

### **3. PROBLEM ANALYSIS AND FORMULATION**

This chapter addresses a linear programming-based approach for solving the problem of determining a trip table using traffic counts on links. The approach falls into the category of  $\Delta$ -equilibrium types of procedures as defined below, with more emphasis on the nature of uncertainty in the data and in the problem. The observed traffic counts are permitted to have some kind of deviations in this approach, by allowing each link volume to be known only over a confidence interval (range). The user-optimality principle is then accordingly modified in a consistent manner. The approach is further extended to handle the case of accommodating prior trip table information to guide the solution trip tables, to treat the case of missing volumes, and to address the case of an unrealized system-optimum.

#### **3.1 Concepts and Notation**

For the traffic system in a given study area, the area is usually divided into several internal and external zones according to the characteristics of the land-use. Accordingly, the transportation system in the area is abstracted as a traffic network (a directed network or a digraph). Each of the zones is represented by a special node, known as a centroid, which is connected to the network through one or several special links called centroid connectors. Intersections between links in the street network are denoted by nodes, and the segments between two intersections, or centroids, are represented as links in the traffic network. Correspondingly, the ground traffic counts on the roads or streets are recognized as the link flows or link

volumes. For consistency, the notation in the paper of Sherali *et al.* [1994a], whenever possible, will be adopted in the following discussion.

Given a traffic network, let  $G(N, A)$  represent the underlying digraph, where  $N$  is the set of nodes (vertices or points) including centroids, and  $A$  is the set of corresponding directed links (arcs, branches, edges, or lines). Let  $OD$  denote the set of O-D pairs that comprise the trip table to be estimated, where  $O$  is the set of possible starting or origin nodes, and  $D$  is the set of ending or destination nodes. Furthermore, let  $T = \{n \in N, n \notin O \cup D\}$  represent the set of all transshipment (intermediate, non-centroid) nodes.

An O-D trip table is a two dimensional matrix in which the rows and columns represent the origin and destination zones, respectively. The cell value, denoted as  $x_{ij}$ , is the number of trips (flows in the network) from the origin zone (centroid)  $i$  to the destination zone (centroid)  $j$ . The problem addressed in this research is to estimate the flows  $x_{ij}$  for each O-D pair  $(i, j)$ , given observed flows  $\bar{f}_a$  on each link  $a \in A$ . The problem is then extended to the case where there are some links on which the link flows are not available.

For any two nodes  $i, j \in N$ , a *path* from node  $i$  to the node  $j$  is a sequence of links (arcs) in which the initial node of each arc is the same as the terminal node of the preceding arc in the sequence, the initial node of the first arc is the node  $i$ , and the terminal node of the last arc is the node  $j$ . A *simple path*  $a$  is a path in which the nodes mentioned above, except possibly  $i$  and  $j$ , are distinct (Bazaraa *et al.* 1990). If  $i = j$  then the simple path is called a *circuit*, or a *cyclic (directed) path*; otherwise, it is called a *noncyclic path* or a *noncircuit*. A path that is not simple is called a *nonsimple path*. In general, a nonsimple path from  $i$  to  $j$ ,  $i \neq j$ , can be decomposed into a simple path plus one or more simple circuits, which are in turn connected. In this research, unless otherwise specified, a path shall always mean a

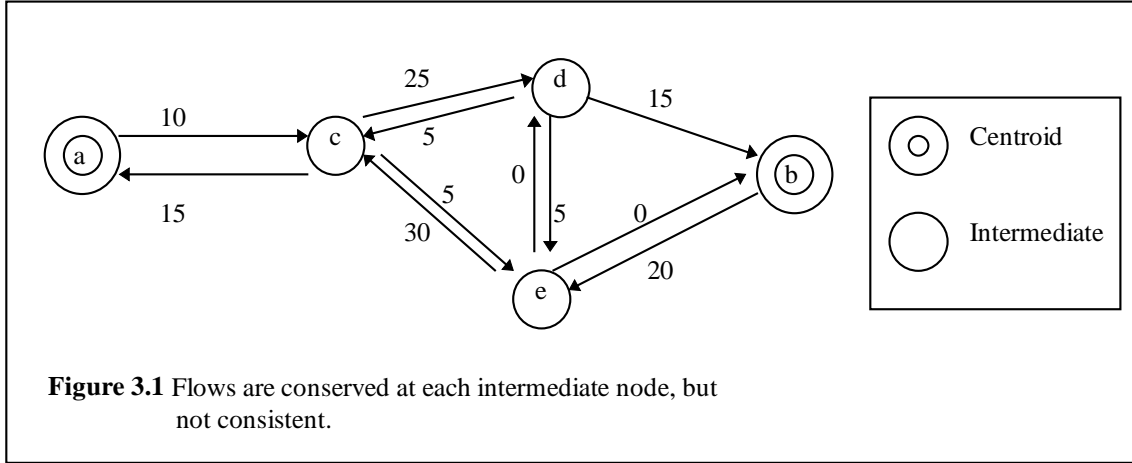
simple noncircuit. In this case, the sequence of the (directed) arcs will be uniquely determined if the set of arcs in the path is given. Therefore, let  $p_{ij}^k$  represent all the paths between  $(i, j) \in OD$ ,  $k = 1, \dots, n_{ij}$ . Note that  $p_{ij}^k$  may be represented by an  $|A|$ -dimensional vector with the  $a$ -th component being unit, for  $a \in A$ , if the corresponding link belongs to the particular path, and zero otherwise. Assume that for each O-D pair  $(i, j) \in OD$ ,  $i \neq j$ , there exists at least one path  $p_{ij}^k$  that connects  $i$  to  $j$ . Assume that each transshipment node can be passed through by at least one of these paths, and each link in the network belongs to at least one of these paths. Let  $x_{ij}^k$  be the contribution of  $x_{ij}$  to the path  $p_{ij}^k$  for each  $k = 1, \dots, n_{ij}$  and let  $\bar{x}$  denote the vector of components  $x_{ij}^k$ . Then  $x_{ij}$  can be decomposed as

$$x_{ij} = \sum_{k=1}^{n_{ij}} x_{ij}^k. \quad [3.1]$$

Furthermore, if the (estimated) flow  $\bar{x}$  indeed reproduces the observed link counts  $\bar{f}_a$ , the following relationship needs to be satisfied,

$$\sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k) x_{ij}^k = \bar{f}, \quad x \geq 0. \quad [3.2]$$

For a given observed link flow vector  $\bar{f}$ , if there exists a (non-negative) O-D matrix  $\bar{x}$  satisfying condition [3.2], the observed flows (or the system) are then called *consistent*; otherwise, the system is called *inconsistent*. The observed flow is said to satisfy the *law of conservation* (or *nodal balance law*) if for each intermediate (transshipment, or non-centroidal) node  $n \in T$ , the total observed flow emanating out of the node  $n$  equals the total observed flow coming into node  $n$ . Obviously, if the observed flows are consistent, then they are conserved for each intermediate node. On the other hand, as indicated in Figure 3.1, conservation of flow at each intermediate node cannot ensure its consistency, because of the possibility of cyclic flows.



Based on the observed flows  $\bar{f}_a$ ,  $a \in A$ , a corresponding link impedance  $\bar{c}_a \equiv c_a(\bar{f}_a)$  can be computed. Since centroid connectors are mostly some artificial links, hence, throughout this research it is assumed that the following holds true.

**Assumption 1.** *The link travel time/cost  $c_a(v_a)$  is a constant if  $a$  is a centroid connector, and is a (strictly) increasing positive function of flow  $v_a$  otherwise.*

Let  $A_C$  indicate the set of centroid connectors and let  $A_T = A \setminus A_C$  indicate the set of other non-centroid connector links. Then, usually,  $c_a(v_a)$  takes the form of (Easa 1991)

$$c_a(v_a) = \begin{cases} c_{\text{delay}} + c_a^F & \text{if } a \in A_C \\ c_{\text{delay}} + c_a^F \left[ 1 + \alpha \left( \frac{v_a}{u_a} \right)^\beta \right] & \text{if } a \in A_T, \end{cases} \quad [3.3]$$

where,  $c_a(v_a)$  is the travel time/cost on link  $a$  as a function of the volume  $v_a$ ,  $c_{\text{delay}}$  and  $c_a^F$  are respectively the (average) delay and free-flow travel time/cost on link  $a$ ,  $v_a$  and  $u_a$  are the flow on link  $a$  and the flow capacity of link  $a$ , respectively, and  $\alpha$  and  $\beta$  are positive constants. Typically, as suggested by the Bureau of Public Roads (BPR) (1964),  $\alpha = 0.15$  and  $\beta = 4$ , and [3.3] is then called the *BPR link cost or impedance equation*.

## 3.2 Model Formulation

For a given complete set of observed flows, if the system is consistent, and hence the flow through each intermediate node is conserved, then the linear programming approach [2.4] introduced in Chapter 2 is equilibrium-based. Furthermore, the observed flow pattern is at an equilibrium if and only if the minimized objective value equals the total observed cost (Sherali *et al.* 1994a).

It should be noted that the perfect conservation of flow, and the consistency of the system, are only an ideal case in theory and commonly cannot be expected to hold in practice. Errors in measurements, approximation in the network representation and, especially, time variations in flow counts may cause this failure. This causes an intrinsic inconsistency with [3.2] and results in an infeasibility of the program [2.4]. Several attempts for handling this inconsistency have been made in the literature, such as manipulating the data in order to make [3.2] feasible (Van Zuylen and Branton 1982), the balancing approach (Barbourt and Fricker 1989) and by penalizing the objective function (Sherali, Hobeika and Sivanandan 1994) as in program [2.5]. The philosophy used in this research is the acceptance of the fact that the observed data may have some kind of uncertainty, and that the actual volume on each link could belong to a (confidence) interval. Note that the widths of these confidence intervals vary in the link number  $a \in A$  for the same criteria (of probability). Therefore, instead of [3.2], the feasible region for the estimated O-D matrix is constrained via

$$\bar{f} \leq \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k) x_{ij}^k \leq \bar{f} + \bar{\Lambda}, \quad x \geq 0 \quad [3.4]$$

where,  $\bar{f}$  is the vector of observed flows on each link, and  $\bar{\Lambda} \geq 0$  is a vector of non-negative (integer) constants.

From the allowance of uncertainty for the observed flows, it is natural to correspondingly modify the inter-zonal accessibility  $K_{ij}$  in [2.4d]. That is, instead of

examining the paths having a minimum cost, paths for which the corresponding cost belongs to a *cost band* are used for representing inter-zonal accessibility for each O-D pair. Again, the widths of these bands depend on both the minimum observed cost for each O-D pair and the corresponding components of vector  $\bar{\Lambda}$  in [3.4]. Hence, they are  $(i, j) \in OD$ -dependent. The introduction of the bands for inter-zonal accessibility could be interpreted in two ways. 1) It represents the inaccuracy of the inter-zonal accessibility corresponding to the inaccuracy in the observed data as derived above; 2) it represents the fact that the costs perceived by users could be *suboptimal* and might have some kind of variation.

According to Assumption I and [3.4], for a given link  $a \in A$ , the estimated impedance or cost on the link depends (nonlinearly) on the total estimated flow passing through that link. Therefore, the objective function, similar to the one in [2.4a], is nonlinearly dependent on the solution  $\bar{x}$ . In this research, however, the following linearized cost hypothesis is adopted.

Assumption II      *The impedance or cost on link  $a \in A$  depends only on the "observed flows"  $\bar{f}_a$  on that link.*

Notationally, let  $c_{ij}^k \equiv \bar{c} \cdot p_{ij}^k$  denote the impedance or cost on route  $k$  between O-D pair  $(i, j)$  for each  $k = 1, \dots, n_{ij}$ ,  $(i, j) \in OD$ , and let  $c_{ij}^* = \min\{c_{ij}^k, k = 1, \dots, n_{ij}\}$ . In terms of the fact that the travel cost is a strictly increasing function of link volumes, a modified inter-zonal accessibility set can be defined as

$$\Pi_{ij} \equiv \{ k \in \{1, \dots, n_{ij}\} : c_{ij}^* \leq c_{ij} \leq c_{ij}^* + \Delta_{ij} \}. \quad [3.5]$$

Given an O-D trip table, let  $\Delta$  denote the vector having components  $\Delta_{ij}$  for each  $(i, j) \in OD$ . An assignment is said to be  $\Delta$ -optimal or at a  $\Delta$ -equilibrium if all the routes between any O-D pair  $(i, j)$  which have positive flows are chosen within the set  $\Pi_{ij}$  defined by [3.5]. An observed link flow pattern is said to be at a  $(\Delta, \Lambda)$ -

*equilibrium*, or in short a  $\Delta$ -*equilibrium*, if there exists an O-D matrix and a  $\Delta$ -*equilibrium* assignment, such that when the O-D matrix is assigned to the network, in a  $\Delta$ -user optimal fashion that satisfies [3.5], it reproduces observed link flows to a  $\Lambda$  tolerance given by [3.4]. Therefore, if an approach can generate an O-D matrix which, whenever the observed link flow pattern is at a  $(\Delta, \Lambda)$ -equilibrium, the O-D matrix can be assigned to the network in the aforementioned fashion, the approach is then called a  $(\Delta, \Lambda)$ -*equilibrium* type of procedure (or a  $\Delta$ -*equilibrium* type of procedure if  $\bar{\Lambda} = 0$  in [3.4]).

$\Delta$ -*optimality* may be regarded as a natural extension of the user-optimum principle to the case of uncertainty. In general, if the system is at a user-optimum, then it is  $\Delta$ -optimal. For a given network, if  $|\Delta|$  and  $|\Lambda|$  are sufficiently small, then user-optimality and  $\Delta$ -optimality can be shown to be the same.

Based on Wardrop's principle, and imposing Assumption I & II, the above discussion leads to the following linear programming model.

$$\text{Minimize} \quad \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} \tilde{c}_{ij}^k x_{ij}^k \quad [3.6a]$$

$$\text{subject to} \quad \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a x_{ij}^k \geq \bar{f}_a, \quad \forall a \in A \quad [3.6b]$$

$$\sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a x_{ij}^k \leq \bar{f}_a + \bar{\Lambda}_a, \quad \forall a \in A \quad [3.6c]$$

$$x \geq 0 \quad [3.6d]$$

where

$$\tilde{c}_{ij}^k = \begin{cases} c_{ij}^k & \text{if } k \in \Pi_{ij} \\ M_1 c_{ij}^k & \text{if } k \notin \Pi_{ij}, \end{cases} \quad [3.6e]$$

and where  $M_1 > 0$  and  $\bar{\Lambda}_a > 0$  are some constants.

The following lemma compares the optimal objective value of the linear program [3.6] with the optimal objective value of the problem defined by [2.4].

**LEMMA 3.1** *If the linear program [2.4] is feasible, then problem [3.6] is feasible. In this case, the optimal objective value  $\tilde{z}$  of [3.6] is at most equal to  $\hat{z}$ , the optimal objective value of [2.4].*

**PROOF.** This follows from the fact that if  $x$  satisfies [2.4b], then it satisfies [3.6b]-[3.6c], and the fact that, from the definition, each of the coefficients  $\hat{c}_{ij}^k$  in [2.4a] is greater than or equal to the correspondent coefficient  $\tilde{c}_{ij}^k$  in [3.6a].

**LEMMA 3.2** *Let  $\bar{C}_{total} = \sum \bar{c}_a \bar{f}_a$  represent the total observed system cost. Then the optimal objective value of the linear program [3.6], if it exists, is at least  $\bar{C}_{total}$ . Furthermore, the optimal objective value equals  $\bar{C}_{total}$  if and only if there exists a feasible solution  $x^*$  satisfying [3.2] and using only the paths in  $\Pi_{ij}$  for all  $(i, j) \in OD$ .*

**PROOF.** For any feasible solution  $x$  of [3.6], it follows from [3.6e] and [3.6b] that

$$\sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} \tilde{c}_{ij}^k x_{ij}^k \geq \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} c_{ij}^k x_{ij}^k = \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (\bar{c} \cdot p_{ij}^k) x_{ij}^k \geq \bar{c} \cdot \bar{f} = \bar{C}_{total}. \quad [3.7]$$

Hence, it follows that the optimal value of the linear program [3.6] is at least  $\bar{C}_{total}$ .

Next, there exists a feasible solution  $x^*$  such that the objective value equals  $\bar{C}_{total}$  if and only if the two inequalities in [3.7] become equalities. Since  $M_1 > 1$ ,

$$\sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} \tilde{c}_{ij}^k x_{ij}^k = \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} c_{ij}^k x_{ij}^k \Leftrightarrow x_{ij}^k = 0, \quad \forall k \notin \Pi_{ij}, i, j \in OD.$$

Moreover, since  $\bar{c}$  is a vector with all positive components and  $x^*$  satisfies [3.6b],



$$\sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (\bar{c} \cdot p_{ij}^k) x_{ij}^k = \bar{c} \cdot \bar{f} \quad \Leftrightarrow \quad \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k) x_{ij}^k = (\bar{f})$$

This concludes the second part of Lemma 3.2.

**Remark 3.1** *Three types of costs, total observed cost  $\bar{C}_{total}$ , assigned cost  $\sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} c_{ij}^k x_{ij}^k$ , and penalizing cost  $\sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (\tilde{c}_{ij}^k - c_{ij}^k) x_{ij}^k$ , have been referred to in the proof of Lemma 3.2. The observed total cost depends only on the observed link flows, while the assigned cost depends on the assigned volume on each link, and the penalizing cost depends on the estimated trip matrix  $\bar{x}$ , as well as on the parameters  $M_1$  and  $\Delta$ . When the observed volume is at a  $\Delta$ -optimum, the linear program [3.6] will determine such a solution since the second and third kinds of costs reach their respective minima in this case. Otherwise, a minimization of the assigned cost attempts to allocate as much flow as possible such that the assigned volumes are close (from above) to the observed volumes on all links, while minimizing the penalizing cost attempts to allocate the flows to the paths in  $\Pi_{ij}$  as far as possible. By varying the parameters  $M_1$  and  $\Delta$ , the corresponding changes in the optimal solution  $x^*$  and in the optimal objective value  $\tilde{z}$  should reflect such relationships.*

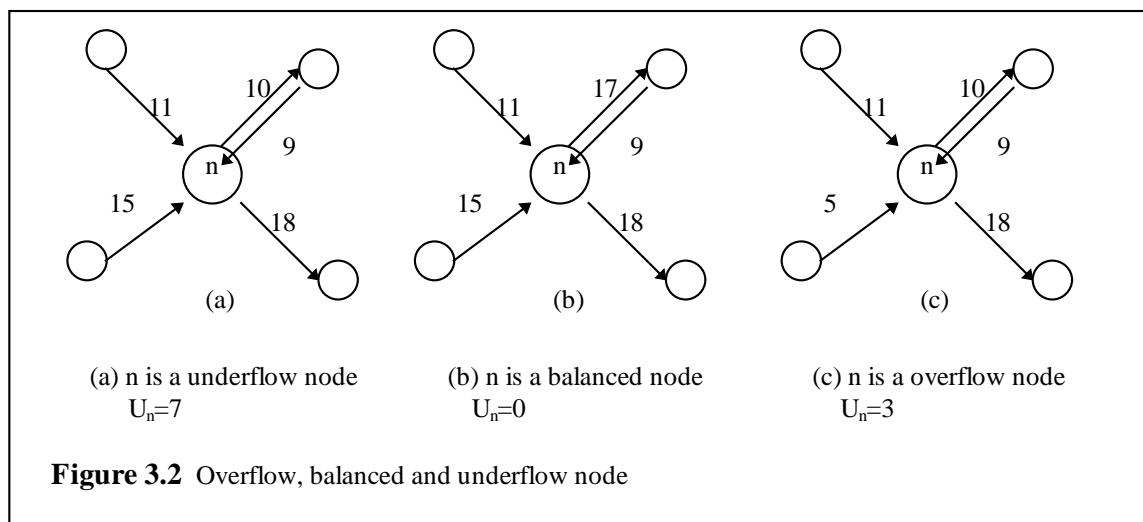
**Remark 3.2** *For the path  $k \in \Pi_{ij}$ , one possible way is to first find the solution of the shortest impedance problem  $c_{ij}^* \equiv \min\{c_{ij}^k : k = 1, \dots, n_{ij}\}$ , as well as the set of user-suboptimal paths  $\Pi_{ij}$ . This can be done at the start of the algorithm by solving  $|O|$  shortest path problems having positive link coefficients  $c_a$ ,  $a \in A$ .*

### 3.3 Existence of Feasible Solutions

Until now, it has been assumed that the linear program [3.6] has at least a feasible solution. Since only simple noncircuit paths are permitted to carry positive flows, this assumption is true if the observed flows are conserved (at each intermediate node) and cyclic flows are absent. On the other hand, if there are some intermediate nodes at which the flows are not conserved, it is questionable if the system possesses any feasible solution. The problem addressed here is to determine the value of  $\Lambda$  so that the linear program assures feasibility.

An intermediate node  $n \in T$  is said to be an *overflow* (*underflow*) node if the total flow out of the node is greater (less) than the total flow into the node. For any intermediate node  $n$ , define the *unbalanced index*, denoted by  $U_n$ , of node  $n$  by

$$U_n \equiv \begin{cases} \text{total incoming flows} - \text{total outgoing flows} & \text{if } n \text{ is an underflow node} \\ 0 & \text{if node } n \text{ is balanced} \\ \text{total outgoing flows} - \text{total incoming flows} & \text{if } n \text{ is an overflow node.} \end{cases}$$



The following lemma ensures the existence of conserved flows for the linear program [3.6].

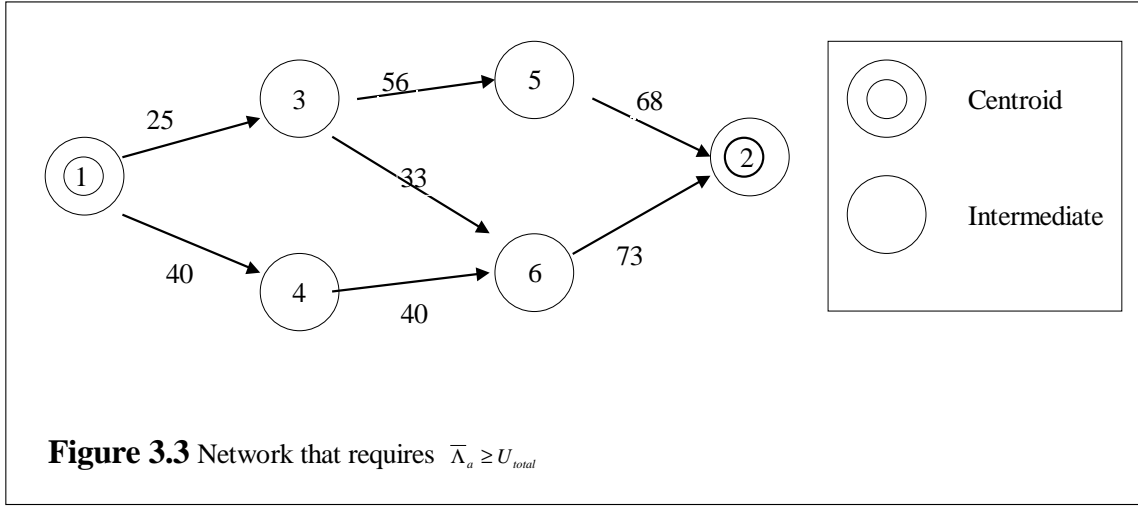
**LEMMA 3.3** Let  $U_{total} = \sum_{n \in T} U_n$  represent the total (observed) system unbalanced index. If  $\min_{a \in A} \bar{\Lambda}_a \geq U_{total}$ , then there exists a set of conserved flows  $\{f_a, a \in A\}$  such that  $\bar{f}_a \leq f_a \leq \bar{f}_a + \bar{\Lambda}_a$ .

**PROOF** The lemma is proven by constructing a set of flows  $f$  as follows. Given any intermediate node  $n$ , by the assumption, there exists at least one path  $p_{ij}^k$  passing through that node. If  $n$  is an underflow node, then increase the flow by  $U_n$  on any link  $a \in A$  which is located in the path  $p_{ij}^k$  in the *forward direction* from node  $n$  onward, and keep the other flows unchanged. Similarly, when  $n$  is an overflow node, then increase the flow on any link  $a \in A$  which is located in the path  $p_{ij}^k$  in the direction leading up to node  $n$  by  $U_n$ , and keep the other flows unchanged. By this method, the unbalanced index of any other transshipment node, except node  $n$ , will remain unchanged, while the node  $n$  will become balanced. Since it is assumed that there are no circuitous flows, the above process will increase the flow on any link by at most  $U_n$ . The lemma then follows if the above procedure is used for each  $n \in T$ .

As shown in the example in Figure 3.3, in certain cases,  $U_{total}$  is an optimal least band for  $\bar{\Lambda}_a$  in order to ensure the existence of feasible conserved flows. However, as seen below, an improved condition is also available.

Recall that the *in-degree* of a node  $n$  is the number of arcs that have  $n$  as their to-node, and the *out-degree* of  $n$  is the number of arcs that have  $n$  as their from-node. Define the *divided unbalance index* of node  $n$ , denoted by  $U_n$ , as the average

(rounded-up integer) of the balance index of  $n$  divided by its out-degree if  $n$  is an underflow node, and by its in-degree if  $n$  is an overflow node. Then the following lemma improves upon the result in Lemma 3.3.



**LEMMA 3.4** Assume that the network has the following property: for each intermediate node  $n \in T$ , there are two subtrees  $T_n^o$  and  $T_n^i$  (of the network) in which the node  $n$  acts as their root. Furthermore, in  $T_n^o$  ( $T_n^i$ ), the order of the parent-child in the tree is adopted from the from-to (to-from) relationship of the network, the leaves of the subtree consist only of some destination (origin) nodes, and all of the arcs that have  $n$  as its from-node (to-node) in the network are links of the subtree. Hence,  $T_n^o$  ( $T_n^i$ ) is a subtree composed of all nodes that are reachable from (can reach) node  $n$ . If  $\min_{a \in A} \bar{\Lambda}_a \geq \sum_{n \in T} U_n$ , then there exist conserved flows  $\{f_a, a \in A\}$  such that  $\bar{f}_a \leq f_a \leq \bar{f}_a + \bar{\Lambda}_a$ .

**PROOF** The approach in this proof is similar to the proof of Lemma 3.3. The only difference is that, instead of “distributing” the overflow at each node  $n$  on just one

path, these overflows are distributed “uniformly” over each branch of the appropriate subtrees rooted at  $n$ . The detailed argument is therefore omitted.

As discussed in the last section, even if the observed flow is conserved at each intermediate node, the system may still be inconsistent, and therefore, the conditions in Lemmas 3.3 and 3.4 do not ensure the existence of a feasible solution for [3.6]. The difficulty here is mainly caused by the occurrence of circuitous flows. The following lemma addresses this situation.

**LEMMA 3.5** (a) *If  $\max_{a \in A} \bar{\Lambda}_a < \max_{n \in T} U_n$ , then the linear program [3.6] has no feasible solutions.*

(b) *If  $\min_{a \in A} \bar{\Lambda}_a \geq \frac{1}{2} \sum_{a \in A} \bar{f}_a + U_{total}$ , then there exists a feasible solution for the linear program [3.6].*

**PROOF** The first half of the lemma follows from the definitions of nodal balance and the constraints in [3.6]. For the second half, by the hypothesis, each link can be passed through by at least one simple path from an origin node to a destination node. When a circuitous flows occurs, decompose the corresponding cyclic path into two or more simple ones between some node pairs. Since each cyclic path consists of at least two links, this decomposition is possible. Next, for each resulting simple path, by approaching its beginning node from an origin node and linking its ending node to a destination node, a simple path from that origin node to that destination node is generated. By using a balancing process on each such resulting simple path as in the proof of Lemma 3.3, the O-D flows can be augmented until all of the cyclic flows are accounted for. This will have increased the flow on any link by at most  $\frac{1}{2} \sum_{a \in A} \bar{f}_a$ . Proceeding now as in Lemma 3.3 for covering the remainder of the observed flows, the flow on any link will be increased by at most an additional amount of  $U_{total}$ . This completes the proof.

In order to ensure feasibility of the linear program, one method might be to enlarge the width of these confidence intervals. This theoretical approach does not fit certain realistic situations. Furthermore, the higher complexity of covering cyclic flows increases the computational demand for large-sized networks. Fortunately, there is another way to deal with this situation, i.e., by introducing extra penalty variables to account for flow deviations only when necessary. Therefore, a vector of artificial variables  $y$  is introduced, with components  $y_a$  for each  $a \in A$ . Accordingly, the objective function is penalized using the term  $M e_A \cdot y$ , where  $e_A$  is a vector of  $|A|$  ones, and  $M$  is a suitable penalty parameter. This modifies the interval linear programming model to the following form (LPI).

$$\text{LPI: Minimize} \quad \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} \tilde{c}_{ij}^k x_{ij}^k + M e_A \cdot y \quad [3.8a]$$

$$\text{subject to} \quad \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a x_{ij}^k + y_a \geq \bar{f}_a, \quad a \in A \quad [3.8b]$$

$$\sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a x_{ij}^k - y_a \leq \bar{f}_a + \bar{\Lambda}_a, a \in A \quad [3.8c]$$

$$x \geq 0, \quad y \geq 0. \quad [3.8d]$$

**Remark 3.3** *The penalty term  $M e_A \cdot y$  in the LPI model plays a two-fold role in practice. First, as mentioned before, it represents a cost for the possible out of range deviation of the link flows. But since  $M$  is assumed to be very large, the optimal solution attempts to allocate as much flow to lie within the confidence range as possible. Secondly, similar to the technique used in the big- $M$  method (Bazaraa et al. 1990), the use of the artificial variables  $y$  can make it easier to choose a starting basic feasible solution for solving the linear program [3.6] via the simplex method. It should be pointed out that the artificial variable  $y$  here essentially differs from the artificial variable used in the big- $M$  method. The former is a*

variable of the linear program itself and is allowed to be kept in the basis even when the system is at an optimum, while the latter is just a device for starting the simplex method and must be ultimately driven out of the optimal basis.

**Remark 3.4** Two constants,  $M_1$  and  $M$ , and two constant vectors,  $\bar{\Lambda}$  and  $\Delta$ , have been introduced. They represent treatments for different aspects of uncertainty within the model. Different combinations of these values can be chosen in the research via sensitivity analyses. In general, the following choices are recommended.  $\bar{\Lambda}$  is set to be ten percent of the observed flow on link  $a \in A$ .  $\Delta_{ij}$  is similarly set to be ten percent of  $c_{ij}^*$  for each O-D pair  $(i, j) \in OD$ .  $M_1$  is chosen as  $M_1 = 2$ . Furthermore, letting  $\bar{c}_* = \max\{\bar{c}_a : a \in A\}$ ,  $M$  will be chosen, as recommended by Sherali et al. (1994) in a similar case, to be  $M = 1 + \bar{c}_* + \bar{C}_{total}$ .

### 3.4 Using a Target Trip Table to Guide Solutions

In this section the proposed model LPI is extended to the case of using a target trip table to “guide” the output solutions. This idea originated from the LINKOD model, and has been used in the LP(TT) model (Sherali et al. 1994) and also in other equilibrium-based models. The purpose of this treatment is to rectify some shortcomings of model LPI, such as the absence of a mechanism to discriminate among alternative optimal O-D trip tables, and the sparse characteristics of the generated O-D matrix due to the nature of extreme point LP optima.

Suppose that there exists a partial prior (target) trip table (TT) with associated O-D flows  $Q_{ij} > 0$  for  $(i, j) \in \overline{OD} \subseteq OD$ , where  $\overline{OD}$  might represent some significant or key O-D pairs. By introducing two vectors of non-negative artificial variables  $Y_{ij}^+$  and  $Y_{ij}^-$ , which represent the positive and negative deviations of the O-D flow  $x_{ij}$

from the target trip table value  $Q_{ij}$ , respectively, and introducing a penalty parameter  $M_\sigma$ , which reflects the relative degree of importance in minimizing the trip table deviations versus the link flow deviations, the following modified version LPI(TT) of model LPI can be formulated.

$$\text{LPI(TT): Minimize } \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} \tilde{c}_{ij}^k x_{ij}^k + M e_A \cdot y + M_\sigma \sum_{(i,j) \in \overline{OD}} (Y_{ij}^+ + Y_{ij}^-) \quad [3.9a]$$

$$\text{subject to } \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a x_{ij}^k + y_a \geq \bar{f}_a, \quad a \in A \quad [3.9b]$$

$$\sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a x_{ij}^k - y_a \leq \bar{f}_a + \bar{\Lambda}_a, a \in A \quad [3.9c]$$

$$\sum_{k=1}^{n_{ij}} x_{ij}^k + (Y_{ij}^+ - Y_{ij}^-) = Q_{ij}, \quad (i,j) \in \overline{OD} \quad [3.9d]$$

$$x \geq 0, \quad y \geq 0, \quad Y^+ \geq 0, \quad Y^- \geq 0. \quad [3.9e]$$

As recommended in (Sherali *et al.* 1994a), the penalty parameter  $M_\sigma$  is chosen as  $M_\sigma = \sigma M$  with  $\sigma \in (0, 1]$ .

Other alternative approaches, similar to those used in the models *NM1* and *NM2* discussed in Chapter 2, can be applied to take into account the effects of target trip tables. For example, consider the following linear programming subproblem.

$$\text{LPISP: Minimize } \sum_{(i,j) \in \overline{OD}} (Y_{ij}^+ + Y_{ij}^-) \quad [3.10a]$$

$$\text{subject to } \sum_{k=1}^{n_{ij}} x_{ij}^k + (Y_{ij}^+ - Y_{ij}^-) = Q_{ij}, \quad (i,j) \in \overline{OD} \quad [3.10b]$$

$$x \text{ solves LPI,} \quad Y^+ \geq 0, \quad Y^- \geq 0. \quad [3.10c]$$

The following lemma states some relationships among the optimal objective values of the foregoing problems.



**LEMMA 3.6** Let  $\tilde{z}_1$ ,  $\tilde{z}_2$ , and  $\tilde{z}_3$  represent the optimal objective values of LPI, LPI(TT) and LPISP, respectively. Then

$$\tilde{z}_1 \leq \tilde{z}_2 \leq \tilde{z}_1 + M_\sigma \tilde{z}_3.$$

The lemma can be easily proven by comparing the corresponding feasible regions and the coefficients in the objective function, and a formal proof is omitted.

**Remark 3.5** Although one motivation for introducing a target trip table in LPI(TT) is to reduce the number of possible alternative (extreme point) optimal solutions of model LPI, there is no theoretical result to ensure that this will indeed be the case. On the other hand, when  $M_\sigma$  is sufficiently small, this discrimination is more likely.

**Remark 3.6** It is known that the linear programming subproblem LPISP can be alternatively rewritten as:

$$\text{Minimize} \quad \sum_{(i,j) \in OD} \left| \sum_{k=1}^{n_{ij}} x_{ij}^k - Q_{ij} \right| \quad [3.11]$$

subject to  $x$  optimally solves LPI.

Similarly, the linear programming model LPI(TT) can be rewritten as

$$\text{Minimize} \quad \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} \tilde{c}_{ij}^k x_{ij}^k + Me_A \cdot y + M_\sigma \sum_{(i,j) \in OD} \left| \sum_{k=1}^{n_{ij}} x_{ij}^k - Q_{ij} \right| \quad [3.12]$$

subject to the constraints (3.9b), (3.9c) and (3.9e).

Let  $z_1^*$  be the objective value of [3.8a] corresponding to some basic feasible solution to LPI so that  $z_1^* > \tilde{z}_1$ , and suppose that there are no other basic feasible solutions to LPI such that the corresponding objective value  $z_1 \in (\tilde{z}_1, z_1^*)$ . Let  $\sigma > 0$  be sufficiently small so that  $M_\sigma \cdot \tilde{z}_3 < z_1^* - \tilde{z}_1$ . Then by using the extreme point optimality property of linear programming, it can be shown that for such a  $\sigma$ , solving the bilevel linear

*programming problem [3.10] is equivalent to solving the linear programming problem [3.9].*

### 3.5 Missing Volumes and Sub-System Optima

In this section, the proposed linear programming model LPI(TT) is enhanced to account for two miscellaneous cases: the problem of missing link volumes in a given network and the problem of system suboptima. Differing from the techniques used in other models, these two problems are treated in this research by using preprocessing and postprocessing, respectively. These ideas can also be used in other equilibrium-based approaches, since the problems themselves are independent of linear programming techniques.

The missing link volume problem occurs in various real network models. For example, since a centroid in a given network is usually a symbolic node for a zone, it is frequently the case that the link volumes on centroid connectors are not available. Different criteria have been employed to treat this subject, such as balancing the node by using shortest path flows (Barbour and Fricker 1990), matrix regeneration techniques (Erlander *et al.* 1985), local user equilibrium techniques (Frank 1994), and updating linear programming cost techniques (Sherali *et al.* 1994b). This research does not attempt to evaluate these approaches, but provides a variety of simpler preprocessing routines for treating the case of missing volumes on some links.

In order to minimize the artificial factor influencing the real link flow data and hence the O-D trip table, the obtained “observed flows”  $\bar{f}_a$ , via the preprocessing routine for those links  $a \in A$  having missing volumes, are *only* used for the purpose of “estimating the observed cost”  $c_a(\bar{f}_a)$ . Moreover, since there is no precise mechanism to ensure what the real observed flows on these links should be, it is natural to consider these link flows as “unconstrained” in the linear programming

model. The inter-zonal accessibility set should be modified in the same manner. More precisely, let  $A_v$  represent the set of links on which the link volume is available and  $A_s$  be the set of links with missing volumes. For an intermediate node  $n \in T$ , let  $A_{iv}^{(n)}$  ( $A_{ov}^{(n)}$ ) denote the set of arcs that have volumes available and have  $n$  as their *to-node* (*from-node*), and similarly, let  $A_{is}^{(n)}$  ( $A_{os}^{(n)}$ ) denote the set of arcs that have missing volumes and that have  $n$  as their *to-node* (*from-node*). Since the capacity of links is a very important factor that influences the user choice of routes, the following type of simple preprocessing routine might be considered to handle missing volumes.

Suppose that on the link  $a \in A$ , which connects some node  $n_1$  to a node  $n_2$ , the link volume is not available. Then *assign* an “observed volume”  $\bar{f}_a$  to  $a$  so that

$$\bar{f}_a = \frac{(u_a)^\gamma}{|A_{ov}^{(n_1)}| + |A_{iv}^{(n_2)}|} \left( \sum_{b \in A_{ov}^{(n_1)}} \frac{\bar{f}_b}{(u_b)^\gamma} + \sum_{b \in A_{iv}^{(n_2)}} \frac{\bar{f}_b}{(u_b)^\gamma} \right) \quad [3.13]$$

where  $\gamma \geq 0$  is a constant which is called a *choice factor*, and  $u_a$  is the capacity of link  $a$ . Note that the choice factor might depend on the attributes of the to-node. However, this is treated as a constant in this research. Once the flow in [3.13] is determined, it is used to update the corresponding cost in the objective function of the linear program. This flow  $\bar{f}_a$  will be assumed to be the “observed” volume temporarily assigned to the link  $a$ , and the link  $a$  will be regarded as belonging to  $A_v$  until all of the costs on the missing volume links are found. By assumption, the network is connected. Therefore, if  $A_v$  is initially not empty, then all of the links with missing volumes in the network will be assigned some “observed” volumes by this approach.

The above preprocessing routine may be called a *capacity-based approach*, because of its assumption that the user’s choice on the route depends only on the capacity of that route. Two special cases will be tested in this research. The first case, to be referred to as the *capacity-proportional approach*, will assume that  $\gamma = 1$ .

The second one, will be referred to as *average-volume approach*, and will assume that  $\gamma = 0$ .

**Remark 3.7** *Although the solution to the equation [3.13] is unique, the assigned “observed” volume in the above approach clearly depends on the order in which the links are considered for assignment. It is reasonable to choose the order of links beginning with the one that has the most information regarding available volumes on its incident nodes.*

Now, assume that the above approach has been used to determine the costs on the links having missing volumes, and hence, the travel times on all of the links are available. The model LPI can be modified to account for the case of missing volumes via the following revised model LPIM.

$$\text{LPIM: Minimize} \quad \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} \tilde{c}_{ij}^k x_{ij}^k + M e_{A_v} \cdot y \quad [3.14a]$$

$$\text{subject to} \quad \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a x_{ij}^k + y \geq \bar{f}_a, \quad a \in A_v \quad [3.14b]$$

$$\sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a x_{ij}^k - y \leq \bar{f}_a + \bar{\Lambda}_a, \quad a \in A_v \quad [3.14c]$$

$$\bar{x} \geq 0, \quad \bar{y} \geq 0, \quad [3.14d]$$

where,

$$(\tilde{c}_{ij}^k)_a = \begin{cases} M_1 c_{ij}^k & k \notin \Pi_{ij} \text{ and } a \in A_v \\ c_{ij}^k & \text{otherwise.} \end{cases} \quad [3.14e]$$

Similarly, if there is a target trip table used to guide the solution, the model LPI(TT) can be modified as follows.

$$\text{LPIM(TT): Minimize} \quad \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} \tilde{c}_{ij}^k x_{ij}^k + M e_{A_v} \cdot y + M_\sigma \sum_{(i,j) \in \overline{OD}} (Y_{ij}^+ + Y_{ij}^-) \quad [3.15a]$$

$$\text{subject to} \quad \sum_{k=1}^{n_{ij}} x_{ij}^k + (Y_{ij}^+ - Y_{ij}^-) = Q_{ij} \quad \forall (i,j) \in \overline{OD} \quad [3.15d]$$

$$\bar{x} \geq 0, \quad \bar{y} \geq 0, \quad Y^+ \geq 0, \quad Y^- \geq 0, \quad [3.15e]$$

and the constraints [3.14b], [3.14c], [3.14e].

**Remark 3.8** *As indicated before, centroidal and centroid connectors have a certain artificial intention. It would be nice if the land-use character could be considered in the capacity-based approach.*

The models formulated thus far are typically based on the system optimum hypothesis, unless if the observed flows, and hence the derived objective coefficients, correspond to a user-equilibrium. In reality, this hypothesis is questionable for some economical and psychological reasons. On one hand, if the users are not very cooperative so that the total system cost differs from the optimal objective value, no sensitivity analysis has been done on how “far” the real solution  $(\bar{x}_r, \bar{y}_r, Y^+, Y^-)$  would be from an optimal one. On the other hand, even if the system optimal hypothesis is accepted, there is no mechanism provided by the models themselves to prevent the case of alternative optimal solutions. Therefore, the obtained optimal solution may depend on the trajectory of the search algorithm used. Therefore, the following postprocessing may be used in addition to solving these linear programs. For the sake of illustration, the following discussion considers LPI as the example model. The idea and techniques are applicable to the other models as well.

Let  $(\bar{x}_r, \bar{y}_r)$  be the real O-D distribution, and  $z(x_r, y_r)$  be the corresponding penalized system cost (objective value in [3.8a]). Instead of the assumption that the system is in its optimal state, assume that  $z(x_r, y_r) \in [\tilde{z}, \tilde{z} + \Theta]$ , where  $\tilde{z}$  is the optimal objective value of LPI, and  $\Theta \geq 0$  is a constant. In this case, the solution  $(\bar{x}_r, \bar{y}_r)$  is called  $\Theta$ -suboptimal. Obviously,  $(\bar{x}_r, \bar{y}_r)$  is in  $\Theta$ -suboptimal if and only if it is a member of the following constraint set:

$$S_{\Theta} \equiv \left\{ (x, y) : \begin{array}{l} z(x, y) \in [\tilde{z}, \tilde{z} + \Theta] \\ \bar{f}_a \leq \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a x_{ij}^k + y_a, a \in A_v \\ \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_a x_{ij}^k - y_a \leq \bar{f}_a + \bar{\Lambda}_a, a \in A_v \\ (x, y) \geq 0 \end{array} \right\}, \quad [3.16a]$$

where

$$z(x, y) = \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} \tilde{c}_{ij}^k x_{ij}^k + M e_A \cdot y \quad . \quad [3.16b]$$

Since the cost function has positive coefficients, the set  $S_{\Theta}$  defined in [3.16a] is bounded, and is a polytope. Theoretically, any point  $(\bar{x}, \bar{y})$  in  $S_{\Theta}$  can be represented as a convex combination of the extreme points of this polytope, and the representation could be interpreted as that the user might choose different trip patterns with certain probabilities. In practice, however, searching for these extreme points is very computational expensive, and furthermore, even if all of the extreme points have been identified, it is still a major issue on how to determine the pattern of this “probability distribution”. Since a prior trip table usually gives a useful guideline on the behavior of the user trip pattern, it is natural to consider the following problem.

$$\text{Minimize} \quad d(x, Q) \quad [3.17a]$$

$$\text{subject to} \quad x \in S_{\Theta} \quad [3.17b]$$

where  $Q$  is a given vector representing the target trip table, and  $d(\bullet, \bullet)$  is a *convex function* that measures the “*distance*” between two (partial) O-D trip tables. The same idea, as discussed in Remark 3.6, has been used in the development of subproblem LPISP defined in [3.10] and the model LPI(TT), where the distance function  $d(\bullet, \bullet)$  is induced via the  $L^1$ -norm and  $\Theta$  is set to be zero in LPISP. Although the  $L^1$ -norm-based optimal problem can be easily converted to a linear programming problem, since its objective function is not strictly convex, the resulting

optimal solution may not be unique over the polytope  $S_\Theta$ . These shortcomings can be overcome by considering the following strictly convex optimization problems.

(a) Least Squares Suboptimal Problem:

$$\begin{array}{ll}
 \text{LSSP} & \text{Minimize} \quad \sum_{(i,j) \in \overline{OD}} \left( \sum_{k=1}^{n_{ij}} x_{ij}^k - Q_{ij} \right)^2 \\
 & \text{subject to} \quad x \in S_\Theta.
 \end{array} \tag{3.18}$$

(b) Most Likely Suboptimal Problem.

$$\begin{array}{ll}
 \text{MLSP} & \text{Minimize} \quad \sum_{(i,j) \in \overline{OD}} \left( \sum_{k=1}^{n_{ij}} x_{ij}^k \right) \cdot \left( \log \left( \frac{\sum_{k=1}^{n_{ij}} x_{ij}^k}{Q_{ij}} \right) - 1 \right) \\
 & \text{subject to} \quad x \in S_\Theta.
 \end{array} \tag{3.19}$$

Since the subset  $S_\Theta$  is non-empty for any non-negative  $\Theta$ , given that LPI has an optimum, and the objective functions of above problems are strictly convex, the following lemma is obtained.

**LEMMA 3.7** *Assume that the target trip table has cell values defined for every feasible O-D pair  $(i, j) \in \overline{OD}$ , that is,  $\overline{OD} \equiv OD$ . Then, for any given  $\Theta \geq 0$ , each of the problem LSSP and MLSP has a unique optimal solution.*

**Remark 3.9** *Similar to the model LINKOD reviewed in Chapter 2, both of the above models, LSSP and MLSP, have a bilevel programming structure. This poses computational difficulties for large-scale networks. An alternative approach could be to accommodate the objective function in [3.18] or [3.19] with the objective function of the linear programming problem LPIM, as done for model LPIM(TT) and [3.12]. Despite the fact that this treatment reduces a bilevel programming structure to a single level programming problem, the complexity of the objective function in the problem, on*

the other hand, is increased. No evidence shows that this simplification indeed improves the computational effort required to solve the problem.

**Remark 3.10** *Theoretically, linear programming problems can be solved in a finite number of steps, and the exact optimal solution may be found only via infinite steps of convergent approximations for nonlinear programming problems. On the other hand, since the objective functions of problems LSSP and MLSP are uniformly strict convex, i.e., the spectrum of their Hessian (matrices) are located in a finite positive interval, the distance from a feasible solution to the optimal set can be estimated via the difference between their objective values. This property would be very useful, if a certain degree of alteration (from the optimal solutions) is allowed, to determine a reasonable termination criterion for approximately solving the problem.*

## 3.6 Summary and Comments

In order to account for uncertainties in the observed link volumes, intervals having band widths  $\bar{\Lambda}$  and artificial variables  $\bar{y}$  are used to represent the possible positive and negative deviations in the flow data. The traditional user-equilibrium principle is modified in a comparable manner, by not penalizing the costs on bands about the shortest paths. This leads to a linear programming approach LPI defined by [3.8] for estimating the O-D trip tables from given counts. An alternate model LPI(TT) is introduced to accommodate a target trip table to guide optimal solutions. This model, when the relative importance factor  $M_{\sigma}$  is sufficient small, is equivalent to another bilevel linear programming problem LPI-LPSP.

Problems considered in Section 3.5 are more philosophical rather than technical. The choice of treating missing volumes on links via preprocessing is motivated by the computational effort required for practical networks. The evaluation of these techniques as compared with other approaches should rely on testing results, rather



than on the theory alone. The introduction of non-system optimal concepts, as well as the use of nonlinear programming procedures in solving the subproblems, is motivated from two considerations. First, it is believed that, in the real-world, the traffic system is mostly in its sub-system optimum state, rather than the conventional user-optimum state hypothesis. Another motivation of this concept comes from the attempt to rectify the “sparse” characteristics of extreme optimal solutions that result from linear programming approaches.

Another very important issue on modeling and analysis is how to implement the model formulated above in practice, including the procedure (algorithm) on how to find optimal solutions, as well as how to convert the algorithm into a computer code. These issues will be considered in the next chapter.