Distributionally Robust Markovian Traffic Equilibrium*

Selin Damla Ahipaşaoğlu[†] Uğur Arıkan[‡] Karthik Natarajan[§]

First Version: September 2017, Second Version: May 2018, Third Version: October 2018, Fourth Version: February 2019

Abstract

In a Markovian traffic equilibrium model, users move towards their destinations by a sequence of successive link choices using a discrete choice model at each node, taking congestion into account. While a convex optimization formulation is available to compute the equilibrium flows for a continuous distribution of link utilities, practical applications have thus far been mainly restricted to the multinomial logit model and its variants. In this paper, we relax the assumption of a complete joint distribution of link utilities to only knowledge on the marginal distributions and propose a new convex optimization formulation for a distributionally robust Markovian traffic equilibrium. The formulation is provably efficiently solvable and has the flexibility of allowing for general marginal distributions, thus capturing different types of non-identical, skewed and heavy tailed distributions at the link level.

Keywords: Markovian traffic equilibrium, distributionally robust, convex optimization

*This project was partly funded by the the MOE Tier 2 grant number MOE2016-T2-2-148 on 'On the Interplay of Choice, Robustness and Optimization in Transportation' and the SUTD-MIT IDC grant number IDG21700101 'Design of the Last Mile Transportation System: What Does the Customer Really Want?'.

[†]Engineering Systems and Design, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372, ahipasaoglu@sutd.edu.sg

[‡]Engineering Systems and Design, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372, ugur_arikan@sutd.edu.sg

[§]Engineering Systems and Design, Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372, karthik_natarajan@sutd.edu.sg

1 Introduction

Traffic equilibrium models are fundamental to the analysis of transportation systems. The inputs to such models are the network topology, origin-destination (OD) pairs, demands of OD pairs and link performance functions. The goal is to estimate the equilibrium link flows. Daganzo and Sheffi (1977) relaxed the perfect information assumption of the user equilibrium (Wardrop, 1952) and proposed a stochastic user equilibrium (SUE) with a discrete choice model to capture randomness in user route choice behavior. However, traffic networks have a huge number of routes. A common approach is to generate a smaller set of plausible routes with respect to some criteria (Bekhor et al., 2006). These algorithms fall in the class of deterministic methods, and ignore the effect of congestion. While there has been much focus on developing route based SUE models,there are computational drawbacks due to the possibility of large choice sets. One approach in this context is to assume a sequential Markovian decision-making process where the route choice of a user is determined by successive link choices that is made independently of how the users reached the current node (see Akamatsu (1996, 1997); Bell (1995)). Evaluating the steady state flows in this Markov chain allows the system planner to evaluate all routes without explicitly generating them. However, this idea has attracted lesser attention than the route based models due to the difficulty in computing the choice probabilities and its implementation has been primarily limited to the multinomial logit (MNL) model and some of its variants. In the traffic assignment literature, the MNL model has received particular criticism for its independently and identically distributed (i.i.d) error term assumption (Sheffi, 1985). Baillon and Cominetti (2008) develop the Markovian choice model to allow for a general distribution on the error terms and propose a convex optimization formulation to compute the traffic equilibrium flows. However, it is not clear if this formulation is efficiently solvable for instances beyond the multinomial logit. Furthermore, an underlying assumption is that the probability distribution of the error terms in the link utilities is known to the system planner, which in practice is rarely the case.

In this paper, we propose a distributionally robust optimization perspective by relaxing the assumption that the distribution of the error terms of the link utilities is known to the system planner. The main contributions of the paper are as follows:

(a) We define a new Markovian marginal distribution model (referred to as M-MDM) and propose a convex optimization formulation to compute the link choice probabilities. This formulation can recreate the entropy formulation that was first developed in Akamatsu (1996, 1997); and the Recursive Logit model (Fosgerau et al., 2013) and Nested Recursive Logit model (Mai et al., 2015), which have been shown to be useful on real traffic networks.

(b) Unlike the logit based models, the new model relaxes the i.i.d error assumption and allows for using new families of marginal distributions such as, logistic, student-t or normal, while preserving computational tractability. Our model also does not need the joint distribution to be completely specified which might itself be complicated to specify, but rather only the marginal distributions of link utilities need to be specified. This results in lower computational and cognitive burden on the system planner. We showcase the flexibility of the model in incorporating heterogeneity at the link level and the shape of the marginal distribution thereby capturing a wider range of user choice behavior.

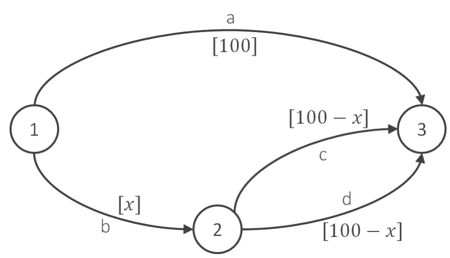
(c) We propose a distributionally robust optimization traffic equilibrium using the marginal distribution model (referred to as MTE-MDM). For this model, we develop a new convex optimization formulation and use optimization methods to compute the equilibrium flows. We provide examples of marginal distributions under which the equilibrium flow is provably efficiently computable. Our experiments indicate that the solution approach is applicable to large traffic networks.

2 Markovian Choice

In a Markovian choice model, a user chooses one of the links emanating from a node irrespective of how she reached that node. This choice behavior is fundamentally different from traditional route choice in the sense of spatial knowledge (see Maher and Hughes (1997)). The Markovian choice fits the situations where the user is capable of linking locations, but lacks an overall understanding of the spatial organization; while route choice represents the behavior of more professional users having a proper spatial knowledge, such as taxi drivers (see also Gould (1989); Kuipers (1978); Mark and McGranaghan (1986); Stern and Leiser (1988)).

Another difference is related to the overlapping route problem. Consider the loop hole network in Figure 1 where the numbers in brackets are the expected link costs. There are three routes with equal travel costs. It is intuitive to expect the choice probability of upper route to be 1/3 when $x = 0$, and 1/2 when $x = 100$ (see Baillon and Cominetti (2008)). In route choice models, this overlapping effect cannot be modeled unless correlation is

Figure 1: Loop hole network.

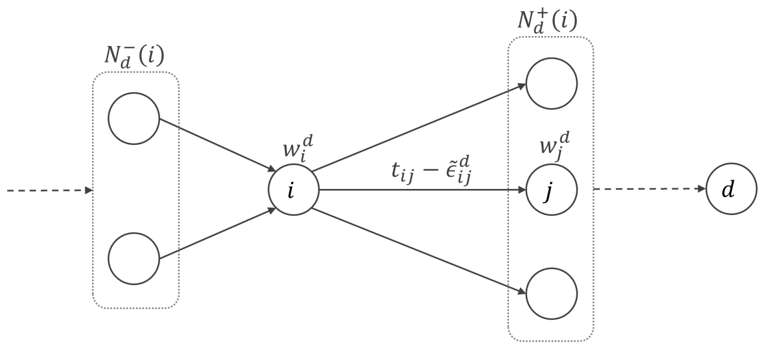


captured. Under Markovian choice, the overlapping problem can be handled by using different choices of distributions of the error terms in link costs. In order to illustrate, consider the same example with given expected costs, but also assume that the observed choice probability of the upper route is 0.40 and that of the lower routes are 0.25 and 0.35. Our model provides the flexibility to capture such choice behavior by relaxing the identical error term assumption and allowing heterogeneity at the link level. Note that the multinomial logit model cannot capture this effect.

2.1 Notation

Let $G = (N, A)$ denote a directed traffic network with a set of nodes N and a set of links A, where W is the set of OD pairs and $D \subseteq N$ is the set of destination nodes in the set of OD pairs. We define $G_d = (N_d \cup \{d\}, A_d)$ as the subnetwork for destination d, where A_d is the subset of links $(i, j) \in A$ for which i is reachable from an origin o such that $(o, d) \in W$ without visiting d, and destination d is reachable from j. The subset N_d contains all tail nodes of links in A_d. Finally, we let $N_d^+(i)$ denote the out-neighbourhood of node i and $N_d^-(i)$ denote the in-neighbourhood of node i in G_d (see Figure 2). The choice at a node is independent of the previous nodes the user visits. The uncertainty in the choice behavior is introduced by modelling the link cost of arc (i, j) as a random variable which is equal to $t_{ij} - \tilde{\epsilon}_{ij}^d$ where t_{ij} is the deterministic cost of travelling on the link and $\tilde{\epsilon}_{ij}^d$ is a random utility error term of using the link towards destination d. That term captures the preference of the user that is unobservable to the system planner. Let w_j^d denote the expected minimum cost of travelling from node j to destination d. A Markovian choice model for the users travelling towards destination d

Figure 2: Illustration of network terminology.



is characterized as the solution to the following system of equations:

$$w_i^d = E_{\theta_{id}} \left[\min_{j \in N_d^+(i)} \left\{ t_{ij} - \tilde{\epsilon}_{ij}^d + w_j^d \right\} \right], \qquad \forall i \in N_d, \tag{1}$$

$$w_d^d = 0, \tag{2}$$

where θ_{id} is the joint distribution of the error terms $\{\tilde{\epsilon}_{ij}^d; j \in N_d^+(i)\}$. The error terms of the links emanating from two different nodes are assumed to be independent, while the error terms of the links exiting a particular node might be correlated. The corresponding conditional link choice probability of selecting link (i,j) at node i is defined as:

$$p_{ij}^d = \mathrm{P}_{\theta_{id}} \left\{ j = \operatorname*{argmin}_{k \in N_d^+(i)} \left\{ t_{ik} - \tilde{\epsilon}_{ik}^d + w_k^d \right\} \right\}, \qquad \forall (i,j) \in A_d.$$

Let R_d be the set of routes with destination d. This set might be countably infinite to allow for the possibility of cycles. Let $\delta_{ij,r}$ be the number of times link (i,j) is used in a route r. Then, route choice probabilities are given by $p_r = \prod_{(i,j) \in A_d} (p_{ij}^d)^{\delta_{ij,r}}, \forall r \in R_d, d \in D$; and the route flows by $y_r = h_{o_r}^d \cdot p_r$, where o_r is the origin of route r and h_o^d is the demand of the OD pair (o,d). Finally, we define $x_{ij}^d = \sum_{r \in R_d : (i,j) \in r} y_r, \forall (i,j) \in A_d$ as the flow towards a designated destination; and $f_{ij} = \sum_{d \in D : (i,j) \in A_d} x_{ij}^d, \forall (i,j) \in A$ as the total link flow.

2.2 Literature Review

Building on the earlier work of Dial (1971), Akamatsu (1996, 1997) and Bell (1995) proposed the Markovian logit model. In this model, the random utility error terms $\widetilde{\epsilon}_{ij}^d$ are assumed to be i.i.d Gumbel random variables with a dispersion parameter β. This model is also sometimes referred to as the Recursive Logit model (see Fosgerau et al. (2013)). In this model, a closed form expression for the conditional link probability expression in the w_j^d variables is given as follows:

$$p_{ij}^d = \frac{\exp\left[-\beta\left(t_{ij} + w_j^d\right)\right]}{\sum_{k \in N_d^+(i)} \exp\left[-\beta\left(t_{ik} + w_k^d\right)\right]}, \qquad \forall (i,j) \in A_d. \tag{3}$$

Solving for the Markovian logit model reduces to solving the following set of nonlinear equations in the w_j^d variables:

$$w_i^d = -\frac{1}{\beta} \ln\left(\sum_{j \in N_d^+(i)} \exp\left[-\beta\left(t_{ij} + w_j^d\right)\right]\right), \qquad \forall i \in N_d, \tag{4}$$

$$w_d^d = 0. \tag{5}$$

A straightforward approach to solve this system of nonlinear equations is to convert it to a system of linear equations through an exponential transformation of the w_j^d variables (see Fosgerau et al. (2013) for such an approach). Recently, Mai et al. (2015) extended this to Nested Recursive Logit model that allows for correlations among the error terms. However this model is much harder to solve as there is no straightforward transformation of the system of nonlinear equations to a system of linear equations.

Baillon and Cominetti (2008) proposed a general Markovian Traffic Equilibrium (MTE) model for the congested case with general distributions. Let $\tau_{ij}(f_{ij})$ be a strictly increasing continuous cost function which defines the cost of a link as a function of its flow f_{ij}. Then the MTE is defined as the solution to a fixed point problem in the variables $(f_{ij}, t_{ij}, w_i^d, x_{ij}^d, n_i^d)$ where

n_i^d is the number of users at node i moving towards destination d as follows:

$$
\begin{aligned}
& t_{ij} = \tau_{ij}\left(f_{ij}\right), && \forall (i,j) \in A, \\
& w_i^d = E_{\theta_{id}}\left[\min_{j \in N_d^+(i)} \left\{t_{ij} - \widetilde{\epsilon}_{ij}^d + w_j^d\right\}\right], && \forall i \in N_d, d \in D, \\
& w_d^d = 0, && \forall d \in D, \\
& n_i^d = h_i^d + \sum_{k \in N_d^-(i)} x_{ki}^d, && \forall i \in N_d, d \in D, \\
& x_{ij}^d = n_i^d \cdot \mathrm{P}_{\theta_{id}}\left\{j = \operatorname*{argmin}_{k \in N_d^+(i)} \left\{t_{ik} - \widetilde{\epsilon}_{ik}^d + w_k^d\right\}\right\}, && \forall (i,j) \in A_d, d \in D, \\
& f_{ij} = \sum_{d \in D : (i,j) \in A_d} x_{ij}^d, && \forall (i,j) \in A.
\end{aligned}
$$

Under mild assumptions on the distribution of the random utilities which is denoted by $\boldsymbol{\theta}$, the authors showed that the MTE exists and is unique. Furthermore, the link costs at equilibrium are obtained as the solution to the following unconstrained convex optimization formulation:

$$
Z\left(\boldsymbol{\theta}\right) = \max_{\mathbf{t}} \ \sum_{d \in D} \sum_{i \in N_d} h_i^d \cdot w_i^d(\mathbf{t}) - \sum_{(i,j) \in A} \int_0^{t_{ij}} \tau_{ij}^{-1}(\omega) d\omega, \qquad (6)
$$

where $w_i^d(\mathbf{t})$ is a concave and smooth function of the link costs and is implicitly defined in (1)-(2). The equilibrium link flow is given as $f_{ij} = \tau_{ij}^{-1}(t_{ij}^*)$ where t_{ij}^* is the unique optimal solution in the convex optimization formulation (6). The application of this model has however been primarily restricted to the logit model. This largely stems from the challenge in computing the choice probabilities for general distributions. Furthermore in link based models, the system planner also needs to solve a set of nonlinear equations in the w_j^d variables unlike the route based models; hence, it is not clear if this formulation is efficiently solvable in theory. Another important assumption that is implicitly made in the model but is itself subject to scrutiny is that the system planner has complete knowledge on the distribution of the random utilities. In this paper, we derive a new class of Markovian traffic equilibrium models which relaxes this assumption while at the same time preserving computational tractability.

A related class of link based methods uses an efficiency criterion similar to the STOCH approach of Dial (1971). Maher (1998) developed an efficient method to find the equilibrium under the logit assumption. The methods in this class allow for efficient stochastic network loading with the cost of being restrictive in routes that can be used by the commuters. Maher (1992) and Maher and Hughes (1997) proposed a stochastic network loading method with normal error terms, referred as the Stochastic Assignment Model (SAM) where users follow a dynamic choice process. This approach has similarities with M-MDM having normal marginals; except that SAM uses Clark's approximation (Clark, 1961) to dynamically build an approximate joint distribution, while M-MDM uses a robust optimization approach.

2.3 Distributionally Robust Markovian Traffic Equilibrium

To provide a distributionally robust perspective on the MTE, we start by reformulating (6) as a constrained convex optimization problem in the decision variables (t_{ij}, w_i^d) as follows:

$$Z(\boldsymbol{\theta}) = \max_{\mathbf{t},\mathbf{w}} \sum_{d \in D} \sum_{i \in N_d} h_i^d \cdot w_i^d - \sum_{(i,j) \in A} \int_0^{t_{ij}} \tau_{ij}^{-1}(\omega) d\omega \tag{7}$$

$$\text{s.t. } w_i^d \le E_{\theta_{id}} \left[\min_{j \in N_d^+(i)} \left\{ t_{ij} - \tilde{\epsilon}_{ij}^d + w_j^d \right\} \right], \quad \forall i \in N_d, d \in D, \tag{8}$$

$$w_d^d = 0, \qquad \forall d \in D. \tag{9}$$

Note that at optimality, the inequalities in (8) will be tight as $h_i^d > 0$.

However an important assumption in the model is that the joint distribution of the error terms is known to the system planner. For a joint distribution θ_{id} with continuous marginals, Sklar's theorem (Sklar, 1959) shows that it can be represented in terms of the univariate marginal distribution functions and a copula which describes the dependence structure between the variables as follows:

$$\theta_{id} = C_{id}(F_{ij}^d; j \in N_d^+(i)),$$

where F_{ij}^d are the marginal distributions of the error terms $\tilde{\epsilon}_{ij}^d$ for $j \in N_d^+(i)$ and $C_{id}(\cdot) : [0,1]^{|N_d^+(i)|} \to [0,1]$ is a copula function. A wide range of copula functions has been developed to represent positive and negative correlation structures and to incorporate tail dependency, which provides modeling flexibility depending on the available data. On the other hand, it is not clear if the corresponding traffic equilibrium problem is efficiently solvable in all these cases. Furthermore, it is sometimes unclear a priori, what the right choice of copula for the link utility error terms is. As a result, it is very common in the literature to assume independence among the link error terms and use the Gumbel marginal distributions to recreate the logit formula for tractability purposes, or generalizations to models such as nested logit have been developed. In this paper, we propose a distributionally robust optimization perspective to deal with this issue where only the marginal distributions are specified, but the copula is not specified. Computationally, our model has the advantage that there is no need to specify a high dimensional copula for the system modeler. Further, similar to the copula based approach, it allows for a very wide range of marginal distribution families, beyond Gumbel distributions. For example, normal distributions have been proposed previously when the number of route users is sufficiently large (see Watling (2006)). Other traffic researchers have proposed that travel time distributions are often skewed (see Polus (1979); Castillo et al. (2013, 2014)) and used distributions such as lognormal or gamma to model the travel link utilities. Furthermore, in a network with a few but serious traffic incidents, travel times might be significantly different from normal, having heavy tails (Fosgerau and Fukuda, 2012). To capture such situations, heavy tailed marginals such as student-t with parameter 2 can be used. Similarly, when different types of users; say business and leisure drivers use the links, it is natural to use a bimodal distribution such as the mixture of two normal distributions to model the link utilities. On the other hand, the limitation of the model is that by accounting for the worst-case distribution, our result might in some cases be over conservative. It is natural to ask for what types of correlation structures, the proposed framework provides a reasonable fit. In our simulations (see Section 6.1.4), we observe that the model provides accurate link flow predictions when there is a negative correlation among the error terms; such as the cases where each individual has a strong preference for a certain link over others and also provides good predictions for independent distributions. On the other hand, in instances where there is a strong positive correlation among the link utilities, the accuracy of our model's predictions is lower. In such cases, it seems more reasonable to use an appropriately chosen copula.

We now provide a distributionally robust formulation for the MTE which relaxes this assumption as follows. Suppose that the joint distribution θ_{id} of the error terms $\{\widetilde{\epsilon}_{ij}^{d}; j \in N_d^+(i)\}$, is not completely known. Rather it is only known to lie in a set of joint distributions Θ_{id}. Then the distributionally robust Markovian traffic equilibrium is formulated as follows:

$$\max_{\mathbf{t},\mathbf{w}} \sum_{d\in D}\sum_{i\in N_d} h_i^d \cdot w_i^d - \sum_{(i,j)\in A}\int_0^{t_{ij}} \tau_{ij}^{-1}(\omega)d\omega \tag{10}$$

$$\text{s.t. } w_i^d \leq E_{\theta_{id}}\left[\min_{j\in N_d^+(i)}\left\{t_{ij} - \widetilde{\epsilon}_{ij}^d + w_j^d\right\}\right], \quad \forall\theta_{id}\in\Theta_{id}, \forall i\in N_d, d\in D, \tag{11}$$

$$w_d^d = 0, \qquad \forall d\in D, \tag{12}$$

where (11) enforces that the constraint is valid for all joint distributions in the set Θ_{id}. Since the seminal work of Scarf (1958) on the minmax newsvendor problem, there has been significant progress in developing robust and distributionally robust optimization techniques for network flow problems. The applications studied include network design (see Atamturk and Zhang (2007)), dynamic empty container repositioning (see Erera and Savelsbergh (2009)), vehicle routing (see Carlsson and Delage (2013)), path based traffic equilibrium models (see Ahipaşaoğlu et al. (2015, 2016) and Ordóñez and Stier-Moses (2010)) and the traveling salesperson problem (see Carlsson and Behroozi (2017)).

3 Markovian Marginal Distribution Model (M-MDM)

In this section, we introduce a new discrete choice model for a single destination with a fixed link cost and extend the model to the congested case in the next section. We make the following assumptions.

A1 Random utility error terms of the links that emanate from two different nodes are independent of each other.

A2 The joint distribution θ_{id} of the random error terms $\{\widetilde{\epsilon}_{ij}^d; j\in N_d^+(i)\}$ that emanate from a node is only known to lie in a set of distributions with given marginal distributions. This set of distributions is designated as Θ_{id}.

A3 The random utility error term of each link has a finite first moment, namely $E_{\theta_{id}}[|\widetilde{\epsilon}_{ij}^d|] < \infty$ for all $\theta_{id}\in\Theta_{id}$. The support of the random error term is either the whole real line $(-\infty,\infty)$, or the semi-infinite interval $(\underline{\epsilon},\infty)$, where $\underline{\epsilon}$ is the minimum value for the corresponding error term. The cumulative distribution function is strictly increasing on the support and is continuous with a probability density function $f(.) > 0$ on the support.

Assumption A3 on the marginals is satisfied by a wide range of distribution families, such as the exponential, normal, student-t, logistic, Gumbel and gamma distributions to name a few. Consider the node i where the link costs t_{ij} and the expected costs to go from nodes $j \in N_d^+(i)$ denoted by w_j^d are known. In MDM, the minimum expected cost at node i is calculated as follows:

$$w_i^d = \min_{\theta_{id} \in \Theta_{id}} E_{\theta_{id}} \left[\min_{j \in N_d^+(i)} \left\{ t_{ij} - \widetilde{\epsilon}_{ij}^d + w_j^d \right\} \right], \tag{13}$$

and $p_{ij}^d = \mathrm{P}_{\theta_{id}^*}\{j = \mathrm{argmin}_{k \in N_d^+(i)} \{t_{ik} - \widetilde{\epsilon}_{ik}^d + w_k^d\}\}$, where θ_{id}^* is the extremal distribution in (13). For fixed values of w_j^d, Mishra et al. (2014) provided an equivalent reformulation as a convex minimization problem over finite dimensional choice probabilities as follows:

$$w_i^d = \min_{\mathbf{p_i^d}} \left\{ \sum_{j \in N_d^+(i)} \left((t_{ij} + w_j^d) p_{ij}^d - \int_{1-p_{ij}^d}^{1} F_{ij}^{d\,(-1)}(\omega) d\omega \right) : \mathbf{p_i^d} \in \Delta_i^d \right\},$$

where Δ_i^d is a unit simplex defined as follows:

$$\Delta_i^d = \left\{ \mathbf{p_i^d} : \sum_{j \in N_d^+(i)} p_{ij}^d = 1, p_{ij}^d \geq 0, \forall j \in N_d^+(i) \right\}.$$

Under the assumptions on the marginal distributions, this is a strictly convex optimization problem over the unit simplex and the choice probabilities are strictly positive. The Karush-Kuhn-Tucker (KKT) optimality conditions are:

$$p_{ij}^d = 1 - F_{ij}^d \left(\lambda_i^d + t_{ij} + w_j^d \right), \qquad \forall j \in N_d^+(i), \tag{14}$$

$$1 = \sum_{j \in N_d^+(i)} \left[1 - F_{ij}^d \left(\lambda_i^d + t_{ij} + w_j^d \right) \right], \tag{15}$$

where λ_i^d is the dual variable corresponding to the equality constraint $\sum_j p_{ij}^d = 1$. The choice probabilities in this case are easily computed by an efficient line search method over the single variable λ_i^d using the condition (15). Alternatively, the minimum expected cost and the optimal value of the dual variable is obtained by solving the following concave maximization problem:

$$w_i^d = \max_{\lambda_i^d} \left\{ -\lambda_i^d - \sum_{j \in N_d^+(i)} \int_{\lambda_i^d + t_{ij} + w_j^d}^{\infty} \left[1 - F_{ij}^d(\omega) \right] d\omega \right\}. \tag{16}$$

It is easy to verify again that under the assumptions, there exists a unique λ_i^d satisfying the normalization condition (15), which is an optimality condition that arises from (16). We now generalize the result to the Markovian Marginal Distribution Model (M-MDM) as follows.

Definition 1. *Given the link cost vector* $\boldsymbol{t}$*, the link choice probabilities of M-MDM, for the users moving towards destination* d *are characterized as the solution of the following system of equations in the variables* w_i^d*:*

$$w_i^d = \min_{\theta_{id} \in \Theta_{id}} E_{\theta_{id}} \left[\min_{j \in N_d^+(i)} \left\{ t_{ij} - \tilde{\epsilon}_{ij}^d + w_j^d \right\} \right] \qquad \forall i \in N_d, \tag{17}$$

$$w_d^d = 0, \tag{18}$$

where:

$$p_{ij}^d = P_{\theta_{id}^*} \left\{ j = \operatorname*{argmin}_{k \in N_d^+(i)} \left\{ t_{ik} - \tilde{\epsilon}_{ik}^d + w_k^d \right\} \right\}, \qquad \forall (i,j) \in A_d, \tag{19}$$

and θ_{id}^* *is the extremal distribution in* (17).

It is important to note that there exists at most one solution for the M-MDM system (see Appendix A.1). Next, we propose an equivalent convex optimization reformulation to solve the system of equations in M-MDM and compute the probabilities.

Proposition 1. *Given a link cost vector* $\boldsymbol{t}$*, the conditional link choice probabilities of the users moving to destination* d *in M-MDM is calculated as:*

$$p_{ij}^d = 1 - F_{ij}^d \left(\lambda_i^d + t_{ij} + w_j^d \right), \qquad \forall (i,j) \in A_d,$$

where the values of λ_i^d *and* w_i^d *are obtained from the optimal solution to the* *following convex optimization problem:*

$$Z_d^* (\boldsymbol{t}) = \max_{\boldsymbol{w}^d, \boldsymbol{\lambda}^d} \sum_{i \in N_d} h_i^d \cdot w_i^d \tag{20}$$

$$s.t.\ w_i^d \leq -\lambda_i^d - \sum_{j \in N_d^+(i)} \int_{\lambda_i^d + t_{ij} + w_j^d}^{\infty} \left[1 - F_{ij}^d (\omega)\right] d\omega, \quad \forall i \in N_d, \tag{21}$$

$$w_d^d = 0. \tag{22}$$

Proofs of propositions and lemmas are presented in Appendix A. By Proposition 1, link choice probabilities under M-MDM can be obtained as the solution to a convex program. Constraint (21) involves one dimensional integrals, which can be efficiently calculated for a wide range of distribution families. One such family is the student-t distribution with two degrees of freedom. In the following lemma, we show that the problem can be solved in polynomial time for this special case.

Lemma 1. *For the special case where the marginals have student-t distribution with two degrees of freedom, the problem in* (20)-(22) *can be reformulated as a second-order cone programming model.*

The KKT optimality conditions of the M-MDM formulation are equivalent to the constraints of a network flow problem where h_i^d units enter the network at node $i \in N_d$, and $\sum_{i \in N_d} h_i^d$ units leave the network from node d. Let x_{ij}^d be the flow on link (i, j), and n_i^d be the outflow from node i. Using the substitution $x_{ij}^d = n_i^d \cdot p_{ij}^d$, the KKT optimality conditions can be rewritten as $\sum_{j \in N_d^+(i)} x_{ij}^d - \sum_{k \in N_d^-(i)} x_{ki}^d = h_i^d, \forall (i, j) \in A_d$. The choice process can be considered as a Markov chain where nodes are the states and destination d is the single absorbing state. Let $\mathbf{Q^d}$ be an $|N_d| \times |N_d|$ matrix where $(i, j)^{th}$ entry of the matrix, $\mathbf{Q^d_{ij}}$, is equal to p_{ij}^d if $(i, j) \in A_d$ and zero otherwise. Since $p_{ij}^d > 0$, $\mathbf{Q^d}$ is sub-stochastic; and $\mathbf{I} - \mathbf{Q^d}$ is nonsingular (Kemeny et al., 1960), where $\mathbf{I}$ is the identity matrix of size $|N_d|$. The fundamental matrix of the Markov chain is $\mathbf{M^d} = \sum_{n=0}^{\infty} (\mathbf{Q^d})^n = (\mathbf{I} - \mathbf{Q^d})^{-1}$, where $\mathbf{M^d_{ij}}$ is the expected number of times a user entering the system at node i visits node j before reaching the destination. Then, n_i^d can be expressed as follows:

$$n_i^d = \sum_{o \in N_d} h_o^d \cdot \mathbf{M^d_{oi}}, \qquad \forall i \in N_d. \tag{23}$$

Now consider the M-MDM recursion (17). Given optimal link choice probabilities p_{ij}^d, the expression can be rewritten using the primal MDM formulation as follows:

$$\underline{w}_i^d = \sum_{j \in N_d^+(i)} \left(t_{ij} \cdot p_{ij}^d - \int_{1-p_{ij}^d}^{1} F_{ij}^{d\,(-1)}(\omega)\, d\omega \right), \qquad \forall i \in N_d,$$

$$w_i^d = \underline{w}_i^d + \sum_{j \in N_d^+(i)} p_{ij}^d \cdot w_j^d, \qquad \forall i \in N_d.$$

Here, $\underline{w}_i^d$ corresponds to the instantaneous cost and the second term calculates the expected cost from the downstream nodes. In matrix form, the system can be expressed as $\mathbf{W^d} = \underline{\mathbf{W}}^{\mathbf{d}} + \mathbf{Q^d W^d}$ which is equivalent to the following expression:

$$w_i^d = \sum_{j \in N_d} \mathbf{M}_{ij}^{\mathbf{d}} \cdot \underline{w}_j^d \qquad \forall i \in N_d. \tag{24}$$

Using equations (23) and (24), the optimal objective function value of M-MDM given the link choice probabilities can be expressed as follows (see Appendix A.4):

$$\sum_{i \in N_d} h_i^d \cdot w_i^d = \sum_{i \in N_d} \sum_{j \in N_d^+(i)} \left(t_{ij} \cdot x_{ij}^d - n_i^d \int_{1-\frac{x_{ij}^d}{n_i^d}}^{1} F_{ij}^{d\,(-1)}(\omega)\, d\omega \right).$$

Using the KKT conditions and the alternative expression of the objective function, we can reformulate the M-MDM problem in flow variables.

Proposition 2. *Given a link cost vector* $\boldsymbol{t}$*, the link choice probabilities of the users moving towards destination* d *in M-MDM is calculated by:*

$$p_{ij}^d = \frac{x_{ij}^d}{n_i^d}, \qquad \forall (i,j) \in A_d,$$

where the unique values of x_{ij}^d *and* n_i^d *are obtained from the optimal solution* *to the following convex optimization problem:*

$$Z_d^*(\boldsymbol{t}) = \min_{\boldsymbol{x^d},\boldsymbol{n^d}} \sum_{(i,j)\in A_d} \left(t_{ij} \cdot x_{ij}^d - n_i^d \int_{1-\frac{x_{ij}^d}{n_i^d}}^{1} F_{ij}^{d\,(-1)}(\omega) d\omega \right) \tag{25}$$

$$s.t. \sum_{j\in N_d^+(i)} x_{ij}^d = n_i^d, \qquad \forall i \in N_d, \tag{26}$$

$$\sum_{k\in N_d^-(i)} x_{ki}^d + h_i^d = n_i^d, \qquad \forall i \in N_d, \tag{27}$$

$$x_{ij}^d \geq 0, \qquad \forall (i,j) \in A_d. \tag{28}$$

3.1 Special Case: Multinomial Logit

In this section, we show that for special instances of the marginal distributions, M-MDM reduces to the multinomial logit model and some of its extensions. We do this by showing the equivalence of a special instance of Proposition 2 with an entropy maximization problem that was first proposed in Akamatsu (1996) for the Markovian logit model with independent and identical Gumbel random variables:

$$Z_d\left(\boldsymbol{\theta}_{\mathbf{d}}^{(G)}, \mathbf{t}\right) = \min_{\mathbf{x}} \sum_{(i,j)\in A_d} x_{ij}^d \cdot t_{ij} + \frac{1}{\beta} \sum_{(i,j)\in A_d} x_{ij}^d \cdot \ln x_{ij}^d - \frac{1}{\beta} \sum_{i\in N_d} \left(\sum_{j\in N_d^+(i)} x_{ij}^d \right) \cdot \ln \left(\sum_{j\in N_d^+(i)} x_{ij}^d \right) \tag{29}$$

$$\text{s.t.} \sum_{j\in N_d^+(i)} x_{ij}^d - \sum_{k\in N_d^-(i)} x_{ki}^d = h_i^d, \qquad \forall i \in N_d, \tag{30}$$

$$\sum_{i\in N_d^-(d)} x_{id}^d = \sum_{i\in N_d} h_i^d, \tag{31}$$

$$x_{ij}^d \geq 0, \qquad \forall (i,j) \in A_d. \tag{32}$$

We now provide an alternative derivation of this model from our result.

Proposition 3. *Let* $\boldsymbol{\theta}_{\boldsymbol{d}}^{(G)}$ *be the joint distribution of independent and identical Gumbel random error terms with dispersion parameter* β*; and* $\boldsymbol{\Theta}_{\boldsymbol{d}}^{(E)}$ *be* *the set of all joint distributions of the random error terms with exponential marginals having location parameter* $-1/\beta$ *and scale parameter* $1/\beta$*, where:*

$$F_{ij}^d(\omega) = 1 - \exp(-1 - \beta\omega), \qquad \forall \omega \geq -\frac{1}{\beta}, (i,j) \in A_d.$$

Then, the M-MDM formulation given in (25)-(28) *reduces to the Markovian logit formulation given in* (29)-(32) *with*

$$Z_d^{*(E)}(\boldsymbol{t}) = Z_d\left(\boldsymbol{\theta}_{\boldsymbol{d}}^{(G)}, \boldsymbol{t}\right).$$

Similarly, it is possible to show the equivalence of the model in the expected cost variables with the formulation for the Markovian logit model studied in Fosgerau et al. (2013) and given by the equations (4)-(5). Let λ_i^d be the optimal dual variable. Substituting the cumulative distribution function, we obtain the following choice probability expression for exponential marginals:

$$p_{ij}^d = \exp\left(-1 - \beta\lambda_i^d\right) \cdot \exp\left(-\beta\left(t_{ij} + w_j^d\right)\right), \qquad \forall (i,j) \in A_d.$$

The normalization condition (15) provides a closed form expression for λ_i^d:

$$\sum_{j \in N_d^+(i)} p_{ij}^d = 1 \Longrightarrow \quad \lambda_i^d = \frac{1}{\beta} \ln\left(\sum_{j \in N_d^+(i)} \exp\left(-\beta\left(t_{ij} + w_j^d\right)\right)\right) - \frac{1}{\beta}.$$

Substituting λ_i^d in the choice probability expression, we obtain:

$$p_{ij}^d = \frac{\exp\left(-\beta\left(t_{ij} + w_j^d\right)\right)}{\sum_{k \in N_d^+(i)} \exp\left(-\beta\left(t_{ik} + w_k^d\right)\right)}, \qquad \forall (i,j) \in A_d.$$

Finally, substituting the cumulative distribution function in the recursive equation, we obtain the following:

$$\begin{aligned} w_i^d &= -\lambda_i^d - \frac{\exp\left(-1 - \beta\lambda_i^d\right)}{\beta} \sum_{j \in N_d^+(i)} \exp\left(-\beta\left(t_{ij} + w_j^d\right)\right) \\ &= -\frac{1}{\beta} \ln\left(\sum_{j \in N_d^+(i)} \exp\left(-\beta\left(t_{ij} + w_j^d\right)\right)\right), \end{aligned}$$

which is exactly the expected minimum cost expression of the Markovian logit model given in (4).

Corollary 1. *When the marginal distribution of the random error term* $\tilde{\epsilon}_{ij}^d$ *is exponential with location parameter* $-1/\beta_i$ *and scale parameter* $1/\beta_i$*, the M-MDM system of equations given in* (17)-(18) *reduces to the system of equations in the Nested Recursive Logit model proposed by Mai et al. (2015).*

The proof follows directly from Proposition 3 by using node specific dispersion parameters β_i instead of identical β. The Nested Recursive Logit model aims to model the correlation among the links by allowing heterogeneity in the node level. M-MDM generalizes this model by relaxing the identically distributed assumption and allows heterogeneity in the link level.

4 Markovian Traffic Equilibrium with MDM

In congested traffic networks, users' choices affect the link costs and hence, link choices towards different destinations are interdependent. The equilibrium flow is then obtained as the solution to a fixed point problem defined as follows.

Definition 2. *Let* $p_{ij}^d(\boldsymbol{t})$ *be the link choice probabilities obtained by M-MDM for given link cost vector* $\boldsymbol{t}$*. The Markovian traffic equilibrium under this model, termed as MTE-MDM, is characterized as the solution to the following fixed point problem in the variables* $(f_{ij}, t_{ij}, x_{ij}^d, n_i^d, p_{ij}^d)$*:*

$$n_i^d = h_i^d + \sum_{k \in N_d^-(i)} x_{ki}^d, \qquad \forall i \in N_d, d \in D, \tag{33}$$

$$x_{ij}^d = n_i^d \cdot p_{ij}^d(\boldsymbol{t}), \qquad \forall (i,j) \in A_d, d \in D, \tag{34}$$

$$f_{ij} = \sum_{d \in D:(i,j) \in A_d} x_{ij}^d, \qquad \forall (i,j) \in A, \tag{35}$$

$$t_{ij} = \tau_{ij}\left(f_{ij}\right), \qquad \forall (i,j) \in A. \tag{36}$$

Conditions (33)-(34) are equivalent to the flow conservation constraints (26)-(27) in M-MDM. Therefore, the M-MDM formulation given in (25)-(28) can be naturally extended to handle congestion by using the link cost functions. We make the following additional assumption in this section for the congested network case.

A4 The link cost function $\tau_{ij}(.)$ is strictly increasing with $\tau_{ij}\left(f\right) > t_{ij}^0 \geq 0$, for any positive flow $f > 0$; where t_{ij}^0 is the free flow travel cost of link (i,j).

One such function that is commonly used is the Bureau of Public Roads (BPR) cost function which accounts for congestion as follows:

$$\tau_{ij}(f_{ij}) = t_{ij}^0 \left[1 + \alpha_{ij}\left(\frac{f_{ij}}{q_{ij}}\right)^{\gamma_{ij}}\right], \quad \forall (i,j) \in A,$$

where q_{ij} is the capacity, and α_{ij} and γ_{ij} are constants. Therefore, f_{ij}/q_{ij} serves as a measure of congestion.

Proposition 4. *Under assumptions A1-A4, the equilibrium link flows for the MTE-MDM is obtained as the unique optimal solution to the following convex optimization problem:*

$$Z^* = \min_{\boldsymbol{x},\boldsymbol{n},\boldsymbol{f}} \sum_{(i,j)\in A} \int_0^{f_{ij}} \tau_{ij}(\omega)d\omega - \sum_{d\in D}\sum_{(i,j)\in A_d} n_i^d \int_{1-\frac{x_{ij}^d}{n_i^d}}^{1} F_{ij}^{d\,(-1)}(\omega)d\omega \tag{37}$$

$$s.t. \sum_{j\in N_d^+(i)} x_{ij}^d = n_i^d, \qquad \forall i \in N_d, d \in D, \tag{38}$$

$$\sum_{k\in N_d^-(i)} x_{ki}^d + h_i^d = n_i^d, \quad \forall i \in N_d, d \in D, \tag{39}$$

$$\sum_{d\in D:(i,j)\in A_d} x_{ij}^d = f_{ij}, \quad \forall (i,j) \in A, \tag{40}$$

$$x_{ij}^d \geq 0, \qquad \forall (i,j) \in A_d, d \in D. \tag{41}$$

Furthermore, the link choice probabilities at equilibrium are:

$$p_{ij}^d = \frac{x_{ij}^d}{n_i^d}, \qquad \forall (i,j) \in A_d, d \in D.$$

The MTE-MDM flows can alternatively be obtained by solving a convex optimization problem using the cost variables in the dual formulation:

$$Z^* = \max_{\mathbf{w},\boldsymbol{\lambda},\mathbf{t}} \sum_{d\in D}\sum_{i\in N_d} h_i^d \cdot w_i^d - \sum_{(i,j)\in A} \int_0^{t_{ij}} \tau_{ij}^{-1}(\omega)\, d\omega \tag{42}$$

$$\text{s.t. } w_i^d \leq -\lambda_i^d - \sum_{j\in N_d^+(i)} \int_{\lambda_i^d + t_{ij} + w_j^d}^{\infty} \left[1 - F_{ij}^d(\omega)\right] d\omega, \quad \forall i \in N_d, d \in D, \tag{43}$$

$$w_d^d = 0, \qquad \forall d \in D. \tag{44}$$

The dual formulation allows to compare MTE-MDM formulation and MTE (Baillon and Cominetti, 2008) formulation given in (7)-(9). The MTE formulation is convex under mild assumptions and can be applied to general error distributions. However, solving it is hard due to the computation of multidimensional integrals in the expectation (constraint (8)). This has limited the application of the model to the multinomial logit with i.i.d Gumbel error terms, for which this expectation has a closed form expression. On the other hand, the constraint (43) involves only one dimensional integrals, which is significantly less complex. This leads to tractable solution algorithms as we discuss next.

5 Solution Algorithm

In this section, we show that a fixed point iteration can be used to find the choice probabilities for a given link cost vector. These iterations include a step to solve the MDM problem using a line search method over the dual variables. The equilibrium flows can be obtained by integrating the proposed stochastic network loading method within the method of successive averages (MSA) algorithm (Powell and Sheffi, 1982). However, MSA might suffer from sublinear convergence rate, and hence, we also present a gradient descent algorithm with an efficient line search.

Consider the Markovian choice model for a given link cost vector. We define the function $\varphi_i^d(\mathbf{w})$ as follows:

$$\varphi_i^d\left(\mathbf{w^d}\right) = \min_{\theta_{id}\in\Theta_{id}} E_{\theta_{id}}\left[\min_{j\in N_d^+(i)}\left\{t_{ij} - \tilde{\epsilon}_{ij}^d + w_j^d\right\}\right], \qquad \forall i \in N_d,$$

where w_d^d is fixed to zero. Since the solution of the M-MDM system is unique, there exists a fixed point $\boldsymbol{\varphi}^d\left(\mathbf{w^{d*}}\right) = \mathbf{w^{d*}}$. Next, consider the sequence $\mathbf{w^{d(n)}}$ defined as follows:

$$\begin{aligned}
w_d^{d(n)} &= 0, && \forall n \in \{0,1,\dots\}, \\
w_i^{d(0)} &= s_i^d, && \forall i \in N_d, \\
w_i^{d(n)} &= \varphi_i\left(\mathbf{w^{d(n\text{-}1)}}\right), && \forall i \in N_d, n \in \{1,2,\dots\},
\end{aligned}$$

where s_i^d is the shortest path distance from node i to destination d with respect to arc costs $t_{ij} - E[\tilde{\epsilon}_{ij}^d]$. The following lemma shows that the sequence monotonically converges to the fixed point.

Lemma 2. *Define the expected cost at* $\boldsymbol{t}$*,* $w_i^d(\boldsymbol{t})$*, as:*

$$w_i^d(\boldsymbol{t}) = -\lambda_i^d(\boldsymbol{t}) - \sum_{j \in N_d^+(i)} \int_{\lambda_i^d(\boldsymbol{t})+t_{ij}+w_j^d(\boldsymbol{t})}^{\infty} \left[1 - F_{ij}^d(\omega)\right] d\omega, \quad \forall i \in N_d, d \in D,$$

and $w_d^d(\boldsymbol{t}) = 0, \forall d \in D$*. Consider the unique fixed point* $\boldsymbol{w^{d*}}$ *and the sequence* $\boldsymbol{w^{d(n)}}$ *defined above. The following inequalities hold:*

$$\begin{aligned} w_i^{d*} &\leq s_i^d, && \forall i \in N_d, \\ w_i^{d(n)} &\leq w_i^{d(n-1)}, && \forall i \in N_d, n \in \{1, 2, \dots\}. \end{aligned}$$

Then, the sequence $\boldsymbol{w^{d(n)}}$ *converges to* $\boldsymbol{w^{d*}}$*.*

Using Lemma 2, we propose a stochastic network loading algorithm for M-MDM (Algorithm 1). The algorithm has three main steps. First, shortest path distances are calculated for the given cost vector. Second, fixed point iterations are carried out in the inner loop in lines 6-13. This step involves solution of the MDM problems, which can be obtained by line search algorithms. An efficient method is the Ridder's method (Ridders, 1979) that does not require the computation of the derivative. For the distributions where the cumulative distribution function values can be efficiently calculated, the method converges quadratically (Kiusalaas, 2013); while simulation techniques might be required for open form distributions. Third, linear systems of size $|N_d|$ need to be solved (line 16). For termination, we use the following stopping condition with sufficiently small ϵ_1:

$$\max_{i \in N_d} \left\{ \frac{w_i^{d(n-1)} - w_i^{d(n)}}{w_i^{d(n-1)}} \right\} \leq \epsilon_1.$$

In the following lemma, we show that computational cost of calculating the objective function value and the gradient of the MDM-MTE formulation (42)-(44) is a single stochastic network loading with Algorithm 1.

Lemma 3. *Consider the following unconstrained reformulation of* (42)-(44)*:*

$$Z^* = \max_{\boldsymbol{t}} Z(\boldsymbol{t}) \triangleq \sum_{d \in D} \sum_{i \in N_d} h_i^d \cdot w_i^d(\boldsymbol{t}) - \sum_{(i,j) \in A} \int_0^{t_{ij}} \tau_{ij}^{-1}(\omega)\, d\omega. \tag{45}$$

Let $\boldsymbol{f}(\boldsymbol{t})$ *be the flow vector obtained by stochastic network loading at* $\boldsymbol{t}$*; and* $\boldsymbol{\tau}^{-1}(\boldsymbol{t}) = (\tau_{ij}^{-1}(t_{ij}))_{(i,j) \in A}$*. The gradient of the problem* (45) *at* $\boldsymbol{t}$ *is given by:*

$$\nabla Z(\boldsymbol{t}) = \boldsymbol{\tau}^{-1}(\boldsymbol{t}) - \boldsymbol{f}(\boldsymbol{t}). \tag{46}$$

Algorithm 1 M-MDM Loading Algorithm

Input: $\mathbf{t}$

procedure M-MDM($\mathbf{t}$)
  Find shortest path distances $s_i^d, \forall i \in N_d, d \in D$ for arc costs $\mathbf{t}$.
  for each $d \in D$ **do**
    $w_i^{d(0)} = s_i^d, \forall i \in N_d$
    $n = 0$
    while stopping condition is not satisfied **do**
      $n = n + 1.$
      $c_{ij}^{d(n)} = t_{ij} + w_j^{d(n-1)}$
      for each $i \in N_d$ **do**
        find $\lambda_i^{d(n)}$ s.t. $\sum_{j \in N_d^+(i)} \left[1 - F_{ij}^d \left(\lambda_i^{d(n)} + c_{ij}^{d(n)}\right)\right] = 1$
        $w_i^{d(n)} = -\lambda_i^{d(n)} - \sum_{j \in N_d^+(i)} \int_{\lambda_i^{d(n)} + c_{ij}^{d(n)}}^{\infty} \left[1 - F_{ij}^d(\omega)\right] d\omega$
      end for
    end while
    $p_{ij}^d = 1 - F_{ij}^d \left(\lambda_i^{d(n)} + t_{ij} + w_j^{d(n)}\right)$
    $\mathbf{Q}_{ij}^{\mathbf{d}} = \begin{cases} p_{ij}^d, & \text{if } (i,j) \in A_d \text{ and } j \neq d, \\ 0, & \text{otherwise.} \end{cases}$
    $\mathbf{M}^{\mathbf{d}} = \left(\mathbf{I} - \mathbf{Q}^{\mathbf{d}}\right)^{-1}$
    $\mathbf{n_d} = \left(\mathbf{M}^{\mathbf{d}}\right)^{\intercal} \mathbf{h_d}$
    $x_{ij}^d = n_i^d \cdot p_{ij}^d, \qquad \forall (i,j) \in A_d$
  end for
  $f_{ij} = \sum_{d \in D:(i,j) \in A_d} x_{ij}^d, \qquad \forall (i,j) \in A$
end procedure

Output: $\mathbf{f} = (f_{ij})_{(i,j) \in A}$

In the numerical examples, we used the BPR function for which a closed form expression for the integral (45) is available. For more complex functions, the partial linearization approach proposed by Lee et al. (2009) can be adopted. Algorithm 1 can be integrated within the MSA framework; however, MSA algorithm is known to suffer from sublinear convergence. Alternatively, using Lemma 3, we consider the projected gradient algorithm given in Algorithm 2. Since the function value and gradient are available after stochastic network loading, an efficient line search can be implemented. The algorithm performs backtracking line search satisfying the sufficient increase (Armijo) condition within lines 8-11 (we used parameters $c = 0.0001$ and $\rho = 0.9$). For the initial step size of the backtracking line search ($\overline{\alpha}$), we obtained the best performance by interpolating a quadratic to current and previous function values, and the directional derivative. The projection operator π projects each t_{ij} into (t_{ij}^0, ∞) where t_{ij}^0 is the free flow travel cost. Finally, we use the following stopping condition in our experiments:

$$rmse = \sum_{(i,j)\in A} \left(f_{ij}^{(n)} - f_{ij}^{(n-1)} \right)^2 \leq \epsilon_2.$$

It is known that the convergence rate of the gradient descent method with backtracking line search is at least linear (Boyd and Vandenberghe, 2004). In our experiments, gradient descent algorithm outperformed MSA (see Section 6.2.1).

6 Computational Experiments

In this section, we investigate the user choice behavior in M-MDM, and equilibrium flows in MTE-MDM for small and large traffic networks[1]. In our experiments we use real networks of Sioux Falls, Winnipeg, and Austin (the topology, link characteristics, and OD demands of these networks are obtained from Bar-Gera's Traffic Assignment Test Problem website[2]). Characteristics of these networks are summarized in Table 1.

[1] All experiments are carried out with the open-source software, seSue, provided in http://people.sutd.edu.sg/~ugur_arikan/seSue/.

[2] See https://github.com/bstabler/TransportationNetworks

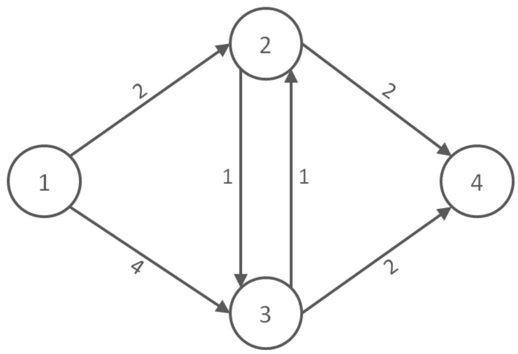
Algorithm 2 MTE-MDM Algorithm

procedure MTE-MDM
  Find a feasible flow vector $\mathbf{f}_0$; set $\mathbf{t}_0 = \boldsymbol{\tau}(\mathbf{f}_0)$ and set $k = 0$.
  Perform M-MDM($\mathbf{t}_k$), get function value Z_k and gradient ∇Z_k at $\mathbf{t}_k$.
  Set $\mathbf{d}_k = \nabla Z_k$.
  while stopping condition is not satisfied **do**
    Set $\alpha = \overline{\alpha}$ and $\mathbf{t}_{k+1} = \pi(\mathbf{t}_k + \alpha_k \mathbf{d}_k)$.
    Perform M-MDM($\mathbf{t}_{k+1}$), get function value Z_{k+1} at $\mathbf{t}_{k+1}$.
    while $Z_{k+1} < Z_k + c\alpha \nabla Z_k^\intercal \mathbf{d}_k$ **do**
      Set $\alpha = \rho\alpha$ and $\mathbf{t}_{k+1} = \pi(\mathbf{t}_k + \alpha \mathbf{d}_k)$.
      Perform M-MDM($\mathbf{t}_{k+1}$), get function value Z_{k+1} at $\mathbf{t}_{k+1}$.
    end while
    Obtain the gradient ∇Z_{k+1} from M-MDM($\mathbf{t}_{k+1}$) result.
    Set $\mathbf{d}_k = \nabla Z_{k+1}$.
    Set $k = k + 1$.
  end while
  Return the flow vector $\mathbf{f}$ obtained by M-MDM($\mathbf{t}_k$) result.
end procedure

Table 1: Characteristics of the networks.

	Sioux Falls	Winnipeg	Austin
Number of nodes	24	1,040	7,388
Number of links	76	2,836	18,961
Number of OD pairs	528	4,344	1,080,603
Number of destinations	24	138	1,117

Figure 3: Four-node network with single OD pair (1,4) where n corresponds to the number of times the middle set of links are used.



Routes (r)	Label	Cost
1-2-n(3-2)-4	a(n)	4 + 2n
1-2-3-n(2-3)-4	b(n)	5 + 2n
1-3-n(2-3)-4	c(n)	6 + 2n
1-3-2-n(3,2)-4	d(n)	7 + 2n

6.1 Choice Behavior under M-MDM

6.1.1 Relaxing the Independence Assumption

Consider the four-node network in Figure 3 taken from Akamatsu (1996) with a single destination node 4, where the numbers next to the arcs represent the deterministic link costs. Note that there are infinitely many routes in this network due to the presence of the cycle between nodes 2 and 3. In a route based model considering all routes, we have:

$$\sum_{n=0}^{\infty} \left(p_{a(n)} + p_{b(n)} + p_{c(n)} + p_{d(n)}\right) = 1,$$

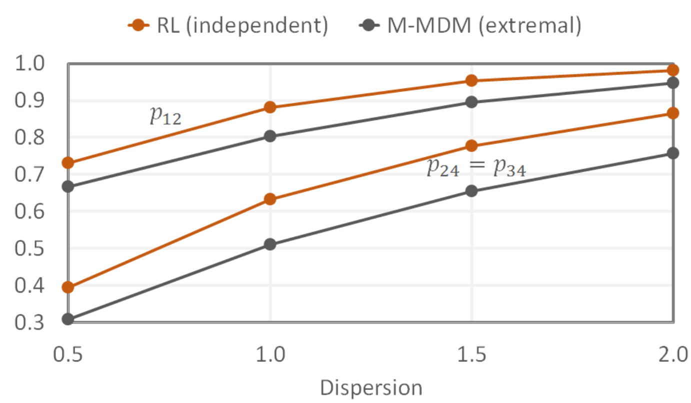
where p_r is the choice probability of route r.

The Recursive Logit (RL) model assumes a joint distribution $\boldsymbol{\theta}^G$ with i.i.d Gumbel random variables. M-MDM does not assume independence; instead it calculates the choice probabilities with respect to an extremal joint distribution minimizing the expected cost over the set of all joint distributions satisfying the given marginals. Let $\boldsymbol{\Theta}^G$ be the set of all joint distributions with Gumbel marginals, i.e.,

$$F_{ij}\left(\omega\right) = \exp\left[-\exp\left(-\gamma - \beta\omega\right)\right],$$

where γ is the Euler's constant. The behavioral difference between the RL and M-MDM with Gumbel marginals can be interpreted analytically. Since $\boldsymbol{\theta}^G \in \boldsymbol{\Theta}^G$, the optimal expected cost of the M-MDM is smaller than that of the RL. The expected costs are bounded above by the deterministic

Figure 4: Effect of independence assumption on link choice probabilities.



shortest path costs, s_i^d. Therefore, RL assignment is closer to deterministic assignment than the M-MDM. We calculate the choice probabilities of the network in Figure 3 for different values of the dispersion parameter, $\beta \in \{0.5, 1.0, 1.5, 2.0\}$. Figure 4 represents the choice probabilities of the links $(1, 2), (2, 4)$, and $(3, 4)$, which are on the shortest paths from nodes 1, 2 and 3 to the destination, respectively. As discussed above, the RL model assigns a greater portion of the flow to the deterministic shortest routes.

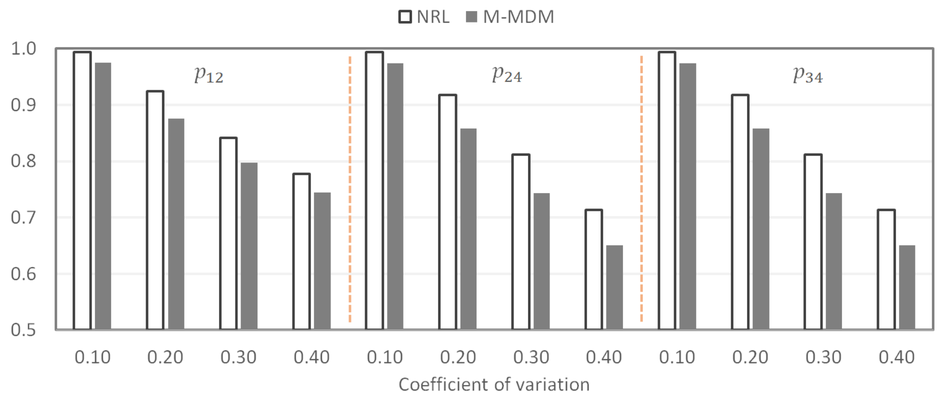
6.1.2 Incorporating Scale Heterogeneity and Overlapping Routes

We first experiment with the model M-MEM defined by exponential marginals where the error variances are scaled with respect to the free flow link costs as follows:

$$F_{ij}(\omega) = 1 - \exp\left(-1 - \frac{\omega}{\nu_i \cdot t_{ij}^0}\right), \qquad \forall \omega \geq \nu_i \cdot t_{ij}^0, (i, j) \in A. \tag{47}$$

The random error term $\widetilde{\epsilon}_{ij}$ has mean zero and standard deviation $\nu_i \cdot t_{ij}^0$. This model is suitable under the assumption that longer links are subject to greater error variance. We use $\nu_i = \nu \in \{0.1, 0.2, 0.3, 0.4\}$, and compare the link choice probabilities in M-MEM with the Nested Recursive Logit model (NRL) where the variances are scaled in the same manner, except that error variance is equal for the links emanating from the same node. Figure 5 illustrates the choice probabilities p_{12}, p_{24} and p_{34} for different values of ν. For small values of ν, the majority of the users choose the shortest route; while the choice probabilities of the links on the shortest routes decrease

Figure 5: Effect of link specific standard deviation on link choice probabilities.



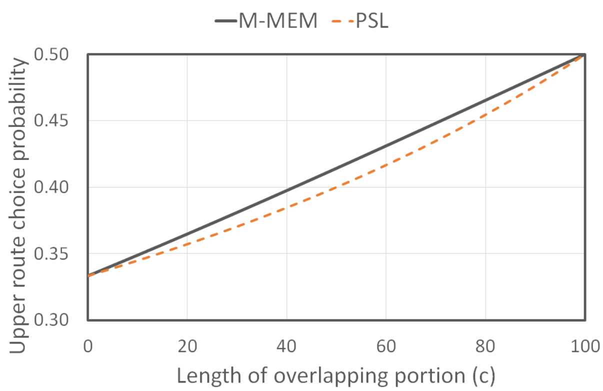
as the coefficient of variation increases. Furthermore, we observe that the probabilities of these links are less in the M-MEM when compared with the logit model. The reason for this is that the logit model scales all error variances at node i with respect to the link with smallest cost due to the identically distributed assumption. This underestimates the variance in the longer links, and hence favors the shortest link. On the other hand, M-MEM assigns the variance of each link separately and provides more accurate variance scaling.

We next discuss how the model can deal with overlapping routes. Let $s_i^{d(0)}$ be the shortest path distance from node i to destination d with respect to free flow link costs. Consider the M-MEM defined by exponential marginals where the error terms are scaled with respect to shortest path distances to the destination as follows:

$$F_{ij}(\omega) = 1 - \exp\left(-1 - \frac{\omega}{\nu_i \cdot \left(t_{ij}^0 + s_j^{d(0)}\right)}\right), \quad \forall \omega \geq \nu_i \cdot t_{ij}^0, (i,j) \in A. \tag{48}$$

It is worth noting that the error distributions are now both link and destination specific. We test the model on the loop hole network presented in Figure 1 and compare the results with the path size logit (PSL) model (Ben-Akiva and Bierlaire, 1999) which handles the overlapping problem by incorporating the path size attributes. Figure 6 plots the upper route choice probabilities with respect to the length of the overlapping section. When $x = 0$ we obtain three independent routes, and both models assign a one

Figure 6: M-MEM and PSL assignment in loop hole network.



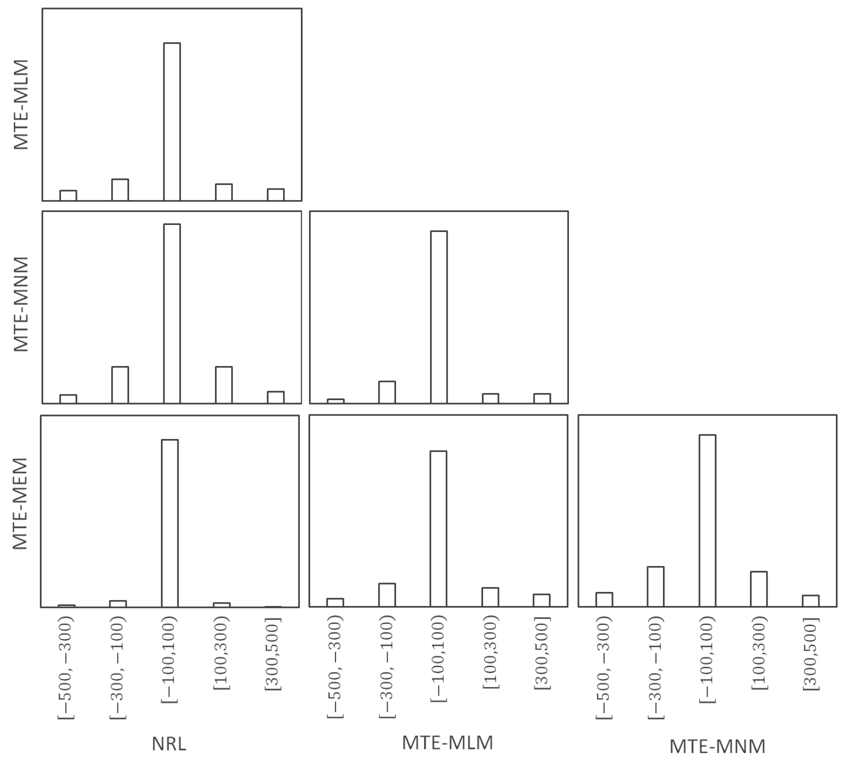
third choice probability to each; as x increases, the overlapping routes are penalized and more users prefer the upper route.

Suppose the observed fraction of users choosing the upper route is 0.40; while those for routes bc and bd are 0.25 and 0.35, respectively. Note that lower routes have equal deterministic costs; and the log likelihood estimation of a model such as NRL (see Mai et al. (2015)) would estimate the route choice probabilities as $(0.40, 0.30, 0.30)$. On the other hand, this behavior can be captured by the M-MDM. In the estimated model, error terms of the links b and c have higher variance than that of link d. This case, where more users choose the link with equal expected cost but a lower variance, may be observed when the users desire to make sure that they arrive at their destination on time.

6.1.3 Incorporating Tail Behavior

We investigate the effect of the shape of the marginal distributions on equilibrium flows on the Winnipeg network using exponential, normal and logistic marginals (MTE-MEM, MTE-MNM and MTE-MLM, respectively) with standard deviation scaled with respect to shortest distance to destination. We also calculate the flows with the NRL model scaled in the same manner, except that all arcs emanating from the same node are assigned equal variance due to the identically distributed assumption. For each pair of these models, we count the number of links with a flow difference falling in the intervals $[-500, -300), [-300, -100), [-100, 100), [100, 300), [300, 500]$. The graph illustrated in Figure 7 represents the similarity in the equilibrium link flows. We observe that the NRL and MTE-MEM result in similar

Figure 7: Pairwise comparison of equilibrium link flows.

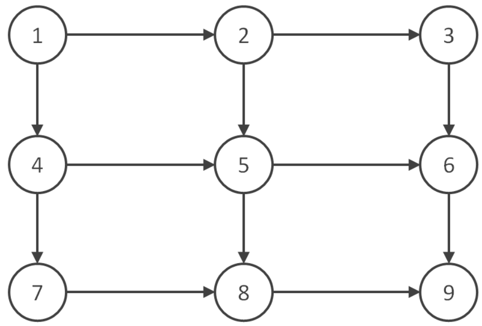


equilibrium flows, while the effect of scale heterogeneity is still observed to some extent. Note that all three MTE-MDM models have equal means and variances, and the differences observed in equilibrium flows stem from the tail behavior. This effect is most significant between exponential and normal marginals; while equilibrium flows with logistic marginals mostly lie between these two models.

6.1.4 Simulated Data Experiment on Grid Network

In order to understand the dependency structure of the extremal joint distribution minimizing the expected minimum utility (13), we use the grid network having a single OD pair $(1, 9)$ presented in Figure 8. Mean and standard deviation of perceived link costs are set to 10 and 3, respectively. We assume that the links emanating from the same node have multivariate normal distribution with correlation ρ. We simulate users link

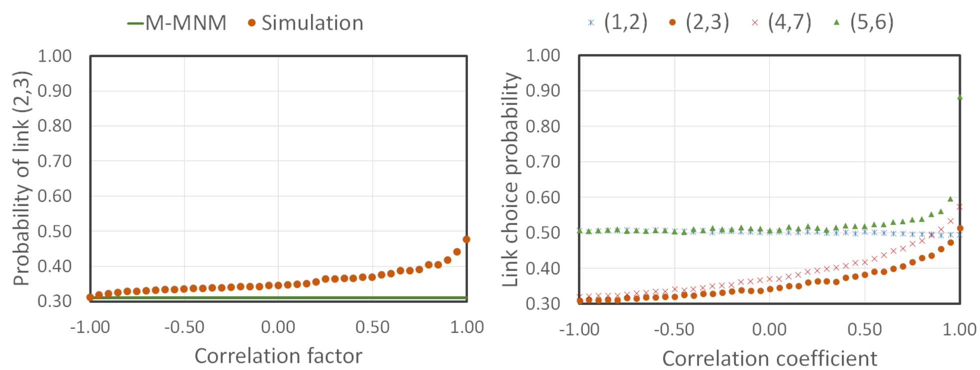
Figure 8: Grid network.



choices and obtain the split to the routes with respect to different dependency structures. Due to the symmetry of the network, we will have $p_{12} = p_{14} = p_{56} = p_{58} = 0.5$, $p_{23} = p_{47}$ and $p_{25} = p_{45}$. Since $p_{25} = 1 - p_{23}$, we can express the system behavior with single link choice probability p_{23}. Figure 9a plots p_{23} against ρ. We predict the split by using the M-MDM with normal marginals (M-MNM) having the same mean and standard deviation. As illustrated in the figure, M-MNM predicts p_{23} very close to the observed choice probabilities when the error terms are negatively correlated, while the gap increases as ρ increases. Next, we repeat the experiment with normal error terms having heterogeneous standard deviations. In this case, we can define the system behavior with p_{12}, p_{23}, p_{47} and p_{56}. Simulation results are illustrated in Figure 9b. M-MNM predictions for this case are $p_{12} = 0.51, p_{23} = 0.30, p_{47} = 0.33, p_{56} = 0.50$. We again observe that the predictions are very close to the observed split when ρ is close to -1, and fairly good for $\rho < 0$. These results suggest that the proposed model is more suitable to the cases where the error terms are believed to be negatively correlated. We can observe such situations when the individual has a strong preference for a certain alternative over the others; i.e., a favorite or familiar alternative.

6.2 Convergence Behavior of the Algorithm for MTE-MDM

In this section, we study the convergence performance of Algorithm 2 to calculate the MTE-MDM flows. We compare the results with MSA integrated with the proposed network loading method (Algorithm 1), and finally investigate the effect of parameters on convergence.

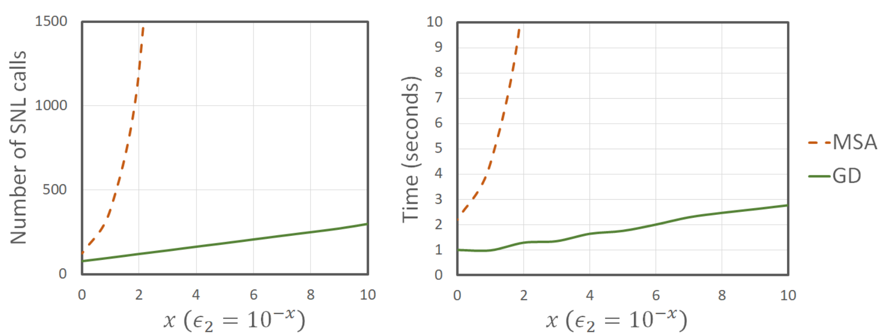


(a) Identical error terms.

(b) Heterogeneous error terms.

Figure 9: Link choice probabilities obtained by simulations.

Figure 10: Convergence behavior of Algorithm 2 on Sioux Falls network.

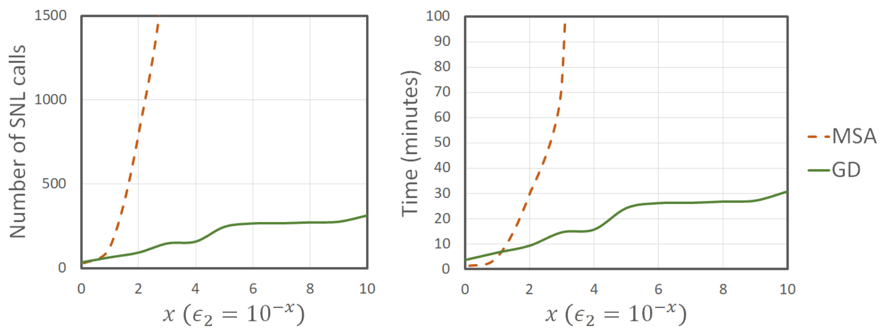


6.2.1 Effect of the Line Search Method

We start by investigating the convergence behavior of the MSA in addition to the gradient descent method (GD) integrated with a backtracking line search and the stochastic network loading method proposed in Algorithm 1. Note that computation time is directly related with the number of SNL calls for both methods.

We first use the small network of Sioux Falls with exponential marginals. Figure 10 plots the number of SNL calls and total time to find the equilibrium against the tolerance ϵ_2. We observe that MSA performs well when $\epsilon_2 \leq 0.01$; however, solution time increases drastically when higher accuracy is required; while GD converges significantly faster. We observe a similar result with the experiments on the Winnipeg network, as illustrated in Figure 11.

Figure 11: Convergence behavior of Algorithm 2 on Winnipeg network.



Next, we test the model on the Austin network with the GD algorithm. Single SNL call (Algorithm 1) on this network takes around 35 minutes. We have been able to obtain the equilibrium flows in 102 SNL calls. The errors against SNL iterations are plotted in Figure 12. The figure also plots iterations of the MSA algorithm. The missing points for GD algorithm correspond to SNL calls within the backtracking algorithm, assignments of which are not accepted due to the Armijo condition. Finally, we investigate the distribution of the solution time to the subprocedures. Recall that the SNL method in Algorithm 1 has three main components: (i) shortest path distance calculation (SPA), (ii) inner loop (IL), and (iii) solving the linear system (LS). Figure 13 presents the distribution of computational time spent for different components. As expected, the overhead cost of the algorithm, presented as other, reduces as the network size increases. We also observe that the shortest path and inner loop calculations scale well. On the other hand, solving the linear system becomes the bottleneck on large networks.

Different approaches need to be considered for a comparison with route based models. One approach is to generate predetermined choice sets for OD pairs, treating these choice sets as if they were the actual choice sets. Considering that the network has more than a million OD pairs, size of the choice sets needs to be kept sufficiently small to obtain equilibrium flows efficiently. However, in this case, the equilibrium flows significantly depend on the predetermined choice sets. Frejinger et al. (2009) also argue that parameter estimates with this approach are not consistent since the values significantly depend on the generated choice sets. On the other hand, Fosgerau et al. (2013) show that the link based RL model can be consistently estimated, which makes the Markovian model theoretically superior to route

Figure 12: Root mean square error (RMSE) against SNL calls in Austin network.

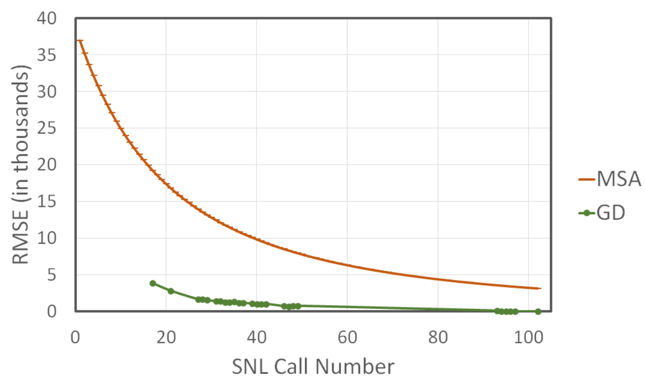
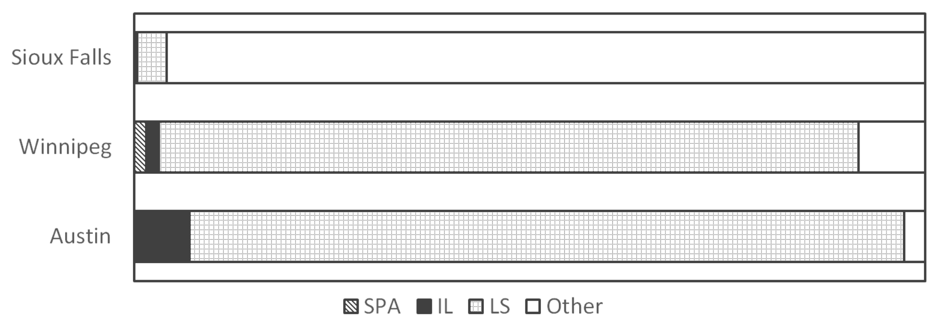


Figure 13: Distribution of solution time to subprocedures.



based models. Furthermore, we observe in Figure 13 that solution time on large networks significantly depends on the computation time of the linear system $\mathbf{M^d} = \left(\mathbf{I} - \mathbf{Q^d}\right)^{-1}$. For the numerical results in this paper, we used the solver in MATLAB. As discussed in Fosgerau et al. (2013), iterative methods can be used for very large networks (Saad and Van Der Vorst, 2001).

Finally, we investigate the effect of solving MDM problems on the computational performance. Note that the RL or NRL models use a closed form expression to calculate choice probabilities and expected costs in the inner loop of Algorithm 1. The M-MDM, on the other hand, uses root finding techniques. Therefore, the additional cost of using the M-MDM that generalizes RL and NRL is directly related with IL component of the algorithm. Figure 13 reveals that this additional computational cost is not significant with respect to the total computational time.

6.2.2 Effect of Parameters

In this section, we experiment the effect of distribution parameters on convergence performance of Algorithm 2. Markovian models might be considerably more complex than the route based SUE models, due to the recursive relation of the expected minimum costs. In particular, the recursion might diverge when the error variance is sufficiently large. We observe in our experiments on the small network that the choice probabilities of cyclic routes increase as the expected minimum costs decrease. On the other hand, execution time of these steps is much faster when the expected costs are close to the deterministic shortest path cost, or when majority of the users follow the deterministic shortest routes. Similar issues are discussed by several authors (see Akamatsu (1996); Wong (1999); Baillon and Cominetti (2008); Fosgerau and Bierlaire (2009); Si et al. (2010)). Therefore, understanding the effect of the distribution parameters on convergence rate is crucial. In our experimental setup, we first find the parameters that result in zero expected minimum cost at each node. This represents the farthest setting from the deterministic choice while we constraint the costs to be nonnegative. Then, we test settings that are closer to the deterministic choice to understand the sensitivity of the solution approach to the parameters. We use the M-MEM choice model within the equilibrium model. Let $\underline{\beta}_i$ be a lower bound on the dispersion parameter that guarantees nonnegative instantaneous expected minimum cost for the logit model:

$$\sum_{j \in N_d^+(i)} \exp\left(-\underline{\beta}_i \cdot t_{ij}^0\right) = 1, \qquad \forall i \in N_d.$$

We set the scale parameter of the exponential marginals to $B_{ij} = B_i = (\alpha_1 \cdot \underline{\beta_i})^{-1}$, and vary α_1 in the range $\{1.25, 1.50, 1.75, 2.00\}$. Note that as α_1 increases, the error variances decrease. We allow non-zero means and calculate upper bounds on the location parameters, $\overline{A}_i$, for every B_i to guarantee nonnegative expected costs as follows:

$$\overline{A}_i = \operatorname*{argmax}_{a \in R} \left\{ \sum_{j \in N_d^+(i)} \exp\left(\frac{a - t_{ij}^0}{B_i}\right) \leq \exp(-1) \right\}, \qquad \forall i \in N_d.$$

We set the location parameters of the exponential marginals to $A_{ij} = A_i = -B_i + \alpha_2(\overline{A}_i + B_i)$, for $\alpha_2 \in \{-1.0, -0.5, 0.0, 0.5\}$. The model reduces to the traffic equilibrium with the Nested Recursive Logit model when $\alpha_2 = 0$. All combinations of α_1 and α_2 guarantee nonnegative costs. Finally, note that as α_2 increases, the expected costs decrease for given variance. Due to the discussion above, we expect the solution time to increase as α_1 decreases and α_2 increases. We use the GD algorithm for NRL model utilizing closed form choice probability and expected minimum cost expressions to compare the convergence performance of the MTE-MDM.

We carry out the experiment on the networks of Sioux Falls and Winnipeg. The results are summarized in Table 2. The columns with header *Iter* represent the number of SNL calls, (s) and (m) represent seconds and minutes, and the rightmost column represents the NRL solution time for the Winnipeg network. The effect of the distribution parameters on the convergence rate is not significant for the small network. On the other hand, we observe an increase in the number of SNL calls and solution time as α_2 increases for the Winnipeg network. Interestingly, the effect of variance on the solution performance is less significant. This is due to the parameter settings that guarantee nonnegative costs which protects the M-MDM algorithm against divergence. Finally, the average solution time of the MTE-MDM for Winnipeg network is 9.42 minutes, while that of the NRL model is 7.31 minutes. Therefore, we conclude that the proposed equilibrium model provides modeling flexibility with a reasonable additional computational burden.

Table 2: Convergence behavior of Algorithm 2.

		Sioux Falls		Winnipeg		Winnipeg (NRL)
α_1	α_2	Iter.	Time(s)	Iter.	Time(m)	Time(m)
2.00	-1.0	72	1.44	59	6.66	
	-0.5	150	3.01	78	6.63	
	0.0	122	2.45	95	9.50	6.50
	0.5	101	2.02	116	11.67	
1.75	-1.0	128	2.57	77	7.56	
	-0.5	136	2.73	88	7.78	
	0.0	103	2.06	98	10.11	6.78
	0.5	93	1.86	110	11.36	
1.50	-1.0	161	3.23	85	9.11	
	-0.5	101	2.02	88	9.41	
	0.0	91	1.82	104	10.60	7.78
	0.5	82	1.64	113	10.03	
1.25	-1.0	79	1.58	93	8.37	
	-0.5	87	1.74	98	10.23	
	0.0	77	1.54	103	9.95	8.18
	0.5	75	1.50	113	11.71	

7 Conclusion

Baillon and Cominetti (2008) proposed the Markovian Traffic Equilibrium model assuming complete knowledge on the distribution of the random utilities. Despite the convexity of the MTE formulation, the application of the model has been primarily restricted to the logit model due to the challenge in computing choice probabilities and expected minimum costs for general choice models. In this paper, we have developed a distributionally robust Markovian choice model (M-MDM) and Markovian traffic equilibrium model (MTE-MDM) under the assumption that the joint distribution of the error terms is not known. We develop an equivalent convex optimization reformulation and an efficient solution algorithm to obtain the M-MDM link choice probabilities and equilibrium flows. We show that the Markovian logit model can be recreated by the M-MDM for a particular choice of the marginals. Moreover, it provides further flexibility by relaxing the i.i.d assumption and incorporating different tail behavior. We present the results of a comprehensive experimentation to illustrate the modeling flexibility, as well as the convergence performance of the MTE-MDM model. The results indicate that the model provides computational tractability for calculating the distributionally robust traffic equilibrium in large traffic networks. We believe that the flexible nature of the model and the practicality of the solution approach makes MTE-MDM interesting for future research.

Acknowledgements
The authors would like to thanks Professor Martin Savelsbergh (editor-in-chief), the Associate Editor and two referees for providing useful comments and suggestions that helped improve the paper.

A Proofs

A.1 Uniqueness of M-MDM Solution

To see the uniqueness of the solution to the M-MDM system (17)-(18), note that the right-hand-side of the M-MDM recursion (17), and hence, w_i^d, is componentwise nondecreasing in $w_j^d, \forall j \in N_d^+(i)$. Under Assumption A3, it is componentwise increasing. Suppose that $\mathbf{w^{d(1)}}$ and $\mathbf{w^{d(2)}}$ are two solutions. Let $\delta_i^d = w_i^{d(1)} - w_i^{d(2)}$, and let $\delta = \max_{i \in N_d}\{\delta_i^d\} \geq 0$. Let $i \in N_d$ be a node where the maximum difference is attained, i.e., $w_i^{d(1)} - w_i^{d(2)} = \delta$. By definition of δ, we have $\delta_j^d \leq \delta, \forall j \in N_d^+(i)$. Strict monotonicity of

w_i^d in w_j^d implies that $\delta_j^d = \delta, \forall j \in N_d^+(i)$. Moving forward in the same manner, we can show that $\delta_j^d = \delta$ for all j reachable from node i including the destination. Then, the boundary condition implies that $\delta = 0$. Thus we have $\delta_i^d = 0, \forall i \in N_d$, implying that there is at most one solution.

A.2 Proof of Proposition 1

Solution. The Lagrangian function of the formulation (20)-(22) is given below:

$$L = \sum_{i \in N_d} h_i^d \cdot w_i^d - \sum_{i \in N_d} w_i^d \cdot n_i^d - \sum_{i \in N_d} \lambda_i^d \cdot n_i^d - \sum_{(i,j) \in A_d} n_i^d \int_{\lambda_i^d + t_{ij} + w_j^d}^{\infty} \left[1 - F_{ij}^d(\omega)\right] d\omega,$$

where n_i^d is the dual variable corresponding to constraint (21). The KKT optimality conditions are as follows:

$$\frac{\partial L}{\partial \lambda_i^d} = 0 \Longrightarrow n_i^d = n_i^d \sum_{j \in N_d^+(i)} \left[1 - F_{ij}^d\left(\lambda_i^d + t_{ij} + w_j^d\right)\right], \qquad \forall i \in N_d,$$

$$\frac{\partial L}{\partial w_i^d} = 0 \Longrightarrow n_i^d = h_i^d + \sum_{k \in N_d^-(i)} n_k^d \left[1 - F_{ki}^d\left(\lambda_k^d + t_{ki} + w_i^d\right)\right], \quad \forall i \in N_d.$$

For any origin node o, i.e., $(o, d) \in W$, we have $h_o^d > 0$ and, therefore, the second KKT condition implies that $n_o^d > 0$. Then, the summation in the right-hand-side of the first KKT condition (for node o) must be equal to one. In other words, λ_o^d solves the MDM optimality condition (15). Since MDM choice probabilities are strictly positive, there will be a positive flow into each node that is adjacent to o. The same reasoning can be used to show that $n_i^d > 0, \forall i \in N_d$ and to conclude that all λ_i^d's solve the MDM optimality conditions (15) for all $\forall i \in N_d$. The right-hand-side of constraint (21) is componentwise increasing in w_j^d. Monotonicity and the fact that we are maximizing the objective function imply that the constraint holds at equality for the optimal solution. Then, the optimal values of w_i^d and λ_i^d solve the M-MDM system of equations (17)-(19). □

A.3 Proof of Lemma 1

Solution. Consider problem (20)-(22) where $F_{ij}^d(t) = 0.5 + 0.5t(2+t^2)^{-0.5}$. In this special case, constraint (21) can be reformulated as follows:

$$\begin{aligned}
2w_i^d &\le -2\lambda_i^d + \sum_{j \in N_d^+(i)} \left(a_{ij}^d - b_{ij}^d\right), && \forall i \in N_d, \\
a_{ij}^d &= \lambda_i^d + t_{ij} + w_j^d, && \forall (i,j) \in A_d, \\
\sqrt{a_{ij}^d + 2} &\le b_{ij}^d, && \forall (i,j) \in A_d,
\end{aligned}$$

and hence, the problem can be represented as a second-order cone programming model. □

A.4 Derivation of Primal M-MDM Objective Function

Given the link choice probabilities, and using (23) and (24), we can express the optimal objective function value of the dual M-MDM formulation (20)-(22) as follows:

$$\begin{aligned}
\sum_{i \in N_d} h_i^d \cdot w_i^d &= \sum_{i \in N_d} h_i^d \sum_{j \in N_d} \mathbf{M}_{ij}^{\mathbf{d}} \cdot \underline{w}_j^d \\
&= \sum_{j \in N_d} \underline{w}_j^d \sum_{i \in N_d} h_i^d \cdot \mathbf{M}_{ij}^{\mathbf{d}} \\
&= \sum_{j \in N_d} n_j^d \cdot \underline{w}_j^d \\
&= \sum_{i \in N_d} n_i^d \sum_{j \in N_d^+(i)} \left(t_{ij} \cdot p_{ij}^d - \int_{1-p_{ij}^d}^{1} F_{ij}^{d\,(-1)}(\omega)\, d\omega \right) \\
&= \sum_{i \in N_d} \sum_{j \in N_d^+(i)} \left(t_{ij} \cdot x_{ij}^d - n_i^d \int_{1-\frac{x_{ij}^d}{n_i^d}}^{1} F_{ij}^{d\,(-1)}(\omega)\, d\omega \right).
\end{aligned}$$

A.5 Proof of Proposition 2

Solution. Let $\varphi(\mathbf{x^d}, \mathbf{n^d}) = \sum_{(i,j) \in A_d} \varphi_{ij}(x_{ij}^d, n_i^d)$ represent the objective function where:

$$\varphi_{ij}(x_{ij}^d, n_i^d) = \left(t_{ij} \cdot x_{ij}^d - n_i^d \int_{1-\frac{x_{ij}^d}{n_i^d}}^{1} F_{ij}^{d\,(-1)}(\omega) d\omega \right).$$

This is the perspective of the function $\varphi_{ij}^0(x)$ defined as:

$$\varphi_{ij}^0(x) = \left(t_{ij} \cdot x - \int_{1-x}^{1} F_{ij}^{d\,(-1)}(\omega) d\omega \right),$$

such that $\varphi_{ij}(x_{ij}^d, n_i^d) = n_i^d \varphi_{ij}^0(x_{ij}^d / n_i^d)$. Since the function $\varphi_{ij}^0(x)$ is strictly convex under the assumptions on the marginal distribution, the perspective of the function denoted by $\varphi_{ij}(x_{ij}^d, n_i^d)$ is also strictly convex. This implies that $\varphi(\mathbf{x^d}, \mathbf{n^d})$ is strictly convex. Furthermore the constraints are convex and hence it is a convex optimization problem with a unique optimal solution. The KKT optimality conditions of the formulation are given below:

$$\frac{x_{ij}^d}{n_i^d} = 1 - F_{ij}^d \left(\lambda_i^d + t_{ij} + w_j^d \right), \qquad \forall (i,j) \in A_d,$$

$$\begin{aligned} w_i^d &= -\lambda_i^d - \sum_{j \in N_d^+(i)} \left(\int_{1-\frac{x_{ij}^d}{n_i^d}}^{1} F_{ij}^{d\,(-1)}(\omega) d\omega - \frac{x_{ij}^d}{n_i^d} F_{ij}^{d\,(-1)} \left(1 - \frac{x_{ij}^d}{n_i^d} \right) \right) \\ &= -\lambda_i^d - \sum_{j \in N_d^+(i)} \int_{\lambda_i^d + t_{ij} + w_j^d}^{\infty} \left[1 - F_{ij}^d(\omega) \right] d\omega, \qquad \forall i \in N_d, \end{aligned}$$

where λ_i^d and w_i^d are the dual variables corresponding to constraints (26) and (27), respectively. By constraint (26), λ_i^d satisfies the normalization condition (15). Therefore, the KKT conditions are equivalent to the M-MDM system of equations. □

A.6 Proof of Proposition 3

Solution. The exponential distribution satisfies the assumption on marginals, A3. Substituting the inverse cumulative distribution function $F_{ij}^{d\,(-1)}(\omega) = -(1 + \ln(1-\omega))/\beta$ in the objective function (25), the integral is calculated as follows:

$$\int_{1-\frac{x_{ij}^d}{n_i^d}}^{1} F_{ij}^{d\,(-1)}(\omega) d\omega = \frac{1}{\beta n_i^d} \left(-x_{ij}^d \cdot \ln x_{ij}^d + x_{ij}^d \cdot \ln n_i^d \right),$$

and the objective function simplifies to the following:

$$\sum_{(i,j) \in A_d} \left(t_{ij} \cdot x_{ij}^d + \frac{x_{ij}^d \cdot \ln x_{ij}^d}{\beta} - \frac{x_{ij}^d \cdot \ln n_i^d}{\beta} \right).$$

Constraints (26) and (27) can be expressed as a single constraint by substituting n_i^d in the objective function. Using substitution $n_i^d = \sum_{j\in N_d^+(i)} x_{ij}^d$, the last term in the objective function can be expressed as follows:

$$
\begin{aligned}
-\frac{1}{\beta}\sum_{(i,j)\in A_d} x_{ij}^d \cdot \ln n_i^d &= -\frac{1}{\beta}\sum_{i\in N_d}\sum_{j\in N_d^+(i)} x_{ij}^d \cdot \ln\left(\sum_{j\in N_d^+(i)} x_{ij}^d\right) \\
&= -\frac{1}{\beta}\sum_{i\in N_d}\left(\sum_{j\in N_d^+(i)} x_{ij}^d\right)\cdot \ln\left(\sum_{j\in N_d^+(i)} x_{ij}^d\right).
\end{aligned}
$$

Then, the M-MDM formulation reduces to the Markovian logit model given in (29)-(32). □

A.7 Proof of Proposition 4

Solution. Let λ_i^d, w_i^d and t_{ij} be the dual variables corresponding to constraints (38), (39) and (40), respectively. Then, the KKT optimality conditions of the convex optimization problem are as follows:

$$
\begin{aligned}
& t_{ij} = \tau_{ij}\left(f_{ij}\right), && \forall (i,j)\in A, \\
& \frac{x_{ij}^d}{n_i^d} = 1 - F_{ij}^d\left(\lambda_i^d + t_{ij} + w_j^d\right), && \forall (i,j)\in A_d, d\in D, \\
& w_i^d = -\lambda_i^d - \sum_{j\in N_d^+(i)} \int_{\lambda_i^d + t_{ij} + w_j^d}^{\infty}\left[1 - F_{ij}^d\left(\omega\right)\right] d\omega, && \forall i\in N_d, d\in D.
\end{aligned}
$$

The last two optimality conditions solve the M-MDM system of equations for a given link cost vector. Then, conditions (35) and (36) are satisfied since $x_{ij}^d/n_i^d = p_{ij}(\mathbf{t})$, where $t_{ij} = \tau_{ij}\left(f_{ij}\right)$ by the first KKT condition. Conditions (33) and (34) are satisfied by the constraints of the model. Therefore, the optimal solution of the formulation (37)-(40) solves the MTE-MDM fixed point problem. The constraints of the model are linear, and the objective function is strictly convex and hence, the solution of the model is unique. □

A.8 Proof of Lemma 2

Solution. The M-MDM recursion (17) implies that $w_i^d \leq t_{ij} - E[\tilde{\epsilon}_{ij}^d] + w_j^d, \forall j \in N_d^+(i)$. This inequality recursively implies that s_i^d is an upper bound for w_i^{d*}. We can make the induction assumption that $w_i^{d(n)} \leq$

$w_i^{d(n-1)}, \forall n \in \{1, \ldots, m-1\}, i \in N_d$, since $w_i^{d(1)} \leq w_i^{d(0)} = s_i^d, \forall i \in N_d$. By this assumption, we have $t_{ij} + w_j^{d(m-1)} \leq t_{ij} + w_j^{d(m-2)}, \forall (i,j) \in A_d$. The monotonicity of w_i^d in $w_j^d, \forall j \in N_d^+(i)$, then implies that $w_i^{d(m)} \leq w_i^{d(m-1)}, \forall i \in N_d$. By induction, the second inequality holds for all n. Thus, the sequence is monotonic nonincreasing, moreover bounded from below provided that the M-MDM solution exists. This proves that the sequence converges to the unique fixed point $\mathbf{w^{d*}}$. □

A.9 Proof of Lemma 3

Solution. We start the proof by defining $Y_{ai}^d(\mathbf{t})$, $Q_{ji}^d(\mathbf{t})$ and $L_{ai}^d(\mathbf{t})$ as follows:

$$Y_{ai}^d(\mathbf{t}) = \frac{\partial w_i^d(\mathbf{t})}{\partial t_a}, \qquad \forall a \in A_d, d \in N_d, d \in D,$$

$$Q_{ji}^d(\mathbf{t}) = \begin{cases} p_{ij}^d(\mathbf{t}) & \text{if } (i,j) \in A_d \\ 0 & \text{othwerise} \end{cases}, \qquad \forall a \in A_d, d \in N_d, d \in D,$$

$$L_{ai}^d(\mathbf{t}) = \begin{cases} p_a^d(\mathbf{t}) & \text{if tail of } a \text{ is } i \\ 0 & \text{othwerise} \end{cases}, \qquad \forall a \in A_d, d \in N_d, d \in D,$$

where $p_{ij}^d(\mathbf{t}) = 1 - F_{ij}^d(\lambda_i^d(\mathbf{t}) + t_{ij} + w_j^d(\mathbf{t}))$. Taking the derivative of the expected cost with respect to link cost, we obtain:

$$\begin{aligned} Y_{ai}^d(\mathbf{t}) &= -\frac{\partial \lambda_i^d(\mathbf{t})}{\partial t_a} + \sum_{j \in N_d^+(i)} \left(\frac{\partial \lambda_i^d(\mathbf{t})}{\partial t_a} + \frac{\partial t_{ij}}{\partial t_a} + Y_{aj}^d(\mathbf{t}) \right) \cdot p_{ij}^d(\mathbf{t}) \\ &= \sum_{j \in N_d^+(i)} \left(\frac{\partial t_{ij}}{\partial t_a} + Y_{aj}^d(\mathbf{t}) \right) \cdot p_{ij}^d(\mathbf{t}) \\ &= L_{ai}(\mathbf{t}) + \sum_{j \in N_d^+(i)} Y_{aj}^d(\mathbf{t}) \cdot Q_{ji}^d(\mathbf{t}). \end{aligned}$$

This leads to the linear system $Y^d(\mathbf{t}) = L^d(\mathbf{t}) \cdot (M^d(\mathbf{t}))^\intercal$ where $M^d(\mathbf{t})$ is the fundamental matrix of the absorbing Markov chain. Then, the first derivative of the objective function (45) is given by the following:

$$\begin{aligned} \frac{\partial Z(\mathbf{t})}{\partial t_a} &= \tau_a^{-1}(t_a) - \sum_{d \in D} \sum_{i \in N_d} h_i^d \cdot Y_{ai}^d(\mathbf{t}), \\ &= \tau_a^{-1}(t_a) - \sum_{d \in D} x_a^d(\mathbf{t}), \\ &= \tau_a^{-1}(t_a) - f_a(\mathbf{t}), \qquad \forall a \in A, \end{aligned}$$

where $x_a^d(\mathbf{t})$ and $f_a(\mathbf{t})$ are the flows obtained by the stochastic network loading method (Algorithm 1) at link cost vector $\mathbf{t}$. This completes the proof. $\square$

References

S. D. Ahipaşaoğlu, R. Meskarian, T. L. Magnanti, and K. Natarajan. Beyond normality: A cross moment-stochastic user equilibrium model. *Transportation Research Part B: Methodological*, 81(2):333–354, 2015.

S. D. Ahipaşaoğlu, U. Arıkan, and K. Natarajan. On the flexibility of using marginal distribution choice models in traffic equilibrium. *Transportation Research Part B: Methodological*, 91:130–158, 2016.

T. Akamatsu. Cyclic flows, Markov process and stochastic traffic assignment. *Transportation Research Part B: Methodological*, 30(5):369–386, 1996.

T. Akamatsu. Decomposition of path choice entropy in general transport networks. *Transportation Science*, 31(4):349–362, 1997.

A. Atamturk and M. Zhang. Two-stage robust network flow and design under demand uncertainty. *Operations Research*, 55(4):662–673, 2007.

J.-B. Baillon and R. Cominetti. Markovian traffic equilibrium. *Mathematical Programming*, 111(1-2):33–56, 2008.

S. Bekhor, M. E. Ben-Akiva, and M. S. Ramming. Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, 144(1):235–247, 2006.

M. G. Bell. Alternatives to Dial's logit assignment algorithm. *Transportation Research Part B: Methodological*, 29(4):287–295, 1995.

M. Ben-Akiva and M. Bierlaire. Discrete choice methods and their applications to short term travel decisions. *Hall, R.W. (Ed.), Handbook of Transportation Science, Kluwer Publishers*, pages 5–33, 1999.

S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

J. G. Carlsson and M. Behroozi. Wasserstein distance and the distributionally robust TSP. *Working Paper*, 2017.

J. G. Carlsson and E. Delage. Robust partitioning for stochastic multivehicle routing. *Operations Research*, 61(3):727–744, 2013.

E. Castillo, P. Jiménez, J. M. Menéndez, and M. Nogal. A bayesian method for estimating traffic flows based on plate scanning. *Transportation*, 40 (1):173–201, 2013.

E. Castillo, A. Calviño, M. Nogal, and H. K. Lo. On the probabilistic and physical consistency of traffic random variables and models. *Computer-Aided Civil and Infrastructure Engineering*, 29(7):496–517, 2014.

C. E. Clark. The greatest of a finite set of random variables. *Operations Research*, 9(2):145–162, 1961.

C. F. Daganzo and Y. Sheffi. On stochastic models of traffic assignment. *Transportation Science*, 11(3):253–274, 1977.

R. Dial. A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research*, 5(2):83–111, 1971.

M. J. C. Erera, A. L. and M. Savelsbergh. Robust optimization for empty repositioning problems. *Operations Research*, 57(2):468–483, 2009.

M. Fosgerau and M. Bierlaire. Discrete choice models with multiplicative error terms. *Transportation Research Part B: Methodological*, 43(5):494–505, 2009.

M. Fosgerau and D. Fukuda. Valuing travel time variability: Characteristics of the travel time distribution on an urban road. *Transportation Research Part C*, 24:83–101, 2012. doi: 10.1016/j.trc.2012.02.008.

M. Fosgerau, E. Frejinger, and A. Karlstrom. A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological*, 56:70–80, 2013.

E. Frejinger, M. Bierlaire, and M. Ben-Akiva. Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological*, 43(10):984–994, 2009.

M. D. Gould. Considering individual cognitive ability in the provision of usable navigation assistance. In *Vehicle Navigation and Information Systems Conference, 1989. Conference Record*, pages 443–447. IEEE, 1989.

J. G. Kemeny, J. L. Snell, et al. *Finite Markov chains*, volume 356. van Nostrand Princeton, NJ, 1960.

J. Kiusalaas. *Numerical methods in engineering with Python 3*. Cambridge university press, 2013.

B. Kuipers. Modeling spatial knowledge. *Cognitive science*, 2(2):129–153, 1978.

D.-H. Lee, Q. Meng, and W. Deng. Origin-based partial linearization method for the stochastic user equilibrium traffic assignment problem. *Journal of Transportation Engineering*, 136(1):52–60, 2009.

M. Maher. Sam-a stochastic assignment model. In *Mathematics in Transport Planning and Control (Institute of Mathematics & Its Applications Conference Series 38)*, 1992.

M. Maher. Algorithms for logit-based stochastic user equilibrium assignment. *Transportation Research Part B: Methodological*, 32(8):539–549, 1998.

M. Maher and P. Hughes. A probit-based stochastic user equilibrium assignment model. *Transportation Research Part B: Methodological*, 31(4): 341–355, 1997.

T. Mai, M. Fosgerau, and E. Frejinger. A nested recursive logit model for route choice analysis. *Transportation Research Part B: Methodological*, 75:100–112, 2015.

D. M. Mark and M. McGranaghan. Effective provision of navigation assistance to drivers: A cognitive science approach. In *Proceedings, Auto-Carto London*, volume 2, pages 399–408, 1986.

V. K. Mishra, K. Natarajan, D. Padmanabhan, C.-P. Teo, and X. Li. On theoretical and empirical aspects of marginal distribution choice models. *Management Science*, 60(6):1511–1531, 2014.

F. Ordóñez and N. E. Stier-Moses. Wardrop equilibria with risk-averse users. *Transportation Science*, 44(1):63–86, 2010.

A. Polus. A study of travel time and reliability on arterial routes. *Transportation*, 8:141–151, 1979.

W. B. Powell and Y. Sheffi. The convergence of equilibrium algorithms with predetermined step sizes. *Transportation Science*, 16(1):45–55, 1982.

C. Ridders. A new algorithm for computing a single root of a real continuous function. *IEEE Transactions on circuits and systems*, 26(11):979–980, 1979.

Y. Saad and H. A. Van Der Vorst. Iterative solution of linear systems in the 20th century. In *Numerical Analysis: Historical Developments in the 20th Century*, pages 175–207. Elsevier, 2001.

H. Scarf. *A min-max solution of an inventory problem in Studies in the Mathematical Theory of Inventory and Production.* Stanford University Press, 1958.

Y. Sheffi. Urban transportation networks: Equilibrium analysis with mathematical programming methods. *Prentice-Hall*, 1985.

B.-F. Si, M. Zhong, H.-Z. Zhang, and W.-L. Jin. An improved Dial's algorithm for logit-based traffic assignment within a directed acyclic network. *Transportation Planning and Technology*, 33(2):123–137, 2010.

A. Sklar. Fonctions de répartition ă n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.

E. Stern and D. Leiser. Levels of spatial knowledge and urban travel modeling. *Geographical Analysis*, 20(2):140–155, 1988.

J. Wardrop. Some theoretical aspects of road traffic research. *Proceedings, Institution of Civil Engineers*, II(1):325–378, 1952.

D. Watling. User equilibrium traffic network assignment with stochastic travel times and late arrival penalty. *European Journal of Operational Research*, 175:1539–1556, 2006.

S. Wong. On the convergence of Bell's logit assignment formulation. *Transportation Research Part B: Methodological*, 33(8):609–616, 1999.