Hindawi Publishing Corporation
Discrete Dynamics in Nature and Society
Volume 2015, Article ID 827094, 13 pages
http://dx.doi.org/10.1155/2015/827094

Research Article

Estimating the Capacity of Urban Transportation Networks with an Improved Sensitivity Based Method

Muqing Du,[1] Xiaowei Jiang,[2] and Lin Cheng[2]

[1] *College of Civil and Transportation Engineering, Hohai University, 1 Xikang Road, Nanjing 210098, China*
[2] *School of Transportation, Southeast University, 35 Jinxianghe Road, Nanjing 210096, China*

Correspondence should be addressed to Muqing Du; dumuqing@gmail.com

Received 7 July 2014; Accepted 18 August 2014

Academic Editor: Yongjun Shen

The throughput of a given transportation network is always of interest to the traffic administrative department, so as to evaluate the benefit of the transportation construction or expansion project before its implementation. The model of the transportation network capacity formulated as a mathematic programming with equilibrium constraint (MPEC) well defines this problem. For practical applications, a modified sensitivity analysis based (SAB) method is developed to estimate the solution of this bilevel model. The high-efficient origin-based (OB) algorithm is extended for the precise solution of the combined model which is integrated in the network capacity model. The sensitivity analysis approach is also modified to simplify the inversion of the Jacobian matrix in large-scale problems. The solution produced in every iteration of SAB is restrained to be feasible to guarantee the success of the heuristic search. From the numerical experiments, the accuracy of the derivatives for the linear approximation could significantly affect the converging of the SAB method. The results also show that the proposed method could obtain good suboptimal solutions from different starting points in the test examples.

1. Introduction

As the rapid growth of urbanization, the population and economy of most cities in the developing countries or regions are going through significant changes. The fast development of transportation infrastructures in these areas gives rise to a quick change on the travel behaviors. Along with the city's expansion, an inevitable thing is to design new transportation system which is capable of meeting the future development in land-using and population growth. However, under fast-changing of travel demand, the conventional forecasting methods (e.g., "four-step") could not provide a straightforward evaluation to the new designed transportation network such as on whether the new network is suitable for the travel demand in the target year or how many trips can be accommodated. Thus, to help the government to make decision on the expansion or construction project or pre-assessment of social benefits, the throughput, or "capacity," of the given transportation network is of practical meaning to be estimated before planning implementation.

Capacity is a commonly used property to represent the maximum flows that can pass through the link or node in transportation system. In an attempt to address the question of what is the maximum attainable throughput of the given network, the concept of capacity is employed as an important measurement for transportation system evaluation. It is able to reflect how much traffic demand can be accommodated by a given transportation system. Thus, efficient policy for land use or traffic restraint and growth can be established in advance. According to the conventional network flow theory, the capacity problem is stated to find the maximum flows that can be sent from a specified source node to another specified sink node without exceeding the capacity of any link. This well-known problem is extended to the multicommodity and is widely used in freight transport. However, when modeling the capacity of urban transportation network, the problem becomes quite complex. Noted by Yang et al. [1], travelers in urban transportation network can choose their routes and their trip costs increase with increasing flow as a result of congestion. Besides, multiple origin and destination (O-D) pairs exist and the flows between distinct O-D pairs cannot be exchanged in passenger transportation system. These differences make the modeling of transportation network capacity complex, and the intriguing problem is also hard to solve.

The most popular formulation of the transportation network capacity is the bilevel programming model, which maximizes the traffic flows under the equilibrium constraint. Wong and Yang [2] first incorporated the *reserve capacity* concept into a traffic signal control network. This concept was widely extended in the study of signal controlled networks [3–5]. The reserve capacity is defined as the largest multiplier applied to a given O-D demand matrix that can be allocated to a network, so the solution is significantly affected by the predetermined O-D matrix. However, it is unrealistic to assume that all O-D flows increase in the same rate, especially for the areas under rapid changing. If the predetermined distribution proportion is far from the future tendency, the solution will be of little use. Consequently, in order to reflect the differences of the future development of each urban subarea, Yang et al. [1] considered that the new increased O-D demand pattern should be variable both in level and in distribution, while the distribution of the current trips would be relatively fixed. Thereby they introduced the equilibrium trip distribution/assignment model with variable destination costs (ETDA-VDC) [6] to capture this characteristic for network capacity estimation. Based on this model, Kasikitwiwat and Chen [7] proposed the concepts and models of the *ultimate* and *practical capacity*. The former is used for the new network without any current flow, while the latter is the same as Yang's model. Then, Chen and Kasikitwiwat [8] used the practical network capacity model to describe the limited flexibility of transportation networks. According to the literatures, the concept of the practical capacity model is more fully functional and preferred, as it takes both the current demand pattern and the variability of future growth into consideration.

In order to solve the various bilevel capacity models, the SAB algorithm is generally employed. This SAB algorithm for bilevel programing was first presented in Friesz et al. [9]. It is heuristic and depends absolutely on the derivative information produced by the sensitivity analysis of the lower-level problem [10]. Benefiting from the rich achievements in the study of the sensitivity analysis for equilibrium models [11–18], SAB algorithm has been widely utilized in the optimization problems of equilibrium network flows, such as traffic signal control [3, 5, 19, 20] and network design [21–23], as well as network capacity [1, 3–5, 24]. But due to the difficulty of the sensitivity analysis for ETDA-VDC model, Yang et al. [1] used an iterative estimation-assignment (IEA) algorithm [11] instead to solve the transportation network capacity problem. Later, Kasikitwiwat and Chen [7] and Chen and Kasikitwiwat [8] selected using a genetic algorithm to solve the problem in a very small network. However, since the complexity of the network capacity problem, the global optimization algorithms (e.g., genetic algorithm or simulated annealing) can hardly find the exact solution to the capacity problem in larger networks, and the computation time could be intolerably long. By contrast, SAB algorithm has the property of fast convergence which makes the computation terminate at a local optimum within a considerable time. Nevertheless, the calculation issue of the matrix inversion still limits the applications of the SAB method. To address this problem, we developed an effective method by simplifying the matrix inversion in the sensitivity analysis approach, which will take much less memory space, so the capacity of the real transportation networks could be estimated.

In this study, Yang's formulation [1] for the transportation network capacity model is employed to describe the practical capacity of the urban road system. In an attempt to estimate the capacity of the real road networks, a series of improvements are taken to the SAB method to make the heuristic search successfully converge to a relatively better suboptimal solution. Firstly, the OB algorithm [25] is modified for the solution of the lower-level ETDA-VDC model. Then, the restriction sensitivity analysis approach for the ETDA-VDC model [18] is employed in and improved on the expressions so as to deal with the large-scale problems. Besides, the solution update strategy is modified on the step-size adaption, which ensures the entire heuristic search to converge to a local optimum. Finally, numerical experiments are implemented to show the efficiency and capability of the proposed SAB method.

2. Road Network Capacity Model

It should be noted that the boldface type of the Notation section represents the corresponding column vectors in the remainder of this paper.

2.1. Model Formulation. Conversional methods, like the reserve capacity model, evaluate the capacity of transportation networks by assuming that the travel demand increases with a determined distribution proportion, which is usually far from the regularity and underestimates the results. In order to evaluate zonal development potential and equilibrium network capacity more appropriately, Oppenheim's definitions on the behaviors of the existing travel demand and the additional demand are introduced as follows.

(i) *The existing demand*, denoted by e_{pq}, has predetermined origins and destinations. The pattern of the existing demand is formed during the past long term, so its distribution is going to be relatively stable and can be regarded as fixed. The existing demand only changes routes to optimize the travel cost.

(ii) *The additional demand*, d_{pq}, is variable. The new generated demands from residential area can decide their daily travels without the constraints of either destination or route choices. But the behavior of the additional demand still follows the rule that selects the destinations which maximize the "utility" of the trips.

The utility could include the destination attractiveness, the cost along traveling route, and other factors. The attractiveness of destination is determined by the congestion at destination and the expenses for the activity in that area.

In network capacity model, the utility from origin p to destination q is formulated as $U_{pq} = -(M_q + \tau_{pq})$, in which τ_{pq} is the travel cost from p to q, and M_q denotes the destination cost which could be a decrease function of the total additional trip attraction, D_q, at destination q. Besides, the destination choices of the travelers at each origin p are assumed to have certain randomness. Thus, the conditional probability that an individual will choose destination p is derived by using the standard logit function, so the O-D travel demand is conducted by

$$d_{pq} = O_p \frac{\exp\left\{-\theta\left(\tau_{pq} + M_q\right)\right\}}{\sum_{k\in Z}\exp\left\{-\theta\left(\tau_{pk} + M_q\right)\right\}}, \quad \forall p \in Z,\ q \in Z_p. \tag{1}$$

Thus, with the objective to maximize the additional demand under the above travel behavior regularity and certain physical constraints, the typical road network capacity model is formulated as the following bilevel programming problem.

Upper-level problem is as follows:

$$\operatorname*{Max}_{\mathbf{O}} \quad \sum_{p\in Z} O_p, \tag{2}$$

$$\text{s.t.} \quad v_a(\mathbf{O}) \le C_a, \quad \forall a \in A, \tag{3}$$

$$O_p = \sum_{q\in Z_p} d_{pq}(\mathbf{O}) \le O_p^{\max} - \overline{O}_p, \quad \forall p \in Z, \tag{4}$$

$$D_q = \sum_{p\in Z_q} d_{pq}(\mathbf{O}) \le D_q^{\max} - \overline{D}_q, \quad \forall q \in Z, \tag{5}$$

$$O_p \ge 0, \quad \forall p \in Z, \tag{6}$$

where $d_{pq}(\mathbf{O})$ and $v_a(\mathbf{O})$ are obtained by solving the ETDA-VDC problem in lower-level problem.

Lower-level problem, ETDA-VDC model, is as follows:

$$\begin{aligned}\operatorname*{Min}_{(\mathbf{f},\mathbf{h},\mathbf{d})} \quad & \sum_a \int_0^{v_a} t_a(x)\,dx + \frac{1}{\theta}\sum_{p\in Z}\sum_{q\in Z_p} d_{pq}\left(\ln d_{pq} - 1\right) \\ & + \sum_{q\in Z}\int_0^{D_q} M_q(y)\,dy,\end{aligned} \tag{7}$$

$$\text{s.t.} \quad \sum_{q\in Z_p} d_{pq} = O_p, \quad \forall p \in Z, \tag{8}$$

$$\sum_{r\in R_{pq}} h_r^{pq} = e_{pq}, \quad \forall p \in Z,\ q \in Z_p, \tag{9}$$

$$\sum_{r\in R_{pq}} f_r^{pq} = d_{pq}, \quad \forall p \in Z,\ q \in Z_p, \tag{10}$$

$$v_a = \sum_p\sum_q\sum_r \left(f_r^{pq} + h_r^{pq}\right)\delta_{a,r}^{pq}, \quad \forall a \in Am, \tag{11}$$

$$d_{pq} \ge 0, \quad \forall p \in Z,\ q \in Z_p m, \tag{12}$$

$$f_r^{pq} \ge 0, \quad \forall p \in Z,\ q \in Z_p,\ r \in R_{pq}, \tag{13}$$

$$h_r^{pq} \ge 0, \quad \forall p \in Z,\ q \in Z_p,\ r \in R_{pq}. \tag{14}$$

The upper-level problem defines a maximal trip production model. The objective is to maximize the summation of the additional trip production at origins. Equation (3) represents that the traffic flow on every link should not exceed its capacity. Constraints (4) and (5) are the limitation of the zonal trip production and attraction. They mean the number of trips generated and attracted at each traffic zone should be limited by some upper bounds, namely, $O_p^{\max}$ and $D_q^{\max}$, respectively.

The lower-level problem is the ETDA-VDC model. The objective function (7) indicates the choice behavior of both the existing and additional travel demand. Constraint (9) shows that the amount of the existing flows is fixed for each O-D, while constraints (8) and (10) show that the additional flows are only restrained at the origin productions. The relationship between the link flow and route flow is represented in (11). All the variables must be nonnegative, that is, constraints (12)–(14). The lower-level problem is a combined distribution and assignment model.

This bilevel model was first presented in work by Yang et al. [1]. Because of the advantages on the formulation of the travel demand growth, it was continually used in later researches as a typical model for the road network capacity concept. The remaining part of this study focuses on the solution of this model in the real-sized road networks.

3. Sensitivity Based Heuristic Algorithm

This section presents an improved version of the SAB algorithm for the solution of the road network capacity model. To overcome the drawbacks of the conventional SAB algorithm [9] that cannot be applied to any real-sized network for capacity estimations, the following improvements are carried out.

(i) The lower-level ETDA-VDC model is fast solved by a modified OB algorithm to produce a high accurate solution.

(ii) The rectified sensitivity analysis method for the ETDA-VDC model is simplified on the calculation of the matrix inverse to be applicable for large-scale problems.

(iii) The solution update of the heuristic search is improved by step-size adaption in order to ensure that the SAB algorithm can converge to a local optimum.

Correspondingly, a series of techniques is proposed in this section, so our modified SAB algorithm will be capable to solve the bilevel road capacity model efficiently.

3.1. Origin-Based Algorithm for the ETDA-VDC Model. In the standard SAB search, the lower-level ETDA-VDC model

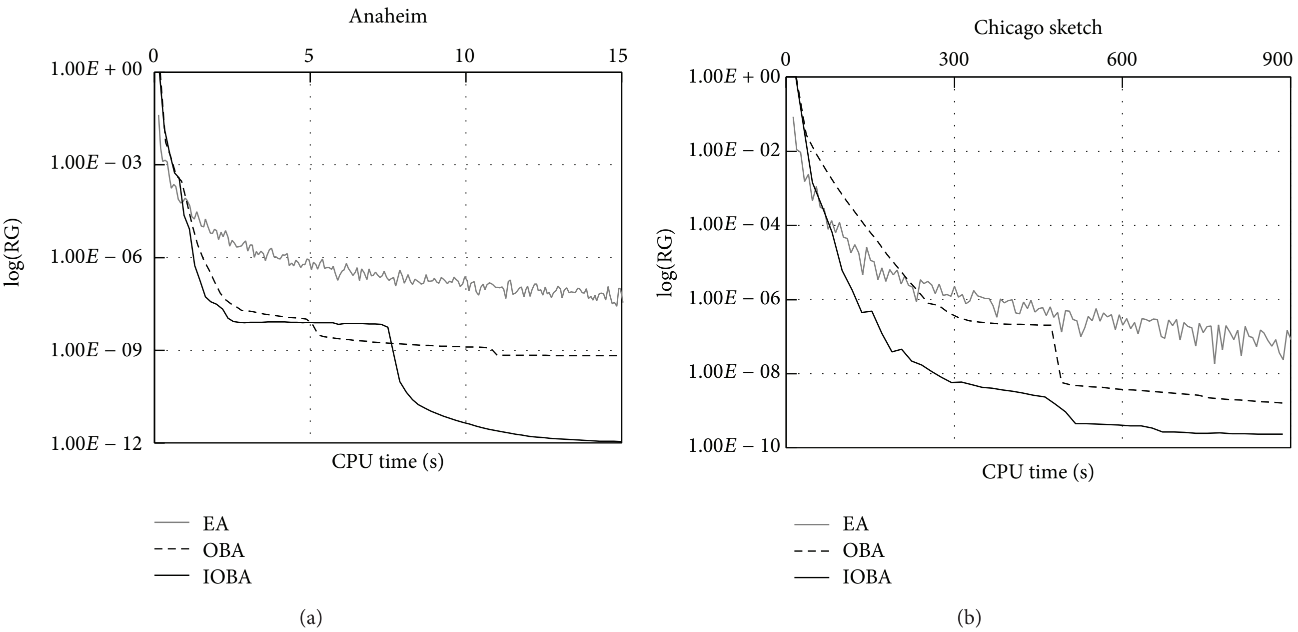

FIGURE 1: Relative gap versus CPU time.

should be solved in every iteration to conduct an equilibrium traffic flow pattern, namely, the solution to the lower-level problem. According to recent researches on the traffic assignment problems, the OB algorithm is demonstrated to be one of the state-of-the-art algorithms [26]. In addition, from the results of the OB algorithm, the set of all equilibrated routes can be easily extracted, which will be a precondition for the restriction sensitivity analysis approach in next step of SAB algorithm.

The OB algorithm uses the *origin-based approach proportions*, $\boldsymbol{\alpha}_p = \{\alpha_{ap}\}$, as its main solution variables, where α_{ap} represents the proportion of flow that comes from origin p through link a. The approach proportions are updated by shifting flows within the restricting subnetwork, A_p. The details of the OB algorithm can be referred to in Bar-Gera [26]. Using the approach proportions, the OB algorithm is able to store the route flow information with significantly less memory than the route-based algorithms and achieve high-accuracy solutions compared to the link-based algorithms.

In the process of the OB algorithm, it starts with trees of minimum cost routes from origins as restricting networks. Then steps of updating restricting network and approach proportions are implemented for each origin separately. An "inner loop" is performed subsequently to accelerate convergence, in which the origin-based link flows are updated while keeping the restricting subnetworks fixed. The inner loop is useful, because the restricting subnetworks tend to stabilize fairly quickly but updating the restricting subnetworks requires more computational effort [26, 27].

The OB algorithm for the combined trip distribution and assignment problems requires an additional step to update the O-D flows while keeping the route proportions fixed. Taking Evans' algorithm as a reference, we modified the original OB algorithm for the solution of the ETDA-VDC model, in which the step size of the O-D flow update is obtained by solving a one-dimension search problem. The proposed valgorithm is summarized as shown in Algorithm 1.

The performance of Algorithm 1 was testified on different test networks. The characteristics of these networks are obtained from http://www.bgu.ac.il/~bargera/tntp/. The converging processes on two well-known networks are demonstrated in Figure 1, where the relative gap is defined as RG $= |F^n - F^{n-1}|/F^n$ (F^n is the value of objective function (7) at nth main loop). It shows that the converging of the improved OB algorithm (IOBA) is faster and more stable than Evans' algorithm (EA) or the original OB algorithm (OBA) steps in Bar-Gera and Boyce [25].

3.2. Sensitivity Analysis of the ETDA-VDC Model. On the basis of an equilibrium solution of the lower-level model, the derivatives of the lower-level decision variables with respect to the upper-level ones should be produced for linearly approximating the whole bilevel model in next step. The derivative expressions can be obtained by employing the restriction sensitivity analysis approach for the ETDA-VDC model [18]. The restriction approach reduces the original network to a restricting one in which the nonuniqueness difficulty will be overcome, and thus the full derivatives of the link flows in the primal problem can be solved. The necessary results from the sensitivity analysis of the ETDA-VDC model are presented in this section without proof.

Initialization:

Set $n := 0$, determine the initial link cost $\mathbf{t}^0$ and destination cost $\mathbf{M}^0$ by setting $\mathbf{x}^0 = 0$ and $\mathbf{d}^0 = 0$, respectively.

For each origin $p \in Z$ do

(i) Find the tree of minimum cost routes rooting from p. Let A_p be the minimum route tree. Denote the minimum route cost to destination $q \in Z_p$ by τ^0_{pq}, and choose a minimum cost route from p to q.

(ii) Compute the initial variable O-D demands by

$$d^0_{pq} = O_p \frac{\exp\left[-\theta\left(\tau^0_{pq} + M^0_q\right)\right]}{\sum_k \exp\left[-\theta\left(\tau^0_{pk} + M^0_k\right)\right]}, \quad \forall q \in Z_p$$

(iii) For $q \in Z_p$, assign the entire O-D demand e_{pq} and d^0_{pq} to the minimum cost route r from p to q, and obtain initial link flow x^0_{ap}.

(iv) Update the link costs using the initial link flows.

(v) Initialize the origin-based approach proportions α^0_{ap}.

Main Loop:

Given the current variable O-D demand $\mathbf{d}^n$ obtained in $(n-1)$th-iteration:

For $n = 1$ to number of main iterations (I_{Main})

for each p in Z do

Update restricting network A_p

Update origin-based approach proportions $\boldsymbol{\alpha}_p$

end for

Inner Loop:

for $m = 1$ to number of inner iterations (I_{inner})

for each p in Z do

Update origin-based approach proportions $\boldsymbol{\alpha}_p$

end for

end for

Update* O-D *flows, retain origin-based approach proportions:

Given the origin-based approach proportions $\boldsymbol{\alpha}^n_p$, link flows $\mathbf{v}^n$ and link costs $\mathbf{t}^n$ obtained in the steps above:

(i) For each $q \in Z$, compute the destination cost M^n_q by O-D demand d^n_{pq} and e_{pq}.

(ii) Find the set of auxiliary trip demands $\hat{d}^n_{pq}$ by solving the following logit distribution model:

$$\hat{d}^n_{pq} = O_p \frac{\exp\left[-\theta\left(\tau^n_{pq} + M^n_q\right)\right]}{\sum_k \exp\left[-\theta\left(\tau^n_{pk} + M^n_k\right)\right]} \quad \forall p \in Z,\ q \in Z_p$$

(iii) Calculate the auxiliary traffic flow $\hat{v}^n_a$ on each link a with the approach proportions α^n_{ap}

(iv) Let $(\mathbf{v}^{\lambda(n)}, \mathbf{d}^{\lambda(n)}) = (1-\lambda)(\mathbf{v}^n, \mathbf{d}^n) + \lambda(\hat{\mathbf{v}}^n, \hat{\mathbf{d}}^n)$ and solve the one-dimensional search problem defined as follows to obtain the step size $\lambda^* \in [0,1]$

$$\min_{0\le\lambda\le1} F(\lambda) = \sum_{a\in A} \int_0^{v_a^{\lambda(n)}} t_a(x)\,dx + (1/\theta) \sum_{p\in Z} \sum_{q\in Z_p} d^{\lambda(n)}_{pq} (\ln d^{\lambda(n)}_{pq} - 1) + \sum_{q\in Z} \int_0^{D_q^{\lambda(n)}} M_q(y)\,dy$$

(v) Set $(\mathbf{v}^{n+1}, \mathbf{d}^{n+1}) := (\mathbf{v}^{\lambda(n)}, \mathbf{d}^{\lambda(n)})$ and check for convergence. Terminate if the convergence criterion is satisfied; otherwise, update total link flows and link costs, set $n := n+1$ and start a new iteration of the main loop.

End for

Algorithm 1

In the restricting problem the derivatives of the model solutions with respect to the input parameters are derived from

$$\nabla_{\mathbf{O}} \mathbf{x}(\boldsymbol{\varepsilon}) = \left[\mathbf{J_x}\right]^{-1} \left[-\mathbf{J_O}\right], \tag{15}$$

where $\mathbf{J_x}$ is an invertible Jacobian matrix with respect to the solution variables $\mathbf{x} = (\mathbf{f}^{BT}, \mathbf{h}^{BT}, \mathbf{d}^T, \boldsymbol{\lambda}^T, \boldsymbol{\mu}^T, \mathbf{u}^T)^T$, in which $\mathbf{f}^B$ and $\mathbf{h}^B$ are the basic variables of the additional and existing route flows, respectively. $\mathbf{J_O}$ is the Jacobian matrix with respect to the additional zonal productions, $\mathbf{O}$. Here, $\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{u}$ are the Lagrange multipliers associated with the conservation constraints (8)–(10). $\mathbf{O}$ is referred to as an input parameter for the ETDA-VDC model. It is in terms of the perturbation $\boldsymbol{\varepsilon}$. The expressions of $\mathbf{J_x}$ and $\mathbf{J_O}$ are as follows:

$$\mathbf{J_x} = \begin{bmatrix} \nabla^2_{\mathbf{f}} L & \nabla_{\mathbf{f,h}} L & \mathrm{O} & \mathrm{O} & \mathrm{O} & -\Lambda^T_{\mathbf{f}} \\ \nabla_{\mathbf{h,f}} L & \nabla^2_{\mathbf{h}} L & \mathrm{O} & \mathrm{O} & -\Lambda^T_{\mathbf{h}} & \mathrm{O} \\ \mathrm{O} & \mathrm{O} & \nabla^2_{\mathbf{d}} L & -\Phi^T & \mathrm{O} & I \\ \mathrm{O} & \mathrm{O} & -\Phi & \mathrm{O} & \mathrm{O} & \mathrm{O} \\ \mathrm{O} & -\Lambda_{\mathbf{h}} & \mathrm{O} & \mathrm{O} & \mathrm{O} & \mathrm{O} \\ -\Lambda_{\mathbf{f}} & \mathrm{O} & I & \mathrm{O} & \mathrm{O} & \mathrm{O} \end{bmatrix}, \tag{16}$$

$$\mathbf{J_O} = \left[\mathrm{O}\ \ \mathrm{O}\ \ \mathrm{O}\ \ I\ \ \mathrm{O}\ \ \mathrm{O}\right]^T, \tag{17}$$

where $\nabla_{\mathbf{f}}^2 L = \Delta_{\mathbf{f}}^T \nabla_{\mathbf{v}} \mathbf{t}(\mathbf{v}, \boldsymbol{\varepsilon}) \Delta_{\mathbf{h}}$, $\nabla_{\mathbf{h},\mathbf{f}} L = \Delta_{\mathbf{h}}^T \nabla_{\mathbf{v}} \mathbf{t}(\mathbf{v}, \boldsymbol{\varepsilon}) \Delta_{\mathbf{f}}$, $\nabla_{\mathbf{f}}^2 L = \Delta_{\mathbf{h}}^T \nabla_{\mathbf{v}} \mathbf{t}(\mathbf{v}, \boldsymbol{\varepsilon}) \Delta_{\mathbf{h}}$, $\nabla_{\mathbf{f},\mathbf{h}} L = \Delta_{\mathbf{f}}^T \nabla_{\mathbf{v}} \mathbf{t}(\mathbf{v}, \boldsymbol{\varepsilon}) \Delta_{\mathbf{h}}$, and $\nabla_{\mathbf{d}}^2 L = (1/\theta)$ $\text{diag}(1/d_{pq}) + \Psi^T \nabla_{\mathbf{d}} \mathbf{M}$; $\text{diag}(1/d_{pq})$ is a diagonal matrix with $1/d_{pq}$ as its diagonal elements. The superscript "T" represents the transposed matrix. "I" indicates the identity matrix. The link cost function, $\mathbf{t}(\cdot)$, should be strongly monotonically defined in $\mathbf{v}$, which guarantees the uniqueness of the equilibrium solution and is also a precondition for the sensitivity analysis approach. Utilizing the restriction approach, the derivative, $\nabla_{\mathbf{O}} \mathbf{x}(\boldsymbol{\varepsilon})$, is first conducted in the restricting problem and then $\nabla_{\mathbf{O}} \mathbf{v}$, which is defined in the original one, can be obtained by

$$\nabla_{\mathbf{O}} \mathbf{v} = \Delta_{\mathbf{f}} \nabla_{\mathbf{O}} \mathbf{f}^B + \Delta_{\mathbf{h}} \nabla_{\mathbf{O}} \mathbf{h}^B. \tag{18}$$

Equation (18) indicates that the variations on the link flows can be just represented by the changes on the basic route flows, which is the rationale of the restriction sensitivity analysis approach.

From (15), the inverse of the matrix $\mathbf{J_x}$ is required to be worked out. However, the calculation of the inverse matrix is costly for both computation time and storage space in numerical computation for large-scale problems. The dimension of $\mathbf{J_x}$ (from (16)) could be about five times of the number of O-D pairs. So the elements of $\mathbf{J_x}$ could totally take approximately $O(25n^4)$ space, where n is the amount of the traffic zones, and thus even a medium network with hundreds of origins will conduct a Jacobian matrix with over one hundred thousand dimensions. For example, in a small network with about 30 traffic zones (900 O-D pairs), nearly 400 MB RAM should be expended on the storage of the Jacobian matrix (if each element uses a double type). Consequently, the conventional sensitivity analysis approach for the combined model is very difficult to be utilized in the large-scale problems. Besides, the direct calculation for the inverse of a large matrix is pretty inefficient and inaccurate in practice.

To avoid the defects of deriving the inverse matrix directly, the usual way to solve the following linear equation as an alternative. Equation (15) which produces the sensitivity results can be rewritten as

$$\mathbf{J_x} \cdot \nabla_{\mathbf{O}} \mathbf{x} = -\mathbf{J_O}. \tag{19}$$

Then, $\nabla_{\mathbf{O}} \mathbf{x}$ is derived by solving the following series of linear equations:

$$\mathbf{b}_i = \mathbf{J_x} \cdot \mathbf{z}_i, \tag{20}$$

where $\mathbf{b}_i$ is the ith column vector in matrix of $[-\mathbf{J_O}]$ and $\mathbf{z}_i$ is the ith column vector in $\nabla_{\mathbf{O}} \mathbf{x}$. Let $\mathbf{e}_i$ be a unit vector with one in the ith position and zeros elsewhere. The length of $\mathbf{e}_i$ is equal to the column number of $\mathbf{J_O}$. Thus,

$$\mathbf{b}_i = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ -\mathbf{e}_i \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \qquad -\mathbf{J_O} = [\ldots, \mathbf{b}_i, \ldots], \tag{21}$$

where $\mathbf{0}$ in bold is zero vectors associate with the zero block in $\mathbf{J_O}$.

To solve (20), $\mathbf{J_x}$ is premultiplied by the matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ in sequence and postmultiplied by $\mathbf{K}_1^T$ and $\mathbf{K}_2^T$ simultaneously, which is equivalent to making the elementary transformation of matrix $\mathbf{J_x}$. Consider

$$\mathbf{K}_1 = \begin{bmatrix} I &&&&& \\ & I &&&& \\ && I &&& \\ &&& I & \Phi & \\ &&&& I & \\ &&&&& I \end{bmatrix},$$
$$\mathbf{K}_2 = \begin{bmatrix} I &&&&& \\ & I &&&& \\ && I &&& \\ &&& I && \\ &&&& I & \\ && -\nabla_{\mathbf{d}}^2 L^{-1} &&& I \end{bmatrix}. \tag{22}$$

Thus,

$$\mathbf{K}_2 \mathbf{K}_1 \mathbf{J_x} \mathbf{K}_1^T \mathbf{K}_2^T = \begin{bmatrix} \nabla_{\mathbf{f}}^2 L & \nabla_{\mathbf{f},\mathbf{h}} L & \mathrm{O} & -\Lambda_{\mathbf{f}}^T \Phi^T & \mathrm{O} & -\Lambda_{\mathbf{f}}^T \\ \nabla_{\mathbf{h},\mathbf{f}} L & \nabla_{\mathbf{h}}^2 L & \mathrm{O} & \mathrm{O} & -\Lambda_{\mathbf{h}}^T & \mathrm{O} \\ \mathrm{O} & \mathrm{O} & \nabla_{\mathbf{d}}^2 L & \mathrm{O} & \mathrm{O} & \mathrm{O} \\ -\Phi \Lambda_{\mathbf{f}} & \mathrm{O} & \mathrm{O} & \mathrm{O} & \mathrm{O} & \mathrm{O} \\ \mathrm{O} & -\Lambda_{\mathbf{h}} & \mathrm{O} & \mathrm{O} & \mathrm{O} & \mathrm{O} \\ -\Lambda_{\mathbf{f}} & \mathrm{O} & \mathrm{O} & \mathrm{O} & \mathrm{O} & -\nabla_{\mathbf{d}}^2 L^{-1} \end{bmatrix}. \tag{23}$$

By using the above operations on (20), we obtain

$$\mathbf{K}_2 \mathbf{K}_1 \mathbf{J_x} \mathbf{K}_1^T \mathbf{K}_2^T \left[\mathbf{K}_2^T\right]^{-1} \left[\mathbf{K}_1^T\right]^{-1} \mathbf{z}_i = \mathbf{K}_2 \mathbf{K}_1 \mathbf{b}_i. \tag{24}$$

Obviously,

$$\mathbf{y} = \left[\mathbf{K}_2^T\right]^{-1} \left[\mathbf{K}_1^T\right]^{-1} \mathbf{z}_i, \qquad \mathbf{K}_2 \mathbf{K}_1 \mathbf{b}_i = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ -\mathbf{e}_i \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{25}$$

Therefore, it is equivalent to solving the equation in terms of $\mathbf{y}$ and then deriving $\mathbf{z}_i$ by

$$\mathbf{z}_i = \mathbf{K}_1^T \mathbf{K}_2^T \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 - \nabla_{\mathbf{d}}^2 L^{-1} \cdot \mathbf{y}_6 \\ \mathbf{y}_4 \\ \mathbf{y}_5 \\ \mathbf{y}_6 + \Phi^T \cdot \mathbf{y}_4 \end{bmatrix}. \tag{26}$$

Since it is only concerned with the first three subvectors of $\mathbf{z}_i$ in the sensitivity result, we just need to find the value of

$\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$, and $\mathbf{y}_6$. From (24), the following group of equations can be obtained:

$$\nabla^2_{\mathbf{f}} L \cdot \mathbf{y}_1 + \nabla_{\mathbf{f},\mathbf{h}} L \cdot \mathbf{y}_2 - \Lambda^T_{\mathbf{f}} \Phi^T \cdot \mathbf{y}_4 - \Lambda^T_{\mathbf{f}} \cdot \mathbf{y}_6 = 0, \quad (27)$$

$$\nabla_{\mathbf{h},\mathbf{f}} L \cdot \mathbf{y}_1 + \nabla^2_{\mathbf{h}} L \cdot \mathbf{y}_2 - \Lambda^T_{\mathbf{h}} \cdot \mathbf{y}_5 = 0, \quad (28)$$

$$\nabla^2_{\mathbf{d}} L \cdot \mathbf{y}_3 = 0, \quad (29)$$

$$-\Phi \Lambda_{\mathbf{f}} \cdot \mathbf{y}_1 = -\mathbf{e}_i, \quad (30)$$

$$-\Lambda_{\mathbf{h}} \cdot \mathbf{y}_2 = 0, \quad (31)$$

$$-\Lambda_{\mathbf{f}} \cdot \mathbf{y}_1 - \nabla^2_{\mathbf{d}} L^{-1} \cdot \mathbf{y}_6 = 0. \quad (32)$$

The above equation systems can be further simplified depending on whether the incidence matrix $\Lambda_{\mathbf{h}}$ is square.

(1) If $\Lambda_{\mathbf{h}}$ is square, consider the following.

Thus, $\Lambda_{\mathbf{h}}$ will be constructed as an identity matrix; the value of $\mathbf{y}_2$ and $\mathbf{y}_3$ can be computed from (29) and (31), and thereby $\mathbf{y}_2 = \mathbf{y}_3 = 0$. From (32),

$$\mathbf{y}_6 = -\nabla^2_{\mathbf{d}} L \cdot \Lambda_{\mathbf{f}} \cdot \mathbf{y}_1. \quad (33)$$

Substituting (33) into (27), we obtain

$$\left(\nabla^2_{\mathbf{f}} L + \Lambda^T_{\mathbf{f}} \nabla^2_{\mathbf{d}} L \Lambda_{\mathbf{f}} \right) \mathbf{y}_1 - \Lambda^T_{\mathbf{f}} \Phi^T \cdot \mathbf{y}_4 = 0. \quad (34)$$

Combined with (30) we get

$$\begin{bmatrix} \nabla^2_{\mathbf{f}} L + \Lambda^T_{\mathbf{f}} \nabla^2_{\mathbf{d}} L \Lambda_{\mathbf{f}} & -\Lambda^T_{\mathbf{f}} \Phi^T \\ -\Phi \Lambda_{\mathbf{f}} & O \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_4 \end{bmatrix} = \begin{bmatrix} 0 \\ -\mathbf{e}_i \end{bmatrix}, \quad (35)$$

where

$$\nabla^2_{\mathbf{f}} L + \Lambda^T_{\mathbf{f}} \nabla^2_{\mathbf{d}} L \Lambda_{\mathbf{f}} = \begin{bmatrix} \Delta^T_{\mathbf{f}} & \Lambda^T_{\mathbf{f}} \end{bmatrix} \begin{bmatrix} \nabla_{\mathbf{v}} \mathbf{t}(\cdot) & \\ & \nabla^2_{\mathbf{d}} L \end{bmatrix} \begin{bmatrix} \Delta_{\mathbf{f}} \\ \Lambda_{\mathbf{f}} \end{bmatrix}. \quad (36)$$

Because the matrix $\begin{bmatrix} \nabla_{\mathbf{v}} \mathbf{t}(\cdot) & \\ & \nabla^2_{\mathbf{d}} L \end{bmatrix}$ is positive definite, the columns of $\begin{bmatrix} \Delta_{\mathbf{f}} \\ \Lambda_{\mathbf{f}} \end{bmatrix}$ are leaner independent from the restriction approach. It can be easily conducted that matrix (36) is invertible. Let

$$\begin{bmatrix} \nabla^2_{\mathbf{f}} L + \Lambda^T_{\mathbf{f}} \nabla^2_{\mathbf{d}} L \Lambda_{\mathbf{f}} & -\Lambda^T_{\mathbf{f}} \Phi^T \\ -\Phi \Lambda_{\mathbf{f}} & O \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & O \end{bmatrix}. \quad (37)$$

Solving (35), we can get

$$\mathbf{y}_1 = \mathbf{A}^{-1} \mathbf{B}^T \left(\mathbf{B} \mathbf{A}^{-1} \mathbf{B}^T \right)^{-1} \cdot \mathbf{e}_i. \quad (38)$$

To this extent, the only effort to produce the desired derivative results is to calculate $\mathbf{y}_1$, in which the inverse of matrices $\mathbf{A}$ and $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ should be derived. The dimension of matrix $\mathbf{A}$ is equal to the numbers of the basic routes used by the additional travel demand, which is a little more than the number of O-D pairs from our observation in computational experiments. The dimension of matrix $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ is the same as the number of origins and is not large.

Consequently, by utilizing (33) and $\mathbf{y}_2 = \mathbf{y}_3 = 0$, the first three subvectors of $\mathbf{z}_i$ can be obtained, which are separately corresponding to the derivative results, $\nabla_{\mathbf{O}} \mathbf{f}^B$, $\nabla_{\mathbf{O}} \mathbf{h}^B$, and $\nabla_{\mathbf{O}} \mathbf{d}$, in $\nabla_{\mathbf{O}} \mathbf{x}$. The expressions are

$$\begin{aligned} \nabla_{\mathbf{O}} \mathbf{f}^B &= \mathbf{A}^{-1} \mathbf{B}^T \left(\mathbf{B} \mathbf{A}^{-1} \mathbf{B}^T \right)^{-1}, \\ \nabla_{\mathbf{O}} \mathbf{h}^B &= 0, \\ \nabla_{\mathbf{O}} \mathbf{d} &= \Lambda_{\mathbf{f}} \cdot \nabla_{\mathbf{O}} \mathbf{f}^B. \end{aligned} \quad (39)$$

Thus,

$$\nabla_{\mathbf{O}} \mathbf{v} = \Delta_{\mathbf{f}} \nabla_{\mathbf{O}} \mathbf{f}^B. \quad (40)$$

(2) If $\Lambda_{\mathbf{h}}$ is not square, consider the following.

In this case, $\mathbf{y}_2 = \nabla_{\mathbf{O}} \mathbf{h}^B$ will not be equal to zero. For further simplification, we rewrite $\Lambda_{\mathbf{h}}$ as

$$\Lambda_{\mathbf{h}} = \begin{bmatrix} \Lambda'_{\mathbf{h}} & \\ & I \end{bmatrix}. \quad (41)$$

Let R' denote the set of routes associated with $\Lambda'_{\mathbf{h}}$ and the flows on these routes are represented by $\mathbf{h}'$. Repeating the derivation from (33) to (35), we can get

$$\begin{bmatrix} \nabla^2_{\mathbf{f}} L + \Lambda^T_{\mathbf{f}} \nabla^2_{\mathbf{d}} L \Lambda_{\mathbf{f}} & \nabla_{\mathbf{f},\mathbf{h}} L' & -\Lambda^T_{\mathbf{f}} \Phi^T & O \\ \nabla_{\mathbf{h},\mathbf{f}} L' & \nabla^2_{\mathbf{h}} L' & O & \Lambda'_{\mathbf{h}} \\ -\Phi \Lambda_{\mathbf{f}} & O & O & O \\ O & \Lambda'_{\mathbf{h}} & O & O \end{bmatrix} \times \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}'_2 \\ \mathbf{y}_4 \\ \mathbf{y}'_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -\mathbf{e}_i \\ 0 \end{bmatrix}, \quad (42)$$

where the superscript "$'$" indicates the matrices and variables corresponding to the routes in R'. Considering that the set of R' is fairly small for network capacity problems, the equation system is slightly larger than (35), which is also easy to be solved. The derivative of the existing route flows is computed by $\mathbf{y}'_2$. Consider

$$\nabla_{\mathbf{O}} \mathbf{f}^B = \mathbf{y}_1, \quad (43)$$

$$\nabla_{\mathbf{O}} \mathbf{h}^B = \mathbf{y}'_2, \quad (44)$$

$$\nabla_{\mathbf{O}} \mathbf{d} = \Lambda_{\mathbf{f}} \cdot \nabla_{\mathbf{O}} \mathbf{f}^B. \quad (45)$$

And thus,

$$\nabla_{\mathbf{O}} \mathbf{v} = \Delta_{\mathbf{f}} \mathbf{y}_1 + \Delta'_{\mathbf{h}} \mathbf{y}'_2. \quad (46)$$

From the above result, it can be noted that the primal problem to calculate the inverse ofthe entire matrix $\mathbf{J}_{\mathbf{x}}$ is replaced by solving a small system of linear equations.

3.3. Restriction of the Equilibrium Network. When the restriction sensitivity analysis approach was proposed [10], the restricting problem was conducted by finding an extreme point of the feasible region of the equilibrium route flow. An equivalent linear programming was provided for solution. However, this method has been verified usually failed, since the linear programming may find a degenerated solution which causes the fact that the reduced restricting problem cannot reserve the essential information and conducts a wrong result [16]. A rectified method is to find a maximum set of the linear independent columns in the coefficient matrix of the following equation system [18]:

$$\begin{bmatrix} \Delta_{\mathbf{f}}^{0} & \Delta_{\mathbf{h}}^{0} \\ \Lambda_{\mathbf{f}}^{0} & \mathrm{O} \\ \mathrm{O} & \Lambda_{\mathbf{h}}^{0} \end{bmatrix} \begin{bmatrix} \mathbf{f}^{0} \\ \mathbf{h}^{0} \end{bmatrix} = \begin{bmatrix} \mathbf{v}^{*} \\ \mathbf{d}^{*} \\ \mathbf{e} \end{bmatrix}. \tag{47}$$

The above equation system defines the feasible region of the equlibirum route flow under unique euqilibrium solution of $\mathbf{v}^*$ and $\mathbf{d}^*$. Note that the notations are distinguishing from the ones in (16). $\mathbf{f}^0$and $\mathbf{h}^0$ correspond to the routes that carry positive flows, which are the equilibrated (or minimum-cost) routes according to Assumption 2 in Du et al. [18], and can be obtained from the results of the OB algorithm. The superscript "0" is associated with the variables in a reduced problem, which only consists of the equilibrated routes.

From the coefficient matrix of (47), the maximum set of the linear independent columns corresponds to the set of the equilibrated and linear independent routes (ELI, denoted by R^B) and can be found using the row echelon form of the coefficient matrix. We can utilize the blocks $\Lambda_{\mathbf{f}}$ and $\Lambda_{\mathbf{h}}$ which are incidence matrices with one in each column (because one route can only serve one O-D). Since there may be more than one equilibrated route between an O-D pair, $\Lambda_{\mathbf{f}}$ and $\Lambda_{\mathbf{h}}$ are generally not column full rank. Thus, the coefficient matrix can be rewritten as

$$\begin{aligned} \begin{bmatrix} \overline{\Lambda}_{\mathbf{f}} & \mathrm{O} & \widetilde{\Lambda}_{\mathbf{f}} & \mathrm{O} \\ \mathrm{O} & \overline{\Lambda}_{\mathbf{h}} & \mathrm{O} & \widetilde{\Lambda}_{\mathbf{h}} \\ \overline{\Delta}_{\mathbf{f}} & \overline{\Delta}_{\mathbf{h}} & \widetilde{\Delta}_{\mathbf{f}} & \widetilde{\Delta}_{\mathbf{h}} \end{bmatrix} &= \begin{bmatrix} I & & \widetilde{\Lambda}_{\mathbf{f}} & \mathrm{O} \\ & I & \mathrm{O} & \widetilde{\Lambda}_{\mathbf{h}} \\ \overline{\Delta}_{\mathbf{f}} & \overline{\Delta}_{\mathbf{h}} & \widetilde{\Delta}_{\mathbf{f}} & \widetilde{\Delta}_{\mathbf{h}} \end{bmatrix} \\ &\longrightarrow \begin{bmatrix} I & \mathrm{O} & \mathrm{O} \\ \mathrm{O} & \widetilde{\Delta}_{\mathbf{f}} - \overline{\Delta}_{\mathbf{f}}\widetilde{\Lambda}_{\mathbf{f}} & \widetilde{\Delta}_{\mathbf{h}} - \overline{\Delta}_{\mathbf{h}}\widetilde{\Lambda}_{\mathbf{h}} \end{bmatrix}, \end{aligned} \tag{48}$$

where $\overline{\Lambda}_{\mathbf{f}}$ and $\overline{\Lambda}_{\mathbf{h}}$ are square matrices, which consist of exactly one equilibrated route for each O-D, so they are equal to identity matrices. Thus, the question is reduced to find and eliminate the linear dependent columns in a submatrix. Consider

$$\left[\widetilde{\Delta}_{\mathbf{f}} - \overline{\Delta}_{\mathbf{f}}\widetilde{\Lambda}_{\mathbf{f}} \quad \widetilde{\Delta}_{\mathbf{h}} - \overline{\Delta}_{\mathbf{h}}\widetilde{\Lambda}_{\mathbf{h}}\right]. \tag{49}$$

The columns of the above matrix are only related to the alternative routes between every O-D. Therefore, the set R^B will include the routes corresponding to $\overline{\Lambda}_{\mathbf{f}}$, $\overline{\Lambda}_{\mathbf{h}}$, and the maximum set of the independent columns in (49).

3.4. Solution of the Maximum Trip Production Problem. When the sensitivity results are derived, the SAB algorithm will use this information to represent the implicit relationships between $\mathbf{v}$, $\mathbf{d}$, and $\mathbf{O}$. Thus, the bilevel problem is first-order approximated at the given point, $\mathbf{O}^* = f^{-1}(\mathbf{v}^*, \mathbf{d}^*)$. Let $\mathbf{v}(\mathbf{O}^*)$ and $\mathbf{d}(\mathbf{O}^*)$, respectively, denote the solutions to the lower-level model at $\mathbf{O}^*$. The relationship can be represented by using the Taylor expansion:

$$\begin{aligned} \mathbf{v}(\mathbf{O}) &\approx \mathbf{v}(\mathbf{O}^*) + \nabla_{\mathbf{O}}\mathbf{v} \cdot (\mathbf{O} - \mathbf{O}^*), \\ \mathbf{d}(\mathbf{O}) &\approx \mathbf{d}(\mathbf{O}^*) + \nabla_{\mathbf{O}}\mathbf{d} \cdot (\mathbf{O} - \mathbf{O}^*), \end{aligned} \tag{50}$$

where the derivatives $\nabla_{\mathbf{O}}\mathbf{v}$ and $\nabla_{\mathbf{O}}\mathbf{d}$ are obtained from the sensitivity analysis of the ETDA-VDC model. Therefore, the bilevel problem can be reformulated as

$$\begin{aligned} \text{Max} \quad & \sum_{p \in Z} O_p \\ \text{s.t.} \quad & \nabla_{\mathbf{O}}\mathbf{v} \cdot \mathbf{O} \le \mathbf{C} - \mathbf{v}^* + \nabla_{\mathbf{O}}\mathbf{v} \cdot \mathbf{O}^* \\ & \Phi \cdot \nabla_{\mathbf{O}}\mathbf{d} \cdot \mathbf{O} \le \mathbf{O}^{\max} - \overline{\mathbf{O}} - \mathbf{O}^* + \Phi \cdot \nabla_{\mathbf{O}}\mathbf{d} \cdot \mathbf{O}^* \\ & \Psi \cdot \nabla_{\mathbf{O}}\mathbf{d} \cdot \mathbf{O} \le \mathbf{D}^{\max} - \overline{\mathbf{D}} - \mathbf{D}^* + \Psi \cdot \nabla_{\mathbf{O}}\mathbf{d} \cdot \mathbf{O}^*. \end{aligned} \tag{51}$$

The solution of this linear programming can be easily derived using the simplex method. However, because the linear programming problem is just locally approximated, the new solution might be infeasible to the original problem. At an infeasible point of $\mathbf{O}^*$ the lower-level trip distribution and assignment results may not satisfy the upper-level constraints. In extreme cases, for example, some link flow v_a^* may be much greater than its capacity C_a, which could cause the capacity constraints to fail for any nonnegative $\mathbf{O}$. In consequence, the new linear approximation will have no solution.

In consideration of this flaw, we modified the conventional SAB algorithm by restraining the solution to be always located within the feasible region. A trip distribution and assignment step is implemented at the solution of the above linear programming, $\mathbf{O}^*$, and then the results are checked with the capacity conditions. If any capacity constraint is violated, the new solution should be updated by a convex combination of the solution from last iteration, $\mathbf{O}^{(n)}$, and $\mathbf{O}^{*(n)}$. Therefore, let $\widehat{\mathbf{O}}$ be the solution to the approximate linear problem above; the maximum step size $\lambda = 2^{-k}$ ($k = 0, 1, 2, \ldots$) is chosen which ensures the capacity constraint is satisfied at any link or traffic zone. The new solution is updated by

$$\mathbf{O}^{(n+1)} := \mathbf{O}^{(n)} + 2^{-\lambda}\left(\widehat{\mathbf{O}}^{(n)} - \mathbf{O}^{(n)}\right). \tag{52}$$

3.5. A Heuristic Solution Process. In summary, the modified SAB algorithm involves an iterative process between the

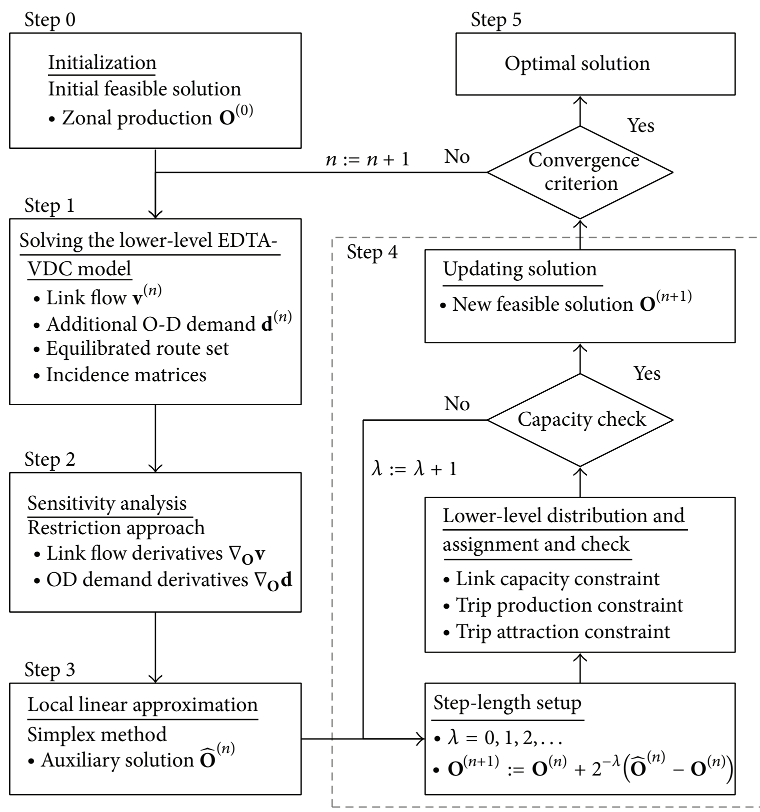

FIGURE 2: Flowchart of sensitivity analysis based algorithm for network capacity model.

upper-level and lower-level problems and can be summarized as follows.

Step 0 (initialization). Determine an initial value of trip production pattern $\mathbf{O}^{(0)}$. Set $n := 0$.

Step 1 (solving lower-level problem). Solve the ETDA-VDC model in lower-level for the given $\mathbf{O}^{(n)}$, by which the equilibrium link flows $\mathbf{v}^{(n)}$ and O-D demand $\mathbf{d}^{(n)}$ are obtained. The set of all equilibrated routes can be derived at the equilibrium point.

Step 2 (sensitivity analysis). Calculate the partial derivatives, $\nabla_{\mathbf{O}}\mathbf{v}$ and $\nabla_{\mathbf{O}}\mathbf{d}$, using the sensitivity method for the ETDA-VDC model.

Step 3 (local linear approximation). Formulate local linear approximations of the upper-level capacity constraints using the derivative information and solve the approximate linear programming problem to produce an auxiliary trip production $\widehat{\mathbf{O}}^{(n)}$.

Step 4 (updating solution). Let $\mathbf{O}^{(n+1)} := \mathbf{O}^{(n)} + 2^{-\lambda}(\widehat{\mathbf{O}}^{(n)} - \mathbf{O}^{(n)})$, where $\lambda = 0, 1, 2, \ldots$, until the capacity constraints are satisfied by solving the ETDA-VDC problem with $\mathbf{O}^{(n+1)}$. Set $n := n + 1$.

Step 5 (convergence criterion). If $|O_p^{(n+1)} - O_p^{(n)}| \le \kappa$ for all $p \in Z$ then stop, where κ is a predetermined tolerance. Otherwise, return to Step 1.

The above process is a modification of the conventional SAB approach. The flowchart is shown in Figure 2. It should be noted that without the step-size adaption in Step 4 the SAB method only works in very small networks but could fail in larger examples. The set of all equilibrated routes is used to apply the restriction approach for sensitivity analysis. In this study, the modified OB algorithm for the ETDA-VDC problem is used for this purpose. In addition, the derivative results of the sensitivity analysis can give a precise local linear approximation to the upper-level capacity constraints, which is very important to make the heuristic search converge to a considerable good solution.

In addition, given the nonconvexity of the bilevel problem, the SAB algorithm will converge to a local optimal point [2]. However, as shown in Yang et al. [28], for MPEC models, if the upper-level objective function is a linear function of its decision variables, the heuristic algorithm can at least find a noninferior optimal solution. So the SAB algorithm is supposed to be able to find a satisfying solution for the road network capacity problem in practice.

4. Numerical Experiments

Numerical experiments are conducted in this section. In the numerical experiments the link cost function employs the Bureau of Public Roads (BPR) function:

$$t_a(v_a) = t_a^{\text{free}}\left[1 + 0.15\left(\frac{v_a}{C_a}\right)^4\right]. \tag{53}$$

The destination cost function is defined, referring to Chen and Kasikitwiwat [8], as follows:

$$M_q(D_q) = k_q\left[\sum_{p\in Z_q}\left(e_{pq} + d_{pq}\right)\right]^{\omega_q} - m_q, \tag{54}$$

where k_q is a scaling factor between demand and service cost; ω_q is a dimensionless parameter related to the severity of congestion; and m_q represents a fixed attraction at destination q.

The experiments were implemented on Intel Core i5 CPU 3.20 GHz, 4 GB RAM, using the Microsoft Windows 7 operating system. All of the coding was carried out in Visual C#. The solution accuracy is measured in each iteration n by the relative error of the trip productions $\mathbf{O}^n$ (RE) as follows:

$$\text{RE}^n = \frac{\left\|\mathbf{O}^{(n)} - \mathbf{O}^{(n-1)}\right\|}{\left\|\mathbf{O}^{(n)}\right\|}. \tag{55}$$

4.1. Convergent Rate. To testify the efficiency of the SAB method, the Sioux Falls and Anaheim networks are used for the solution of the network capacity model. The IEA algorithm [1] is also employed as a reference in the test. The Sioux Falls is an aggregated network with 24 zones, 24 nodes, and 76 links. The Anaheim is a medium network that consists of 38 zones, 416 nodes, and 914 links (http://www.bgu.ac.il/~bargera/tntp/). The exiting O-D matrix is adapted to make the current traffic flow be unsaturated. The parameters of the destination cost function are given by default. The two algorithms, SAB and IEA, are implemented on these two networks to compare their convergence performances, and the results are listed in Table 1. The stopping tolerance is set to be 10^{-7}.

From the tests, although SAB algorithm shows a slower speed to reach a local optimum, it always produces much better solutions compared to IEA algorithm. In the tests on the Anaheim network, IEA has not converged and tended to zigzag after a few cycles of computation. Since IEA can quickly produce the approximate derivatives at current point, the CPU time taken in each iteration is much less than SAB algorithm. However, IEA could easily stop at a nonoptimal solution or not converge. This indicates that the accuracy of the derivative results has a significant influence on the convergence of the heuristic search for the network capacity model. The converging processes of the two algorithms on the two networks are illustrated in Figure 3. Note that the conventional SAB algorithm failed in both networks in our tests.

Table 1: Computational result test networks.

Test networks	Observations	SAB algorithm	IEA algorithm
Sioux Falls	CPU time (sec)	320	162
	Iterations	18	37
	F^1	158498.9	126093.8
Anaheim	CPU time (sec)	650	$>650^2$
	Iterations	20	$>163^2$
	F^1	137853.3	122039.0^2

[1] F is the value of the objective function when the algorithm terminates.
[2] IEA does not meet the convergence criteria on the Anaheim network.

4.2. Convergent Stability. Since the characteristic of the heuristic search, little can be said theoretically about the convergence of the SAB algorithm. In this study, we implement SAB algorithm from different start points to observe the converging process. On both the Sioux Falls and Anaheim networks, SAB algorithm can find a sufficient good solution given any arbitrary initial solution. We select the converging processes from five different start points on Anaheim network and plot the results in Figure 4. All the computations were implemented in 20 times of iterations. The final solutions from the five computations are shown in Figure 5.

The experiment result in Figure 4 indicates that the solutions by our SAB algorithm can reach a sufficient high precision. All the solutions in Figure 5 are very close, which reflects that the SAB algorithm is comparatively stable for the network capacity problem. Consequently, the proposed SAB algorithm is testified to probably converge to a suboptimal solution from any appropriate feasible start point for the network capacity model. Since the bilevel problem is not convex, the suboptimal solution can be regarded as a satisfying approximation. This solution might be quite close to the global optimum according to the aforementioned characteristics of the network capacity model. Since it only takes a few (less than 30) times of iterations to reach the local optimum, the whole computation of the SAB algorithm can be completed in a considerable CPU time, which is quite superior to the global optimization, for example, genetic algorithm. Thus, the modified SAB method could be a valuable tool for the estimation of the road network capacity in practice.

5. Conclusions

This paper has presented an effective SAB method for the solution of the transportation network capacity problem. An OB algorithm for ETDA-VDC problem is presented to solve the lower-level problem rapidly and efficiently. The primal problem of sensitivity analysis has been simplified, so the inverse calculation of a full-size matrix has been converted to solving corresponding linear systems of equations. To ensure the heuristic search can proceed properly, the solution from

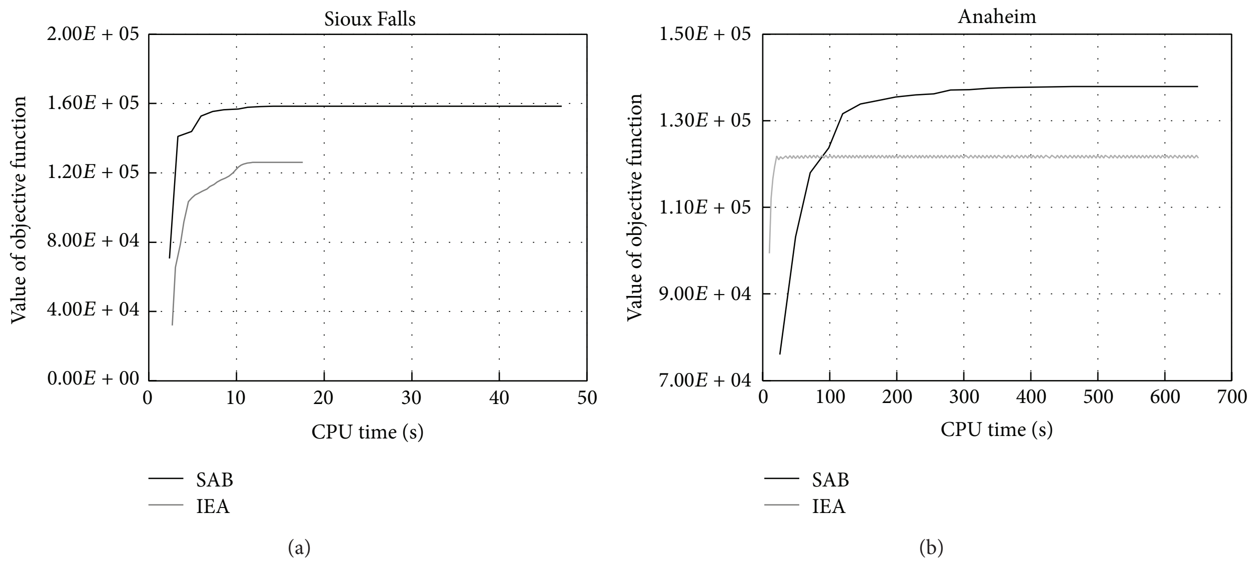

Figure 3: Value of objective function versus CPU time.

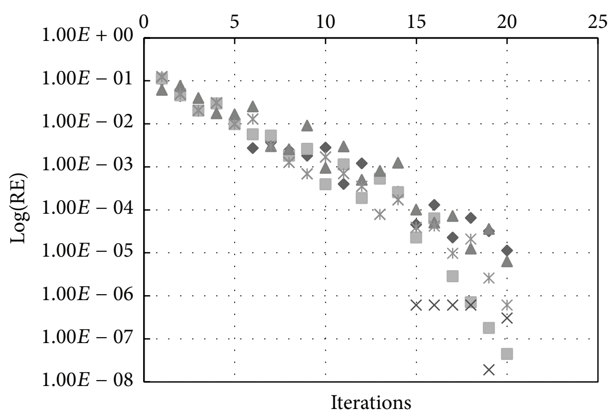

Figure 4: Relative errors of trip production versus iteration from five start points on Anaheim.

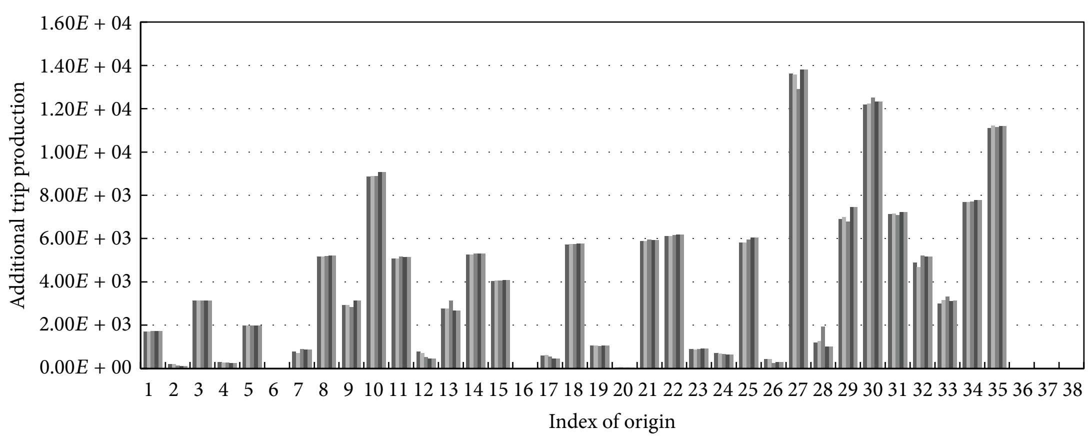

Figure 5: Solutions of additional trip production from five start points on Anaheim.

the local linear approximation is rectified to be restrained in its feasible region. Based on these improvements, the SAB method is able to be applied to estimate the capacity of the large-scale real networks. The performance of the proposed SAB method has been demonstrated on two experimental networks.

Based on the method in this paper, the network capacity model could be used to provide an ideal travel demand pattern for a given transportation system, by which the traffic resources can be utilized in maximum. From the model results, the traffic engineers can evaluate the design scheme of a transportation network more accurately before its construction. The results can also be used as a reference for the layout of land use in the new developing areas. In addition, the traffic flow pattern at full saturation may reflect the potential bottleneck of the given network. Further applications of the network capacity model could be detected based on the proposed estimation method.

Notation

A: Set of nodes in the network
A_p: Restricting subnetwork for origin p
C_a: Capacity of link a
$\overline{D}_q$: Existing trip attraction at destination q
D_q: Additional trip attraction at destination q
$D_q^{\max}$: Upper limit of trip attraction at destination q
M_q: Cost at destination q
$\overline{O}_p$: Existing trip production at origin p
O_p: Additional trip production at origin p
$O_p^{\max}$: Upper limit of trip production at origin p
R_{pq}: Set of routes from p to q
Z: Set of zones in the network
Z_p: Set of all destination for origin p
Z_q: Set of all origins for destination q
a: Link index, $a \in A$
d_{pq}: Additional demand from origin p to destination q
e_{pq}: Existing demand from origin p to destination q
f_r^{pq}: Flow on route r from p to q associated with elastic demand
h_r^{pq}: Flow on route r from p to q associated with fixed demand
p: Origin index, $p \in Z$
q: Destination index, $q \in Z$
r: Route index, $r \in R_{pq}$
t_a: Travel cost on link a and the function of all link flows $\mathbf{x}$
t_a^{free}: Free flow travel time on link a
v_a: Flow on link a
v_{ap}: Traffic flow on link a from origin p
α_{ap}: Approach proportion of link a from origin p
θ: Impedance parameter for trip distribution
τ_{pq}: The minimum route cost from p to q
$\delta_{a,r}^{pq}$: Link/route incidence indicator, 1 if link a is on route r from origin p to destination q; 0, otherwise
Δ: Link/route incidence matrix; $\Delta_{\mathbf{f}}$ and $\Delta_{\mathbf{h}}$ correspond to route flows $\mathbf{f}$ and $\mathbf{h}$, respectively
Λ: O-D/route incidence matrix; $\Lambda_{\mathbf{f}}$ and $\Lambda_{\mathbf{h}}$ correspond to route flows $\mathbf{f}$ and $\mathbf{h}$, respectively
Φ: Origin/O-D incidence matrix
Ψ: Destination/O-D incidence matrix.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research is supported by "the Natural Science Foundation of China (no. 51078085)" and "the Fundamental Research Funds for the Central Universities (no. 2014B04514)."

References

[1] H. Yang, M. G. H. Bell, and Q. Meng, "Modeling the capacity and level of service of urban transportation networks," *Transportation Research Part B: Methodological*, vol. 34, no. 4, pp. 255–275, 2000.

[2] S. C. Wong and H. Yang, "Reserve capacity of a signal-controlled road network," *Transportation Research Part B: Methodological*, vol. 31, no. 5, pp. 397–402, 1997.

[3] G. Ziyou and S. Yifan, "A reserve capacity model of optimal signal control with user-equilibrium route choice," *Transportation Research B: Methodological*, vol. 36, no. 4, pp. 313–323, 2002.

[4] H. Ceylan and M. G. H. Bell, "Reserve capacity for a road network under optimized fixed time traffic signal control," *Journal of Intelligent Transportation Systems*, vol. 8, no. 2, pp. 87–99, 2004.

[5] S. Chiou, "Reserve capacity of signal-controlled road network," *Applied Mathematics and Computation*, vol. 190, no. 2, pp. 1602–1611, 2007.

[6] N. Oppenheim, "Equilibrium trip distribution/assignment with variable destination costs," *Transportation Research B: Methodological*, vol. 27, no. 3, pp. 207–217, 1993.

[7] P. Kasikitwiwat and A. Chen, "Analysis of transportation network capacity related to different system capacity concepts," *Journal of the Eastern Asia Society of Transportation Studies*, vol. 6, pp. 1439–1454, 2005.

[8] A. Chen and P. Kasikitwiwat, "Modeling capacity flexibility of transportation networks," *Transportation Research Part A: Policy and Practice*, vol. 45, no. 2, pp. 105–117, 2011.

[9] T. L. Friesz, R. L. Tobin, H. J. Cho, and N. J. Mehta, "Sensitivity analysis based heuristic algorithms for mathematical programs with variational inequality constraints," *Mathematical Programming*, vol. 48, no. 2, pp. 265–284, 1990.

[10] R. L. Tobin and T. L. Friesz, "Sensitivity analysis for equilibrium network flow," *Transportation Science*, vol. 22, no. 4, pp. 242–250, 1988.

[11] H. Yang, "Heuristic algorithms for the bilevel origin-destination matrix estimation problem," *Transportation Research B: Methodological*, vol. 29, no. 4, pp. 231–242, 1995.

[12] H. Yang, "Sensitivity analysis for the elastic-demand network equilibrium problem with applications," *Transportation Research Part B: Methodological*, vol. 31, no. 1, pp. 55–70, 1997.

[13] S. D. Clark and D. P. Watling, "Probit-based sensitivity analysis for general traffic networks," *Transportation Research Record*, vol. 1733, no. 1, pp. 88–95, 2000.

[14] S. D. Clark and D. P. Watling, "Sensitivity analysis of the probit-based stochastic user equilibrium assignment model," *Transportation Research Part B: Methodological*, vol. 36, no. 7, pp. 617–635, 2002.

[15] J. Q. Ying and T. Miyagi, "Sensitivity analysis for stochastic user equilibrium network flows—a dual approach," *Transportation Science*, vol. 35, no. 2, pp. 124–133, 2001.

[16] H. Yang and M. G. H. Bell, "Sensitivity analysis of network traffic equilibria revisited: the corrected approach," in *Proceedings of the 4th IMA International Conference on Mathematics in Transport*, London, UK, 2007.

[17] C. Yang and A. Chen, "Sensitivity analysis of the combined travel demand model with applications," *European Journal of Operational Research*, vol. 198, no. 3, pp. 909–921, 2009.

[18] M. Du, L. Cheng, and H. Rakha, "Sensitivity analysis of combined distribution-assignment model with applications," *Transportation Research Record*, no. 2284, pp. 10–20, 2012.

[19] H. Yang and S. Yagar, "Traffic assignment and signal control in saturated road networks," *Transportation Research A: Policy and Practice*, vol. 29, no. 2, pp. 125–139, 1995.

[20] S. W. Chiou, "Optimization of area traffic control for equilibrium network flows," *Transportation Science*, vol. 33, no. 3, pp. 279–289, 1999.

[21] Q. Meng, H. Yang, and M. G. H. Bell, "An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem," *Transportation Research B: Methodological*, vol. 35, no. 1, pp. 83–105, 2001.

[22] Z. Gao, H. Sun, and L. L. Shan, "A continuous equilibrium network design model and algorithm for transit systems," *Transportation Research Part B: Methodological*, vol. 38, no. 3, pp. 235–250, 2004.

[23] S.-W. Chiou, "Bilevel programming for the continuous transport network design problem," *Transportation Research B: Methodological*, vol. 39, no. 4, pp. 361–383, 2005.

[24] M. L. Tam and W. H. K. Lam, "Maximum car ownership under constraints of road capacity and parking space," *Transportation Research Part A: Policy and Practice*, vol. 34, no. 3, pp. 145–170, 2000.

[25] H. Bar-Gera and D. Boyce, "Origin-based algorithms for combined travel forecasting models," *Transportation Research B: Methodological*, vol. 37, no. 5, pp. 405–422, 2003.

[26] H. Bar-Gera, "Origin-based algorithm for the traffic assignment problem," *Transportation Science*, vol. 36, no. 4, pp. 398–417, 2002.

[27] M. Xu, A. Chen, and Z. Gao, "An improved origin-based algorithm for solving the combined distribution and assignment problem," *European Journal of Operational Research*, vol. 188, no. 2, pp. 354–369, 2008.

[28] H. Yang, S. Yagar, Y. Iida, and Y. Asakura, "An algorithm for the inflow control problem on urban freeway networks with user-optimal flows," *Transportation Research Part B*, vol. 28, no. 2, pp. 123–139, 1994.