Primal Method for Determining the Most Likely Route Flows in Large Road Networks

Hillel Bar-Gera*

*Department of Industrial Engineering and Management, Ben-Gurion University
P.O.B. 653
Be'er-Sheva, 84105 Israel
bargera@bgumail.bgu.ac.il

1 Introduction

An important concern in most travel forecasting models is the way travelers choose their routes. The user-equilibrium (UE) model (Wardrop, 1952) assumes that all used routes are minimum cost routes under prevailing (congested) traffic conditions. Further assuming that link costs are separable monotonically increasing functions of link flows allows to determine uniquely total flows on all links in the network. Total link flows determine link costs, route costs, and the set of minimum cost routes, referred to as the set of UE routes. Although the set of UE routes is unique, finding it computationally is not a trivial task, as discussed below. A greater challenge is to determine the distribution of flows between UE routes, as it requires additional behavioral assumptions.

The ability to find the set of used routes and the flows on these routes is highly desirable from both theoretical and practical perspectives. In many practical applications route flows are essential to the analysis. For example: assessing the appropriate contributions of different communities to a transportation improvement project (Rossi et al. 1989); deriving OD flows in a sub-region from a regional model (Hearn et al. 1984); estimating emissions in consideration of cold starts; and more. In addition, practitioners examine the set of used routes to gain understanding about the model behavior and to verify its validity. Recently, researchers began to explore the structure of the set of UE routes as well. Harris (2002) investigated the geographic patterns created by the set of UE routes. Xiong (2002) explored the connection between the number of alternative routes for each OD pair and their location. The sets of routes analyzed in these studies were found by the method described in this paper.

To date, several researchers have addressed the non-uniqueness of route flows in the UE model. Rossi et al. (1989) suggested that among all route flow patterns that satisfy the UE condition, the entropy-maximizing pattern is the most likely one. Preliminary behavioral interpretation for entropy maximization has been presented by Bar-Gera and Boyce (1999). Janson (1993) presented a link-based computational procedure for the Maximum Entropy User Equilibrium (MEUE) problem, using successive stochastic user-equilibrium (SUE) approximations of the original problem. Dual methods for MEUE consist of two stages, in the first stage total link flows and the set of UE routes are determined by a traffic assignment algorithm; in the second stage route flows are determined using either iterative balancing (Bell and Iida, 1997) or by conjugate gradient (Larsson et al. 1998). It is important to note that in these dual methods all routes chosen in the first stage will receive a positive (non-zero) amount of flow in the second stage. Therefore, it is essential to make a good choice of routes in the first stage. In addition, during the iterative process the dual solution is typically primal infeasible, meaning that total link flows are different from those obtained in the first stage. These mismatches imply loss of precision in meeting the UE conditions, which is only regained as the solution converges, possibly doubling the overall computation time (according to results by Larsson et al. ,1998). Linearly dependent link flow constraints, as demonstrated by example in Bell and Iida (1997), create additional difficulty for dual methods, as they lead to non-unique dual solutions and potentially to convergence problems. To summarize, while dual methods are perhaps easier to implement, they suffer from several essential shortcomings that may be better resolved by a primal method.

The main goal of this research is to develop an efficient primal method for the MEUE problem. In this method, once the UE problem has been solved to sufficient accuracy, route flows are modified without changing total links flows at all, thus maintaining the UE precision achieved in the first stage throughout the iterative process. Direct consideration of the MEUE optimality conditions to derive ascent directions guarantees convergence. Their special structure is utilized to maximize algorithm efficiency. An intermediate goal of this research is to develop a coherent practical methodology for identifying the set of UE routes, which is in fact an important goal by itself as discussed above. Interestingly, the proposed conditions for choosing an appropriate set of routes are closely related to the optimality conditions of the MEUE problem. Satisfying these conditions is therefore critical in both primal and dual methods for the MEUE problem.

2 The set of user-equilibrium routes

The UE traffic assignment problem, like most other non-linear problems, can only be solved to a limited level of precision. At the exact equilibrium there are often several UE routes with equal costs; however, in an approximate solution, even if it is a very good one, the approximated costs of these UE routes are usually not exactly equal. In many cases, for each OD pair, only one of the routes attains the minimum among the approximated costs, while the cost of other UE routes is slightly higher than the minimum. Similar to previous studies (e.g. Larsson et al. 1998), we suggest to choose all the routes that are within a certain *acceptance gap*, $g_a > 0$, from the minimum. To be more precise, suppose that $\mathbf{t}$ is the vector of link costs according to the approximate solution obtained at the assignment stage, $c_r(\mathbf{t})$ is the implied cost of route r, and $C_{pq}(\mathbf{t}) = \min\{c_r(\mathbf{t}) : r \in R_{pq}\}$ is the minimum cost from origin p to destination q under these conditions; define the *excess cost* of route r from origin p to destination q as $ec_r(\mathbf{t}) = c_r(\mathbf{t}) - C_{pq}(\mathbf{t})$ and choose all routes such that $ec_r(\mathbf{t}) < g_a$. Note that if $\mathbf{t}^*$ is the vector of exact equilibrium link costs then by definition $ec_r(\mathbf{t}^*) = 0$ for all UE routes and $ec_r(\mathbf{t}^*) > 0$ for all other routes. Since the number of a-cyclic routes is finite, we can define a strictly positive *equilibrium rejection gap*, $g_r^* > 0$, as the minimum equilibrium excess cost on all non-UE routes. Therefore, if the solution is sufficiently well converged, it is possible

Network	Zones	Nodes	Links	Accept	Reject	Consistency	# routes
Sioux Falls	24	24	76	1.35E-13	0.699	5.2E+12	770
Barcelona	110	1020	2522	1.95E-14	5.20E-07	2.7E+7	11,295
Winnipeg	147	1052	2836	2.34E-13	3.09E-05	1.3E+8	9,880
Tucson	646	3603	9619	7.47E-12	5.33E-06	7.1E+5	1,568,387
Chicago sketch	387	933	2950	1.71E-13	2.42E-04	1.4E+9	127,248
Chicago regional	1790	12,982	39,018	2.64E-12	9.32E-10	345	93,026,894

Table 1: Sets of routes obtained according to the consistency principle.

to find the exact set of UE routes by choosing an acceptance gap $0 < g_a < g_r^*$ such that for all UE routes $ec_r(\mathbf{t}) < g_a$ and for all non-UE routes $ec_r(\mathbf{t}) > g_a$. The remaining challenge is to choose the acceptance gap and to determine if the solution is sufficiently well converged.

Unfortunately, we do not have an ultimate answer for these questions, but we do offer considerations that should be taken into account, and demonstrate their usefulness in large-scale numerical experiments. The first and obvious consideration is the set of routes actually used in the approximate solution, especially if residual flows from suboptimal routes are eliminated by the assignment algorithm, which is typically the case for route-based and origin-based algorithms. The inclusion of all used routes in the set of chosen routes guarantees the existence of a feasible solution to the entropy maximization problem.

The second consideration, which is not as obvious, is based on a fundamental property of the set of UE routes, as any other set of minimum cost routes, that it guarantees consistent consideration of alternative route segments. Meaning, that if a pair of alternative route segments is considered by some travelers, then other travelers with the same priorities (same class), but possibly traveling from a different origin to a different destination, either consider both alternative segments or none of them. Formally, let K_{ra} indicate the number of times route r traverses link a, and let $\mathbf{K}_r$ be the vector of K_{ra} for all the links. We say that a set of routes R^0 is n-consistent if for every sequence of n OD pairs $p_i q_i$ with two alternative routes $r_i, r'_i \in R_{p_i q_i}$ for each OD pair, such that $r_i \in R^0$ for $i = 1 \ldots n$ and $\sum_{i=1}^{n}(\mathbf{K}_{r_i} - \mathbf{K}_{r'_i}) = 0$ then $r'_i \in R^0$. We show that the condition of 2-consistency is equivalent to the intuitive condition of consistent consideration of alternative segments, and discuss the implications of higher levels of consistency using graphical examples.

Our interest in consistent sets of routes stems from the fact that any set of minimum cost routes, and particularly the set of UE routes, is *completely consistent*, meaning that it is n-consistent for all n. Therefore, a set of routes with higher level of consistency is more likely to be the set of UE routes. We show by example that any predetermined acceptance gap may lead to an inconsistent set of routes. However, we show that if for a given vector of link costs $\mathbf{t}$, there exist an acceptance gap $g_a > 0$ and a rejection gap $g_r > n \cdot g_a$ such that there are no routes with $g_a < ec_r(\mathbf{t}) < g_r$, then the set of routes with $ec_r(\mathbf{t}) < g_a$ is n-consistent. Numerical results for several networks obtained according to the consistency principle are presented in Table 1, including the acceptance gap g_a, the rejection gap g_r, the guaranteed level of consistency and the number of routes. Note that all of these solutions are extremely well converged, with less than a nanosecond maximum excess cost (acceptance gap). The high levels of consistency suggest that these sets of routes are quite possibly the exact sets of UE routes.

We explore several non-trivial theoretical issues regarding the above-mentioned definition of consistency. We show by graphical examples that for any value of n there are sets of routes that are n-consistent but not completely consistent. We demonstrate that it is not always possible to extend an existing set of routes to a completely consistent one, using examples of finite (small) sets of a-cyclic routes where the condition of 2-consistency requires adding an infinite number of cyclic routes. The last and perhaps most surprising result, based on Gallager (1977), is that any finite completely consistent set of routes is the set of minimum cost routes under some (arbitrary) strictly positive values of link costs.

Once we choose a set of routes R^0, we may be interested to verify whether it is consistent or not by searching for consistency conditions that *are not* satisfied by R^0; however, there does not seem to be any apparent efficient way of doing that. We may also be interested in searching for consistency conditions that *are* satisfied by R^0, that is sequences of pairs of alternative routes $r_i, r'_i \in R^0_{p_i q_i}$ such that $\sum_{i=1}^{n}(\mathbf{K}_{r_i} - \mathbf{K}'_{r_i}) = 0$. The characterization of the consistency conditions that are satisfied by R^0 provides interesting insights about the structure of R^0 and about the frequency of higher order conditions of consistency. This characterization is also useful for solving the MEUE problem as discussed in the next section.

We aim to find a set of basic consistency conditions for R^0, such that all other consistency conditions satisfied by R^0 would be equivalent to sums of basic ones. Note that indeed any sum of consistency conditions is also a consistency condition. A consistency condition is trivial if it corresponds to reordering of the routes, that is if $r_i = r'_{g(i)}$ for some permutation $g : \{1 \dots n\} \to \{1 \dots n\}$. Consistency conditions are equivalent if the difference between them is trivial. Now, consider the difference vector $\mathbf{d} = \mathbf{K}_r - \mathbf{K}'_r$ for $r, r' \in R_{pq}$ as a connection between the routes r and r'. (A connection by $\mathbf{d}$ or by $-\mathbf{d}$ is considered to be the same.) Suppose that a set of difference vectors D connects R^0 in the sense that every pair of alternative routes is connected through a sequence of vectors in D. Formally, we assume that for every OD pair pq, there exists a spanning tree $\left\{R^0_{pq}, T^0_{pq}\right\}$ such that every edge $\{r, r'\} \in T^0_{pq}$ is directly connected by $\mathbf{K}_r - \mathbf{K}_{r'} = \mathbf{d} \in D$. It can be shown that any consistency condition is equivalent to one that uses only route pairs that are directly connected by D. Note that in 2-consistency conditions that use edges of T^0, both r_1, r'_1 and r_2, r'_2 must be connected by the same vector $\mathbf{d} \in D$. Therefore we choose for every vector $\mathbf{d} \in D$ a global representative edge, $E(\mathbf{d})$, as well as origin-specific representatives $E_p(\mathbf{d})$ for every origin where such representative exists. The 2-consistency condition created by an edge and the relevant origin representative is considered basic within origin. The 2-consistency condition created by an origin representative and the relevant global representative is considered basic between origins. To complete the analysis we find an integer basis for the linear dependencies between vectors in D, and from every element of this basis we construct an additional basic consistency condition using the global representatives of the participating vectors. We show that every consistency condition is equivalent to sums of basic conditions.

The usefulness of this analysis depends mainly on the choice of D. We present a method for choosing D and the resulting breakdown of basic consistency conditions for different networks, as shown in Table 2. As discussed in the next section, these are associated with MEUE optimality conditions. Note that consistency conditions of level higher than 2 must imply linear dependency in D. The small numbers of these linear dependencies demonstrate that higher-level consistency conditions are relatively rare in real life networks.

Network	optimality conditions			constraints	
	2-consistency				independent
	within origins	between origins	other (in D)	OD flows	link flows
Sioux Falls	138	70	5	528	29
Barcelona	2,268	962	38	7,922	119
Winnipeg	2,430	2,942	2	4,345	161
Tucson	1,179,105	22,911	8	366,087	276
Chicago sketch	27,261	6,204	4	93,513	266
Chicago regional	89,822,183	901,656	92	2,297,945	5019

Table 2: Breakdown of route flow solution dimension

3 Maximizing route flow entropy

To obtain primal optimality conditions for the MEUE problem one must first identify the subspace of feasible directions, that is the set of flow changes over the routes in the chosen set R^0 that do not change OD flows or total link flows. To maintain OD flows, the change must be a sum of shifts of flows between alternative routes. A shift of flow between routes $r, r' \in R^0_{pq}$ can be viewed as a sequence of shifts along the edges of T^0_{pq}. The change in total link flows as a result of shifting flow δ along the edge $r, r' \in T^0_{pq}$ is $\delta \cdot (\mathbf{K}_r - \mathbf{K}_{r'})$ where $\mathbf{K}_r - \mathbf{K}_{r'} = \mathbf{d} \in D$. Therefore, any feasible direction corresponds to a linear dependency in D. Given the integer basis for these linear dependencies discussed above, it is clear that any feasible direction is associated with a consistency condition. This fact emphasizes the importance of the proposed condition of consistency in choosing the set of routes R^0; since if certain routes that are needed according to a consistency condition are not included in R^0, then clearly it will not be possible to satisfy the corresponding MEUE optimality condition. (As discussed above, adding more routes than necessary, that is adding non-UE routes, is also problematic as these routes will get positive flows, thus violating the UE condition.)

The characterization of consistency conditions presented above can therefore be viewed as a characterization of feasible directions, providing the foundation for a primal method for the MEUE problem. A general strategy can be to consider all basic conditions/directions repeatedly, in some order, and to maximize the entropy by line search along each of these directions. This strategy guarantees that total link flows will not change at all during the iterative process, thus maintaining the precision of the UE solution obtained in the assignment stage. It also guarantees that entropy will never descend. Convergence can be proven by showing that the algorithmic map is closed (in fact continuous).

Computational efficiency of the proposed primal approach can be substantially enhanced using an origin-based representation of the solution (Bar-Gera, 2002). Any a-cyclic origin-based solution has an immediate route flow interpretation that satisfies the entropy maximization optimality conditions within each origin (Bar-Gera and Boyce, 1999). The optimality conditions associated with a single difference vector $\mathbf{d} \in D$ imply that the same proportions of flows between the two alternative route segments represented by $\mathbf{d}$ apply to all route pairs connected directly by $\mathbf{d}$. All of the optimality conditions for a single vector $\mathbf{d}$ can be handled simultaneously by computing the total flow on the two segments and applying the resulting ratio to all origins. A special data structure has been developed that allows performing these computations efficiently by considering only the links that are part of the alternative segments. The optimality conditions associated with non-trivial linear dependencies in D require additional effort; however, given their negligible number, it seems that for all practical purposes good approximations of the entropy maximizing solution can be found even if these conditions are ignored. The precision in satisfying the optimality conditions that are considered can be measured by the amount of flow that must be shifted to satisfy them (one at a time). Without the proposed algorithm, necessary flow shifts reach 100-1000vph in the different networks. The algorithm finds solutions with maximum flow shift of 1vph for the different networks in 0.01, 0.1, 0.25, 2.65, 0.3, and 310 seconds respectively.

4 Conclusions

This paper presents a new formal condition of consistency for sets of routes, which is satisfied by any set of user-equilibrium routes. The main implication of the proposed condition is consistent consideration of alternative route segments. A methodology for choosing a consistent set of routes from an approximate traffic assignment solution is presented as well. The connection between primal optimality conditions and consistency conditions provides the basis for a new primal algorithm for the maximum entropy user-equilibrium problem. As a primal algorithm, feasibility is maintained throughout the iterative process, which means that total link flows remain without any modification in this process, thus maintaining the precision of the solution to the user-equilibrium problem obtained by the traffic assignment algorithm. Numerical examples demonstrate that the proposed approach performs well in a variety of networks, including real-life large-scale networks. This algorithm is important for practical applications such as sub-region models, impact fee assessment and more.

5 References

H. Bar-Gera, “Origin-based Algorithm for the Traffic Assignment Problem,” *Transportation Science* 36 (4), 398–417, (2002).

H. Bar-Gera and D. Boyce, “Route flow entropy maximization in origin-based traffic assignment,” In *Transportation and Traffic Theory*, Proceedings of the 14th International Symposium on Transportation and Traffic Theory, Jerusalem, Israel, 1999. A. Ceder, ed., Elsevier Science, Oxford, UK, 397–415, (1999).

M.G.H. Bell and Y. Iida, *Transportation Network Analysis*, Wiley, New York, (1997).

R.G. Gallager, “Loops in Multicommodity Flows,” In *Proceedings of the Tenth IEEE Conference on Decision and Control*, New Orleans, LA, 819–825, (1977).

B. Harris, “Multiple Paths Through a Network,” Presented at the 42nd Congress of the European Regional Science Association, Dortmund, Germany, (2002).

D.W. Hearn, “Practical and theoretical aspects of aggregation problems in transportation planning models,” In *Transportation Planning Models*, Proceedings of the course given at The International Center for Transportation Studies (ICTS), Amalfi, Italy, October 11-16, 1982, M. Florian, ed., North-Holland, Amsterdam, 257–287, (1984).

B.N. Janson, “Most likely origin-destination link uses from equilibrium assignment,” *Transportation Research* 27B, 333–350, (1993).

T. Larsson, J. Lundgren, M. Patriksson and C. Rydergren, "Most likely traffic equilibrium route flows - analysis and computation," In *Optimization Methods for Analysis of Transportation Networks*, C. Rydergren, Theses No. 702, Linkopings Universitet, Sweden, (1998).

J.G. Wardrop, "Some Theoretical Aspects of Road Traffic Research," In *Proceedings of the Institution of Civil Engineers*, Part II, 1, 325–378, (1952).

T.F. Rossi, S. McNeil and C. Hendrickson, "Entropy model for consistent impact fee assessment," *Journal of Urban Planning and Development/ASCE*, 115, 51–63, (1989).

Q. Xiong, Comparison of User-Optimal and System-Optimal Traffic Assignments for the Chicago Regional Road Network, M.Sc. Thesis, University of Illinois at Chicago, U.S.A., (2002).