

# Application of Quasi-Newton Methods to Capacitated UE Assignment

## 容量制約条件つき交通量配分への準ニュートン方法の適用

by Lin Cheng\*, Yasunori Iida\*\* and Nobuhiro Uno\*\*\*  
程 琳\*, 飯田恭敬\*\*, 宇野伸宏\*\*\*

### 1. Introduction

The conventional user equilibrium (UE) model<sup>1)</sup> is widely used to solve both theoretical and practical traffic problems, because of the simple, understandable formulation of the model and its theoretical principles. In a congested network, however, it is necessary to estimate the waiting time due to congestion at bottlenecks in a road network. To obtain a more reasonable description of network flow, it has been suggested that upper bounds be placed on link flows in the conventional UE model. Such a capacitated equilibrium model has been studied<sup>2-5)</sup>, but has received little attention. One important reason for the lack of attention is the underdevelopment of efficient algorithms to solve the capacitated equilibrium model.

The efficiency of the algorithm used to solve the conventional equilibrium model is critical because most capacitated network problems are ultimately converted into a conventional network problem without any constraints on link capacity using some mathematical transformation. For instance, Hearn and Ribera<sup>6)</sup> proposed that explicit link flow capacities be included in an extended objective function by means of an exterior penalty function method, thereby obtaining a sequence of conventional traffic assignment sub-problems, which can be solved by the Frank-Wolfe method. Inoue<sup>2,3)</sup> used the Frank-Wolfe method to solve the conventional assignment sub-problems produced by transforming a capacitated traffic assignment using an interior penalty method. Larsson and Patriksson<sup>4)</sup> introduced a Lagrange term into the extended objective function through dual representation of the capacity constraints, thereby creating an augmented Lagrange dual function, and they combined the techniques of Dantzig-Wolfe decomposition<sup>7)</sup>, column generation<sup>8)</sup>, and reduced gradient<sup>9)</sup> in an iterative augmented Lagrange scheme. Prashjer and Toledo<sup>10)</sup> reported that the orthogonal projection method<sup>11,12)</sup> could be applied to a capacitated network assignment by restricting the increment of path flows approaching the shortest path, as in Inoue's scheme.

In this regard, this paper is not different from previous work. In order to integrate column generation and penalization / dual representation of capacity constraints, this paper proposes two path-based methods for the capacitated user equilibrium assignment problem, which is a general non-linear optimization problem with explicit link capacity constraints, called the Quasi-Newton method. The first method, called the Quasi-Newton method with a straight gradient, uses the gradient of the objective function directly to obtain an improved feasible direction scaled from the second-order derivatives, and a line search is executed in terms of both path flow and link flow. The second method, called the Quasi-Newton method with a Rosen gradient, adopts the average cost of the paths used, which is called a Rosen gradient<sup>13)</sup>, to obtain an improved direction, and the same line search is applied. Both methods are combined using the interior penalty technique, to achieve the capacitated equilibrium solution efficiently. Both methods are path-based and of the Newton type, but they differ in feasible direction and step size. They are used to solve the capacitated network assignment problem that augments knowledge of a congested network. Numerical experiments are used to confirm their characteristics and usefulness.

When link capacity constraints are not considered in an equilibration procedure, Quasi-Newton methods provide excellent path-based solutions to the conventional UE assignment problem, which still include the disaggregate simplicial decomposition algorithm<sup>4)</sup> and the orthogonal gradient projection method<sup>11,12)</sup>. Unlike the link-based Frank-Wolfe method, where only one shortest path is used in every OD pair, the path-based Quasi-Newton methods assume that multiple shortest paths might exist in every OD pair, and rationally disperse the OD flow over the paths used. These multiple shortest paths saved at every iteration facilitate

---

Keyword: Traffic Assignment, Network Flow, Traffic Capacity, Quasi-Newton Method

\*student member of JSCE, Department of Civil Engineering, School of Engineering, Kyoto University  
(Yoshida Honcho, Sakyo-ku, Kyoto, Japan, TEL 075-753-5126, FAX 075-7535907  
cheng@urbanfac.kuciv.kyoto-u.ac.jp

\*\*Fellow of JSCE, Ph.D., Professor, Department of Civil Engineering, Kyoto University

\*\*\*Regular Member of JSCE, Ph.D., Associate Professor, Department of Civil Engineering, Kyoto University

convergence and readily provide a reference path that can be used in the equilibrium flow sensitivity analysis<sup>16</sup>. Quasi-Newton methods use the second-order derivatives directly to achieve equilibration and have greater theoretical efficiency than the disaggregate simplicial decomposition algorithm. Theoretically, Quasi-Newton methods parallel the orthogonal gradient projection method in computational efficiency because they are all based on the Newton formula. However, since all the tentative solutions and the final solution obtained by Quasi-Newton methods always exist in the feasible region, it is not necessary to employ orthogonal projection, which might be difficult for a traffic engineer to understand. Moreover the proposed Quasi-Newton methods offer alternative feasible directions, and this flexibility is useful in choosing suitable solution algorithms.

This paper consists of seven sections. The capacitated equilibrium model and its properties are briefly described in Section 2. In Section 3, a penalty function is used to translate the capacitated model into an analogue of the conventional UE. Section 4 deduces the movement direction, scaling, and step size for the two Quasi-Newton methods. Section 5 presents the procedure used to solve the capacitated model. Section 6 solves numerical examples using the proposed methods. Finally, conclusions are drawn in Section 7.

## 2. Capacitated Traffic Assignment Model

Given transportation network  $G(A, N)$ , where  $A$  and  $N$  are the sets of links and nodes, respectively, each directed link  $a \in A$  is associated with a positive travel time  $t_a(x_a)$  as an increasing function of link flow  $x_a$ . Also  $C_a$  represents the capacity of link  $a$ .  $W$  is the set of origin-destination (OD) pairs, such that for each pair  $w \in W$ , there is a given traffic demand  $q^w$ . The capacitated user equilibrium assignment problem is formulated as follows:

$$\text{minimize } \sum_a \int_0^{x_a} t_a(x) \cdot dx \quad (1a)$$

$$\text{subject to } \sum_k f_k^w = q^w \quad \forall w \in W \quad (1b)$$

$$x_a = \sum_w \sum_k f_k^w \cdot \delta_{ak}^w \quad \forall a \in A \quad (1c)$$

$$x_a \leq c_a \quad \forall a \in A \quad (1d)$$

$$f_k^w \geq 0 \quad \forall k \in K^w, w \in W \quad (1e)$$

where  $f_k^w$  denotes the flow on path  $k$  within OD pair  $w$  and  $K^w$  is the path set within OD pair  $w$ .  $\delta_{ak}^w = 1$  if link  $a$  belongs to path  $k$  for OD pair  $w$ ; otherwise,  $\delta_{ak}^w = 0$ .

This problem becomes the conventional UE assignment if the link capacity constraints are dropped (1d). Since all the path flow variables of interest are positive and the nonnegative constraints (1e) are not binding, no negativity in terms of path flow may be omitted hereafter without affecting the assignment solution. The positive path flows are acquired via column generation, as illustrated in Section 4. In this study, the path set includes only the positive variables, *i.e.*, the paths used.

Consistent with the necessary and sufficient conditions of optimality, the equilibrium condition for every OD pair is stated as

$$\left. \begin{aligned} \sum_a (t_a + \mu_a) \cdot \delta_{ak}^w &= \tau^w \quad \text{if } f_k^w > 0 \\ \sum_a (t_a + \mu_a) \cdot \delta_{ak}^w &\geq \tau^w \quad \text{if } f_k^w = 0 \end{aligned} \right\} \quad \forall k, w \quad (2)$$

where  $\mu_a$  is the Lagrange multiplier of the link capacity constraint on link  $a$ ,  $\tau^w$  is the shortest path travel time within OD pair  $w$ .  $\mu_a$  is the Lagrange multiplier of the corresponding equation (1d), and is positive if the equation is active and zero otherwise.  $\delta_{ak}^w = 1$  if link  $a$  belongs to path  $k$ ; otherwise,  $\delta_{ak}^w = 0$ . The solution of the capacitated UE assignment model is considered to have the characterization of a Wardrop principle when the travel time is articulated in terms of running and waiting times. This generalized travel time is, in fact, the cost that individual travelers in a congested network generally seek to minimize. The waiting time or queuing delay is equivalent to the Lagrange multiplier associated with the capacity constraint on a given link<sup>2,3</sup>. Then, the equilibrium flow pattern and generalized time over a network can be obtained once equations (1a)-(1e) are solved.

### 3. Integrating Capacity Constraints into the Objective

It is desirable to apply some useful properties of the conventional equilibrium model to solving the capacitated traffic assignment problem. This study tries to utilize some familiar knowledge, such as the shortest route search and the all-or-nothing load of traffic. For this purpose, capacity constraints are incorporated into the objective function, thereby obtaining an extended objective function with additional penalty items for every link. Consequently, a barrier function<sup>9)</sup> is defined with respect to link  $a$ .

$$\psi_a(x_a) = -\log \frac{c_a - x_a}{c_a} \quad (3)$$

Let  $\gamma$  ( $\gamma > 0$ ) denote a penalty parameter, then an extended objective function is obtained:

$$\begin{aligned} z(\mathbf{x}, \gamma) &= \sum_a \int_0^{x_a} t_a(x) \cdot dx + \gamma \cdot \sum_a \psi_a(x_a) \\ &= \sum_a \int_0^{x_a} \{t_a(x) + \gamma \cdot \psi'_a(x)\} \cdot dx \\ &= \sum_a \int_0^{x_a} \left\{ t_a(x) + \frac{\gamma}{c_a - x} \right\} \cdot dx \end{aligned} \quad (4)$$

The capacitated traffic assignment problem can be transformed into an analogue of the conventional UE assignment, by extending the objective function. Alternatively, the following model can be solved to approximate a solution of the capacitated model (1a)-(1e).

$$\text{minimize } z(\mathbf{x}, \gamma) \quad (5a)$$

$$\text{subject to } \sum_k f_k^w = q^w \quad \forall w \quad (5b)$$

$$x_a = \sum_w \sum_k f_k^w \cdot \delta_{ak}^w \quad \forall a \quad (5c)$$

Setting penalty parameter  $\gamma$  ( $\gamma > 0$ ) to have decreasing values produces a series of optimization problems. As  $\gamma$  approaches 0, the solution of the link flow vector  $\mathbf{x} = \{x_a\}$ , converges on the optimal solution of the original problem. In parallel, the penalty item ( $\gamma \cdot \psi'_a(x)$ ) added to the link cost function is associated with the Lagrange multiplier for a given penalized capacity constraint, and is interpreted as the waiting time at the link exit. The relationship between the penalty item and the delay is explicitly described by the following relationship:

$$\mu_a = \lim_{\gamma \rightarrow 0^+} \frac{\gamma}{c_a - x_a} \quad (6)$$

Since the model can be regarded as a conventional UE problem with the augmented cost function  $\bar{t}_a(x) = t_a(x) + \gamma \cdot \psi'_a(x)$ , existing methods can be used to achieve the equilibrium flow, such as the Frank-Wolfe algorithm<sup>2,3)</sup>, or the orthogonal gradient projection method<sup>10)</sup>. To obtain a quickly converging accurate solution, and to facilitate operation of the algorithm and understanding, this study presents two feasible descent direction methods for solving equation (5), which originate from the Newton method and column generation.

### 4. Alternative Quasi-Newton Methods

With respect to the flow vectors of the paths used by traffic, the proposed methods solve equation (5) iteratively from a feasible point to an improved feasible point. Given a feasible point, or say path flow vector  $\mathbf{f}^n$ , a direction  $\Delta \mathbf{f}^n$  is determined such that for sufficiently small  $\lambda$  ( $\lambda > 0$ ), the following properties are satisfied: 1)  $\mathbf{f}^n + \lambda \Delta \mathbf{f}^n$  is feasible, and 2) the value of the objective function at  $\mathbf{f}^n + \lambda \Delta \mathbf{f}^n$  is better than that at  $\mathbf{f}^n$ . After such a direction is obtained, one-dimensional optimization is executed to determine how far to proceed along  $\Delta \mathbf{f}^n$ . This procedure leads to a new feasible point  $\mathbf{f}^{n+1}$ , and the process is repeated until convergence.

$$\mathbf{f}^{n+1} = \mathbf{f}^n + \lambda \cdot \Delta \mathbf{f}^n \quad (7a)$$

$$\mathbf{x}^{n+1} = \mathbf{x}^n + \lambda \cdot \Delta \mathbf{x}^n \quad (7b)$$

Figure 1 shows that a convex objective function of traffic assignment is defined over a feasible region in terms of path flows. In the unconstrained minimization case, the optimal solution can be represented by the lowest point on the objective function, where the gradient vector becomes the zero vector. For constrained minimization in traffic assignment, however, it is possible to have a minimum where the gradient vector does not become the zero vector. The shaded area in Figure 1 is the feasible region. When the minimum point exists on the boundary of the feasible region, all the elements of the gradient vector are greater than or equal to a constant for each OD pair, compatible with the Wardrop principle. Remembering that the gradient of the objective function represents the vector of path travel time in the domain of path flow, we conjecture that the constant for each OD pair is obviously equivalent to the shortest path time within that pair. The non-uniqueness of path flow signifies that the feasible region is uncertain with respect to path flow. In path-based methods, however, the uniqueness of the path travel time vector can be used to find a descent direction that leads to an efficient solution algorithm due to utilization of the second-order information of the objective function.

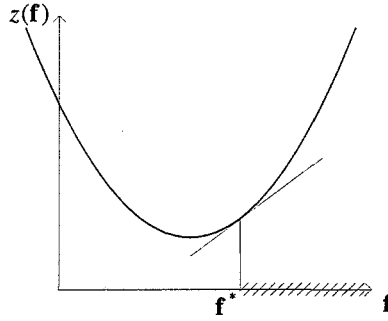


Figure 1 Optimum of capacitated UE assignment

Let  $\tau(\mathbf{f})$  denote a path travel time vector, which is a gradient of the objective function (5a) in the domain of path flows, and  $\tau'(\cdot)$  denote its derivative. According to the Newton formula, the direction of improvement along the path domain becomes

$$\Delta \mathbf{f}^n = -[\mathbf{t}(\mathbf{f}^n) - \mathbf{t}(\mathbf{f})] \cdot [\mathbf{t}'(\mathbf{f}^n)]^{-1} \quad (8)$$

From this equation, the direction of movement in the domain of link flows can be readily obtained through their linear incidence relationship, i.e.,

$$\Delta \mathbf{x}^n \equiv \mathbf{A} \cdot \Delta \mathbf{f}^n \quad (9)$$

where  $\mathbf{A} = [\delta_{ak}]$ , is the link-path incidence matrix. Consequently, it is critical to determine the direction of movement and the scaling factor, which are expressed by the first and second terms on the right hand side of Equation (8), respectively, and step size  $\lambda$  in Equation (7).

Note that in the standard Newton method, the step size is assumed to be one. However, here, a proper step size  $\lambda$ , which does not exceed 1, is adopted to guarantee that the link capacity constraints are not violated. Therefore, the methodology described in equations (7) and (8) belongs to the framework of Quasi-Newton methods. The different strategies used to solve the model can be adapted to different policies for gradient and step size.

#### 4.1 Moving Direction

In moving toward the optimum, the flows on all the paths used decrease in the direction expressed in Equation (8). The decreased flows on the paths used within each OD pair correspond to the increased flow on the shortest path. The shortest path  $k^{w*}$ , if it does not exist in the current path set, will join path set  $K^w(n)$  in the next iteration, i.e.,  $K^w(n+1) := K^w(n) \cup k^{w*}$ . According to the conservation of flow in every OD pair, the sum of the decreased flows must equal the increased flow along the shortest path. The change in path flows can be described as

$$\Delta \mathbf{f}^n = -[\mathbf{t}(\mathbf{f}^n) - \tau(f_{k^{w*}}^w) \cdot \mathbf{E}] \cdot [\mathbf{t}'(\mathbf{f}^n)]^{-1} \quad (10a)$$

where  $\mathbf{E}$  is an identity matrix of appropriate dimension.

Note that Equation (10a) contains only the initially updated path flows, exclusive of the coming shortest path. The path flow along the shortest path that now joins the path set can be expressed by summing the decreased flows.

$$\Delta f_{k^{**}}^w(n) = - \sum_{k \in K^w(n)} \Delta f_k^w(n) \quad (10b)$$

The combination of path flows described in equations (10a) and (10b) will be used to update the path flow vector using Equation (7).

In contrast to this strategy, in which path flows rush into the shortest path only, the second strategy assumes that the change in path flow depends on the average cost level within every OD pair, which is verified as a Rosen gradient in Inoue<sup>13)</sup>. Specifically, the path set is first augmented by adding the shortest path,  $K^w(n+1) := K^w(n) \cup k^{**}$ . If the cost of the path concerned is higher than the average cost within the given OD pair, then its flow decreases; if the cost of the path concerned is lower than the average cost within the given OD pair, then its flow increases. There is no change if the costs are equal, *i.e.*,

$$\Delta \mathbf{f}^n = - \left[ \mathbf{t}(\mathbf{f}^n) - \frac{1}{|K^w(n+1)|} \cdot \sum_{k \in K^w(n+1)} \tau(f_k^w) \cdot \mathbf{E} \right] \cdot [\tau'(\mathbf{f}^n)]^{-1} \quad (11)$$

where  $\mathbf{E}$  is an identity matrix of appropriate dimension, and  $|K^w(\cdot)|$  denotes the number of paths within pair  $K^w(\cdot)$ .

#### 4.2 Scaling

The two resulting directions must be scaled in order to coordinate the directions of the path flows used within the OD pair and the directions among different OD pairs, which are expressed by the second term on the right hand side of equations (10) and (11). The scaling is determined by the derivative of path costs, which is in fact the Hessian matrix of Equation (5a), and is a diagonal matrix due to the assumed irrelevance of paths. In the first strategy, the Hessian is defined in the domain of used paths exclusive of the shortest paths, and its diagonal elements are calculated by

$$\frac{\partial^2 z(\mathbf{f})}{(\partial f_k^w)^2} = \sum_{a \in A} \bar{t}_a' (\delta_{ak}^w - \delta_{ak^{**}}^w)^2, \quad \forall k \neq k^{**}, \forall w$$

where  $\bar{t}_a'$  denotes the derivative of the augmented travel time function of link  $a$ .

In the second strategy, the Hessian is defined in the domain of used paths inclusive of the shortest paths, and its diagonal elements are replaced by the inverse of the corresponding OD flow<sup>2)</sup>, which could decrease the overhead associated with the inversion of the Hessian matrix.

$$\frac{\partial^2 z(\mathbf{f})}{(\partial f_k^w)^2} \equiv (q^w)^{-1} \quad \forall k, \forall w$$

#### 4.3 Step Size

In the standard Newton method, the step size is defined as 1 in the unconstrained optimization. For traffic equilibrium problems with demand conservation, non-negativity of decision variables, and capacity constraints, a step size of 1 might be inappropriate, because the feasibility of the solution might be violated. For this reason, the step size should be restricted using various constraints and optimized by the line search method.

First, the non-negativity of path flows requires

$$f_k^w + \lambda_{path} \Delta f_k^w \geq 0, \quad \forall k \in K^w, w \in W \quad (12a)$$

which is rewritten as

$$\lambda_{path} \leq -\frac{f_k^w}{\Delta f_k^w}, \quad \forall k \in K^w, w \in W \quad (12b)$$

In terms of all the OD pairs, the step size of the flow on every path must be coordinated and the optimal step size for all paths must satisfy the following relationship.

$$\lambda_{path} = \min \left( \min_{k \in K^w} \left\{ -\frac{f_k^w}{\Delta f_k^w} \mid \Delta f_k^w < 0 \right\}, 1 \right) \quad (12c)$$

Second, we try to decide the step size required by link flows in the sense of mathematical programming. In terms of the current link flow vector,  $\mathbf{x}^n$ , the second-order approximation of the function is expressed as:

$$z(\mathbf{x}) \cong z(\mathbf{x}^n) + \lambda \cdot \Delta \mathbf{x}^T \cdot z'(\mathbf{x}^n) + \frac{1}{2} \lambda^2 \cdot \Delta \mathbf{x}^T \cdot z''(\mathbf{x}^n) \cdot \Delta \mathbf{x} \quad (13a)$$

where,  $\lambda$  is the feasible step size,  $\Delta \mathbf{x}$  is the vector of increased link flows, and  $z'$ ,  $z''$  are the first and second order derivatives of objective function (5a), respectively. Substitute

$$\Delta \mathbf{x}^n \equiv \mathbf{A} \cdot \Delta \mathbf{f}^n \quad (\text{Incidence equation of moving flows between a link and path})$$

and

$$z''(\mathbf{x}) = \bar{\mathbf{t}}' \quad (\text{The second derivative equations of the function})$$

into Equation (13a), then

$$z(\mathbf{x}) \cong z(\mathbf{x}^n) + \lambda \cdot \sum_{a \in A} [\Delta x_a \cdot \bar{t}_a]_{\mathbf{x}^n} + \frac{1}{2} \lambda^2 \sum_{a \in A} [\Delta x_a^2 \cdot \bar{t}_a']_{\mathbf{x}^n} \quad (13b)$$

Due to the optimality condition, the step size of link flows may be calculated by solving the equation  $z'(\lambda) = 0$ . With respect to each link flow, the step size must simultaneously coordinate and again must not be larger than  $\lambda_{path}$ . Accordingly, the feasible link flow step should satisfy

$$\lambda_{link} \cong \min \left\{ - \frac{\sum_{a \in A} \Delta x_a \cdot \bar{t}_a}{\sum_{a \in A} \Delta x_a^2 \cdot \bar{t}_a'}, \lambda_{path} \right\} \quad (13c)$$

For the capacitated network model, the step size should also be limited by the link capacity constraints.

Third, it is necessary to derive the feasible step size subject to the capacitated network flow. The corresponding direction of movement of link flow can be determined using Equation (9). In order to keep the solution within the feasible region, the step size  $\lambda$  must satisfy the following inequality.

$$\mathbf{x}^n + \lambda \cdot \Delta \mathbf{x}^n \leq \mathbf{c} \quad (14a)$$

Negotiating the possible step size in every link flow and every path flow, the optimal step size becomes

$$\lambda^* = \min \left( \min_{a \in A} \left\{ \frac{c_a - x_a}{\Delta x_a} \mid \Delta x_a > 0 \right\}, \lambda_{link} \right) \quad (14b)$$

This study has presented two strategies for determining movement direction and step size for the Quasi-Newton methods stated in equations (7) and (8). The first strategy puts flow on the shortest path; the second increases the flow on the lower cost path and decreases the flow on the higher cost path. The different strategies lead to their own optimal step size  $\lambda^*$ , thereby creating two path-based methods. Unlike link-based methods like the Frank-Wolfe method and the simplicial decomposition method, path-based methods take advantage of the availability of information on the second-order derivative of the objective function and column generation, where the flow on used paths is saved during iterations. The path-based methods are based on the Newton formula (8), which is superior to Frank-Wolf type methods in convergence and accuracy. The weakness of path-based methods is that they require a large amount of memory, both to save path information and for calculating the second-order derivatives for scaling or step size.

Unlike our step-restriction scenario, Prashjer and Toledo<sup>10)</sup> proposed a different algorithm, in which the flow on used, but non-shortest, paths is shifted to the shortest path for a capacitated UE model involving the orthogonal projection method, in order to ensure that the flow on the links on the shortest path does not violate capacity constraints. It appears to be difficult to determine which path, and how much flow affects the critical link of the shortest path, since the flow at the critical link is a composite of the flows of several paths. Furthermore, it is difficult to realize the feasibility of the solution in Prashjer and Toledo's modified projection function.

The Quasi-Newton methods and the augmented Lagrangean dual algorithm<sup>4)</sup> have different penalty techniques and solutions compared to conventional traffic assignment sub-problems. In the augmented Lagrangean dual algorithm, Larsson and Patriksson combine exterior penalty methods with Lagrangean dual schemes to transform the capacitated traffic problem into a sequence of conventional sub-problems. The main difference between these methods is the efficiency of the calculation process used to obtain the solutions to the conventional sub-problems, which governs the performance of the solution scheme. Larsson and Patriksson solve the sub-problems by using the disaggregate simplicial decomposition algorithm, in which the equilibration procedure is a hybrid first- and second-order approximation of the objective function, called the reduced gradient method and the diagonalized Newton method. Solving the alleged disaggregate master problem for each OD-pair might require onerous calculation, because it is very difficult to determine an appropriate point or say which parameter to switch from the reduced gradient method to the diagonalized Newton method. The equilibration process of our Quasi-Newton methods adopts the second-order

approximation of the objective function directly. The efficiency of the equilibration procedure theoretically results in Quasi-Newton methods being able to identify the used path set more quickly than the augmented Lagrangean dual algorithm.

## 5. Solution Procedures

Two Quasi-Newton methods were developed to solve the capacitated UE traffic assignment by using first and second order derivatives of the extended objective function, which should be incorporated into the outer calculation loop for the interior penalty method. Figure 2 shows the flowchart of the path-based method for capacitated traffic assignment. The Quasi-Newton inner loop, which should be placed in the outer calculation loop for the interior penalty method, consists of four procedures: initialization, column generation, equilibration, and convergence. Column generation includes finding the shortest paths and augmenting the path set to update the path set. Equilibration consists of updating the movement direction and step size to reach a convergent state. The two path-based methods differ only in equilibration, and adopt a straight gradient and Rosen gradient respectively. The detailed steps of the algorithm for these path-based methods are provided below.

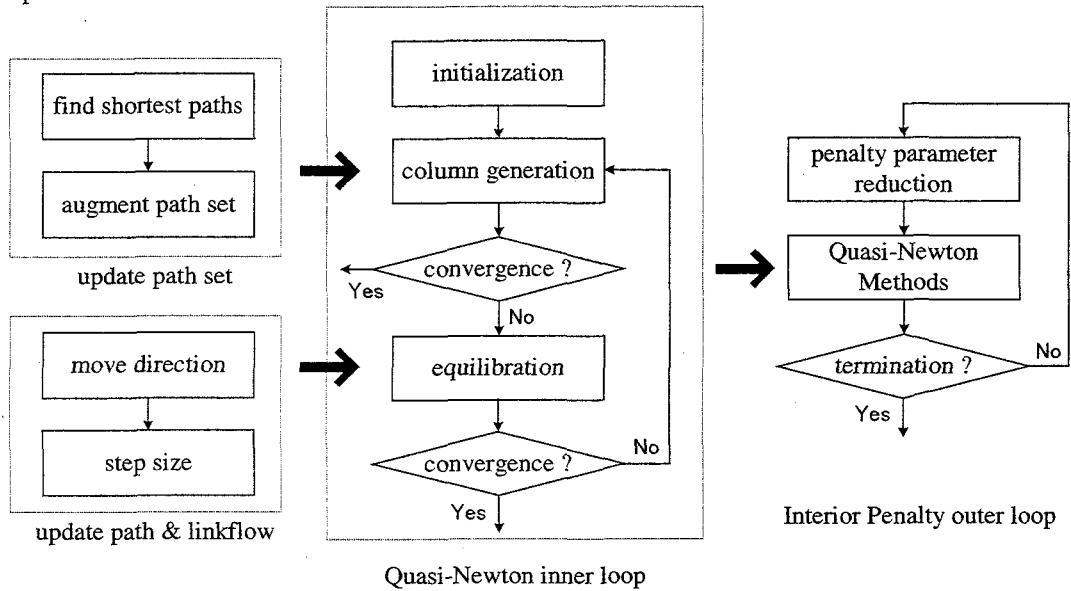


Figure 2 Flowchart of path-based methods for capacitated traffic assignment

### Initialization

- set the iteration counter  $n=0$
- solve the shortest path problem and create an initial path set  $K^w(0)$ ,  $\forall w \in W$
- perform All-or-Nothing assignment and obtain an initial path and link flow

### Column Generation

- increment the iteration counter:  $n:=n+1$
- update the cost and solve the shortest path problem
- record a set of shortest paths  $k^{w*}(n)$  and augment the path set  
 set  $K^w(n) = K^w(n-1) \cup k^{w*}(n)$ , if  $k^{w*}(n) \notin K^w(n-1)$ ; otherwise, set  $K^w(n) = K^w(n-1)$

### Equilibration

- calculate the movement direction of the path flow vector
- if  $f_k^w = 0$ , and  $\Delta f_k^w < 0$ , drop path  $k$ :  $K^w(n) = K^w(n) \setminus k$ <sup>1</sup>
- optimize the step size

<sup>1</sup>  $K^w(n) \setminus k$  means that path  $k$  is excluded from path set  $K^w(n)$

- update the flow pattern of the path and link

*Convergence*

- if  $\max_w \sum_k \frac{f_k^w(n)}{q^w} \left( \frac{\tau_k^w(n) - \tau_{k^{w*}}(n)}{\tau_k^w(n)} \right) \leq \varepsilon$ , terminate; otherwise, go to *Column Generation*

Concerning the outer calculation loop used for the interior penalty method, if the penalty parameter  $\gamma$  becomes sufficiently small that penalty items converge on the Lagrange multiplier for every link, the calculation should be terminated.

## 6. Numerical Experiments

The test network, shown in Figure 3, consists of 4 links and 3 nodes. The origin-destination demand and link performance are listed in Table 1. The standard BPR function is used to estimate the link travel time. In the first example, two Quasi-Newton methods are applied to the conventional traffic assignment to illustrate their different solution strategies. In the second example, the Quasi-Newton methods are used to solve the capacitated traffic assignment and to compare their efficiency with the Frank-Wolfe algorithm. The third example demonstrates the diversification of path flow patterns and its potential significance.

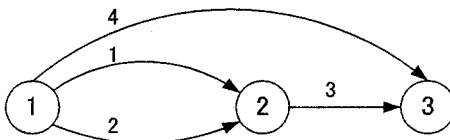


Figure 3 Network

Table 1. Link Performance

link	$t_a^0$	$c_a$	$q$
1	10	600	$q_{OD(1,2)}=600$
2	17	500	$q_{OD(1,3)}=400$
3	9	800	$q_{OD(1,3)}=400$
4	60	400	$q_{OD(2,3)}=600$

### Example 1

Conventional traffic assignment (without explicit link capacity constraints) can be solved using the Quasi-Newton methods proposed in this study, where the first and second methods adopt straight and Rosen gradients, respectively, to find the suitable movement direction. Both methods employ one-dimensional optimization to update the step size of flow movement. Of course, in the conventional model, the line search required by Equation (14) and the outer calculation loop used for the interior penalty method are unnecessary. Since they advance towards the optimal point using different direction strategies, the step size consequently differs. For convenience in describing the movement direction, Table 2 shows only the step size and the value of the objective function when the problem is solved in 5 iterations and termination criterion  $\varepsilon$  is set at 0.001. Although both methods have the same values for the objective function, their step sizes differ at the same iteration. The same phenomenon was verified in other examples. The proposed Quasi-Newton methods improve the feasible solutions although they use different calculations.

Table 2 Comparison of two Quasi-Newton methods in conventional UE assignment

Iterations	Quasi-Newton with Straight Gradient		Quasi-Newton with Rosen Gradient	
	Step Size	Objective	Step Size	Objective
1		0		0
2	0.02160	21974	0.04320	21974
3	0.02937	21727	0.05875	21727
4	0.03115	21721	0.06230	21721
5	0.03121	21721	0.06241	21721

### Example 2

In the second example, the capacitated traffic assignment problem is solved by Quasi-Newton methods. Initially, the penalty parameter  $\gamma$  is set at 1000, and is reduced at the rate 0.1 ( $\gamma := 0.1\gamma$ ); the termination criterion  $\varepsilon$  is set at  $10^{-4}$ . The initial feasible solution is obtained by Daganzo's technique<sup>14)</sup>, where a sufficiently large capacity is set first and the Quasi-Newton inner loop is repeated with decreasing link capacities until every link capacity equals its original value. A detailed explanation can be found in references [2-4]. The link flows of the capacitated UE assignment are shown in Table 3; the link solutions of the conventional UE assignment are included for comparison.



In the capacitated model, congestion occurs at links 1 and 3, which have delays of 5.57 and 33.15, respectively. At a congested link, the travel time is expressed as the sum of the running time and the delay. The difference between the capacitated and conventional models is in their link travel times, especially for congested links. For example, for link 3, the conventional model predicts a travel time of 12.30, while the capacitated model predicts a running time of 10.35 plus a delay of 33.15 caused by the saturated capacity at link 3. Another difference between these two models lies in the relationship between the estimated link flows and link capacity. As can be seen in Table 3, the capacitated UE model never provides estimated link flows that violate the link capacity constraints, while the conventional UE model allows the estimated link flows to exceed the corresponding link capacity.

Table 3 Link solutions for the capacitated and conventional UE problems

Link No.	Capacity	Capacitated UE				Conventional UE	
		Flow	Running Time	Delay	Travel Time	Flow	Travel Time
1	600	600	11.50	5.57	17.07	882.11	17.01
2	500	200	17.07	0	17.07	117.89	17.01
3	800	800	10.35	33.15	43.50	1000	12.30
4	400	200	60.56	0	60.56	0	60.00

Table 4 Path solutions for capacitated UE

Pairs	Path Flow		Link Makeup	Running Time	Delay	Travel Time
	Straight*	Rosen**				
(1, 2)	399.99	403.56	1	11.50	5.57	17.07
	200.01	196.44	2	17.07	0	17.07
(1, 3)	200.00	196.42	1_3	21.85	38.72	60.57
	200.00	200.00	4	60.56	0	60.56
		3.57	2_3	27.42	33.15	60.57
(2, 3)	600.00	600.00	3	10.35	33.15	43.40

Straight\* means Quasi-Newton method with straight gradient

Rosen\*\* means Quasi-Newton method with Rosen gradient

In terms of the generalized path travel time, the Wardrop principle is also satisfied with the estimated path flows, and can be verified by the path travel time illustrated in Table 4. For example, the two shortest paths between OD pair (1, 2) have the same travel times of 17.07; the three shortest paths between OD pair (1, 3) have the same travel times of 60.57. Two different path flow patterns are produced in Table 4 because the two methods use completely different direction strategies.

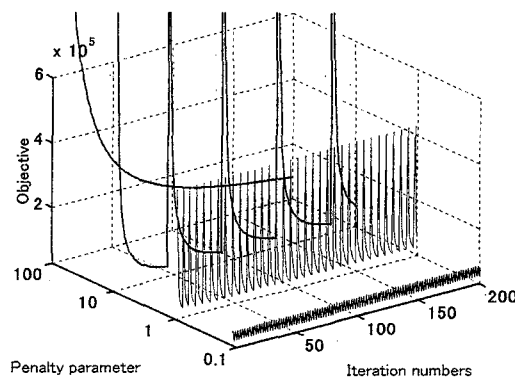
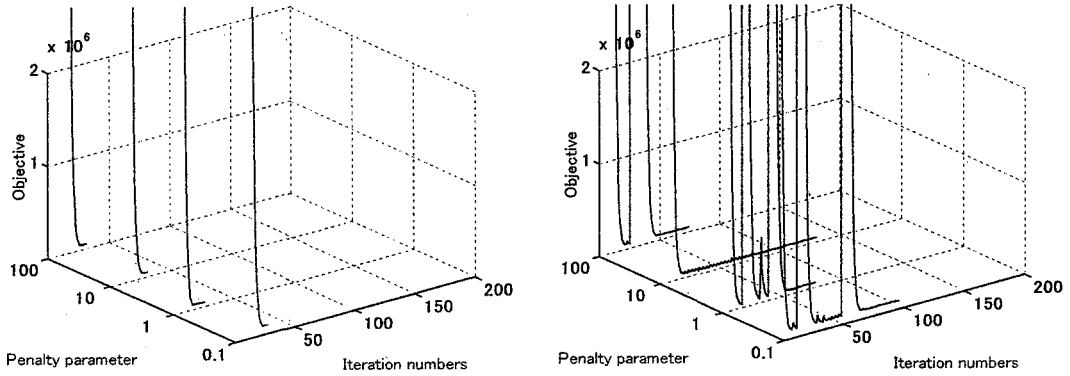


Figure 4 Convergence of the Frank-Wolfe method in capacitated UE assignment

Next, we illustrate the convergence procedure of the Quasi-Newton methods and compare it with that of the Frank-Wolfe method. Figure 4 shows the convergence procedure of the capacitated UE assignment solved using the Frank-Wolfe method. It cannot provide a precise solution because of its known zigzag phenomenon, which occurs at every penalty process. Moreover, the decrease in the objective value is slow

when the penalty parameter is larger. This means the early stage of the outer calculation loop for the inner penalty method might require time-consuming calculation. Inoue<sup>3)</sup> terminated the Frank-Wolfe procedure using a predetermined number. However, determining an appropriate upper limit is also a tricky task. Contrary to the Frank-Wolfe method, Quasi-Newton methods provide an effective algorithm to guarantee precise, converging solutions to the capacitated model with limited repetitions.



(a) Quasi-Newton method with a straight gradient (b) Quasi-Newton method with a Rosen gradient  
Figure 5 Convergence of Quasi-Newton methods in capacitated UE assignment

Figure 5 illustrates the convergence procedure of the Quasi-Newton methods with straight and Rosen gradients. Both methods are superior to the Frank-Wolfe method in accuracy and convergence at the expense of path flow reservation. The method with the straight gradient is so perfect that it converges continuously and rapidly in every penalty outer loop. By contrast, despite the high performance of the method with the Rosen gradient, the objective value is not continuous and sometimes jumps abruptly. These abrupt changes in the objective value occur because the movement direction generated can change abruptly when the movement direction is the zero vector or some link capacity becomes active. The average path cost of every OD pair in Equation (11) is originally defined by a Rosen projection and its direction finding map is not closed. The mathematical explanation refers to [9]. Overall, the method with the straight gradient is more efficient and more stable than that with the Rosen gradient.

### Example 3

Two different path flow vectors were produced using two Quasi-Newton methods in *Example 2*. In fact, if different feasible initial solutions are given, either path-based method can produce diverse path flow vectors. In our experiments, all the path flow patterns in *Example 2* can be described by the solutions in Figure 6(a). Simple calculations verify that every path flow vector for  $s \in [0, 200]$  observes flow conservation. The functional relation of path flows in Figure 6(a) shows that path flows are not capricious, but are subject to some intrinsic relationship. For example, we can foretell that the flow on path 1 is  $400 \leq f_1 \leq 600$ , while the flow on path 5 is  $200 \leq f_5 \leq 600$ , etc. Judging from the calculation results shown in Table 4, the case (b) in Figure 6 corresponds with the solution obtained by the Quasi-Newton Method with straight gradient.

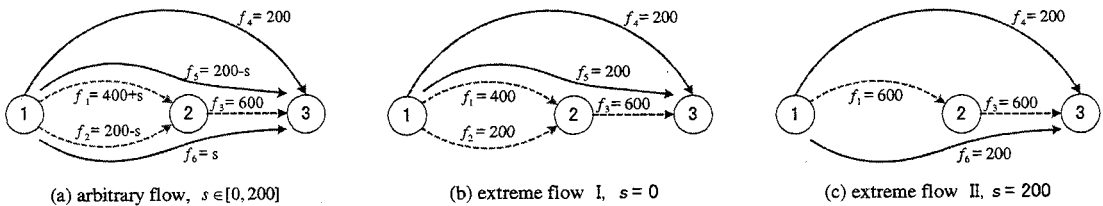


Figure 6 Path flow patterns in a capacitated network

Additionally, in transportation network problems, sensitivity<sup>15)</sup> of link flow to OD flow is needed for traffic control or network design. Sensitivity analysis<sup>16)</sup> requires a previous set of reference path flows, from which a particular path flow pattern can be abstracted. Unfortunately, it is not clear how to obtain the reference path flow pattern. For a small network, all the paths could easily be enumerated as the reference set, but for a large network, path enumeration is unimaginable. The Quasi-Newton methods are alternative ways to provide a reference path set for flow sensitivity analysis. For example, although two extreme points of path flows exist in the capacitated network, shown in Figure 6(b) and 6(c), only extreme point I is useful for sensitivity analysis in our experiment.

## 7. Conclusions and Future Research

In this study, we proposed two Quasi-Newton methods to deal with traffic assignment in a capacitated network. The methods combine the Newton formula, column generation, and penalty techniques. The first method uses the gradient of the objective function to obtain an improving feasible direction scaled using the second-order derivatives. The second uses a Rosen gradient to obtain an improving direction scaled using the corresponding origin-destination demand. Both methods use a line search to obtain an optimal step size to guarantee feasibility of either path or link flow. The proposed methods converge quickly, with high accuracy, at the expense of requiring memory to save path information. Numerical examples verified their efficiency and stability, as well as the usefulness of the saved path flow patterns. The Quasi-Newton method with a straight gradient is more stable than that with a Rosen gradient for capacitated traffic assignment.

Quasi-Newton methods typically yield a satisfactory solution to traffic assignment and are superior to the Frank-Wolfe method. The zigzag phenomenon that often occurs in the Frank-Wolfe method does not exist in Quasi-Newton methods. When a network is heavily congested, there is a considerable difference between the Quasi-Newton solution, in which several paths are used, and the Frank-Wolfe solution, in which only one path is used in every OD pair. In an equilibrium network, the state with multiple shortest paths might better describe the real world. Furthermore, Quasi-Newton methods use the second-order approximation of the objective function to adjust the flow movement among the paths used.

The network described by the capacitated model has a capacity that depends on its component capacities, such as the capacities of links or intersections. One future research topic is to employ Quasi-Newton methods to evaluate the capacity of an entire network subject to component capacities that are compatible with the existing origin-destination demand pattern.

Quasi-Newton methods form a large family and there are many other variants. For instance, the Hessian matrix in the straight gradient strategy can also be approximated using the corresponding OD flow, as with the Rosen gradient strategy. Conversely, the Rosen gradient strategy can be directly scaled using its Hessian matrix if the derivative of the Rosen gradient is available. It is recommended that the OD flow be used to approximate the elements of the Hessian matrix, since this saves the overhead associated with the Hessian. Generally speaking, orthogonal gradient projection (with a step size of one) can also be affiliated with the Quasi-Newton method family, on the same basis as the Newton formula. Such orthogonal projection can be extended to the Rosen gradient strategy and produce a new combination. In this sense, Quasi-Newton methods exhibit much flexibility in movement direction, scaling, and step size. A detailed evaluation of comprehensive Quasi-Newton methods is summarized in Tobin and Friesz<sup>16)</sup>.

Finally, storing and handling the vast amount of path data used in Quasi-Newton methods are still challenging tasks for a large transportation network. The remaining task for practical applications is to integrate Quasi-Newton methods into Dantzig-Wolfe decomposition.

## References

- 1) Sheffi, Y.: *Urban Transportation Networks*, Prentice-Hall, 1985
- 2) Inoue, H.: Traffic assignment on a network with congested flows, *Proceedings of JSCE*, No.365/IV-4, pp.125-133, 1986a
- 3) Inoue, H.: Traffic equilibrium and its solution in congested road networks, *Proceedings of IFAC Control in Transportation Systems*, pp.267-272, 1986b
- 4) Larsson, T. and Patriksson, M.: An augmented Lagrange dual algorithm for link capacity side constrained traffic assignment problems, *Transportation Research*, 29B, pp.433-455, 1995
- 5) Bell, M. G. H. and Iida, Y.: *Transportation Network Analysis*, John Wiley & Sons, pp.114-148, 1997
- 6) Hearn, D.W. and Ribera, J.: Bounded flow equilibrium problems by penalty methods, *Proceedings of the 1980 IEEE International Conference on Circuits and Computers*, pp.162-166, 1980

- 7) Dantzig, G.B. and Wolfe, P.: The decomposition algorithm for linear programming, Operations Research, Vol.8, pp101-111, 1960
- 8) Leventhal, T. *et al.*: A column generation algorithm for optimal traffic assignment, Transportation Science, Vol.7, 168-176, 1973
- 9) Bazaraa, H.D. and Shetty, C.M.: Non-Linear Programming: Theory and Algorithms, John and Wiley, 1979
- 10) Prashjer, J. N. and Toledo: Adaptation of the gradient projection algorithm for the traffic assignment problem with side constraints, 80th Transportation Research Board, CD-ROM, 2000
- 11) Bertsekas, D. and Gallager: Data Networks, Prentice-Hall, NJ, pp.365-478, 1992
- 12) Jayakrishnan, R. *et al.*: A faster path-based algorithm for traffic assignment, Transportation Research Record 1443, pp.75-83, 1994
- 13) Inoue, H.: A computation method for equilibrium traffic assignment, Proceedings of JSCE, No.313, pp.125-133, 1981
- 14) Daganzo, C. F.: On the traffic assignment problem with flow dependent costs—I, II, Transportation Research, Vol.11, 433-441, 1977
- 15) Yang, H. and Yagar, S.: Traffic assignment and traffic control in general freeway arterial corridor systems. Transportation Research, 28B, pp.463-486, 1994
- 16) Tobin, R. L. and Friesz, T. L.: Sensitivity analysis for equilibrium network flow, Transportation Science, Vol.22, pp.242-249, 1988
- 17) Cheng, L.: Road network reliability based on traffic flow equilibrium, Ph.D. thesis of Kyoto University, Chapter 4, 57-97, 2002

---

## Application Quasi-Newton Methods to Capacitated UE Assignment

By Lin Cheng Yasunori Iida Nobuhiro Uno

In this study we proposed two Quasi-Newton methods to deal with traffic assignment in the capacitated network. The methods combine Newton formula, column generation and penalty techniques. The first method employ the gradient of the objective function to obtain an improving feasible direction scaled by the second-order derivatives. The second one is to employ Rosen gradient to obtain an improving direction scaled by the corresponding origin-destination demand. Both methods make line search to obtain an optimal step size to guarantee feasibility of either path or link flow. The proposed methods are of fast convergence and high accuracy at the expense of saving path information. Numerical examples verify their efficiency and stability, as well as usefulness of the path flow pattern reserved. The Quasi-Newton method with straight gradient demonstrates more stability than that with Rosen gradient for capacitated traffic assignment.

## 容量制約条件つき交通量配分への準ニュートン法の適用

程 琳・飯田 恭敬・宇野 伸宏

本稿においてはリンク容量制約条件を明白に考慮した交通量配分の解析法として、準ニュートン法を基礎した二つの方法を提案した。第1の方法は目的関数の勾配を直接に利用して探索方向を定め、一一次次元最適化するものである。第2の方法は目的関数のRosen勾配を利用して探索方向を定める点が第1の方法と異なる。両者とも列生成法と内ペナルティー法と組み合わせて、経路交通量に基づき計算する方法である。制約つき交通量配分問題において提案された準ニュートン法が収束性と解の精度の点で優れていることを計算例で示した。また前者はより後者より安定性、迅速性の点で優れているといえよう。準ニュートン法の副産物としての経路情報はネットワーク分析に利用可能でもある。