1} - A_1}{2}$	$\frac{\sqrt{A_2^2 + B_2} - A_2}{2}$	$2q_a(1 - \exp(\frac{-\zeta(\sqrt{A_2^2 + B_2} - A_2)}{4q_a}))$	$\frac{\sqrt{A_3^2 + B_3} - A_3}{2}$

where $A_1 = \dfrac{(1 - x_a)(\mu_a(t - t^i))^2 + (1 - L_a^{r+o}(t^i))\mu_a(t - t^i) - 2(1 - C)(L_a^{r+o}(t^i) + x_a\mu_a(t - t^i))}{\mu_a(t - t^i) + 1 - C}$

$$B_1 = \frac{4[L_a^{r+o}(t^i) + x_a\mu_a(t - t^i)][\mu_a(t - t^i) - (1 - C)(L_a^{r+o}(t^i) + x_a\mu_a(t - t^i))]}{\mu_a(t - t^i) + 1 - C}$$

$$A_2 = \frac{(1 - x_a)(\frac{\mu_a}{2}(t - t^i))^2 + (1 - L_a^{r+o}(t^i))\frac{\mu_a}{2}(t - t^i) - 2(1 - C)(L_a^{r+o}(t^i) + x_a\frac{\mu_a}{2}(t - t^i))}{\frac{\mu_a}{2}(t - t^i) + 1 - C}$$

$$B_2 = \frac{4\,[L_a^{r+o}(t^i) + x_a\frac{\mu_a}{2}(t - t^i)][\frac{\mu_a}{2}(t - t^i) - (1 - C)(L_a^{r+o}(t^i) + x_a\frac{\mu_a}{2}(t - t^i))]}{\frac{\mu_a}{2}(t - t^i) + 1 - C}$$

$$A_3 = \frac{(1 - x_a)(t - t^i)}{2} - \frac{L_a^{r+o}(t^i) - C + 2}{\mu_a}$$

$$B_3 = \frac{2[(1 - x_a)\mu_a(t - t^i) + C\,x_a\mu_a(t - t^i) - 2(1 - C)(L_a^{r+o}(t^i) + 1)]}{\mu_a^2}$$

where $t \in \boldsymbol{T}^i$, *and* C is a given constant.

2. Deterministic solutions

Given the timeslice T^i *and* $t \in T^i$, for each link a in L with $x_a \geq 1$, let $L_a^o(t)$, $D_a^o(t)$, $d_a^o(t)$ be respectively the number of queuing vehicles at time t , the rate of delay over the period $[\, t^i , t \,]$ and average delay to a vehicle in $[\, t^i , t \,]$ on link a for oversaturation when $x_a \geq 1$. We then have the following expressions for the indicators in the absence of randomness in arrivals or departures.

$$L_a^o(t) = L_a^o(t^i) + \mu_a(x_a - 1)(t - t^i) \tag{3.17}$$

$$D_a^o(t) = L_a^o(t^i) + \frac{\mu_a(x_a - 1)(t - t^i)}{2} \tag{3.18}$$

$$d_a^o(t) = \frac{L_a^o(t^i) + 1}{\mu_a} + \frac{(x_a - 1)(t - t^i)}{2} \tag{3.19}$$

3. Sheared delay formulae

For each link a in L and for $t \in T^i$, let $L_a^{r+o}(t)$, $D_a^{r+o}(t)$, $d_a^{r+o}(t)$ be respectively the random plus oversaturation expressions for the number of queuing vehicles at time t , the rate of delay over the period $[\, t^i , t \,]$ and average delay to a vehicle in $[\, t^i , t \,]$ on link a. We then have the following results.

$$L_a^{r+o}(t) = \frac{\sqrt{A_1^2 + B_1} - A_1}{2} \tag{3.20}$$

$$D_a^{r+o}(t) = \frac{\sqrt{A_2^2 + B_2} - A_2}{2} \tag{3.21}$$

$$d_a^{r+o}(t) = \frac{\sqrt{A_3^2 + B_3} - A_3}{2} \tag{3.22}$$

$$S_a^{r+o}(t) = 2\, q_a \left(1 - \exp\left(\frac{-D_a^{r+o}(t)\, \zeta}{2\, q_a}\right)\right) \tag{3.23}$$

where

$$A_1 = \frac{(1 - x_a)(\mu_a(t - t^i))^2 + (1 - L_a^{r+o}(t^i))\mu_a(t - t^i)}{\mu_a(t - t^i) + 1 - C} -$$

$$\frac{2(1 - C)(L_a^{r+o}(t^i) + x_a\mu_a(t - t^i))}{\mu_a(t - t^i) + 1 - C}$$

$$B_1 = \frac{4[L_a^{r+o}(t^i) + x_a\mu_a(t - t^i)][\mu_a(t - t^i)}{\mu_a(t - t^i) + 1 - C} -$$

$$\frac{(1 - C)(L_a^{r+o}(t^i) + x_a\mu_a(t - t^i))]}{\mu_a(t - t^i) + 1 - C}$$

$$A_2 = \frac{(1 - x_a)(\frac{\mu_a}{2}(t - t^i))^2 + (1 - L_a^{r+o}(t^i))\frac{\mu_a}{2}(t - t^i)}{\frac{\mu_a}{2}(t - t^i) + 1 - C} -$$

$$\frac{2(1 - C)(L_a^{r+o}(t^i) + x_a \frac{\mu_a}{2}(t - t^i))}{\frac{\mu_a}{2}(t - t^i) + 1 - C}$$

$$B_2 = \frac{4\ [L_a^{r+o}(t^i) + x_a\frac{\mu_a}{2}(t - t^i)][\frac{\mu_a}{2}(t - t^i)}{\frac{\mu_a}{2}(t - t^i) + 1 - C} -$$

$$\frac{(1 - C)(L_a^{r+o}(t^i) + x_a\frac{\mu_a}{2}(t - t^i))]}{\frac{\mu_a}{2}(t - t^i) + 1 - C}$$

$$A_3 = \frac{(1 - x_a)(t - t^i)}{2} - \frac{L_a^{r+o}(t^i) - C + 2}{\mu_a}$$

$$B_3 = \frac{2[(1 - x_a)\mu_a(t - t^i) + C\ x_a\mu_a(t - t^i) - 2(1 - C)(L_a^{r+o}(t^i) + 1)]}{\mu_a^2}$$

Expression (3.23), for the number of stops per unit time is as used in TRANSYT (Vincent *et al.*, 1980).

In the expressions (3.20)-(3.23), for A_n , B_n ; $n = 1, 2, 3$, the quantities that depend on the signal timings in timeslice T^i are

$$\mu_a = \Lambda_a s_a$$

$$x_a = \frac{q_a}{\Lambda_a s_a}$$

3.4.3.2. Uniform component

Because the timeslice during which particular timing plans regulate the operation of the signals are not in general multiples of the relevant cycle times and changes between successive timing plans are not simultaneous for all signal groups, link-based time slices are defined below within which the uniform component of the indicators of traffic conditions can be calculated at the basis of whole cycles. Therefore, the calculation of the uniform component of the indicators for queuing and delay on each link in a given timeslice is based on whole cycles which begin and end at the starts of effective red times. However, by doing so, this will leave out end-effects which will occur on each link as a result of different timing plans operating in successive time slices as we have recognized but not attempted to investigate in this work. Regarding to the specific form of the link-based timeslice, a revised time period is defined for each link as follows. For each link a in L , let $T_a = \left[t_a^0 , t_a^n \right]$ be a revised time period within the timeslice T^i , where t_a^0 *and* t_a^n are given by

$$t_a^0 = \frac{\bar{\theta}_a + \Lambda_a + n_{\min}}{\zeta}$$

$$t_a^n = \frac{\bar{\theta}_a + \Lambda_a + n_{\max}}{\zeta} \qquad (3.24)$$

where $0 \le \theta_a \le 1$, n is an integer and

$$n_{\min} = \textit{Minimise} \left\{ n : \frac{\bar{\theta}_a + \Lambda_a + n}{\zeta} \in T^i \right\}$$
$$n_{\max} = \textit{Maximise} \left\{ n : \frac{\bar{\theta}_a + \Lambda_a + n}{\zeta} \in T^i \right\} \tag{3.25}$$

The uniform component of the indicators of traffic conditions with respect to each link a in $\boldsymbol{L}$ will be calculated on the basis of $\boldsymbol{T}_a$.

By adopting the traffic flow patterns described in the traffic model from TRANSYT as cyclic flow profiles, we can make the estimation for the uniform component of the indicators for each link a in $\boldsymbol{L}$ during $\boldsymbol{T}_a$ on a single cycle basis because any accumulated queue for link a has been taken into account in the random plus oversaturation component. Furthermore, we assume that $I_a(t) \leq s_a$, $\forall \; t \in \boldsymbol{T}$, $\forall \; a \in \boldsymbol{L}$ so that the flow approaching the downstream signal cannot exceed the saturation flow at any instant in the whole time span.

Within a given timeslice $\boldsymbol{T}^i$ and for each link a in $\boldsymbol{L}$, the uniform component for the indicators of traffic conditions are calculated as averages over the period $\boldsymbol{T}_a$ for that link. The corresponding mathematical expressions are given below according to the conditions of the degree of saturation and initial queue length at the start of this given time period for each link. Three groups of links are discussed, that is, for the oversaturated links with $x_a \geq 1$, for the undersaturated links with $x_a < 1$ and without accumulated queues, which can be identified as those with $L_a^{r+o}(t_a^0) \leq L_a^e$, and for undersaturated links with accumulated queues, which can be identified as those with $L_a^{r+o}(t_a^0) > L_a^e$.

1. For oversaturated links

$$Let \quad \bar{E}_a(t) = \bar{E}_a(t_a^0) + q_a(t - t_a^0) \quad , \qquad for \quad t_a^0 \le t \le t_a^n$$

$$where \qquad q_a = \frac{\bar{I}_a(t_a^n) - \bar{I}_a(t_a^0)}{t_a^n - t_a^0}$$

According to expression (a.9) in *Appendix A*, $\bar{I}_a(t)$ and $\bar{E}_a(t)$ are two non-decreasing continuous fitted functions over $\boldsymbol{T}_a$. Furthermore, $\bar{I}_a(t)$ is a <u>*Cyclic Accumulative Function*</u> (see <*Definition a.5*> *in Appendix A*) on $\mathbb{R}^+$ (with period $\frac{1}{\zeta}$) over $\boldsymbol{T}_a$ with $t_a^n - t_a^0 = \frac{n}{\zeta}$ and by the results from <*Lemma a.6*> <u>in</u> *Appendix A*, $\int_{\bar{I}_a(t_a^0)}^{\bar{I}_a(t_a^n)} \bar{I}_a^{-1}(w)\, dw = \int_{\bar{I}_a(t_a^0)}^{\bar{I}_a(t_a^n)} \bar{E}_a^{-1}(w)\, dw$. In relation to the calculations of the uniform component of the number of queueing vehicles, L_a^U , and the rate of delay, D_a^U , we suppose that the effect of the cyclic variation in the cumulative arrivals $\bar{I}_a(t)$ on L_a^U and D_a^U is zero; therefore the uniform component for L_a^U and D_a^U can be calculated on the basis of the difference between the cyclic cumulative departure graph from TRANSYT model and the uniform departure rate μ_a as represented by the shaded areas in Figure 3.1, in which a selected interval is illustrated, i.e. $\left[t_a^0 \, , \, t_a^0 + \frac{1}{\zeta} \right] \subset \boldsymbol{T}_a$, and the results are given in expressions (3.26)-(3.27). In addition, the uniform component of the number of vehicle stops per

unit time, S_a^U , can be calculated on the basis of one stop per vehicle as $S_a^U = \dfrac{\bar{I}_a(t_a^n) - \bar{I}_a(t_a^0)}{t_a^n - t_a^0} = q_a$ as shown in expression (3.28), and all other stops are included in S_a^{r+o} as shown in expression (3.23).

Furthermore, in relation to the calculation for the uniform component of average delay d_a^U , the possible effect of the cyclic variation in the cumulative arrivals $\bar{I}_a(t)$ can be demonstrated also to be zero by the <*Lemma a.4*> in *Appendix A*, so calculation for d_a^U is also based on the shaded areas as shown in Figure 3.1. Therefore, the uniform component of the indicators of traffic conditions for each link a in $\boldsymbol{L}$ in the time period $\boldsymbol{T}_a$ can be expressed in the following ways.

$$L_a^U = \frac{\mu_a (1 - \Lambda_a)}{2\zeta} \tag{3.26}$$

$$D_a^U = \frac{\mu_a (1 - \Lambda_a)}{2\zeta} \tag{3.27}$$

$$S_a^U = q_a \tag{3.28}$$

$$d_a^U = \frac{(1 - \Lambda_a)}{2\zeta} \tag{3.29}$$

2. For undersaturated links and without accumulated queues

According to the cyclic IN-profile, $\bar{I}_a(t)$ is a continuous <u>*Cyclic Accumulative Function*</u> over each timeslice $\boldsymbol{T}^i$, and for each link a in $\boldsymbol{L}$ we can calculate the uniform component of the indicators of traffic conditions as averaged over the time

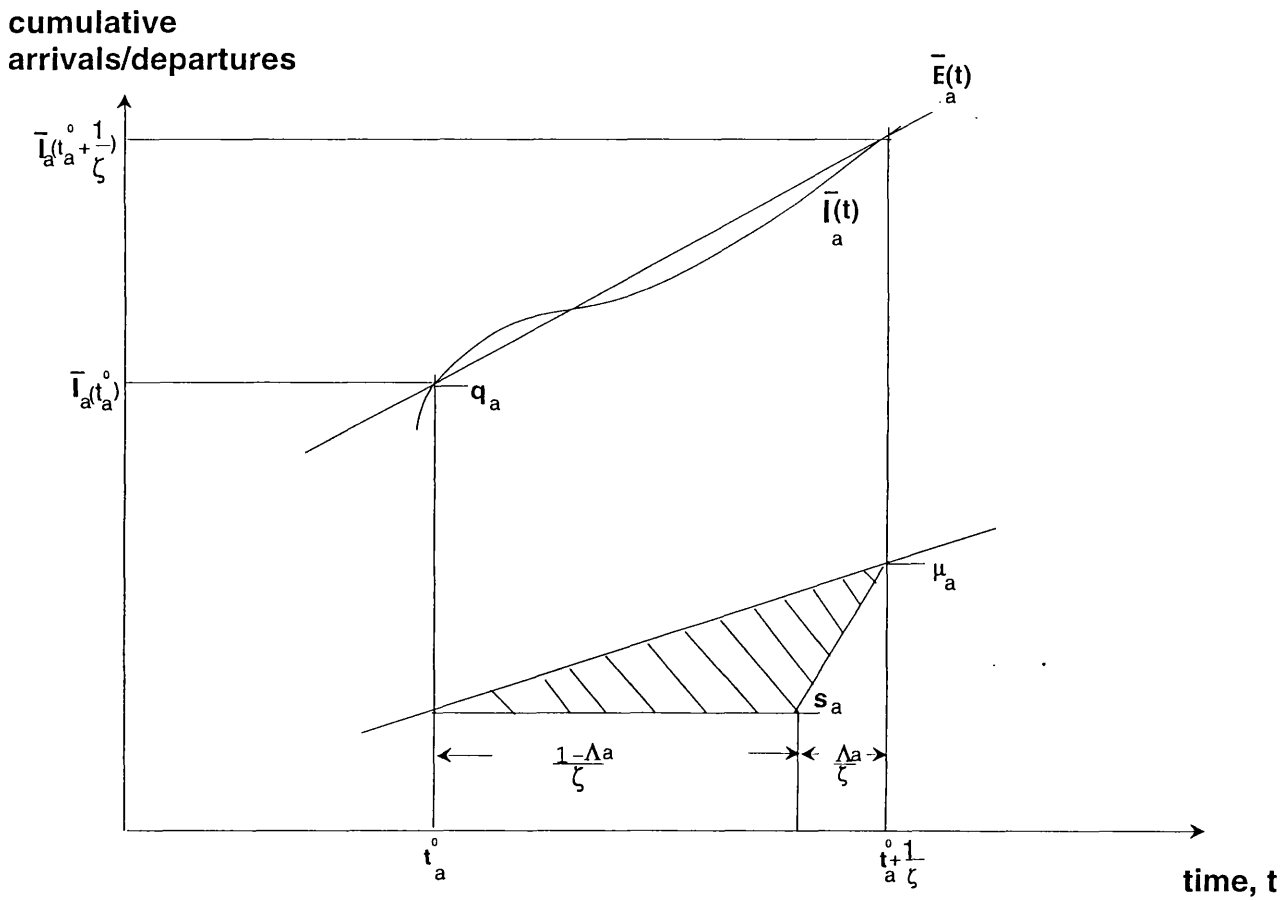


Figure 3.1 Illustration for the uniform component of oversaturated link

period T_a in the following way.

$$D_a^U = \frac{1}{t_a^n - t_a^0} \left[\sum_{k=1}^{\zeta (t_a^n - t_a^0)} \left[\int_{t_a^0 + \frac{(k-1)}{\zeta}}^{t_a^0 + \frac{(k-\Lambda_a)}{\zeta}} \bar{I}_a(t) - \bar{I}_a(t_a^0 + \frac{(k-1)}{\zeta})\, dt + \right.\right.$$

$$\left.\left. \int_{t_a^0 + \frac{(k-\Lambda_a)}{\zeta}}^{z_a + \frac{(k-1)}{\zeta}} \bar{I}_a(t) - \bar{I}_a(t_a^0 + \frac{(k-1)}{\zeta}) - s_a\,(t - (t_a^0 + \frac{(k-\Lambda_a)}{\zeta}))\, dt \right]\right] \tag{3.30}$$

where we suppose the traffic queues develop at each start of effective red

$t_a^0 + \frac{(k-1)}{\zeta}$, and clear at $z_a + \frac{(k-1)}{\zeta}$, $k = 1,2,\ldots,\zeta\,(t_a^n - t_a^0)$

where z_a is the time at which the cumulative arrivals meet the cumulative departures and is the solution of the equation

$$z_a = \frac{\bar{I}_a(z_a) - \bar{I}_a(t_a^0 + \frac{k-1}{\zeta})}{s_a} + (t_a^0 + \frac{k - \Lambda_a}{\zeta}) \tag{3.30a}$$

In the right hand side of the equation the first term shows the duration required for the queue that forms in $\left[t_a^0 + \frac{k-1}{\zeta} , t_a^0 + \frac{k}{\zeta} \right]$ to clear and the second term shows the start of effective green during that period.

If we choose the first cycle of interval T_a , i.e. $\left[t_a^0 , t_a^0 + \frac{1}{\zeta} \right]$, as the basis for the expression of $\bar{I}_a$ on T_a , then

$$D_a^U = \zeta[\int_{t_a^0}^{z_a} \bar{I}_a(t) - \bar{I}_a(t_a^0)dt - \int_{t_a^0 + \frac{(1-\Lambda_a)}{\zeta}}^{z_a} s_a(t - t_a^0 - \frac{(1-\Lambda_a)}{\zeta})dt] \tag{3.31}$$

In a similar way, for each link a in $\boldsymbol{L}$ during the corresponding time period $\boldsymbol{T}_a$, the uniform component of the indicators of traffic conditions can be expressed as follows.

$$L_a^U = D_a^U \tag{3.32}$$

$$S_a^U = \zeta\ (\bar{I}_a(z_a) - \bar{I}_a(t_a^0)) \tag{3.33}$$

$$d_a^U = \frac{D_a^U}{q_a} \tag{3.34}$$

3. For undersaturated links and with accumulated queues

When the initial queue length of link a is greater than the equilibrium queue length for time period $\boldsymbol{T}_a$, that is, $L_a^{r+o}\ (t_a^0) > L_a^e$, let τ_a be the duration needed for the accumulated queue of link a at the start of $\boldsymbol{T}_a$ to be dissipated. Following the derivations by Kimber and Hollis (LR 909, Appendix 2, 1979) but supposing as a simplification that the linear decrease continues until equilibrium mean queue length, the following expression is used.

$$\tau_a = \frac{L_a^{r+o}\ (t_a^0) - L_a^e}{\mu_a\ (\hat{x}_a - x_a)} \tag{3.35}$$

where $L_a^{r+o}(t_a^0)$, L_a^e are respectively the initial and equilibrium queue length for $\boldsymbol{T}_a$, $x_a = \dfrac{q_a}{\mu_a}$ by definition, and $\hat{x}_a$ is the degree of saturation for which $L_a^{r+o}(t_a^0)$ is the corresponding equilibrium queue length.

According to expression (3.14),

$$L_a^{r+o}(t_a^0) = \frac{C\,\hat{x}_a^2}{1 - \hat{x}_a} + \hat{x}_a \tag{3.14}$$

$$\hat{x}_a = \frac{L_a^{r+o}(t_a^0) + 1 - \sqrt{(L_a^{r+o}(t_a^0) + 1)^2 - 4(1 - C)L_a^{r+o}(t_a^0)}}{2(1 - C)} \tag{3.35a}$$

(the larger solution is not relevant)

In expression (3.35), it is assumed that the accumulated queue length decreases linearly at a rate which is the difference between the arrival rates corresponding to degrees of saturation $\hat{x}_a$ and x_a until queue length approaches the equilibrium queue for this time period. In relation to the computation of the dissipation time for the accumulated queue length within a particular time period, the expression (3.35) provides a practical way to calculate it, however, using it will underestimate the time taken by the accumulated queue to clear. In the following paragraphs, on the basis of expression (3.35), we will discuss the uniform component of the indicators of traffic conditions in the following two parts.

For each link a in $\boldsymbol{L}$, $x_a < 1$ *and* $L_a^{r+o}(t_a^0) > L_a^e$, let

$t_a^L = t_a^n - t_a^0$, *where* $T_a = \left[t_a^0 , t_a^n \right]$, $\forall\, a \in \boldsymbol{L}$, according to (3.35)

(1). if $\tau_a \geq t_a^L$, the uniform component of the indicators of traffic conditions on link a can be taken as those of an oversaturated link as in expressions (3.26) - (3.29).

(2). if $\tau_a < t_a^L$, the uniform component of indicators of traffic conditions on link a can be expressed as linear combinations of those for an oversaturated link and those for an undersaturated link without accumulated queues.

$$D_a^U = \frac{C_1\,(\tau_a - t_a^0) + D_1\,(t_a^n - \tau_a)}{t_a^L} \tag{3.36}$$

$$L_a^U = D_a^U \tag{3.37}$$

$$S_a^U = \frac{C_2\,(\tau_a - t_a^0) + D_2\,(t_a^n - \tau_a)}{t_a^L} \tag{3.38}$$

$$d_a^U = \frac{C_3\,\mu_a\,(\tau_a - t_a^0) + D_3\,q_a\,(t_a^n - \tau_a)}{\mu_a\,(\tau_a - t_a^0) + q_a\,(t_a^n - \tau_a)} \tag{3.39}$$

where C_x , D_x , $x = 1, 2, 3$ are respectively the expressions (3.27)-(3.29) and (3.31), (3.33)-(3.34).

3.5 The Lower Level Problem

In relation to the lower level problem in the bi-level problem (3.0), $q = q^\star(\psi)$, the user equilibrium traffic assignment problem can be formulated as a variational inequality problem in the following way.

To find a value $q^\star(\psi) \in \Omega(\psi)$ of $q(\psi)$ such that

$$c(q^\star(\psi))\,(\,q(\psi) - q^\star(\psi))^T \geq 0$$

$$\forall\, q(\psi) \in \Omega(\psi) = \left\{ q(\psi):\, q(\psi)^T = \delta\, f(\psi)^T,\ \Delta\, f(\psi)^T = \hat{D}^T,\ f(\psi) \geq \mathbf{0} \right\} \tag{3.40}$$

3.5.1 Link travel time function

Following the mathematical expressions for the indicators of traffic conditions, by adopting the traffic model from TRANSYT as an assessment tool for the estimation of the conditions to the performance index for each link, we suppose that the corresponding link travel time function for a particular time period can be expressed as the sum of the undelayed travel time along the link and the estimated average delay, which is incurred by vehicles on the link leading to the downstream signal-controlled junction in that time period.

For each link a in L during a given timeslice T^i , by definition the link travel time

is $$c_a(q_a\ ,\ \psi) = c_a^0 + d_a(t^{i+1}) \tag{3.41}$$

Since $d_a(t^{i+1}) = d_a^U + d_a^{r+o}(t^{i+1})$

the link travel time function for that time period can be expressed as

$$c_a(q_a\ ,\ \psi) = c_a^0 + d_a^U + d_a^{r+o}(t^{i+1}) \tag{3.42}$$

where the relevant mathematical expressions are (3.22), (3.29), (3.34), (3.39), and the link travel time is a separable function of q_a .

In relation to the uniqueness of the solution for equilibrium link flows, the following inequalities are assumed throughout this work.

$$\frac{\partial c_a}{\partial q_a} > 0\ , \quad \forall\, a \in \boldsymbol{L} \tag{3.43}$$

where $\dfrac{\partial c_a}{\partial q_a} = \dfrac{\partial d_a^U}{\partial q_a} + \dfrac{\partial d_a^{r+o}(t^{i+1})}{\partial q_a}$, $\forall\, a \in \boldsymbol{L}$ according to expression (3.42).

Detailed expressions for the derivatives of average delay in terms of the uniform and random plus oversaturation delay with respect to the average flow will be discussed in Chapter 4.

3.6 Conclusions

In this Chapter the bi-level programming approach has been used in formulating the optimisation problem of area traffic control for equilibrium flows. The fundamental terminology and the corresponding notation throughout this thesis for the area traffic control optimisation have been respectively given in Section 3.2 and 3.3. The upper level and the lower level problems of the bi-level formulation have been given respectively in Section 3.4 and 3.5, in which the corresponding objective function and the constraint set have been respectively stated. In the following Chapters, the way of searching for the feasible solutions to this bi-level problem of area traffic control optimisation is discussed on the basis of the results provided by this Chapter.

Chapter 4 Sensitivity Analysis

4.0 Introduction

The bi-level formulation for the optimisation of area traffic control leads to a non-convex problem in Chapter 3 due to the non-convexity of the objective function and the non-linearity of the condition for equilibrium network flows. Therefore, only local optimal values for the signal timings can be found for the solution to the bi-level formulation. In this Chapter, the sensitivity analysis approach will be used as an appropriate tool to obtain information about the derivatives of the rate of delay and the number of stops with respect to the current signal timings. These derivatives will be used as a basis for the solution method developed later in this work for the bi-level formulation of the optimisation of area traffic control for equilibrium network flows. The arrangement of this Chapter can be described as follows. In the next section, an introduction for the sensitivity analysis approach to the bi-level formulation will be given. The specification of each element in the derivatives for the bi-level formulation will be discussed respectively in the following sections. In Section 4.2, the derivatives for all links with respect to the common cycle time at current signal timings will be discussed. In Section 4.3, the derivatives for the upstream links which are controlled by a particular signal group at a particular junction with respect to the start and duration of green will be discussed. In Chapter 3, the objective function adopted was the performance index in terms of a weighted linear combination of the rate of delay and the number of stops, which was used in the traffic model from TRANSYT, in which cyclic flow profiles are used to represent vehicle platooning. Accordingly, for a change in the signal timings of one particular signal group at the corresponding junction, not only the primary effects will be caused on the upstream links by the change in the signal timings but consequent effects will also be caused on the downstream and further downstream links since the IN profiles for the downstream and further downstream links are affected by the OUT profiles for the upstream links which are controlled by that signal group. Therefore, in Section 4.4, the derivatives for the downstream links with respect to the start and duration of green for that signal group will be discussed, where the downstream links are the immediate ones away

from the corresponding signal group. Furthermore, in Section 4.5, the derivatives for the further downstream links which are the immediate links away from the downstream links with respect to the start and duration of green for that signal group will be discussed accordingly. Conclusions for this Chapter will be given in Section 4.6.

4.1 Sensitivity Analysis for the Bi-level Formulation

As we have learned in Chapter 3 that the upper level problem (3.1) is a non-linear and non-convex problem with respect to the signal setting variables ζ , θ , ϕ while the lower level problem (3.40) is the user equilibrium assignment in terms of the signal timings, in which the relationship of the equilibrium flows and signal timings is also non-linear, the feasible set of the solutions to the bi-level problem (3.0) fails to be expressed as a closed form. To achieve a better solution for this bi-level problem (3.0), the sensitivity analysis approach which uses the derivatives of the indicators of traffic conditions with respect to the signal setting variables in searching for a local optimum is developed as follows.

The bi-level problem (3.0) can be reexpressed as to

$$\begin{aligned} &\underset{\psi}{Minimise} \quad P = P_0(\psi , q) \\ &subject\ \ to \\ &\qquad q = q^{\star}(\psi) \\ &\qquad \psi \in S_0 \end{aligned} \tag{4.0}$$

where P_0 is the performance index function of the signal setting variables,

$\psi = \{ \zeta , \theta , \phi \}$ which has been expressed in the equation (3.2), and the corresponding notation is as defined in Chapter 3, and $q^{\star}(\psi)$ is formulated as the variational inequality as given in (3.40).

Therefore, the derivatives of the performance index for problem (4.0) with respect to the signal setting variables, ζ , θ , ϕ can be expressed in the following way.

$$\frac{\partial P}{\partial \zeta} = \frac{\partial P_0}{\partial \zeta} + \frac{\partial P_0}{\partial q^\star} \frac{\partial q^\star}{\partial \zeta} \tag{4.1}$$

$$\frac{\partial P}{\partial \theta} = \frac{\partial P_0}{\partial \theta} + \frac{\partial P_0}{\partial q^\star} \frac{\partial q^\star}{\partial \theta} \tag{4.2}$$

$$\frac{\partial P}{\partial \phi} = \frac{\partial P_0}{\partial \phi} + \frac{\partial P_0}{\partial q^\star} \frac{\partial q^\star}{\partial \phi} \tag{4.3}$$

According to the definition of the objective function for P_0 in expression (3.2), the derivatives of the performance index for (4.1)-(4.3) in terms of those of the rate of delay, D_a , and the number of stops, S_a , and the equilibrium link flows, $\boldsymbol{q}$, can be expressed as follows (where D_a , S_a are short for $D_a(t)$, $S_a(t)$ over the interval $[\, t^i \, , \, t \,]$).

$$\begin{aligned} \frac{\partial P}{\partial \zeta} = \sum_{a \in L} [& (\frac{\partial D_a}{\partial \zeta} + \sum_{l \in L} \frac{\partial D_a}{\partial q_l^\star} \frac{\partial q_l^\star}{\partial \zeta}) W_{aD} M_D + \\ & (\frac{\partial S_a}{\partial \zeta} + \sum_{l \in L} \frac{\partial S_a}{\partial q_l^\star} \frac{\partial q_l^\star}{\partial \zeta}) W_{aS} M_S] \end{aligned} \tag{4.4}$$

$$\nabla_\theta P = \left[\frac{\partial P}{\partial \theta_{jm}} \; ; \; \forall j \in \boldsymbol{P}_m \, , \; \forall m \in \boldsymbol{N} \right] \tag{4.5}$$

$$\nabla_\phi P = \left[\frac{\partial P}{\partial \phi_{jm}} \; ; \; \forall j \in \boldsymbol{P}_m \, , \; \forall m \in \boldsymbol{N} \right] \tag{4.6}$$

For each signal group j in $\boldsymbol{P}_m$ at any one signal controlled junction m in $\boldsymbol{N}$, let

$\boldsymbol{A}_{jm}$ = {a; a is the traffic stream such that $\hat{b}_{ajm}$ = 1 }

$\boldsymbol{B}_{jm}$ = { b; b is a downstream link for some link $a \in \boldsymbol{A}_{jm}$ }

$\boldsymbol{C}_{jm}$ = { c; c is a further downstream link for some link $b \in \boldsymbol{B}_{jm}$ }

and $\boldsymbol{L}_{jm} = \boldsymbol{L} \cap (\boldsymbol{A}_{jm} \cup \boldsymbol{B}_{jm} \cup \boldsymbol{C}_{jm})$, then the elements of (4.5)-(4.6) are

$$\frac{\partial P}{\partial \theta_{jm}} = \sum_{a \in \boldsymbol{L}_{jm}} [(\frac{\partial D_a}{\partial \theta_{jm}} + \sum_{l \in \boldsymbol{L}} \frac{\partial D_a}{\partial q_l^\star} \frac{\partial q_l^\star}{\partial \theta_{jm}}) W_{aD} M_D$$

$$+ (\frac{\partial S_a}{\partial \theta_{jm}} + \sum_{l \in \boldsymbol{L}} \frac{\partial S_a}{\partial q_l^\star} \frac{\partial q_l^\star}{\partial \theta_{jm}}) W_{aS} M_S] \qquad (4.5a)$$

$$\frac{\partial P}{\partial \phi_{jm}} = \sum_{a \in \boldsymbol{L}_{jm}} [(\frac{\partial D_a}{\partial \phi_{jm}} + \sum_{l \in \boldsymbol{L}} \frac{\partial D_a}{\partial q_l^\star} \frac{\partial q_l^\star}{\partial \phi_{jm}}) W_{aD} M_D$$

$$+ (\frac{\partial S_a}{\partial \phi_{jm}} + \sum_{l \in \boldsymbol{L}} \frac{\partial S_a}{\partial q_l^\star} \frac{\partial q_l^\star}{\partial \phi_{jm}}) W_{aS} M_S] \qquad (4.6a)$$

The gradient of equilibrium flow with respect to the signal timings, i.e. $\nabla_\psi \, \boldsymbol{q}^\star(\psi)$

$\psi = (\zeta , \theta , \phi)$ can be expressed in terms of the gradient of the link travel time with respect to the signal setting variables, i.e. $\nabla_\psi \, \boldsymbol{c}(\boldsymbol{q} , \psi)$, using the results obtained by Tobin and Friesz (1988)as quoted in pp. 59-60 .

$$(\nabla_\psi \, \boldsymbol{q}^\star)^T = - \delta \, \boldsymbol{B} \, \delta^T \nabla_\psi \, \boldsymbol{c}^T \qquad (4.7)$$

$$\boldsymbol{B} = (\delta^T \nabla_{\boldsymbol{q}} \, \boldsymbol{c} \, \delta)^{-1} (\boldsymbol{I} - \Delta^T (\Delta (\delta \, \nabla_{\boldsymbol{q}} \, \boldsymbol{c} \, \delta^T)^{-1} \Delta^T)^{-1} \Delta (\delta^T \nabla_{\boldsymbol{q}} \, \boldsymbol{c} \, \delta)^{-1})$$

where $\boldsymbol{I}$ is the identity matrix in dimension of the number of paths, and the notation used has been given in Chapter 3, and each element in expressions (4.4)-(4.7) will be derived in the following sections.

The mathematical expressions for the indicators of traffic conditions for all links have been expressed as follows: for the random plus oversaturation component in the expressions (3.21)-(3.23) with t^i replaced by t_a^0 for each link a in $\boldsymbol{L}$, and for the uniform component: for the oversaturated links in expressions (3.27)-(3.29), for the undersaturated links without accumulated queues in expressions (3.31), (3.33)-(3.34), and for the undersaturated links with accumulated queues in expressions (3.36), (3.38)-(3.39). Some derivatives can concurrently be discussed for all types of link together, and this is done in Section 4.2. Others are discussed separately for upstream, downstream and further downstream links in Sections 4.3-4.5 respectively.

4.2 Derivatives for All Links

The derivatives of the indicators of traffic conditions with respect to ζ, the reciprocal of the common cycle time, and with respect to the link flow, are discussed for links of all types in this Section. Details of derivation are given in Appendix B.

1. The random plus oversaturation component.

In relation to expression (4.4), the derivatives of the indicators of traffic conditions with respect to ζ for all links a in $\boldsymbol{L}$ are as follows.

$$\frac{\partial D_a^{r+o}}{\partial \zeta} = \frac{\partial D_a^{r+o}}{\partial A_2}\frac{\partial A_2}{\partial \zeta} + \frac{\partial D_a^{r+o}}{\partial B_2}\frac{\partial B_2}{\partial \zeta} \tag{4.8}$$

$$\frac{\partial S_a^{r+o}}{\partial \zeta} = \exp\left(\frac{-D_a^{r+o}\zeta}{2q_a}\right)\left(D_a^{r+o} + \zeta\frac{\partial D_a^{r+o}}{\partial \zeta}\right) \tag{4.9}$$

$$\frac{\partial d_a^{r+o}}{\partial \zeta} = \frac{\partial d_a^{r+o}}{\partial A_3}\frac{\partial A_3}{\partial \zeta} + \frac{\partial d_a^{r+o}}{\partial B_3}\frac{\partial B_3}{\partial \zeta} \tag{4.10}$$

For each signal group j in $\boldsymbol{P}_m$ at any one signal controlled junction m in $\boldsymbol{N}$, for each link l in $\boldsymbol{A}_{jm}$, the derivatives of the indicators of traffic conditions with respect to q_l, $\forall\, l \in \boldsymbol{A}_{jm}$ are

$$\frac{\partial D_l^{r+o}}{\partial q_l} = \frac{\partial D_l^{r+o}}{\partial A_2}\frac{\partial A_2}{\partial q_l} + \frac{\partial D_l^{r+o}}{\partial B_2}\frac{\partial B_2}{\partial q_l} \tag{4.11}$$

$$\frac{\partial S_l^{r+o}}{\partial q_l} = 2 - 2\exp\left(\frac{-\zeta D_l^{r+o}}{2q_l}\right) - \zeta\exp\left(\frac{-\zeta D_l^{r+o}}{2q_l}\right)\left[\frac{D_l^{r+o}}{q_l} - \frac{\partial D_l^{r+o}}{\partial q_l}\right] \tag{4.12}$$

$$\frac{\partial d_l^{r+o}}{\partial q_l} = \frac{\partial d_l^{r+o}}{\partial A_3}\frac{\partial A_3}{\partial q_l} + \frac{\partial d_l^{r+o}}{\partial B_3}\frac{\partial B_3}{\partial q_l} \tag{4.13}$$

Also, for each link a in $\boldsymbol{B}_{jm} \cup \boldsymbol{C}_{jm}$ and in $\boldsymbol{L} \setminus \boldsymbol{L}_{jm}$, the derivatives with respect to q_l, $\forall\, l \in \boldsymbol{A}_{jm}$ are

$$\frac{\partial D_a^{r+o}}{\partial q_l} = \frac{\partial S_a^{r+o}}{\partial q_l} = \frac{\partial d_a^{r+o}}{\partial q_l} = 0 \tag{4.14}$$

where in (4.8)-(4.13), let $x = a, l$; $\forall\, a \in \boldsymbol{L}$, $\forall\, l \in \boldsymbol{A}_{jm}$

$$\frac{\partial D_x^{r+o}}{\partial A_2} = \frac{1}{2}\left[\frac{A_2}{\sqrt{A_2^2 + B_2}} - 1\right]$$

$$\frac{\partial D_x^{r+o}}{\partial B_2} = \frac{1}{4\sqrt{A_2^2 + B_2}}$$

$$\frac{\partial d_x^{r+o}}{\partial A_3} = \frac{1}{2}\left[\frac{A_3}{\sqrt{A_3^2 + B_3}} - 1 \right]$$

$$\frac{\partial d_x^{r+o}}{\partial B_3} = \frac{1}{4\sqrt{A_3^2 + B_3}}$$

and in (4.8)-(4.10),

$$\frac{\partial A_n}{\partial \zeta} = e_a\, s_a \frac{\partial A_n}{\partial \mu_a} = e_a \frac{\partial A_n}{\partial \phi_{jm}}, \quad n = 2, 3$$

$$\frac{\partial B_n}{\partial \zeta} = e_a\, s_a \frac{\partial B_n}{\partial \mu_a} = e_a \frac{\partial B_n}{\partial \phi_{jm}}, \quad n = 2, 3$$

$$\frac{\partial A_2}{\partial \phi_{jm}} = \frac{s_a(t - t_a^0)[(1+ x_a)(\frac{\mu_a}{2}(t - t_a^0))^2 + (2 - 2C - x_a\frac{\mu_a}{2}(t - t_a^0))\frac{\mu_a}{2}(t - t_a^0)]}{2[\frac{\mu_a}{2}(t - t_a^0)+ 1 - C]^2}$$

$$+ \frac{s_a(t - t_a^0)\ (1 - C)\ (1 + L_a^{r+o}(t_a^0) + x_a\frac{\mu_a}{2}(t - t_a^0))}{2[\frac{\mu_a}{2}(t - t_a^0) + 1 - C]^2}$$

$$\frac{\partial B_2}{\partial \phi_{jm}} = \frac{2s_a(t - t_a^0)(1 - C)(1+ L_a^{r+o}(t_a^0)+ x_a\frac{\mu_a}{2}(t - t_a^0))(L_a^{r+o}(t_a^0)+ x_a\frac{\mu_a}{2}(t - t_a^0))}{[\frac{\mu_a}{2}(t - t_a^0)+ 1 - C]^2}$$

$$\frac{\partial A_3}{\partial \phi_{jm}} = \frac{x_a(t - t_a^0)}{2\Lambda_a} + \frac{L_a^{r+o}(t_a^0) - C + 2}{\Lambda_a \mu_a}$$

$$\frac{\partial B_3}{\partial \phi_{jm}} = \frac{2(t - t_a^0)}{\Lambda_a \mu_a} - \frac{4[(1 - x_a)\mu_a(t - t_a^0) + Cx_a\mu_a(t - t_a^0) - 2(1 - C)(L_a^{r+o}(t_a^0)+1)]}{\Lambda_a \mu_a^2}$$

and in (4.11)-(4.13),

$$\frac{\partial A_2}{\partial q_l} = \frac{\frac{-\mu_l}{4}(t - t_l^0)^2 - (1 - C)(t - t_l^0)}{\frac{\mu_l}{2}(t - t_l^0) + 1 - C}$$

$$\frac{\partial B_2}{\partial q_l} = \frac{\mu_l(t - t_l^0)^2 - 4(t - t_l^0)(1 - C)(L_l^{r+o}(t_l^o) + x_l\frac{\mu_l}{2}(t - t_l^0))}{\frac{\mu_l}{2}(t - t_l^0) + 1 - C}$$

$$\frac{\partial A_3}{\partial q_l} = \frac{-(t - t_l^0)}{2\mu_l}$$

$$\frac{\partial B_3}{\partial q_l} = \frac{-2(1 - C)(t - t_l^0)}{\mu_l^2}$$

2. The uniform component.

(1). The derivatives of the indicators of traffic conditions with respect to ζ for all links a in L, when $x_a \geq 1$, can be expressed as follows.

$$\frac{\partial D_a^U}{\partial \zeta} = \frac{-\mu_a(1 - \Lambda_a) + e_a s_a \zeta(1 - 2\Lambda_a)}{2\zeta^2} \tag{4.15}$$

$$\frac{\partial S_a^U}{\partial \zeta} = 0 \tag{4.16}$$

$$\frac{\partial d_a^U}{\partial \zeta} = \frac{-(1 - \Lambda_a + e_a\zeta)}{2\zeta^2} \tag{4.17}$$

For each signal group j in $\boldsymbol{P}_m$ at any one signal controlled junction m in $\boldsymbol{N}$, for each link l in $\boldsymbol{A}_{jm}$, the derivatives of the indicators of traffic conditions with respect to q_l, $\forall\, l \in \boldsymbol{A}_{jm}$ are

$$\frac{\partial D_l^U}{\partial q_l} = 0 \tag{4.18}$$

$$\frac{\partial S_l^U}{\partial q_l} = 1 \tag{4.19}$$

$$\frac{\partial d_l^U}{\partial q_l} = 0 \tag{4.20}$$

Also, for each link a in $\boldsymbol{B}_{jm} \cup \boldsymbol{C}_{jm}$ and in $\boldsymbol{L} \setminus \boldsymbol{L}_{jm}$, the derivatives with respect to q_l, $\forall\, l \in \boldsymbol{A}_{jm}$ are

$$\frac{\partial D_a^U}{\partial q_l} = \frac{\partial S_a^U}{\partial q_l} = \frac{\partial d_a^U}{\partial q_l} = 0 \tag{4.21}$$

(2). The derivatives of the indicators of traffic conditions with respect to ζ for all links a in $\boldsymbol{L}$, when $x_a < 1$, $L_a^{r+o}(t_a^0) < L_a^e$, can be expressed as follows.

$$\frac{\partial D_a^U}{\partial \zeta} = \int_{t_a^0}^{t_a^0 + \frac{(1-\Lambda_a)}{\zeta}} \bar{I}_a(t) - \bar{I}_a(t_a^0)\, dt +$$

$$\int_{t_a^0 + \frac{(1-\Lambda_a)}{\zeta}}^{z_a} \bar{I}_a(t) - \bar{I}_a(t_a^0) - s_a(t - (t_a^0 + \frac{(1-\Lambda_a)}{\zeta}))\, dt + \tag{4.22}$$

$$\zeta[\int_{t_a^0}^{z_a} \frac{\partial \bar{I}_a(t)}{\partial \zeta} dt - \int_{t_a^0 + \frac{1-\Lambda_a}{\zeta}}^{z_a} s_a(\frac{1 - \Lambda_a + e_a\zeta}{\zeta^2})\, dt\ ]$$

$$\frac{\partial S_a^U}{\partial \zeta} = (\bar{I}_a(z_a) - \bar{I}_a(t_a^0)) + \zeta \frac{\partial \bar{I}_a(z_a)}{\partial \zeta} \tag{4.23}$$

$$\frac{\partial d_a^U}{\partial \zeta} = \frac{1}{q_a} \frac{\partial D_a^U}{\partial \zeta} \tag{4.24}$$

For each signal group j in $\boldsymbol{P}_m$ at any one signal controlled junction m in $\boldsymbol{N}$, for each link l in $\boldsymbol{A}_{jm}$, the derivatives of the indicators of traffic conditions with respect to q_l , $\forall\, l \in \boldsymbol{A}_{jm}$ are

$$\frac{\partial D_l^U}{\partial q_l} = \zeta \int_{t_l^0}^{z_l} \frac{\partial \bar{I}_l(w)}{\partial q_l}\, dw \tag{4.25}$$

$$\frac{\partial S_l^U}{\partial q_l} = \zeta\ [\int_{t_l^0}^{z_l} \frac{\partial I_l(t)}{\partial q_l} dt + \frac{\partial z_l}{\partial q_l} I_l(z_l)\ ] \tag{4.26}$$

$$\frac{\partial d_l^U}{\partial q_l} = \frac{1}{q_l}\frac{\partial D_l^U}{\partial q_l} - \frac{D_l^U}{q_l^2} \tag{4.27}$$

Also, for each link a in $\boldsymbol{B}_{jm} \cup \boldsymbol{C}_{jm}$ and in $\boldsymbol{L} \setminus \boldsymbol{L}_{jm}$, the derivatives with respect to q_l, $\forall\, l \in \boldsymbol{A}_{jm}$ are

$$\frac{\partial D_a^U}{\partial q_l} = \frac{\partial S_a^U}{\partial q_l} = \frac{\partial d_a^U}{\partial q_l} = 0 \tag{4.28}$$

(3)(i). The derivatives of the indicators of traffic conditions with respect to ζ for all links a in $\boldsymbol{L}$, when $x_a < 1$, $L_a^{r+o}(t_a^0) \geq L_a^e$ *and* $\tau_a \geq t_a^L$, are the same as in expressions (4.15)-(4.17); furthermore for each signal group j in $\boldsymbol{P}_m$ at any one signal controlled junction m in $\boldsymbol{N}$, for each link l in $\boldsymbol{A}_{jm}$, the derivatives of the indicators of traffic conditions with respect to q_l, $\forall\, l \in \boldsymbol{A}_{jm}$ are the same as shown in expressions (4.18)-(4.20), and for each link a in $\boldsymbol{B}_{jm} \cup \boldsymbol{C}_{jm}$, and in $\boldsymbol{L} \setminus \boldsymbol{L}_{jm}$ the derivatives with respect to q_l, $\forall\, l \in \boldsymbol{A}_{jm}$ are the same as shown in expression (4.21).

(ii). The derivatives of the indicators of traffic conditions with respect to ζ for all links a in $\boldsymbol{L}$, when $x_a < 1$, $L_a^{r+o}(t_a^0) \geq L_a^e$ *and* $\tau_a < t_a^L$, can be expressed as linear combinations of the derivatives for the oversaturated links and those for the undersaturated links in the following way.

$$\frac{\partial D_a^U}{\partial \zeta} = \frac{1}{t_a^L}[\ \frac{\partial C_1}{\partial \zeta}(\tau_a - t_a^0) + \frac{\partial D_1}{\partial \zeta}(t_a^n - \tau_a) + \frac{\partial \tau_a}{\partial \zeta}(C_1 - D_1)\ ] \tag{4.29}$$

$$\frac{\partial S_a^U}{\partial \zeta} = \frac{1}{t_a^L}[\ \frac{\partial C_2}{\partial \zeta}(\tau_a - t_a^0) + \frac{\partial D_2}{\partial \zeta}(t_a^n - \tau_a) + \frac{\partial \tau_a}{\partial \zeta}(C_2 - D_2)\ ] \tag{4.30}$$

$$\frac{\partial d_a^U}{\partial \zeta} = \frac{(\frac{\partial C_3}{\partial \zeta}\mu_a(\tau_a - t_a^0) + C_3 e_a s_a(\tau_a - t_a^0) + C_3\mu_a\frac{\partial \tau_a}{\partial \zeta})}{\mu_a(\tau_a - t_a^0) + q_a(t_a^n - \tau_a)} +$$

$$\frac{(\frac{\partial D_3}{\partial \zeta}q_a(t_a^n - \tau_a) - D_3 q_a\frac{\partial \tau_a}{\partial \zeta})}{\mu_a(\tau_a - t_a^0) + q_a(t_a^n - \tau_a)} -$$

$$\frac{(C_3\mu_a(\tau_a - t_a^0) + D_3 q_a(t_a^n - \tau_a))(e_a s_a(\tau_a - t_a^0) + \mu_a\frac{\partial \tau_a}{\partial \zeta} - q_a\frac{\partial \tau_a}{\partial \zeta})}{(\mu_a(\tau_a - t_a^0) + q_a(t_a^n - \tau_a))^2}$$

(4.31)

For each signal group j in $\boldsymbol{P}_m$ at any one signal controlled junction m in $\boldsymbol{N}$, for each link l in $\boldsymbol{A}_{jm}$, the derivatives of the indicators of traffic conditions with respect to q_l , $\forall\ l \in \boldsymbol{A}_{jm}$ are

$$\frac{\partial D_l^U}{\partial q_l} = \frac{1}{t_l^L}[\ \frac{\partial D_1}{\partial q_l}(t_l^n - \tau_l) + \frac{\partial \tau_l}{\partial q_l}(C_1 - D_1)\ ] \tag{4.32}$$

$$\frac{\partial S_l^U}{\partial q_l} = \frac{1}{t_l^L}[\ (\tau_l - t_l^0) + \frac{\partial D_2}{\partial q_i}(t_l^n - \tau_l) + \frac{\partial \tau_l}{\partial q_l}(C_2 - D_2)\ ] \tag{4.33}$$

$$\frac{\partial d_l^{\,U}}{\partial q_l} = \frac{(C_3\mu_l\frac{\partial\tau_l}{\partial q_l} + \frac{\partial D_3}{\partial q_l}q_l(t_l^{\,n} - \tau_l) + D_3(t_l^{\,n} - \tau_l) - D_3 q_l\frac{\partial\tau_l}{\partial q_l})}{\mu_l(\tau_l - t_l^{\,0}) + q_l(t_l^{\,n} - \tau_l)} -$$

$$\frac{(C_3\mu_l(\tau_l - t_l^{\,0}) + D_3 q_l(t_l^{\,n} - \tau_l))(\mu_l\frac{\partial\tau_l}{\partial q_l} + (t_l^{\,n} - \tau_l) - q_l\frac{\partial\tau_l}{\partial q_l})}{(\mu_l(\tau_l - t_l^{\,0}) + q_l(t_l^{\,n} - \tau_l))^2} \tag{4.34}$$

Also, for each link a in $\boldsymbol{B}_{jm} \cup \boldsymbol{C}_{jm}$ and in $\boldsymbol{L} \setminus \boldsymbol{L}_{jm}$, the derivatives with respect to q_l, $\forall\, l \in \boldsymbol{A}_{jm}$ are

$$\frac{\partial D_a^{\,U}}{\partial q_l} = \frac{\partial S_a^{\,U}}{\partial q_l} = \frac{\partial d_a^{\,U}}{\partial q_l} = 0 \tag{4.35}$$

where in (4.29)-(4.31), and (4.32)-(4.34),

$$\frac{\partial\tau_a}{\partial\zeta} = \frac{-e_a x_a}{\Lambda_a\mu_a(x_a - \hat{x}_a)}[1 + \frac{2Cx_a - Cx_a^{\,2}}{(1 - x_a)^2}] + \frac{e_a s_a\hat{x}_a(L_a^{\,e} - L_a^{\,r+o}(t_a^{\,0}))}{(\mu_a(x_a - \hat{x}_a))^2}$$

$$\frac{\partial\tau_l}{\partial q_l} = \frac{1}{\mu_l^2(x_l - \hat{x}_l)}[\,1 + \frac{2C\,x_l - C\,x_l^{\,2}}{(1 - x_l)^2}\,] - \frac{L_l^{\,e} - L_l^{\,r\,+o}(t_l^{\,0})}{(\mu_l\,(x_l - \hat{x}_l))^2}$$

$\frac{\partial C_1}{\partial\zeta}$, $\frac{\partial C_2}{\partial\zeta}$ *and* $\frac{\partial C_3}{\partial\zeta}$ can be obtained from expressions (4.15)-(4.17);

$\frac{\partial D_1}{\partial\zeta}$, $\frac{\partial D_2}{\partial\zeta}$ *and* $\frac{\partial D_3}{\partial\zeta}$ can be obtained from expressions (4.22)-(4.24).

3. Derivatives of equilibrium flows

According to expression (4.7), the derivatives of equilibrium flows with respect to ζ can be expressed as

$$(\nabla_\zeta \, q^\star)^T = - \delta \, \boldsymbol{B} \, \delta^T \, \nabla_\zeta \, c^T \tag{4.36}$$

$$\boldsymbol{B} = (\delta^T \nabla_q \, c \, \delta)^{-1} (\boldsymbol{I} - \Delta^T (\Delta (\delta \, \nabla_q \, c \, \delta^T)^{-1} \Delta^T)^{-1} \Delta (\delta^T \nabla_q \, c \, \delta)^{-1})$$

and by definition for each link a in $\boldsymbol{L}$ over $\boldsymbol{T}^i$ according to expression (3.42)

$$\frac{\partial c_a(q_a \, , \, \psi)}{\partial \zeta} = \frac{\partial d_a^U}{\partial \zeta} + \frac{\partial d_a^{r+o}}{\partial \zeta} \tag{4.36a}$$

$$\frac{\partial c_a(q_a \, , \, \psi)}{\partial q_a} = \frac{\partial d_a^U}{\partial q_a} + \frac{\partial d_a^{r+o}}{\partial q_a} \tag{4.36b}$$

where $\psi = [\, \zeta \, , \, \theta \, , \, \phi \,]$ and d_a^{r+o} is short for $d_a^{r+o}(t^{i+1})$ over $\boldsymbol{T}^i$.

Also, $\frac{\partial d_a^{r+o}}{\partial \zeta}$, $\frac{\partial d_a^U}{\partial \zeta}$ respectively have been given in expression (4.10) for the random plus oversaturation component and in expressions (4.17), (4.24) and (4.31) for the uniform component on the various types of link; also $\frac{\partial d_a^{r+o}}{\partial q_a}$, $\frac{\partial d_a^U}{\partial q_a}$ have been given in expressions (4.13)-(4.14) for the random plus oversaturation component and in expressions (4.20)-(4.21), (4.27)-(4.28) and (4.34)-(4.35) for the uniform component on the corresponding types of link.

4.3 Derivatives for Upstream Links

Derivatives for the indicators of traffic conditions in relation to expressions (4.5a)-(4.6a) with respect to the start and duration of green time are discussed for the upstream links which are directly controlled by a particular signal group at the corresponding junction. The mathematical expressions for the indicators of the performance index for all links have been expressed as follows: for the random plus oversaturation component in the expressions (3.21)-(3.23) with t^i replaced by t_a^0 for each link a in $\boldsymbol{L}$, and for the uniform component: for the oversaturated links in expressions (3.27)-(3.29), for the undersaturated links without accumulated queues in expressions (3.31), (3.33)-(3.34), and for the undersaturated links with accumulated queues in expressions (3.36), (3.38)-(3.39). The derivatives of the indicators of traffic conditions with respect to θ, ϕ for the upstream links are discussed in this Section. Details of derivation are given in Appendix B.

For each signal group j in $\boldsymbol{P}_m$ at any one signal controlled junction m in $\boldsymbol{N}$, the upstream links which are controlled by the signal group j at the corresponding junction m form the set

$$\boldsymbol{A}_{jm} = \{a;\ a \text{ is the traffic stream such that } \hat{b}_{ajm} = 1\}$$

1. The random plus oversaturation component

In relation to expression (4.5a)-(4.6a), the derivatives of the indicators of traffic conditions with respect to θ_{jm}, ϕ_{jm} for each upstream link a in $\boldsymbol{A}_{jm}$ are

$$\frac{\partial D_a^{r+o}}{\partial \theta_{jm}} = \frac{\partial S_a^{r+o}}{\partial \theta_{jm}} = \frac{\partial d_a^{r+o}}{\partial \theta_{jm}} = 0 \tag{4.37}$$

$$\frac{\partial D_a^{r+o}}{\partial \phi_{jm}} = \frac{\partial D_a^{r+o}}{\partial A_2}\frac{\partial A_2}{\partial \phi_{jm}} + \frac{\partial D_a^{r+o}}{\partial B_2}\frac{\partial B_2}{\partial \phi_{jm}} \tag{4.38}$$

$$\frac{\partial S_a^{r+o}}{\partial \phi_{jm}} = \zeta \exp\left(\frac{-\zeta D_a^{r+o}}{2q_a}\right) \frac{\partial D_a^{r+o}}{\partial \phi_{jm}} \tag{4.39}$$

$$\frac{\partial d_a^{r+o}}{\partial \phi_{jm}} = \frac{\partial d_a^{r+o}}{\partial A_3}\frac{\partial A_3}{\partial \phi_{jm}} + \frac{\partial d_a^{r+o}}{\partial B_3}\frac{\partial B_3}{\partial \phi_{jm}} \tag{4.40}$$

2. The uniform component.

(1). The derivatives of the indicators of traffic conditions with respect to θ_{jm} , ϕ_{jm} for each link a in A_{jm} when $x_a \geq 1$, can be expressed as follows.

$$\frac{\partial D_a^U}{\partial \theta_{jm}} = \frac{\partial S_a^U}{\partial \theta_{jm}} = \frac{\partial d_a^U}{\partial \theta_{jm}} = 0 \tag{4.41}$$

$$\frac{\partial D_a^U}{\partial \phi_{jm}} = \frac{s_a(1 - 2\Lambda_a)}{2\zeta} \tag{4.42}$$

$$\frac{\partial S_a^U}{\partial \phi_{jm}} = 0 \tag{4.43}$$

$$\frac{\partial d_a^U}{\partial \phi_{jm}} = \frac{-1}{2\zeta} \tag{4.44}$$

(2). The derivatives of the delay per unit time, i.e. D_a^U and stops per unit time, i.e. S_a^U with respect to θ_{jm} , ϕ_{jm} for each link a in A_{jm} , when

$x_a < 1$, $L_a^{r+o}(t_a^0) < L_a^e$, were derived by Wong (1993, pp 251-254) and are, when expressed in the notation of this thesis, as follows. The derivative of the delay per vehicle, i.e. d_a^U takes the form shown because $x_a < 1$.

$$\frac{\partial D_a^U}{\partial \theta_{jm}} = s_a \left(z_a - \left(t_a^0 + \frac{1 - \Lambda_a}{\zeta}\right)\right) - I_a(t_a^0)(z_a - t_a^0) \tag{4.45}$$

$$\frac{\partial S_a^U}{\partial \theta_{jm}} = \frac{s_a(I_a(z_a) - I_a(t_a^0))}{s_a - I_a(z_a)} \tag{4.46}$$

$$\frac{\partial d_a^U}{\partial \theta_{jm}} = \frac{1}{q_a} \frac{\partial D_a^U}{\partial \theta_{jm}} \tag{4.47}$$

$$\frac{\partial D_a^U}{\partial \phi_{jm}} = - I_a(t_a^0) (z_a - t_a^0) \tag{4.48}$$

$$\frac{\partial S_a^U}{\partial \phi_{jm}} = - \frac{s_a I_a(t_a^0)}{s_a - I_a(z_a)} \tag{4.49}$$

$$\frac{\partial d_a^U}{\partial \phi_{jm}} = \frac{1}{q_a} \frac{\partial D_a^U}{\partial \phi_{jm}} \tag{4.50}$$

(3)(i). The derivatives of the indicators of traffic conditions with respect to θ_{jm} , ϕ_{jm} for each link a in A_{jm} , when $x_a < 1$, $L_a^{r+o}(t_a^0) \geq L_a^e$, and $\tau_a \geq t_a^L$ are the same as in expressions (4.41)-(4.44).

(ii). The derivatives of the indicators of traffic conditions with respect to θ_{jm} , ϕ_{jm} for each link a in A_{jm} , when $x_a < 1$, $L_a^{r+o}(t_a^0) \geq L_a^e$, and $\tau_a < t_a^L$ can be expressed as linear combinations of the derivatives for the oversaturated links and those for the undersaturated links in the following way.

$$\frac{\partial D_a^U}{\partial \theta_{jm}} = \frac{1}{t_a^L}\left[\frac{\partial D_1}{\partial \theta_{jm}}(t_a^n - \tau_a) \right] \tag{4.51}$$

$$\frac{\partial S_a^U}{\partial \theta_{jm}} = \frac{1}{t_a^L}\left[\frac{\partial D_2}{\partial \theta_{jm}}(t_a^n - \tau_a) \right] \tag{4.52}$$

$$\frac{\partial d_a^U}{\partial \theta_{jm}} = \frac{\frac{\partial D_3}{\partial \theta_{jm}} q_a(t_a^n - \tau_a)}{\mu_a(\tau_a - t_a^0) + q_a(t_a^n - \tau_a)} \tag{4.53}$$

$$\frac{\partial D_a^U}{\partial \phi_{jm}} = \frac{1}{t_a^L}\left[\frac{\partial C_1}{\partial \phi_{jm}}(\tau_a - t_a^0) + \frac{\partial D_1}{\partial \phi_{jm}}(t_a^n - \tau_a) + \frac{\partial \tau_a}{\partial \phi_{jm}}(C_1 - D_1) \right] \tag{4.54}$$

$$\frac{\partial S_a^U}{\partial \phi_{jm}} = \frac{1}{t_a^L}\left[\frac{\partial C_2}{\partial \phi_{jm}}(\tau_a - t_a^0) + \frac{\partial D_2}{\partial \phi_{jm}}(t_a^n - \tau_a) + \frac{\partial \tau_a}{\partial \phi_{jm}}(C_2 - D_2) \right] \tag{4.55}$$

$$\begin{aligned}\frac{\partial d_a^U}{\partial \phi_{jm}} = & \frac{\left(\frac{\partial C_3}{\partial \phi_{jm}}\mu_a(\tau_a - t_a^0) + C_3 s_a(\tau_a - t_a^0) + C_3\mu_a\frac{\partial \tau_a}{\partial \phi_{jm}}\right)}{\mu_a(\tau_a - t_a^0) + q_a(t_a^n - \tau_a)} + \\ & \frac{\left(\frac{\partial D_3}{\partial \phi_{jm}} q_a(t_a^n - \tau_a) - D_3 q_a\frac{\partial \tau_a}{\partial \phi_{jm}}\right)}{\mu_a(\tau_a - t_a^0) + q_a(t_a^n - \tau_a)} - \\ & \frac{(C_3\mu_a(\tau_a - t_a^0) + D_3 q_a(t_a^n - \tau_a))\left(s_a(\tau_a - t_a^0) + \mu_a\frac{\partial \tau_a}{\partial \phi_{jm}} - q_a\frac{\partial \tau_a}{\partial \phi_{jm}}\right)}{(\mu_a(\tau_a - t_a^0) + q_a(t_a^n - \tau_a))^2}\end{aligned} \tag{4.56}$$

where

$$\frac{\partial \tau_a}{\partial \phi_{jm}} = \frac{-\ x_a}{\Lambda_a \mu_a (x_a - \hat{x}_a)}[1 + \frac{2C\ x_a - C\ x_a^2}{(1 - x_a)^2}] + \frac{s_a \hat{x}_a (L_a^e - L_a^{r+o}(t_a^0))}{(\mu_a (x_a - \hat{x}_a))^2}$$

$\frac{\partial C_1}{\partial \phi_{jm}}$, $\frac{\partial C_2}{\partial \phi_{jm}}$ *and* $\frac{\partial C_3}{\partial \phi_{jm}}$ can be obtained from expressions (4.42)-(4.44), and

$\frac{\partial D_1}{\partial x}$, $\frac{\partial D_2}{\partial x}$ *and* $\frac{\partial D_3}{\partial x}$, $x = \theta_{jm}$, ϕ_{jm} can be obtained from expressions (4.45)-(4.50).

3. Derivatives of equilibrium flows

According to expression (4.7), the derivatives of equilibrium flows with respect to θ_{jm} , ϕ_{jm} for all links a in $\boldsymbol{L}$ can be expressed as

$$(\nabla_{\theta_{jm}} \boldsymbol{q}^\star)^T = -\ \delta\ \boldsymbol{B}\ \delta^T\ \nabla_{\theta_{jm}} \boldsymbol{c}^T \tag{4.57}$$

$$(\nabla_{\phi_{jm}} \boldsymbol{q}^\star)^T = -\ \delta\ \boldsymbol{B}\ \delta^T\ \nabla_{\phi_{jm}} \boldsymbol{c}^T \tag{4.58}$$

$$\boldsymbol{B} = (\ \delta^T\ \nabla_{\boldsymbol{q}}\ \boldsymbol{c}\ \delta\ )^{-1}(\ \boldsymbol{I} - \Delta^T(\ \Delta\ (\ \delta\ \nabla_{\boldsymbol{q}}\ \boldsymbol{c}\ \delta^T)^{-1}\Delta^T\ )^{-1}\Delta\ (\ \delta^T\ \nabla_{\boldsymbol{q}}\ \boldsymbol{c}\ \delta)^{-1}\ )$$

and by definition for each link a in $\boldsymbol{L}$ over $\boldsymbol{T}^i$ according to expression (3.42)

$$\frac{\partial c_a(q_a\ ,\ \psi)}{\partial \theta_{jm}} = \frac{\partial d_a^U}{\partial \theta_{jm}} + \frac{\partial d_a^{r+o}}{\partial \theta_{jm}} \tag{4.57a}$$

$$\frac{\partial c_a(q_a\ ,\ \psi)}{\partial \phi_{jm}} = \frac{\partial d_a^U}{\partial \phi_{jm}} + \frac{\partial d_a^{r+o}}{\partial \phi_{jm}} \tag{4.58a}$$

where $\psi = [\ \zeta\ ,\ \theta\ ,\ \phi\ ]$ and d_a^{r+o} is short for $d_a^{r+o}(t^{i+1})$ over T^i .

Also, $\dfrac{\partial d_a^{r+o}}{\partial \theta_{jm}}$, $\dfrac{\partial d_a^{U}}{\partial \theta_{jm}}$ respectively have been given in expression (4.37) for the random plus oversaturation component and in expressions (4.41), (4.47) and (4.53) for the uniform component on the various kinds of link; also $\dfrac{\partial d_a^{r+o}}{\partial \phi_{jm}}$, $\dfrac{\partial d_a^{U}}{\partial \phi_{jm}}$ have been given in expressions (4.40) for the random plus oversaturation component and in expressions (4.44), (4.50) and (4.56) for the uniform component on the corresponding types of link.

4.4 Derivatives for Downstream Links

Derivatives for the indicators of traffic conditions with respect to the start and duration of green time for each signal group at each junction are obtained in this section for the downstream links which are the immediate links taking traffic away from the links controlled by that signal group at the corresponding junction. The mathematical expressions for the indicators of traffic conditions have been expressed as follows: for the random plus oversaturation component those given by the expressions (3.21)-(3.23) in Chapter 3, and for the uniform component: for the oversaturated links, those given by expressions (3.27)-(3.29), for the undersaturated links without accumulated queues, those given by expressions (3.31), (3.33)-(3.34), and for the undersaturated links with accumulated queues, those given by expressions (3.36), (3.38)-(3.39). The derivatives of the indicators of traffic conditions with respect to θ_{jm} , ϕ_{jm} for the downstream links are now discussed. Details of derivation are given in Appendix B.

For each signal group j in $\boldsymbol{P}_m$ at any one signal controlled junction m in $\boldsymbol{N}$, the downstream links for that signal group j at the corresponding junction m form the set

$\boldsymbol{B}_{jm}$ = {b; b is the downstream link for some link $a \in \boldsymbol{A}_{jm}$ },

where A_{jm} = {a; a is the traffic stream such that $\hat{b}_{ajm}$ = 1 }

1. The random plus oversaturation component

In relation to expression (4.5a)-(4.6a), the derivatives of the indicators of traffic conditions with respect to θ_{jm} , ϕ_{jm} for each downstream link b in $\boldsymbol{B}_{jm}$ are

$$\frac{\partial D_b^{r+o}}{\partial \theta_{jm}} = \frac{\partial S_b^{r+o}}{\partial \theta_{jm}} = \frac{\partial d_b^{r+o}}{\partial \theta_{jm}} = 0 \tag{4.59}$$

$$\frac{\partial D_b^{r+o}}{\partial \phi_{jm}} = \frac{\partial S_b^{r+o}}{\partial \phi_{jm}} = \frac{\partial d_b^{r+o}}{\partial \phi_{jm}} = 0 \tag{4.60}$$

2. The uniform component.

(1). The derivatives of the indicators of traffic conditions with respect to θ_{jm} , ϕ_{jm} , for each downstream link b in $\boldsymbol{B}_{jm}$, when $x_b \geq 1$, are

$$\frac{\partial D_b^{U}}{\partial \theta_{jm}} = \frac{\partial S_b^{U}}{\partial \theta_{jm}} = \frac{\partial d_b^{U}}{\partial \theta_{jm}} = 0 \tag{4.61}$$

$$\frac{\partial D_b^{U}}{\partial \phi_{jm}} = \frac{\partial S_b^{U}}{\partial \phi_{jm}} = \frac{\partial d_b^{U}}{\partial \phi_{jm}} = 0 \tag{4.62}$$

(2). The derivatives of the delay per unit time, i.e. D_b^U , and stops per unit time, i.e. S_b^U with respect to θ_{jm} , ϕ_{jm} , for each downstream link b in $\boldsymbol{B}_{jm}$, when

$x_b < 1$, $L_b^{r+o}(t_b^0) < L_b^e$, were derived by Wong (1993, pp269-271) and are, when expressed in the notation of this thesis, as follows. The derivative of the delay per vehicle, i.e. d_b^U takes the form shown because $x_b < 1$.

$$\frac{\partial D_b^U}{\partial \theta_{jm}} = \zeta \left(\int_{t_b^0}^{z_b} \frac{\partial I_b(t)}{\partial \theta_{jm}} (z_b - t)dt \right) \tag{4.63}$$

$$\frac{\partial S_b^U}{\partial \theta_{jm}} = \left[\frac{\zeta s_b}{s_b - I_b(z_b)} \right] \int_{t_b^0}^{z_b} \frac{\partial I_b(w)}{\partial \theta_{jm}} dw \tag{4.64}$$

$$\frac{\partial d_b^U}{\partial \theta_{jm}} = \frac{1}{q_b} \frac{\partial D_b^U}{\partial \theta_{jm}} \tag{4.65}$$

$$\frac{\partial D_b^U}{\partial \phi_{jm}} = \zeta \left(\int_{t_b^0}^{z_b} \frac{\partial I_b(t)}{\partial \phi_{jm}} (z_b - t)dt \right) \tag{4.66}$$

$$\frac{\partial S_b^U}{\partial \phi_{jm}} = \left[\frac{\zeta s_b}{s_b - I_b(z_b)} \right] \int_{t_b^0}^{z_b} \frac{\partial I_b(w)}{\partial \phi_{jm}} dw \tag{4.67}$$

$$\frac{\partial d_b^U}{\partial \phi_{jm}} = \frac{1}{q_b} \frac{\partial D_b^U}{\partial \phi_{jm}} \tag{4.68}$$

where $\frac{\partial I_b(t)}{\partial \theta_{jm}}$, $\frac{\partial I_b(t)}{\partial \phi_{jm}}$ are given by expressions (b.36)-(b.37), (b.38) and (b.42) when the degree of saturation for the upstream link is taken into account.

(3)(i). The derivatives of the indicators of traffic conditions with respect to θ_{jm} , ϕ_{jm} , for each downstream link b in $\boldsymbol{B}_{jm}$ when $x_b < 1$, $L_b^{r+o}(t_b^0) \geq L_b^e$, and $\tau_b \geq t_b^L$ are the same as in expressions (4.61)-(4.62).

(ii). The derivatives of the indicators for the performance index with respect to θ_{jm} , ϕ_{jm} , for each downstream link b in $\boldsymbol{B}_{jm}$ when $x_b < 1$, $L_b^{r+o}(t_b^0) \geq L_b^e$, and $\tau_b < t_b^L$, can be expressed as linear combinations of the derivatives (each zero) for the oversaturated links and those of indicators D_1 , D_2 *and* D_3 for the undersaturated links in the following way.

$$\frac{\partial D_b^U}{\partial \theta_{jm}} = \frac{1}{t_b^L} \left[\frac{\partial D_1}{\partial \theta_{jm}} (t_b^n - \tau_b) \right] \tag{4.69}$$

$$\frac{\partial S_b^U}{\partial \theta_{jm}} = \frac{1}{t_b^L} \left[\frac{\partial D_2}{\partial \theta_{jm}} (t_b^n - \tau_b) \right] \tag{4.70}$$

$$\frac{\partial d_b^U}{\partial \theta_{jm}} = \frac{\frac{\partial D_3}{\partial \theta_{jm}} q_b (t_b^n - \tau_b)}{\mu_b(\tau_b - t_b^0) + q_b(t_b^n - \tau_b)} \tag{4.71}$$

where $\frac{\partial D_1}{\partial \theta_{jm}}$, $\frac{\partial D_2}{\partial \theta_{jm}}$ *and* $\frac{\partial D_3}{\partial \theta_{jm}}$ are equal to expressions (4.63)-(4.65) respectively.

$$\frac{\partial D_b^U}{\partial \phi_{jm}} = \frac{1}{t_b^L} \left[\frac{\partial D_1}{\partial \phi_{jm}} (t_b^n - \tau_b) \right] \tag{4.72}$$

$$\frac{\partial S_b^U}{\partial \phi_{jm}} = \frac{1}{t_b^L}\left[\frac{\partial D_2}{\partial \phi_{jm}}(t_b^n - \tau_b) \right] \tag{4.73}$$

$$\frac{\partial d_b^U}{\partial \phi_{jm}} = \frac{\frac{\partial D_3}{\partial \phi_{jm}} q_b(t_b^n - \tau_b)}{\mu_b(\tau_b - t_b^0) + q_b(t_b^n - \tau_b)} \tag{4.74}$$

where $\frac{\partial D_1}{\partial \phi_{jm}}$, $\frac{\partial D_2}{\partial \phi_{jm}}$ *and* $\frac{\partial D_3}{\partial \phi_{jm}}$ are equal to expressions (4.66)-(4.68) respectively.

4.5 Derivatives for Further Downstream Links

Derivatives for the indicators of traffic conditions in relation to expressions (4.5a)-(4.6a) with respect to the start and duration of green time are discussed in this section for the further downstream links which are the immediate links after the downstream links. The mathematical expressions for the indicators of traffic conditions have been expressed as follows: for the random plus oversaturation component those given by expressions (3.21)-(3.23) in Chapter 3, and for the uniform component: for the oversaturated links, those given by expressions (3.27)-(3.29), for the undersaturated links without accumulated queues, those are given by expressions (3.31), (3.33)-(3.34), and for the undersaturated links with the accumulated queues, those are given by expressions (3.36), (3.38)-(3.39). The derivatives of the indicators of traffic conditions with respect to θ_{jm} , ϕ_{jm} for the further downstream links are now discussed. Details of derivation are given in Appendix B.

For each signal group j in $\boldsymbol{P}_m$ at any one signal controlled junction m in $\boldsymbol{N}$, the further downstream links for that signal group j at the corresponding junction m form the set

$\boldsymbol{C}_{jm}$ = {c; c is the immediate downstream link for some link $b \in \boldsymbol{B}_{jm}$ },

where $\boldsymbol{B}_{jm}$ = {b; b is the downstream link for some link $a \in \boldsymbol{A}_{jm}$ }

and $\boldsymbol{A}_{jm}$ = {a; a is the traffic stream such that $\hat{b}_{ajm} = 1$ }

1. The random plus oversaturation component

In relation to expression (4.5a)-(4.6a), the derivatives of the indicators of traffic conditions with respect to θ_{jm} , ϕ_{jm} for each further downstream link c are

$$\frac{\partial D_c^{r+o}}{\partial \theta_{jm}} = \frac{\partial S_c^{r+o}}{\partial \theta_{jm}} = \frac{\partial d_c^{r+o}}{\partial \theta_{jm}} = 0 \tag{4.75}$$

$$\frac{\partial D_c^{r+o}}{\partial \phi_{jm}} = \frac{\partial S_c^{r+o}}{\partial \phi_{jm}} = \frac{\partial d_c^{r+o}}{\partial \phi_{jm}} = 0 \tag{4.76}$$

2. The uniform component.

(1). The derivatives of the indicators of traffic conditions with respect to θ_{jm} , ϕ_{jm} , for each further downstream link c in $\boldsymbol{C}_{jm}$ when $x_c \geq 1$ are

$$\frac{\partial D_c^{U}}{\partial \theta_{jm}} = \frac{\partial S_c^{U}}{\partial \theta_{jm}} = \frac{\partial d_c^{U}}{\partial \theta_{jm}} = 0 \tag{4.77}$$

$$\frac{\partial D_c^{U}}{\partial \phi_{jm}} = \frac{\partial S_c^{U}}{\partial \phi_{jm}} = \frac{\partial d_c^{U}}{\partial \phi_{jm}} = 0 \tag{4.78}$$

(2). (i). The derivatives of the indicators of traffic conditions with respect to θ_{jm} , ϕ_{jm} , for each further downstream link c in $\boldsymbol{C}_{jm}$ when

$x_c < 1$, $L_c^{r+o}(t_c^0) < L_c^e$, and $x_a < 1$, $\forall\, a \in \boldsymbol{A}_{jm}$ are

$$\frac{\partial D_c^U}{\partial \theta_{jm}} = \frac{\partial S_c^U}{\partial \theta_{jm}} = \frac{\partial d_c^U}{\partial \theta_{jm}} = 0 \tag{4.79}$$

$$\frac{\partial D_c^U}{\partial \phi_{jm}} = \frac{\partial S_c^U}{\partial \phi_{jm}} = \frac{\partial d_c^U}{\partial \phi_{jm}} = 0 \tag{4.80}$$

(ii). The derivatives of the indicators of traffic conditions with respect to θ_{jm} , ϕ_{jm} ,

for each further downstream link c in $\boldsymbol{C}_{jm}$ when

$x_c < 1$, $L_c^{r+o}(t_c^0) < L_c^e$, and $x_a \geq 1$, $\forall\, a \in \boldsymbol{A}_{jm}$ are

$$\frac{\partial D_c^U}{\partial \theta_{jm}} = \frac{\partial S_c^U}{\partial \theta_{jm}} = \frac{\partial d_c^U}{\partial \theta_{jm}} = 0 \tag{4.81}$$

$$\frac{\partial D_c^U}{\partial \phi_{jm}} = \zeta \int_{t_c^0}^{z_c} \frac{\partial I_c(t)}{\partial \phi_{jm}} (z_c - t)dt \tag{4.82}$$

$$\frac{\partial S_c^U}{\partial \phi_{jm}} = \left[\frac{\zeta\, s_c}{s_c - I_c(z_c)} \right] \int_{t_c^0}^{z_c} \frac{\partial I_c(w)}{\partial \phi_{jm}} dw \tag{4.83}$$

$$\frac{\partial d_c^U}{\partial \phi_{jm}} = \frac{1}{q_c} \frac{\partial D_c^U}{\partial \phi_{jm}} \tag{4.84}$$

where $\dfrac{\partial I_c(t)}{\partial \phi_{jm}}$ are given by expression (b.42).

(3)(i). The derivatives of the indicators of traffic conditions with respect to

θ_{jm} , ϕ_{jm} , for each further downstream link c in $\boldsymbol{C}_{jm}$ when

$x_c < 1$, $L_c^{r+o}(t_c^0) \geq L_c^e$, $\tau_c \geq t_c^L$. The results are the same as in expressions (4.77)-(4.78).

(ii) The derivatives of the indicators for the performance index with respect to

θ_{jm} , ϕ_{jm} , for each further downstream link c in $\boldsymbol{C}_{jm}$ when

$x_c < 1$, $L_c^{r+o}(t_c^0) \geq L_c^e$, $\tau_c < t_c^L$, can be expressed as linear combinations of the derivatives for the oversaturated links and those for the undersaturated links and therefore if $x_a < 1$, $\forall a \in \boldsymbol{A}_{jm}$

$$\frac{\partial D_c^U}{\partial \theta_{jm}} = \frac{\partial S_c^U}{\partial \theta_{jm}} = \frac{\partial d_c^U}{\partial \theta_{jm}} = 0 \tag{4.85}$$

$$\frac{\partial D_c^U}{\partial \phi_{jm}} = \frac{\partial S_c^U}{\partial \phi_{jm}} = \frac{\partial d_c^U}{\partial \phi_{jm}} = 0 \tag{4.86}$$

and if $x_a \geq 1$, $\forall a \in \boldsymbol{A}_{jm}$

$$\frac{\partial D_c^U}{\partial \theta_{jm}} = \frac{\partial S_c^U}{\partial \theta_{jm}} = \frac{\partial d_c^U}{\partial \theta_{jm}} = 0 \tag{4.87}$$

$$\frac{\partial D_c^U}{\partial \phi_{jm}} = \frac{1}{t_c^L}\left[\frac{\partial D_1}{\partial \phi_{jm}}(t_c^n - \tau_c) \right] \tag{4.88}$$

$$\frac{\partial S_c^U}{\partial \phi_{jm}} = \frac{1}{t_c^L}\left[\frac{\partial D_2}{\partial \phi_{jm}}(t_c^n - \tau_c) \right] \tag{4.89}$$

$$\frac{\partial d_c^U}{\partial \phi_{jm}} = \frac{\frac{\partial D_3}{\partial \phi_{jm}} q_c(t_c^n - \tau_c)}{\mu_c(\tau_c - t_c^0) + q_c(t_c^n - \tau_c)} \tag{4.90}$$

where $\frac{\partial D_1}{\partial \phi_{jm}}$, $\frac{\partial D_2}{\partial \phi_{jm}}$ *and* $\frac{\partial D_3}{\partial \phi_{jm}}$ can be obtained from expressions (4.82)-(4.84).

4.6 Conclusions

In this Chapter, a sensitivity analysis approach has been used in obtaining the relevant derivatives of the indicators of traffic conditions with respect to the signal setting variables and link flows. According to the characteristics of the signal setting variables, three groups have been classified for the corresponding derivatives with respect to the signal setting variables: the derivatives for all links with respect to the common cycle time and link flows which have been given in Section 4.2, the derivatives for the upstream links which are controlled by a particular signal group at the corresponding junction which have been given in Section 4.3, and the derivatives for the downstream and further downstream links which are indirectly affected by that particular signal group and the corresponding results have been given in Section 4.4 and 4.5 respectively.

Using the derivatives with respect to the signal timings and link flows provided by the sensitivity analysis, gradient based solution methods can be discussed. In the following chapter, the solution method for the bi-level formulation of the area traffic control for equilibrium flows will be proposed on the basis of the derivatives obtained in this Chapter.

Chapter 5 Solution Methods

5.0 Introduction

Using the derivatives obtained by the sensitivity analysis in Chapter 4, the solution method for the bi-level formulation of area traffic control optimisation for equilibrium network flows is given in the following sections. Firstly, in Section 5.1, a single level optimisation problem subject to linear constraints for the bi-level problem given in Chapter 3 is expressed as a matrix form. Secondly, in Section 5.2, the gradient based solution method which is in terms of the derivatives of the indicators for the performance index with respect to the signal timings is discussed accordingly, in which the gradient projection method is used as an appropriate tool in deciding the descent direction for the constrained optimisation problem at current signal timings. Thirdly, in Section 5.3, the determination of the step length along the descent direction for current signal timings is discussed. As we have mentioned earlier, only a local optimum can be found in this way for the constrained optimisation problem due to the non-convex characteristics of the objective function. To find a better local optimum for the constrained optimisation problem, a global search for the step length of equal and simultaneous junction-specific changes in the starts of green times at the various junctions is made by solving an unconstrained optimisation problem for the offset variables. A simultaneous and equal change in the starts of green which corresponds to the change in the offsets can help us to find a region containing a good local solution to the optimisation problem. Therefore in Section 5.4 a mixed search procedure is proposed for the optimisation problem. Conclusion is given in Section 5.5.

5.1 Problem Formulation

Recall the bi-level problem which has been expressed in (4.0), that is

$$\underset{\zeta\,,\,\theta\,,\,\phi}{Minimise} \quad P = P_0(\zeta\ ,\ \theta\ ,\ \phi\ ,\ \boldsymbol{q})$$

subject to the constraints specified in expressions (3.3)-(3.6), and to $\boldsymbol{q} = \boldsymbol{q}^{\star}(\zeta\ ,\ \theta\ ,\ \phi)$.

Let $K = 2\sum_{m=1}^{N_J} N_{pm} + 1$ be the dimension of all signal control variables, and

$\boldsymbol{\psi} = [\ \zeta\ ,\ \boldsymbol{\theta}_1\ ,\ \boldsymbol{\phi}_1\ ,\ \theta_2\ ,\ \boldsymbol{\phi}_2\ ,\ \ldots\ ,\ \boldsymbol{\theta}_m\ ,\ \boldsymbol{\phi}_m\ ,\ \ldots\ ,\ \boldsymbol{\theta}_{N_J}\ ,\ \boldsymbol{\phi}_{N_J}\ ]$ be the row vector of all signal control variables in dimension of K , where $\boldsymbol{\theta}_m = [\theta_{1\ m}, \theta_{2\ m}, \ldots, \theta_{N_{pm}\ m}]$ and $\boldsymbol{\phi}_m = [\phi_{1\ m}, \phi_{2\ m}, \ldots, \phi_{N_{pm}\ m}]$ are respectively the row vectors of the starts and durations of green for all signal groups at junction m.

Let $\boldsymbol{A}' = [\ \boldsymbol{A}_{m\ ,\ 1 \le m \le N_J}\ ]$ be a diagonal supermatrix with matrices $\boldsymbol{A}_m$, $1 \le m \le N_J$ as its non-zero submatrices, where $\boldsymbol{A}_m$ represents the coefficient matrix in relation to the signal control variables $\boldsymbol{\theta}_m$, $\boldsymbol{\phi}_m$ at junction m in the constraints (3.4)-(3.6), and

$\boldsymbol{a} = \begin{bmatrix} \boldsymbol{a}_1 \\ \cdot \\ \boldsymbol{a}_m \\ \cdot \\ \cdot \\ \boldsymbol{a}_{N_J} \end{bmatrix}$ be the column vector of $\boldsymbol{a}_m$, $1 \le m \le N_J$, where $\boldsymbol{a}_m$ represents the coupling vector of coefficients of ζ at junction m in the constraints (3.4) and (3.6). Let $\boldsymbol{A}$ be the matrix of coefficients of signal timings ψ in the constraints that result from reducing matrix $\begin{bmatrix} \boldsymbol{a}_0 & \boldsymbol{0} \\ \boldsymbol{a} & \boldsymbol{A}' \end{bmatrix}$ to full rank without changing the feasible region by deleting any constraint that is linearly dependent on others and by omitting all but the strictest of any set of constraints with identical left hand sides, where $\boldsymbol{a}_0$ represents the coupling vector of coefficients of ζ in the constraint (3.3), and $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_0 \\ \cdot \\ \boldsymbol{b}_m \\ \cdot \\ \cdot \\ \boldsymbol{b}_{N_J} \end{bmatrix}$ be a column vector of $\boldsymbol{b}_0$ and $\boldsymbol{b}_m$, $1 \le m \le N_J$, where $\boldsymbol{b}_m$ represents the constant vector for the constraints (3.4)-(3.6) at junction m as it remains after reducing matrix $\boldsymbol{A}$ to full rank and $\boldsymbol{b}_0$ represents the

constant vector for the cycle time constraints (3.3).

Thus the original bi-level problem can be transformed into a single level constrained optimisation problem with respect to the signal timing ψ :

$$\underset{\psi}{Minimise} \quad P = P_0(\psi\ ,\ q^{\star}(\psi)) = P_1(\psi) \tag{5.1}$$

subject to the constraints (3.3)-(3.6).

A matrix form for expression (5.1) is given as to:

$$\begin{aligned} &\underset{\psi}{Minimise} \quad P = P_1(\ \psi\ ) \\ &subject\ \ to \\ &\qquad\qquad A\ \psi^{T} \leq \boldsymbol{B} \end{aligned} \tag{5.2}$$

Furthermore, for given feasible signal timings ψ^0 which are used as initial values, let $\boldsymbol{\psi}^0 = \left[\ \zeta^0\ ,\ \boldsymbol{\theta}_1^0\ ,\ \boldsymbol{\phi}_1^0\ ,\ \boldsymbol{\theta}_2^0\ ,\ \boldsymbol{\phi}_2^0\ ,\ \ldots\ ,\ \boldsymbol{\theta}_m^0\ ,\ \boldsymbol{\phi}_m^0\ ,\ \ldots\ ,\ \boldsymbol{\theta}_{N_J}^0\ ,\ \boldsymbol{\phi}_{N_J}^0\ \right]$ be the row vector of initial values of signal control variables, where $\boldsymbol{\theta}_m^0 = \left[\ \theta_{1\ m}^0\ ,\ \theta_{2\ m}^0\ ,\ \ldots\ldots\ \theta_{N_{pm}\ m}^0\ \right]$ and $\boldsymbol{\phi}_m^0 = \left[\ \phi_{1\ m}^0\ ,\ \phi_{2\ m}^0\ ,\ \ldots\ldots,\ \phi_{N_{pm}\ m}^0\ \right]$ are respectively the row vectors of initial values of starts and durations of green for all signal groups at junction m, $1 \leq m \leq N_J$,

According to the techniques used in expressions (4.1)-(4.3) given in Chapter 4 to obtain the first partial derivatives of $P_1(\psi)$, for the given signal timings ψ^0 , which satisfy the constraints $A\ (\psi^0)^{T} \leq \boldsymbol{B}$, we have the followings.

<u>*<Theorem 5.1>*</u> *<finding the gradients of $P_1(\psi)$ at signal timings ψ^0 >*

Since for $A\ \psi^{T} \leq \boldsymbol{B}$, $P = P_0(\psi\ ,\ q)$, which is defined in expression (3.2) is a C^1 function on $\mathbb{R}^{K+N_L} \rightarrow \mathbb{R}^{+}$ and $q = q^{\star}(\psi)$ given by expression (3.40), is a

i.e. a function whose first derivative exists and is continuous.

C^1 *function on* $\mathbb{R}^K \rightarrow \mathbb{R}^{N_L}$, $P_1(\psi)$ *is a* C^1 *function on* $\mathbb{R}^K \rightarrow \mathbb{R}^+$, *that is, all the first partial derivatives of* $P_1(\psi)$ *with respect to* $\boldsymbol{\psi} = (\zeta\ ,\ \boldsymbol{\theta}\ ,\ \boldsymbol{\phi})$ *exist and are continuous at any given point* ψ^0 *in* $\mathbb{R}^K$ *that satisfies* $\boldsymbol{A}\ (\boldsymbol{\psi}^0)^T \leq \boldsymbol{B}$. *If* $\boldsymbol{q}^0 = \boldsymbol{q}^\star(\boldsymbol{\psi}^0)$ *then*

$$\frac{\partial P_1(\psi^0)}{\partial \zeta} = \frac{\partial P_0(\psi^0,\ q^0)}{\partial \zeta} + \frac{\partial P_0(\psi^0,\ q^0)}{\partial q^\star}\ \frac{\partial q^\star(\psi^0)}{\partial \zeta} \tag{5.3}$$

$$\frac{\partial P_1(\psi^0)}{\partial \theta} = \frac{\partial P_0(\psi^0,\ q^0)}{\partial \theta} + \frac{\partial P_0(\psi^0,\ q^0)}{\partial q^\star}\ \frac{\partial q^\star(\psi^0)}{\partial \theta} \tag{5.4}$$

$$\frac{\partial P_1(\psi^0)}{\partial \phi} = \frac{\partial P_0(\psi^0,\ q^0)}{\partial \phi} + \frac{\partial P_0(\psi^0,\ q^0)}{\partial q^\star}\ \frac{\partial q^\star(\psi^0)}{\partial \phi} \tag{5.5}$$

■

<u>*<Corollary 5.2>*</u>*<following Theorem 5.1>*

For each $j \in \boldsymbol{P}_m$ *and* $m \in \boldsymbol{N}$, *let*

$\boldsymbol{A}_{jm} = \{$ *a; a is a traffic stream such that* $\hat{b}_{ajm} = 1$ $\}$

$\boldsymbol{B}_{jm} = \{$ *b; b is a downstream link for some link* $a \in \boldsymbol{A}_{jm}$ $\}$

$\boldsymbol{C}_{jm} = \{$ *c; c is a further downstream link for some link* $a \in \boldsymbol{B}_{jm}$ $\}$

and $\boldsymbol{L}_{jm} = \boldsymbol{L} \cap (\boldsymbol{A}_{jm} \cup \boldsymbol{B}_{jm} \cup \boldsymbol{C}_{jm})$, *for a given point* $(\boldsymbol{\psi}^0,\ \boldsymbol{q}^0)$ *that satisfies* $\boldsymbol{A}\ (\boldsymbol{\psi}^0)^T \leq \boldsymbol{B}$ *and* $\boldsymbol{q}^0 = \boldsymbol{q}^\star(\boldsymbol{\psi}^0)$ *according to expressions (4.4)-(4.6), let* $\nabla_{\boldsymbol{\theta}}\ P_1(\boldsymbol{\psi}^0)$ *and* $\nabla_{\boldsymbol{\phi}}\ P_1(\boldsymbol{\psi}^0)$ *be two vectors with elements* $\dfrac{\partial P_1(\boldsymbol{\psi}^0)}{\partial \theta_{jm}}$ *and* $\dfrac{\partial P_1(\boldsymbol{\psi}^0)}{\partial \phi_{jm}}$, $1 \leq j \leq N_{pm}$, $1 \leq m \leq N_J$.

Then the derivatives (5.3)-(5.5) can be expressed in the following way.

$$\frac{\partial P_1(\boldsymbol{\psi}^0)}{\partial \zeta} = \sum_{a \in L} \left[\left(\frac{\partial D_a}{\partial \zeta} + \sum_{l \in L} \frac{\partial D_a}{\partial q_l^\star}\ \frac{\partial q_l^\star}{\partial \zeta} \right) W_{aD} M_D + \left(\frac{\partial S_a}{\partial \zeta} + \sum_{l \in L} \frac{\partial S_a}{\partial q_l^\star}\ \frac{\partial q_l^\star}{\partial \zeta} \right) W_{aS} M_S \right] \tag{5.6}$$

$$\frac{\partial P_1(\boldsymbol{\psi}^0)}{\partial \boldsymbol{\theta}} = \nabla_\theta P_1(\boldsymbol{\psi}^0) = \left[\frac{\partial P_1(\boldsymbol{\psi}^0)}{\partial \theta_{jm}};\ j \in \boldsymbol{P}_m,\ 1 \le m \le N_J \right] \tag{5.7}$$

where each element $\dfrac{\partial P_1(\boldsymbol{\psi}^0)}{\partial \theta_{jm}}$ *can be expressed as*

$$\frac{\partial P_1(\boldsymbol{\psi}^0)}{\partial \theta_{jm}} = \sum_{a \in L_{j[m]}} \left[\left(\frac{\partial D_a}{\partial \theta_{jm}} + \sum_{l \in L} \frac{\partial D_a}{\partial q_l^\star} \frac{\partial q_l^\star}{\partial \theta_{jm}} \right) W_{aD} M_D + \left(\frac{\partial S_a}{\partial \theta_{jm}} + \sum_{l \in L} \frac{\partial S_a}{\partial q_l^\star} \frac{\partial q_l^\star}{\partial \theta_{jm}} \right) W_{aS} M_S \right] \tag{5.7a}$$

and

$$\frac{\partial P_1(\boldsymbol{\psi}^0)}{\partial \boldsymbol{\phi}} = \nabla_\phi P_1(\boldsymbol{\psi}^0) = \left[\frac{\partial P_1(\boldsymbol{\psi}^0)}{\partial \phi_{jm}};\ j \in \boldsymbol{P}_m,\ 1 \le m \le N_J \right] \tag{5.8}$$

where each element $\dfrac{\partial P_1(\boldsymbol{\psi}^0)}{\partial \phi_{jm}}$ *can be expressed as follows.*

$$\frac{\partial P_1(\boldsymbol{\psi}^0)}{\partial \phi_{jm}} = \sum_{a \in L_{j[m]}} \left[\left(\frac{\partial D_a}{\partial \phi_{jm}} + \sum_{l \in L} \frac{\partial D_a}{\partial q_l^\star} \frac{\partial q_l^\star}{\partial \phi_{jm}} \right) W_{aD} M_D + \left(\frac{\partial S_a}{\partial \phi_{jm}} + \sum_{l \in L} \frac{\partial S_a}{\partial q_l^\star} \frac{\partial q_l^\star}{\partial \phi_{jm}} \right) W_{aS} M_S \right] \tag{5.8a}$$

■

The following theorem is related to Taylor's theorem for a real-valued function of m variables. (which is referred to Baxandall and Liebeck (1986), Theorem 3.11.7, pp164)

<u>*<5.3 Taylor's Theorem>*</u>

Let $\boldsymbol{p}$ *and* $\boldsymbol{p} + \boldsymbol{h}$ *be two distinct points in an open subset* $\boldsymbol{D}$ *of* $\mathbb{R}^m$ *such that the straight line segment joining* $\boldsymbol{p}$ *and* $\boldsymbol{p} + \boldsymbol{h}$ *lies in* $\boldsymbol{D}$ *. Let* $\boldsymbol{u}$ *be the unit vector in direction* $\boldsymbol{h}$ *. Consider a function* $f : \boldsymbol{D} \subseteq \mathbb{R}^m \to \mathbb{R}$ *such that, for some* $n \in \mathbb{N}$ *, the functions* f *,* $\dfrac{\partial f}{\partial u}$ *, ...,* $\dfrac{\partial^{n-1} f}{\partial u^{n-1}}$ *are all differentiable in* $\boldsymbol{D}$ *. Then there exists* $0 < \lambda < 1$ *such that*

$$f(\boldsymbol{p} + \boldsymbol{h}) = f(\boldsymbol{p}) + \| \boldsymbol{h} \| \frac{\partial f}{\partial \boldsymbol{u}} (\boldsymbol{p}) + \frac{\| \boldsymbol{h} \|^2}{2!} \frac{\partial^2 f}{\partial \boldsymbol{u}^2}(\boldsymbol{p}) + \ldots\ldots +$$

$$\frac{\| \boldsymbol{h} \|^{n-1} \partial^{n-1} f}{(n-1)!\, \partial \boldsymbol{u}^{n-1}}(\boldsymbol{p}) + \frac{\| \boldsymbol{h} \|^n \partial^n f}{n!\, \partial \boldsymbol{u}^n}(\boldsymbol{p} + \lambda \boldsymbol{h}) \tag{5.9}$$

where $\| \boldsymbol{h} \|$ *is given by* $\boldsymbol{h} = \| \boldsymbol{h} \| \boldsymbol{u}$ and $\frac{\partial f(p)}{\partial u} = u \nabla f(p)^T$ & $\frac{\partial^n f(p)}{\partial u^n} = (u \nabla)^n f(p)^T$ ■

<u><Corollary 5.4></u> *<Linear approximation of* $\boldsymbol{q}$ *at* ψ^0 *>*

$\boldsymbol{q} = \boldsymbol{q}^\star(\psi)$ *is a* C^1 *function* $\mathbb{R}^K \rightarrow \mathbb{R}^{N_L}$ *where* N_L *is the number of links in* $\boldsymbol{G}$ *.*

Let ψ^0 *and* $\psi^0 + \boldsymbol{h}$ *be two distinct points in an open set* $\boldsymbol{D}$ *such that the straight line segment joining* ψ^0 *and* $\psi^0 + \boldsymbol{h}$ *in* $\boldsymbol{D}$ *and let* $\boldsymbol{q}^0 = \boldsymbol{q}^\star(\psi^0)$ *. Then the linear approximation of* $\boldsymbol{q}$ *can be expressed below as a Taylor polynomial of degree 1 at* ψ^0 *.*

$$\boldsymbol{q}^\star(\psi^0 + \boldsymbol{h})^T \simeq (\boldsymbol{q}^0)^T + \nabla_\psi \boldsymbol{q}^\star(\psi^0)\, \boldsymbol{h}^T$$ ■

<u><Corollary 5.5></u> *<Linear approximation of* $P_1(\psi)$ *at* ψ^0 *>*

Since $P = P_1(\psi)$ *is a* C^1 *function:* $\boldsymbol{D} \subseteq \mathbb{R}^K \rightarrow \mathbb{R}^+$ *. Let* ψ^0 *and* $\psi^0 + \boldsymbol{h}$ *be two distinct points in an open subset* $\boldsymbol{D}$ *such that the straight line segment joining* ψ^0 *and* $\psi^0 + \boldsymbol{h}$ *lies in* $\boldsymbol{D}$ *. Let* $\boldsymbol{u}$ *be the unit vector in direction* $\boldsymbol{h}$ *. Then the linear approximation can be expressed below as a Taylor polynomial of degree 1 at* ψ^0 *.*

$$\begin{aligned} P_1(\psi^0 + \boldsymbol{h}) &\simeq P_1(\psi^0) + \| \boldsymbol{h} \| \nabla_\psi P_1(\psi^0)\, \boldsymbol{u}^T \\ &= P_1(\psi^0) + \nabla_\psi P_1(\psi^0)\, \boldsymbol{h}^T \end{aligned} \tag{5.10}$$

where

$$\nabla_{\psi} P_1(\boldsymbol{\psi}^0) = \left[\frac{\partial P_1(\boldsymbol{\psi}^0)}{\partial \zeta}, \nabla_{\theta_1} P_1(\boldsymbol{\psi}^0), \nabla_{\phi_1} P_1(\boldsymbol{\psi}^0), \nabla_{\theta_2} P_1(\boldsymbol{\psi}^0), \nabla_{\phi_2} P_1(\boldsymbol{\psi}^0), \ldots, \right.$$

$$\left. \nabla_{\theta_m} P_1(\boldsymbol{\psi}^0), \nabla_{\phi_m} P_1(\boldsymbol{\psi}^0), \ldots, \nabla_{\theta_{N_J}} P_1(\boldsymbol{\psi}^0), \nabla_{\phi_{N_J}} P_1(\boldsymbol{\psi}^0) \right]$$

and each element in $\nabla_{\psi} P_1(\psi^0)$ *is given in (5.6)-(5.8).* ■

Recall the original constrained optimisation problem which has been expressed in (5.2) as

$$\begin{aligned} &\underset{\psi}{\textit{Minimise}} \quad P = P_1(\psi) \\ &\textit{subject to} \\ &\qquad \boldsymbol{A}\, \psi^T \leq \boldsymbol{B} \end{aligned} \tag{5.2}$$

For a given point ψ^0, there exists a point $\psi^0 + \boldsymbol{h}$ in the direction $\boldsymbol{h}$, that satisfies <*Corollary 5.5*>, giving the following linear approximation for $P_1(\psi)$ *at* ψ^0.

$$P_1(\psi^0 + \boldsymbol{h}) \simeq P_1(\psi^0) + \nabla_{\psi} P_1(\psi^0)\, \boldsymbol{h}^T$$

Therefore the problem (5.2) can be approximated linearly in the following way.

$$\begin{aligned} &\underset{\boldsymbol{h}}{\textit{Minimise}} \quad P_1(\psi^0) + \nabla_{\psi} P_1(\psi^0)\, \boldsymbol{h}^T \\ &\textit{subject to} \\ &\qquad \boldsymbol{A}\, (\psi^0 + \boldsymbol{h})^T \leq \boldsymbol{B} \end{aligned} \tag{5.11}$$

Since $P_1(\psi^0)$ is a constant in expression (5.11), let $\bar{\boldsymbol{B}} = \boldsymbol{B} - \boldsymbol{A}\, (\psi^0)^T$, the linear approximation for $P = P_1(\psi)$ at ψ^0 in the direction $\boldsymbol{h}$ can be further stated as to

$$\begin{aligned} &\underset{\boldsymbol{h}}{\textit{Minimise}} \quad \nabla_{\psi} P_1(\psi^0)\, \boldsymbol{h}^T \\ &\textit{subject to} \\ &\qquad \boldsymbol{A}\, \boldsymbol{h}^T \leq \bar{\boldsymbol{B}} \end{aligned} \tag{5.12}$$

5.2 Gradient Projection Method

In expression (5.12) a local linear approximation has been given for the constrained optimisation problem (5.2) at ψ^0 . The way of finding the descent direction $\boldsymbol{h}$ at current signal timings ψ^0 such that by repeated steps of the same kind a local optimum can be found is discussed as follows. In particular, we refer to the *gradient projection method* as stated by Luenberger (1989, pp330-337) and focus on two points: firstly the way of deciding an improving feasible direction, and secondly how the descent direction can be decided by the *gradient projection method* for the optimisation problem (5.12).

<u>*<Definition 5.6>*</u> *<Feasible direction for $P_1(\psi)$ at ψ^0 >*

In problem (5.2)

$$\begin{aligned} &\underset{\psi}{\text{Minimise}} \quad P = P_1(\psi) \\ &\text{subject to} \\ &\qquad A\ \psi^T \le \boldsymbol{B} \end{aligned} \tag{5.2}$$

for a given point ψ^0 such that $\boldsymbol{A}\ (\psi^0)^T \le \boldsymbol{B}$, a non-zero vector $\boldsymbol{h}$ is defined as a feasible direction for $P_1(\psi^0)$ in (5.12) if there exists a $\delta > 0$ such that

$$\boldsymbol{A}\ (\psi^0 + \alpha\ \boldsymbol{h})^T \le \boldsymbol{B}\ , \quad \forall\ \alpha \in (0\ ,\ \delta) \tag{5.13}$$

■

<u>*<Definition 5.7>*</u> *<Improving feasible direction for $P_1(\psi)$ at ψ^0 >*

In problem (5.2)

$$\begin{aligned} &\underset{\psi}{\text{Minimise}} \quad P = P_1(\psi) \\ &\text{subject to} \\ &\qquad A\ \psi^T \le \boldsymbol{B} \end{aligned} \tag{5.2}$$

for a given point ψ^0 *such that* $\boldsymbol{A}\ (\psi^0)^T \leq \boldsymbol{B}$ *, a feasible direction* $\boldsymbol{h}$ *satisfying expression (5.13) is defined as an improving feasible direction for* $P_1(\psi^0)$ *in (5.12) if there exists* $\delta > 0$ *such that* $P_1(\psi^0 + \alpha\ \boldsymbol{h}) < P_1(\psi^0)$, $\forall\ \alpha \in (0\ ,\ \delta)$ ■

<u>*<Theorem 5.8>*</u> *<Following Definition 5.6>*

In problem (5.2)

$$\begin{aligned} &\underset{\psi}{\text{Minimise}} \quad P = P_1(\psi) \\ &\text{subject to} \\ &\qquad \boldsymbol{A}\ \psi^T \leq \boldsymbol{B} \end{aligned} \tag{5.2}$$

for a given point ψ^0 *such that* $\boldsymbol{A}\ (\psi^0)^T \leq \boldsymbol{B}$ *, if* $\boldsymbol{A}^T$ *and* $\boldsymbol{B}^T$ *are decomposed into* $\left[\ \boldsymbol{A}_b{}^T\ ,\ \boldsymbol{A}_{nb}{}^T\ \right]$ *and* $\left[\ \boldsymbol{B}_b{}^T\ ,\ \boldsymbol{B}_{nb}{}^T\ \right]$ *such that* $\boldsymbol{A}_b\ (\psi^0)^T = \boldsymbol{B}_b$ *and* $\boldsymbol{A}_{nb}\ (\psi^0)^T < \boldsymbol{B}_{nb}$ *, then a non-zero vector* $\boldsymbol{h}$ *is a feasible direction for* $P_1(\psi^0)$ *in (5.12) if and only if* $\boldsymbol{A}_b\ \boldsymbol{h}^T \leq \boldsymbol{0}$.

<Proof>

By the decomposition of constraints into binding and non-binding constraints at ψ^0

$\boldsymbol{A}_b\ (\psi^0)^T = \boldsymbol{B}_b$ and $\boldsymbol{A}_{nb}\ (\psi^0)^T < \boldsymbol{B}_{nb}$, for any a non-zero vector $\boldsymbol{h}$ there exists

a $\delta > 0$ such that $\boldsymbol{A}_{nb}\ (\psi^0 + \alpha\ \boldsymbol{h})^T \leq \boldsymbol{B}_{nb}$, $\forall\ \alpha \in (0\ ,\ \delta)$ (5.14)

Therefore, according to *<Definition 5.6>* a non-zero vector $\boldsymbol{h}$ is a feasible direction for

$P_1(\psi)$ at ψ^0 if and only if $\boldsymbol{A}_b\ (\psi^0 + \alpha\ \boldsymbol{h})^T \leq \boldsymbol{B}_b$, $\forall\ \alpha \in (0\ ,\ \delta)$

i.e. if & only if $\boldsymbol{A}_b\ \boldsymbol{h}^T \leq \boldsymbol{0}$ (5.15)

■

<Theorem 5.9> *<Following Definition 5.7>*

In problem (5.2)

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

$$\text{subject to} \quad \boldsymbol{A}\ \psi^T \leq \boldsymbol{B} \tag{5.2}$$

for a given point ψ^0 *such that* $\boldsymbol{A}\ (\psi^0)^T \leq \boldsymbol{B}$ *if there exists a feasible direction* $\boldsymbol{h}$ *satisfying expression (5.15) such that* $\nabla_\psi P_1(\psi^0)\ \boldsymbol{h}^T < 0$ *then* $\boldsymbol{h}$ *is an improving feasible direction for* $P_1(\psi^0)$ *in (5.12).*

<Proof>

Suppose that $\boldsymbol{h}$ satisfies the stated conditions. The condition that

$$\nabla_\psi P_1(\psi^0)\ \boldsymbol{h}^T < 0 \tag{5.16}$$

can be restated in the following way:

$$\underset{\alpha \to 0}{\text{limit}} \ \frac{P_1(\psi^0 + \alpha\ \boldsymbol{h}) - P_1(\psi^0)}{\alpha} < 0 \tag{5.17}$$

Following the *<Definition 5.7>*, a feasible direction $\boldsymbol{h}$ is defined as an improving feasible direction for $P_1(\psi^0)$ in (5.12) if and only if there exists a $\delta > 0$ such that

$$\forall\ \alpha \in (0\ ,\ \delta)\ ,\ \boldsymbol{A}\ (\psi^0 + \alpha\ \boldsymbol{h})^T \leq \boldsymbol{B} \text{ and } P_1(\psi^0 + \alpha\ \boldsymbol{h}) < P_1(\psi^0)$$

But by expression (5.17) and the condition that $\boldsymbol{h}$ satisfies expression (5.15) this holds for sufficiently small δ , and $\boldsymbol{h}$ is an improving feasible direction for $P_1(\psi^0)$ in (5.12).

■

For the problem (5.2) at signal timings ψ^0 , a descent direction $\boldsymbol{h}$ satisfying conditions (5.15) and (5.16), which improves the performance index P for the objective function

$P_1(\psi)$ and meanwhile maintains the feasibility for signal timing variables, can be decided by the *gradient projection method* (see Luenberger, 1989, pp330-337 and Bazaraa, Sherali and Shetty, 1993, pp448-454) when the first derivatives of $P_1(\psi)$ with respect to the signal timings ψ *at* ψ^0 are available.

<u><Corollary 5.10></u> *<finding a descent direction for $P_1(\psi^0)$ in (5.12)>*

$P = P_1(\psi)$ is a C^1 function $: D \subseteq \mathbb{R}^K \rightarrow \mathbb{R}^+$ which has been defined in (5.1). For any feasible point ψ^0 which satisfies $\boldsymbol{A}\,(\psi^0)^T \leq \boldsymbol{B}$ such that $\boldsymbol{A}_b\,(\psi^0)^T = \boldsymbol{B}_b$ and $\boldsymbol{A}_{nb}\,(\psi^0)^T < \boldsymbol{B}_{nb}$ where $\boldsymbol{A}^T = \left[\boldsymbol{A}_b^T, \boldsymbol{A}_{nb}^T\right]$, $\boldsymbol{B}^T = \left[\boldsymbol{B}_b^T, \boldsymbol{B}_{nb}^T\right]$, and the problem (5.2) then can be reexpressed below as the corresponding form (c.4) given in <Theorem c.5> in Appendix C, with the set of equality constraints $\boldsymbol{H}(\psi)$ empty.

$$\begin{aligned} &\underset{\psi}{\text{Minimise}} \quad P = P_1(\psi) \\ &\text{subject to} \quad \boldsymbol{G}(\psi) = \boldsymbol{A}\,\psi^T - \boldsymbol{B} \leq 0 \end{aligned} \tag{5.2}$$

Let $\boldsymbol{G}_b(\psi^0) = \boldsymbol{A}_b\,(\psi^0)^T - \boldsymbol{B}_b = \boldsymbol{0}$, if $\boldsymbol{M} = \boldsymbol{A}_b$ has full rank, i.e. the set of the binding constraints at any feasible point ψ^0 are linearly independent, and if $\boldsymbol{g}$ is of the form $\boldsymbol{g} = \boldsymbol{I} - \boldsymbol{M}^T(\boldsymbol{M}\boldsymbol{M}^T)^{-1}\boldsymbol{M}$ such that $\boldsymbol{g}\,\nabla_\psi P_1(\psi^0)^T \neq \boldsymbol{0}$ following the results of <Lemma c.8> in Appendix C, then $\boldsymbol{h}$ is an improving feasible direction for $P_1(\psi^0)$ in (5.12) and is given by $\boldsymbol{h}^T = -\,\boldsymbol{g}\,\nabla_\psi P_1(\psi^0)^T$ (5.18)

■

Given a point ψ^0 , a descent direction can be decided for the problem (5.12) by the *gradient projection method* as given in (5.18) when $g\ \nabla_\psi P_1(\psi^0)^T \neq \mathbf{0}$.

If $g\ \nabla_\psi P_1(\psi^0)^T = \mathbf{0}$ a new projection matrix $\hat{g}$ can be identified such that a new feasible descent direction $-\ \hat{g}\ \nabla_\psi P_1(\psi^0)^T$ can be determined.

<u>*<Corollary 5.11>*</u>

<finding a descent direction for P_1 in (5.12) when $g\ \nabla_\psi P_1(\psi^0)^T = \mathbf{0}$ >

Following <Corollary 5.10> with $g\ \nabla_\psi P_1(\psi^0)^T = \mathbf{0}$, suppose that the constraints $G(\psi) \leq \mathbf{0}$ satisfy the following Linear Independence Constraint Qualification:

(i). $G_{nb}(\psi) = A_{nb}\,(\psi)^T - B_{nb} < \mathbf{0}$ is continuous at ψ^0 .

(ii). $\nabla_\psi G_b(\psi^0)$ are linearly independent.

For any feasible point ψ^0 which satisfies $A\ (\psi^0)^T \leq B$ and does not solve problem (5.2) locally (see <Theorem c.5>), if $\bar{A}_b$ is obtained from A_b by deleting the rows of A_b corresponding to negative Lagrange multipliers at ψ^0 and $\hat{M} = \bar{A}_b$ then following the results of <Lemma c.9> in Appendix C, when $g\ \nabla_\psi P_1\ (\psi^0)^T = \mathbf{0}$ an improving feasible direction for $P_1(\psi^0)$ in (5.12) can be expressed as

$$-\ \hat{g}\ \nabla_\psi P_1(\psi^0)^T \tag{5.19}$$

where $\hat{g}$ is the projection matrix $\hat{g} = I - \hat{M}^T\,(\hat{M}\hat{M}^T)^{-1}\,\hat{M}$ ■

<Corollary 5.12>

<finding a new feasible point ψ_{new} *>*

A new feasible point ψ_{new} *for* $P = P_1(\psi)$ *at current point* ψ^0 *can be decided when the direction of the change in the signal timings is* $\boldsymbol{h}$ *. Following the results obtained from Corollaries 5.10 and 5.11, a new feasible point* ψ_{new} *can be identified as follows.*

$$\psi_{new} = \psi^0 + \alpha \boldsymbol{h}$$

where α *is a small multiplier along the direction* $\boldsymbol{h}$ *.* ■

<Corollary 5.13> <Stopping condition>

Let ψ_{opt} *locally solve the problem* $P_1(\psi)$ *in (5.2) and the constraints in (5.2) satisfy the Linear Independence Constraint Qualification then* ψ_{opt} *is a KKT point for problem* (see Theorem C.5, pp. 226) $P_1(\psi)$ *in (5.2). Following the gradient projection method that is given in Corollaries 5.10 and 5.11 in determining the feasible descent direction for problem* $P_1(\psi)$ *in (5.12) at current point* ψ_{opt} *this search procedure will terminate only at a KKT point according to the <Lemma c.10> in Appendix C, i.e. in a set of points that contains all locally optimal points, which is regarded as a plausible heuristic for solving this non-linear and non-convex problem (5.2).* ■

5.3 Determination of Step Length

Following the foregoing discussions, suppose now we have a given point ψ^0 and an improving feasible direction $\boldsymbol{h}$ for $P_1(\psi^0)$ in (5.12) then the distance in this direction for which the better improvement can be achieved is given by the solution of the one-dimensional search problem. In the following subsections, firstly a way in deciding the maximum step length along the feasible descent direction is discussed for which the

feasibility of the solution for the problem (5.12) is maintained. Secondly, given by the maximum step length obtained from section 5.3.1, a good step length can be determined for which the better improvement along the feasible direction for problem (5.12) can be found. Two approaches of finding the good step length are used to solve the one-dimensional search problem in the following sections: one approach uses the derivative information of the objective function and then the bisection method is used; the other performs a uniform search and evaluates the objective function over the whole feasible interval without using the derivative information and the minimum value of the objective function in this direction can thus be located approximately.

5.3.1 Determination of the maximum step length

Given a feasible point ψ^0 and an improving feasible direction $\boldsymbol{h}$ for $P_1(\psi^0)$ in (5.12), a one-dimensional search problem for (5.12) can be expressed as

$$\begin{aligned} &\underset{\alpha \geq 0}{\textit{Minimise}} \quad P = P_1(\psi^0 + \alpha\, \boldsymbol{h}) = Z_1(\alpha) \\ &\textit{subject to} \quad \boldsymbol{A}\,(\psi^0 + \alpha\, \boldsymbol{h})^T \leq \boldsymbol{B} \end{aligned} \tag{5.20}$$

<u>*<Theorem 5.14>*</u>

<finding the maximum step length along the descent direction>

In (5.20) let $\boldsymbol{A}^T$ *and* $\boldsymbol{B}^T$ *be decomposed into* $\left[\, \boldsymbol{A}_b^T , \boldsymbol{A}_{nb}^T \,\right]$ *and* $\left[\, \boldsymbol{B}_b^T , \boldsymbol{B}_{nb}^T \,\right]$

whose components respectively represent the binding and non-binding constraints at ψ^0

such that $\boldsymbol{A}_b\,(\psi^0)^T = \boldsymbol{B}_b$ *and* $\boldsymbol{A}_{nb}\,(\psi^0)^T < \boldsymbol{B}_{nb}$ *. Let* $\boldsymbol{C} = \boldsymbol{B}_{nb} - \boldsymbol{A}_{nb}(\psi^0)^T$,

$\boldsymbol{D} = \boldsymbol{A}_{nb}\boldsymbol{h}^T$ *and the ith rows of* $\boldsymbol{C}$, $\boldsymbol{D}$ *be* $C_i = B_i - \sum_{j=1}^{K} A_{ij}\psi_j^0$, $D_i = \sum_{j=1}^{K} A_{ij}h_j$

where B_i *and* A_{ij} *are respectively the ith, ijth element of* $\boldsymbol{B}_{nb}$ *and* $\boldsymbol{A}_{nb}$ *. Then (5.20) can be expressed as follows.*

$$\underset{\alpha}{Minimise} \quad P = Z_1(\alpha)$$

$$subject \ to \quad 0 \le \alpha \le \alpha_{max} \tag{5.21}$$

$$where \ \alpha_{max} = \begin{cases} \infty & if \ \boldsymbol{D} \le \boldsymbol{0} \\ Min \ \{ \ \dfrac{C_i}{D_i} : D_i > 0 \ \} & otherwise \end{cases}$$

<Proof>

As for the binding constraints, since $\boldsymbol{A}_b \, \boldsymbol{h}^T \le \boldsymbol{0}$ from *<Theorem 5.8>*, we have

$$\boldsymbol{A}_b \, (\psi^0 + \alpha \, \boldsymbol{h})^T \le \boldsymbol{B}_b \ , \ \forall \ \alpha \ge 0 \ .$$

Therefore, only the non-binding constraints need to be taken into account in setting a limit on α , that is, the constraints $\boldsymbol{A}_{nb} \, (\psi^0 + \alpha \, \boldsymbol{h})^T \le \boldsymbol{B}_{nb}$, and we require

$$\alpha \, \boldsymbol{A}_{nb} \, \boldsymbol{h}^T \le \boldsymbol{B}_{nb} - \boldsymbol{A}_{nb} \, (\psi^0)^T \ , \text{ where } \ \boldsymbol{B}_{nb} - \boldsymbol{A}_{nb} \, (\psi^0)^T > \boldsymbol{0}$$

If $\boldsymbol{D} \le \boldsymbol{0}$ this is true for all $\alpha \ge 0$ and $\alpha_{max} = \infty$. (5.22)

Otherwise, if $D_i > 0$ for any i then the corresponding constraint is satisfied so long as $\alpha \, D_i \le C_i$; and hence $\alpha_{max} = Min \ \{ \ \dfrac{C_i}{D_i} : D_i > 0 \ \}$ (5.23)

Therefore (5.20) can be reexpressed as follows.

$$\underset{\alpha}{Minimise} \quad P = Z_1(\alpha)$$

$$subject \ to \quad 0 \le \alpha \le \alpha_{max}$$

where α_{max} is given either by expression (5.22) or (5.23) according as $\boldsymbol{D} \le \boldsymbol{0}$ or not.

■

5.3.2 Determination of a good step length

Given a feasible point ψ^0 and the corresponding value of the feasible descent direction at ψ^0 for problem Z_1 in (5.21), a good step length along the feasible direction, α_{good} which makes a corresponding improvement in Z_1 can be decided as follows.

1. Bisection method

Let $Z_1'(\alpha)$ be the total derivative of P in (5.20) with respect to α , i.e. $\frac{dZ_1}{d\alpha}$ can be expressed as

$$\frac{dZ_1}{d\alpha} = \frac{dP_1}{d\psi}\left(\frac{d\psi}{d\alpha}\right)^T = \nabla_\psi P_1 \, \boldsymbol{h}^T \tag{5.24}$$

For $\psi = \psi^0 + \alpha \boldsymbol{h}$ and $0 \le \alpha \le \alpha_{max}$, following the results of (5.24) the total derivative of P with respect to α when α is evaluated at $\alpha = 0$ and $\alpha = \alpha_{max}$ can be respectively expressed as follows.

$$\begin{aligned} Z_1'(\alpha)\Big|_{\alpha=0} &= \frac{d\,Z_1(0)}{d\alpha} \\ &= \frac{d\,P_1(\psi^0)}{d\psi}\left(\frac{d\psi}{d\alpha}\right)^T \\ &= \nabla_\psi P_1(\psi^0)\,\boldsymbol{h}^T \end{aligned} \tag{5.24a}$$

$$\begin{aligned} Z_1'(\alpha)\Big|_{\alpha=\alpha_{max}} &= \frac{d\,Z_1(\alpha_{max})}{d\alpha} \\ &= \frac{d\,P_1(\psi_{max})}{d\psi}\left(\frac{d\psi}{d\alpha}\right)^T \\ &= \nabla_\psi P_1(\psi_{max})\,\boldsymbol{h}^T \end{aligned} \tag{5.24b}$$

where $\psi_{max} = \psi^0 + \alpha_{max}\,\boldsymbol{h}$

then for each α_0 in $[\ 0\ ,\ \alpha_{\max}\ ]$, $Z_1'(\alpha)\Big|_{\alpha = \alpha_0}$, the total derivative of P in (5.20) with respect to α when α has the value α_0 can be expressed as

$$
\begin{aligned}
Z_1'(\alpha)\Big|_{\alpha = \alpha_0} &= \frac{d\ Z_1(\alpha_0)}{d\alpha} \\
&= \frac{d\ P_1(\psi_0)}{d\psi}\left(\frac{d\psi}{d\alpha}\right)^T \qquad (5.24c) \\
&= \nabla_{\psi}\ P_1(\psi_0)\ \boldsymbol{h}^T
\end{aligned}
$$

where $\psi_0 = \psi^0 + \alpha_0\ \boldsymbol{h}$

The bisection method for $Z_1(\alpha)$ along the feasible direction $\boldsymbol{h}$ at current signal settings ψ^0 starting with $Z_1'(0) < 0$ can be carried out in the following steps.

<u>Step 0</u>. Set the search interval for problem Z_1 over $[\ 0\ ,\ \alpha_{\max}\ ]$ 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 $\alpha^{(k)} = \dfrac{(\alpha^l + \alpha^r)}{2}$.

<u>Step 1</u>. If $k = N_{\max}$ then take $\alpha^{(k)}$ as the good step length, i.e. $\alpha_{good} = \alpha^{(k)}$, along the feasible direction $\boldsymbol{h}$ at current signal settings ψ^0 and the new signal settings is $\psi_{new} = \psi^0 + \alpha_{good}\ \boldsymbol{h}$ and the corresponding equilibrium flows are $\boldsymbol{q}_{new} = \boldsymbol{q}^{\star}(\psi_{new})$ via the reassignment process (3.40).

Otherwise, perform evaluations of $Z_1(\alpha^{(k)})$ and of $Z_1'(\alpha^{(k)})$ by (5.24c), where

$\alpha^{(k)} = \dfrac{(\alpha^l + \alpha^r)}{2}$. Various possibilities then need to be considered.

(1). If $Z_1'(\alpha^{(k)}) > 0$, it is not necessary to search for the good step length over the right half interval, i.e. $[\ \alpha^{(k)}\ ,\ \alpha^r\ ]$ 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 $Z_1'(\alpha^{(k)}) < 0$, we evaluate the values of $Z_1(\alpha)$ when $\alpha = \alpha^l$ and $\alpha = \alpha^{(k)}$, where the signal settings is $\psi = \psi^0 + \alpha\ \boldsymbol{h}$ and the corresponding equilibrium flows are in the linear approximation form as given by <*Corollary 5.4*>, i.e.

$\boldsymbol{q} = \boldsymbol{q}^0 + \boldsymbol{h}\ \nabla_\psi\ \boldsymbol{q}^\star(\psi^0)^T$ and the following two cases are discussed.

(i).If $Z_1(\alpha^{(k)}) \geq Z_1(\alpha^l)$, then since for all sufficiently small $\delta > 0$ in the neighbourhood of α^l , we have $Z_1(\delta) < Z_1(\alpha^l)$ 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 $Z_1(\alpha^{(k)}) < Z_1(\alpha^l)$, then in the right half interval we may find a good step length that can keep decreasing the value of Z_1 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 $Z_1'(\alpha^{(k)}) = 0$, we evaluate the values of $Z_1(\alpha)$ when $\alpha = \alpha^l$ and $\alpha = \alpha^{(k)}$, where the signal settings is $\psi = \psi^0 + \alpha\ \boldsymbol{h}$ and the corresponding equilibrium flows are in the linear approximation form as given by <*Corollary 5.4*>, i.e.

$\boldsymbol{q} = \boldsymbol{q}^0 + \boldsymbol{h}\ \nabla_\psi\ \boldsymbol{q}^\star(\psi^0)^T$ and the following two cases are discussed.

(i).If $Z_1(\alpha^{(k)}) \geq Z_1(\alpha^l)$, then since for all sufficiently small $\delta > 0$ in the neighbourhood of α^l , we have $Z_1(\delta) < Z_1(\alpha^l)$ 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 $Z_1(\alpha^{(k)}) < Z_1(\alpha^l)$, we choose the left half interval, i.e. $[\alpha^l , \alpha^{(k)}]$ as our new search interval because we are looking for the local optimum in the neighbourhood of current signal settings; let $k = k+1$ and return to Step 1.

2. Uniform search

A uniform search for finding a near-minimum value of $Z_1(\alpha)$ over the search interval $[0 , \alpha_{max}]$ can be decided in the following steps.

Step 0. Set the maximal number of iterations as N_{max} , for which the objective function $Z_1(\alpha)$ is evaluated, where the new signal settings is $\psi = \psi^0 + \alpha \boldsymbol{h}$ and the corresponding equilibrium flows are decided by reassignment process via $\boldsymbol{q} = \boldsymbol{q}^\star(\psi)$.

Step 1. Decide the evaluated interval γ , $\gamma = \dfrac{\alpha_{max}}{N_{max}}$ 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 $k = 0 , 1 , 2 ,..., N_{max} - 1$ and choose the minimum of $Z_1(\alpha)$ as the $Z_1(\alpha_{opt})$ where α_{opt} is the approximately optimal step length for $Z_1(\alpha)$ along the feasible direction $\boldsymbol{h}$ at ψ^0 .

Step 3. A new point can be decided below as the feasible point for next iteration.

$\psi_{new} = \psi^0 + \alpha_{opt} \boldsymbol{h}$ and the corresponding equilibrium flow $\boldsymbol{q}_{new} = q^\star(\psi_{new})$ can be

decided by reassignment process via (3.40).

As we have discussed in the preceding chapters, only local optima for the constrained optimisation problem (5.2) can be found by the procedures of Sections 5.2 and 5.3, due to the non-convexity of the objective function. In the following section, a practical way to find progressively better local optima for the constrained optimisation problem (5.2) can be performed by using a global search for the step length of the offset variables when the common cycle time and durations of greens are specified, which corresponds to solving an unconstrained optimisation problem because there is no feasible constraint for the offset variables in the signal settings. Furthermore, a mixed search procedure is proposed, in which three type of steps are taken into account: the first and most general type of step is to change the common cycle time and the start and duration of green time for each signal group at each junction, which can be expressed as *problem* S_1 , and the second type of step is to change the start and duration of green time for each signal group at each junction for a given common cycle time, which can be expressed as *problem* S_2 ; both the problems S_1 and S_2 are formulated as constrained optimisation problems and can be solved by the techniques given in Sections 5.2 and 5.3. The third type of step is to change the starts of green times for all signal groups at each junction by the same amount specific to that junction for a given common cycle time and given durations of green times of all signal groups, which can be expressed as *problem* S_3 . The problem S_3 uses a global search for the step length of offset variables which is an unconstrained optimisation problem and the search strategy will be discussed accordingly in the following section. Furthermore, the purpose for the first type of the step is to allow each element of the signal settings to vary within the feasible region and then the local solution for the problem (5.2) can be found; similarly, the purpose for the second type of the step is to enable the starts and durations of green times at various junctions to be improved while the common cycle time for the whole road network is specified so that the need for the lengthy evaluations of $\frac{\partial P_1}{\partial \zeta}$ is avoided. Additionally, the purpose of the third type of step, which is particularly proposed in this section, is to help us to identify another feasible region containing better local optimum than the one in the neighbourhood of current point which has been found

by solving the first and second type of steps. The third type of step is constructed by making at each of the junctions simultaneous and equal changes in the starts of greens of all signal groups at that junction, which corresponds to making a change in the offset at that junction. A global search for all step lengths along the descent direction for the *problem* S_3 with respect to the offset variables is carried out because there is no practical constraint on the offset variables.

5.4 Mixed Search Procedure

In this section, a mixed search procedure is proposed to perform the implementation of the solution to problem (5.2). Firstly, three type of steps which can be formulated as *Problem* S_1 , *Problem* S_2 and *Problem* S_3 , are discussed accordingly in section 5.4.1. Because of the non-convexity of the objective function in (5.2), only local solutions can be found. Secondly, we propose the mixed search procedure to solve the problem (5.2) practically in section 5.4.2.

5.4.1 Three type of problems

The three type of steps in solving the bi-level problem can be formulated as follows.

1. Problem S_1

$$\begin{aligned} &\underset{\alpha}{Minimise} \quad P = P_1(\psi) = Z_1(\alpha) \\ &subject\ to \quad \boldsymbol{A}\ \psi^T \le \boldsymbol{B} \\ &and \quad \psi = \psi^0 + \alpha\ \boldsymbol{h} \end{aligned} \tag{5.25}$$

where ψ^0 is the current solution and $\boldsymbol{h}$ is a change in ψ which is a particular descent direction decided by the gradient projection method.

Thus *Problem* S_1 is a particular highly constrained case of the optimisation problem (5.2) and therefore notation used in *Problem* S_1 is the same as given in expression (5.2).

For given signal timings

$\boldsymbol{\Psi}^0 = \left[\ \zeta^0\ ,\ \boldsymbol{\theta}_1^0\ ,\ \boldsymbol{\phi}_1^0\ ,\ \boldsymbol{\theta}_2^0\ ,\ \boldsymbol{\phi}_2^0\ ,\ \ldots\ ,\ \boldsymbol{\theta}_m^0\ ,\ \boldsymbol{\phi}_m^0\ ,\ \ldots\ ,\ \boldsymbol{\theta}_{N_J}^0\ ,\ \boldsymbol{\phi}_{N_J}^0\ \right]$, the value of $\boldsymbol{h}$ is the

corresponding descent direction and is decided by (5.18) or (5.19) according to *Corollaries 5.10 and 5.11* and therefore the good step length α_{good} is subject to the maximum step length $\alpha_{\max}$ that can be decided by (5.22) or (5.23) according to <*Theorem 5.14*>. The problem of finding α_{good} is a one-dimensional search that can be written as follows.

$$\underset{0 \le \alpha \le \alpha_{\max}}{Minimise} \qquad P = Z_1(\alpha) \tag{5.25a}$$

Because Z_1 is not in general convex over $[\, 0 \,,\, \alpha_{\max} \,]$, α_{good} can be found only by a globally search, but a locally optimal α such that $Z_1(\alpha) < Z_1(0)$ can be found by the bisection method which is given in section 5.3.2.

2. *Problem* S_2

Let $\boldsymbol{\psi} = [\, \boldsymbol{\theta}_1 \,,\, \boldsymbol{\phi}_1 \,,\, \boldsymbol{\theta}_2 \,,\, \boldsymbol{\phi}_2 \,,\, \ldots \,,\, \boldsymbol{\theta}_m \,,\, \boldsymbol{\phi}_m \,,\, \ldots \,,\, \boldsymbol{\theta}_{N_J} \,,\, \boldsymbol{\phi}_{N_J} \,]$ be the row vector of dimension $K - 1$ and of signal control variables other than ζ , and $\boldsymbol{B}' = \begin{bmatrix} \boldsymbol{b}_1' \\ \vdots \\ \boldsymbol{b}_m' \\ \vdots \\ \boldsymbol{b}_{N_J}' \end{bmatrix}$ be a column vector of $\boldsymbol{b}_m'$, $1 \le m \le N_J$, where $\boldsymbol{b}_m'$ represents the constant vector for the constraints (3.4)-(3.6) at junction m when the common cycle time for the whole road network is specified and the appropriate multiples of its reciprocal are included in its constants.

For given signal timing $\boldsymbol{\psi}^0 = (\, \zeta^0 \,,\, \boldsymbol{\psi}^0 \,)$ where

$\boldsymbol{\psi}^0 = [\, \boldsymbol{\theta}_1^0 \,,\, \boldsymbol{\phi}_1^0 \,,\, \boldsymbol{\theta}_2^0 \,,\, \boldsymbol{\phi}_2^0 \,,\, \ldots \,,\, \boldsymbol{\theta}_m^0 \,,\, \boldsymbol{\phi}_m^0 \,,\, \ldots \,,\, \boldsymbol{\theta}_{N_J}^0 \,,\, \boldsymbol{\phi}_{N_J}^0 \,]$, the problem is to

$$\underset{\alpha}{Minimise} \quad P = P_1(\zeta^0 , \hat{\psi}) = Z_2(\alpha)$$

$$subject \ to \quad A' \hat{\psi}^T \leq B' \tag{5.26}$$

$$and \quad \hat{\psi} = \hat{\psi}^0 + \alpha \, h$$

where ψ^0 is the current solution and $\boldsymbol{\psi}^0 = (\zeta^0 , \boldsymbol{\psi}^0)$, and h is a change in $\hat{\psi}$ which is a particular direction and determined by the gradient projection method. The descent direction for *Problem* S_2 at ψ^0 can be decided similarly by (5.18) or (5.19) according to *Corollaries 5.10 and 5.11* and the revised forms are as follows.

(1). For $g \, \nabla_{\hat{\psi}} P_1(\hat{\psi}^0)^T \neq \mathbf{0}$: the required feasible descent direction is

$$- g \, \nabla_{\hat{\psi}} P_1(\hat{\psi}^0)^T \tag{5.18a}$$

where $g = I - M^T (MM^T)^{-1} M$ and $M = A'_b$

(2). For $g \, \nabla_{\hat{\psi}} P_1(\hat{\psi}^0)^T = \mathbf{0}$: the required feasible descent direction is

$$- \hat{g} \, \nabla_{\hat{\psi}} P_1(\hat{\psi}^0)^T \tag{5.19a}$$

where $\hat{g} = I - \hat{M}^T (\hat{M}\hat{M}^T)^{-1} \hat{M}$ and $\hat{M} = \bar{A}'_b$

and $\bar{A}'_b$ is obtained from A'_b by deleting the rows of A'_b corresponding to the negative Lagrange multipliers at $\hat{\psi}^0$.

As in *Problem* S_1 , a locally optimal step length for $\alpha \in [0 , \alpha_{max}]$ can be decided by the bisection method as given in Section 5.3.2.

3. *Problem* S_3

Let $\boldsymbol{\Theta} = [\boldsymbol{\Theta}_m ; 1 \leq m \leq N_J]$ be a row vector of signal control variables of dimension

$\frac{K-1}{2}$, in which $\boldsymbol{\Theta}_m = h_m [\ 1, 1, ..., 1\ ]$ is a row vector of N_{pm} equal components specific to junction m , $1 \le m \le N_J$. For given signal timings

$\boldsymbol{\psi}^0 = \left[\ \zeta^0 , \boldsymbol{\theta}_1^0 , \boldsymbol{\phi}_1^0 , \boldsymbol{\theta}_2^0 , \boldsymbol{\phi}_2^0 , ... , \boldsymbol{\theta}_m^0 , \boldsymbol{\phi}_m^0 , ... , \boldsymbol{\theta}_{N_J}^0 , \boldsymbol{\phi}_{N_J}^0 \right]$, the problem is to

$$\underset{\alpha}{Minimise} \quad P = P_1(\zeta^0 , \theta_1^0 + \alpha\ \Theta_1 , \phi_1^0 , \theta_2^0 + \alpha\ \Theta_2 , \phi_2^0 , ... ,$$

$$\theta_m^0 + \alpha\ \Theta_m , \phi_m^0 , ... , \theta_{N_J}^0 + \alpha\ \Theta_{N_J} , \phi_{N_J}^0) \qquad (5.27)$$

$$= Z_3(\alpha)$$

where Θ is the feasible direction for *Problem* S_3 . Because in problem (5.27) there is no feasible constraint for the start of green of all signal groups at each junction, it is an unconstrained optimisation problem with respect to α for interval $[\ 0 , \alpha_{max}\ ]$, and in which the search direction Θ can be decided by the steepest descent direction and expressed as $- \nabla_\Theta P_1(\psi^0)$. (5.27a)

where $\nabla_\Theta P_1(\psi^0)$ can be obtained in the following way.

In the *Problem* S_3 , since it is an unconstrained optimisation problem, a global search for the signal timings can be conducted by a simultaneous and equal change in the starts of greens for all signal groups at any one junction m , $1 \le m \le N_J$ by setting

$$\theta_{jm} = \theta_{jm}^0 + h_m \quad , \forall j \in \boldsymbol{P}_m$$

therefore $\frac{\partial}{\partial h_m} = \sum_{j=1}^{N_{pm}} \frac{\partial}{\partial \theta_{jm}}$ and we have $\nabla_{\boldsymbol{\Theta}} P_1(\boldsymbol{\psi}^0) = \left[\frac{\partial P_1(\boldsymbol{\psi}^0)}{\partial \boldsymbol{\Theta}_m} ;\ 1 \le m \le N_J \right]$

where $\frac{\partial}{\partial \boldsymbol{\Theta}_m} = \left[\frac{\partial}{\partial h_m} , ... , \frac{\partial}{\partial h_m} \right]$ with dimension N_{pm} .

Because $P = Z_3(\alpha)$ in (5.27) is periodic with period 1 with respect to each $\alpha\ h_m$, the optimal step length for *Problem* S_3 along the resulting descent direction Θ at current signal timings ψ^0 can be decided as follows.

$$\begin{aligned} &\underset{\alpha}{Minimise} \quad P = Z_3(\alpha) \\ &subject\ \ to \quad 0 \le \alpha \le \alpha_{\max} = \underset{m}{Max} \left\{ \ | h_m |^{-1} \right\} \end{aligned} \tag{5.27b}$$

In (5.27b) a uniform search for α_{opt} is conducted throughout $[0\ ,\ \alpha_{\max}]$ so that an approximately minimal value of Z_3 is found as decided in section 5.3.2.

5.4.2 The mixed search steps

According to the gradient projection method given in *Corollaries 5.10 and 5.11*, the three type of problems will terminate only at the *KKT points* (see <*Corollary 5.13*>), i.e. within a set known to contain all locally optimal points. In problem (5.2) because of the non-convexity of the objective function, only local *KKT* solutions are expected. In this section, following the three type of problem formulated as *Problem* S_1 , S_2 and S_3 , and the corresponding solution methods described in section 5.4.1, a basic strategy including the three type of problems is summarized and the implementation heuristic is proposed below by which a better local *KKT* solution can be found for the optimisation problem (5.2).

<*Basic Strategy*>

For current signal settings, a practical way to search for a better KKT solution using three type of problems (5.25)-(5.27) are as follows.

Type I. Implement the locally optimal search by solving the first type of problem (5.25) and a KKT point can be identified.

Type II Implement the locally optimal search by solving the second type of problem (5.26) and a local KKT point can be found subject to the unchanged value ζ^0 *of* ζ .

Type III. Implement the global search in the direction determined by certain changes of offset by solving the third type of problem (5.27) starting from a local KKT point decided by Type I or Type II, to find a starting point for a search for a better local KKT point in another part of the feasible region.

<Implementation heuristic>

Following the basic strategy given in types I-III, the implementation heuristic searching for better local KKT solutions for problem (5.2) at initial feasible point ψ^0 *can be carried out in the following mixed sequence of steps.*

Step 1. Use type III to determine a good initial solution for subsequent searches.
Step 2. Use type I to determine the appropriate common cycle time for the whole network and the corresponding duration of green for each signal group.
Step 3. Use type III again to localize the good solution for subsequent searches.
Step 4. Use type II to determine the 'optimal' duration of green for each signal group in the network by which the values of the performance index is minimised.
Step 5. Use type III again to localize the 'best' signal timings by carrying out a global search for the offsets.
Step 6. Use type II again to fine-tuning the resulting signal timings until the difference of the values of the performance index between successive iterations is negligible.

5.5 Conclusions

In this Chapter the gradient projection method has been used in deciding the descent direction for the area traffic control optimisation problem in section 5.2. The determination of the step length along the descent direction has been given in Section 5.3. A mixed search procedure has been proposed in Section 5.4, in which three type of steps in search for a better local optimum have been given and also the implementation heuristic has been proposed. A global search for the optimal step length along the descent direction with respect to simultaneous and equal changes in the starts of green for all signal groups at any one junction has been particularly taken into account in this mixed search procedure. In the following Chapter two numerical examples are used to illustrate the mixed search procedure proposed in this Chapter.

Chapter 6 Numerical Calculations

6.0 Introduction

Following the solution method to the bi-level problem of the area traffic control optimisation given in Chapter 5, the numerical calculations are illustrated on the two example road networks given in this Chapter. Three kind of solution methods for the bi-level problem are applied to the example road networks: the Hooke and Jeeves' method, the mutually consistent calculations for TRANSYT-optimal signal settings and equilibrium flows, and the method proposed in this thesis and given in Chapter 5- the mixed search procedure. The remainder of this Chapter is organized as follows. In the next section, the exact solution method for the area traffic control optimisation is given by the application of the Hooke and Jeeves' method which has been reviewed in Chapter 2. The calculation of mutually consistent TRANSYT-optimal signal settings and equilibrium flows is briefly described in Section 6.2 as applies to this bi-level problem. In Section 6.3, the proposed solution method in terms of the mixed search procedure for the bi-level problem given in Chapter 5 is briefly stated as it applies to the numerical calculations. In Section 6.4, the descriptions of the two example test road networks are given together with the corresponding input data information. In Section 6.5, the calculation results performed by the solution methods to the bi-level problem on the two example networks are discussed respectively; furthermore, the effectiveness of the corresponding solution method and the comparisons of the computation efforts and the values of the indicators of the performance index are investigated as well. Conclusions for this Chapter is given in Section 6.6.

6.1 Hooke and Jeeves' Method

The application of the Hooke and Jeeves' method, which has been described in Chapter 2, to the bi-level problem of area traffic control optimisation can be described as follows. Recall the bi-level problem of the area traffic control optimisation given in (4.0):

$$\begin{aligned} &\underset{\zeta\,,\,\theta\,,\,\phi \,\in\, S_0}{\textit{Minimise}} \quad P = P_0(\zeta\,,\,\theta\,,\,\phi\,,\,q) \\ &\textit{subject to} \\ &\qquad\qquad q = q^{\star}(\zeta\,,\,\theta\,,\,\phi) \end{aligned} \tag{6.0}$$

where S_0 is the constraint set of the signal setting variables $(\zeta\,,\,\theta\,,\,\phi)$ in (3.3)-(3.6).

In (6.0) $$P_0(\zeta\ ,\ \theta\ ,\ \phi) = \sum_{a \in L} (D_a W_{aD} M_D + S_a W_{aS} M_S)$$

which has been given in (3.2) and can be evaluated by (3.14)-(3.39), and $q = q^{\star}(\zeta,\theta,\phi)$ can be formulated as a variational inequality problem which has been given in (3.40):

to find values $q^{\star}(\psi) \in \Omega(\psi)$ of $q(\psi)$ such that

$$c(q^{\star}(\psi))\ (q(\psi) - q^{\star}(\psi))^T \geq 0 \tag{6.1}$$

$$\forall\ q(\psi) \in \Omega(\psi) = \left\{ q(\psi)\ ;\ q(\psi)^T = \delta\ f(\psi)^T\ ,\ \Delta\ f(\psi)^T = \hat{D}^T\ ,\ f(\psi) \geq \mathbf{0} \right\}$$

the problem (6.1) can be solved by the convex combination method given in Section 2.2.2 and stated below.

<u>*Step 6.0*</u>: Carry out the all-or-nothing traffic assignment for each origin-destination pair with link costs corresponding to zero flow and specify the resulting link flows as $x^{(n)}$ where $n = 0$.

<u>*Step 6.1*</u>: Carry out the all-or-nothing traffic assignment again but with $c^{(n)}\ (x^{(n)})$ then specify the resulting auxiliary link flows as $y^{(n)}$.

<u>*Step 6.2*</u>: Find the one dimension optimal move size $\alpha^{(n)}$ such that

$$c^{(n)}(v^{(n)})\ (y^{(n)} - x^{(n)})^T = 0$$

where $v^{(n)} = x^{(n)} + \alpha^{(n)}\ (y^{(n)} - x^{(n)})$

and update the link flows as $x^{(n+1)} = v^{(n)}$

<u>*Step 6.3*</u>: Let u_w be the minimum travel time for the origin-destination pair w, and ε , $\varepsilon \geq 0$ be a predetermined threshold value. If the inequality is satisfied

$$\sum_{w \in W} \frac{|\ u_w^{(n)} - u_w^{(n-1)}\ |}{u_w^{(n)}} \leq \varepsilon$$

for each origin-destination w in W , then the link flows $x^{(n+1)}$ are taken as the equilibrium flows; otherwise go back to *Step 6.1* and perform the procedure again.

In (6.1) the equilibrium link travel time function c_a , $\forall\, a \in L$ has been given in (3.41)

$$c_a = c_a^0 + d_a = c_a^0 + d_a^U + d_a^{r+o} \tag{6.1a}$$

where d_a^U and d_a^{r+o} have been given $\forall\, a \in L$ in (3.22), (3.29), (3.34) and (3.39). For a given solution $\boldsymbol{\psi}^0 = (\zeta^0, \boldsymbol{\theta}^0, \boldsymbol{\phi}^0)$, the application of Hooke and Jeeves' method given in Section 2.3.3 to the bi-level problem in (6.0) can be described as follows.

<u>*Step 1. Initialization:*</u>

1.1 For signal settings $\boldsymbol{\psi} = (\zeta, \boldsymbol{\theta}, \boldsymbol{\phi})$ in $K = 2\sum_{m=1}^{N_J} N_{pm} + 1$ dimensions for problem (6.0), choose an initial solution point $\psi^{(0)}$ satisfying the constraint set S_0 in (6.0), and let $\psi = \psi^{(0)}$.

1.2 Set an initial step size α , $\alpha > 0$ for the *exploratory search* in *Step 2* and the acceleration factor β for the *pattern search* in *Step 3*; set indicator $\eta = 1$

1.3 Set indices $j = 1$, $k = 0$

<u>*Step 2. Exploratory search:*</u>

2.0 If $j = K + 1$, go to *Step 3*

2.1 Let e_j , $j = 1, \ldots, K$ be a vector with 1 in the j th component of the decision variable vector and 0 elsewhere, and set $\psi' = \psi + \alpha\, \eta\, e_j$.

Evaluate the objective function F_1 at ψ' along its positive coordinate direction with predetermined step move size α .

2.2 If $P_l(\psi') < P_l(\psi)$ put $\psi = \psi'$, $j = j + 1$, $\eta = 1$ and go to *Step 2.0*; otherwise go to *Step 2.3*.

2.3 If $\eta = 1$, put $\eta = -1$ and go to *Step 2.1*, otherwise if $\eta = -1$ put $j = j + 1$, $\eta = 1$ and go to *Step 2.0*.

<u>*Step 3. Pattern search:*</u> when the evaluation of the objective function with respect to each component of the signal settings has been conducted in the *exploratory search*, we need to compare P_l at ψ' *and* $\psi^{(k)}$

3.1 If $P_l(\psi') < P_l(\psi^{(k)})$, put $\psi^{(k+1)} = \psi'$ and $\psi = \psi^{(k)} + \beta(\psi^{(k+1)} - \psi^{(k)})$ and $j = 1$ then go to *Step 2*, otherwise go to *Step 3.2*.

3.2 If the step size α is sufficiently small within the predetermined threshold, then the process is complete with $\psi^{(k)}$ as the resulting estimate of the optimal signal settings; otherwise put $\alpha = 0.5\ \alpha$, $j = 1$ *and* $\psi = \psi^{(k)}$ and return to *Step 2*.

6.2 Mutually Consistent Calculations

The mutually consistent calculations of signal timings and equilibrium flows for the bi-level problem (6.0) can be performed in the following steps.

<u>*Step 0*</u>. Set $k = 0$ for given signal timings $\psi^{(0)}$, find the corresponding equilibrium flows $q^{\star}(\psi^{(0)})$ by means of the convex combination method given in *Steps 6.0- 6.3*.

<u>*Step 1*</u>: Run TRANSYT programme to obtain the TRANSYT- optimal signal timings $\psi^{(k+1)}$ for the flows $q^{\star}(\psi^{(k)})$.

<u>*Step 2*</u>. Update the travel time function (6.1a) of all links to obtain

$$c(q\ ,\ \psi^{(k+1)}) = c^0 + d(q\ ,\ \psi^{(k+1)})$$

where $d(q\ ,\ \psi^{(k+1)})$ can be estimated by (3.22), (3.29), (3.34) and (3.39).

Step 3. Calculate the corresponding equilibrium flows $q^{\star}(\psi^{(k+1)})$ by means of the convex combination method given in *Steps 6.0- 6.3*.

Step 4. Run TRANSYT programme again to obtain the optimal signal timings $\psi^{(k+2)}$ given by the equilibrium flows $q^{\star}(\psi^{(k+1)})$.

Step 5. Compare the values of $\psi^{(k+2)}$ *and* $\psi^{(k+1)}$, if there is no change between $\psi^{(k+1)}$ *and* $\psi^{(k+2)}$ then go to *Step 6*; otherwise, increase k by 1 and go to *Step 2*.

Step 6. Stop: $\psi^{(k+1)}$ *and* $q^{\star}(\psi^{(k+1)})$ are the mutually consistent signal timings and equilibrium flows.

6.3 Mixed Search Procedure

The proposed solution method for the bi-level problem in (6.0) has been given in Chapter 5. For any one given solution, the proposed solution method can be described as a mixed search procedure in terms of the following three type of steps:

Type I. Search for the optimal signal setting variables with respect to ζ , θ , ϕ over the whole feasible region 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 one dimensional bisection method, both of which have been discussed in problem (5.25).

Type II. Similarly as in type I but search for the optimal 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 one dimensional bisection method, both of which have been discussed in problem (5.26).

Type III. Search for the optimal signal setting variables with respect to 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 optimisation problem (5.27) 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, both of which have been discussed in problem (5.27).

The purpose of the mixed search procedure for the bi-level problem given in (6.0) 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 effort 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 not only will less computation effort be incurred but also the search in this restricted set of directions will be extended over the whole feasible region. Adopting an equal and simultaneous change in the starts of green times for all signal groups at any one junction can help us to reach a more favourable point of another part of the feasible region along the restricted set of directions because no practical constraint applies with respect to the offset variables. In summary, the rationale for this mixed search procedure is to alternately apply the three type of search steps to finding the near-optimal in many parts of the feasible region and locating them within the corresponding neighbourhoods.

Two variants of the mixed search procedure obtained by applying various sequences of the three types of step are stated as follows. Firstly *method a* starts from any initial

solution with negative values of the total derivatives for the performance index with respect to the move size α 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 bi-level formulation is found and located. Secondly, *method b* adopts the optimal search technique used in TRANSYT programme which is carried out by means of the mixed combination of the small and large steps for the starts of green times. The *method b* starts with the search step of type III by simultaneous and equal changes of the starts of greens at any one junction and therefore in searching many parts of the feasible region 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. Stopping criterion for performing the mixed search will be met if a predetermined threshold is satisfied between successive iterations.

6.4 Test Road Networks

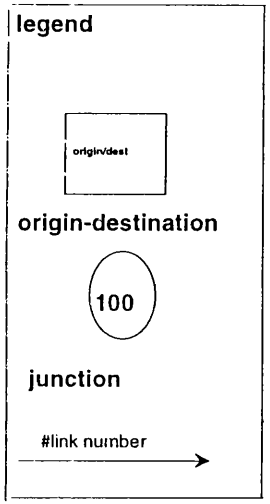
In this Section, two test road networks are introduced for use in testing the effectiveness of various solution methods to the bi-level problem (6.0) discussed above. Each test road network includes the following three items: representation of network layouts for use in the signal settings and traffic assignment, feasible constrained set of signal control variables and given constants, and the test methods. For the first test network, all three kind of solution methods are illustrated. For the second test network, the two last solution methods are illustrated: the mutually consistent calculations and the variants of mixed search procedure. For each test network, the measure of the performance index is in terms of veh-h/h where we assume the monetary value of veh-h is 495 pence and that of 100 veh-stops is 85 pence and therefore 582.4 veh-stops are approximately equivalent to one veh-h by comparing the monetary value of 495 pence for one veh-h and that of 0.85 pence for one veh-stop; furthermore, the effectiveness of the corresponding solution methods to the bi-level problem (6.0) is discussed and comparison of the computation burden is also made.

6.4.1 Two-junction Network

The first test network is a network containing two signal-controlled junctions, which has been modified from Braess's road network (Murchland, 1970; Pas and Principio 1997). The two-junction network consists of one origin and one destination, two signal-controlled junctions and eight TRANSYT links. Link travel times are decided by the sum of undelayed travel time along this link and the average delay incurred by traffic at the downstream junction. For each signal controlled junction, the travel times on entering links will be affected by the changes of the corresponding signal timings, and for non-signal controlled links the travel times are constant throughout this computation process and for those links have greens over the whole cycle the travel times are constant through this computation. Detailed layouts for the two-junction network for use in signal settings and in the traffic assignment are respectively given in Figures 6.1a, and 6.1c, together with the basic information about the signal control variables given in Figure 6.1b. Using typical values found in practice, the minimum green times for each group is 7 seconds, and the clearance times between incompatible signal groups are 5 seconds. The minimum and maximum common cycle times are specified as 24 and 90 seconds, and the travel demand flows are specified as a rather low value as compared to the values of saturation flows at each signal-controlled link because in this case we simply want to know the effectiveness of the three kind of solution methods when they are implemented in a lightly loaded traffic network. The corresponding information of input data is also given in Tables 6.1a and 6.1b.

6.4.2 Allsop and Charlesworth's Network

Another example network is illustrated based upon the one used by Allsop and Charlesworth (1977). Basic layouts of the network for use in signal settings and in traffic assignment are respectively given in Figures 6.2a and 6.2 c, where Figure 6.2c is adapted from Charlesworth (1977, p554) and the correspondence with links used in Figure 6.2a is set out in Table 6.2a, where the signal-delayed links are particularly mentioned in which the average delay is incurred by downstream signal-controlled junction, those links joining nodes in Figure 6.2c are numbered by xy, where x is the label of the upstream node and y the label of the downstream node. The allocations for signal groups at each junction together with the basic information about signal control variables are given in Figure 6.2b. Travel demands for each pair of origin and



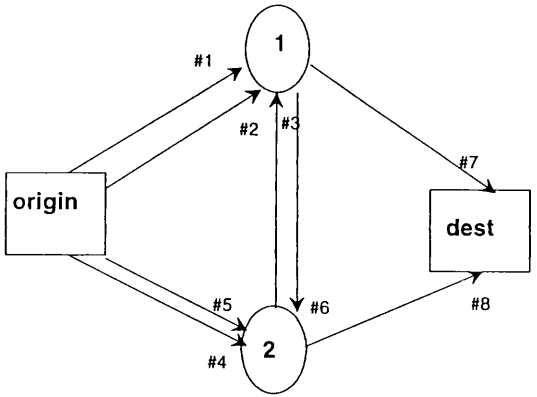


Figure 6.1a TRANSYT links for two-junction network

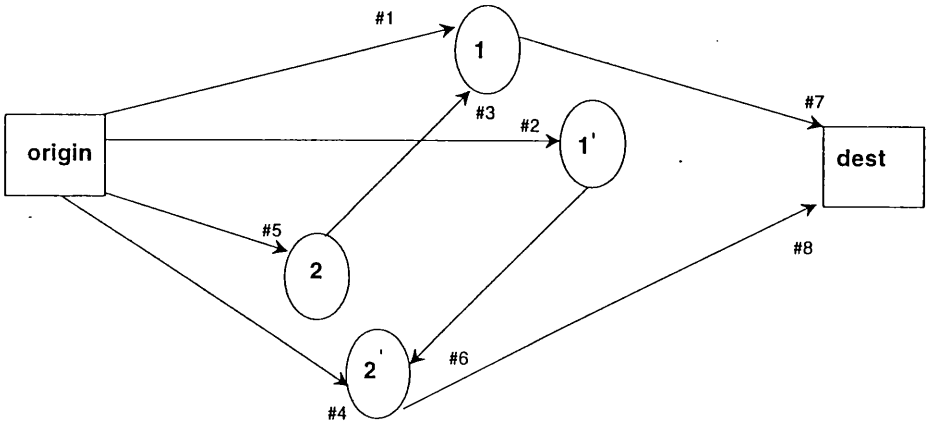


Figure 6.1c Layout of two-junction network represented for use of traffic assignment

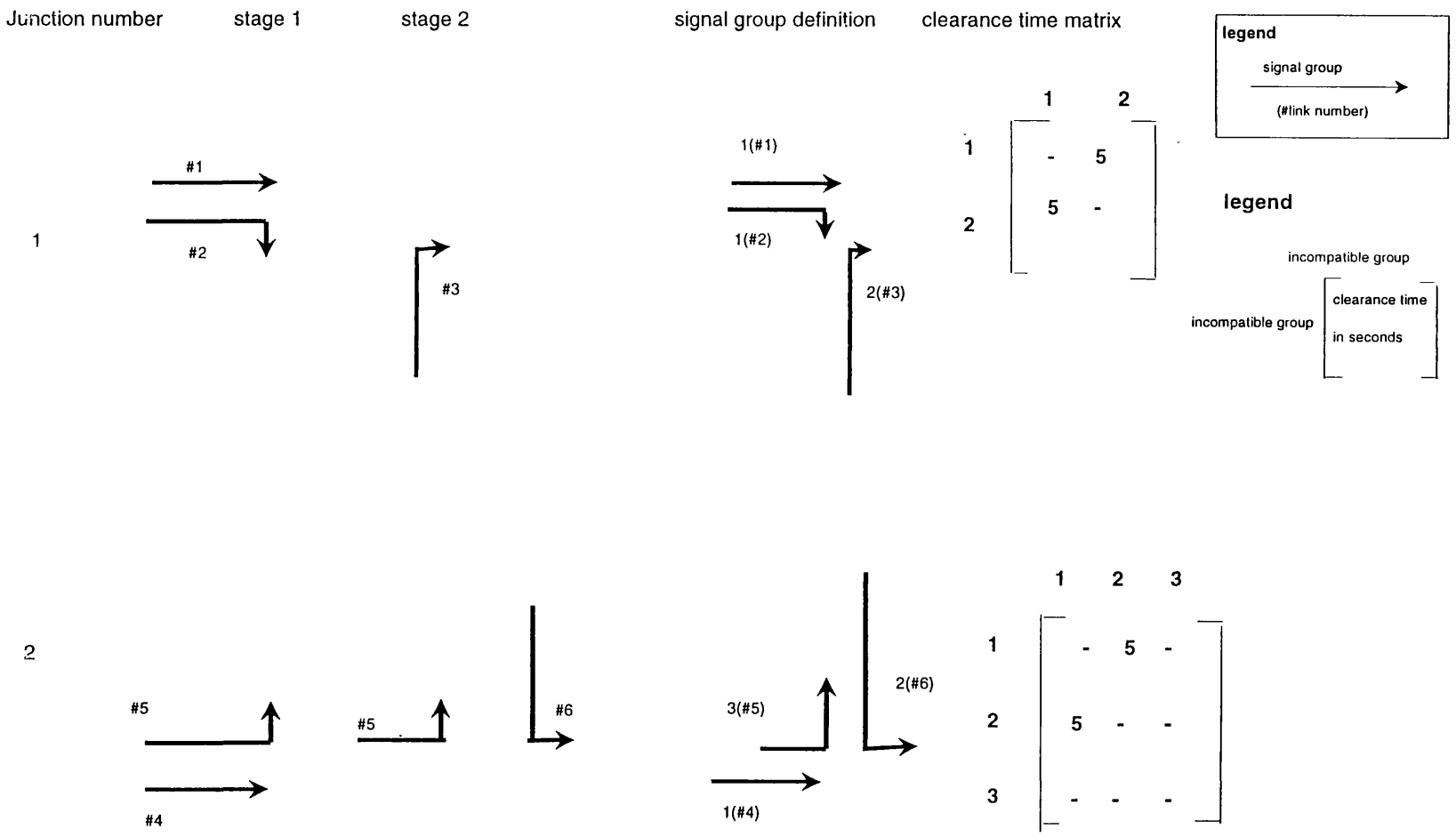


Figure 6.1b Signal groups and clearance time matrices for two-junction network

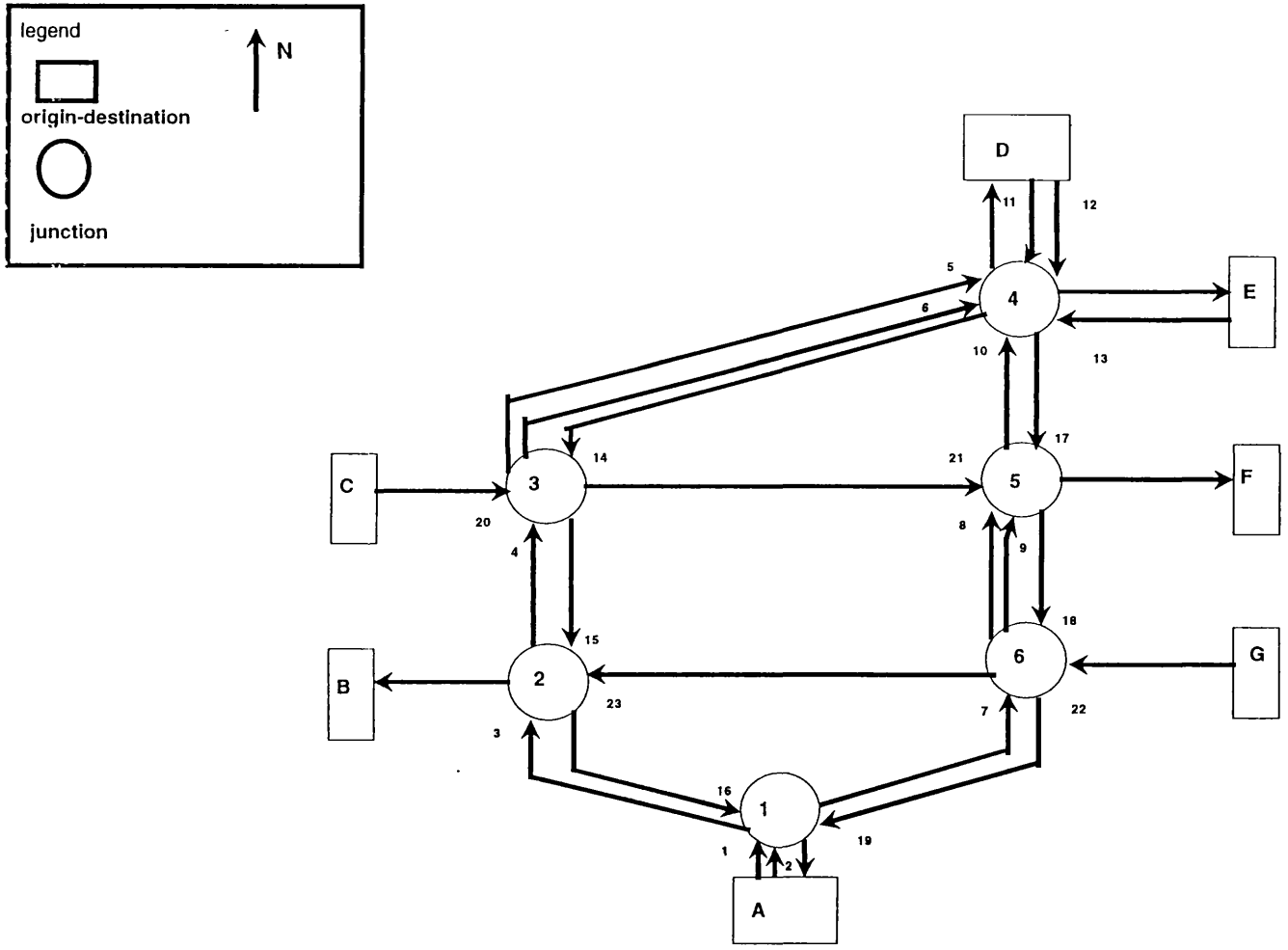


Figure 6.2a Layout for Allsop & Charlesworth's network

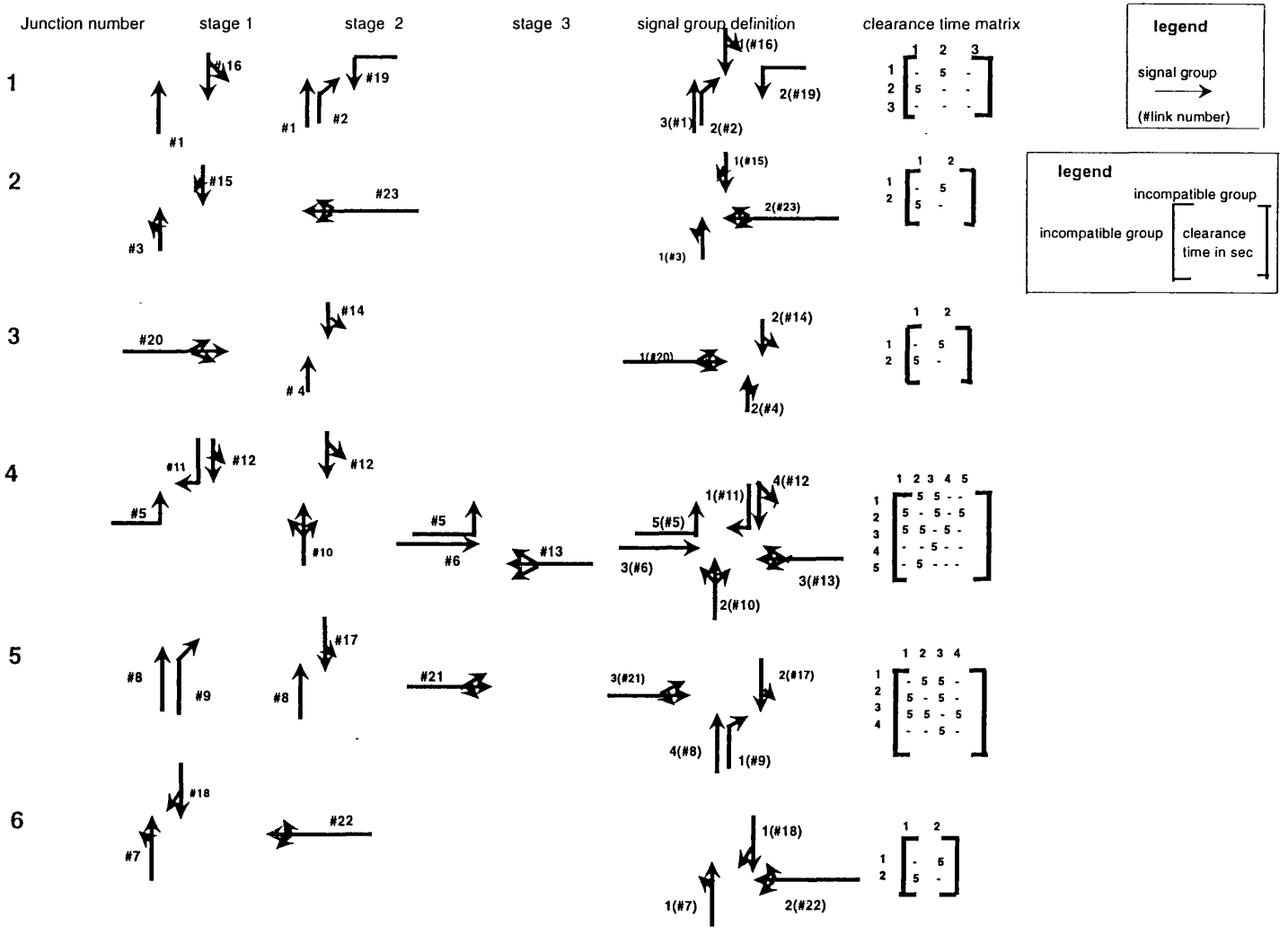


Figure 6.2b Configuration for Allsop & Charlesworth's (1977) road network

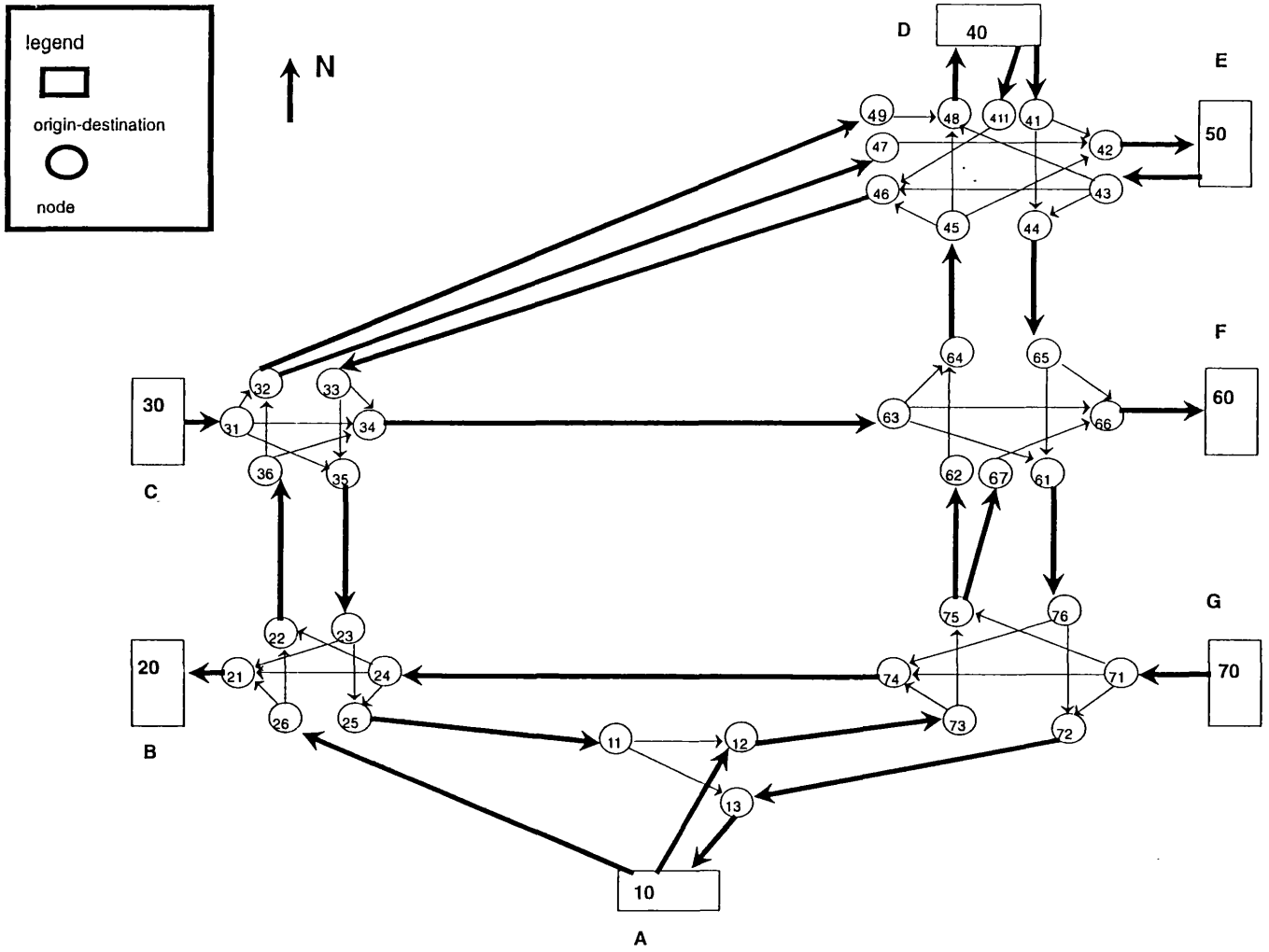


Figure 6.2c Representation for traffic assignment use in nodes and links for Allsop&Charlesworth's network

Table 6.1a Input data for two-junction road network

Junction	Link	c^0	s	Junction	Link	c^0	s
1	1	20	1800	2	4	20	1800
	2	20	1800		5	20	1800
	3	20	1800		6	20	1800

where c^0 , s are respectively the undelayed travel time in seconds and saturation flow in vehicles/hour, the non-signal controlled links 7,8 are given travel time of 10 seconds, the OD demand is set as 1500 veh/h.

Table 6.2a Relationship between links used in figure 6.2a and links used in figure 6.2c

link number in fig 6.2a	signal-delayed links the counterpart in fig 6.2c	zero travel time links in fig 6.2c
1(*)		
2	1012	
3	1026	2621,2622
4	2236	3632,3636
5	3249	4948
6	3247	4742
7	1273	7374,7375
8	7562	6264
9	7567	6766
10	6445	4542,4546,4548
11	40411	41146
12	4041	4142,4144
13	5043	4344,4346,4348
14	4633	3334,3335
15	3523	2321,2325
16	2511	1112,1113
17	4465	6561,6566
18	6176	7672,7674
19	7213	7213
20	3031	3132,3134,3135
21	3463	6361,6364,6366
22	7071	7172,7174,7175
23	7424	2421,2422,2425

link number 1 has green over the whole cycle therefore no traffic delay incurred we give a constant travel time throughout this computation; for correspondence links, given constant travel times are assumed throughout this computation and for signal delayed links, average travel delays are calculated by the equations given in Chapter 3.

Table 6.1b Constraints for two-junction road network

Recall constraints for signal groups in (3.3)-(3.6) where in this computation the capacity constraint (3.5) is assumed relaxed for all signal-controlled traffic streams, we have constraints (3.3), (3.4) and (3.6) as follows.

$$\zeta_{\min} \le \zeta \le \zeta_{\max}$$

$$g_{jm}\zeta \le \phi_{jm} \le 1 \ , \ \forall j \in P_m \ , \ m \in N$$

$$\theta_{jm} + \phi_{jm} + \hat{c}_{jlm}\zeta \le \theta_{lm} + \Omega_m (j \ , \ l) \ , j \ne l \ , \ \forall j \ , \ l \in P_m \ , \ m \in N$$

for constraint (3.6), two cases take into account:
(i) if signal group j precedes signal group l at junction m:

by definition $\Omega_m (j \ , \ l) = 0$ then (3.6) becomes $\phi_{jm} + \hat{c}_{jlm}\zeta \le \theta_{lm} - \theta_{jm}$ (3.6a)

(ii) if signal group l precedes signal group j at junction m:

by definition $\Omega_m (j \ , \ l) = 1$ then (3.6) becomes

$$\theta_{jm} + \phi_{jm} + \hat{c}_{jlm}\zeta \le \theta_{lm} + 1 \qquad (3.6b)$$

In this computation, the minimum green for each signal group g_{jm} is 7 sec and the clearance time between incompatible signal groups $\hat{c}_{jlm}$ is 5 sec for the constraint set $\boldsymbol{A}\,\boldsymbol{\psi}^T \le \boldsymbol{B}$. In the first signal-controlled junction we have two signal groups and each of which is incompatible with the other, and in the second signal-controlled junction we have three signal groups and only the first two are incompatible groups to each other. It follows

$$\begin{bmatrix} -1 & \mathbf{0} & \mathbf{0} \\ \boldsymbol{a}_1 & \boldsymbol{A}_1 & \mathbf{0} \\ \boldsymbol{a}_2 & \mathbf{0} & \boldsymbol{A}_2 \end{bmatrix} \begin{bmatrix} \zeta \\ \boldsymbol{\psi}_1 \\ \boldsymbol{\psi}_2 \end{bmatrix} \le \begin{bmatrix} b_0 \\ \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix} \text{ where } b_0 = \frac{-1}{90} \ ,$$

and $\boldsymbol{a}_1 = \begin{bmatrix} 5 \\ 5 \\ 7 \\ 7 \end{bmatrix}, \boldsymbol{A}_1 = \begin{bmatrix} 1 & -1 & 1 & 0 \\ -1 & 1 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}, \boldsymbol{b}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{\psi}_1 = \begin{bmatrix} \theta_{11} \\ \theta_{21} \\ \phi_{11} \\ \phi_{21} \end{bmatrix}$ and

$$\boldsymbol{a}_2 = \begin{bmatrix} 5 \\ 5 \\ 7 \\ 7 \\ 7 \end{bmatrix}, \boldsymbol{A}_2 = \begin{bmatrix} 1 & -1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}, \boldsymbol{b}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{\psi}_2 = \begin{bmatrix} \theta_{12} \\ \theta_{22} \\ \phi_{12} \\ \phi_{22} \\ \phi_{32} \end{bmatrix}$$

destination are those used by Charlesworth (1977, p556) and also given in Table 6.2b. Fixed data for undelayed travel time and saturation flow for each link entering each junction are also adopted from Charlesworth (1977, p555) and given in Tables 6.3a and 6.3b, where we assume that the travel times on the non-signal controlled exit links are zero because these values are constant throughout the computation process, and for those links have greens over the whole cycle the travel times are constant throughout this computation. The corresponding coefficient and matrices and constant vectors for the constraints on signal timings are expressed on a junction basis as shown in Table 6.3b. This numerical test includes 20 origin-destination pairs, 23 links, 40 feasible routes and 18 signal groups at 6 signal-controlled junctions.

6.5 Computation Results

In this section, we will show the computation results for the tested methods on the two test road networks discussed above. Computation results for each tested method contain the following three parts: the initial values of the signal settings and the corresponding equilibrium link flows, the intermediate computation results for signal timings and the corresponding equilibrium link flows from iteration to iteration and the final results for the performance index and the degree of saturation of each link.

6.5.1 Two-junction Network

In the first network, all the three kind of solution methods to the bi-level problem (6.0) were illustrated with the same arbitrary initial signal timings for the two connected signal-controlled junctions. Comprehensive results are shown in Tables 6.4-6.7 respectively for each kind of solution method from the initial values to the final results from iteration to iteration.

Firstly, as it is seen in Table 6.4, the results from Hooke and Jeeves' method showed the value of the performance index reduced to 3.08 veh-h/h from the initial value of 7.14 veh-h/h with the resulting cycle time 60 seconds after 3 full iterations of alternate exploratory and pattern searches which are summarized in Section 6.1, which improved the system performance by 57%. The corresponding signal timings and the equilibrium flows at each iteration are also shown in Table 6.4.

Table 6.2b Travel demand for Allsop and Charlesworth's road network in vehicles/hour

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

* where the travel demands between OD pair D & E are not included in this numerical test which can be allocated directly via links 12 & 13.

Table 6.3a Fixed data for Allsop and Charlesworth's road network

Junction	Link	c^0	s	Junction	Link	c^0	s
1	1	0	2000	4	5	20	1800
	2	0	1600		6	20	1850
	16	10	2900		10	10	2200
	19	10	1500		11	0	2000
					12	0	1800
					13	0	2200
2	3	10	3200	5	8	15	1850
	15	15	2600		9	15	1700
	23	15	3200		17	10	1700
					21	15	3200
3	4	15	3200	6	7	10	1800
	14	20	3200		18	15	1700
	20	0	2800		22	0	3600

where c^0, s are respectively the undelayed travel time in seconds and saturation flow in vehicles/hour, and the non-signal controlled links entry the network are given travel time of zero.

Table 6.3b Constraints for Allsop and Charlesworth's road network

Following similarly procedure given in Table 6.1b, in relation to the constraint set for $\boldsymbol{A}\ \boldsymbol{\psi}^T \leq \boldsymbol{B}$, it follows

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_1 & A_1 & 0 & 0 & 0 & 0 & 0 \\ a_2 & 0 & A_2 & 0 & 0 & 0 & 0 \\ a_3 & 0 & 0 & A_3 & 0 & 0 & 0 \\ a_4 & 0 & 0 & 0 & A_4 & 0 & 0 \\ a_5 & 0 & 0 & 0 & 0 & A_5 & 0 \\ a_6 & 0 & 0 & 0 & 0 & 0 & A_6 \end{bmatrix} \begin{bmatrix} \zeta \\ \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \\ \psi_5 \\ \psi_6 \end{bmatrix} \leq \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix} \quad \text{where } b_0 = \frac{-1}{130} ,$$

and for junction 1 $a_1 = \begin{bmatrix} 5 \\ 5 \\ 7 \\ 7 \\ 7 \end{bmatrix}$, $A_1 = \begin{bmatrix} 1 & -1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$, $b_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, $\psi_1 = \begin{bmatrix} \theta_{11} \\ \theta_{21} \\ \phi_{11} \\ \phi_{21} \\ \phi_{31} \end{bmatrix}$

for junctions 2 and 3

$$a_2 = \begin{bmatrix} 5 \\ 5 \\ 7 \\ 7 \end{bmatrix}, A_2 = \begin{bmatrix} 1 & -1 & 1 & 0 \\ -1 & 1 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}, b_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \psi_2 = \begin{bmatrix} \theta_{12} \\ \theta_{22} \\ \phi_{12} \\ \phi_{22} \end{bmatrix}$$

$$a_3 = \begin{bmatrix} 5 \\ 5 \\ 7 \\ 7 \end{bmatrix}, A_3 = \begin{bmatrix} 1 & -1 & 1 & 0 \\ -1 & 1 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}, b_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \psi_3 = \begin{bmatrix} \theta_{13} \\ \theta_{23} \\ \phi_{13} \\ \phi_{23} \end{bmatrix}$$

(cont)

for junction 4 $a_4 = \begin{bmatrix} 5 \\ 5 \\ 5 \\ 5 \\ 5 \\ 5 \\ 5 \\ 5 \\ 5 \\ 5 \\ 7 \\ 7 \\ 7 \\ 7 \\ 7 \end{bmatrix}$, $A_4 = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$, $b_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, $\psi_4 = \begin{bmatrix} \theta_{14} \\ \theta_{24} \\ \theta_{34} \\ \theta_{44} \\ \theta_{54} \\ \phi_{14} \\ \phi_{24} \\ \phi_{34} \\ \phi_{44} \\ \phi_{54} \end{bmatrix}$

for junction 5 $a_5 = \begin{bmatrix} 5 \\ 5 \\ 5 \\ 5 \\ 5 \\ 5 \\ 5 \\ 5 \\ 7 \\ 7 \\ 7 \\ 7 \end{bmatrix}$, $A_5 = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$, $b_5 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, $\psi_5 = \begin{bmatrix} \theta_{15} \\ \theta_{25} \\ \theta_{35} \\ \theta_{45} \\ \phi_{15} \\ \phi_{25} \\ \phi_{35} \\ \phi_{45} \end{bmatrix}$

and for junction 6 $a_6 = \begin{bmatrix} 5 \\ 5 \\ 7 \\ 7 \end{bmatrix}$, $A_6 = \begin{bmatrix} 1 & -1 & 1 & 0 \\ -1 & 1 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$, $b_6 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, $\psi_6 = \begin{bmatrix} \theta_{16} \\ \theta_{26} \\ \phi_{16} \\ \phi_{26} \end{bmatrix}$; where the maximum common cycle time is 130 sec, the minimum green is 7 sec.

Secondly, as it is seen in Table 6.5, results from mutually consistent calculations for TRANSYT-optimal signal timings and the corresponding equilibrium flows showed the value of the performance index reduced to a similar value of 3.09 veh-h/h from the initial 7.14 veh-h/h with the resulting cycle time 75 seconds after alternately performing 10 full iterations in TRANSYT runs and solving traffic assignments, which approximately achieved the same performance as did the Hooke and Jeeves' method in this case. The corresponding signal timings and the equilibrium flows at each iteration are shown in Table 6.5.

Thirdly, as it is seen in Tables 6.6&6.7, results from the two variants of mixed search procedure were obtained for combinations of steps of type I-III: the *method a* and the *method b* as introduced in Section 6.3. The result from *method a*, as seen in Table 6.6, it showed the value of the performance index reduced from the initial value of 7.14 veh-h/h to 3.11 veh-h/h with the resulting cycle time 60 seconds. After carrying out the steps of type I in iterations 1-3, which improved the system performance by 48.04%, carrying out the step of type III in iteration 4 decreased the value of the performance index by a further 1.89%; which means a better starting point for further searching had been found in another part of the feasible region. Then carrying out the step of type I in iteration 5 again achieved the final value of the performance index for this computation of 3.11 veh-h/h and total improvement of the system performance achieved by *method a* was 56.44%. The total number of iterations in performing *method a* of the mixed search procedure in this case is 5 full iterations. Also, as it is seen in Table 6.7, for the result from *method b*, the value of the performance index was reduced to 3.12 veh-h/h from the initial value of 7.14 veh-h/h with same resulting cycle time 60 seconds after carrying out the steps of type III in iterations 1, 3 and 5, and alternately followed by carrying out type I or II in iterations 2, 4 and 6. The total improvement of the system performance achieved by *method b* is 56.30%, which is almost the same as that for *method a,* and the total number of iterations in performing the *method b* of mixed search procedure in this case is 6 full iterations.

As it is seen in Tables 6.4-6.7, the three solution methods implemented in the two-junction network improve the values of system performance to effectively the same extent, which provides a background to the subsequent numerical analysis using the

mutually consistent calculations and mixed search procedure when a medium-sized road network is considered.

Table 6.8 summarizes the computation efforts for the three solution methods in terms of the number of times of solving the equilibrium traffic assignment problems and the corresponding convex combination method. The number of times of solving the equilibrium traffic assignment in the full computations of Hooke and Jeeves' method is much greater than those in the mutually consistent calculations and in the mixed search method. For each iteration in calculating the total derivatives of the performance index with respect to the signal setting variables in the mixed search procedure, the computation efforts have been executed in the C++ programming integrated development environment and performed on PC 486SX 25/33 Zenith machine: each iteration for this numerical example was carried out in less than 20 seconds of CPU time and total computation efforts for the complete run of the mixed search procedure did not exceed 10 minutes of CPU time on that machine.

6.5.2 Allsop and Charlesworth's Network

We continue to illustrate the effectiveness of the solution methods discussed in Sections 6.2-6.3 on a medium-sized road network which has been used by Allsop and Charlesworth (1977). In this numerical test, as it has been introduced in Section 6.4.2, we have 18 signal groups for 6 signal-controlled junctions and 23 TRANSYT links and 40 feasible paths for 20 OD pairs, where the total number of decision variables is 37. As it has been reviewed in Chapter 2, the major disadvantage of Hooke and Jeeves' method lies in its intensive computation efforts, and also as is shown in Table 6.8 a great number of solving the traffic assignment problem has arisen when this method applied even to the two-junction network. Therefore in this numerical test we perform only the other two solution methods - the mixed search variants and the mutually consistent calculations on the Allsop & Charlesworth's network.

Because the non-convexity of the bi-level problem (6.0), only a local optimum is expected to be obtained in this numerical test. For the reasons of understanding the effectiveness and robustness of the proposed solution method in this test, two distinct sets of initial signal timings are specified. The first set of initial signal timings are

iteration number	cycle time ($1/\xi$) sec	PI veh-h/h	start of green in seconds junction 1 θ_{11}/ξ	θ_{21}/ξ	junction 2 θ_{12}/ξ	θ_{22}/ξ	θ_{32}/ξ	duration of green in seconds junction 1 ϕ_{11}/ξ	ϕ_{21}/ξ	junction 2 ϕ_{12}/ξ	ϕ_{22}/ξ	ϕ_{32}/ξ
1	60.0	7.14	0.0	40.0	0.0	46.0	0.0	35.0	15.0	41.0	9.0	60.0
2	60.0	3.56	3.0	51.0	0.0	46.0	3.0	43.0	7.0	41.0	9.0	60.0
3	60.0	3.08	12.0	60.0	1.0	49.0	12.0	43.0	7.0	43.0	7.0	60.0

iteration number	link flow in veh/h q_1	q_2	q_3	q_4	q_5	q_6
1	615.0	45.0	225.0	615.0	225.0	45.0
2	969.5	1.0	5.0	524.5	5.0	1.0
3	750.0	0.0	0.0	750.0	0.0	0.0
	x_1	x_2	x_3	x_4	x_5	x_6
final value of degree of saturation	0.57	0.00	0.00	0.57	0.00	0.00

Table 6.5 Results on two-junction network for mutually consistent calculations

iteration number	cycle time ($1/\xi$) sec	PI veh-h/h	start of green in seconds junction 1 θ_{11}/ξ	θ_{21}/ξ	junction 2 θ_{12}/ξ	θ_{22}/ξ	θ_{32}/ξ	duration of green in seconds junction 1 ϕ_{11}/ξ	ϕ_{21}/ξ	junction 2 ϕ_{12}/ξ	ϕ_{22}/ξ	ϕ_{32}/ξ
1	60.0	7.14	0.0	40.0	0.0	46.0	0.0	35.0	15.0	41.0	9.0	60.0
2	60.0	6.30	8.0	49.0	0.0	48.0	0.0	36.0	14.0	43.0	7.0	60.0
3	72.0	5.98	10.0	62.0	0.0	60.0	0.0	47.0	15.0	55.0	7.0	72.0
4	72.0	5.19	62.0	46.0	52.0	40.0	52.0	51.0	11.0	55.0	7.0	72.0
5	72.0	4.44	3.0	62.0	62.0	50.0	62.0	54.0	8.0	55.0	7.0	72.0
6	75.0	3.96	11.0	74.0	0.0	63.0	0.0	58.0	7.0	58.0	7.0	75.0
7	75.0	3.98	11.0	74.0	0.0	63.0	0.0	58.0	7.0	58.0	7.0	75.0
8	75.0	3.31	0.0	63.0	74.0	62.0	74.0	58.0	7.0	58.0	7.0	75.0
9	75.0	3.09	0.0	63.0	0.0	63.0	0.0	58.0	7.0	58.0	7.0	75.0
10	75.0	3.09	0.0	63.0	0.0	63.0	0.0	58.0	7.0	58.0	7.0	75.0

iteration number	link flow in veh/h q_1	q_2	q_3	q_4	q_5	q_6
2	620.0	40.0	220.0	620.0	220.0	40.0
3	635.0	25.0	205.0	635.0	205.0	25.0
4	705.0	25.0	135.0	635.0	135.0	25.0
5	705.0	25.0	85.0	685.0	85.0	25.0
6	705.0	25.0	55.0	715.0	55.0	25.0
7	715.0	15.0	55.0	715.0	55.0	25.0
8	735.0	15.0	15.0	715.0	35.0	15.0
9	750.0	0.0	0.0	750.0	0.0	0.0
10	750.0	0.0	0.0	750.0	0.0	0.0
	x_1	x_2	x_3	x_4	x_5	x_6
final value of degree of saturation	0.53	0.00	0.00	0.53	0.00	0.00

Table 6.6 Results on two-junction network for method a

iteration number	step type	PI veh-h/h	grad_PI wrt alpha	alpha_ max	alpha_ ratio	start of green in seconds: junction 1 θ_{11}/ξ	θ_{21}/ξ	junction 2 θ_{12}/ξ	θ_{22}/ξ	θ_{32}/ξ	duration of green in seconds: junction 1 ϕ_{11}/ξ	ϕ_{21}/ξ	junction 2 ϕ_{12}/ξ	ϕ_{22}/ξ	ϕ_{32}/ξ	cycle time sec $(\frac{1}{\xi})$
1	I/II	7.14	-63.50	0.17	0.25	0.0	40.0	0.0	46.0	0.0	35.0	15.0	41.0	9.0	60.0	60.0
2	I/II	4.58	-11.20	0.65	0.12	0.1	45.0	59.9	46.3	59.9	39.9	10.1	41.3	8.7	60.0	60.0
3	I/II	3.71	-2.14	0.50	1.00	0.0	47.8	60.0	46.0	60.0	42.8	7.2	41.0	8.9	60.0	60.0
4	III	3.64	-0.84	0.50	0.90	0.0	48.0	60.0	48.0	60.0	43.0	7.0	43.0	7.0	60.0	60.0
5	I/II	3.13	-0.50	0.05	0.20	59.9	47.9	59.9	47.9	59.9	43.0	7.0	43.0	7.0	60.0	60.0
6		3.11				0.0	47.9	60.0	48.0	60.0	42.9	7.1	43.0	7.0	60.0	60.0

iteration number	improvement rate	link flow in veh/h q_1	q_2	q_3	q_4	q_5	q_6
1	35.85%	615.0	45.0	225.0	615.0	225.0	45.0
2	19.00%	707.0	21.8	131.6	639.6	131.6	21.8
3	1.89%	771.7	1.8	63.5	663.0	63.5	1.8
4	14.01%	746.7	11.5	1.1	740.6	1.1	11.5
5	0.64%	750.0	0.0	0.0	750.0	0.0	0.0
6		750.3	0.1	0.1	749.6	0.1	0.1
		x_1	x_2	x_3	x_4	x_5	x_6
final value of degree of saturation		0.57	0.00	0.00	0.57	0.00	0.00

Table 6.7 Results on two-junction network for method b

iteration number	step type	PI veh-h/h	grad_PI wrt alpha	alpha_ max	alpha_ ratio	start of green in seconds: junction 1 θ_{11}/ξ	θ_{21}/ξ	junction 2 θ_{12}/ξ	θ_{22}/ξ	θ_{32}/ξ	duration of green in seconds: junction 1 ϕ_{11}/ξ	ϕ_{21}/ξ	junction 2 ϕ_{12}/ξ	ϕ_{22}/ξ	ϕ_{32}/ξ	cycle time sec $(\frac{1}{\xi})$
1	III	7.14	-1.80	0.85	0.50	0.0	40.0	0.0	46.0	0.0	35.0	15.0	41.0	9.0	60.0	60.0
2	I/II	6.37	-3.18	0.52	0.50	1.1	41.1	0.6	46.6	0.6	39.9	10.1	41.3	8.7	60.0	60.0
3	III	5.55	-2.37	0.75	0.45	60.0	45.5	0.0	39.6	0.0	40.5	9.5	34.6	15.4	60.0	60.0
4	I/II	4.75	-7.38	0.07	1.00	0.1	45.6	59.9	39.5	59.9	40.5	9.5	34.6	15.4	60.0	60.0
5	III	4.51	-2.68	0.65	0.75	60.0	48.0	0.0	46.7	0.0	43.0	7.0	41.6	8.4	60.0	60.0
6	I/II	3.20	-1.23	0.65	0.10	59.9	47.9	0.0	46.7	0.0	43.0	7.0	41.6	8.4	60.0	60.0
7		3.12				0.0	45.5	0.0	45.8	0.0	40.5	9.5	40.8	9.2	60.0	60.0

iteration number	improvement rate	link flow in veh/h q_1	q_2	q_3	q_4	q_5	q_6
1	10.78%	615.0	45.0	225.0	615.0	225.0	45.0
2	12.87%	625.0	35.0	200.0	640.0	200.0	35.0
3	14.41%	770.1	62.1	43.1	624.7	43.1	62.1
4	5.05%	770.0	40.0	30.0	660.0	30.0	40.0
5	29.05%	770.7	39.9	6.5	682.9	6.5	39.9
6	2.50%	750.0	10.0	10.0	730.0	10.0	10.0
7		752.2	6.3	2.8	738.7	2.8	6.3
		x_1	x_2	x_3	x_4	x_5	x_6
final value of degree of saturation		0.60	0.02	0.00	0.59	0.00	0.02

Table 6.8 Computation efforts for Hooke and Jeeves' method (HJ), mutually consistent calculations(mc) and mixed search variants a&b (mixa, mixb) with respect to the number of times of solving the equilibrium traffic assignment (ETA) and corresponding number of executions of the convex combination method (CCM)

	HJ	mc	mixa	mixb
Number of times of solving ETA	37	11	4	6
Number of executions of CCM	45	33	9	12

specified with the start times for signal groups equally distributed over the cycle at each junction, while the second set of initial signal timings are specified to favour the roads via junctions 5 and 6 for north-south and south-north traffic flows. The corresponding initial equilibrium link flows are respectively given. In this numerical test, as in the two-junction network the minimum green duration is 7 seconds for each signal group, the clearance times between each pair of the incompatible signal group at each junction is 5 seconds and the maximum common cycle time is specified as 130 seconds. In the following subsections, the two variants of mixed search procedure and the mutually consistent calculation are performed on the basis of the same two sets of initial signal settings; furthermore, in order to understand the characteristics of the non-optimal mutually consistent signal settings and equilibrium flows, a re-run of the mutually consistent calculation is carried out starting from the solutions obtained by the mixed search methods.

6.5.2.1 Results for mixed search procedure

Computation results for Allsop and Charlesworth's road network are shown in Tables 6.9-6.12 for the variants of the proposed method using the first and second sets of initial signal timings. As is seen in Tables 6.9-6.12, the two sets of initial signal timings gave rise to substantially different initial performance index values of 499 and 661 veh-h/h.

Firstly, for the first set of initial signal settings in Table 6.9, *method a* starts from the initial solution with negative values of the total derivatives for the performance index with respect to the move size α and follows the sequence of steps of types I-II, III

and I-II of the mixed search procedure to achieve a good local optimum for the bi-level formulation. As we can see, the steps of type I/II apply to the iterations 1-3, which reduces the value of performance index to 43 veh-h/h and improves the system performance by approximately 90%. Following the step of type III applied in iteration 4 a better starting point is located at which the system performance is improved by 1.9%; since a global change has been made by applying the step of type III, a re-assignment process is carried out and thereafter a local optimal search is made in iteration 5 by applying the step of type II and the resulting value of the performance index is reduced by about 92% to 41.39 veh-h/h. The maximum final value of degree of saturation is 0.88. The number of full iterations carrying out the mixed search in this numerical test is 5 and the predetermined thresholds in the system performance improvement rate for the steps of type III is less than 2% and for the steps of type I or II is less than 1%.

Furthermore, the *method b* adopting the optimal search technique used in TRANSYT which is carried out by combining the small and large steps for the starts of green times starts from type III by simultaneous and equal changing the starts of greens at any one junction and therefore searching many parts of the feasible region a good initial solution can be localized as a new starting point for the subsequent search process carried out by types I-II of the mixed search procedure. As is seen in Table 6.10 the steps of type III apply to the iterations 1, 3 and 6 and alternately the steps of type I/II apply to the iterations 2, 4, 5 and 7. After each iteration of carrying out the step of type III, a re-assignment process is performed and a good local point in another part of the feasible region is identified as the next input for the subsequent local optimal search. The number of iterations of this mixed search in this numerical test is 7 and the predetermined thresholds in the system performance improvement rate for the steps of type III is less than 2% and for the steps of type I or II is less than 1%. The final value of the performance index is 38.62 veh-h/h and total improvement of system performance achieved by the *method b* is about 92%. The maximum final value of degree of saturation is 0.83.

Secondly, for the second set of initial signal settings, the results are shown in Tables 6.11-6.12. Similarly as described for the first initial signal settings, the *method a*

performs 5 full iterations of mixed search and the final value of the performance index is reduced to 39.57 veh-h/h from the initial value of 661.04 veh-h/h, which improves the system performance by 94% and the maximum final value of degree of saturation is 0.86. The predetermined thresholds in the system performance improvement rate for the steps of type III is less than 2.2% and for the steps of type I or II is less than 1%. Again in this numerical test the *method b* performs 6 full iterations of mixed search procedure and the final value of the performance index is reduced to 39.37 veh-h/h, which improves the system performance by 94% and the maximum final value of degree of saturation is 0.87.

In relation to the resulting path flows, using the mixed search procedure has transferred the initial path flows between the OD pairs of AD, AE, AF, DA, DB, EA, EB, GA and GD via junctions 5,6 to the path flows via junctions 2,3 and via junctions 3,5 and 6,2, which can be found in the changes of link flows from iteration to iteration in Tables 6.9-6.12.

6.5.2.2 Results for mutually consistent calculations

Computation results performed by the mutually consistent calculations with the same two sets of initials as those in the mixed search are respectively given in Tables 6.13-6.14. For the first set of initial signal settings, as is seen in the first iteration in Table 6.13 after performing TRANSYT with the corresponding equilibrium flows the value of performance index reduced to 99.78 veh-h/h from the initial value of 498.76 veh-h/h, which achieves the greatest improvement of system performance in one iteration by 80%. Thereafter alternately carrying out the two separate procedures of traffic assignment and TRANSYT optimisation of signal timings, the final value of the performance index is reduced to 89.13 veh-h/h and the total system performance is improved by 82.12%. Fluctuation of the value of the performance index from iteration to iteration is obvious, which shows the non-optimal characteristics of the mutually consistent signal settings and equilibrium flows for the solution of the bi-level problem. The total number of iterations in performing the mutually consistent calculations is 7 and the maximum final value of the degree of saturation is 0.83.

For the second set of initial signal settings, as is seen in the first iteration in Table

Table 6.9 Results on Allsop & Charlesworth's network at 1st initial signal settings for method a

iteration number	step type	start of green in seconds																	cycle time ($\frac{1}{s}$)
		junction 1			junction 2		junction 3		junction 4					junction 5				junction 6	
		θ_{11}/s	θ_{21}/s	θ_{31}/s	θ_{12}/s	θ_{22}/s	θ_{13}/s	θ_{23}/s	θ_{14}/s	θ_{24}/s	θ_{34}/s	θ_{44}/s	θ_{54}/s	θ_{15}/s	θ_{25}/s	θ_{35}/s	θ_{45}/s	θ_{16}/s	θ_{26}/s	
1	I/II	0.0	30.0	0.0	0.0	30.0	0.0	30.0	0.0	20.0	40.0	0.0	40.0	0.0	20.0	40.0	0.0	0.0	30.0	60.0
2	I/II	9.8	58.9	9.8	83.7	46.6	84.3	36.2	0.0	28.8	52.8	0.0	52.8	0.6	24.1	46.8	0.6	1.0	36.6	85.0
3	II	16.4	86.9	16.4	117.3	66.8	118.6	46.3	0.0	41.0	63.9	0.0	63.9	1.4	25.2	48.5	1.4	2.1	46.5	120.0
4	III	17.1	87.9	17.1	117.2	67.2	118.5	45.5	0.0	41.0	62.3	0.0	62.3	1.4	24.2	46.9	1.4	2.2	46.2	120.0
5	II	82.0	32.8	82.0	116.8	66.8	0.5	47.4	0.5	41.5	62.8	0.5	62.8	1.4	24.2	46.9	1.4	2.5	46.5	120.0
6		57.7	8.5	57.7	117.2	67.2	3.4	50.3	1.2	42.2	63.4	1.2	63.4	1.4	24.2	46.9	1.4	3.0	47.0	120.0

iteration number	duration of green in seconds																	
	junction 1			junction 2		junction 3		junction 4					junction 5				junction 6	
	ϕ_{11}/s	ϕ_{21}/s	ϕ_{31}/s	ϕ_{12}/s	ϕ_{22}/s	ϕ_{13}/s	ϕ_{23}/s	ϕ_{14}/s	ϕ_{24}/s	ϕ_{34}/s	ϕ_{44}/s	ϕ_{54}/s	ϕ_{15}/s	ϕ_{25}/s	ϕ_{35}/s	ϕ_{45}/s	ϕ_{16}/s	ϕ_{26}/s
1	25.0	25.0	60.0	25.0	25.0	25.0	25.0	15.0	15.0	15.0	35.0	35.0	15.0	15.0	15.0	35.0	25.0	25.0
2	44.1	30.9	85.0	42.9	32.1	31.9	43.1	23.8	19.0	27.2	47.8	55.0	18.5	17.7	33.8	41.2	30.7	44.3
3	65.5	44.5	120.0	64.5	45.5	42.8	67.2	36.0	17.9	51.2	58.9	92.2	18.9	18.3	67.8	42.2	39.3	70.7
4	65.8	44.2	120.0	65.0	45.0	41.9	68.1	36.0	16.2	52.8	57.2	93.8	17.8	17.7	69.5	40.5	39.0	71.0
5	65.8	44.2	120.0	65.0	45.0	41.9	68.1	36.0	16.2	52.8	57.2	93.8	17.8	17.7	69.5	40.5	39.0	71.0
6	65.8	44.2	120.0	65.0	45.0	41.9	68.1	36.0	16.2	52.8	57.2	93.8	17.8	17.7	69.5	40.5	39.0	71.0

iteration number	link flow in veh/h															
	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9	q_{10}	q_{11}	q_{12}	q_{13}	q_{14}	q_{15}	q_{16}
1	950.0	230.0	950.0	840.0	880.0	150.0	230.0	250.0	90.0	260.0	460.0	290.0	450.0	740.0	790.0	610.0
2	1075.4	104.6	1075.4	934.8	974.1	147.7	104.6	156.9	88.4	168.2	460.5	289.5	450.0	794.0	841.9	660.9
3	1090.7	89.3	1090.7	943.4	973.4	152.5	89.4	152.4	84.6	164.6	588.0	162.0	450.0	952.6	983.3	794.0
4	1086.8	93.2	1086.8	939.6	971.1	152.8	93.3	154.6	86.4	166.6	596.8	153.2	450.0	968.4	994.5	820.8
5	1081.3	98.7	1081.3	937.4	973.3	150.5	98.7	154.7	88.4	166.7	597.5	152.5	450.0	970.2	995.9	824.6
6	1075.9	104.1	1075.9	935.2	975.8	148.1	104.1	154.7	90.7	166.7	598.5	151.5	450.0	971.9	997.1	823.0

iteration number	q_{17}	q_{18}	q_{19}	q_{20}	q_{21}	q_{22}	q_{23}
1	460.0	350.0	680.0	1290.0	1050.0	1250.0	810.0
2	406.1	298.2	629.1	1290.0	1055.1	1250.0	778.4
3	247.9	157.2	496.0	1290.0	1076.8	1250.0	763.5
4	232.1	146.0	469.2	1290.0	1079.6	1250.0	779.1
5	230.3	144.6	465.4	1290.0	1077.9	1250.0	784.8
6	228.7	143.5	467.0	1290.0	1076.1	1250.0	785.3

final value of degree of saturation

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
0.58	0.17	0.61	0.82	0.72	0.18	0.17	0.28	0.34	0.39	0.88	0.19	0.46	0.85	0.70	0.51

x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}
0.86	0.25	0.83	0.80	0.57	0.58	0.64

avg	max
0.54	0.88

Table 6.9a Results on Allsop & Charlesworth's network at 1st initial signal settings for method a (cont)

Iteration number	PI veh-h/h	grad_PI wrt alpha	Improvement rate	alpha_max	alpha_good/alpha_max
1	498.76	-1040.48	63.90%	0.59	0.50
2	180.05	-320.53	73.31%	0.48	0.65
3	48.05	-98.91	11.34%	0.38	0.15
4	42.60	-4.89	1.88%	0.49	0.35
5	41.80	-4.42	0.98%	0.63	0.15
6	41.39				

Table 6.10a Results on Allsop & Charlesworth's network at 1st initial signal settings for method b (cont)

Iteration number	PI veh-h/h	grad_PI wrt alpha	Improvement rate	alpha_max	alpha_good/alpha_max
1	498.76	-45.43	4.61%	4.90	0.10
2	475.79	-1144.89	78.04%	0.47	0.73
3	104.49	-74.08	4.95%	0.30	0.25
4	99.32	-121.70	57.15%	0.60	0.78
5	42.56	-71.39	6.88%	0.80	0.05
6	39.63	-1.07	1.99%	1.48	0.50
7	38.84	-1.05	0.57%	1.47	0.15
8	38.62				

Table 6.11a Results on Allsop & Charlesworth's network at 2nd initial signal settings for method a (cont)

Iteration number	PI veh-h/h	grad_PI wrt alpha	Improvement rate	alpha_max	alpha_good/alpha_max
1	661.04	-815.27	75.88%	0.89	0.70
2	159.41	-209.27	53.05%	0.74	0.55
3	74.84	-91.70	45.84%	0.92	0.40
4	40.53	-2.44	2.17%	0.59	0.61
5	39.85	-0.20	0.20%	0.91	0.40
6	39.57				

Table 6.12a Results on Allsop & Charlesworth's network at 2nd initial signal settings for method b (cont)

Iteration number	PI veh-h/h	grad_PI wrt alpha	Improvement rate	alpha_max	alpha_good/alpha_max
1	661.04	-15.29	0.92%	1.02	0.38
2	654.95	-761.39	79.69%	0.83	0.80
3	133.02	-65.86	3.71%	0.21	0.35
4	128.08	282.61	68.70%	0.59	0.52
5	40.09	-6.44	1.27%	0.11	0.75
6	39.58	-0.65	0.53%	0.57	0.56
7	39.37				

Table 6.10 Results on Allsop & Charlesworth's network at 1st initial signal settings for method b

start of green in seconds

iteration number	step type	junction 1 θ_{11}/s	θ_{21}/s	θ_{31}/s	junction 2 θ_{12}/s	θ_{22}/s	junction 3 θ_{13}/s	θ_{23}/s	junction 4 θ_{14}/s	θ_{24}/s	θ_{34}/s	θ_{44}/s	θ_{54}/s	junction 5 θ_{15}/s	θ_{25}/s	θ_{35}/s	θ_{45}/s	junction 6 θ_{16}/s	θ_{26}/s	cycle time $(\frac{1}{s})$
1	III	0.0	30.0	0.0	0.0	30.0	0.0	30.0	0.0	20.0	40.0	0.0	40.0	0.0	20.0	40.0	0.0	0.0	30.0	60.0
2	I/II	36.3	6.3	36.3	22.6	52.6	56.2	26.2	0.6	20.6	40.6	0.6	40.6	59.9	19.9	39.9	59.9	4.0	34.0	60.0
3	III	58.8	18.8	58.8	34.1	83.1	84.3	31.6	0.9	31.5	55.2	0.9	55.2	89.8	24.3	45.6	89.8	6.0	47.6	90.0
4	I/II	86.3	46.4	86.3	37.3	86.3	85.2	32.5	1.1	31.7	55.4	1.1	55.4	89.8	24.3	45.6	89.8	6.6	48.2	90.0
5	II	121.1	64.1	121.1	51.6	120.4	117.4	41.5	1.6	43.7	73.9	1.6	73.9	123.7	33.4	62.5	123.7	9.2	64.8	124.0
6	III	122.2	64.9	122.2	51.6	121.4	117.4	41.8	1.6	43.7	73.0	1.6	73.0	123.7	33.3	62.3	123.7	9.2	64.4	124.0
7	II	1.4	68.0	1.4	51.6	121.4	117.4	41.8	1.6	43.7	73.0	1.6	73.0	123.7	33.3	62.3	123.7	8.6	63.9	124.0
8		4.0	70.7	4.0	51.6	121.4	117.4	41.8	1.6	43.7	73.0	1.6	73.0	123.7	33.2	62.3	123.7	8.1	63.4	124.0

duration of green in seconds

iteration number	junction 1 ϕ_{11}/s	ϕ_{21}/s	ϕ_{31}/s	junction 2 ϕ_{12}/s	ϕ_{22}/s	junction 3 ϕ_{13}/s	ϕ_{23}/s	junction 4 ϕ_{14}/s	ϕ_{24}/s	ϕ_{34}/s	ϕ_{44}/s	ϕ_{54}/s	junction 5 ϕ_{15}/s	ϕ_{25}/s	ϕ_{35}/s	ϕ_{45}/s	junction 6 ϕ_{16}/s	ϕ_{26}/s
1	25.0	25.0	60.0	25.0	25.0	25.0	25.0	15.0	15.0	15.0	35.0	35.0	15.0	15.0	15.0	35.0	25.0	25.0
2	25.0	25.0	60.0	25.0	25.0	25.0	25.0	15.0	15.0	15.0	35.0	35.0	15.0	15.0	15.0	35.0	25.0	25.0
3	45.0	35.0	90.0	43.9	36.1	32.4	47.6	25.5	18.7	30.8	49.2	61.3	19.5	16.3	39.2	40.8	36.5	43.5
4	45.0	35.0	90.0	43.9	36.1	32.4	47.6	25.5	18.7	30.8	49.2	61.3	19.5	16.3	39.2	40.8	36.5	43.5
5	62.0	52.0	124.0	63.8	50.2	43.2	70.8	37.1	25.2	46.7	67.3	88.8	28.7	24.1	56.2	57.8	50.6	63.4
6	61.7	52.3	124.0	64.8	49.2	43.4	70.6	37.2	24.3	47.6	66.5	89.8	28.6	24.0	56.4	57.6	50.3	63.7
7	61.7	52.3	124.0	64.8	49.2	43.4	70.6	37.2	24.3	47.6	66.5	89.8	28.6	24.0	56.4	57.6	50.3	63.7
8	61.7	52.3	124.0	64.8	49.2	43.4	70.7	37.2	24.3	47.6	66.5	89.8	28.5	24.0	56.5	57.6	50.3	63.7

link flow in veh/h

iteration number	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9	q_{10}	q_{11}	q_{12}	q_{13}	q_{14}	q_{15}	q_{16}
1	950.0	230.0	950.0	840.0	880.0	150.0	230.0	250.0	90.0	260.0	460.0	290.0	450.0	740.0	790.0	610.0
2	960.4	219.6	960.4	845.5	881.8	154.1	219.6	244.1	90.4	254.2	460.0	290.0	450.0	746.8	793.4	610.0
3	1041.2	138.8	1041.2	890.6	907.2	155.6	138.8	215.7	73.9	227.3	534.7	215.3	450.0	850.6	873.9	706.4
4	1048.1	131.9	1048.1	895.6	912.9	157.9	131.9	207.7	76.9	219.2	534.5	215.5	450.0	852.5	875.8	707.6
5	1057.2	122.8	1057.2	903.4	925.0	157.8	122.8	195.7	80.9	207.2	591.6	158.4	450.0	965.9	993.8	927.7
6	1059.1	120.9	1059.1	904.2	926.0	158.0	120.9	194.5	81.3	206.0	558.8	191.2	450.0	938.5	966.6	945.7
7	1057.7	122.3	1057.7	902.9	927.0	156.6	122.3	194.9	82.2	206.4	560.7	189.3	450.0	940.5	968.5	946.8
8	1061.0	119.0	1061.0	902.1	925.9	157.7	119.0	195.0	83.0	206.4	562.2	187.8	450.0	942.3	970.3	947.7

iteration number	q_{17}	q_{18}	q_{19}	q_{20}	q_{21}	q_{22}	q_{23}
1	460.0	350.0	680.0	1290.0	1050.0	1250.0	810.0
2	453.2	346.7	680.0	1290.0	1053.1	1250.0	801.8
3	349.5	266.2	583.6	1290.0	1094.4	1250.0	781.8
4	347.5	264.3	582.4	1290.0	1091.5	1250.0	779.2
5	234.2	146.2	362.3	1290.0	1082.6	1250.0	880.1
6	261.5	173.4	344.3	1290.0	1082.0	1250.0	924.2
7	259.5	171.5	343.2	1290.0	1081.2	1250.0	923.5
8	257.7	169.7	342.3	1290.0	1080.4	1250.0	918.4

final value of degree of saturation

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
0.57	0.17	0.62	0.79	0.74	0.22	0.16	0.24	0.20	0.42	0.83	0.21	0.52	0.82	0.70	0.65

x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}
0.75	0.24	0.53	0.80	0.73	0.67	0.71

avg	max
0.53	0.83

Table 6.11 Results on Allsop & Charlesworth's network at 2nd initial signal settings for method a

iteration number	step type	start of green in seconds junction 1 θ_{11}/s	θ_{21}/s	θ_{31}/s	junction 2 θ_{12}/s	θ_{22}/s	junction 3 θ_{13}/s	θ_{23}/s	junction 4 θ_{14}/s	θ_{24}/s	θ_{34}/s	θ_{44}/s	θ_{54}/s	junction 5 θ_{15}/s	θ_{25}/s	θ_{35}/s	θ_{45}/s	junction 6 θ_{16}/s	θ_{26}/s	cycle time $(\frac{1}{s})$
1	I/II	0.0	32.5	0.0	0.0	32.5	0.0	42.5	0.0	20.0	55.0	0.0	55.0	0.0	25.0	50.0	0.0	0.0	42.5	75.0
2	I/II	15.9	66.1	15.9	90.3	46.4	0.0	39.3	0.0	26.3	61.2	0.0	61.2	93.2	27.0	47.4	93.2	2.9	43.5	96.0
3	I/II	34.5	106.1	34.5	111.0	59.0	0.1	48.8	0.0	34.2	65.5	0.0	65.5	115.5	31.3	51.0	115.5	4.6	44.3	120.0
4	II	45.8	120.7	45.8	119.6	58.0	0.3	55.1	0.1	47.3	72.7	0.1	72.7	124.2	26.0	45.2	124.2	5.2	46.4	129.0
5	II	50.7	125.7	50.7	119.9	58.2	0.7	55.4	0.2	47.4	72.8	0.2	72.8	124.2	26.0	45.2	124.2	5.4	46.6	129.0
6		51.1	126.1	51.1	120.2	53.9	1.1	56.2	0.3	48.0	73.1	0.3	73.1	124.2	22.7	41.7	124.2	5.6	44.3	129.0

iteration number	duration of green in seconds junction 1 ϕ_{11}/s	ϕ_{21}/s	ϕ_{31}/s	junction 2 ϕ_{12}/s	ϕ_{22}/s	junction 3 ϕ_{13}/s	ϕ_{23}/s	junction 4 ϕ_{14}/s	ϕ_{24}/s	ϕ_{34}/s	ϕ_{44}/s	ϕ_{54}/s	junction 5 ϕ_{15}/s	ϕ_{25}/s	ϕ_{35}/s	ϕ_{45}/s	junction 6 ϕ_{16}/s	ϕ_{26}/s
1	27.5	37.5	75.0	27.5	37.5	37.5	27.5	15.0	30.0	15.0	50.0	35.0	20.0	20.0	20.0	45.0	37.5	27.5
2	45.3	40.7	96.0	47.1	38.9	34.3	51.7	21.3	29.9	29.8	56.2	56.1	24.8	15.4	40.8	45.2	35.6	50.4
3	66.6	43.4	120.0	63.0	47.0	43.7	66.3	29.2	26.3	49.5	60.5	83.7	30.7	14.7	59.6	50.4	34.7	75.3
4	70.0	49.0	129.0	62.4	56.6	49.8	69.2	42.2	20.3	51.4	67.5	98.6	25.8	14.1	74.1	44.9	36.2	82.8
5	70.0	49.0	129.0	62.4	56.6	49.8	69.2	42.2	20.3	51.4	67.5	98.6	25.8	14.1	74.1	44.9	36.2	82.8
6	69.9	49.1	129.0	57.6	61.4	50.1	68.9	42.6	20.1	51.3	67.7	98.9	22.5	14.0	77.5	41.5	33.7	85.3

iteration number	link flow in veh/h q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9	q_{10}	q_{11}	q_{12}	q_{13}	q_{14}	q_{15}	q_{16}
1	810.0	370.0	810.0	770.0	700.0	140.0	470.0	350.0	210.0	500.0	380.0	370.0	450.0	640.0	700.0	620.0
2	830.6	349.4	830.6	784.2	705.0	144.2	449.0	340.4	205.6	491.4	469.7	280.3	450.0	826.8	881.3	791.9
3	969.7	210.3	969.7	887.4	807.9	143.1	247.3	214.2	143.1	366.6	529.5	220.5	450.0	970.0	970.2	941.3
4	1086.8	93.2	1086.8	939.6	971.1	152.8	93.3	154.6	86.4	166.6	596.8	153.2	450.0	968.4	994.5	820.8
5	1011.5	168.5	1011.5	923.9	898.3	143.4	190.2	100.0	183.0	253.4	585.5	164.5	450.0	1036.2	1019.1	1034.5
6	1012.3	167.7	1012.3	924.5	897.9	143.4	189.4	100.2	181.9	253.7	583.9	166.1	450.0	1034.4	1017.5	1032.7

iteration number	q_{17}	q_{18}	q_{19}	q_{20}	q_{21}	q_{22}	q_{23}
1	510.0	590.0	670.0	1290.0	1160.0	1250.0	1080.0
2	423.8	408.9	498.1	1290.0	1170.5	1250.0	1063.8
3	257.7	234.5	348.7	1290.0	1226.1	1250.0	1025.9
4	192.4	174.5	303.9	1290.0	1206.7	1250.0	1029.4
5	168.9	147.7	255.5	1290.0	1189.2	1250.0	1049.5
6	170.6	149.2	257.3	1290.0	1190.1	1250.0	1049.1

final value of degree of saturation

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
0.54	0.27	0.70	0.73	0.68	0.19	0.39	0.19	0.59	0.64	0.78	0.19	0.50	0.82	0.86	0.65

x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}
0.86	0.33	0.44	0.85	0.61	0.52	0.68

avg	max
0.57	0.86

Table 6.12 Results on Allsop & Charlesworth's network at 2nd initial signal settings for method b

iteration number	step type	start of green in seconds																	cycle time
		junction 1			junction 2		junction 3		junction 4				junction 5				junction 6		
		θ_{11}/ξ	θ_{21}/ξ	θ_{31}/ξ	θ_{12}/ξ	θ_{22}/ξ	θ_{13}/ξ	θ_{23}/ξ	θ_{14}/ξ	θ_{24}/ξ	θ_{34}/ξ	θ_{44}/ξ	θ_{54}/ξ	θ_{15}/ξ	θ_{25}/ξ	θ_{35}/ξ	θ_{45}/ξ	θ_{16}/ξ	θ_{26}/ξ	$(\frac{1}{\xi})$
1	III	0.0	32.5	0.0	0.0	32.5	0.0	42.5	0.0	20.0	55.0	0.0	55.0	0.0	25.0	50.0	0.0	0.0	42.5	75.0
2	I/II	32.8	65.3	32.8	1.2	33.7	74.8	42.3	0.1	20.1	55.1	0.1	55.1	74.8	24.8	49.8	74.8	1.0	43.5	75.0
3	III	65.6	2.2	65.6	2.4	66.8	127.6	59.9	0.2	34.6	72.6	0.2	72.6	127.6	31.3	56.2	127.6	2.0	68.2	128.0
4	II	81.2	17.8	81.2	11.1	75.5	127.6	59.8	0.6	35.1	73.0	0.6	73.0	127.2	30.9	55.8	127.2	5.2	71.5	128.0
5	III	80.5	24.0	80.5	11.4	76.8	127.6	51.3	0.6	47.2	71.8	0.6	71.8	127.2	30.2	50.6	127.2	5.3	67.6	128.0
6	II	82.8	26.4	82.8	12.9	78.3	127.9	51.7	0.8	47.4	72.0	0.8	72.0	127.2	30.2	50.6	127.2	5.8	68.1	128.0
7		82.9	26.4	82.9	13.4	78.9	127.7	51.5	0.8	47.5	72.1	0.8	72.1	127.2	30.2	50.6	127.2	5.6	67.8	128.0

iteration number	duration of green in seconds																
	junction 1			junction 2		junction 3		junction 4				junction 5				junction 6	
	ϕ_{11}/ξ	ϕ_{21}/ξ	ϕ_{31}/ξ	ϕ_{12}/ξ	ϕ_{22}/ξ	ϕ_{13}/ξ	ϕ_{23}/ξ	ϕ_{14}/ξ	ϕ_{24}/ξ	ϕ_{34}/ξ	ϕ_{44}/ξ	ϕ_{54}/ξ	ϕ_{15}/ξ	ϕ_{25}/ξ	ϕ_{35}/ξ	ϕ_{45}/ξ	ϕ_{16}/ξ	ϕ_{26}/ξ
1	27.5	37.5	75.0	27.5	37.5	37.5	27.5	15.0	30.0	15.0	50.0	35.0	20.0	20.0	20.0	45.0	37.5	27.5
2	27.5	37.5	75.0	27.5	37.5	37.5	27.5	15.0	30.0	15.0	50.0	35.0	20.0	20.0	20.0	45.0	37.5	27.5
3	59.6	58.4	128.0	59.4	58.6	55.3	62.7	29.5	33.0	50.6	67.5	85.1	26.7	19.9	66.4	51.6	61.3	56.7
4	59.6	58.4	128.0	59.4	58.6	55.3	62.7	29.5	33.0	50.6	67.5	85.1	26.7	19.9	66.4	51.6	61.3	56.7
5	66.5	51.5	128.0	60.4	57.6	46.7	71.3	41.6	19.6	51.8	66.2	98.4	26.0	15.4	71.6	46.4	57.3	60.7
6	66.5	51.5	128.0	60.4	57.6	46.7	71.3	41.6	19.6	51.8	66.2	98.4	26.0	15.4	71.6	46.4	57.3	60.7
7	66.5	51.5	128.0	60.4	57.6	46.7	71.3	41.6	19.6	51.8	66.2	98.4	25.9	15.4	71.6	46.4	57.3	60.7

iteration number	link flow in veh/h															
	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9	q_{10}	q_{11}	q_{12}	q_{13}	q_{14}	q_{15}	q_{16}
1	810.0	370.0	810.0	770.0	700.0	140.0	470.0	350.0	210.0	500.0	380.0	370.0	450.0	640.0	700.0	620.0
2	815.3	364.7	815.3	771.2	700.9	140.1	463.6	348.9	208.7	498.9	387.4	362.6	450.0	651.4	712.5	624.6
3	976.0	204.0	976.0	902.5	850.7	139.1	296.1	208.5	220.9	360.1	457.2	292.8	450.0	850.9	882.0	852.6
4	988.4	191.6	988.4	907.4	852.0	140.3	282.6	203.8	217.6	355.4	467.9	282.1	450.0	867.0	898.5	851.9
5	1015.0	165.0	1015.0	918.0	944.7	144.0	181.6	133.5	151.5	207.7	628.3	121.7	450.0	1040.8	1011.6	1033.1
6	1015.6	164.5	1015.6	917.6	945.1	144.1	177.2	133.5	148.4	207.5	623.9	126.1	450.0	1034.9	1001.8	1029.6
7	1015.3	164.7	1015.3	917.6	945.1	144.1	177.4	133.5	148.3	207.6	621.7	128.3	450.0	1032.3	999.1	1023.3

iteration number							
	q_{17}	q_{18}	q_{19}	q_{20}	q_{21}	q_{22}	q_{23}
1	610.0	590.0	670.0	1290.0	1160.0	1250.0	1080.0
2	598.5	576.4	665.4	1290.0	1159.2	1250.0	1066.9
3	409.0	410.0	437.4	1290.0	1171.7	1250.0	1089.2
4	390.7	390.2	438.1	1290.0	1173.5	1250.0	1063.4
5	165.6	151.4	256.9	1290.0	1148.4	1250.0	1041.1
6	171.8	157.7	260.4	1290.0	1151.4	1250.0	1042.6
7	174.4	160.1	266.7	1290.0	1151.6	1250.0	1039.1

final value of degree of saturation

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
0.55	0.25	0.66	0.77	0.71	0.19	0.22	0.22	0.41	0.48	0.85	0.15	0.50	0.87	0.80	0.67

x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}
0.80	0.21	0.43	0.82	0.63	0.72	0.71

avg	max
0.55	0.87

6.14 after performing TRANSYT with the corresponding equilibrium flows the value of performance index reduced to 216.12 veh-h/h from the initial value of 661.04 veh-h/h, which achieves the greatest improvement of system performance in one iteration by nearly 70%. Thereafter alternately carrying out the two separate procedures of traffic assignment and TRANSYT optimisation of signal timings, the final value of the performance index is reduced to 105.17 veh-h/h and the total system performance is improved by 84.1%. Fluctuation of the value of the performance index from iteration to iteration during this numerical test is rather obvious in this numerical test. The total number of iterations in performing the mutually consistent calculations is 9 and the maximum final value of the degree of saturation is 0.90.

6.5.2.3 Results for re-run mutually consistent calculation followed by mixed search

Computation results for the re-run mutually consistent calculations starting with the equilibrium flows given by the mixed search method are given in Tables 6.15-6.16. In this numerical test, we use the flows produced by the *method b* of the mixed search procedure as the initial values for re-run mutually consistent calculations because the results produced by *method b* are slightly better as those did *method a* as it has shown in Tables 6.9-6.12.

As is seen in the first iteration in Table 6.15, the value of the performance index increases to a higher value of 93.78 veh-h/h from the initial value of 38.62 veh-h/h, and the similar situation appears also in the first iteration in Table 6.16 when the value of the performance index increases to 100.77 veh-h/h from the initial value of 39.37 veh-h/h. In the subsequent iterations, in the Table 6.15 the values of the performance index continue increasing in iterations 1-2, and 3-5 and settles down to a resulting value of 94.53 veh-h/h, which is much worse in terms of system performance than the values given by the optimisation method. In Table 6.16 the values of the performance index continue increasing in iterations 1-2, and 3-4 and settle down to a final value of 99.58 veh-h/h, which is again much worse in terms of system performance than those given by the optimisation method. All these results show the non-optimal characteristics of the mutually consistent calculations when apply to the bi-level problem.

Table 6.13 Results on Allsop&Charlesworth's network at 1st initial signal settings for mutually consistent calculations

iteration number	cycle time (s)	PI veh-h/h	start of green in seconds																
			junction 1			junction 2		junction 3		junction 4				junction 5				junction 6	
			θ_{11}/s	θ_{21}/s	θ_{31}/s	θ_{12}/s	θ_{24}/s	θ_{13}/s	θ_{23}/s	θ_{14}/s	θ_{34}/s	θ_{14}/s	θ_{44}/s	θ_{54}/s	θ_{15}/s	θ_{25}/s	θ_{35}/s	θ_{45}/s	θ_{16}/s	θ_{26}/s
1	60.0	198.76	0.0	30.0	0.0	0.0	30.0	0.0	30.0	0.0	20.0	40.0	0.0	40.0	0.0	20.0	40.0	0.0	0.0	30.0
2	60.0	99.78	46.0	8.0	46.0	32.0	6.0	3.0	27.0	22.0	46.0	0.0	22.0	0.0	11.0	23.0	45.0	11.0	35.0	59.0
3	75.0	88.24	54.0	3.0	54.0	2.0	48.0	0.0	32.0	27.0	58.0	74.0	27.0	74.0	10.0	22.0	49.0	10.0	72.0	22.0
4	90.0	87.96	62.0	6.0	62.0	86.0	50.0	80.0	28.0	87.0	39.0	57.0	87.0	57.0	82.0	5.0	38.0	82.0	13.0	44.0
5	90.0	89.05	62.0	2.0	62.0	87.0	49.0	84.0	30.0	89.0	37.0	59.0	89.0	59.0	80.0	3.0	37.0	80.0	8.0	42.0
6	90.0	89.13	65.0	2.0	65.0	88.0	52.0	85.0	30.0	89.0	36.0	59.0	89.0	59.0	79.0	2.0	36.0	79.0	7.0	41.0
7	90.0	89.13	65.0	2.0	65.0	88.0	52.0	85.0	30.0	89.0	36.0	59.0	89.0	59.0	79.0	2.0	36.0	79.0	7.0	41.0

iteration number	duration of green in seconds																	
	junction 1			junction 2		junction 3		junction 4				junction 5				junction 6	
	ϕ_{11}/s	ϕ_{21}/s	ϕ_{31}/s	ϕ_{12}/s	ϕ_{24}/s	ϕ_{13}/s	ϕ_{23}/s	ϕ_{14}/s	ϕ_{34}/s	ϕ_{34}/s	ϕ_{44}/s	ϕ_{54}/s	ϕ_{15}/s	ϕ_{25}/s	ϕ_{35}/s	ϕ_{45}/s	ϕ_{16}/s	ϕ_{26}/s
1	25.0	25.0	60.0	25.0	25.0	25.0	25.0	15.0	15.0	15.0	35.0	35.0	15.0	15.0	15.0	35.0	25.0	25.0
2	17.0	33.0	60.0	29.0	21.0	19.0	31.0	19.0	9.0	17.0	33.0	41.0	7.0	17.0	21.0	29.0	19.0	31.0
3	19.0	46.0	75.0	41.0	24.0	27.0	38.0	26.0	11.0	23.0	42.0	54.0	7.0	22.0	31.0	34.0	20.0	45.0
4	29.0	51.0	90.0	49.0	31.0	33.0	47.0	37.0	13.0	25.0	55.0	67.0	8.0	28.0	39.0	41.0	26.0	54.0
5	25.0	55.0	90.0	47.0	33.0	31.0	49.0	33.0	17.0	25.0	55.0	63.0	8.0	29.0	38.0	42.0	29.0	51.0
6	22.0	58.0	90.0	49.0	31.0	30.0	50.0	32.0	18.0	25.0	55.0	62.0	8.0	29.0	38.0	42.0	29.0	51.0
7	22.0	58.0	90.0	49.0	31.0	30.0	50.0	32.0	18.0	25.0	55.0	62.0	8.0	29.0	38.0	42.0	29.0	51.0

iteration number	link flow in veh/h															
	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9	q_{10}	q_{11}	q_{12}	q_{13}	q_{14}	q_{15}	q_{16}
1	950.0	220.0	950.0	840.0	880.0	150.0	230.0	250.0	90.0	260.0	460.0	290.0	450.0	740.0	790.0	610.0
2	955.0	215.0	955.0	845.0	885.0	150.0	225.0	250.0	85.0	260.0	470.0	280.0	450.0	730.0	780.0	600.0
3	960.0	220.0	960.0	850.0	900.0	150.0	220.0	230.0	100.0	240.0	480.0	270.0	450.0	820.0	860.0	550.0
4	960.0	220.0	960.0	850.0	900.0	150.0	220.0	230.0	100.0	240.0	460.0	290.0	450.0	800.0	840.0	650.0
5	880.0	300.0	880.0	770.0	820.0	150.0	300.0	310.0	100.0	320.0	460.0	290.0	450.0	770.0	810.0	580.0
6	860.0	320.0	860.0	750.0	800.0	150.0	320.0	320.0	100.0	340.0	460.0	290.0	450.0	770.0	810.0	530.0
7	860.0	320.0	860.0	750.0	800.0	150.0	320.0	320.0	100.0	340.0	460.0	290.0	450.0	770.0	810.0	530.0

iteration number	q_{17}	q_{18}	q_{19}	q_{20}	q_{21}	q_{22}	q_{23}
	460.0	350.0	680.0	1290.0	1050.0	1250.0	810.0
1	460.0	360.0	680.0	1290.0	1050.0	1250.0	810.0
2	380.0	280.0	740.0	1290.0	1050.0	1250.0	680.0
3	400.0	300.0	640.0	1290.0	1050.0	1250.0	800.0
4	430.0	330.0	710.0	1290.0	1050.0	1250.0	760.0
5	430.0	330.0	760.0	1290.0	1050.0	1250.0	710.0
6	430.0	330.0	760.0	1290.0	1050.0	1250.0	710.0
7	430.0	330.0	760.0	1290.0	1050.0	1250.0	710.0

final value of degree of saturatio	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
	0.46	0.31	0.49	0.70	0.64	0.29	0.55	0.37	0.66	0.77	0.65	0.26	0.74	0.72	0.57	0.75
	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}									
	0.78	0.60	0.79	0.83	0.78	0.61	0.64									
	avg	max														
	0.61	0.83														

Table 6.14 Results on Allsop&Charlesworth's network at 2nd initial signal settings for mutually consistent calculations

Iteartion number	cycle time (ξ)	PI veh-h/h	start of green in seconds																	
			junction 1			junction 2		junction 3		junction 4					junction 5				junction 6	
			θ_{11}/ξ	θ_{21}/ξ	θ_{31}/ξ	θ_{12}/ξ	θ_{22}/ξ	θ_{13}/ξ	θ_{23}/ξ	θ_{14}/ξ	θ_{24}/ξ	θ_{34}/ξ	θ_{44}/ξ	θ_{54}/ξ	θ_{15}/ξ	θ_{25}/ξ	θ_{35}/ξ	θ_{45}/ξ	θ_{16}/ξ	θ_{26}/ξ
1	75.0	661.04	0.0	32.5	0.0	0.0	32.5	0.0	42.5	0.0	20.0	55.0	0.0	55.0	0.0	25.0	50.0	0.0	0.0	42.5
2	75.0	216.12	11.0	41.0	11.0	0.0	38.0	71.0	26.0	74.0	23.0	49.0	74.0	49.0	64.0	2.0	32.0	64.0	3.0	40.0
3	90.0	134.04	16.0	49.0	16.0	85.0	39.0	85.0	30.0	5.0	40.0	66.0	5.0	66.0	77.0	89.0	36.0	77.0	4.0	46.0
4	105.0	114.93	56.0	84.0	56.0	102.0	55.0	101.0	36.0	4.0	45.0	75.0	4.0	75.0	88.0	101.0	39.0	88.0	7.0	56.0
5	105.0	104.38	0.0	22.0	0.0	103.0	57.0	101.0	37.0	5.0	49.0	72.0	5.0	72.0	90.0	104.0	44.0	90.0	9.0	57.0
6	105.0	121.37	5.0	24.0	5.0	1.0	57.0	0.0	40.0	62.0	93.0	10.0	62.0	10.0	88.0	101.0	45.0	88.0	5.0	58.0
7	105.0	120.97	5.0	24.0	5.0	2.0	60.0	0.0	40.0	62.0	93.0	10.0	62.0	10.0	88.0	101.0	45.0	88.0	5.0	58.0
8	105.0	105.17	104.0	23.0	104.0	102.0	55.0	101.0	37.0	4.0	50.0	72.0	4.0	72.0	90.0	104.0	44.0	90.0	8.0	56.0
9	105.0	105.17	104.0	23.0	104.0	102.0	55.0	101.0	37.0	4.0	50.0	72.0	4.0	72.0	90.0	104.0	44.0	90.0	8.0	56.0

Iteartion number	duration of green in seconds																	
	junction 1			junction 2		junction 3		junction 4					junction 5				junction 6	
	ϕ_{11}/ξ	ϕ_{21}/ξ	ϕ_{31}/ξ	ϕ_{12}/ξ	ϕ_{22}/ξ	ϕ_{13}/ξ	ϕ_{23}/ξ	ϕ_{14}/ξ	ϕ_{24}/ξ	ϕ_{34}/ξ	ϕ_{44}/ξ	ϕ_{54}/ξ	ϕ_{15}/ξ	ϕ_{25}/ξ	ϕ_{35}/ξ	ϕ_{45}/ξ	ϕ_{16}/ξ	ϕ_{26}/ξ
1	27.5	37.5	75.0	27.5	37.5	37.5	27.5	15.0	30.0	15.0	50.0	35.0	20.0	20.0	20.0	45.0	37.5	27.5
2	25.0	40.0	75.0	33.0	32.0	25.0	40.0	19.0	21.0	20.0	45.0	44.0	8.0	25.0	27.0	37.0	32.0	33.0
3	28.0	52.0	90.0	39.0	41.0	30.0	50.0	30.0	21.0	24.0	56.0	59.0	7.0	32.0	36.0	44.0	37.0	43.0
4	23.0	72.0	105.0	53.0	42.0	35.0	60.0	36.0	25.0	29.0	66.0	70.0	8.0	38.0	44.0	51.0	44.0	51.0
5	17.0	78.0	105.0	54.0	41.0	36.0	59.0	39.0	18.0	33.0	62.0	77.0	9.0	40.0	41.0	54.0	43.0	52.0
6	14.0	81.0	105.0	51.0	44.0	35.0	60.0	26.0	17.0	47.0	48.0	78.0	8.0	44.0	38.0	57.0	48.0	47.0
7	14.0	81.0	105.0	53.0	42.0	35.0	60.0	26.0	17.0	47.0	48.0	78.0	8.0	44.0	38.0	57.0	48.0	47.0
8	19.0	76.0	105.0	53.0	42.0	36.0	59.0	41.0	17.0	32.0	63.0	78.0	9.0	40.0	41.0	54.0	43.0	52.0
9	19.0	76.0	105.0	53.0	42.0	36.0	59.0	41.0	17.0	32.0	63.0	78.0	9.0	40.0	41.0	54.0	43.0	52.0

Iteartion number	link flow in veh/h															
	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9	q_{10}	q_{11}	q_{12}	q_{13}	q_{14}	q_{15}	q_{16}
1	810.0	370.0	810.0	770.0	700.0	140.0	470.0	350.0	210.0	500.0	380.0	370.0	450.0	640.0	700.0	720.0
2	820.0	360.0	820.0	780.0	710.0	140.0	460.0	350.0	200.0	500.0	390.0	360.0	450.0	650.0	710.0	720.0
3	870.0	310.0	870.0	810.0	750.0	140.0	310.0	250.0	120.0	400.0	380.0	370.0	450.0	590.0	560.0	620.0
4	870.0	310.0	870.0	810.0	750.0	140.0	310.0	250.0	120.0	400.0	360.0	390.0	450.0	620.0	590.0	470.0
5	880.0	300.0	880.0	840.0	870.0	180.0	300.0	220.0	120.0	240.0	360.0	390.0	450.0	620.0	590.0	330.0
6	880.0	300.0	880.0	840.0	870.0	180.0	300.0	220.0	120.0	240.0	300.0	450.0	450.0	520.0	490.0	280.0
7	900.0	280.0	900.0	840.0	870.0	180.0	280.0	220.0	120.0	240.0	300.0	450.0	450.0	520.0	490.0	280.0
8	900.0	280.0	900.0	840.0	870.0	180.0	280.0	220.0	120.0	240.0	400.0	350.0	450.0	620.0	590.0	380.0
9	900.0	280.0	900.0	840.0	870.0	180.0	280.0	220.0	120.0	240.0	400.0	350.0	450.0	620.0	590.0	380.0

Iteartion number							
	q_{17}	q_{18}	q_{19}	q_{20}	q_{21}	q_{22}	q_{23}
1	610.0	550.0	650.0	1290.0	1160.0	1250.0	1080.0
2	610.0	540.0	650.0	1290.0	1160.0	1250.0	1080.0
3	610.0	575.0	670.0	1290.0	1240.0	1250.0	1110.0
4	580.0	550.0	820.0	1290.0	1240.0	1250.0	930.0
5	580.0	550.0	960.0	1290.0	1110.0	1250.0	810.0
6	680.0	650.0	1010.0	1290.0	1110.0	1250.0	860.0
7	680.0	650.0	1010.0	1290.0	1110.0	1250.0	840.0
8	580.0	550.0	910.0	1290.0	1110.0	1250.0	840.0
9	580.0	550.0	910.0	1290.0	1110.0	1250.0	840.0

final value of degree of saturation

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
0.47	0.24	0.56	0.77	0.65	0.32	0.38	0.23	0.83	0.68	0.51	0.32	0.67	0.56	0.45	0.72

x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}
0.90	0.79	0.84	0.82	0.89	0.70	0.66

avg	max
0.61	0.90

Table 6.15 Results on Allsop & Charlesworth's network at 1st initial signal settings for re-run mutually consistent calculations

iteration number	cycle time	PI veh-h/h	start of green in seconds																
			junction 1			junction 2		junction 3		junction 4					junction 5				junction 6	
			θ_{11}/ξ	θ_{21}/ξ	θ_{31}/ξ	θ_{12}/ξ	θ_{22}/ξ	θ_{13}/ξ	θ_{23}/ξ	θ_{14}/ξ	θ_{24}/ξ	θ_{34}/ξ	θ_{44}/ξ	θ_{54}/ξ	θ_{15}/ξ	θ_{25}/ξ	θ_{35}/ξ	θ_{45}/ξ	θ_{16}/ξ	θ_{26}/ξ
1	124.0	93.78	7.0	82.0	7.0	0.0	69.0	120.0	50.0	4.0	65.0	87.0	4.0	87.0	9.0	27.0	63.0	9.0	36.0	73.0
2	124.0	94.81	34.0	117.0	34.0	1.0	71.0	120.0	51.0	2.0	63.0	85.0	2.0	85.0	10.0	32.0	61.0	10.0	3.0	29.0
3	124.0	93.60	34.0	117.0	34.0	0.0	70.0	120.0	51.0	2.0	63.0	85.0	2.0	85.0	10.0	32.0	60.0	10.0	3.0	28.0
4	124.0	94.10	35.0	123.0	35.0	4.0	75.0	120.0	52.0	4.0	65.0	87.0	4.0	87.0	13.0	36.0	62.0	13.0	4.0	28.0
5	124.0	94.57	36.0	2.0	36.0	4.0	75.0	121.0	53.0	5.0	66.0	88.0	5.0	88.0	14.0	34.0	62.0	14.0	2.0	28.0
6	124.0	94.53	33.0	1.0	33.0	2.0	72.0	119.0	51.0	2.0	63.0	85.0	2.0	85.0	13.0	34.0	61.0	13.0	122.0	26.0
7	124.0	94.53	33.0	1.0	33.0	2.0	72.0	119.0	51.0	2.0	63.0	85.0	2.0	85.0	13.0	34.0	61 0	13.0	122.0	26.0

iteration number	duration of green in seconds																
	junction 1			junction 2		junction 3		junction 4					junction 5				junction 6	
	ϕ_{11}/ξ	ϕ_{21}/ξ	ϕ_{31}/ξ	ϕ_{12}/ξ	ϕ_{22}/ξ	ϕ_{13}/ξ	ϕ_{23}/ξ	ϕ_{14}/ξ	ϕ_{24}/ξ	ϕ_{34}/ξ	ϕ_{44}/ξ	ϕ_{54}/ξ	ϕ_{15}/ξ	ϕ_{25}/ξ	ϕ_{35}/ξ	ϕ_{45}/ξ	ϕ_{16}/ξ	ϕ_{26}/ξ
1	70.0	44.0	124.0	64.0	50.0	49.0	65.0	56.0	17.0	36.0	78.0	97.0	13.0	31.0	65.0	49.0	32.0	82.0
2	78.0	36.0	124.0	65.0	49.0	50.0	64.0	56.0	17.0	36.0	78.0	97.0	17.0	24.0	68.0	46.0	21.0	93.0
3	78.0	36.0	124.0	65.0	49.0	50.0	64.0	56.0	17.0	36.0	78.0	97.0	17.0	23.0	69.0	45.0	20.0	94.0
4	83.0	31.0	124.0	66.0	48.0	51.0	63.0	56.0	17.0	36.0	78.0	97.0	18.0	21.0	70.0	44.0	19.0	95.0
5	85.0	29.0	124.0	66.0	48.0	51.0	63.0	56.0	17.0	36.0	78.0	97.0	15.0	23.0	71.0	43.0	21.0	93.0
6	87.0	27.0	124.0	66.0	48.0	51.0	63.0	56.0	17.0	36.0	78.0	97.0	16.0	22.0	71.0	43.0	23.0	91.0
7	87.0	27.0	124.0	66.0	48.0	51.0	63.0	56.0	17.0	36.0	78.0	97.0	16.0	22.0	71.0	43.0	23.0	91.0

iteration number	link flow in veh/h															
	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9	q_{10}	q_{11}	q_{12}	q_{13}	q_{14}	q_{15}	q_{16}
1	1061.0	119.0	1061.0	902.1	925.9	157.7	119.0	195.0	83.0	206.4	562.2	187.8	450.0	942.3	970.3	947.7
2	1060.0	120.0	1060.0	899.0	928.0	158.0	120.0	189.0	80.0	209.0	570.0	180.0	450.0	1030.0	1061.0	1000.0
3	1060.0	120.0	1060.0	900.0	930.0	160.0	120.0	190.0	80.0	210.0	570.0	180.0	450.0	1055.0	1085.0	1000.0
4	1095.0	85.0	1060.0	935.0	965.0	160.0	85.0	155.0	80.0	210.0	570.0	180.0	450.0	1055.0	1085.0	1015.0
5	1125.0	55.0	1090.0	965.0	965.0	160.0	55.0	155.0	50.0	210.0	570.0	180.0	450.0	1055.0	1085.0	1035.0
6	1125.0	55.0	1090.0	965.0	965.0	160.0	55.0	155.0	50.0	210.0	570.0	180.0	450.0	1055.0	1085.0	1075.0
7	1125.0	55.0	1090.0	965.0	965.0	160.0	55.0	155.0	50.0	210.0	570.0	180.0	450.0	1055.0	1085.0	1075.0

iteration number	q_{17}	q_{18}	q_{19}	q_{20}	q_{21}	q_{22}	q_{23}
1	257.7	169.7	342.3	1290.0	1080.4	1250.0	918.4
2	170.0	80.0	291.0	1290.0	1081.0	1250.0	879.0
3	145.0	55.0	290.0	1290.0	1080.0	1250.0	855.0
4	145.0	55.0	240.0	1290.0	1080.0	1250.0	905.0
5	145.0	85.0	220.0	1290.0	1080.0	1250.0	925.0
6	145.0	85.0	180.0	1290.0	1080.0	1250.0	965.0
7	145.0	85.0	180.0	1290.0	1080.0	1250.0	965.0

final value of degree of saturation

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
0.59	0.16	0.65	0.73	0.69	0.30	0.16	0.24	0.23	0.70	0.70	0.25	0.70	0.80	0.80	0.53

x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}
0.48	0.27	0.55	0.91	0.59	0.47	0.76

avg	max
0.53	0.91

Table 6.16 Results on Allsop & Charlesworth's network at 2nd initial signal settings for re-run mutually consistent calculations

iteration number	cycle time $\frac{1}{(\xi)}$	PI veh-h/h	start of green in seconds																
			junction 1			junction 2		junction 3		junction 4					junction 5				junction 6	
			θ_{11}/ξ	θ_{21}/ξ	θ_{31}/ξ	θ_{12}/ξ	θ_{22}/ξ	θ_{13}/ξ	θ_{23}/ξ	θ_{14}/ξ	θ_{24}/ξ	θ_{34}/ξ	θ_{44}/ξ	θ_{54}/ξ	θ_{15}/ξ	θ_{25}/ξ	θ_{35}/ξ	θ_{45}/ξ	θ_{16}/ξ	θ_{26}/ξ
1	128.0	100.77	38.0	125.0	38.0	3.0	71.0	126.0	54.0	4.0	67.0	90.0	4.0	90.0	13.0	40.0	66.0	13.0	123.0	31.0
2	128.0	100.87	37.0	127.0	37.0	4.0	71.0	126.0	54.0	5.0	69.0	91.0	5.0	91.0	14.0	40.0	66.0	14.0	123.0	31.0
3	128.0	99.72	37.0	0.0	37.0	4.0	73.0	126.0	55.0	4.0	69.0	90.0	4.0	90.0	14.0	43.0	65.0	14.0	125.0	31.0
4	128.0	101.68	35.0	3.0	35.0	4.0	72.0	126.0	55.0	3.0	68.0	89.0	3.0	89.0	16.0	44.0	66.0	16.0	125.0	34.0
5	128.0	100.10	34.0	6.0	34.0	0.0	70.0	123.0	53.0	1.0	66.0	87.0	1.0	87.0	21.0	33.0	59.0	21.0	34.0	76.0
6	128.0	99.58	34.0	10.0	34.0	1.0	71.0	123.0	53.0	2.0	67.0	87.0	2.0	87.0	22.0	34.0	59.0	22.0	33.0	77.0
7	128.0	99.58	34.0	10.0	34.0	1.0	71.0	123.0	53.0	2.0	67.0	87.0	2.0	87.0	22.0	34.0	59.0	22.0	33.0	77.0

iteration number	duration of green in seconds																	
	junction 1			junction 2		junction 3		junction 4					junction 5				junction 6	
	ϕ_{11}/ξ	ϕ_{21}/ξ	ϕ_{31}/ξ	ϕ_{12}/ξ	ϕ_{22}/ξ	ϕ_{13}/ξ	ϕ_{23}/ξ	ϕ_{14}/ξ	ϕ_{24}/ξ	ϕ_{34}/ξ	ϕ_{44}/ξ	ϕ_{54}/ξ	ϕ_{15}/ξ	ϕ_{25}/ξ	ϕ_{35}/ξ	ϕ_{45}/ξ	ϕ_{16}/ξ	ϕ_{26}/ξ
1	82.0	36.0	128.0	63.0	55.0	51.0	67.0	58.0	18.0	37.0	81.0	100.0	22.0	21.0	70.0	48.0	31.0	87.0
2	85.0	33.0	128.0	62.0	56.0	51.0	67.0	59.0	17.0	37.0	81.0	101.0	21.0	21.0	71.0	47.0	31.0	87.0
3	86.0	32.0	128.0	64.0	54.0	52.0	66.0	60.0	16.0	37.0	81.0	102.0	24.0	17.0	72.0	46.0	29.0	89.0
4	87.0	27.0	128.0	63.0	55.0	52.0	66.0	60.0	16.0	37.0	81.0	102.0	23.0	17.0	73.0	45.0	32.0	86.0
5	95.0	23.0	128.0	65.0	53.0	53.0	65.0	60.0	17.0	38.0	82.0	103.0	7.0	21.0	85.0	33.0	37.0	81.0
6	99.0	19.0	128.0	65.0	53.0	53.0	65.0	60.0	15.0	38.0	80.0	103.0	7.0	20.0	86.0	32.0	39.0	79.0
7	99.0	19.0	128.0	65.0	53.0	53.0	65.0	60.0	15.0	38.0	80.0	103.0	7.0	20.0	86.0	32.0	39.0	79.0

iteration number	link flow in veh/h															
	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9	q_{10}	q_{11}	q_{12}	q_{13}	q_{14}	q_{15}	q_{16}
1	1015.3	164.7	1015.3	917.6	945.1	144.1	177.4	133.5	148.3	207.6	621.7	128.3	450.0	1032.3	999.1	1023.3
2	1015.0	165.0	1015.0	930.0	955.0	145.0	170.0	125.0	140.0	200.0	620.0	130.0	450.0	1030.0	990.0	1055.0
3	1015.0	165.0	1015.0	930.0	955.0	145.0	170.0	125.0	140.0	200.0	660.0	90.0	450.0	1085.0	1045.0	1075.0
4	1015.0	165.0	1015.0	930.0	955.0	145.0	170.0	125.0	140.0	200.0	660.0	90.0	450.0	1085.0	1045.0	1155.0
5	1140.0	40.0	1140.0	1055.0	955.0	145.0	45.0	125.0	15.0	200.0	660.0	90.0	450.0	1085.0	1045.0	1155.0
6	1140.0	40.0	1140.0	1055.0	955.0	165.0	45.0	125.0	15.0	180.0	660.0	90.0	450.0	1085.0	1065.0	1175.0
7	1140.0	40.0	1140.0	1055.0	955.0	165.0	45.0	125.0	15.0	180.0	660.0	90.0	450.0	1085.0	1065.0	1175.0

iteration number	q_{17}	q_{18}	q_{19}	q_{20}	q_{21}	q_{22}	q_{23}
1	174.4	160.1	266.7	1290.0	1151.6	1250.0	1039.1
2	175.0	160.0	235.0	1290.0	1160.0	1250.0	1085.0
3	120.0	105.0	215.0	1290.0	1160.0	1250.0	1050.0
4	120.0	105.0	135.0	1290.0	1160.0	1250.0	1130.0
5	120.0	105.0	135.0	1290.0	1285.0	1250.0	1130.0
6	120.0	85.0	115.0	1290.0	1245.0	1250.0	1130.0
7	120.0	85.0	115.0	1290.0	1245.0	1250.0	1130.0

final value of degree of saturation

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
0.59	0.17	0.70	0.80	0.66	0.30	0.08	0.27	0.16	0.70	0.70	0.08	0.69	0.82	0.81	0.52

x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}
0.45	0.16	0.52	0.91	0.58	0.56	0.85

avg	max
0.53	0.91

6.5.2.4 Comparisons for mixed search procedure

In this sub-section, self-comparisons for the value of the performance index of the variants of the mixed search procedure are made on the basis of the results obtained in Section 6.5.1.1. Firstly, for the first set of initial signal settings, the difference in the final values of the performance index for the *method a* and *method b* is less than 7% while the maximum final values of degree of saturation for *method a* and *method b* are appreciably different: 0.88 and 0.83. The corresponding final signal settings and the equilibrium link flows are also different. Similarly, for the second set of initial signal settings, the difference in the final values of the performance index for the *method a* and *method b* is less than 0.5% while the maximum final values of degree of saturation for *method a* and *method b* are slightly different: 0.86 and 0.87, and the corresponding final signal settings and the equilibrium link flows are again different.

Secondly, for the system performance obtained by *method a*, the difference in the final values of the performance index starting from the first and second set of initial signal settings is less than 5% while the maximum final values of degree of saturation for the two sets of initial signal settings are slightly different: 0.88 and 0.86, and the corresponding final signal settings and the equilibrium link flows are also different. Similarly, for the system performance obtained by *method b*, the difference in the final values of the performance index starting from the first and second set of initial signal settings is less than 2% while the maximum final values of degree of saturation for the two sets of initial signal settings are appreciably different: 0.83 and 0.87, and the corresponding final signal settings and the equilibrium link flows are also different.

Thirdly, in comparison of the performance obtained by the two variants of the mixed search procedure in this numerical test: *method a* and *method b*, in the first initial signal settings we can find the performance obtained by *method b* is better by 7% than that obtained by *method a*, and in the second initial signal settings, the performance made by *method b* is slightly better by 0.5% than that obtained by *method a.*

6.5.2.5 Comparisons for mixed search and mutually consistent calculation

In this sub-section, a comparison of the values of performance index obtained by the mixed search and those did by the mutually consistent calculations is made as follows.

For the first set of initial signal settings, the system performance achieved by the *method b* of the mixed search variants is better by 56% than that achieved by the mutually consistent calculations; similarly for the second set of initial signal settings, the system performance achieved by the *method b* is better by 63% than that obtained by the mutually consistent calculations.

As for the computation efforts for the mixed search procedure performed on PC 486SX 25/33 Zenith machine, each iteration for this numerical example was performed in less than 30 seconds of CPU time in the C++ integrated development environment. Total computation efforts for complete run of the mixed search procedure did not exceed 10 minutes of CPU time on that machine.

6.6 Conclusions

In this Chapter, three kind of solution methods to the bi-level problem of area traffic control optimisation given in (6.0) have been briefly stated in Sections 6.1-6.3. Two example test road network have been given in Section 6.4. The calculation results for the solution methods illustrated on the two junctions based road network and Allsop and Charlesworth's example network have been given in Section 6.5. In the two junctions based road network, three kind of the solution methods showed good progress in decreasing the values of the indicators of the performance index by 90% approximately. Furthermore, Allsop and Charlesworth's example network has been used as an illustrative example for showing the effectiveness of the mixed search procedure, in which two distinct sets of signal timings were specified as initial values for the numerical test. In terms of resulting values of performance index for the whole network and degree of saturation for selected links, the robustness of the proposed method in solving the non-convex bi-level problem has been shown in that the resulting values of performance index were within negligible difference and none of the resulting values of the degree of saturation for selected links was over 90%. For example, the *method b* adopted the optimal search technique used by TRANSYT model and starting types I and II from a good localized solution given by type III of the mixed search procedure, has been shown in achieving the neighbourhood of a unknown global optimum with very close values of the performance index given by

substantially different initial signal timings. This can be shown in Tables 6.9-6.11 the initial difference in the values of performance index for the two sets of signal timings was as high as 32.54%, whilst the resulting values implemented by *method b* declined to 1.94%. Moreover, the proposed *method b* improved in the values of performance index by 57% and 63% approximately in comparison with those given by the non-optimising mutually consistent calculations. Also, when a re-run of mutually consistent calculations was carried out starting with the equilibrium flows and cycle time obtained from the *method b*, again the results have shown the non-optimising mutually consistent calculations result in substantially higher values of the performance indices than did the proposed *method b*.

Chapter 7 Conclusions and Recommendations for Future Research

7.0 Introduction

In this Chapter, comprehensive conclusions from Chapter 1 to Chapter 6 of this thesis are given in Section 7.1. Topics of interest for future research are identified in Section 7.2.

7.1 Conclusions

In Chapter 1, the general context for the combined problem of area traffic control optimisation and equilibrium network flows have been given in Section 1.1. The potential difficulties in dealing with this combined problem which have been indicated the use of a bi-level formulation were also pointed out. The objectives and structure for this thesis have been set out in Sections 1.2 and 1.3.

In Chapter 2, detailed reviews of literature relevant to this combined problem of area traffic control optimisation and equilibrium traffic assignment have been made in Sections 2.1 and 2.2. Because the equilibrium link flow is affected by the changes of the signal setting variables, such a combined problem for the area traffic control optimisation with equilibrium network flows can be regarded as a particular case of the equilibrium network design problem. Therefore, in Section 2.3, both the formulations and the corresponding solution methods for this particular equilibrium network design problem have been discussed. In particular, the bi-level programming technique which has been used for the equilibrium network design problem has been suggested as an appropriate tool in dealing with the combined problem for area traffic control optimisation with equilibrium network flows; in this bi-level programme the area traffic control optimisation has been regarded as the upper level problem whilst the user equilibrium traffic assignment has been regarded as the lower level problem. In Section 2.4, the literature concerning sensitivity analysis for the variational inequality problem for determination of the equilibrium network flows has been reviewed. The application of the sensitivity analysis technique to the area traffic control optimisation problem with fixed flows has also been reviewed.

In Chapter 3, the formulation of the bi-level programme for the area traffic control optimisation with equilibrium network flows has been discussed. The fundamental terminology and the corresponding notation used in the area traffic control optimisation and user equilibrium traffic assignment problems have been given respectively in Sections 3.2

and 3.3. Problem formulation for the upper level problem has been given in Section 3.4, in which explicit mathematical expressions for the objective function and the constraint set for the area traffic control optimisation with respect to the signal setting variables have been given. Problem formulation for the lower level problem has been given in Section 3.5, in which a general expression for the user equilibrium traffic assignment as a variational inequality problem has been used and the corresponding link travel time function has been defined as a sum of the undelayed travel time along the link and the average traffic delay which is incurred at the end of the link controlled by the downstream signals.

In Chapter 4, the sensitivity analysis for the bi-level problem has been discussed and extended. The derivatives for the chosen performance index for the area traffic control optimisation with respect to the signal setting variables have been given in Sections 4.1 to 4.5. In Section 4.1, the derivatives of the performance index with respect to the signal setting variables have been expressed as the sum of two components: one component arising directly from the immediate effects caused by the changes in the signal setting variables and the other component indirectly contributed by the reassignment process for equilibrium network flows caused mainly by the changes of the downstream controlling signal setting variables. In the following Sections 4.2-4.5, detailed expressions for each term of the derivatives of the performance index with respect to the signal setting variables have been adopted from previous work or newly derived. The derivatives for all links with respect to the common cycle time have been given in Section 4.2, and the derivatives for upstream links which are immediately controlled by the signal groups at each junction with respect to the start and duration of green time for the relevant group have been given in Section 4.3; furthermore the derivatives for the downstream and further downstream links with respect to the starts and durations of green times for the upstream links have been given in Sections 4.4 and 4.5 respectively.

In Chapter 5, a solution method for the bi-level problem of area traffic control optimisation with equilibrium flows has been discussed. For any given signal settings, a feasible descent direction in which the values of the performance index decreases has been decided by using the gradient projection method given in Section 5.2. In relation to the calculation of the gradients of the performance index for the bi-level problem, the results set out in Chapter 4 have been used. The way of finding a favourable step length in the descent direction at each iteration to reduce the value of the objective function has been set out in Section 5.3.

Since we have discussed in Chapter 3 that only local optimal solution can be obtained in this bi-level problem for the area traffic control optimisation due to the non-convexity of the objective function and of the constraint set, a method of searching for one of the better local optima to the bi-level problem has been given in Section 5.4.

In Chapter 6, two test road networks have been selected for the numerical calculations on the following three methods: Hooke & Jeeves' method, the calculation of mutually consistent signal settings and equilibrium flows and the mixed search optimisation procedure proposed in Chapter 5, each of which has been summarized in Sections 6.1-6.3 respectively. In Section 6.4, the following two test road networks have been described: a two-junction network which was modified from the network used to illustrate Braess's paradox, and one that was used by Allsop and Charlesworth (1977) consisting of 20 origin-destination pairs, 6 signal-controlled junctions and 23 traffic links. The calculation results for the two-junction network have shown that the three methods all achieved good decreases in the value of the performance index from its initial value.

For Allsop and Charlesworth's example network, two kind of solution methods were illustrated, in which two distinct sets of signal timings were specified as the initial values for the numerical test. In terms of the resulting values of performance index for the whole network and degree of saturation for selected links, the robustness of the mixed search procedure in solving the non-convex bi-level problem has been shown in that the resulting values of performance index were within negligible difference and none of the resulting values of the degree of saturation for selected links was over 90%. As was to be expected, values resulting from the proposed *method b* for the mixed search procedure did much better in terms of the values of performance index than did the mutually consistent calculations. Also, when a re-run of mutually consistent calculation was carried out starting with the equilibrium flows and cycle time obtained from the *method b*, again the non-optimising mutually consistent calculations resulted in substantially higher values of the performance indices than did the proposed *method b*. Therefore, these encouraging results confirm that, as expected, the mixed search procedure can lead to substantial improvements in performance over the mutually consistent calculations.

7.2 Recommendations for Future Research

In this thesis, a bi-level formulation for area traffic control with equilibrium network flows

has been established and the corresponding solution method has been proposed by means of the mixed search procedure to obtain a good local optimum for this bi-level problem; furthermore numerical calculations have been performed for two test road networks and the effectiveness of the proposed solution method has been compared with that of other feasible solution methods. To pave the way for future applications of the mixed search procedure for the bi-level problem of area traffic control with equilibrium network flows, the following recommendations are made for further research in this field.

1. Investigate further the search procedure in the mixed search method in terms of different sequences of steps of types I-II and type III, to find sequences by means of which good local optimum for the performance index for the bi-level problem can be achieved with fewer iterations.

2. Test the stability and robustness of the mixed search procedure in road networks with higher levels of traffic demand flows in order to improve the search procedure for congested road networks.

3. Explore the feasibility of applying the new method to the system-optimised network design problem in which the total network travel cost is to be minimised with respect to the signal setting variables and network flows by using the gradient projection method to decide the descent direction for any given signal settings and system-optimised link flows; furthermore, compare the optimal values for the performance index at system-optimised signal settings and link flows with those obtained by means of the bi-level formulation for the user-equilibrium network design problem in order to assess the scope for improving on the best conditions that can be obtained subject to user-equilibrium route choice.

4. Apply this bi-level formulation for area traffic control optimisation with equilibrium network flows to various practical-sized road networks in order to demonstrate the stability and robustness of the proposed solution method and to establish confidence in its practical use by traffic engineers in urban traffic management.

5. Investigate joint optimisation in which sequential time slices are considered together, so that the choice of decision variables within each timeslice should take into account the effects that carry over into the next timeslice through boundary values of queue-length. The problem formulated in Chapter 3 in terms of timeslices assumes that for a given specified

timeslice the signal settings is fixed; still further requirements of this problem formulation and optimisation resulting from changes in timings between consecutive time slices need to be take into account.

6. Investigate the non-separable equilibrium traffic assignment in the lower level problem, when a more realistic non-separable link cost function is considered for opposed turning traffic streams. In this thesis we have simply considered the use of separable traffic assignment in which the average delay at the downstream signal-controlled junction of each link is estimated as a function of the flow on that link only, whereas use of a non-separable link cost function is allowed this delay to be influenced by the traffic flows on other links as in the case for opposed turning traffic streams. We need to re-consider the corresponding formulation and solution methods for this non-separable traffic assignment problem in the bi-level formulation.

References

Abdulaal, M. and LeBlanc, L.J. (1979) Continuous equilibrium network design models. *Transportation Research* **13 B**(1), 19-32.

Allsop, R.E. (1971) Delay-minimising settings for fixed-time traffic signals at a single road junction, *Journal of the Institute of Mathematics and its Applications*, **8**(2), 164-185.

Allsop, R.E. (1974) Some possibilities of using traffic control to influence trip distribution and route choice. In: *Transportation and Traffic Theory (Proceedings of the 6th International Symposium on Transportation and Traffic Theory)*, Elsevier, Amsterdam, 1974, 345-374.

Allsop, R. E. and J. A. Charlesworth (1977), Traffic in a signal-controlled road network: an example of different signal timings inducing different routeings, *Traffic Engineering Control,* **18**(5), 262-264.

Allsop, R.E. (1992) Evolving application of mathematical optimisation in design and operation of individual signal-controlled road junctions. *Mathematics in Transport Planning and Control.* (Griffths J.D. Ed), Oxford University Press. Oxford, 1-25.

Allsop, R. E. (1993) Derivation and Use of Sheared Expressions for Queue Length and Delay," lecture note T8(5) of the Intercollegiate MSc Course in Transport, University College London (Unpublished).

Allsop, R.E. (1995) Representation of road junctions by sets of nodes and links so that traffic assignment models can represent effects of traffic management measures on traffic route choice, lecture note T8(19) of the Intercollegiate MSc Course in Transport, University College London (Unpublished).

Allsop, R.E. (1996) Calculation of signal timings for coordinated signal control, lecture note T8(16) of the Intercollegiate MSc Course in Transport, University College London (Unpublished).

Allsop, R.E. (1997) Private communications.

Baxandall, P. and H. Liebeck (1986) *Vector Calculus*, Clarendon Press, Oxford.

Bazaraa, M.S. , H.D. Sherali and C.M. Shetty (1993) *Nonlinear Programming Theory and Algorithms*. 2nd edition, John Wiley & Sons, New York.

Beckmann, M.J., McGuire, C.B. and Winsten C.B. (1956) *Studies in Economics of Transportation.* Yale University Press, New Haven.

Charlesworth, J.A. (1977) The calculation of mutually consistent signal settings and traffic assignment for a signal-controlled road network. In*: the Proceedings of the 7th International Symposium on Transportation and Traffic Theory,* Elsevier, Kyoto, 1977, 545-569.

Chiou, S-W (1997 a) Optimisation of area traffic control subject to user equilibrium traffic assignment, *25th European Transport Forum, Proceedings of Seminar F, Transportation Planning Methods: Volume II,* PTRC, London, 53-64.

Chiou, S-W. (1997 b) Derivatives in an area traffic control problem subject to user equilibrium traffic assignment, *Universities Transport Study Group 29th Annual Conference,* Bournemouth University, January 1997. (Unpublished)

Dafermos, S.C. (1980) Traffic equilibrium and variational inequalities, *Transportation Science,* **14**(1) , 42-54.

Dickson, T. J. (1981) A note on traffic assignment and signal timings in a signal-controlled road network, *Transportation Research,* **15 B**(4), 267-271.

Fiacco, A.V. (1976) Sensitivity analysis for nonlinear programming using penalty methods, *Mathematical Programming*, **10**, 287-311.

Fiacco, A.V. (1983) *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming,* Academic Press, New York, 1983.

Friesz, T. L., R. L. Tobin, H-J. Cho and N. J. Mehta (1990) Sensitivity analysis based heuristic algorithms for mathematical programs with variational inequality constraints. *Mathematical Programming*, **48**, 265-284.

Gallivan, S and B.G. Heydecker (1988) Optimising the control performance of traffic signals at a single junction, *Transportation Research*, **22B** (5), 357-370.

Gartner, N. (1976), Area traffic control and network equilibrium. *In Proceedings of the International Symposium on Traffic Equilibrium Methods, held at the Universite de Montreal, November, 1974. Lecture Notes in Economics and Mathematical Systems*, **118.** Ed. (M.A. Florian) Springer-Verlag, Berlin, 274-297.

Gershwin, S. B. and H. N. Tan (1979) 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.

Heydecker, B.G. (1986) The equilibrium network design problem: a critical review. Presented to *the NATO Advanced Research Workshop on Flow Control of Congested Networks*, Capri, Italy, 12-18 October, 1986.

Heydecker, B.G. and Dudgeon, I.W. (1987) Calculation of signal settings to minimise delay at a junction. In: *the Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, MIT, July 1987, Elsevier. New York, 159-178

Heydecker, B. G. and T. K. Khoo (1990) The equilibrium network design problem, *Proceedings of AIRO'90 Conference on Models and Methods for Decision Support,* Sorrento, 587-602.

Heydecker, B.G. (1996) A decomposed approach for signal optimisation in road networks, *Transportation Research* 30 **B** (2), 99-114.

Hooke, R and T.A. Jeeves (1961) Direction search solution of numerical and statistical problems, *Journal of Association Computer Machinery*, **8**, 212-229.

Kimber, R.M. and E.M. Hollis (1979) Traffic queues and delays at road junctions, *TRRL Report* , **LR 909**, Transportation and Road Research Laboratory, Crowthrone.

Luenberger, D. G. (1989) *Linear and Nonlinear Programming*. 2nd edition, Addison-Wesley, Massachusetts.

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

Marcotte, P. and Marquis, G. (1992) Efficient implementation of heuristics for the continuous network design problem. *Annals of Operations Research* **34**, 163-176.

Murchland, J.D. (1970) Braess' paradox of traffic flow. *Transportation Research* **4**(4), 391-394.

Pas, E.I. and S.L. Principio (1997) Braess' paradox: some new insights, *Transportation Research* **31 B** (3), 265-276.

Robertson, D.I. (1969) TRANSYT: a traffic network study tool, *RRL Report*, **LR 253,** Transport and Road Research Laboratory, Crowthrone.

Sheffi, Y. (1985) *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods.* MIT. Prentice-Hall, Inc. New Jersey.

Silcock, J.P., and Sang, A.P. (1989) SIGSIGN: a phase-based optimisation program for individual signal-controlled junctions, *Traffic Engineering and Control*, **31**(5), 291-298.

Silcock, J.P. (1992) Phase-based optimisation of isolated signal-controlled junctions: sensitivity analysis and a treatment of double green periods. *Mathematics in Transport Planning and Control.* (Griffths J.D. Ed) Oxford University Press. Oxford, 45-57.

Silcock, J.P. (1997) Designing signal-controlled junctions for group-based operation. *Transportation Research* **31 A**(2), 157-173.

Smith, M.J. (1979) The existence, uniqueness and stability of traffic equilibria, *Transportation Research,* **13 B**(4), 295-304.

Suh, S. and Kim, T.J. (1992) Solving nonlinear bilevel programming models of the equilibrium network design problem: a comparative review. *Annals of Operations Research* **34,** 203-218.

Suwansirikul, C., Friesz, T.L. and Tobin, R.L. (1987) Equilibrium decomposed optimization: a heuristic for the continuous equilibrium network design problem. *Transportation Science* **21**(4), 254-263.

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

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

Van Vliet, D. (1987) The frank-wolfe algorithm for equilibrium traffic assignment viewed as a variational inequality, *Transportation Research* **21 B** (1) 87-89.

Vincent, R.A., A.I. Mitchell and D.I. Robertson (1980) User guide to TRANSYT version 8, *TRRL Report*, **LR888**, Transport and Road Research Laboratory, Crowthrone.

Wardrop, J.G. (1952) Some theoretical aspects of road research. *Proc. Inst. Civ. Engineers (Part II),* **1** (2), 325-362.

Wong, S.C. (1993) *Phase-based Optimisation of Signal Timings for Area Traffic Control.* Ph.D. Thesis, University of London.

Wong, S.C. (1995) Derivatives of the performance index for the traffic model from TRANSYT. *Transportation Research,* **29 B** (5), 303-327.

Wong, S.C. (1996), Group-based optimisation of signal timings using the TRANSYT traffic model, *Transportation Research* **30 B**(3), 217-244.

Yang, H and Yagar, S. (1994) Traffic assignment and traffic control in general freeway-arterial corridor systems. *Transportation Research* **28B** (6), 463-486.

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

Yang, H (1997) Sensitivity analysis for the elastic-demand network equilibrium problem with applications, *Transportation Research* **31 B**(1) , 55-70.

Appendix A Mathematical Preliminaries

<u>*<Definition a.1>*</u> <u>*<The Fitted Function>*</u>

Let two non-decreasing continuous functions f_1 *and* f_2 *on the positive half of the real line be fitted for each other if and only if for some interval is contained in the positive half of the real line the integral of* f_1 *is equal to that of* f_2 *on the interval.*

That is, let two non-decreasing continuous functions $f_1 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ *and* $f_2 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ *be fitted for each other if and only if there exists an interval* $[\, x_1 \,,\, x_2 \,] \subset \mathbb{R}^+$ *such that* $\int_{x_1}^{x_2} f_1(w)dw = \int_{x_1}^{x_2} f_2(w)dw$.

Then f_1 *is described as a fitted function for* f_2 *on interval* $[\, x_1 \,,\, x_2 \,]$ *and* f_2 *is described as a fitted function for* f_1 *on interval* $[\, x_1 \,,\, x_2 \,]$ *as well.* ■

<u>*<Proposition a.2>*</u>

Let f *be a non-decreasing continuous function defined as* $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ *and* g *be a non-decreasing continuous mapping defined as* $g : f(\mathbb{R}^+) \rightarrow \mathbb{R}^+$ *. For any interval* $[\, x_1 \,,\, x_2 \,] \subseteq \mathbb{R}^+$ *such that*

$$f : [\, x_1 \,,\, x_2 \,] \rightarrow f\,(\,[\, x_1 \,,\, x_2 \,]\,)$$

$$g : f\,(\,[\, x_1 \,,\, x_2 \,]\,) \rightarrow [\, x_1 \,,\, x_2 \,]$$

it follows that

$$\int_{x_1}^{x_2} f(w)\ dw = x_2\, f(x_2) - x_1\, f(x_1) - \int_{f(x_1)}^{f(x_2)} g(w)\ dw \ .$$

<proof>

For any function f that is integrable over the interval $[\, x_1 \,,\, x_2 \,]$,

$$\int_{x_1}^{x_2} f(w)\ dw = \int_{0}^{x_2} f(w)\ dw - \int_{0}^{x_1} f(w)\ dw \tag{a.1}$$

As in expression (a.1) by the given conditions on the function f and mapping g ,

$\int_{0}^{x_2} f(w)\ dw$, $\int_{0}^{x_1} f(w)\ dw$ can be expressed as follows.

$$\int_{0}^{x_2} f(w)dw = x_2\, f(x_2) - \int_{0}^{f(x_2)} g(w)dw \tag{a.2a}$$

$$\int_{0}^{x_1} f(w)dw = x_1\, f(x_1) - \int_{0}^{f(x_1)} g(w)dw \tag{a.2b}$$

We complete this proof by substituting expressions (a.2a) and (a.2b) into (a.1). ■

<Proposition a.3>

Let $f_1 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ *and* $f_2 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ *be two non-decreasing continuous functions fitted for each other on interval* $[\, x_1 \,,\, x_2 \,]$ *in the positive half of the real line.*

Let $g_1 : f_1(\mathbb{R}^+) \rightarrow \mathbb{R}^+$ *and* $g_2 : f_2(\mathbb{R}^+) \rightarrow \mathbb{R}^+$ *be two non-decreasing continuous mappings such that the conditions on* g_1 *and* g_2 *correspond to the conditions in <Proposition a.2>. Then if there exists an interval* $[\, \hat{x}_1 \,,\, \hat{x}_2 \,] \subseteq \mathbb{R}^+$ *such that* $f_1(x_1) = f_2(\hat{x}_1)$ *and* $f_1(x_2) = f_2(\hat{x}_2)$ *, so that*

$g_2 : [\, f_1(x_1) \,,\, f_1(x_2) \,] \rightarrow [\, \hat{x}_1 \,,\, \hat{x}_2 \,]$

and if also

$$[f_1(x_2)\ \hat{x}_2 - \int_{x_2}^{\hat{x}_2} f_2(w)\ dw] - [f_1(x_1)\ \hat{x}_1 - \int_{x_1}^{\hat{x}_1} f_2(w)\ dw] = f_1(x_2)\ x_2 - f_1(x_1)\ x_1$$

then

$$\int_{f_1(x_1)}^{f_1(x_2)} g_1(w)\ dw = \int_{f_1(x_1)}^{f_1(x_2)} g_2(w)\ dw\ .$$

<proof>

For the interval $[\, x_1 \,,\, x_2 \,] \subseteq \mathbb{R}^+$, by definition we have

$$\begin{aligned} &f_1 : [\, x_1 \,,\, x_2 \,] \rightarrow f_1\ (\, [\, x_1 \,,\, x_2 \,]\,) \\ &f_2 : [\, x_1 \,,\, x_2 \,] \rightarrow f_2\ (\, [\, x_1 \,,\, x_2 \,]\,) \\ &g_1 : f_1\ (\, [\, x_1 \,,\, x_2 \,]\,) \rightarrow [\, x_1 \,,\, x_2 \,] \\ &g_2 : f_2\ (\, [\, x_1 \,,\, x_2 \,]\,) \rightarrow [\, x_1 \,,\, x_2 \,] \end{aligned} \tag{a.3}$$

As f_1 *and* f_2 are fitted functions on $[\, x_1 \,,\, x_2 \,] \subseteq \mathbb{R}^+$, it follows that

$$\int_{x_1}^{x_2} f_1(w)\ dw = \int_{x_1}^{x_2} f_2(w)\ dw \tag{a.4}$$

Applying the result of *<Proposition a.2>* to the two cases of (a.4) gives

$$\int_{x_1}^{x_2} f_1(w)\ dw = x_2\, f_1(x_2) - x_1\, f_1(x_1) - \int_{f_1(x_1)}^{f_1(x_2)} g_1(w)\ dw \tag{a.5}$$

$$\int_{x_1}^{x_2} f_2(w)\ dw = x_2\, f_2(x_2) - x_1\, f_2(x_1) - \int_{f_2(x_1)}^{f_2(x_2)} g_2(w)\ dw \tag{a.6}$$

In expression (a.6), $\int_{f_2(x_1)}^{f_2(x_2)} g_2(w)\ dw$ can be rearranged as follows.

$$\int_{f_2(x_1)}^{f_2(x_2)} g_2(w)\ dw = \int_{f_2(x_1)}^{f_1(x_1)} g_2(w)\ dw + \int_{f_1(x_1)}^{f_1(x_2)} g_2(w)\ dw + \int_{f_1(x_2)}^{f_2(x_2)} g_2(w)\ dw$$

$$= \int_{f_2(x_1)}^{f_1(x_1)} g_2(w)\ dw + \int_{f_1(x_2)}^{f_2(x_2)} g_2(w)\ dw + \int_{f_1(x_1)}^{f_1(x_2)} g_2(w)\ dw$$

$$= [\ f_1(x_1)\ g_2(f_1(x_1)) - f_2(x_1)\ g_2(f_2(x_1)) - \int_{g_2(f_2(x_1))}^{g_2(f_1(x_1))} f_2(w)\ dw] +$$

$$[f_2(x_2)\ g_2(f_2(x_2)) - f_1(x_2)\ g_2(f_1(x_2)) - \int_{g_2(f_1(x_2))}^{g_2(f_2(x_2))} f_2(w)\ dw] +$$

$$\int_{f_1(x_1)}^{f_1(x_2)} g_2(w)\ dw \qquad \text{(a.6a)}$$

(by applying the result of <*Proposition a.2*> again)

And by (a.3) and the given fact that $g_2 : [\ f_1(x_1)\ ,\ f_1(x_2)\ ] \rightarrow [\ \hat{x}_1\ ,\ \hat{x}_2\ ]$

the expression (a.6a) can be rewritten as

$$\int_{f_2(x_1)}^{f_2(x_2)} g_2(w)dw = [f_1(x_1)\hat{x}_1 - f_2(x_1)x_1 - \int_{x_1}^{\hat{x}_1} f_2(w)dw] + [f_2(x_2)x_2 - f_1(x_2)\hat{x}_2 - \int_{\hat{x}_2}^{x_2} f_2(w)dw] +$$

$$\int_{f_1(x_1)}^{f_1(x_2)} g_2(w)dw \qquad \text{(a.7)}$$

Now expression (a.6) can be reexpressed using the expression (a.7) as

$$\int_{x_1}^{x_2} f_2(w)dw = [f_1(x_2)\hat{x}_2 - \int_{x_2}^{\hat{x}_2} f_2(w)dw] - [f_1(x_1)\hat{x}_1 - \int_{x_1}^{\hat{x}_1} f_2(w)dw] - \int_{f_1(x_1)}^{f_1(x_2)} g_2(w)dw \qquad \text{(a.8)}$$

Therefore by expressions (a.4) and (a.5) and the given fact that

$$[f_1(x_2)\ \hat{x}_2 - \int_{x_2}^{\hat{x}_2} f_2(w)\ dw] - [f_1(x_1)\ \hat{x}_1 - \int_{x_1}^{\hat{x}_1} f_2(w)\ dw] = f_1(x_2)\ x_2 - f_1(x_1)\ x_1$$

it follows that $\int_{f_1(x_1)}^{f_1(x_2)} g_1(w)\ dw = \int_{f_1(x_1)}^{f_1(x_2)} g_2(w)\ dw$. ■

Then results will now be applied for any link a to the functions $\bar{I}_a(t)$ *and* $\bar{E}_a(t)$ defined in Section 3.3.3 and 3.4.3.2 respectively over the interval $\boldsymbol{T}_a$ defined in Section 3.4.3.2, making use also of the facts that $\bar{I}_a(t)$ is integral of a cyclic function and $\boldsymbol{T}_a$ is an integer number of the period.

$\bar{I}_a(t)$ *and* $\bar{E}_a(t)$ are two non-decreasing continuous fitted functions over $\boldsymbol{T}_a$ such that

$$\int_{t_a^0}^{t_a^n} \bar{I}_a(t)\ dt = \int_{t_a^0}^{t_a^n} \bar{E}_a(t)\ dt \qquad \text{(a.9)}$$

and $\bar{E}_a(t)$ is a linear function with the slope q_a on $\boldsymbol{T}_a$, where

$$q_a = \frac{\bar{I}_a(t_a^n) - \bar{I}_a(t_a^0)}{t_a^n - t_a^0}$$

let

$$\bar{I}_a^{-1} : \bar{I}_a\left(\left[t_a^0, t_a^n\right]\right) \rightarrow \left[t_a^0, t_a^n\right]$$

$$\bar{E}_a^{-1} : \bar{E}_a\left(\left[t_a^0, t_a^n\right]\right) \rightarrow \left[t_a^0, t_a^n\right]$$

be two non-decreasing continuous mappings onto $\boldsymbol{T}_a$.

Let $\hat{t}_a^0$ *and* $\hat{t}_a^n$ be the solutions of the linear equations

$$\bar{E}_a(\hat{t}_a^0) = \bar{I}_a(t_a^0)$$

$$\bar{E}_a(\hat{t}_a^n) = \bar{I}_a(t_a^n)$$

Then we have $\bar{E}_a^{-1} : [\ \bar{I}_a(t_a^0)\ ,\ \bar{I}_a(t_a^n)\ ] \rightarrow [\ \hat{t}_a^0\ ,\ \hat{t}_a^n\ ]$ (a.10)

<u>*<Lemma a.4>*</u>

if $(t_a^n - \hat{t}_a^n)\ (\bar{E}_a(t_a^n) - \bar{I}_a(t_a^n)) = (t_a^0 - \hat{t}_a^0)\ (\bar{E}_a(t_a^0) - \bar{I}_a(t_a^0))$

$$\textit{then} \int_{\bar{I}_a(t_a^0)}^{\bar{I}_a(t_a^n)} \bar{I}_a^{-1}(w)\ dw = \int_{\bar{I}_a(t_a^0)}^{\bar{I}_a(t_a^n)} \bar{E}_a^{-1}(w)\ dw$$

<proof>

Applying *<Proposition a.3>* to the fitted functions $\bar{I}_a(t)$ *and* $\bar{E}_a(t)$ on interval $\boldsymbol{T}_a$,

if the condition

$$[\bar{I}_a(t_a^n)\ \hat{t}_a^n - \int_{t_a^n}^{\hat{t}_a^n} \bar{E}_a(t)dt] - [\bar{I}_a(t_a^0)\ \hat{t}_a^0 - \int_{t_a^0}^{\hat{t}_a^0} \bar{E}_a(t)dt] = \bar{I}_a(t_a^n)\ t_a^n - \bar{I}_a(t_a^0)\ t_a^0 \tag{a.11}$$

holds then

$$\int_{\bar{I}_a(t_a^0)}^{\bar{I}_a(t_a^n)} \bar{I}_a^{-1}(w)\ dw = \int_{\bar{I}_a(t_a^0)}^{\bar{I}_a(t_a^n)} \bar{E}_a^{-1}(w)\ dw$$

Since $\bar{E}_a(t) = \bar{E}_a(t_a^0) + q_a(t - t_a^0)$, it follows

$$\int_{t_a^0}^{\hat{t}_a^0} \bar{E}_a(t)\, dt = \frac{(\hat{t}_a^0 - t_a^0)\,(\bar{E}_a(\hat{t}_a^0) + \bar{E}_a(t_a^0))}{2}$$

$$\int_{t_a^n}^{\hat{t}_a^n} \bar{E}_a(t)\, dt = \frac{(\hat{t}_a^n - t_a^n)\,(\bar{E}_a(\hat{t}_a^n) + \bar{E}_a(t_a^n))}{2}$$

And since also

$$\bar{E}_a(\hat{t}_a^0) = \bar{I}_a(t_a^0)$$

$$\bar{E}_a(\hat{t}_a^n) = \bar{I}_a(t_a^n)$$

the condition (a.11) can be rearranged as

$$[\bar{I}_a(t_a^n)t_a^n + \frac{(t_a^n - \hat{t}_a^n)(\bar{E}_a(t_a^n) - \bar{I}_a(t_a^n))}{2}] - [\bar{I}_a(t_a^0)t_a^0 + \frac{(t_a^0 - \hat{t}_a^0)(\bar{E}_a(t_a^0) - \bar{I}_a(t_a^0))}{2}]$$

$$= \bar{I}_a(t_a^n)t_a^n - \bar{I}_a(t_a^0)t_a^0$$

and it follows immediately that if

$$(t_a^n - \hat{t}_a^n)\,(\bar{E}_a(t_a^n) - \bar{I}_a(t_a^n)) = (t_a^0 - \hat{t}_a^0)\,(\bar{E}_a(t_a^0) - \bar{I}_a(t_a^0))$$

then condition (a.11) holds and hence by <*Proposition a.3*>, we have

$$\int_{\bar{I}_a(t_a^0)}^{\bar{I}_a(t_a^n)} \bar{I}_a^{-1}(w)\, dw = \int_{\bar{I}_a(t_a^0)}^{\bar{I}_a(t_a^n)} \bar{E}_a^{-1}(w)\, dw \qquad \blacksquare$$

<u><*Definition a.5*><*Cyclic Accumulative Function*></u>

F is a cyclic accumulative function with period c on the positive half of the real line $F : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ if and only if for given c *and any* x_1 *and* $x \in \mathbb{R}^+$ and any $n \in \mathbb{N}$; it follows $F(x + n\,c) = F(x) + n\,Q$

where $Q = F(x_1 + c) - F(x_1)$ $\blacksquare$

By definition $\bar{I}_a(t)$ is a *cyclic accumulative function* on $\mathbb{R}^+$, with period $c = \frac{1}{\zeta}$

and $Q = q_a\, c$. Also by definition of $\boldsymbol{T}_a$, $t_a^n - t_a^0 = n\, c$

<u>*<Lemma a.6>*</u>

$$\int_{\bar{I}_a(t_a^0)}^{\bar{I}_a(t_a^n)} \bar{I}_a^{-1}(w)\, dw = \int_{\bar{I}_a(t_a^0)}^{\bar{I}_a(t_a^n)} \bar{E}_a^{-1}(w)\, dw$$

<proof>

$$\bar{I}_a(t_a^n) = \bar{I}_a(t_a^0) + q_a(t_a^n - t_a^0)$$

$$\bar{E}_a(t_a^n) = \bar{E}_a(t_a^0) + q_a(t_a^n - t_a^0)$$

Hence $\bar{E}_a(t_a^n) - \bar{E}_a(t_a^0) = \bar{I}_a(t_a^n) - \bar{I}_a(t_a^0)$ (a.12)

Since

$$\bar{E}_a(\hat{t}_a^0) = \bar{I}_a(t_a^0)$$

$$\bar{E}_a(\hat{t}_a^n) = \bar{I}_a(t_a^n)$$

it follows

$$\bar{E}_a(\hat{t}_a^n) = \bar{I}_a(t_a^n) = \bar{I}_a(t_a^0 + n\, c) = \bar{I}_a(t_a^0) + n\, c\, q_a$$

$$\bar{E}_a(\hat{t}_a^n) = \bar{E}_a(\hat{t}_a^0) + n\, c\, q_a$$

but $\bar{E}_a(\hat{t}_a^n) = \bar{E}_a(\hat{t}_a^0) + q_a(\hat{t}_a^n - \hat{t}_a^0)$

Hence $\hat{t}_a^n - \hat{t}_a^0 = nc = t_a^n - t_a^0$ (a.13)

From (a.12) and (a.13) the condition of *<Lemma a.4>* is satisfied and *<Lemma a.6>* is complete. ■

Appendix B Derivatives

B.0 Introduction

For a change in the signal timings, using the sensitivity analysis approach in terms of the derivatives can help us to estimate the corresponding changes on the indicators of traffic conditions. In Chapter 4, the results for the derivatives of the indicators of traffic conditions have been given. In this Appendix, the details of the calculations for the derivatives shown in Chapter 4 are given.

For a particular signal group j at the junction m, where for a link a which is controlled by the signal group j at that junction, following the notation defined in Chapter 3, the relationships between the signal setting variables can be expressed as follows.

Let $a \in \boldsymbol{L}$, $j \in \boldsymbol{P}_m$, $m \in \boldsymbol{N}$ be such that $\hat{b}_{ajm} = 1$, then the start and duration of effective green time therefore can be described as

$$\frac{1}{\zeta}\,\bar{\theta}_a = \sum_{j=1}^{N_{pm}} \hat{b}_{ajm}\,(\frac{1}{\zeta}\,\theta_{jm}) + e_1 \tag{b.1a}$$

$$\frac{1}{\zeta}\,\Lambda_a = \sum_{j=1}^{N_{pm}} \hat{b}_{ajm}\,(\frac{1}{\zeta}\,\phi_{jm}) + e \tag{b.1b}$$

where $\bar{\theta}_a$, Λ_a are respectively the effective start and duration of green times given in proportion of the common cycle time, e_1 , e are respectively the green start lag and the extra effective green for link a and θ_{jm} , ϕ_{jm} are respectively the actual start and duration of green times given in the proportion of the common cycle time.

For a change in the common cycle time, it can be represented in terms of $\frac{\partial P}{\partial \zeta}$, where ζ is the reciprocal common cycle time; for a change in the start of actual

green time, the corresponding changes on the performance index P is determined by

$$\frac{\partial P}{\partial\,(\frac{1}{\zeta}\,\theta_{jm})} \tag{b.2a}$$

and for a change in the duration of actual green time, the corresponding changes on the performance index P is determined by

$$\frac{\partial P}{\partial\,(\frac{1}{\zeta}\,\phi_{jm})} \tag{b.2b}$$

Applying expressions (b.1a) to (b.2a) and (b.1b) to (b.2b) respectively, the corresponding changes in the performance index P caused by the changes in the actual start and duration of green times can be seen to be equal to those caused by the changes in the start and duration of effective green time when the green start lag and extra effective green time are constant. That is:

$$\frac{\partial P}{\partial\,(\frac{1}{\zeta}\,\theta_{jm})} = \frac{\partial P}{\partial\,(\frac{1}{\zeta}\,\bar{\theta}_a)}$$

$$\frac{\partial P}{\partial\,(\frac{1}{\zeta}\,\phi_{jm})} = \frac{\partial P}{\partial\,(\frac{1}{\zeta}\,\Lambda_a)}$$

Accordingly, we have

$$\frac{\partial P}{\partial\,\theta_{jm}} = \frac{\partial P}{\partial\,\bar{\theta}_a} \tag{b.3a}$$

$$\frac{\partial P}{\partial\,\phi_{jm}} = \frac{\partial P}{\partial\,\Lambda_a} \tag{b.3b}$$

It follows that for given green start lag and extra effective green time e_1 , e for all links a in L which are controlled by the corresponding signal groups j at the

corresponding junctions m, $\forall j \in \boldsymbol{P}_m$, $\forall m \in \boldsymbol{N}$, the corresponding changes in the performance index P caused by the changes in the signal timings, i.e. the common cycle time, the start and duration of green time are,

$$\frac{\partial P}{\partial \zeta}, \frac{\partial P}{\partial \theta_{jm}}, \frac{\partial P}{\partial \phi_{jm}}, \quad \forall j \in \boldsymbol{P}_m, \quad \forall m \in \boldsymbol{N}$$

can be represented as

$$\frac{\partial P}{\partial \zeta}, \frac{\partial P}{\partial \bar{\theta}_a}, \frac{\partial P}{\partial \Lambda_a}, \quad \forall a \in \boldsymbol{L} .$$

In the following sections, specifications for the calculations of

$$\frac{\partial P}{\partial \zeta}, \frac{\partial P}{\partial \bar{\theta}_a}, \frac{\partial P}{\partial \Lambda_a}, \quad \forall a \in \boldsymbol{L}$$

will be discussed.

B.1 Derivatives for All Links

Derivatives of the indicators of traffic conditions with respect to ζ can be obtained in the following way.

For the indicator $X_a = X_0(\zeta , \Lambda_a(\zeta))$, $\forall \ a \in \boldsymbol{L}$, the derivatives with respect to

$$\zeta \text{ are } \frac{dX_a}{d\zeta} = \frac{\partial X_0}{\partial \zeta} + \frac{\partial X_0}{\partial \Lambda_a}\frac{d\Lambda_a}{d\zeta} \tag{b.4}$$

and by expression (b.1b), $\dfrac{d\Lambda_a}{d\zeta} = e_a$

Moreover, for each link a in $\boldsymbol{L}$, since $x_a = \dfrac{q_a}{\mu_a}$, $\mu_a = s_a \Lambda_a$, so

$$\frac{\partial X_0}{\partial \Lambda_a} = s_a \frac{\partial X_0}{\partial \mu_a} \tag{b.5}$$

and for each link a in A_{jm} , $\forall j \in P_m$, $\forall m \in N$

$$\frac{\partial X_0}{\partial \phi_{jm}} = \frac{\partial X_0}{\partial \Lambda_a} . \tag{b.6}$$

Therefore, the derivatives of the indicators of traffic conditions with respect to ζ for all links in the random plus oversaturation component can be directly obtained in expressions (4.8)-(4.14) and in the uniform component for oversaturated links can be obtained in expressions (4.15)-(4.21). For the uniform component of the undersaturated links without cumulative queues which have been given in expressions (3.31)-(3.34), and according to the following expression, the derivatives are discussed as follows.

$$F = \int_{a(x)}^{b(x)} f(x , t)dt$$

$$\frac{dF}{dx} = \frac{db(x)}{dx} f(x , b(x)) - \frac{da(x)}{dx} f(x , a(x)) + \int_{a(x)}^{b(x)} \frac{\partial f(x , t)}{\partial x} dt \tag{b.7}$$

It follows that

$$\frac{\partial D_a^U}{\partial \zeta} = \int_{t_a^0}^{t_a^0 + \frac{(1 - \Lambda_a)}{\zeta}} \bar{I}_a(t) - \bar{I}_a(t_a^0)\, dt +$$

$$\int_{t_a^0 + \frac{(1 - \Lambda_a)}{\zeta}}^{z_a} \bar{I}_a(t) - \bar{I}_a(t_a^0) - s_a(t - (t_a^0 + \frac{(1 - \Lambda_a)}{\zeta}))\, dt + \tag{b.8}$$

$$\zeta[\int_{t_a^0}^{z_a} \frac{\partial \bar{I}_a(t)}{\partial \zeta}\, dt - \int_{t_a^0 + \frac{1 - \Lambda_a}{\zeta}}^{z_a} s_a\, (\frac{1 - \Lambda_a + e_a\zeta}{\zeta^2})\, dt\ ]$$

$$\frac{\partial S_a^U}{\partial \zeta} = (\bar{I}_a(z_a) - \bar{I}_a(t_a^0)) + \zeta \frac{\partial \bar{I}_a(z_a)}{\partial \zeta} \tag{b.9}$$

$$\frac{\partial d_a^U}{\partial \zeta} = \frac{1}{q_a}\frac{\partial D_a^U}{\partial \zeta} \tag{b.10}$$

Similarly, the derivatives with respect to the equilibrium flows can be obtained and those have been given in expressions (4.25)-(4.28).

For the uniform component on link a with $x_a < 1$, $\forall a \in \boldsymbol{L}$ but with accumulated queues for which the clearance time is beyond the specified time period, the derivatives are given in expressions (4.15)-(4.21). On the other hand, for the uniform component on such links with accumulated queues for which the clearance time lies within the specified time period in expressions (3.36), (3.38)-(3.39), the derivatives are obtained as follows.

$$\frac{\partial D_a^U}{\partial \zeta} = \frac{1}{t_a^L}\left[\frac{\partial C_1}{\partial \zeta}(\tau_a - t_a^0) + \frac{\partial D_1}{\partial \zeta}(t_a^n - \tau_a) + \frac{\partial \tau_a}{\partial \zeta}(C_1 - D_1) \right] \tag{b.11}$$

$$\frac{\partial S_a^U}{\partial \zeta} = \frac{1}{t_a^L}\left[\frac{\partial C_2}{\partial \zeta}(\tau_a - t_a^0) + \frac{\partial D_2}{\partial \zeta}(t_a^n - \tau_a) + \frac{\partial \tau_a}{\partial \zeta}(C_2 - D_2) \right] \tag{b.12}$$

$$\begin{aligned}\frac{\partial d_a^U}{\partial \zeta} = {} & \frac{\left(\frac{\partial C_3}{\partial \zeta}\mu_a(\tau_a - t_a^0) + C_3 e_a s_a(\tau_a - t_a^0) + C_3\mu_a\frac{\partial \tau_a}{\partial \zeta}\right)}{\mu_a(\tau_a - t_a^0) + q_a(t_a^n - \tau_a)} + \\ & \frac{\left(\frac{\partial D_3}{\partial \zeta}q_a(t_a^n - \tau_a) - D_3 q_a\frac{\partial \tau_a}{\partial \zeta}\right)}{\mu_a(\tau_a - t_a^0) + q_a(t_a^n - \tau_a)} - \\ & \frac{(C_3\mu_a(\tau_a - t_a^0) + D_3 q_a(t_a^n - \tau_a))(e_a s_a(\tau_a - t_a^0) + \mu_a\frac{\partial \tau_a}{\partial \zeta} - q_a\frac{\partial \tau_a}{\partial \zeta})}{(\mu_a(\tau_a - t_a^0) + q_a(t_a^n - \tau_a))^2}\end{aligned} \tag{b.13}$$

$$\frac{\partial \tau_a}{\partial \zeta} = \frac{\frac{\partial L_a^e}{\partial \zeta}}{\mu_a(x_a - \hat{x}_a)} + \frac{e_a s_a \hat{x}_a \, (L_a^e - L_a^{r+o}(t_a^0))}{(\mu_a(x_a - \hat{x}_a))^2}$$

$$\frac{\partial L_a^e}{\partial \zeta} = \frac{-e_a x_a}{\Lambda_a} \left[1 + \frac{2C\, x_a - C\, x_a^2}{(1 - x_a)^2}\right]$$

because $\mu_a = \Lambda_a s_a$, $x_a = \dfrac{q_a}{\Lambda_a s_a}$.

$\dfrac{\partial C_1}{\partial \zeta}$, $\dfrac{\partial C_2}{\partial \zeta}$ *and* $\dfrac{\partial C_3}{\partial \zeta}$ have been obtained in expressions (4.15)-(4.17), and

$\dfrac{\partial D_1}{\partial \zeta}$, $\dfrac{\partial D_2}{\partial \zeta}$ *and* $\dfrac{\partial D_3}{\partial \zeta}$ have been obtained in expressions (4.22)-(4.24).

As for the derivatives with respect to the equilibrium flows, resulting expressions can be obtained in a similar way and those have been given in expressions (4.32)-(4.35).

B.2 Derivatives for Upstream Links

Derivatives of the indicators for the upstream links with respect to $\bar{\theta}_a$, Λ_a can be obtained in the following way.

For each signal group j in $\boldsymbol{P}_m$ at any one signal controlled junction m in $\boldsymbol{N}$, the upstream links which are controlled by the signal group j at that junction form the set

$\boldsymbol{A}_{jm} = \{a;\ a$ is the traffic stream such that $\hat{b}_{ajm} = 1\ \}$.

For the indicators given in Chapter 3 as briefly stated below:

$$X_a = X_0(\theta_a , \Lambda_a) , \forall\ a \in \boldsymbol{A}_{jm} , \ \forall j \in \boldsymbol{P}_m , \forall m \in \boldsymbol{N} \qquad \text{(b.14)}$$

the derivatives therefore can be obtained in the following ways.

For the random plus oversaturation component, because differentiation with respect

to $\bar{\theta}_a$ and Λ_a is equivalent to differentiation with respect to θ_{jm} and ϕ_{jm} respectively, the derivatives $\frac{\partial X_a}{\partial \bar{\theta}_a}$, $\frac{\partial X_a}{\partial \Lambda_a}$ are given in expressions (4.37), (4.38)-(4.40) respectively. For the uniform component on the oversaturated links, according to expressions given in (3.27)-(3.29), the derivatives can be obtained directly.

$$\frac{\partial D_a^U}{\partial \bar{\theta}_a} = \frac{\partial S_a^U}{\partial \bar{\theta}_a} = \frac{\partial d_a^U}{\partial \bar{\theta}_a} = 0 \tag{b.15}$$

$$\frac{\partial D_a^U}{\partial \Lambda_a} = \frac{s_a(1 - 2\Lambda_a)}{2\zeta} \tag{b.16}$$

$$\frac{\partial S_a^U}{\partial \Lambda_a} = 0 \tag{b.17}$$

$$\frac{\partial d_a^U}{\partial \Lambda_a} = \frac{-1}{2\zeta} \tag{b.18}$$

For the uniform component of the undersaturated links, the results obtained by Wong (1993, pp 251-254) are quoted in expressions (4.45)-(4.46), and (4.48)-(4.49).

For the uniform component on the undersaturated links with accumulated queues and the clearance time extending beyond the specified time period, the resulting derivatives are the same as in expressions (b.15)-(b.18) for the oversaturated links. On the other hand, for the undersaturated links with accumulated queues but the clearance time within the specified time period, the resulting derivatives are obtained as follows. According to the expressions given in (3.36), (3.38)-(3.39), the derivatives

$\frac{\partial X_a}{\partial y}$, $y = \bar{\theta}_a$, Λ_a can be obtained directly.

$$\frac{\partial D_a^U}{\partial \bar{\theta}_a} = \frac{1}{t_a^L}\left[\frac{\partial C_1}{\partial \bar{\theta}_a}(\tau_a - t_a^0) + \frac{\partial D_1}{\partial \bar{\theta}_a}(t_a^n - \tau_a)\right] \tag{b.19}$$

$$\frac{\partial S_a^U}{\partial \bar{\theta}_a} = \frac{1}{t_{a_L}}\left[\frac{\partial C_2}{\partial \bar{\theta}_a}(\tau_a - t_a^0) + \frac{\partial D_2}{\partial \bar{\theta}_a}(t_a^n - \tau_a)\right] \tag{b.20}$$

$$\frac{\partial d_a^U}{\partial \bar{\theta}_a} = \frac{\frac{\partial C_3}{\partial \bar{\theta}_a}\mu_a(\tau_a - t_a^0) + \frac{\partial D_3}{\partial \bar{\theta}_a}q_a(t_a^n - \tau_a)}{\mu_a(\tau_a - t_a^0) + q_a(t_a^n - \tau_a)} \tag{b.21}$$

$$\frac{\partial D_a^U}{\partial \Lambda_a} = \frac{1}{t_a^L}\left[\frac{\partial C_1}{\partial \Lambda_a}(\tau_a - t_a^0) + \frac{\partial D_1}{\partial \Lambda_a}(t_a^n - \tau_a) + \frac{\partial \tau_a}{\partial \Lambda_a}(C_1 - D_1)\right] \tag{b.22}$$

$$\frac{\partial S_a^U}{\partial \Lambda_a} = \frac{1}{t_a^L}\left[\frac{\partial C_2}{\partial \Lambda_a}(\tau_a - t_a^0) + \frac{\partial D_2}{\partial \Lambda_a}(t_a^n - \tau_a) + \frac{\partial \tau_a}{\partial \Lambda_a}(C_2 - D_2)\right] \tag{b.23}$$

$$\begin{aligned}\frac{\partial d_a^U}{\partial \Lambda_a} = {} & \frac{\left(\frac{\partial C_3}{\partial \Lambda_a}\mu_a(\tau_a - t_a^0) + C_3 s_a(\tau_a - t_a^0) + C_3\mu_a\frac{\partial \tau_a}{\partial \Lambda_a}\right)}{\mu_a(\tau_a - t_a^0) + q_a(t_a^n - \tau_a)} + \\ & \frac{\left(\frac{\partial D_3}{\partial \Lambda_a}q_a(t_a^n - \tau_a) - D_3 q_a\frac{\partial \tau_a}{\partial \Lambda_a}\right)}{\mu_a(\tau_a - t_a^0) + q_a(t_a^n - \tau_a)} - \\ & \frac{(C_3\mu_a(\tau_a - t_a^0) + D_3 q_a(t_a^n - \tau_a))\left(s_a(\tau_a - t_a^0) + \mu_a\frac{\partial \tau_a}{\partial \Lambda_a} - q_a\frac{\partial \tau_a}{\partial \Lambda_a}\right)}{(\mu_a(\tau_a - t_a^0) + q_a(t_a^n - \tau_a))^2}\end{aligned} \tag{b.24}$$

where $\frac{\partial C_x}{\partial y}$, $\frac{\partial D_x}{\partial y}$, $x =$ 1,2,3 have been given in expressions (b.15)-(b.18), and (4.45)-(4.50), and

$$\frac{\partial \tau_a}{\partial \Lambda_a} = \frac{\dfrac{\partial L_a^e}{\partial \Lambda_a}}{\mu_a(x_a - \hat{x}_a)} + \frac{s_a \hat{x}_a \, (L_a^e - L_a^{r+o}(t_a^0))}{(\mu_a(x_a - \hat{x}_a))^2}$$

$$\frac{\partial L_a^e}{\partial \Lambda_a} = \frac{-x_a}{\Lambda_a} \left[1 + \frac{2C\,x_a - C\,x_a^2}{(1 - x_a)^2}\right]$$

because $\mu_a = \Lambda_a s_a$, $x_a = \dfrac{q_a}{\Lambda_a s_a}$.

Using the facts that

$$\frac{\partial}{\partial \theta_{jm}} = \frac{\partial}{\partial \bar{\theta}_a} , \quad \frac{\partial}{\partial \phi_{jm}} = \frac{\partial}{\partial \Lambda_a} , \quad \frac{\partial C_x}{\partial \bar{\theta}_a} = 0 \ , \ x = 1, 2, 3$$

in expressions (b.19)-(b.24) yields the expressions (4.51)-(4.56).

B.3 Derivatives for Downstream Links

Derivatives of the indicators for the downstream links with respect to $\bar{\theta}_a$, Λ_a can be obtained in the following way.

For each signal group j in $\boldsymbol{P}_m$ at any one signal controlled junction m in $\boldsymbol{N}$, the downstream links for that signal group j at that junction form the set

$$\boldsymbol{B}_{jm} = \{b;\ b \text{ is the downstream link for some link } a \in \boldsymbol{A}_{jm}\},$$

where $\boldsymbol{A}_{jm} = \{a;\ a \text{ is the traffic stream such that } \hat{b}_{ajm} = 1\}$.

For the indicators given in Chapter 3, the expressions for the random plus oversaturated components take the form

$$X_b = X_0(q_b \ , \ x_b(q_b)) \ , \ \forall \ b \in \boldsymbol{B}_{jm} \ , \ \forall j \in \boldsymbol{P}_m \ , \ \forall m \in \boldsymbol{N} \qquad \text{(b.25)}$$

and those for the uniform component on saturated links take the form

$$Y_b = Y_0(\theta_b \ , \ \Lambda_b \ , \ q_b) \ , \ \forall b \in \boldsymbol{B}_{jm} \ , \ x_b \geq 1 \ , \ \forall j \in \boldsymbol{P}_m \ , \ \forall m \in \boldsymbol{N} \qquad \text{(b.25a)}$$

the derivatives therefore can be obtained as follows.

For the random plus oversaturation component, in expression (b.25), where

$x_b = \dfrac{q_b}{\Lambda_b s_b}$, since $\dfrac{\partial X_b}{\partial \theta_{jm}}$ *and* $\dfrac{\partial X_b}{\partial \phi_{jm}}$ measure rates of change with respect to

θ_{jm} , ϕ_{jm} subject to no change in other elements of $[\ \zeta\ ,\ \theta\ ,\ \phi\ ,\ \boldsymbol{q}\ ]$, and X_b

depends only on q_b *and* Λ_b , therefore

$$\frac{\partial X_b}{\partial \theta_{jm}} = \frac{\partial X_b}{\partial \phi_{jm}} = 0\ ,\quad \forall\, b \in \boldsymbol{B}_{jm}\ ,\quad \forall\, j \in \boldsymbol{P}_m\ ,\ \forall\, m \in \boldsymbol{N}$$

as in expressions (4.59)-(4.60).

For the uniform component of oversaturated links, since there are no quantities appearing in expression (b.25a) which are affected by the changes in signal timings

θ_{jm} , ϕ_{jm} when all other elements of $[\ \zeta\ ,\ \theta\ ,\ \phi\ ,\ \boldsymbol{q}\ ]$ remain unchanged, therefore

the derivatives

$$\frac{\partial Y_b}{\partial \theta_{jm}} = \frac{\partial Y_b}{\partial \phi_{jm}} = 0\ ,\quad \forall\, b \in \boldsymbol{B}_{jm}\ ,\quad \forall\, j \in \boldsymbol{P}_m\ ,\ \forall\, m \in \boldsymbol{N}$$

as in expressions (4.61)-(4.62).

For the uniform component on the undersaturated links, the derivatives for the downstream links with respect to θ_{jm} , ϕ_{jm} are affected by the degree of saturation for the upstream links. In the following discussions, we adopted the derivatives of Wong (1993, pp269-271) with respect to $\bar{\theta}_a$ for both undersaturated and oversaturated upstream links and the derivatives with respect to Λ_a for undersaturated upstream links only as quoted in expressions (4.63)-(4.64) and (4.66)-(4.67). The derivatives with respect to Λ_a when the upstream links are oversaturated will be investigated further in the latter part of this Appendix; because, unlike Wong, we require

q_b , $\forall\, b \in \boldsymbol{B}_{jm}$, $\forall\, j \in \boldsymbol{P}_m$, $\forall\, m \in \boldsymbol{N}$ to remain unchanged when we vary Λ_a , and this leads to a different form for the integrand in expressions (4.66)-(4.67).

For the uniform component on the undersaturated links with accumulated queues and the clearance time giving beyond the specified time period, the resulting derivatives are the same as those of the oversaturated links. On the other hand, for the undersaturated links with accumulated queues but the clearance time within the specified time period, the resulting derivatives are discussed as follows. According to the expressions given in (3.36), (3.38)-(3.39), $\forall\ b \in \boldsymbol{B}_{jm}$, $\forall\, j \in \boldsymbol{P}_m$, $\forall\, m \in \boldsymbol{N}$

$$D_b^U = \frac{C_1\,(\tau_b - t_b^0) + D_1\,(t_b^n - \tau_b)}{t_b^L} \tag{3.36}$$

$$S_b^U = \frac{C_2\,(\tau_b - t_b^0) + D_2\,(t_b^n - \tau_b)}{t_b^L} \tag{3.38}$$

$$d_b^U = \frac{C_3\,\mu_b\,(\tau_b - t_b^0) + D_3\,q_b\,(t_b^n - \tau_b)}{\mu_b\,(\tau_b - t_b^0) + q_b\,(t_b^n - \tau_b)} \tag{3.39}$$

where $\tau_b = \dfrac{L_b^e - L_b^{r+o}\,(t_b^0)}{\mu_b\,(x_b - \hat{x}_b)}$

$$\hat{x}_b = \frac{L_b^{r+o}\,(t_b^0) + 1 - \sqrt{(L_b^{r+o}\,(t_b^0) + 1)^2 - 4\,(1 - C)\,L_b^{r+o}\,(t_b^0)}}{2\,(1 - C)}$$

As before, the derivatives with respect to θ_{jm} *and* ϕ_{jm} are subject to no change in other elements of $[\ \zeta\ ,\ \theta\ ,\ \phi\ ,\ q\ ]$, and hence to no change in x_b and hence

$$\frac{\partial \tau_b}{\partial \theta_{jm}} = \frac{\partial \tau_b}{\partial \phi_{jm}} = 0$$

therefore, for $x = \bar{\theta}_a$, Λ_a ,

$$\frac{\partial D_b^U}{\partial x} = \frac{1}{t_b^L} [\frac{\partial C_1}{\partial x} (\tau_b - t_b^0) + \frac{\partial D_1}{\partial x}(t_b^n - \tau_b)] \tag{b.26}$$

$$\frac{\partial S_b^U}{\partial x} = \frac{1}{t_b^L} [\frac{\partial C_2}{\partial x}(\tau_b - t_b^0) + \frac{\partial D_2}{\partial x}(t_b^n - \tau_b)] \tag{b.27}$$

$$\frac{\partial d_b^U}{\partial x} = \frac{\frac{\partial C_3}{\partial x}\mu_b(\tau_b - t_b^0) + \frac{\partial D_3}{\partial x}q_b(t_b^n - \tau_b)}{\mu_b(\tau_b - t_b^0) + q_b(t_b^n - \tau_b)} \tag{b.28}$$

where $\frac{\partial C_y}{\partial x}$, $\frac{\partial D_y}{\partial x}$, $x = \bar{\theta}_a$, Λ_a ; $y = 1,2,3$ are respectively the derivatives for the oversaturated in expressions (4.61)-(4.62) and undersaturated downstream links in expressions (4.63)-(4.68).

B.4 Derivatives for Further Downstream Links

Derivatives for the indicators of the further downstream links with respect to $\bar{\theta}_a$, Λ_a can be obtained in the following way.

For each signal group j in $\boldsymbol{P}_m$ at any one signal controlled junction m in $\boldsymbol{N}$, the further downstream links for that signal group j at that junction form a set

$\boldsymbol{C}_{jm} = \{c;\ c$ is the immediate downstream link for some link $b \in \boldsymbol{B}_{jm}\}$,

where $\boldsymbol{B}_{jm} = \{b;\ b$ is the downstream link for some link $a \in \boldsymbol{A}_{jm}\}$, and

$\boldsymbol{A}_{jm} = \{a;\ a$ is the traffic stream such that $\hat{b}_{ajm} = 1\}$

For the random plus oversaturation component and the uniform component of the oversaturated links, for the same reasons as applied to the downstream links, the derivatives of the indicators of traffic conditions with respect to $\bar{\theta}_a$, Λ_a can be seen to be zeros, leading to expressions (4.75)-(4.78). For the uniform component for the undersaturated links, the derivatives are obtained as follows

Following Wong (1993, pp 279), suppose that the changes in the indicators of traffic conditions produced by changing θ_{jm} for link a and by changing ϕ_{jm} for undersaturated link a, $\forall \ a \in \boldsymbol{A}_{jm}$ are negligible when the flows have dispersed beyond one link; then the corresponding derivatives can be taken as zero as shown in expressions (4.79)-(4.81). Furthermore, because the coupling effects (which will be discussed in Section B.5) for the oversaturated upstream links a, $\forall \ a \in \boldsymbol{A}_{jm}$, the derivatives for the further downstream links c, $\forall \ c \in \boldsymbol{C}_{jm}$ with respect to Λ_a are by analogy with the downstream links as follows.

$$\frac{\partial D_c^U}{\partial \Lambda_a} = \zeta \int_{t_c^0}^{z_c} \frac{\partial \bar{I}_c(t)}{\partial \Lambda_a} dt = \zeta \int_{t_c^0}^{z_c} \int_{t_c^0}^{t} \frac{\partial I_c(w)}{\partial \Lambda_a} dw \, dt = \zeta \int_{t_c^0}^{z_c} \frac{\partial I_c(t)}{\partial \Lambda_a} (z_c - t) dt \qquad \text{(b.29)}$$

$$\frac{\partial S_c^U}{\partial \Lambda_a} = \zeta \left[I_c(z_c) \frac{\partial z_c}{\partial \Lambda_a} + \int_{t_c^0}^{z_c} \frac{\partial I_c(t)}{\partial \Lambda_a} dt \right] = \left[\frac{\zeta \, s_c}{s_c - I_c(z_c)} \right] \int_{t_c^0}^{z_c} \frac{\partial I_c(w)}{\partial \Lambda_a} dw \qquad \text{(b.30)}$$

$$\frac{\partial d_c^U}{\partial \Lambda_a} = \frac{1}{q_c} \frac{\partial D_c^U}{\partial \Lambda_a} \qquad \text{(b.31)}$$

where $\frac{\partial I_c(t)}{\partial \Lambda_a}$ can be evaluated by the techniques used for expression (b.42).

For the uniform component for the undersaturated links with accumulated queues and the clearance time giving beyond the specified time period, the resulting derivatives

are the same as those for the oversaturated links. On the other hand, for undersaturated links with accumulated queues but the clearance time within the specified time period, the resulting derivatives are for the same reasons as for the downstream links.

$$\frac{\partial D_c^U}{\partial \Lambda_a} = \frac{1}{t_c^L}\left(\frac{\partial D_1}{\partial \Lambda_a}(t_c^n - \tau_c)\right) \tag{b.32}$$

$$\frac{\partial S_c^U}{\partial \Lambda_a} = \frac{1}{t_c^L}\left(\frac{\partial D_2}{\partial \Lambda_a}(t_c^n - \tau_c)\right) \tag{b.33}$$

$$\frac{\partial d_c^U}{\partial \Lambda_a} = \frac{\frac{\partial D_3}{\partial \Lambda_a} q_c(t_c^n - \tau_c)}{\mu_c(\tau_c - t_c^0) + q_c(t_c^n - \tau_c)} \tag{b.34}$$

where $\frac{\partial D_1}{\partial \Lambda_a}$, $\frac{\partial D_2}{\partial \Lambda_a}$ *and* $\frac{\partial D_3}{\partial \Lambda_a}$ are given by expressions (b.29)-(b.31).

B.5. Coupling Effects

The coupling effects between the upstream, downstream and further downstream links that affect the derivatives of indicators of traffic conditions on downstream and further downstream links with respect to the upstream signal timings will now be discussed. The changed input flow profiles on the downstream and further downstream links can be obtained from the changed output flow profiles of the upstream links when the signal timings at the upstream junction are changed. Basically, we refer to the results of Wong (1993, pp69-72) and additionally take a detailed look at the changed output flow profiles for the oversaturated upstream links when the duration of green at the corresponding junction is changed while the average link flows are fixed.

Let

O_a be the output flow pattern for link a , $\forall\ a \in \boldsymbol{L}$.

$\bar{\delta}$ be the Dirac's delta function with unit area such that

$$\bar{\delta}(t - t_a^0) = \begin{cases} \infty , & if \ t = t_a^0 \\ 0 , & otherwise \end{cases}$$

T_x be the undelayed travel time on the link x , and

$$y_x^1 = (t_a^0 + T_x) \ mod \ (\frac{1}{\zeta})$$

$$y_x^2 = (t_a^0 + \frac{1 - \Lambda_a}{\zeta} + T_x) \ mod \ (\frac{1}{\zeta})$$

$$y_x^3 = (z_a + T_x) \ mod \ (\frac{1}{\zeta})$$

F be a modified smoothing factor, and

$$n \in \{ 1, 2, 3, ..., n_{\max} \}$$

$$n_{\max} = Maximise\{ p : p \ integer \ and \ p \leq \zeta \ (t_x^n - y_x^k) \}$$

for $x = b \ or \ c$, $b \in \boldsymbol{B}_{jm}$ *and* $c \in \boldsymbol{C}_{jm}$, $\forall j \in \boldsymbol{P}_m$, $\forall m \in \boldsymbol{N}$, let

$$R(y_x^k, t) = \frac{F \ \exp (- F (t - y_x^k))}{1 - \exp (\frac{- F}{\zeta})} \tag{b.35}$$

$$for \ y_x^k + \frac{n - 1}{\zeta} \leq t \leq y_x^k + \frac{n}{\zeta} , \ k = 1, 2, 3$$

Therefore the changed output flow profile of link a in $\boldsymbol{A}_{jm}$, $\forall j \in \boldsymbol{P}_m$, $\forall m \in \boldsymbol{N}$ and the subsequent changed input flow profiles on the downstream and further downstream links b and c , $b \in \boldsymbol{B}_{jm}$, $c \in \boldsymbol{C}_{jm}$, $\forall j \in \boldsymbol{P}_m$, $\forall m \in \boldsymbol{N}$ are

(i). for the undersaturated upstream link a in $\boldsymbol{A}_{jm}$, $\forall j \in \boldsymbol{P}_m$, $\forall m \in \boldsymbol{N}$,

$$\frac{\partial O_a(t)}{\partial \bar{\theta}_a} = \frac{1}{\zeta}[\ I_a(t_a^0)\bar{\delta}(t - t_a^0) - s_a\bar{\delta}(t - (t_a^0 + \frac{1 - \Lambda_a}{\zeta})) + (s_a - I_a(t_a^0))\bar{\delta}(t - z_a)\ ]$$

$$\frac{\partial O_a(t)}{\partial \Lambda_a} = \frac{1}{\zeta}\ I_a(t_a^0)[\ \bar{\delta}(t - t_a^0) - \bar{\delta}(t - z_a)\ ]$$

Therefore the subsequent changed output flow profiles are

$$\frac{\partial I_x(t)}{\partial \theta_a} = \frac{f_{ax}}{\zeta}\ [\ I_a(t_a^0)\ R(y_x^1, t) - s_a\ R(y_x^2, t) - (s_a - I_a(t_a^0))\ R(y_x^3, t)] \qquad \text{(b.36)}$$

$$\frac{\partial I_x(t)}{\partial \Lambda_a} = \frac{f_{ax}\ I_a(t_a^0)}{\zeta}\ [\ R(y_x^1, t) - R(y_x^3, t)\ ] \qquad \text{(b.37)}$$

where f_{ax} , $x = b$, c is the proportion of the flow on the upstream link a turns to downstream link b or further downstream link c.

(ii). for the oversaturated upstream link a in $\boldsymbol{A}_{jm}$, $\forall j \in \boldsymbol{P}_m$, $\forall m \in \boldsymbol{N}$,

$$\frac{\partial O_a(t)}{\partial \bar{\theta}_a} = \frac{s_a}{\zeta}\ [\ \bar{\delta}(t - t_a^0) - \bar{\delta}(t - (t_a^0 + \frac{1 - \Lambda_a}{\zeta}))\ ]$$

$$\frac{\partial I_x(t)}{\partial \bar{\theta}_a} = \frac{f_{ax}\ s_a}{\zeta}\ [\ R(y_x^1, t) - R(y_x^2, t)\ ] \qquad \text{(b.38)}$$

As we evaluate the derivatives respect to ζ , θ , ϕ , it is assumed that all equilibrium link flows remain unchanged for the whole time period. But as we evaluate the derivatives with respect to Λ_a , $\forall a \in L$ for downstream and further downstream links when the corresponding upstream link a which is controlled by signal group j at junction m is oversaturated, this assumption may be violated. To maintain the unchanged link flows, we need additional discussions.

As the upstream link a is oversaturated, the OUT profile in TRANSYT for this link corresponds to the average exit flow μ_a , which is the product of saturation flow

rate s_a and effective green represented as proportion of cycle time Λ_a . To keep the exit flow of upstream link a unchanged in each time period while the duration of green at the exit from the link changes, the saturation flow rate on this link is required to change.

For any upstream link a in $\boldsymbol{L}$, with saturation flow rate s_a , which is controlled by signal group j at junction m, where $x_a \geq 1$, we require

$$(s_a + \Delta s_a)(\Lambda_a + \Delta \Lambda_a) = s_a \Lambda_a \tag{b.39}$$

In the limit as $\Delta \Lambda_a \rightarrow 0$, $$\frac{\partial s_a}{\partial \Lambda_a} = \frac{-s_a}{\Lambda_a} \tag{b.40}$$

Let

$$j \in \{ 1, 2, ..., j_{\max} \}$$

$$j_{\max} = \textit{Maximise} \left\{ p : p \textit{ integer and } p \leq \zeta (t_a^n - t_a^0) \right\}$$

The original output flow profile over a whole cycle is given by

$$O_a(t) = \begin{cases} 0 , & \textit{for } t_a^0 + \dfrac{j-1}{\zeta} \leq t < t_a^0 + \dfrac{j-\Lambda_a}{\zeta} \\ s_a , & \textit{for } t_a^0 + \dfrac{j-\Lambda_a}{\zeta} \leq t < t_a^0 + \dfrac{j}{\zeta} \end{cases}$$

Following the expression introduced by Wong for the case where Λ_a changes but s_a does not change, and according to the required relation in expression (b.39) the change in the output flow profile for upstream link a, i.e. $\Delta O_a(t)$, due to the corresponding changes in Λ_a and s_a can be expressed in the following way.

For $t_a^0 + \frac{j-1}{\zeta} \le t < t_a^0 + \frac{j-\Lambda_a}{\zeta}$

$$\Delta O_a(t) = (s_a + \Delta s_a) \frac{\Delta \Lambda_a}{\zeta} \bar{\delta} \left(t - t_a^0 - \frac{(j-1)}{\zeta}\right)$$

to first order, and

$$\approx \frac{s_a \Delta \Lambda_a}{\zeta} \bar{\delta} \left(t - t_a^0 - \frac{(j-1)}{\zeta}\right)$$

for $t_a^0 + \frac{j-\Lambda_a}{\zeta} \le t < t_a^0 + \frac{j}{\zeta}$

$$\Delta O_a(t) = \frac{s_a \Delta \Lambda_a}{\zeta} \bar{\delta} \left(t - t_a^0 - \frac{(j-1)}{\zeta}\right) + \Delta s_a$$

$$= \frac{s_a \Delta \Lambda_a}{\zeta} \bar{\delta} \left(t - t_a^0 - \frac{(j-1)}{\zeta}\right) - \frac{s_a \Delta \Lambda_a}{\Lambda_a}$$

when expression (b.40) is used.

Thus the following results can be obtained immediately.

$$\frac{\partial O_a(t)}{\partial \Lambda_a} = \begin{cases} \dfrac{s_a \bar{\delta} \left(t - t_a^0 - \frac{(j-1)}{\zeta}\right)}{\zeta} & \text{for } t_a^0 + \frac{j-1}{\zeta} \le t < t_a^0 + \frac{j-\Lambda_a}{\zeta} \\ \dfrac{s_a \bar{\delta} \left(t - t_a^0 - \frac{(j-1)}{\zeta}\right)}{\zeta} - \dfrac{s_a}{\Lambda_a} & \text{for } t_a^0 + \frac{j-\Lambda_a}{\zeta} \le t < t_a^0 + \frac{j}{\zeta} \end{cases} \tag{b.41}$$

Consider the dispersed flow on the downstream link b and further downstream link c: by using expression (b.35), we shall show that

$$\frac{\partial I_x(t)}{\partial \Lambda_a} = \frac{1}{\zeta} [f_{ax} s_a R(y_x^1, t)] + K(t) \tag{b.42}$$

where $x = b , c ; \ \forall b \in \boldsymbol{B}_{jm} , \forall c \in \boldsymbol{C}_{jm} , \forall j \in \boldsymbol{P}_{jm} , \forall m \in \boldsymbol{N}$

and $K(t)$ can be determined in the following way.

Let $k(t)$ be the partial derivative with respect to Λ_a of the OUT profile for link a resulting from the change in the saturation flow on the upstream link a in T_a ; therefore following expression (b.41)

$$k(t) = \begin{cases} 0 , \ for \ t_a^0 + \dfrac{j - 1}{\zeta} < t < t_a^0 + \dfrac{j - \Lambda_a}{\zeta} \\ \dfrac{- s_a}{\Lambda_a} , \ for \ t_a^0 + \dfrac{j - \Lambda_a}{\zeta} \leq t < t_a^0 + \dfrac{n}{\zeta} \\ \dfrac{s_a \ \bar{\delta}(0)}{\zeta} , \ for \ t = t_a^0 + \dfrac{j - 1}{\zeta} \end{cases} \tag{b.43}$$

Following the platoon dispersal equation used in TRANSYT but as a first order differential equation (Wong 1993, pp283-285), let $\boldsymbol{q}(t)$ be the OUT profiles for the upstream links at time t, and $\bar{\boldsymbol{q}}(t)$ be the IN profiles for corresponding downstream or further downstream links at time t,

$$\bar{\boldsymbol{q}}'(t) + F \ \bar{\boldsymbol{q}}(t) = F \ \boldsymbol{q}(t - \mathrm{T}) \tag{b.44}$$

and it can be solved as

$$\bar{\boldsymbol{q}}(t) = F \exp (- F \ t) \int_{-\infty}^{t} \exp (F \ w) \ \boldsymbol{q}(w - \mathrm{T}) \ dw \tag{b.45}$$

We now apply this platoon dispersal equation expression (b.45) to (b.43).

Let T_b , T_c be respectively the undelayed travel time on the downstream and further downstream links b and c, $\forall b \in \boldsymbol{B}_{jm} , \forall c \in \boldsymbol{C}_{jm} , \forall j \in \boldsymbol{P}_{jm} , \forall m \in \boldsymbol{N}$, and

for $x = b\ ,\ c$

$$n \in \{\ 1,\ 2,\ 3,\ \ldots,\ n_{\max}\ \}$$

$$n_{\max} = Maximise\left\{\ p : p\ integer\ \ and\ \ p \le \zeta\ (t_x^{\,n} - y_x^{\,1})\ \right\}$$

$$y_x^{\,1} = (t_a^{\,0} + \mathrm{T}_x)\ mod\ (\frac{1}{\zeta})$$

let f_{ab} , f_{ac} be respectively the proportion of traffic on link a which flows into the corresponding downstream and further downstream links b and c.

The partial derivatives $K(t)$ with respect to Λ_a of the dispersed flow on downstream link b and on further downstream link c, arising from $k(t)$, which is the partial derivative with respect to Λ_a of the OUT profile of changed saturation flow on upstream link a, can be obtained respectively by summing up all the partial derivatives of periodic dispersed flows on downstream and further downstream links as follows.

(i). For $y_x^{\,1} + \dfrac{n\ -\ 1}{\zeta} < t \le y_x^{\,1} + \dfrac{n\ -\ \Lambda_a}{\zeta}$,

the partial derivative $K(t)$ with respect to Λ_a of the dispersed flow are

$$K(t) = f_{ax}\ F\ \exp(\ -F\ t)\ [\ \sum_{u\ =\ 0}^{\infty}\ \int_{y_x^{\,1} + \frac{n\ -\ u\ -\ 1\ -\ \Lambda_a}{\zeta}}^{y_x^{\,1} + \frac{n\ -\ u\ -\ 1}{\zeta}} \exp(\ F\ w)\ (\frac{-\ s_a}{\Lambda_a})dw\ ]$$

$$= (-\ \frac{f_{ax}\ s_a}{\Lambda_a})\ [\frac{\exp(\ F\ (y_x^{\,1} + \frac{n\ -\ 1}{\zeta}))\ (1\ -\ \exp(\frac{-\ F\ \Lambda_a}{\zeta}))}{1\ -\ \exp(\ -\ \frac{F}{\zeta})}\ ]\ \exp(\ -F\ t)$$

(b.46)

(ii). for $y_x^1 + \frac{n - \Lambda_a}{\zeta} \leq t < y_x^1 + \frac{n}{\zeta}$,

$$K(t) = f_{ax} F \exp(-F t) \{ [\sum_{u=0}^{\infty} \int_{y_x^1 + \frac{n-u-1-\Lambda_a}{\zeta}}^{y_x^1 + \frac{n-u-1}{\zeta}} \exp(F w)(\frac{-s_a}{\Lambda_a}) dw] +$$

$$\int_{y_x^1 + \frac{n-\Lambda_a}{\zeta}}^{t} \exp(F w)(\frac{-s_a}{\Lambda_a}) dw \}$$

$$= (\frac{-f_{ax} s_a}{\Lambda_a}) [1 + \exp(-F t)$$

$$[\frac{\exp(F(y_x^1 + \frac{n-\Lambda_a}{\zeta}))(\exp(-F(\frac{1-\Lambda_a}{\zeta})) - 1)}{1 - \exp(\frac{-F}{\zeta})}]]$$

(b.46a)

Furthermore, in expressions (b.46) the derivatives with respect to Λ_a of the total dispersed flow on the downstream and further downstream links over a whole cycle can be calculated in the following way.

$$\int_{t}^{t+\frac{1}{\zeta}} K(w)\, dw = \frac{-f s_a}{\zeta} , \forall \ t \in T \tag{b.47}$$

where f is the appropriate flow fraction factor.

Appendix C Some results from nonlinear programming

Consider a nonlinear problem of minimizing $F(\boldsymbol{x})$, *where* $F : \mathbb{R}^n \rightarrow \mathbb{R}$ and $\boldsymbol{x}$ is a row vector of dimension n . There are some relevant definitions and theorems stated below.

<u><*Definition c.1*></u>*<Descent direction of* F *at* $\boldsymbol{x}^0$ *>*

Let $F : \mathbb{R}^n \rightarrow \mathbb{R}$ *be differentiable at* $\boldsymbol{x}^0$ *. If there is a vector* $\boldsymbol{h}$ *such that*

$\nabla F(\boldsymbol{x}^0)\, \boldsymbol{h}^T < 0$ *then vector* $\boldsymbol{h}$ *is called a descent direction of* F *at* $\boldsymbol{x}^0$ *.* ■

<u><*Theorem c.2*></u> *<Steepest descent direction of* F *at* $\boldsymbol{x}^0$ *>*

Let $F : \mathbb{R}^n \rightarrow \mathbb{R}$ *be differentiable at* $\boldsymbol{x}^0$ *and* $\nabla F(\boldsymbol{x}^0) \neq \boldsymbol{0}$ *. Let* $\boldsymbol{h}$ *be a direction at* $\boldsymbol{x}^0$ *then the solution to the problem to minimize* $\nabla F(\boldsymbol{x}^0)\, \boldsymbol{h}^T$ *subject to* $\| \boldsymbol{h} \| \leq 1$ *is given by the steepest descent direction* $\boldsymbol{h}^0 = \dfrac{-\nabla F(\boldsymbol{x}^0)}{\| \nabla F(\boldsymbol{x}^0) \|}$.

<proof>

The problem is to

$$\begin{aligned} &\underset{\boldsymbol{h}}{\textit{Minimise}} \quad \nabla F(\boldsymbol{x}^0)\, \boldsymbol{h}^T \\ &\textit{subject to} \\ &\qquad \| \boldsymbol{h} \| \leq 1 \end{aligned} \tag{c.1}$$

According to *Cauchy-Schwarz inequality*, for any vectors $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{R}^n$ we have

$$| \boldsymbol{x}\, \boldsymbol{y}^T | \leq \| \boldsymbol{x} \| \, \| \boldsymbol{y} \| \tag{c.2}$$

Applying to expression (c.1) we have

$$\| \nabla F(\boldsymbol{x}^0)\, \boldsymbol{h}^T | \leq \| \nabla F(\boldsymbol{x}^0) \| \, \| \boldsymbol{h} \|$$

$$\Rightarrow \nabla F(\boldsymbol{x}^0)\, \boldsymbol{h}^T \geq - \| \nabla F(\boldsymbol{x}^0) \| \, \| \boldsymbol{h} \| = - \| \nabla F(\boldsymbol{x}^0) \| \quad \textit{when } \|\boldsymbol{h}\| = 1$$

with equality holding throughout if and only if $\boldsymbol{h} = \boldsymbol{h}^0 = \dfrac{-\nabla F(\boldsymbol{x}^0)}{\| \nabla F(\boldsymbol{x}^0) \|}$ (c.3)

■

<u>*<Definition c.3>*</u> *<Feasible direction of F at $\boldsymbol{x}^0$ >*

Consider now the constrained nonlinear problem of minimizing $F(\boldsymbol{x})$ subject to $\boldsymbol{x} \in S_0$ where S_0 is a nonempty set in $\mathbb{R}^n$. For a given $\boldsymbol{x}^0 \in S_0$, a vector $\boldsymbol{h}$ is called a feasible direction of F at $\boldsymbol{x}^0$ if there exists a $\delta > 0$ such that

$\boldsymbol{x}^0 + \lambda \boldsymbol{h} \in S_0$, *for each* $\lambda \in (0 , \delta)$. ■

<u>*<Definition c.4>*</u>*<Improving feasible direction of F at $\boldsymbol{x}^0$ >*

A feasible direction $\boldsymbol{h}$ of F at $\boldsymbol{x}^0$ is called an improving feasible direction, if there exists a $\delta > 0$ such that

$F(\boldsymbol{x}^0 + \lambda \boldsymbol{h}) < F(\boldsymbol{x}^0)$, *for each* $\lambda \in (0 , \delta)$. ■

<u>*<Theorem c.5>*</u>*<Karush-Kuhn- Tucker necessary conditions for local optimality>*

$F : \mathbb{R}^n \to \mathbb{R}$, $G_i : \mathbb{R}^n \to \mathbb{R}$, *for* $i = 1 ,..., m$, $H_i : \mathbb{R}^n \to \mathbb{R}$, *for* $i = 1 ,..., l$

let S_0 be a nonempty open set in $\mathbb{R}^n$ then a nonlinear problem P is to

$$\underset{\boldsymbol{x}}{\text{Minimise}} \quad F(\boldsymbol{x})$$

subject to

$$\begin{aligned} & G_i(\boldsymbol{x}) \le 0 \quad , \text{ for } i = 1, ..., m \\ & H_i(\boldsymbol{x}) = 0 \quad , \text{ for } i = 1, ..., l \\ & \boldsymbol{x} \in S_0 \end{aligned} \tag{c.4}$$

Let $\boldsymbol{x}^0$ locally solve problem P and let $\boldsymbol{E} = \{ i : G_i(\boldsymbol{x}^0) = 0 \}$. Suppose F , G_i

for $i = 1, ..., m$ *and* H_i , *for* $i = 1, ..., l$ *are differentiable at* x^0 *; furthermore, suppose that the constraints in (c.4) satisfy the* <u>*Linear Independence Constraint Qualification*</u> *as follows.*

(i). $G_i(x)$ *for each* $i \notin E$ *is continuous at* x^0 .

(ii). $H_i(x)$ *for* $i = 1 , ... , l$ *is continuously differentiable at* x^0 *, and*

(iii). $\nabla G_i(x^0)$ *for each* $i \in E$ *and* $\nabla H_i(x^0)$ *for* $i = 1 , ... , l$ *are linearly independent.*

Then x^0 *is a KKT point, that is, there exist scalars* $u_i \geq 0$, $i \in E$ *and* v_i *, for* $i = 1,...,l$ *such that*

$$\nabla F(x^0) + \sum_{i \in E} u_i \nabla G_i(x^0) + \sum_{i=1}^{l} v_i \nabla H_i(x^0) = \mathbf{0} \qquad \text{(c.5)}$$

<proof>

A proof can be found in Bazaraa, Sherali and Shetty (1993, pp192). ■

The following definitions and theorems are related to a *gradient projection method* (Luenberger 1989, pp330-337), and are adopted from Bazaraa, Sherali and Shetty, (1993, pp448-454).

<u>*<Definition c.6>*</u> *<Projection matrix* g *>*

An $n \times n$ *matrix* g *is called a projection matrix if* $g = g^T$ *and* $g\,g = g$. ■

<u>*<Lemma c.7>*</u>*<adopted from Bazaraa et al. (1993, pp448)*

Let g *be an* $n \times n$ *matrix then the following two statements are true:*

1. If g *is a projection matrix then* g *is positive semidefinite.*

2. g *is a projection matrix if and only if* $I - g$ *is a projection matrix, where* I *is an identity matrix.*

■

<u>*<Lemma c.8>*</u> *<adopted from Bazaraa et al. (1993, pp448-454)*

Consider the problem P *in expression (c.4) subject to a linear constraint set. Let* x^0 *be a feasible solution for problem* P *and let* $E = \{ i : G_i(x^0) = 0 \}$ *. Let* M *be the matrix whose rows are* $\nabla G_b = \left[\nabla G_i(x^0) , \ i \in E \right]$ *and*

$\nabla H = \left[\nabla H_i(x^0) , \ i = 1 ,..., l \right]$ *. Suppose* F *is differentiable at* x^0 *. If* g *is a projection matrix such that* $g \nabla F(x^0)^T \neq 0$ *then* h *given by*

$h^T = - g \nabla F(x^0)^T$ *is an improving feasible direction of* F *at* x^0 *. Moreover if*

$M^T = \left[\nabla G_b^T , \nabla H^T \right]$ *has full rank and if* g *is of the form*

$$g = I - M^T (MM^T)^{-1} M \tag{c.6}$$

then h *is an improving feasible direction.* ■

Now suppose that $h^T = - g \nabla F(x^0)^T = 0$ and let $w = - (MM^T)^{-1} M \nabla F(x^0)^T$ and $w^T = [u^T , v^T]$ when partitioned to match $M^T = \left[\nabla G_b^T , \nabla H^T \right]$, then

$$\begin{aligned} 0 &= g \nabla F(x^0)^T \\ &= [I - M^T(MM^T)^{-1} M] \nabla F(x^0)^T \\ &= \nabla F(x^0)^T + M^T w \\ &= \nabla F(x^0)^T + \nabla G_b^T u + \nabla H^T v \end{aligned}$$

<u>*<Lemma c.9>*</u> *<when* $h^T = - g \nabla F(x^0)^T = 0$ *> <adopted from Bazaraa et al. (1993, pp448-454)>*

If $u \geq 0$ *then* x^0 *is a KKT point; otherwise an improving feasible direction* h *can be*

constructed as follows. Since $\boldsymbol{u}$ *is not* $\geq \mathbf{0}$ *, there is at least one negative component of* $\boldsymbol{u}$ *. Let* $\hat{M}^T = \left[\nabla \bar{G}_b^T , \nabla H^T \right]$ *where* $\nabla \bar{G}_b^T$ *is obtained from* ∇G_b *by deleting the rows of* ∇G_b *corresponding to the negative components* u_j *, and*

$$\hat{g} = I - \hat{M}^T (\hat{M}\hat{M}^T)^{-1} \hat{M} .$$

Then $h^T = - \hat{g} \nabla F(x^0)^T$ *is an improving feasible direction.* (c.7)

■

Following the results in lemmas *c.8 & c.9*, the search process for solution for problem P in (c.4) either can be terminated at a KKT point or a new improving feasible direction at current solution can be generated; we conclude this appendix with the following lemma.

<u>*<Lemma c.10>*</u> *<Stopping conditions>*

If x^0 *is a KKT point for problem* P *in (c.4) then the search process may stop;*

otherwise a new improving feasible direction at x^0 *can be generated according to <Lemma c.8> & <Lemma c.9>.* ■