Transportation Research Part B 179 (2024) 102853

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Statistical inference of travelers' route choice preferences with system-level data

Pablo Guarda [a], Sean Qian [a,b,*]

[a] Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
[b] Heinz College, Carnegie Mellon University, Pittsburgh, PA, USA

ARTICLE INFO

Keywords:
Network models
Stochastic user equilibrium
Route choice models
Travel behavior
Utility function
Multinomial logit model
Hypothesis testing
Pseudo-convexity
Normalized gradient descent
Traffic count data

ABSTRACT

Traditional network models encapsulate travel behavior among all origin–destination pairs based on a simplified and generic travelers' utility function. Typically, the utility function consists of travel time solely, and its coefficients are equated to estimates obtained from discrete choice models and stated preference data. While this modeling strategy is reasonable, the inherent sampling bias in individual-level experimental data may be further amplified over network flow aggregation, leading to inaccurate flow estimates. In addition, individual-level data must be collected from surveys or travel diaries, which may be labor-intensive, costly, and limited to a small time horizon. To address these limitations, this study extends classical bi-level formulations to estimate travelers' utility functions with multiple attributes using system-level data. This data tends to be less subject to sampling bias than individual-level data, it is cheaper to collect and it has become increasingly diverse and available. To leverage system-level data, we formulate a methodology grounded on non-linear least squares to statistically infer travelers' utility function in the network context using traffic counts, traffic speeds, the number of traffic incidents, and sociodemographic information obtained from the US Census, among other attributes. The analysis of the mathematical properties of the optimization problem and its pseudo-convexity motivates the use of normalized gradient descent, an algorithm developed in the machine learning community that is suitable for pseudo-convex programs. More importantly, we develop a hypothesis test framework to examine the statistical properties of coefficients attached to utility terms and to perform attribute selection. Experiments on synthetic data show that the travelers' utility function coefficients can be consistently recovered and that hypothesis tests are reliable statistics to identify which attributes are determinants of travelers' route choices. Besides, a series of Monte-Carlo experiments showed that statistical inference is robust to various levels of sensor coverage and to noises in the Origin-Destination matrix and the traffic count measurements. The methodology is also deployed at a large scale using real-world multi-source data in Fresno, CA, collected before and during the COVID-19 outbreak.

1. Introduction

System-level data, as opposed to individual-level data, measures characteristics of aggregated flow in transportation networks, and it is a valuable source of information for studying travel demand. An example is the use of traffic count data for estimating origin–destination (O-D) matrices (Fisk, 1989; Yang et al., 1992; Cascetta and Postorino, 2001; Ma and Qian, 2018b; Krishnakumari

* Corresponding author at: Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA, USA.
E-mail address: seanqian@cmu.edu (S. Qian).

https://doi.org/10.1016/j.trb.2023.102853
Received 22 April 2022; Received in revised form 24 July 2023; Accepted 30 October 2023
Available online 1 December 2023
0191-2615/

et al., 2020). Interestingly, little attention has been given to using system-level data to estimate travelers' utility functions in route choice models. These models are crucial to depicting travel demand in transport planning applications, and their parameters are usually estimated with data collected from individual-level surveys and experiments. Modeling route choices at the system level using individual-level data may be inaccurate and expensive due to the inherent sampling bias and high collection cost of experimental data. The increase in availability and diversity of system-level data offers opportunities to overcome those limitations and to understand the impact of a broader set of factors that may influence travelers' route choice decisions. Some new sources of system-level data include crashes, weather and pavement conditions, land use characteristics, socio-demographics information from Census, and travel time measurements collected from probe vehicles. These factors may determine how a traveler chooses a route, but surveying all this information is infeasible. Thus, this paper proposes a statistical method to model travelers' route choices using system-level data only.

Researchers have already leveraged system-level data within traditional network models to estimate travelers' route choice preferences (Robillard, 1974; Fisk, 1977; Daganzo, 1977; Anas and Kim, 1990; Liu and Fricker, 1996; Yang et al., 2001; Lo and Chan, 2003; García-Ródenas and Marín, 2009; Russo and Vitetta, 2011; Wang et al., 2016). However, their primary focus has been on estimating a utility function dependent on travel time only. Furthermore, most existing algorithms have been deployed to networks of relatively small size, which raises questions about their potential to scale up to large transportation networks and contribute to real-world applications. The main difficulty in solving this estimation problem is that traditional network models are not data-driven. Discrete choice models, the gold standard to analyze individual-level data in travel behavior studies, may be considered a promising alternative to solve this problem, given their data-driven nature. However, these models cannot capture the endogeneity of travel times arising from the interdependence between travelers' decisions and the effect of traffic congestion in transportation networks.

With the goal of estimating the coefficients of travelers' utility functions with multiple attributes obtained from system-level data, this paper enhances traditional network models with optimization algorithms developed in the machine learning literature. Mathematical properties of the optimization problem and its pseudo-convexity motivate using normalized gradient descent, a first-order method suitable for pseudo-convex optimization. Our solution algorithm is tailored to perform a robust estimation of the coefficients of the travelers' utility function and reproduce the traffic conditions observed at the system level. To ease the identification of the determinants of travelers' route choices with system-level data, we formulate hypothesis tests on the utility function coefficients.

The paper is structured as follows. We first survey the existing literature and identify the main contributions of our work. The following three sections describe the formulation of our methodology on a general transportation network, analyze the mathematical properties of the optimization problem and propose a solution algorithm to solve the problem. Subsequently, we describe the framework to perform statistical inference on the utility function coefficients, and we present the results of numerical experiments conducted in networks of small and medium size. Next, we show the estimation results obtained in a large-scale network using real-world system-level data. Finally, we describe our main conclusions and suggest avenues for further research. The mathematical notation used for the remainder of the paper is included in Appendix A.1.

2. Literature review and research gaps

The computation of network equilibrium requires specifying a travelers' route choice model. The choice of route choice model defines the class of network equilibria (Prashker and Bekhor, 2004). The simplest and most traditional class of equilibria is known as deterministic user equilibrium (DUE). At DUE, no traveler has incentives to change to alternative routes, and travelers are assumed to pick, in a deterministic manner, the route that maximizes their utility. The multinomial logit model (MNL) remains a gold standard in travel behavior research to depict travelers' decision-making (McFadden, 1973). A well-known application of the MNL model in network modeling is for the computation of Stochastic User Equilibrium with Logit assignment (SUELOGIT). In contrast to other probabilistic models of travelers' choices, the logit model exhibits a better compromise between behavioral realism and mathematical tractability. Similar to the MNL model, the SUELOGIT enjoys this mathematical convenience in the network modeling context. For instance, under mild conditions, SUELOGIT has solution uniqueness in the link and path flow space, whereas DUE only has solution uniqueness in the link flow space. Besides, the existence of a closed form expression for the choice probabilities in the multinomial logit model can be leveraged for the computation of SUELOGIT, e.g., the Dial logit algorithm (Dial, 1971; Bell, 1995) or solution methods for the ODE problem (Ma and Qian, 2018a).

A stream of research that brought to our attention concerns the estimation of the travelers' utility function from traffic count data, which induces a traffic assignment consistent with SUELOGIT (Robillard, 1974; Fisk, 1977; Daganzo, 1977; Anas and Kim, 1990; Liu and Fricker, 1996; Yang et al., 2001; Lo and Chan, 2003; García-Ródenas and Marín, 2009; Russo and Vitetta, 2011; Wang et al., 2016). The solution to this problem and of network equilibrium, in general, requires knowing a priori the coefficients of the travelers' utility function of the route choice model. Hence, the standard practice is to set the values of these coefficients equal to estimates obtained in previous travel behavior studies. Nevertheless, there are multiple advantages to estimating these coefficients. First, it avoids searching for external estimates, which may be cumbersome and provide inaccurate coefficients. Second, it can improve the generalization performance on the estimated O-D matrix when, for instance, the coefficients fitted from existing count data become close to the population coefficients. The following sections describe the relevant literature that has studied the abovementioned problem.

2.1. Overview of the Logit Utility Estimation (LUE) problem

Seminal work in the literature focused on the problem of estimating the coefficient $\hat{\theta}$ of a utility function dependent solely on travel time and where both the O-D matrix and travel costs among links/paths are assumed exogenous. To our knowledge, the LUE problem was first studied by Robillard (1974), who estimated the coefficient $\hat{\theta}$ using the link and path flows obtained from a traffic assignment consistent with the Dial (1971) method. The estimator $\hat{\theta}$ was assumed to be Gaussian distributed and obtained via maximum likelihood estimation (MLE). A primary limitation of this work was the requirement of knowing the traffic counts at every link/path in the network. The problem of estimating $\hat{\theta}$ from a subset of traffic count in the network was then addressed by Fisk (1977). Another major limitation of this work was assuming full knowledge of the average cost of the paths connecting each O-D pair, which may be plausible in settings where costs are equated with travel times but arguably unrealistic in real-world scenarios where travelers make route choices based on multiple attributes such as monetary cost and travel time reliability.

Daganzo (1977) focused on a more general setting where link flows observations were consistent with SUE but not necessarily with a logit assignment. In contrast to Robillard (1974), the distribution of the estimator $\hat{\theta}$ was derived by imposing distributional properties on the link flows and relying on the statistical properties of the MLE estimator. This work is the first that conducts hypothesis testing on the coefficient of a traveler's utility function dependent on a single attribute. However, similar to previous research, it assumes exogenous link costs, which limits the application to uncongested networks or settings where link costs at equilibrium are known. Under the assumption that link flow data is consistent with SUELOGIT, Anas and Kim (1990) extended prior work by allowing for endogenous travel costs. The authors studied the impact of estimating $\hat{\theta}$ without accounting for the endogeneity of the travel costs. They observed that estimates of $\hat{\theta}$ become biased and that statistical inference is less consistent in this setting.

2.2. Overview of the origin-destination and Logit Utility Estimation (ODLUE) problem

In the following decades, the research interest shifted to addressing the joint estimation of the O-D matrix and $\hat{\theta}$ from traffic count data following SUELOGIT, which we coined as O-D and Logit Utility estimation (ODLUE) problem[1]. Interestingly, the extension of the LUE problem to settings where the utility function was dependent on multiple attributes besides travel time did not gain the same attention. By this time, the ODE literature had made significant progress in estimating O-D matrices with traffic count data consistent with SUELOGIT. These methodological advances could be directly leveraged to solve the LUE problem, which include sensitivity analysis (Patriksson, 2004) or the alternating optimization of the upper and lower levels of the classic ODE bilevel formulation (Ma and Qian, 2017).

To our knowledge, Liu and Fricker (1996) conducted the first study that studied the ODLUE problem and presented an application using system-level data collected from a real-world transportation network. Traffic count and travel time data were collected from a small network at the Purdue University campus, and estimates of $\hat{\theta}$ were obtained for different periods of the day. The authors proposed a two-stage calibration method that leveraged the closed form of the logit route choice probabilities to solve for $\hat{\theta}$ using the secant method. The O-D matrix was estimated via a least square minimization. Fundamental limitations of this work were to not account for the congestion effects in the transportation network and to assume that traffic counts and travel time data were available for the full set of links. This assumption could be mild in small networks where traffic counts can be collected at every link but implausible in larger networks where this data is typically available for a small proportion of the links.

Yang et al. (2001) overcame some of the limitations mentioned above by solving a bilevel optimization problem that resembled the traditional mathematical program with equilibrium constraints (MPEC) that had been used for solving the ODE problem. A main contribution of this work was to estimate both $\hat{\theta}$ and the O-D matrix while also accounting for the impact of traffic congestion. A Sequential Quadratic Programming (SQP) algorithm was used to find the parameters of interest and to minimize the gap between observed and predicted traffic counts. The optimization problem also included constraints to ensure that the SUELOGIT equilibrium conditions were satisfied over iterations. The authors also derived for the first time the analytical expression of the gradients of the outer level objective with respect to the parameters associated with the O-D matrix and to the travelers' utility function. Lo and Chan (2003) proposed a similar framework where a MLE instead of a non-linear least squares (NLLS) problem was solved at the upper level of the bilevel formulation. The authors performed ODLUE on synthetic and real-world data collected from the Tuen Mun Corridor network in Hong Kong. Later, Wang et al. (2016) implemented a similar approach using real-world data gathered from a small network in Seattle, WA. They also performed experiments on synthetic data to study the robustness of the parameter estimates to noise in the ground truth O-D matrix and in the parameter $\hat{\theta}$ used to generate synthetic traffic count data.

2.3. Contributions of this research

This paper extends existing formulations to solve the LUE problem with the goal of statistically inferring the coefficients of travelers' utility functions with multiple attributes using system-level data. A bilevel optimization program is formulated to minimize the gap between estimated and observed traffic counts and to account for the endogenous effect of traffic congestion on travelers' route choices. In addition, a hypothesis test framework is proposed to examine the statistical properties of the estimated coefficients.

Below are summarized the contributions of this paper to existing literature:

[1] We denominated this problem as O-D Logit Utility estimation (ODLUE) because the O-D matrix is estimated in addition to the travelers' utility function coefficients as it is done in LUE.

1. It presents a methodology to statistically infer the coefficients of travelers' utility functions consisting of multiple attributes and using system-level data, which enables learning travel behavior from diverse datasets at the system level. This methodology not only avoids the need to use individual-level data collected from surveys or experiments to estimate route choice models but also allows to induce link flows that (i) satisfy Stochastic User Equilibrium under Logit assignment (SUELOGIT) and (ii) reproduce traffic counts that are observed at the system level.
2. It conducts a mathematical analysis of the non-convexity of the LUE problem with respect to the utility function coefficients estimated from traffic counts. For a single-attribute utility function and a general network, we provide sufficient conditions for the pseudo-convexity of the LUE problem, which guarantees that the solution is unique and that the gradients are always pointing toward the global optima. For a multi-attribute utility function and a general network, we prove that the problem is non-convex. In line with the theoretical results found for single-attribute and multi-attribute utility functions, experiments conducted in networks that have been studied in the prior literature show that the problem is coordinate-wise pseudo-convex.
3. It shows that integrating first-order and second-order optimization methods outperforms existing methods to solve the LUE problem. The use of first-order optimization methods is supported by the realization that the LUE problem can be, at best, pseudo-convex. Experiments with networks studied in the prior literature show that our optimization algorithm improves the recovery of the utility function coefficients.
4. It develops a statistical framework grounded in non-linear least squares (NLLS) to perform hypothesis testing and attribute selection on the coefficients of multi-attribute utility functions. Extensive Monte-Carlo experiments show that hypothesis testing is reliable under various levels of random noise and sensor coverage. To our knowledge, this is the first time that a hypothesis test framework has been developed for the LUE problem in the network context. A thorough analysis of the NLLS assumptions is conducted to ensure that hypothesis tests are valid for the LUE problem.
5. It solves the LUE problem in a large-scale transportation network and with real-world system-level data. To make the estimation scalable, we adapt prior column-generation algorithms to solve path-based SUELOGIT efficiently. We also derive, in vectorized form, the analytical gradients of the objective function with respect to the utility function coefficients. Using analytical gradients instead of numerical differentiation significantly reduces the computational burden of solving LUE at a large scale. To our knowledge, this is the first time that analytical gradients in vectorized form are provided for a path-based formulation of the LUE problem.

Our methodology is open source to ease adoption and further extension by the transportation community (Section 11). To our knowledge, there are no other open-source tools available to solve the LUE problem.

3. Estimation of travelers' utility function coefficients with system level data

This section discusses in detail our methodology to estimate the travelers' utility function coefficients using system-level data. We first introduce our main assumptions and then we describe the components of the bilevel optimization program formulated to estimate the utility function coefficients.

3.1. Assumptions

Assumption 1 (*Equilibrium Principle*). Network traffic flow follows stochastic user equilibrium with logit assignment (SUELOGIT)

With the motivation of bridging transportation network analysis and the study of travel behavior with discrete choice models, we focused on the problem of estimating the travelers' utility function from link flow measurements consistent with stochastic user equilibrium under logit assignment (SUELOGIT). On one hand, SUELOGIT is one of the many alternative representations of user equilibrium used in the networking modeling community, hence it may be considered a strong assumption. On the other hand, SUELOGIT is in line with the state of the practice in travel behavior research, where the logit model remains as one of the gold standards for depicting individual route choices. The good compromise between behavioral realism and the mathematical tractability of the logit model is a key property for the estimation in discrete choice models, and that will also be leveraged in the formulation and solution of our problem.

Assumption 2 (*Exogenous and Deterministic O-D Matrix*). The origin–destination (O-D) demand matrix is deterministic and exogenous

O-D demand estimation (ODE) is a complex problem that has been largely studied in the network modeling literature. As an underdetermined problem, there is no guarantee of solution uniqueness for the estimated O-D matrix and this makes difficult to verify the consistency of the solution with a ground truth O-D matrix. The ODLUE literature has extended ODE to estimate the travelers' utility function coefficients on top of the O-D matrix. The ODLUE problem is harder to solve due to the additional degrees of freedom introduced by an endogenous O-D matrix. Therefore, as ODE, it also suffers from solution non-uniqueness.

The assumption of an exogenous and deterministic O-D matrix allows us to focus on the estimation of the utility function coefficients, a first step to understanding the mathematical properties of our problem. In the context of the existing literature, our methodology can be seen as a case where a reasonably accurate reference O-D matrix is available and assumed as the ground truth O-D matrix. As we will show later, we can still obtain satisfactory results if the exogenous O-D demand is noisy but not too far from the ground truth O-D. Note also that our methodology can be extended to an iterative process where ODE and LUE alternate if the goal is to simultaneously estimate both the travelers' utility function coefficients and the O-D matrix.

Assumption 3 (*Linearity and Homogeneity of Utility Function*)**.** The utility function is a linear weight of attributes and coefficients, and the coefficients are common/homogeneous for all individuals traveling in the transportation network.

Choosing a linear-in-parameters utility function with homogeneous coefficients is a standard assumption for estimating discrete choice models in travel behavior studies. The preference for linear over non-linear specifications of the utility function is motivated by parsimony arguments and the convenience of dealing with a concave likelihood function in the estimation of multinomial logit models (MNL) models. In our case, this assumption significantly facilitates the proof of theoretical guarantees in our optimization problem. Note that this assumption could be relaxed by allowing a variation of the utility function coefficients at the O-D pair level or by letting the coefficients be drawn from some arbitrary probability density function. Both modeling strategies can be seen as analogies to the systematic taste variation used in MNL models and to the mixing distributions used in Mixed Logit models (Train, 2003).

Assumption 4 (*Properties of Link Performance Functions*)**.** The performance function of a link is monotonically increasing with respect to the traffic flow at that link, and the parameters of this function are exogenous

The exogeneity of the link performance parameters is a standard assumption in transportation network analysis. Following the state of the practice in prior literature, we adopt the Bureau of Public Roads (BPR) link travel time function. For each link, the travel time monotonically increases with respect to the link flow. For the sake of simplicity, we also assume that there is no interaction between links, and hence, a link's travel time depends on the traffic flow in that link only. A direct consequence of this assumption is the uniqueness of the link flow solution of SUELOGIT. Some work in the literature has proposed methods to estimate the parameters of the link performance function, namely, relaxing the exogeneity assumption. Although it is feasible to estimate these parameters on top of the travelers' utility function coefficients within our modeling framework, the introduction of these additional degrees of freedom may cause identifiability issues, or it could make the statistical inference of the utility function coefficients less reliable.

Assumption 5 (*Path Set Generation*)**.** The set of paths between each O-D pair has a fixed size and it is dynamically updated over iterations with column-generation methods

Our model solves a path-based formulation of SUELOGIT. To improve our prior about the true composition of the path set, the path sets are dynamically updated via column generation methods (Damberg et al., 1996; García-Ródenas and Marín, 2009). The size of the consideration sets is treated as a hyperparameter and it is constrained to small values to reduce the computational burden. Previous literature suggests generating paths that do not overlap with existing paths during the column generation phase (Damberg et al., 1996). However, this approach requires solving a combinatorial problem on all possible subsets of paths and it implicitly assumes that travelers' consideration sets have paths with few or no overlap. To address this challenge, we leverage existing methods to correct path-overlapping in path-based formulations (Ben-Akiva and Bierlaire, 1999).

Remark 1. Most prior work on the LUE problem assumes that the consideration set in an O-D pair is equal to the set of all reasonable paths in that pair (Yang et al., 2001; Wang et al., 2016; Lo and Chan, 2003; Liu and Fricker, 1996). Under this assumption, each path choice probability can be decomposed into a weight of exponentials of their link utilities. This approach leads to the link-based formulation of SUELOGIT. On one hand, this formulation avoids the enumeration of all paths in the network and it can significantly speed up the computation of SUELOGIT that underlies the estimation of the travelers' utility function coefficients. On the other hand, it implicitly assumes that travelers consider all paths to travel between an O-D pair. This assumption induces higher overlapping among paths, which is known to affect the computation of path choice probabilities in multinomial logit models (MNL) and it has been identified as a limitation of algorithms that rely on the set of all reasonable paths.

3.2. Stochastic user equilibrium with logit assignment (SUELOGIT)

In a congested network, travel times are endogenous and dependent on the traffic flows. At equilibria, both traffic flows and travel times are expected to reach a stationary point where travelers have no incentives to switch to alternative paths. Depending on the underlying behavioral representation used to model travelers' route choices, different types of network equilibrium are induced. In Deterministic User Equilibrium (DUE), travelers' choices are deterministic, meaning that the modeler fully knows the specification of the traveler's utility function. In contrast, in stochastic user equilibrium (SUE), travelers' choices are assumed to be probabilistic, and modelers may ignore a set of unobservable components of the traveler's utility function. In the network context, at SUE, no user believes he can improve his travel cost by unilaterally changing routes (Daganzo and Sheffi, 1977). Under SUE with logit assignment (SUELOGIT), travelers' are assumed to make choices consistent with a multinomial logit model (MNL).

3.2.1. Multinomial logit (MNL) route choice model

The logit model remains as the gold standard to model route choices due to its good compromise between behavioral realism and mathematical tractability (McFadden, 1973). Consider a traveler $l \in \mathcal{L}$ choosing a path i within her consideration set $\mathcal{J}_l$ and according to the latent utility U_{jl} of each alternative path $j \in \mathcal{J}_l$. Suppose that the modeler knows the observable component V_{jl} of the travelers' latent utility but ignores a component e_{jl}, which can be safely assumed to be stochastic. Therefore, if any attribute relevant to route choice decision is unobserved, travelers' choices could look probabilistic but not deterministic from the modelers' perspective. In particular, if $e_{jl} \overset{\text{i.i.d}}{\sim} \text{EV}(0,\mu)$ the probability p_i that a traveler l chooses a path $i \in \mathcal{J}_l$ will have the following closed form:

$$p_i = \frac{\exp(\mu V_i)}{\sum_{j\in\mathcal{J}_l}\exp(\mu V_j)} \tag{1}$$

where $\mu > 0$ is a scale parameter of the Extreme Value (EV) Type 1 distribution and is set to 1 for convenience and identification purposes. μ is inversely proportional to the standard deviation of the random component, and thus, it is expected to be lower as smaller is the unobserved component of the latent utility. At an extreme case where the latent utility function is fully observed, $\mu \to \infty$ and the choice probabilities reduce to an indicator function taking the value one if the utility of an alternative is the highest within a choice set and zero otherwise, i.e., the DUE case.

Remark 2. To capture the correlation between paths in the consideration set due to shared link segments, path utilities can be corrected with the same principle used in the path size logit (PSL) model (Ben-Akiva and Bierlaire, 1999). Path utilities incorporate the logarithm of a factor that increases with the amount of overlapping among paths within the same consideration set. A path with no overlapping links needs no utility adjustment since the PSL factor has a size of one. Thus, the PSL correction reduces the chances of a violation of the Independence of Irrelevant Alternatives (IIA) assumption of the MNL model.

3.2.2. SUE-logit with multi-attribute utility function

The original formulation of the SUELOGIT problem entails that route choices are made based on travel cost/time only (Fisk, 1980). Utility, however, may depend on additional attributes such as travel time reliability, waiting time, or monetary costs (e.g. toll fees). To make a more explicit bridge between the travelers' utility function specified in route choice models and to extend the analysis for a multi-attribute case, we reformulated the SUELOGIT problem as follows:

$$\begin{aligned} \max_{\mathbf{x},\mathbf{f}} \quad & \sum_{a\in\mathcal{A}} \int_0^{x_a} v_a(u,\boldsymbol{\theta})du - \langle \mathbf{f}, \ln \mathbf{f}\rangle \\ \text{s.t.} \quad & \mathbf{M}\mathbf{f} = \mathbf{q} \\ & \mathbf{D}\mathbf{f} = \mathbf{x} \\ & \mathbf{x},\mathbf{f} \geq \mathbf{0} \end{aligned} \tag{2}$$

where $\mathbf{x} \in \mathbb{R}^{|\mathcal{A}|}_{\geq 0}, \mathbf{f} \in \mathbb{R}^{|\mathcal{H}|}_{\geq 0}, \mathbf{q} \in \mathbb{R}^{|\mathcal{W}|}_{+}, \mathbf{M} \in \mathbb{R}^{|\mathcal{W}|\times|\mathcal{H}|}_{\{0,1\}}$, $\mathbf{D} \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{H}|}_{\{0,1\}}$. Besides, $v_a(u,\boldsymbol{\theta}) = \theta_t t_a(u) + \sum_{k\in\mathcal{K}_\mathbf{Z}} \theta_k \cdot Z_{ak}$ is the utility associated with link a at a traffic flow level u. Z_{ak} is the value of the exogenous attribute $k \in \mathcal{K}_\mathbf{Z}$ at link a, $\mathcal{K}_\mathbf{Z}$ is the set of exogenous attributes at each link, $\boldsymbol{\theta}_Z \in \mathbb{R}^{|\mathcal{K}_\mathbf{Z}|}$ and $\boldsymbol{\theta} = \begin{bmatrix}\theta_t & \boldsymbol{\theta}_Z\end{bmatrix} \in \mathbb{R}^{|\mathcal{K}|}$ are the vector of coefficients associated to the exogenous attributes and all attributes, respectively. Exogenous attributes may include the number of traffic lights, street intersections, or the level of income of the area where a link is located. From the specification of $v_a(\cdot)$, it follows that travel time (t) is the only endogenous attribute in the utility function, and it is linearly weighting the preference coefficient θ_t. The first order optimality condition of Problem 2 gives the following path flow solution:

$$f_h^\star = q_w \frac{\exp\left(\sum_{a\in\mathcal{A}} v_a^\star \delta_{ah}\right)}{\sum_{j\in\mathcal{H}_w}\exp\left(\sum_{a\in\mathcal{A}} v_a^\star \delta_{aj}\right)} \tag{3}$$

where $\delta_{ah} = \mathbb{I}(\text{link } a \in \text{ path } h)$ and $v_a^\star$ is the link utility of link $a \in \mathcal{A}$ at SUELOGIT. As expected, the path flow vector $\mathbf{f}_h^\star$ at SUELOGIT follows a logit distribution. From a micro-level travel behavior standpoint, q_w individuals traveling in the O-D pair $w \in \mathcal{W}$ are making route choices according to a logit model (Eq. (1)). Hence, path flows $\mathbf{f}_h^\star$ are the aggregate of these individual decisions (Eq. (3)). A detailed derivation of the extension of the Fisk (1980) formulation from a single to a multi-attribute case is presented in Appendix A.2.

Remark 3. In line with the rules of parameterization used to derive the MNL, the specification of the utility function used to compute SUELOGIT should account for the fact that the utility function coefficients are scaled by a factor $\mu \in \mathbb{R}_+$ that is inversely proportional to the standard deviation of the unobservable component of the utility function, i.e., $\boldsymbol{\theta} = \mu\tilde{\boldsymbol{\theta}}$ where $\tilde{\boldsymbol{\theta}}$ is the unscaled vector of logit coefficients and which is not identifiable. A direct consequence in the single attribute case is that only the sign but not the magnitude of the coefficients should be assumed to be known by the modeler. This is particularly relevant in cases where estimates from external travel behavior studies are used to set the utility function coefficients of the route choice model defined for the computation of SUELOGIT. Therefore, accounting for the scale factor of the logit parameters is critical for a correct economic and behavioral interpretation of the coefficients of multi-attribute utility functions. However, this is often overlooked by studies in the LUE and ODLUE literature.

3.2.3. Stochastic network loading

Suppose the vector of the travelers' preferences $\boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{K}|}$, the values of the matrix of exogenous attributes $\mathbf{Z}$, and the travel times at SUELOGIT are known and that the goal is to find the resulting path/link flows in the transportation network. This is the problem that stochastic network loading (SNL) aims to solve. Assumptions about the composition of the consideration set can significantly speed up the computation of SNL. A well-known example is the Dial algorithm, which, under the assumption of path sets containing all reasonable paths, allows decomposing path probabilities into link-level weights that are used to recursively obtain the resulting link flows in the transportation network. Alternatively, and in line with Assumption 5, path sets can be dynamically updated via column generation methods and SNL can be solved at the path level. The pseudo-code of our SNL method is shown in Algorithm 2, Appendix C.1.

3.2.4. Solution methods for SUELOGIT

If travel times were insensitive to variation of link flows, namely, exogenous, a single computation of SNL would suffice to obtain the link and path flows at SUELOGIT. In practice, travel times are endogenous variables, and thus, the computation of SUELOGIT requires performing SNL multiple times.

The method of successive averages (MSA) is one of the prominent algorithms used to compute SUELOGIT in the ODLUE literature. At the initial iteration, SNL is computed to generate a feasible link flow solution. For each of the following iterations, the convex combination of the SNL link flow solutions at the current and the previous MSA iteration is computed to obtain a new feasible solution. MSA defines a step size $\lambda_i = 1/(1+i)$ to set the weights λ_i and $1-\lambda_i$ that are used to compute the convex combination of solutions at iteration i. Because the feasible set of the SUELOGIT is convex, the convex combination necessarily gives a feasible point. The process is repeated until some convergence criterion has been achieved, such as the difference between the current and previous feasible link flow solution.

Frank–Wolfe (F–W) is a general-purpose method for constrained convex optimization that can outperform MSA by allowing a smarter choice of the step size parameter λ. A key thing to notice is that the convex combination of two feasible link flows solutions obtained via SNL provides a feasible descent direction for the objective function of the traffic equilibrium problem (Eq. 2). Thus, the value of λ used for the convex combination of solutions is chosen to maximize the improvement of the SUELOGIT objective. This solution strategy for SUELOGIT was originally implemented by Chen and Alfa (1991) with a utility function dependent on travel time only, and it is also integrated into the disaggregate simplicial decomposition (DSD) algorithm developed by Damberg et al. (1996). A distinct feature of DSD is the introduction of a column generation phase that dynamically update the consideration sets among O-D pairs. Consideration sets can be augmented with new paths according to different criteria, such as the choice probabilities or the level of dissimilarity of the set of candidate paths with respect to the existing paths among consideration sets.

3.3. Nonlinear least squares

Non-linear least squares (NLLS) is a standard estimation method within the ODE and ODLUE literature. In contrast to maximum likelihood estimation, there is no need to make distributional assumptions about the data-generating process, and there is a variety of specialized algorithms to minimize the NLLS optimization objective (Bazaraa et al., 2006). Note that when the response function is linear, NLLS reduces to ordinary least squares (OLS), and thus, the NLLS solution could be proved to be unique under mild conditions. A non-linear response function may induce multiple local minima and saddle points that could make the optimization to be sensitive to the starting points for optimization and converge toward a local but not the global minima. Therefore, the performance of the optimization algorithms heavily depends on the class of non-linearity of the response function with respect to the parameters of interest.

3.3.1. Problem formulation

Consider the regression equation:

$$\mathbf{y} = m(\boldsymbol{\beta}, \mathbf{X}) + \mathbf{u} \tag{4}$$

where $m(\cdot)$ is the response function, $\mathbf{X} \in \mathbb{R}^{|N| \times |\mathcal{K}_{\mathbf{Z}}|}$ is the matrix of values of a set $\mathcal{K}_{\mathbf{Z}}$ of exogenous attributes, $\mathbf{y} \in \mathbb{R}^{|N|}$ is the vector of values of the dependent variable, $\boldsymbol{\beta} \in \mathbb{R}^{|\mathcal{K}|}$ is the true parameter vector and $\boldsymbol{u} \in \mathbb{R}^{|N|}$ is a vector of random perturbations of some arbitrary distribution.

The NLLS problem consists in finding the NLLS estimator $\hat{\boldsymbol{\beta}}_{NLLS}$ that minimizes the residuals, namely, the deviation between the vector of observed measurements $\bar{\mathbf{y}}$ and the predictions of the response function:

$$\hat{\boldsymbol{\beta}}^{\star} = \arg\min_{\hat{\boldsymbol{\beta}}} \|\bar{\mathbf{y}} - m(\hat{\boldsymbol{\beta}}, \mathbf{X})\|_2^2 \tag{5}$$

The application of the NLLS formulation to our problem is direct. Let $\hat{\boldsymbol{\theta}} \in \mathbb{R}^{|\mathcal{K}|}$ be the vector of estimated coefficients in the travelers' utility function, $\mathbf{t} \in \mathbb{R}^{|\mathcal{A}|}$ the vector of link travel times, $\mathbf{Z} \in \mathbb{R}^{|\mathcal{A}| \times |\mathcal{K}_{\mathbf{Z}}|}$ the matrix of exogenous link attributes and $\mathbf{x}(\hat{\boldsymbol{\theta}}, \mathbf{Z}, \mathbf{t})$ the vector with the link flow functions of any link $a \in \mathcal{A}$. In particular, when all attributes of the travelers' utility function are exogenous and known, such that $\mathbf{t} = \bar{\mathbf{t}}$, $\mathbf{x}(\hat{\boldsymbol{\theta}}, \mathbf{Z}, \bar{\mathbf{t}})$ becomes a vector-valued function that resembles the response function $m(\cdot)$ in NLLS and it has the following closed form:

$$\mathbf{x}(\hat{\boldsymbol{\theta}}, \mathbf{Z}, \bar{\mathbf{t}}) = \mathbf{D}\mathbf{f} = \mathbf{D}\left((\mathbf{M}^{\top}\mathbf{q}) \circ \mathbf{p}(\hat{\boldsymbol{\theta}}, Z, \bar{\mathbf{t}})\right) \tag{6}$$

where the vector $\mathbf{p}(\theta, \mathbf{Z}, \bar{\mathbf{t}})$ of path choice probabilities is:

$$\mathbf{p}(\hat{\boldsymbol{\theta}}, \mathbf{Z}, \bar{\mathbf{t}}) = \exp\left(\mathbf{D}^{\top}\mathbf{v}_x(\hat{\boldsymbol{\theta}}, \mathbf{Z}, \bar{\mathbf{t}})\right) \oslash \left(\mathbf{M}^{\top}\mathbf{M}\exp(\mathbf{D}^{\top}\mathbf{v}_x(\hat{\boldsymbol{\theta}}, \mathbf{Z}, \bar{\mathbf{t}}))\right) \tag{7}$$

and $\oslash$ is an operator for element-wise division. The vector $\mathbf{v}_x(\theta, \mathbf{Z}, \bar{\mathbf{t}})$ of link utilities is:

$$\mathbf{v}_x(\hat{\boldsymbol{\theta}}, \mathbf{Z}, \bar{\mathbf{t}}) = \bar{\mathbf{t}}\hat{\theta}_t + \mathbf{Z}\hat{\boldsymbol{\theta}}_{\mathcal{K}_{\mathbf{Z}}} = \begin{bmatrix}\bar{\mathbf{t}} & \mathbf{Z}\end{bmatrix}\begin{bmatrix}\hat{\theta}_t \\ \hat{\theta}_{\mathcal{K}_{\mathbf{Z}}}\end{bmatrix} = \begin{bmatrix}\bar{\mathbf{t}} & \mathbf{Z}\end{bmatrix}\hat{\boldsymbol{\theta}} \tag{8}$$

Finally, if $\bar{\mathbf{x}} \in \mathbb{R}^{|\mathcal{A}^o|}$ is the vector of observed traffic counts, the NLLS estimator $\hat{\theta}^{\star}$ of $\theta \in \mathbb{R}^{|\mathcal{K}|}$ can be obtained as:

$$\hat{\theta}^{\star} = \arg\min_{\hat{\theta}} \|\mathbf{x}(\hat{\theta}, \mathbf{Z}, \bar{\mathbf{t}}) - \bar{\mathbf{x}}\|_2 \tag{9}$$

3.3.2. Solution methods

Second-order optimization methods remain a gold standard to estimate the travelers' utility function from traffic counts in the LUE and ODLUE literature. One of the earliest examples in the ODLUE literature is found in Liu and Fricker (1996), which uses the secant method, a Quasi-Newton method for unidimensional optimization, to estimate the coefficients of a utility function dependent on travel time only. More recent examples are Yang et al. (2001) and Wang et al. (2016), who solved the outer-level problem of the ODLUE via Sequential Quadratic Programming (SQP), a variant of the Newton method for constrained optimization. Applications of second-order methods are also found in travel behavior research where Broyden–Fletcher–Goldfarb–Shannon (BFGS) remains as a preferred optimizer to find the maximum likelihood estimates in discrete choice models (Train, 2003).

Second-order optimization methods can achieve a superquadratic or superlinear convergence in convex problems. They are appealing when datasets are of moderate size and, thus, when the computational cost for the calculation or approximation of the Hessian matrix is reasonable. However, their convergence guarantees heavily rely on how well the curvature of the objective function informs about the direction of its steepest descent/ascent. Besides, in non-convex problems, their performance strongly depends on how close the initial points for optimization are to a local optimum and whether these points are located within a locally convex region of the objective function. Furthermore, the frequent sign changes of the curvature in non-convex functions may cause that second-order optimization methods do not converge to local minima or that get stuck at saddle points or flat regions of the optimization landscape.

First-order methods have become a gold standard for non-convex optimization due in part to the huge computational gains that are attained when avoiding Hessian matrix computation. The machine learning community has made continued efforts to develop multiple variants of these methods that are able to accelerate the optimization without compromising computational costs and that can perform reasonably well in highly non-convex optimization landscapes (Staib et al., 2019). Some optimizers that remain in state-of-the-art applications include stochastic gradient descent (SGD) (Robbins and Monro, 1951) and the Adagrad (Duchi et al., 2011) and Adam (Kingma and Ba, 2015) optimizers.

In the context of quasiconvex optimization problems, Hazan and Levy (2015) proposed a modified version of gradient descent (GD) in which the gradient direction is normalized by its norm. The existence of flat regions in quasiconvex problems causes the gradients to become small despite the fact that they may be pointing in the right direction of improvement. In this setting, the normalization of the gradient is helpful to keep improving the objective function in flat regions and avoids gradient explosion in sharp regions of the feasible space. The application of normalized gradient descent (NGD) in unidimensional unconstrained optimization problems generates parameter updates according to the sign of the first derivative of the objective function.

Interestingly, the ODLUE and LUE literature has not explored the use of first-order optimization methods to estimate the utility function coefficients using system-level data. Further sections of this paper will show some theoretical and empirical results to support the choice of NGD over the standard second-order optimization methods in previous work.

3.4. Bilevel optimization

3.4.1. Problem formulation

The sole application of the NLLS method to estimate the travelers' utility function coefficients from system-level data requires that travel times are known and exogenous. In uncongested networks, this assumption may be enforced by setting the travel times equal to the free flow travel times. In congested networks, this assumption is not plausible; hence, it is necessary to also solve the SUELOGIT problem to account for the equilibrium condition in the network. One strategy is to solve the following bilevel optimization problem:

$$\begin{aligned}
&\min_{\hat{\theta}} \ \ell(\hat{\theta}) = \|\mathbf{x}(\hat{\theta}) - \bar{\mathbf{x}}\|_2^2 \\
&\quad \text{s.t. } \mathbf{x}(\hat{\theta}) \in \arg\max_{\mathbf{x}} g(\hat{\theta}, \mathbf{x}, \mathbf{f}) \\
&\qquad \max_{\mathbf{x},\mathbf{f}} \ g(\hat{\theta}, \mathbf{x}, \mathbf{f}) = -\langle \mathbf{f}, \ln \mathbf{f} \rangle + \sum_{a \in \mathcal{A}} x_a \sum_{k \in \mathcal{K}_{\mathbf{Z}}} \theta_k \cdot Z_{ak} + \sum_{a \in \mathcal{A}} \int_0^{x_a} \theta_t t_a(u)\, du \\
&\qquad\quad \text{s.t. } \mathbf{Mf} = \mathbf{q} \\
&\qquad\qquad \mathbf{Df} = \mathbf{x} \\
&\qquad\qquad \mathbf{x}, \mathbf{f} \geq \mathbf{0}
\end{aligned} \tag{10}$$

where $\ell(\hat{\theta})$ and $g(\hat{\theta}, \mathbf{Z}, \mathbf{x}, \mathbf{f})$ are the objective functions at the upper and lower level and $\hat{\theta} \in \mathbb{R}^{|\mathcal{K}|}$ is the vector of estimated utility function coefficients. The idea behind expressing the objective functions in terms of $\hat{\theta}$ can be understood as follows. At the inner level, a given value of $\hat{\theta} = \tilde{\theta}$ will lead to a different assignment of path/link flows in the network and thus, to a different value of the objective function g. Likewise, the link flow obtained from the inner-level problem when $\hat{\theta} = \tilde{\theta}$ will induce another value of the objective function ℓ of the upper-level problem. Furthermore, thanks to the SUELOGIT property (Eq. (3)), $\mathbf{x}(\hat{\theta})$ can be approximated as an analytical function of $\hat{\theta}$ in the upper-level problem.

Remark 4. Note that Problem (10) could also be expressed as a single-level formulation. If SUELOGIT is formulated as a fixed-point problem, the inner-level problem can be cast in a set of $|\mathcal{A}|$ non-linear equations that depend on the link flow variables. Then, this set of equations could be added as constraints of the upper-level problem. On the one hand, solving a single-level problem circumvents the need to use alternating optimization schemes to solve bilevel formulations. On the other hand, the feasible space of the problem can become highly non-convex with the addition of the fixed-point equations. In addition, each iteration solves a larger number of decision variables, which can also be detrimental to optimization performance. In contrast, in the bilevel formulation, the feasible spaces of the upper and inner levels are unconstrained and convex (Proposition 2), respectively, which allows leveraging specialized algorithms that optimize for a subset of the decision variables at each level (Section 3.2.4).

3.4.2. Solution methods

A brute-force strategy to solve the bilevel formulation in Problem (10) would consist of performing a grid search on all feasible values of $\hat{\theta}$. Therefore, for each $\hat{\theta}$, SUELOGIT would be first computed at the inner level. Then the objective function at the upper level would be evaluated using the closed form expression derived in Eq. (6), Section 3.3.1. Lastly, a candidate solution would be the value of θ that minimizes $\ell(\hat{\theta})$. However, this solution strategy becomes intractable with a high dimensional multi-attribute travelers' utility function.

A standard heuristic employed to solve bilevel formulations in the ODE and ODLUE literature is to perform an alternating optimization between the inner-level and outer-level problems. Over iterations, this heuristic is expected to converge toward a stationary point that, at the outer level, minimizes the gap between observed and predicting flows and, at the inner level, induces a link/path flow solution that satisfies traffic equilibria. The solution strategy in our problem boils down to performing gradient-based updates of the utility function coefficients at the outer level and computing SUELOGIT at the inner level using the coefficients $\hat{\theta}$ previously updated at the outer level. The resulting travel times from the SUELOGIT computation are then fed to the outer level, and the process is repeated until some criterion is convergence is met. There are many variants of solution methods in the literature, but most can be summarized with the following steps:

1. Initialization of estimated parameters (e.g., utility function coefficients or O-D matrix)
2. Inner-level problem: solve SUELOGIT and update travel times in the network
3. Outer-level problem: solve NLLS with new travel times and update estimated parameters
4. Repeat 2 and 3 until fulfilling some convergence criteria

In our application and similar to the LUE literature, the utility function coefficients in $\hat{\theta}$ are the only free parameters.

Remark 5. A key advantage of SUELOGIT in network modeling applications is the existence of a closed form and unique solution at the path and link flow spaces. This property can significantly ease the solution of the typical bilevel optimization structure that arises in ODLUE problems. Note that during the alternating optimization, the closed-form solution is valid but only for the set of travel times obtained in the inner-level problem. Thus, it is convenient to perform small updates of $\hat{\theta}$ at the outer-level problem such that the inner-level solution remains valid. In this respect, using the closed form solution of the SUELOGIT also guarantees that the change of the SUELOGIT solution is continuous and, thus, that a slight change of $\hat{\theta}$ will not change the inner solution drastically.

4. Mathematical properties of the optimization problem

This section studies the mathematical properties of our optimization problem. Note that our problem falls into the LUE class, and since LUE is a subclass of ODLUE, the analysis of the mathematical properties of our problem is also relevant to the ODLUE problem.

4.1. Exogenous case

Assuming that the utility function is dependent on exogenous attributes can ease the analysis of the mathematical properties. This assumption is plausible in an uncongested network where travel times are close to the free flow travel times or when there is access to travel time measurements for every link in the network. Alternatively, this assumption may be enforced by either setting the travel time coefficient of the utility function or the parameter α of the cost performance to zero, but this is arguably less realistic. This assumption reduces the bi-level formulation to the outer-level objective only because the travel times do not need to be estimated from the inner-level problem.

4.1.1. Non-convexity

The non-convexity of the LUE problem with respect to the travelers' utility function coefficients has been discussed in prior literature but via counterexamples only (Yang et al., 2001). To our knowledge, no formal proof of the non-convexity has been given yet. A key observation to prove the non-convexity of this problem is that its objective function f is upper-bounded with respect to the utility function coefficients. In a one-dimensional case, the upper bounds of f correspond to the solutions of UE when $\theta \to \infty^+$ and $\theta \to \infty^-$. Proposition 1 extends this idea to a multi-dimensional case where $\theta \in \mathbb{R}^{|\mathcal{K}|}$. The non-convexity of the LUE problem is then proven in Proposition 13.

Proposition 1. *The objective function of the* LUE *problem is lower-bounded and upper-bounded*

Proof. The objective function $\ell : \mathbb{R}^{|\mathcal{K}|} \to \mathbb{R}$ of the LUE problem can be expressed as:

$$\ell(\boldsymbol{\theta}) = \|\mathbf{x}(\theta) - \bar{x}\|_2^2 = \sum_{i \in N} \left(x_i(\boldsymbol{\theta}) - \bar{\mathbf{x}}_i\right)^2$$

where N is the set of observed traffic count measurements. Each traffic function $x_i, \forall i \in N$ can be bounded as:

$$0 \leq x_i(\theta) \leq \sum_{w \in \mathcal{W}} q_w = Q \tag{11}$$

where $Q \in \mathbb{R}_+$ is the total demand, namely, the sum of all cells in the O-D matrix $\mathbf{Q} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$. Then, the lower and upper bounds of the objective function can be found as follows:

$$\begin{aligned}
0 &\leq x_i^2(\theta) \leq Q^2 \\
-2\bar{x}_i + \bar{x}_i^2 &\leq x_i^2(\theta) - 2\bar{x}_i + \bar{x}_i^2 \leq Q^2 - 2\bar{x}_i + \bar{x}_i^2 \\
\bar{x}_i(\bar{x}_i - 2) &\leq (x_i(\theta) - \bar{x}_i)^2 \leq (Q - \bar{x}_i)^2 \\
\sum_{i \in N} \bar{x}_i(\bar{x}_i - 2) &\leq \sum_{i \in N} (x_i(\theta) - \bar{x}_i)^2 \leq \sum_{i \in N} (Q - \bar{x}_i)^2 \\
(\bar{\mathbf{x}}^\top - 2 \cdot \mathbf{1}^\top)\bar{\mathbf{x}} &\leq \|\mathbf{x}(\theta) - \bar{\mathbf{x}}\|_2^2 \leq \|\left(\mathbf{q}^\top \mathbf{1}\right)\mathbf{1}^\top - \bar{\mathbf{x}}\|_2^2
\end{aligned} \tag{12}$$

which completes the proof. □

Corollary 1 (*Non-Convexity of* LUE *Problem*)**.** *Suppose the coefficients of the travelers' utility function in the* LUE *problem are identifiable. Then, the* LUE *problem is non-convex*

Proof. The proof of Corollary 1 follows from the fact that upper-bounded functions are non-convex. A detailed proof of the non-convexity of the LUE problem is presented in Appendix A.6. □

Convexity is a sufficient but not necessary condition for global optimality. Therefore, before ruling out the global optimality of the LUE problem, it is convenient to study a generalization of convexity known as a pseudo-convexity that imposes weaker conditions on the optimization problem and that provides sufficient conditions for global optimality.

4.1.2. Pseudo-convexity

Pseudo-convex functions are a subclass of quasi-convex functions. Hence, the properties of pseudo-convex functions are also valid for quasi-convex functions, while the converse is not true (Mangasarian, 1965). Pseudo-convexity is a generalization of convexity that, as quasi-convexity, extends the notion of unimodality to higher dimensional spaces. In contrast to quasi-convex functions but similar to convex functions, the first-order necessary optimality condition of pseudo-convex functions suffices for global optimality (Crouzeix and Ferland, 1982). This property will be key later to prove the existence of a global minimizer in our problem. A formal definition of pseudo-convexity is the following (Bazaraa et al., 2006):

Definition 1 (*Pseudo-Convexity*)**.** f is a *pseudo-convex* function on the feasible set S iff $\forall x_1, x_2 \in \mathcal{S}$:

$$f(x_2) < f(x_1) \implies \nabla f(x_1)^\top (x_2 - x_1) < 0 \tag{13}$$

or equivalently,

$$\nabla f(x_1)^\top (x_2 - x_1) \geq 0 \implies f(x_2) \geq f(x_1) \tag{14}$$

Remark 6. The definition of pseudo-convexity can be assessed via the augmented Hessian of the objective function (Bazaraa et al., 2006). Formally, f is pseudoconvex if there exists a scalar $v,\ 0 \leq v < \infty$ such that its augmented Hessian $H(x) + v\nabla f(x)\nabla f(x)^T$ is positive semidefinite for all $x \in \mathcal{X}$ (Crouzeix and Ferland, 1982). This proposition can be also evaluated in terms of the signs of the principal minors of the bordered Hessian (Mereau and Paquet, 1974).

4.1.3. Existence and uniqueness of global minima

In the absence of measurement error in the traffic counts, the proofs for the existence and uniqueness of global optimality are direct (Propositions 9 and 10, Appendix A.5.1). While the absence of measurement error can be enforced in experiments with synthetic data, it is implausible in real-world problems where traffic count data is expected to be noisy. Fortunately, the pseudo-convexity of the objective function in our problem can be leveraged to extend the proof of existence and uniqueness in the presence of measurement error. A key property of pseudo-convex functions that also holds in convex functions is that every local minimum is also a global minimum (Mangasarian, 1965). This relaxes the positive (semi)definite assumption of the Hessian of the objective function used to prove (strict) global optimality in convex problems. Following the proof in Bazaraa et al. (2006) for pseudo-convex functions, Proposition 11, Appendix A.5.2 proves the existence of global minima in the LUE problem when the travelers' utility function depends on a single exogenous attribute. Proposition 12, Appendix A.5.2 proves the uniqueness of the global optima in our problem by leveraging the strict quasi-convexity of differentiable pseudo-convex functions proven by Mangasarian (1965). While this seems restrictive, it addresses the standard setting studied in prior literature where utility functions depend on travel time only. Proposition 8, Appendix A.4 provides a sufficient condition for the pseudo-convexity of the LUE problem.

4.2. Endogenous case

In a congested network, travel times are endogenous and this requires solving a bilevel formulation of the LUE problem (see Section 3.4). Even if the inner and outer-level problems of the bilevel formulation were convex, the analysis of global optimality is complex (Dempe and Franke, 2016). Because convex functions are a subclass of the pseudo-convex functions, the pseudo-convexity of the outer problem is expected to make the analysis even harder. Despite the above limitations, this section proves two properties of the inner-level and outer-level problems of the LUE bilevel formulation that can improve the convergence of an alternating optimization algorithm.

4.2.1. Inner-level problem

The SUELOGIT problem with a utility function dependent on travel time only is known to be strictly convex in the path (and link) flow space under link performance functions that are monotonically increasing and dependent on the traffic flow on the links only. Proposition 2 extends the aforementioned property to multi-attribute utility functions.

Proposition 2 (*Strict Convexity of* SUELOGIT *and Uniqueness of Path Flow Solution with a Multi-Attribute Utility Function*)**.** *Suppose that the travelers' utility function is linear-in-parameters and dependent on travel time and on a set of exogenous attributes. Assume that the performance function in each link is monotonically increasing and dependent on the traffic flow on that link only. Then, if the marginal utility of travel time is negative, the* SUELOGIT *problem has a unique solution in path flow space.*

Proof. Problem 2, Section 3.2.2 solves SUELOGIT for performance functions that are dependent on the traffic flow of each link only. This problem can be transformed into a minimization problem dependent on path flows variables only and with the following objective function:

$$
\begin{aligned}
g(\mathbf{x},\mathbf{f}) &= -\sum_{a\in\mathcal{A}}\int_0^{x_a} v_a(u,\boldsymbol{\theta})du - \langle \mathbf{f}, \ln \mathbf{f}\rangle \\
&= -\sum_{a\in\mathcal{A}}\int_0^{\sum_{h\in H}\delta_{ah}f_h}\left(\theta_t t_a(u) - \sum_{k\in\mathcal{K}_\mathbf{Z}}\theta_k\cdot Z_{ak}\right)du - \langle \mathbf{f}, \ln \mathbf{f}\rangle \\
&= g_1(\mathbf{f}) + g_2(\mathbf{f})
\end{aligned}
$$

where the link flow variables in Problem 2 are expressed in terms of path flow variables using the set of link-path flow conservation constraints $x_a = \sum_{h\in H}\delta_{ah}f_h, \forall a\in\mathcal{A}$. Note that the non-negativity constraints associated with the link flow variables in Problem 2 are redundant because they are directly satisfied in the presence of the non-negativity constraints for the path flow variables. Because the matrix $\mathbf{Z}\in\mathbb{R}^{|\mathcal{A}|\times|\mathcal{K}_\mathbf{Z}|}$ of link attributes is exogenous, $g_1(\mathbf{f})$ can be rewritten as:

$$
g_1(\mathbf{f}) = -\theta_t\sum_{a\in\mathcal{A}}\int_0^{\sum_{h\in H}\delta_{ah}f_h} t_a(u)du - \sum_{a\in\mathcal{A}}\left(\sum_{h\in\mathcal{H}}\delta_{ah}f_h\right)\left(\sum_{k\in\mathcal{K}_\mathbf{Z}}\theta_k\cdot Z_{ak}\right)
$$

By assumption, each link performance function $t_a, \forall a\in\mathcal{A}$ is monotonically increasing, and thus, their integral is an increasing function, which is convex. Since the summation of the integrals of the performance functions is also convex and the marginal utility is negative, i.e., $-\theta_t > 0$, the first term of g_1 is convex. The second term in g_1 is an affine function with respect to $\mathbf{f}$, which is convex. As a result, g_1 is a sum of convex functions, which is convex. From Theorem 1 in Evans (1973), it follows that $g_2(\mathbf{f}) = -\langle \mathbf{f}, \ln \mathbf{f}\rangle$ is strictly convex with respect to the path flows variables on the feasible set $\mathcal{S}$ of Problem 2. As a consequence, $g(\mathbf{x},\mathbf{f}) = g(\mathbf{f})$ is a sum of a convex and strictly convex function, which is strictly convex. Furthermore, all constraints in Problem 2 are affine, and thus, they define a convex set. Finally, Problem 2 is strictly convex, and thereby, it has a unique global minimum, which completes the proof. □

4.2.2. Outer-level problem

A pseudo-convex function can be convex and concave in its feasible space. Despite the above, the gradient of a pseudo-convex function always points out to the global descent direction (Hazan and Levy, 2015). Proposition 3 shows that the property of the gradient also holds in the outer-level objective of the LUE problem when the problem is pseudo-convex. This property is key to justify the convenience of using first-order instead of second-order optimization methods to minimize the outer-level objective function of the LUE problem.

Proposition 3. *Assume that the objective function ℓ of the* LUE *problem is dependent on a single exogenous attribute and that ℓ is pseudo-convex with respect to the coefficient weighting that attribute in the travelers' utility function. If a global optimum exists, the negative first derivative of ℓ with respect to the utility function coefficient always points out to the global descent direction*

Proof. The proof of Proposition 3 follows from leveraging the gradient property of pseudo-convex functions in the unidimensional case. A detailed proof is included in Appendix A.7. Sufficient conditions for the pseudo-convexity of the LUE problem are provided in Proposition 8, Appendix A.4. □

Remark 7. This proposition addresses a traveler's utility function with a single exogenous attribute. Finding sufficient conditions for the pseudo-convexity of the objective function of the LUE problem in a multi-attribute case is challenging and left for further research. Furthermore, accounting for the endogeneity of the travel time for the analysis of pseudo-convexity is also complex. Suppose the travelers' utility function is a linear weight between travel time and a coefficient θ_t and that the transportation network is congested. Each value of θ_t would induce a different traffic assignment at the inner level of the bilevel formulation. Consequently, link travel times would change between iterations of an alternating optimization algorithm. Then, the objective function of the outer-level problem will be pseudo-convex with respect to θ_t only in a small neighborhood where travel times remain similar.

5. Solution algorithm and convergence guarantees

5.1. Inner-level optimization

The solution method for the inner-level problem of the bilevel formulation is described in Algorithm 3, Appendix C.2. The stochastic network loading (SNL) is performed according to Algorithm 2, Appendix C.1. The column generation phase is adapted from Damberg et al. (1996), and it is performed once per each iteration of the alternating optimization algorithm (Step 1, Algorithm 3, Appendix C.2). Note that in the original implementation of DSD, the column generation phase is performed at each iteration of SUELOGIT, but in the context of a bilevel optimization, this path set augmentation strategy would not select good paths if the current solution for $\boldsymbol{\theta}$ is far from the global optima. The best convex combination of solutions at each iteration of SUELOGIT (Step c, Algorithm 3, Appendix C.2) is searched over a uniform grid in $[0,1]$ and with an arbitrary granularity depending on the desired level of accuracy. To reduce the computational cost, the maximum size of the augmented path set is set to k_g, and the path set augmentation is only performed in a proportion ρ_W of the O-D pairs at each iteration. The sample of O-D pairs is selected according to the level of demand. To capture the correlation between paths in the consideration set due to overlapped link segments, we correct path utilities using the path size logit (PSL) factor (Ben-Akiva and Bierlaire, 1999). The utility term associated with the PSL factor is weighted by a coefficient β that can be estimated (Bierlaire and Frejinger, 2008). For the scope of this paper, the coefficient β is treated as a hyperparameter with a default value equal to one. The utility term associated with the path size correction is updated before generating new paths and after selecting paths depending if the composition of the path set changes when performing steps 1 and 3 of the inner-level optimization (Algorithm 3, Appendix C.2),

5.2. Outer-level optimization

To study the convenience of using first or second-order methods, the outer-level objective function of the bilevel formulation is optimized with a hybrid algorithm (see Algorithm 4, Appendix C.3). In a *non-refined stage*, the algorithm uses Normalized Gradient Descent (NGD) (Algorithm 5, Appendix C.3.4). Subsequently, the best solution obtained from the *non-refined stage* is used as a starting point for the *refined stage*, where second-order optimization methods are used to obtain a more accurate solution (Algorithm 6, Appendix C.3.4). Two specialized second-order methods were initially considered. The first is Gauss–Newton (GN), one of the standard optimizers in the Econometrics literature on non-linear regression (Del Pino, 1989; Amemiya, 1983; Gallant, 1975a) and which can be seen as an unconstrained version of the sequential quadratic programming (SQP) method. The second is the Levenberg–Marquardt (LM) algorithm which, by interpolating between GN and gradient descent, can increase the robustness of GN in flat regions of the optimization landscape (Marquardt, 1963). Given that GN is a particular case of the LM method (Algorithm 6, Appendix C.3.4), only LM is used in the *refined stage*.

5.3. Bilevel optimization

Algorithm 1, Section 5.3 shows the pseudocode for the alternating optimization of the inner and outer levels of our bilevel formulation (Problem (10), Section 3.4). The algorithm returns the vector of utility function coefficients $\hat{\boldsymbol{\theta}} \in \mathbb{R}^{|\mathcal{K}|}$ that minimizes the discrepancy between predicted and observed link flows. To guarantee that the predicted link flows follow SUELOGIT, the last iteration of the alternating algorithm only solves the inner-level problem using the value of $\hat{\boldsymbol{\theta}}$ obtained in the previous iteration of the outer-level problem.

5.4. Convergence guarantees

Our bilevel formulation can be cast as a mathematical program with equilibrium constraints (MPEC). For this class of problems, the feasible region is typically non-convex (Boyles et al., 2020), making it difficult to find theoretical guarantees even for convergence toward local optima (Sabach and Shtern, 2017; Dempe and Franke, 2016). To ease the analysis, we found it convenient to analyze a case where the utility function is dependent on a single exogenous attribute (Section 4.1) and where we could leverage the pseudo-convexity of the optimization problem (Section 4.1.2).

Algorithm 1 `BilevelOptimization`

Input: Iterations I, initial vector of utility function coefficients θ_0, inputs of `InnerLevelOptimization` $T, \mathbf{Z}, \mathbf{t}^0, \mathbf{q}, \boldsymbol{\lambda}_{FW}, \rho_W$, inputs of `OuterLevelOptimization` $T_1, T_2, \eta, \delta_{\text{LM}}, \bar{\mathbf{x}}, \mathbf{x}, \mathbf{p}$

Step 0: Initialization:

$$\theta := \theta_0$$

Step 1: Alternating optimization

for $i = 1 \dots I$ **do**
  $\mathbf{x}, \mathbf{p}, \tilde{\mathbf{t}} \leftarrow$ `InnerLevelOptimization`$(\theta, T, \mathbf{Z}, \mathbf{t}^0, \mathbf{q}, \boldsymbol{\lambda}_{FW}, \rho_W)$

  if $i < I$ **then**
    $\theta \leftarrow$ `OuterLevelOptimization`$(\mathbf{x}, \mathbf{p}, \bar{\mathbf{x}}, \tilde{\mathbf{t}}, \theta, T_1, T_2, \eta, \delta_{\text{LM}})$
  end if
end for

return $\theta^\star = \text{argmin}_{\{\theta_0,\dots,\theta_{I-1}\}} \ell_i(\theta_i) := \|\bar{\mathbf{x}} - \mathbf{x}(\theta_i)\|^2$

5.4.1. *Theoretical ϵ-convergence under exogenous travel times*

The analysis of mathematical properties in Section 4 provided sufficient conditions for the existence and uniqueness of a global optimum. Now it remains to prove that the global optimum is attainable when running normalized gradient descent (`NGD`) on a finite number of steps. We will again consider the `LUE` problem with a utility function dependent on a single exogenous attribute and where only the no-refined stage of the outer-level algorithm is conducted.

Hazan and Levy (2015) proved the convergence of `NGD` to a global minimum for functions that are both strictly quasi-convex and *locally-Lipschitz*. These results are relevant to our problem since pseudo-convex functions are also quasi-convex (Mangasarian, 1965). *Locally-Lipschitz* functions have the property of having their first derivative bounded by an arbitrary positive constant, and they are required to be *Lipschitz* in a small region around the optimum. Generally, the *Lipschitz constant* is relevant to study the theoretical convergence speed of first optimization methods. Proposition 15, Appendix A.7 proves the Lipschitzness of the `LUE` objective function when the travelers' utility function depends on a single exogenous attribute. Then, Proposition 4 provide guarantees for the convergence of `NGD` in the following problem setting:

Proposition 4. *Assume that the objective function of the* `LUE` *problem is pseudo-convex and that the travelers' utility function depends on a single exogenous attribute. If the optimization problem is solved with normalized gradient descent (*`NGD`*),* `NGD` *converges to an ϵ-optimal solution in a finite number of iterations (poly($1/\epsilon$))*

Proof. By assumption, the objective function $\ell : \mathbb{R} \to \mathbb{R}$ is pseudo-convex and thus strictly quasi-convex. Under exogenous travel times, we can use Proposition 15, Appendix A.7 to conclude that ℓ is locally Lipschitz. Thus, all conditions from Theorem 4.2, Hazan and Levy (2015) are satisfied, and hence the proof is complete. Sufficient conditions for the pseudo-convexity of the `LUE` problem are provided in Proposition 8, Appendix A.4. □

Remark 8. This proposition does not address the cases of a multi-attribute utility function or a congested network where travel time is endogenous. In these cases, our current results give no theoretical guarantees on the existence of a unique global optimum and the convergence of our algorithm toward a global optimum. A good practice to optimize non-convex problems is using different parameter initializations and selecting the solution that reduces the objective function the most. The convergence criterion will depend on the choice of the optimization algorithm.

6. Statistical inference on the parameters estimated from system level data

This section derives closed-form expressions of the statistical tests used to analyze the coefficients of the travelers' utility function estimated with the methodology presented in Section 3. Non-linear-least squares (`NLLS`) estimation is preferred over Maximum Likelihood Estimation (`MLE`) because the former relies on weaker distributional assumptions (Cameron and Trivedi, 2005).

6.1. *The* `NLLS` *estimator*

The nonlinear regression formulation of the outer-level objective of our problem defines each traffic count measurement $x_i, \forall i \in N$ in the true data generating process as a scalar dependent random variable with conditional mean:

$$\mathbb{E}(x_i|\mathbf{Z}, \bar{\mathbf{t}}) = x_i(\mathbf{Z}, \bar{\mathbf{t}}, \theta) \tag{15}$$

where x_i is the traffic count (response) function (Eq. (6), Section 3.3.1) associated to observation i, $\bar{\mathbf{t}}$ is the set of link travel times and $\mathbf{Z}$ is the matrix of exogenous attributes. Note that $\bar{\mathbf{t}}$ is assumed exogenous in the NLLS problem solved at the outer level, but it is iteratively updated via the alternating optimization with the inner-level problem (Section 5.3). The regression equation associated with the traffic count measurement x_i is defined as:

$$x_i = \mathbb{E}(x_i|\mathbf{Z}, \bar{\mathbf{t}}) + u_i \tag{16}$$

where u_i is the random error from the true data-generating process. The differentiation of the objective function of the NLLS problem (Appendix C.3.2) leads to the following first-order necessary optimality condition:

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{2}{|N|} \sum_{i \in N} \frac{\partial x_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} (x_i - x_i(\boldsymbol{\theta})) = \mathbf{0} \implies [D_{\boldsymbol{\theta}}\ x(\boldsymbol{\theta})]^{\top} \boldsymbol{u} = \mathbf{0} \tag{17}$$

where $|N|$ is the number of observations and $D_{\boldsymbol{\theta}}\ x(\boldsymbol{\theta}) \in \mathbb{R}^{|N| \times |\mathcal{K}|}$ is a Jacobian matrix corresponding to the stacked gradient vectors for each observation $n \in N$. Also, for the ease of notation and the remainder of the paper, $\mathbf{x}(\boldsymbol{\theta}) = \mathbf{x}(\mathbf{Z}, \mathbf{t}, \boldsymbol{\theta})$. Note that Eq. (17) defines a set of d equations, namely, one equation for each element of the vector of utility function coefficients $\boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{K}|}$, and it restricts the residual to be orthogonal to $D_{\boldsymbol{\theta}}\ x(\boldsymbol{\theta})$. A key difference of NLLS with respect to OLS is that the matrix of exogenous attributes of the regression equation is replaced by the Jacobian matrix $D_{\boldsymbol{\theta}}\ x(\boldsymbol{\theta})$ (Appendix A.8). Under this realization, the statistical properties of the NLLS problem can be derived analogously to the OLS case.

6.2. Assumptions

The proof of the consistency of the NLLS estimator relies on a series of assumptions that are well-established in Econometrics literature (Wu, 1981). An accessible reference on these assumptions is found in Cameron and Trivedi (2005). A thorough analysis of these assumptions in the context of the LUE problem is shown in Appendix B.1.

6.3. Statistical properties of the NLLS estimator

Suppose $\theta_0 \in \mathbb{R}^{|\mathcal{K}|}$ satisfies the first-order necessary optimality condition defined in Eq. (17). Then, under the assumptions stated in Appendix B.1, we can establish the consistency of the NLLS estimator $\hat{\theta}$ at θ_0 (see Proposition 5.6, Cameron and Trivedi, 2005) and that

$$\sqrt{|N|}(\hat{\theta} - \boldsymbol{\theta_0}) \xrightarrow{d} \mathcal{N}[0, \mathbf{A_0}^{-1}\mathbf{B}_0\mathbf{A_0}^{-1}] \tag{18}$$

where

$$\mathbf{A_0} = \text{plim} \frac{1}{|N|} \left[D_{\theta}\ x(\theta)\Big|_{\theta_0} \right]^{\top} \left[D_{\theta}\ x(\theta)\Big|_{\theta_0} \right], \quad \mathbf{B}_0 = \text{plim} \frac{\sigma^2}{|N|} \left[D_{\theta}\ x(\theta)\Big|_{\theta_0} \right]^{\top} \left[D_{\theta}\ x(\theta)\Big|_{\theta_0} \right] = \sigma^2 \mathbf{A_0}$$

By the non-singularity of $\mathbf{A_0}$ (Assumption 10, Appendix B.1):

$$\mathbf{A_0}^{-1}\mathbf{B}_0\mathbf{A_0}^{-1} = \mathbf{A_0}^{-1}\sigma^2\mathbf{A_0}\mathbf{A_0}^{-1} = \sigma^2\mathbf{A_0}^{-1}$$

By rearranging terms in Eq. (18), the asymptotic distribution of $\hat{\theta}$ becomes:

$$\hat{\theta} \xrightarrow{a} \mathcal{N}\left[\boldsymbol{\theta_0}, \text{Var}(\hat{\boldsymbol{\theta}})\right] \tag{19}$$

and where

$$\text{Var}(\hat{\boldsymbol{\theta}}) = \sigma^2 \left[D_{\theta}\ x(\theta)\Big|_{\theta_0} \right]^{\top} \left[D_{\theta}\ x(\theta)\Big|_{\theta_0} \right] \tag{20}$$

is the asymptotic variance matrix of the NLLS estimator. Note that the distribution of the NLLS estimator resembles OLS, except that the design matrix $\mathbf{X}$ is replaced by the Jacobian matrix $\tilde{\mathbf{X}} = D_{\theta=\theta_0}\ x(\theta)$ (Appendix A.8). Also, similar to the OLS case, the statistical behavior of the NLLS estimator and the variance of the error terms in large samples can be expressed in terms of $\boldsymbol{u}$ and approximated as (Gallant, 1975a):

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + (\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T\boldsymbol{u} \tag{21}$$

$$s^2 = (n-p)^{-1}\boldsymbol{u}^{\top}(\mathbf{I} - \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^{\top}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^{\top})\boldsymbol{u} \tag{22}$$

The variance $\sigma^2 \in \mathbb{R}_{\geq 0}$ of the error terms (Eq. (20)) can be consistently estimated as:

$$\hat{\sigma}^2 = \frac{\text{RSS}(\hat{\theta})}{|N| - |\mathcal{K}|} = \frac{\|\bar{\mathbf{x}} - x(\hat{\theta})\|_2^2}{|N| - |\mathcal{K}|} \tag{23}$$

where the numerator is the residual sum of squares (RSS) of the model and $|\mathcal{K}|$ is the dimension of the NLLS estimator.

6.4. Hypothesis testing and confidence intervals

Based on the asymptotic normality of the NLLS estimator $\hat{\theta}$ proved in the previous section, we can derive confidence intervals and hypothesis tests of $\hat{\theta}$. The $1-\alpha\%$ confidence interval of $\hat{\theta}$ can be found by using the percentile $\mathbf{1}-\alpha/2$ of a t-variate with $n-|\mathcal{K}|$ degrees of freedom $T_{n-|\mathcal{K}|,1-\alpha/2}$ and the variance of the NLLS estimator as shown in the following formula (Gallant, 1975a):

$$\mathtt{CI}(\hat{\theta})_{1-\alpha/2} = \hat{\theta} \pm t_{n-|\mathcal{K}|,1-\alpha/2}\sqrt{\mathrm{diag}(\mathrm{Var}(\hat{\theta}|\mathbf{X}))} \tag{24}$$

The hypothesis $H_0 : \theta_d = \theta_{H_0}$ can be contrasted at the $\alpha\%$ confidence level by computing the t-statistic:

$$\tilde{T}_{d,H_0} = \frac{\hat{\theta}_d - \theta_{H_0}}{\sqrt{\mathrm{diag}(\mathrm{Var}(\hat{\theta}))_d}} \tag{25}$$

and H_0 is rejected when $|\tilde{T}_{d,H_0}| > |T_{\mathbf{1}-\alpha/2}|$.

The F-test is another useful statistic for model comparison in non-linear regression and allows to contrast hypotheses on multiple model parameters (Gallant, 1975b). Let us define $\theta_1, \theta_2 \in \mathbb{R}^{|\mathcal{K}|}$ as the vector of coefficients in the restricted (or nested) and unrestricted models. Then, the F-statistic $\tilde{F}_{1,2}$ is:

$$\tilde{F}_{1,2} = \frac{\dfrac{\mathtt{RSS}(\theta_1) - \mathtt{RSS}(\theta_2)}{\|\theta_2\|_0 - \|\theta_1\|_0}}{\dfrac{\mathtt{RSS}(\theta_2)}{|N| - \|\theta_2\|_0}} \tag{26}$$

where $\|\theta_1\|_0$ and $\|\theta_2\|_0$ are the number of non-zero entries in the vectors of parameters of the non-restricted and restricted models. The null hypothesis is that the models are statistically equivalent. It is rejected when $\tilde{F}_{1,2} > F_{\|\theta_2\|_0-\|\theta_{12}\|_0,|N|-\|\theta_2\|_0,\alpha}$, where $F_{\|\theta_2\|_0-\|\theta_{12}\|_0,|N|-\|\theta_2\|_0,\alpha}$ is the value of a central F-distribution with $\|\theta_2\|_0 - \|\theta_1\|_0$ numerator degrees of freedom and $|N| - \|\theta_2\|_0$ denominator degrees of freedom and at the α-th upper percentile.

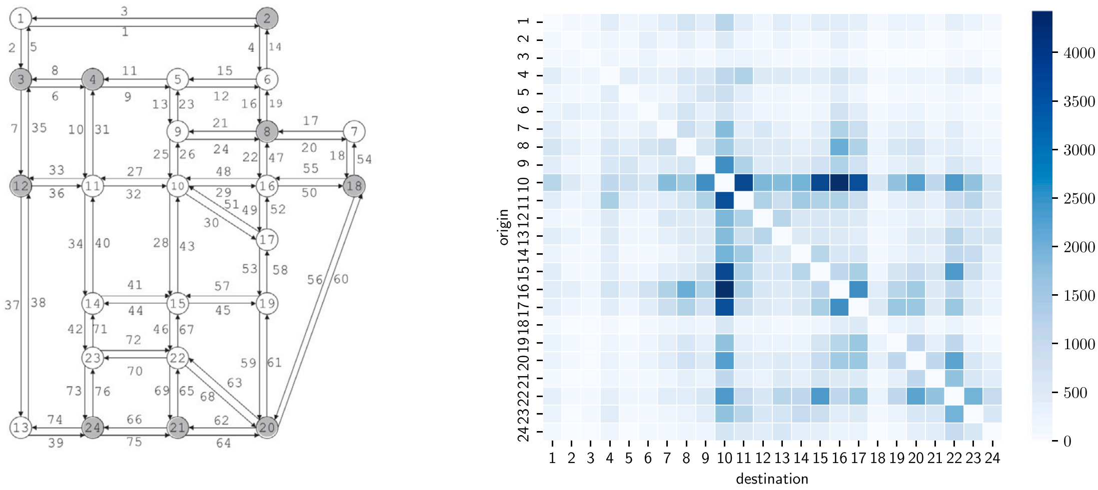
Remark 9. A key consideration for hypothesis testing concerns the identifiability of the parameters under the null hypothesis (Gallant, 1975a). Assume the travelers' utility function is $v = \theta_t t^{\alpha}$ where θ_t, α are the estimated coefficients and t is the travel time attribute. If one tests the hypothesis $\theta_t = 0$, α is not identifiable. This violates the Assumption 10, Section 3.1 on the rank condition that is required to prove the consistency and asymptotic normality of the NLLS estimator. Conversely, under a homogeneous linear-in-parameters utility function (Assumption 3, Section 3.1), the identification of the parameters on a restricted model can be guaranteed under mild conditions (Appendix D.2).

7. Numerical experiments

To study the performance of the proposed algorithm and some of the mathematical properties of the optimization problem reviewed in Section 4, we conduct experiments with synthetic data generated from four small networks (Appendix D.3). Subsequently, we study the statistical properties of the non-linear least squares (NLLS) estimators of the utility function coefficients via a series of Monte Carlo experiments conducted in the Sioux Falls, SD network. We employ a validation framework where a hypothetical O-D matrix and a vector of the utility function coefficients are assumed as the ground truth and then used to generate synthetic traffic count measurements consistent with SUELOGIT. In line with the LUE literature, the utility function coefficients are estimated under the assumption of an exogenous O-D matrix. All the experiments in this section are conducted on a MacBook Pro with Intel Core i5 CPU 2.7 GHz × 2, 1867 MHz 8 GB RAM, 256 GB SSD. The running time of the experiments is 18 h.

7.1. A medium-scale network: Sioux falls

The Sioux Falls network is a standard testing bed used by transportation researchers in network modeling studies (Shen and Wynter, 2012). The network comprises 24 nodes and 76 links (Fig. 1(a)). We generate 1,584 paths corresponding to the three paths with the shortest distance to travel between each O-D pair. The O-D matrix is obtained from TNTP (2021), and a heatmap with its values is presented in Fig. 1(b). The setup to generate the synthetic data is the same as for the small networks (Appendix D.3) except that the exogenous attributes for the monetary cost c and the number of intersections s at each link/path are incorporated in the utility function. The attributes c and s are generated as continuous and discrete random variables, respectively, and the utility function coefficients are set to $\theta_t = -1,\ \theta_c = -6,\ \theta_s = -3$. If the units of the travel time and monetary cost attributes are minutes and US dollars, the value of time (VOT) becomes equal to 10 USD$ per hour. By Assumption 3, Section 3.1, the set of coefficients $\theta_t, \theta_c, \theta_s \in \mathbb{R}$ are common among travelers and linearly weighting the attributes t, c, s in the utility function. The true path set is assumed to be known, and thus, the experiments do not perform the column generation phase of the inner-level optimization algorithm (Step 1, Algorithm 3, Appendix C.2)



(a) Network topology

(b) O-D matrix

Fig. 1. Topology and O-D matrix of Sioux Falls network.

7.2. Data generating process (DGP)

The data generating process (DGP) to conduct the experiments assumes the modeler has perfect information about the link performance functions, the network topology, the path sets, and the ground truth coefficients and attributes of the travelers' utility function. For each network and replicate of the experiment, we generate a set of synthetic traffic counts under an ideal scenario where link flows perfectly match `SUELOGIT`. To introduce randomness, we add *i.i.d* errors to each set of traffic count measurements. This *i.i.d* sampling strategy (see Remark 14) ensures that Assumptions 7 and 8, Appendix B.1 are satisfied and thus, the statistical inference for the `NLLS` estimator remains consistent. For convenience, we choose Gaussian errors, but other choices of the probability distribution are also reasonable, including the Poisson or Negative Binomial distributions. The Gaussian errors are generated with a standard deviation equal to 10% of the average value of the traffic counts. As a reference, the standard deviation of the error term used in the Monte Carlo experiments conducted by Gallant (1975a) was equal to the 3% of the average value of the dependent variable reported in his synthetic dataset.

7.3. Pseudo-convexity of the objective function

To have a detailed characterization of the coordinate-wise pseudo-convexity of the outer level optimization problem, Fig. 2 shows plots of the objective function ℓ (top left), the first derivative of ℓ (top right), the sign of the first derivative of ℓ (bottom left) and the sign of the second derivative of ℓ (bottom right) respect to a coefficient of the utility function. The vectorized expressions to compute the first and second derivatives of the objective function are presented in Appendices C.3.2 and C.3.3. To avoid an excess of overlap of multiple curves when included in the same figure, we only analyze the curves associated with the travel time and cost coefficients. The true values of the coefficients are represented with vertical dashed lines in each figure, and the curve associated with each coefficient was generated by fixing the value of the other coefficient to its true value. To satisfy the exogeneity assumption required for coordinate-wise pseudo-convexity in the objective function (Section 4.1), travel times are assumed known and equal to the travel times obtained when generating the synthetic traffic counts at `SUELOGIT`.

Similar to the results obtained for the small networks (Appendix D.3.1), we observe that the objective function is locally convex and minimized at a point close to $\theta_t = -1, \theta_c = -6$. In addition, the sign of the first derivatives always points toward the global optima, which illustrates this fundamental property of pseudo-convex functions. In contrast to the results obtained in small networks, the sign of the second derivative changes more frequently. The latter may be associated with the higher complexity in larger networks which tends to increase the non-linearity of the traffic flow functions and hence, of the objective function, with respect to the utility function coefficients.

7.4. Convergence with a multi-attribute utility function

The coordinate-wise pseudo-convexity of the outer level objective function provides theoretical guarantees for the convergence of `NGD` toward global optima but only when the utility function depends on a single exogenous attribute (Section 5.4.1). Interestingly, our experimental results suggest the ground truth coefficients of a multi-attribute utility function can also be recovered in the endogenous case. The top and bottom plots in Fig. 3 show the values of the ratio between the travel time and cost coefficients, i.e., the

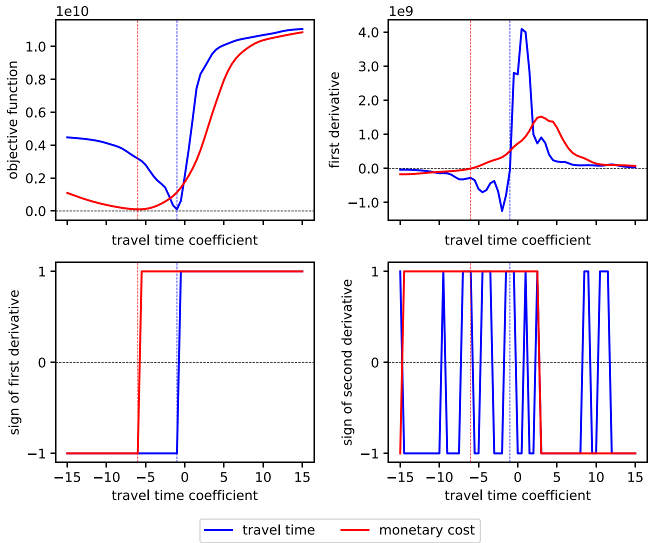


Fig. 2. Pseudoconvexity of the objective function in Sioux Falls network.

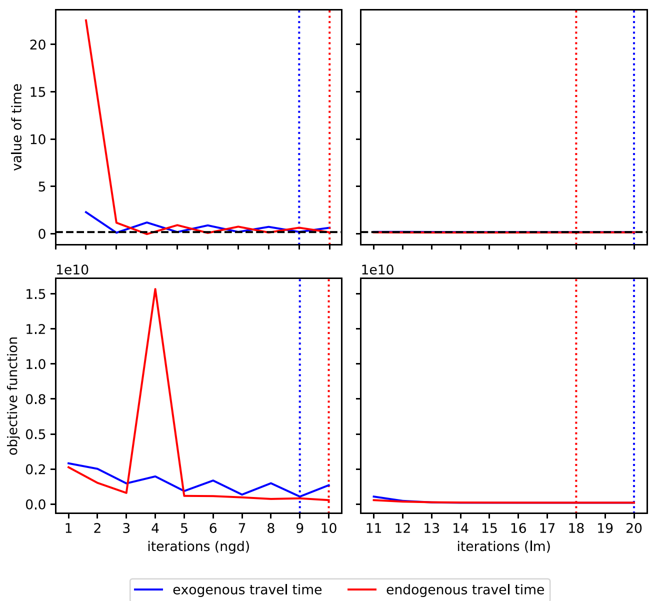
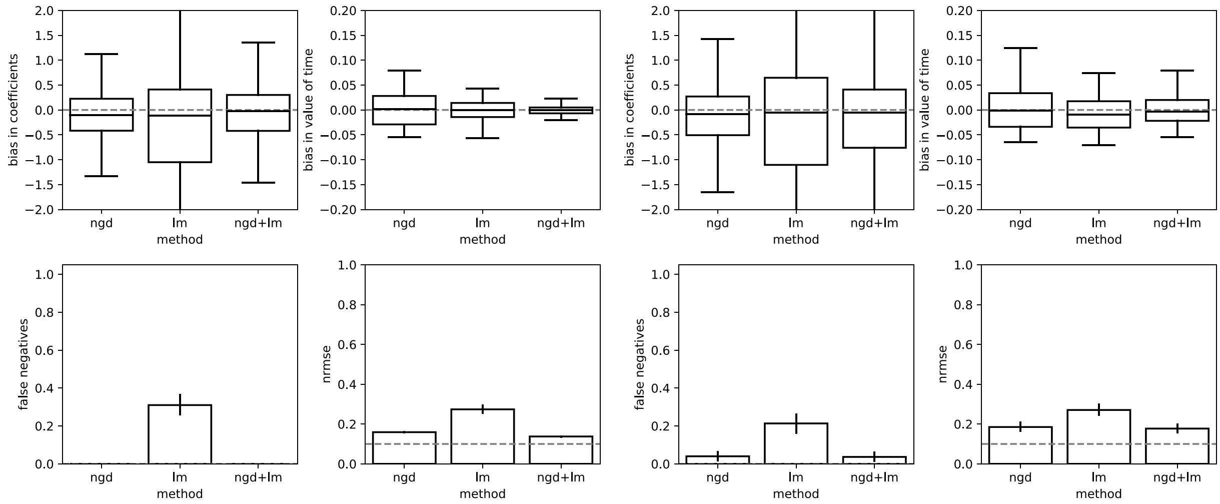


Fig. 3. Convergence to the ground truth value of time in Sioux Falls network.

value of time, and the value of the objective function over iterations, respectively. The non-refined and refined optimization stages performed ten iterations of NGD and LM, respectively. The starting points for optimization of the utility function coefficients were set to zero. In the exogenous and endogenous cases, we observe that the value of time is perfectly recovered and that the value of the objective function is close to zero. In the exogenous case, travel times are assumed to be known, and thus, each iteration reduces to minimize the outer level objective function only. In the endogenous case, the convergence is less stable due to the additional computation of SUELOGIT at each iteration of the bilevel optimization.

7.5. Impact of endogeneity of travel times on statistical inference

The discrete choice modeling community often uses Monte Carlo experiments to empirically study the properties of estimators in travel behavior models (Train, 2003). We conduct 100 replicates of two Monte Carlo experiments to study the impact of the endogeneity of travel times on the statistical inference of the utility function coefficients. We analyze the impact on statistical



(a) Exogenous travel times

(b) Endogenous travel times

Fig. 4. Consistency in coefficients recovery.

inference of three different setups of the optimization algorithm; (i) NGD is used in the no-refined stage, (ii) LM is used in the no-refined stage, (iii) NGD and LM are used in the no-refined and refined stages, respectively. In scenarios (i) and (ii), no iterations of the optimization methods are performed in the refined stage. The significance level for all hypotheses tests is set at $\alpha = 0.1$.

Each replicate of the experiment draws a new sample of Gaussian distributed errors (Section 7.2) and hence of new traffic counts. For each replicate, we compute the bias of the utility function coefficients and of the value of time, the normalized root mean squared error (NRMSE), and the amount of false negatives. The NRMSE is defined as the ratio between the root mean squared error (RMSE) and the mean of the observed traffic counts. Because the standard deviation σ of the random error is defined as a proportion of the mean of true traffic counts (Section 7.2), the NRMSE should be close to $\sigma = 0.1$. Since the ground truth coefficients of the utility function are set to values different than zero, a false negative equates to a non-rejection of the null hypothesis $H_0 = 0$. The initial values for optimization of the utility function coefficients are set with a uniform distribution centered at the true values of the coefficients and with a width of two. Thus, the initial value of the travel time coefficient θ_t is randomly chosen within the interval $[-3, 1]$.

The experiment results show that the estimates were roughly unbiased in all optimization setups and under exogenous and endogenous travel times. This result is interesting because the NLLS assumptions prove only consistency but not unbiasedness of the NLLS estimators in large samples. A potential explanation is related to the pseudo-convexity of the outer-level objective and the existence of a unique global optimum in the neighborhood around the initial values for optimization. Furthermore, our evidence strongly supports the use of NGD + LM over the standalone application of first or second-order optimization methods. The higher standard error of the estimates obtained with LM unveils the problems of numerical stability of second-order methods when the initial values for optimization are in regions where the objective function is not convex. The latter is also associated with the large proportion of false negatives reported by LM in both experimental scenarios. In contrast, the standalone application of NGD reports less than 5% of false negatives in both scenarios, which equates to a statistical power of 95%. This is an encouraging result given the reference threshold of 80% typically used for experimental studies.

In the endogenous case, we observe an increase of the difference between the NRMSE and its expected value, given by the dashed bar shown in the bottom right plots of Figs. 4(a) and 4(b). Similarly, the bias of the estimated coefficients and the value of time increases consistently in the three setups of optimization methods compared to the exogenous case. The computation of equilibria in the endogenous case may result in inaccurate approximations of the travel times at equilibria. From an econometric standpoint, this introduces measurement error in the travel time attribute, and it can add bias to the parameter estimates. Another problem in the endogenous case is the significant increase in computational burden due to solving traffic equilibria at each replicate of the experiment.

To ensure good convergence over replicates and that the assumptions for consistency of the NLLS estimator are satisfied, the following experiments in this section are performed under the scenario of exogenous travel times only. The initial values for optimization of the utility function coefficients are also set with a uniform distribution centered at the true values of the coefficients and with a width of two. This choice of width allows the optimization to start in a region close to the ground truth values of the utility function coefficients and where the objective function is, at best, convex or, at worst, pseudo-convex. A tighter width makes it difficult to assess the reduction of the objective function with different setups of the optimization method. The significance level for all hypotheses tests is also set at $\alpha = 0.1$.

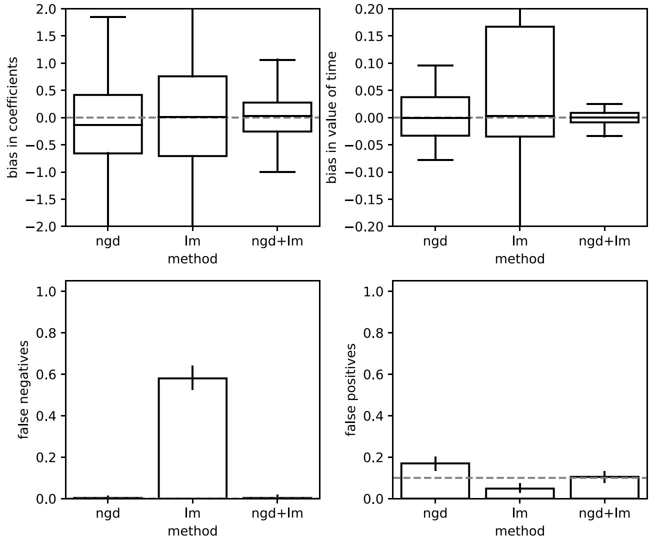


Fig. 5. Impact of irrelevant attributes on statistical inference.

7.6. Irrelevant attributes

The NLLS theory suggests that the NLLS estimator is asymptotically consistent. Therefore, the coefficients associated with irrelevant attributes included in the utility function should converge to zero, and they should not impact the statistical inference on coefficients weighting relevant attributes. However, with finite samples, including irrelevant attributes in the utility function can reduce the efficiency of coefficients of relevant attributes. Consequently, the t-tests of other coefficients may decrease, which could artificially increase the number of false negatives, i.e., lower rejections of the null hypothesis for coefficients different than zero.

Fig. 5 shows the impact of the inclusion of irrelevant attributes on the bias of the estimated coefficients and of the value of time and on hypothesis testing. In contrast to the Monte Carlo experiments conducted in Section 7.5, the utility function used to generate synthetic traffic counts includes six irrelevant attributes. Each irrelevant attribute is generated as a standard Gaussian random variable and weighted by a coefficient equal to zero in the utility function. Hence, the hypothesis tests are not expected to reject the null for these coefficients. To quantify the power of the hypothesis tests to detect irrelevant attributes, we compute the proportion of false positives, namely, the proportion of times that the null hypotheses for coefficients weighting the set of irrelevant attributes are rejected. By classic statistical theory, p-values are uniformly distributed under the null hypothesis and, hence, a significance level of $\alpha = 0.1$ should result in a 10% of false positives. To analyze the impact of the inclusion of irrelevant attributes on the statistical inference of relevant attributes, we also report the proportion of false negatives.

Similar to the results obtained in Section 7.5, integrating NGD and LM significantly reduces the number of false negatives and the bias of the estimated coefficients and of the value of time. Including irrelevant attributes in the utility function increases the amount of false negatives but only in the case of LM. Notably, the integration of NGD and LM perfectly matches the theoretical 10% level of false positives that are expected at a significance level of $\alpha = 0.1$. Overall, these results reaffirm that statistical inference is more reliable when first-order and second-order optimization methods are integrated. Given this evidence, the following experiments are conducted using NGD + LM only.

7.7. Sensor coverage

A lower sensor coverage reduces the sample size, which is expected to increase the standard error in the estimates of the utility function coefficients. Consequently, t-tests may be overestimated, and thus, the amount of false negatives may increase. Fig. 6 shows the impact on hypothesis testing when varying the link coverage at three levels; 25%, 50%, and 75%. As expected, the amount of false negatives and the standard error of the coefficients decrease with the level of sensor coverage. Notably, for coverages of only 50% (N = 38), the statistical power already surpasses the desired threshold of 80%. For all coverage levels, the proportion of false positives does not surpass the theoretical level of 10%, and the NRMSE closely matches its expected value at 10%.

7.8. Error in traffic count measurements

The standard error of the NLLS estimators is expected to increase when the variance of the random perturbation is higher. Thus, hypothesis testing should tend to reject less frequently the null for coefficients associated with relevant attributes, and hence, the amount of false negatives could artificially increase. Fig. 7 shows the impact of increasing the standard deviation of the Gaussian perturbation on statistical inference. The standard deviation is defined as a percentage of the average value of the synthetic traffic

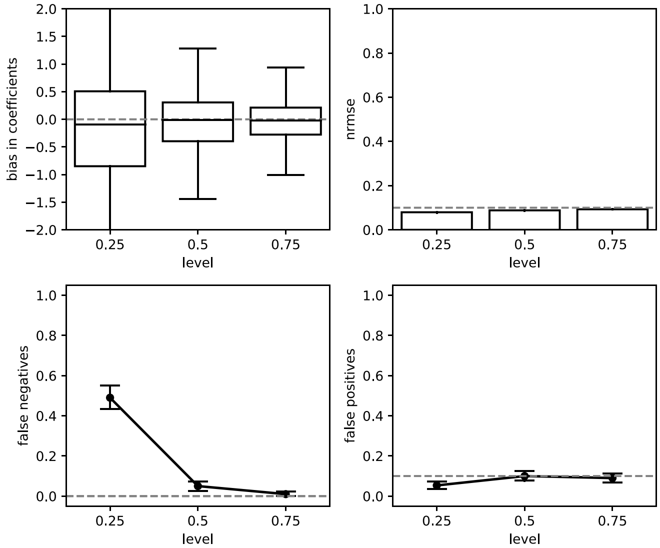


Fig. 6. Impact of sensor coverage on statistical inference.

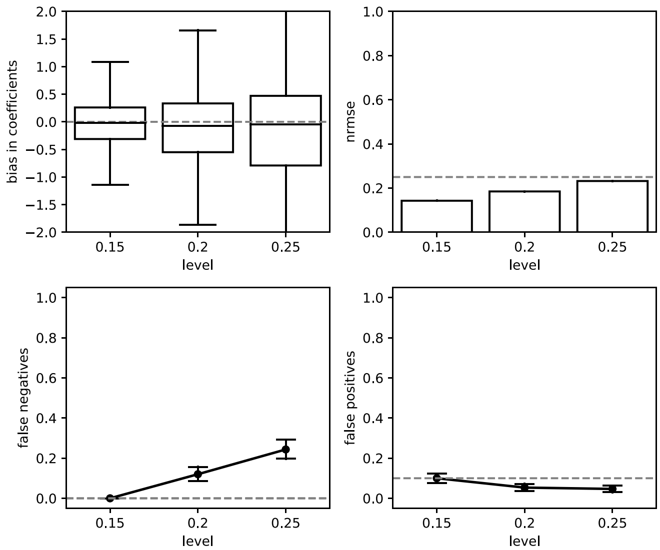


Fig. 7. Impact of noise in traffic counts on statistical inference.

counts (see Section 7.2). As expected, the amount of false negatives and the standard error of the estimated coefficients increase with the level of variability in the random perturbation. The values of the NRMSE closely matches the error levels, and the proportion of false positives does not surpass the theoretical level of 10%. In real applications and in line with the sensor coverage experiment (Fig. 6), a larger sample size should compensate for an increase in the variance of the random perturbation.

7.9. Non-deterministic O–D matrix

Our methodology assumes that the O-D matrix is exogenous and deterministic (Assumption 2, Section 3.1). While it is feasible to have an accurate reference O-D matrix available, its cells may be subject to random perturbations. Any gap between the true and reference O-D matrix may add measurement error in the model, which could bias or distort the t-tests of the coefficients. Fig. 8 shows the impact of introducing a different noise level in the O-D matrix. The reference O-D matrix is assumed to be a Gaussian random variable with expected value $\boldsymbol{\mu}_{\mathbf{Q}} \in \mathbb{R}_{\geq 0}^{|\mathcal{V}|\times|\mathcal{V}|}$ equal to the values of the true O-D matrix $\mathbf{Q}$ and with a uniform standard deviation $\sigma_{\mathbf{Q}} \in \mathbb{R}_{+}$ defined as a percentage of the average value of $\mathbf{Q}$. To compare the impact of adding noise in either the traffic counts (Section 7.8) or the O-D matrix, we set the same levels of the standard deviation of the random perturbation in both experiments.

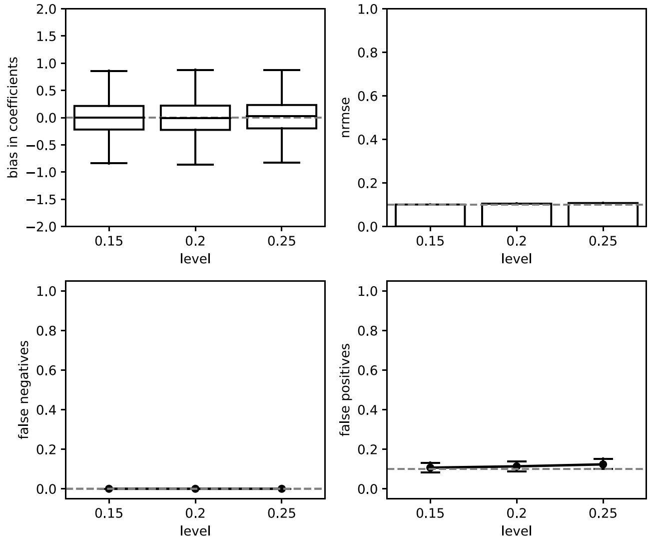


Fig. 8. Impact of noise in O-D matrix on statistical inference.

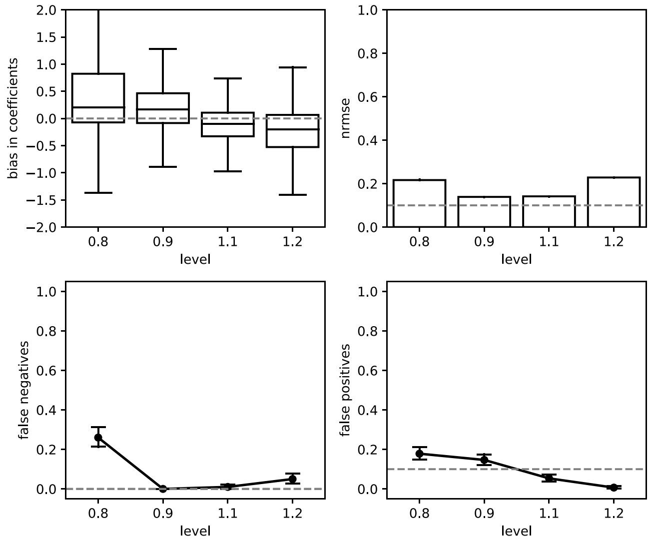


Fig. 9. Impact of a bad scaled O-D matrix on statistical inference.

Surprisingly, our results show that the noise in the O-D matrix does not significantly impact statistical inference. Therefore, the standard error of the estimated coefficients and the NRMSE remains invariant for all levels of random noise.

7.10. Ill scaled O–D matrix

An ill-scaled O-D matrix is expected to increase the gap between predicted and observed traffic counts, which may induce an overestimation of the variance of the random error. As a consequence, the t-tests and the false negatives could increase. Fig. 9 shows the impact of an ill-scaled O-D matrix. The level represents the true value of the scale or factor that should multiply the reference O-D matrix to recover the true O-D matrix. Therefore, values lower or higher than 1 represent cases of overestimation or underestimation of the true O-D matrix, respectively. Compared to the previous experiments that assume a well-scaled O-D matrix, we observe that the NRMSE is far from its expected value, which suggests that there is no good convergence toward the ground truth values of the utility function coefficients. The estimated coefficients are biased upwards and downwards when the true O-D matrix is overestimated or underestimated, respectively. The latter is directly associated with the underestimation and overestimation of the false positives with respect to the theoretical level of 10%. The false negatives surpass the 20% when the true O-D matrix is underestimated, which significantly reduces the statistical power to detect the effect of relevant attributes in the utility function.

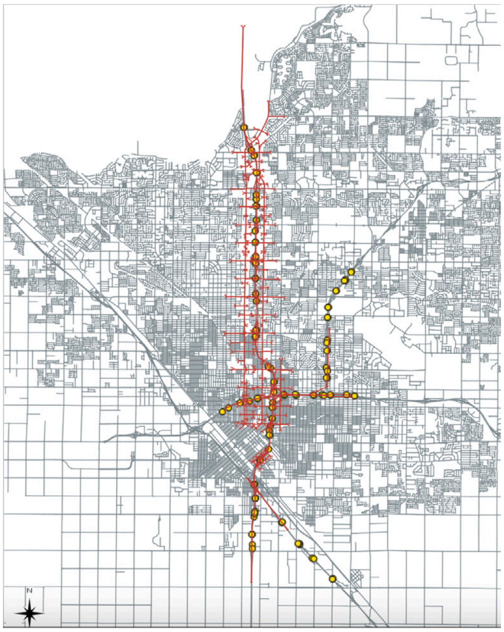


Fig. 10. SR-R1 corridor in Fresno, CA.

8. A large-scale network with multiple attributes: California SR-41 corridor

The proposed methodology was also applied to a large-scale transportation network in the City of Fresno, California (Fig. 10). This network primarily covers major roads and highways around the SR-41 corridor and comprises 1789 nodes and 2413 links (Ma and Qian, 2017, 2018a). The following sections describe the use of various data sources to estimate our model in the Fresno network. Subsequently, we describe the attributes of the utility function, the models' specifications, the estimation procedure, and the results.

8.1. Traffic counts

The Caltrans Performance Measurement System (PeMS, 2021) provides open-source data with georeferenced stations that record information on traffic counts, speeds, and travel times on major highways in California. Using geoprocessing tools, 141 links of the Fresno network were matched to PeMS stations, which equates to a 5.8% of sensor coverage. We only use traffic count data collected during the first Tuesday of October 2019 and October 2020 between 4pm and 5pm, which correspond to periods before and during COVID-19. The average traffic counts for the selected periods in 2019 and 2020 are 2213.6 and 2113.3 vehicles per hour, respectively.

8.2. Travel demand

A dynamic O-D demand matrix calibrated with support of the PARAMICS estimator tool for the SR-R1 corridor is adopted in our study (Ma and Qian, 2017; Zhang et al., 2008; Liu et al., 2006). This O-D demand is provided with a 15-minute resolution for the period between 4pm and 6pm on a typical weekday. To have consistency in the temporal resolutions of the O-D demand and the traffic counts data, both data sources are aggregated in a one-hour window starting at 4pm. The aggregation in vehicles per hour is also consistent with the temporal resolution of the link performance functions.

Fig. 11 show the cumulative demand with respect to the number of O-D pairs. There are 6970 O-D pairs that reported trips, which gives a total of 66,266.3 trips. The dashed bars in the figures suggest that approximately 30% of the O-D pairs with the highest demand cover 85% of all trips. To account for the reduction in travel demand in the selected periods of 2019 and 2020, we scale the O-D matrix in 2019 by a constant factor of 0.9546, which is equal to the ratio between the average traffic counts in each year (2113.3/2213.6). Note that this factor is similar but more conservative than the ratio between the traffic volumes of October 2020 and October 2019 in all roads and streets in the United States (FHA, 2020).

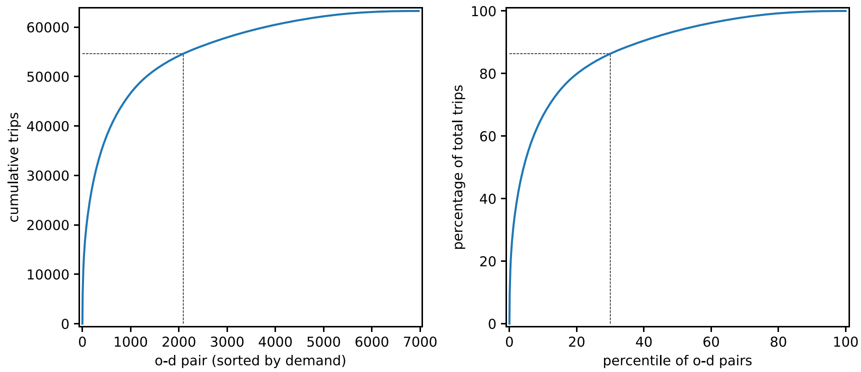


Fig. 11. Cumulative travel demand in Fresno, CA.

8.3. System level data

Below are the sources of system-level data used to compute the attributes of the travelers' utility function in every link of the Fresno network:

- Speed and travel times
  INRIX traffic speed data have been successfully used in previous transportation studies, and it is considered a reliable data source for our analyses (Yao and Qian, 2020; INRIX, 2021). Our raw data includes a shapefile with a collection of georeferenced line segments representing the major roads in Fresno, CA. It also contains `csv` files with traffic speeds for every line segment and a 5 min resolution. Attributes in the dataset include a unique identifier for each line segment, a time stamp, observed speed, average speed, and reference speed. Using the length of each line segment, the speed attributes were transformed into travel time attributes in minute units. The line segments of the INRIX shapefile and the links of the Fresno network are matched according to the orientation and proximity of the streets. For the 15.5% links that were not successfully matched, we imputed the mean value of the attributes that were matched in the remaining links.
- Traffic incidents
  The Statewide Integrated Traffic Records System (SWITRS) is one of the two official sources of traffic incidents data in California (Waetjen and Shilling, 2021). SWITRS is managed by the California Highway Patrol (CHP) and includes post-processed and georeferenced data on incidents causing human injury or death. A codebook of the dataset and the incidents reported between 2016 and 2021 are publicly available in SWITRS (2021) and Kaggle (2021). A buffer of 50 ft is used to match the links in the Fresno network with the location of the incidents. In 2019 and 2020, 1601 and 2020 incidents are matched to 276 and 382 links of the Fresno network, respectively.
- Streets characteristics
  Street characteristics are hypothesized to be a predictor of travelers' route choices. Using an open repository maintained by the GIS Division City of Fresno (2021), we processed shapefiles with the georeferenced positions of bus stops and street intersections in Fresno. A buffer of 50 ft is used to match the links in the Fresno network with the location of the stops and intersections. 362 and 2115 bus stops and street intersections are matched to 300 and 1139 links, respectively.
- Socio-demographics from Census data
  A shapefile with the most recent sociodemographic data at the block level was retrieved from US Census Bureau (2019). We only use the layers with data about the population proportions by gender, age, and income level at each block. There are other layers with relevant features to explain route choice preference, such as one related to commuting patterns. However, this data is unavailable for the city of Fresno or only at the tract and county levels.

8.4. Attributes of the utility function

Travel time is the only endogenous attribute in the utility function, and it is equal to the output of the links' performance functions. The free flow travel times defined in the link performance functions are obtained from INRIX data. The speed attribute used to compute free flow travel times corresponds to the speed driven on a road when it is wide open and does not necessarily match the legal speed limit. Using the system-level data described in Section 8.3, we also compute the following set of exogenous attributes for every link of the Fresno network:

- *Std. Speed*: Standard deviation of speed [miles/hour]
- *Incidents*: Total incidents in the current year
- *Intersections*: Number of streets intersections
- *Bus stops*: Number of bus stops

- *Median income*: Median household income in the U.S. Census block [USD/year]

Std. Speed is computed using historical data provided by INRIX for the period between 4 pm and 5 pm on a weekday, and it is assumed to negatively impact the utility of choosing a path. There are alternative ways to capture the effect of travel time variability on travelers' route choices, such as the standard deviation or coefficient of variation of travel time. The main advantage of the standard deviation of speed is that it measures time variability independent of the length of the link segments. The rationale for aggregating the number of incidents in the current year is to have a proxy of the drivers' perception of the road safety of a link segment. *Median income* is used as a proxy of the perception of the level of crime in surrounding neighborhoods, and it is hypothesized to be positively associated with the utility of choosing a path. A larger number of bus stops or street intersections in a link segment should encourage drivers to use alternative paths. Thus, all attributes, except for *Median income*, are expected to have a negative coefficient in the utility function.

To test the non-linear effects of the attributes on the utility function, we binarized the exogenous attributes as follows. *Reliable speed* and *Low income* take the value 1 if the values of *Std. speed* and *Median income* are below the 25th percentile of their distributions, respectively, and 0 otherwise. *Bus stop*, *Intersection* and *Incident* takes the value one if *Bus stops*, *Intersections* and *Incidents* are greater than zero, respectively, and 0 otherwise. Under this new specification of the attributes, all coefficients of the utility function, except for the coefficient weighting *Reliable speed*, are expected to be negative.

Note that the network contains 656 links (38.8%) that are centroid connectors. These connectors usually do not have physical counterparts in the real network (Qian and Zhang, 2012) and should not impact travelers' route choices. Thus, we set the values of the attributes in these links to zero, regardless if the attributes are continuous or binary. The path size correction factor is set to its default value of one in all model specifications.

8.5. Descriptive statistics

Fig. 12 shows the correlations among traffic counts and system-level attributes of the Fresno network during the first Tuesdays of October 2019 and 2020. We include free flow speed instead of free flow travel times which normalizes the attribute by the link length. The Pearson correlations shown in the top row of the plots indicate that the free flow speed is negatively correlated with the number of intersections and bus stops in the two selected periods. Also, as expected, the number of incidents is positively associated with the free flow speed. An interesting result is a negative correlation between median income and free flow speed. The latter is probably explained by the higher rate of motorization in high-income areas, which should be associated with lower speeds. The standard deviation of speed positively correlates with free flow speed but only during 2020. The sign of correlation between traffic flows and free flow speed is unclear. Overall, the correlational analysis shows sensitive associations between the attributes of the utility function and suggests that the spatial matching of attributes conducted in Section 8.3 was reasonably accurate.

8.6. Model specifications

We estimate three model specifications using the data collected on each date. Following the state of the practice in previous literature, we first estimate a *Baseline model* including *travel time* only. We also estimate a *Full model* including the exogenous attributes in a continuous scale; *Std. Speed*, *Median income*, *Incidents*, *Intersections*, *Bus stops*. Finally, a *Binarized model* includes the exogenous attributes in a binary scale; *Reliable speed*, *Low income*, *Incident*, *Intersection*, *Bus stop* (Section 8.4). To assess the gains in the explanatory and predictive power of the *Full model* and *Binarized model* compared to the *Baseline model*, these models also include *Travel time*. By assumption, the utility function is defined as a linear weight between attributes and their respective coefficients. *Travel time*, the only endogenous attribute in all models, is iteratively updated during the bilevel optimization.

8.7. Estimation

To prevent that numerical issues associated with the scale of the attributes impact the quality of the statistical inference, we normalize each exogenous attribute, including the free flow travel time defined in the link performance functions, by their maximum values. Because all these attributes are non-negative, their new ranges are between 0 and 1. Figure Fig. 5, Appendix E.1 shows descriptive statistics of the attributes after this normalization.

The models are estimated on Amazon Web Services on two `t2.2xlarge` instances equipped with Intel Xeon CPU 3.3 GHz x 8, 32 GB RAM (AWS, 2022). The three model specifications are run in parallel using the data collected in either 2019 or 2020. Because `SUELOGIT` is path-based, the computational and memory complexity is a function of the number of paths. The CPU processor and RAM of the `t2.2xlarge` instance allow handling 33,430 paths. In contrast, a `t2.xlarge` instance with Intel Xeon CPU 3.3 GHz x 4, 16 GB RAM, handles 22,015 paths without reporting memory overflow. The memory bottleneck of the algorithm is on the computation of the analytical gradients performed to update the utility function coefficients during the outer-level optimization. The computation time of each notebook is about six hours.

All models are estimated using `NGD` in the no-refined stage with a learning rate of $\eta_1 = 0.5$ and using `LM` in the refined stage, respectively. Each stage performs ten iterations of the optimization methods. The starting points for optimization in the non-refined stage are set to zero for the three model specifications. The best estimate of the utility function coefficients in the non-refined stage is chosen as the starting point for optimization in the refined stage. The coefficients are projected to zero during the iterations of the bilevel optimization if their sign is inconsistent with our expectation (Section 8.4). If the coefficients of the attributes are zero at the end of the non-refined stage, they are excluded from the refined stage. This strategy is used to reduce computational burden and accelerate the convergence of the optimization algorithm to a local optimum.

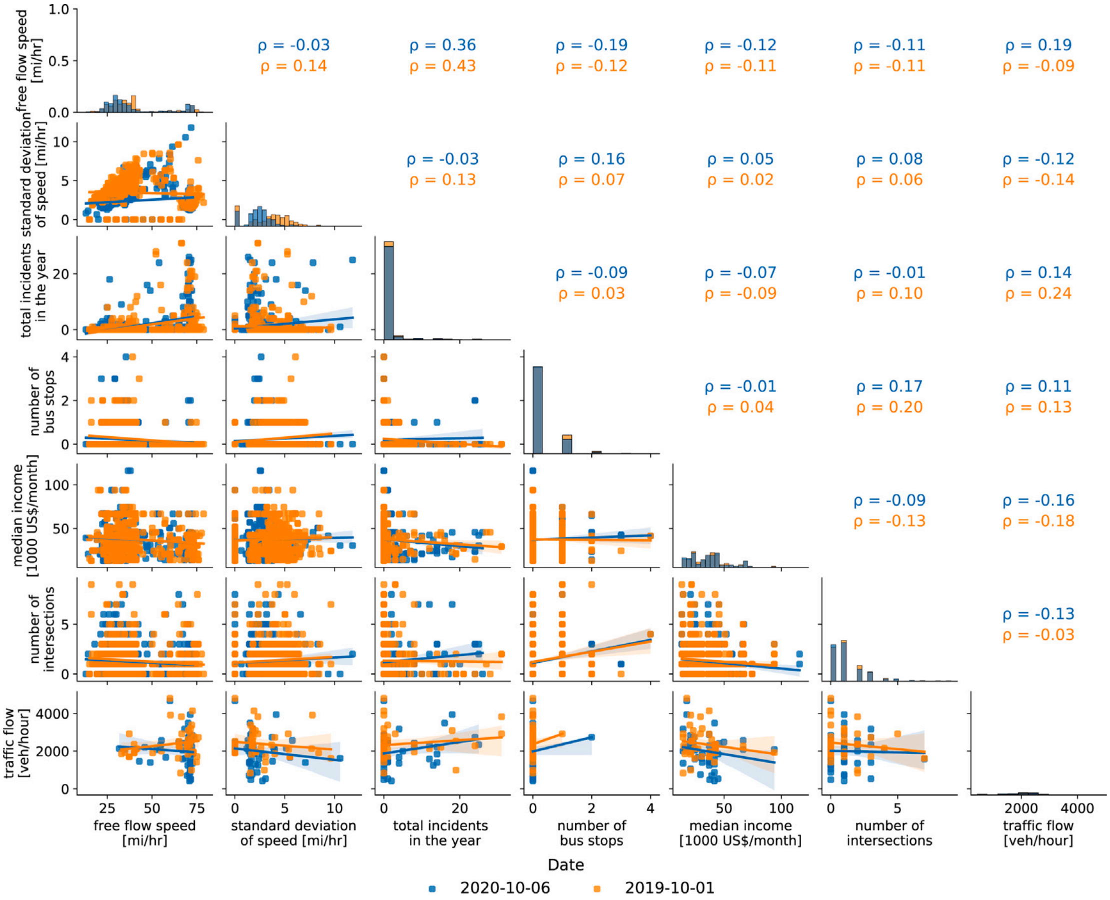


Fig. 12. Correlations among system-level attributes in Fresno, CA.

8.8. Path generation

We generate an initial set of 13,905 paths using the two shortest paths between the 6970 O-D pairs that report trips (Fig. 11, Section 8.2). The initial path set of the three model specifications is the same. The total number of paths is odd because some O-D pairs are only connected by a single path. The column generation method described in Step 1, Algorithm 3, Appendix C.2 is used to update the path sets during the bilevel optimization. To guide the paths' exploration, new paths are generated in the 30% of O-D pairs with the highest demand only, which covers approximately 85% of the total trips in the Fresno network (Fig. 11, Section 8.2). Based on the current estimate of the utility function coefficients and the link attributes, the column generation step generates the 40 shortest paths among 3% of the OD pairs at each iteration of the bilevel optimization. The selected O-D pairs at each iteration change sequentially according to their level of demand, and the process continues until the target of 30% of O-D pairs is reached in the last iteration. The sequential selection of O-D pairs reduces the computational burden of generating new paths in all O-D pairs and at every iteration without compromising path exploration significantly. Subsequently, the inner-level optimization algorithm solves SUELOGIT, and the path utilities are updated according to the new travel times at equilibria. Finally, it selects the ten paths with the highest utility for every O-D pair.

8.9. Indicators for model comparison

For each model specification and date, we compute the value of the objective function (i.e., the sum of squared errors), the root mean squared error (RMSE), the normalized RMSE (NRMSE), the F-test (Section 6.4) and the adjusted pseudo R^2. The F-test compares the sum of squared errors of the model with its null version, where all coefficients of the utility function were set to zero, and that represents a scenario where travelers make equilikely choices among paths. F-tests with p-values lower than α reject the null hypothesis that two models are statistically equivalent at the $1-\alpha$ % confidence level. Hence, this is evidence that supports the selection of the augmented model.

Our adjusted pseudo R^2 is analogous to the McFadden Pseudo-R^2 (McFadden, 1973) and is adjusted by the number of model coefficients (Ben-Akiva et al., 1985). It is defined as one minus the ratio between the SSE of a model minus the number of coefficients

Table 1
Point estimates and summary statistics of models fitted with data collected between 4pm and 5pm during the first Tuesdays of October 2019 and October 2020 in Fresno, CA.

Attribute (t-test)	First Tuesday of October 2019 (Before COVID-19)			First Tuesday of October 2020 (During COVID-19)		
	Baseline model	Full model	Binarized model	Baseline model	Full model	Binarized model
Travel time (θ_t)	−2.000** (−2.5)	−1.904** (−2.3)	−0.760*** (−3.1)	−2.500** (−2.1)	−1.889* (−1.9)	−0.458* (−1.8)
Std. speed	–	0.000 (0.0)	–	–	0.000 (0.0)	–
Incidents	–	−2.409** (−2.2)	–	–	−2.597 (−1.3)	–
Median income	–	0.864** (2.0)	–	–	0.447 (0.8)	–
Intersections	–	0.000 (0.0)	–	–	−0.708 (−0.7)	–
Bus stops	–	0.000 (0.0)	–	–	0.000 (0.0)	–
Reliable speed	–	–	0.000 (0.0)	–	–	0.000 (0.0)
Incident	–	–	−0.118 (−0.7)	–	–	−0.249 (−1.5)
Low income	–	–	0.000 (0.0)	–	–	0.000 (0.0)
Intersection	–	–	0.000 (0.0)	–	–	0.000 (0.0)
Bus stop	–	–	0.000 (0.0)	–	–	0.000 (0.0)
Obs. (coverage)	141 (5.8%)	141 (5.8%)	141 (5.8%)	141 (5.8%)	141 (5.8%)	141 (5.8%)
Initial objective	278,924,807	278,924,807	278,924,807	273,069,736	273,069,736	273,069,736
Final objective	192,568,627	168,917,573	186,985,714	188,694,982	158,156,498	178,166,886
RMSE	1168.6	1094.5	1151.6	1094.5	1059.1	1124.1
NRMSE	0.528	0.494	0.520	0.547	0.501	0.532
Adjusted pseudo R^2	0.311	0.395	0.331	0.316	0.422	0.345
F-test (p-value)	71.954*** (0.000)	11.224** (0.000)	8.047*** (0.000)	99.482*** (0.000)	4.157*** (0.002)	9.166*** (0.000)

Note: Significance levels: *p<0.1; **p<0.05; ***p<0.01.

and the SSE of the model with all coefficients set to zero. Since traffic counts are an aggregate of individual path choices, our model's adjusted pseudo R^2 is directly comparable to those obtained from discrete choice models. This indicator is an absolute measure of the predictive ability of discrete choice models, and it is helpful to generate benchmark values against which researchers can evaluate (Parady et al., 2021). It tends to be considerably lower than the R^2 index used in ordinary regression analysis, and values of 0.2 to 0.4 represent an excellent fit (McFadden, 1977).

8.10. Results

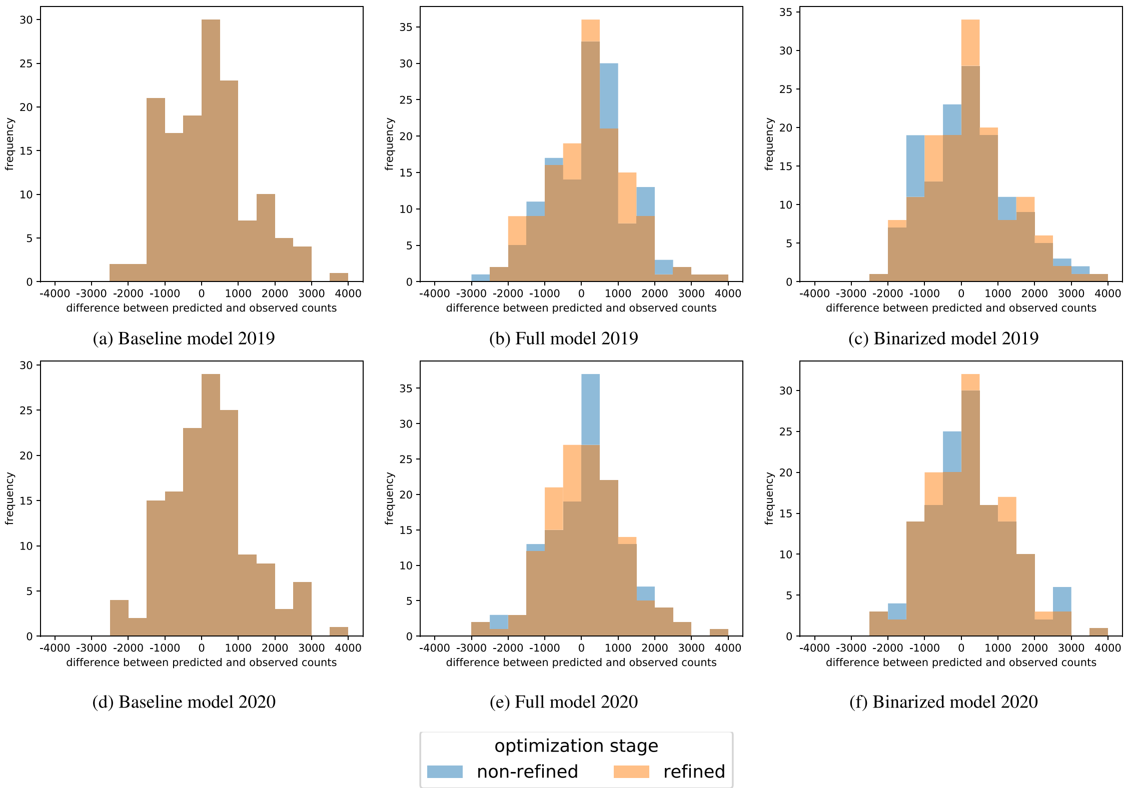
Table 1 shows the estimation results obtained with traffic count data collected during the first Tuesday of October 2019 (before COVID-19) and October 2020 (during COVID-19) and for three model specifications (Section 8.6). Besides the point estimates and t-tests of each coefficient of the utility function, the table also includes multiple indicators for model comparison (Section 8.9).

Fig. 13 shows histograms with the distribution of errors obtained after performing the non-refined and refined optimization stages. Figs. 21 and 22, Appendix E.2 show a comparison of the convergence of the full and binarized models against the baseline model. The top plots of the figures show the number of paths generated during column generation and added after performing the paths selection step during the inner-level optimization. The significant decrease in the number of paths added in the bilevel iterations of the refined stage suggests that the path exploration over the selected set of O-D pairs in the non-refined stage was reasonably exhaustive.

Notably, our results show that the adjusted pseudo R^2 obtained in all models are in the order of 0.3–0.56, which is considered an excellent fit in travel behavior studies (McFadden, 1977; Bovy et al., 2008; Bierlaire and Frejinger, 2008; Bwambale et al., 2019). The F-tests in the table suggest that all models are statistically different from the null model at the 99% confidence level. The F-tests comparing the *Full model* with the *Baseline model* are equal to 3.7804 and 4.3697 when using data collected before and during COVID, respectively. Because they are lower than their critical value of $F_{6-1,141-6,99\%}$, we conclude that the *Full model* is not statistically equivalent to the *Baseline model*. The indicators for model comparison of the *Binarized model* were lower than the *Full model* in all cases. Thus, the *Binarized model* is preferred over the *Full model*.

The NRMSE of the *Full model* is in the order of 0.5, meaning that the standard deviation of the traffic counts is approximately 50% of their sample mean. The experiments conducted in the Sioux Falls network find that an NRMSE of 0.25 results in more than 20% false negatives. Although the sample size in the Fresno network is two times higher than the number of links in the Sioux Falls network, it is unlikely to compensate for the difference in NRMSE. Other factors that could contribute to the decrease of the t-test of the coefficients in the *Full model* are associated with the high correlation of travel time with some exogenous attributes and also to numerical issues arising from the estimation of a richer utility function. Overall, this evidence anticipates a low statistical power to detect the effect of attributes that significantly impact travelers' utility in the Fresno network.

As expected, the travel time coefficient is significant at the 90% confidence level in all the models. Note that the estimated coefficients of travel times in the baseline models are exact multiples of the learning rate of NGD because there was no improvement of the objective function in the refined stage. In contrast, the estimations of the full and binarized models report improvements during the refined stage. The estimation of the *Full model* with the data collected before COVID-19 suggests that the total incidents during the year and the monthly household income of the neighborhoods nearby the link segments have a negative and positive effect on the travelers' utility, respectively. Both effects were significant at the 95% level of confidence. The estimation results of



(a) Baseline model 2019 (b) Full model 2019 (c) Binarized model 2019

(d) Baseline model 2020 (e) Full model 2020 (f) Binarized model 2020

Fig. 13. Distribution of errors in the non-refined and refined stages of the estimation of the baseline, full and binarized models.

the *Full model* with the data collected during COVID-19 also found a negative effect on the number of street intersections, but the coefficients of all attributes, except for travel time, were not significant at the 90% confidence level. As discussed earlier, the high level of noise in the Fresno network can explain the difficulty in detecting a significant effect of some attributes. Although it is not required in non-linear regression models, the distribution errors of the models resemble Gaussian distributions with zero means (Fig. 13). Finally, regarding the impact of COVID-19, we conclude that there are no significant differences in the attributes that determine travelers' route choices.

9. Conclusions

The network modeling community has studied the problem of estimating the travelers' utility function coefficients using traffic counts and travel time measurements. However, research has been limited to utility functions dependent on travel time only, and methods have been tested on networks of relatively small size. Under the assumption of a known exogenous O-D demand matrix, we enhance existing methods to estimate the coefficients of utility functions with multiple attributes by using traffic counts consistent with stochastic user equilibrium with logit assignment (`SUELOGIT`). We refer to this problem as Logit Utility Estimation (`LUE`). To perform attribute selection, we conduct hypothesis tests on the coefficients of a multi-attribute utility function. Furthermore, a rigorous analysis of the non-convexity and mathematical properties of the `LUE` problem is conducted to inform the design of our solution algorithm and to derive some theoretical guarantees about convergence toward local optima.

The realization of the pseudo-convex nature of the optimization problem motivates the use of normalized gradient descent (`NGD`), a first-order method developed in the machine learning community that is suitable for pseudo-convex optimization. The integration of `NGD` with the Levenberg–Marquardt (`LM`) algorithm outperforms the standalone application of second-order optimization methods used in previous literature in many ways. First, the estimates of the utility function coefficients become less sensitive to the starting points for optimization, which is a common issue in non-convex problems. Second, `NGD` improves the convergence toward global optima, making the statistical inference more reliable. Third, using first-order methods reduces computational cost because it only requires computing the gradient of the objective function with respect to the utility function coefficients. To our knowledge, this is the first time that a paper presents vectorized expressions to perform the gradient computation, which is critical to accelerating the optimization and scaling up our methodology to the largest transportation network studied to date within the `LUE` literature.

The analysis of the mathematical properties of the problem identifies the coordinate-wise monotonicity of the traffic flow functions as a primary driver of the coordinate-wise pseudo-convexity of the objective function. Experiments in networks used in previous studies show that the traffic flow functions are generally monotonic and that the objective function of the `LUE` problem is pseudo-convex with respect to the coefficient of a utility function dependent on travel time only. Results in the Sioux Falls network support the coordinate-wise pseudo-convexity of the objective function with respect to the coefficients of a multi-attribute utility function. A series of Monte Carlo experiments show that statistical inference on the utility function coefficients is robust to traffic congestion and different levels of sensor coverage and noises in the O-D matrix and traffic counts. The amount of false negatives and false positives obtained in these experiments are generally well-aligned with statistical theory. For instance, the amount of false positives perfectly matches the value expected for an arbitrary significance level. A higher amount of noise in the traffic counts increases false negatives, reducing the statistical power to identify the effect of relevant attributes in the utility function. Surprisingly, the statistical inference is resilient to high noise levels in the cells of the reference O-D matrix.

Our solution algorithm is deployed on a large-scale network in Fresno, CA. It gave reasonable results for the estimated values and the hypothesis tests of the utility function coefficients. The models report excellent goodness of fit based on standard metrics of model comparison used in travel behavior research. As expected, travel time is identified as a primary determinant of travelers' route choices in all model specifications. This partially supports the standard practice in the network modeling community of assuming utility functions dependent on travel time only. However, incorporating additional system-level attributes in the utility function significantly increases the goodness of fit of the baseline model that uses travel time only. These conclusions are robust when using data collected before and during COVID-19. Among the exogenous attributes, the total incidents during the year and the median income are the most relevant predictors of travelers' route choices. The coefficients of these attributes are significant at the 95% level of confidence but only when using data collected before COVID-19.

Implementing our methodology in a large-scale network is challenging mainly due to the high computational cost and the level of noise of real-world data. On the one hand, accounting for the interdependence between travelers' choices in transportation networks and the endogeneity of travel times requires computing traffic equilibria. On the other hand, estimating the travelers' utility function coefficients requires solving a regression problem. The computational complexity of the inner and outer level problems is a function of the number of paths. Thus, the alternating optimization of the problems becomes intractable with a large number of O-D pairs. We learned that using column generation methods for updating path sets in SUELOGIT provides a good compromise between computational cost and prediction error. The strategy of selecting the O-D pairs according to their demand level is also crucial to controlling the exploration of new paths over iterations of the bilevel optimization and identifying the paths that most decrease the prediction error. To our knowledge, this strategy is not integrated into previous column generation algorithms, and we strongly encourage its use in the further application of our methodology. To control for the effect of path correlation/overlapping, we correct path utilities according to the Path Size Logit model.

10. Further research

Further research could enhance our methodology to incorporate multi-day data and to perform joint estimation of the utility function coefficients and the O-D matrix. The increase in sample size should help to identify all parameters of interest and to improve the quality of the statistical inference. The use of cloud computing is beneficial to handle a larger number of paths and, thus, for achieving a larger reduction of prediction error. However, the cost of estimating a model with data collected in multiple periods can become high. We expect that using deep learning models, computational graphs, and automatic differentiation tools will help handle large amounts of multi-day data while keeping the computational costs reasonable. Regarding statistical inference, our study lacks a more careful treatment of the endogeneity that arises from the computation of travel time over the iterations of the bilevel optimization. Therefore, using Two-Stage Least Squares (2SLS) and instrumental variables may contribute to correct bias and inconsistency in the coefficient estimates caused by the presence of endogeneity (Gallant and Jorgenson, 1979).

We are also interested in relaxing the assumption of a homogeneous and linear-in-parameters utility function. Using a non-homogeneous utility function and fixed effects may contribute to capturing the heterogeneity of preferences among individuals traveling between different O-D pairs. Besides, the impact of time variability may be better captured with a non-linear specification of the utility function as the one used in prospect theory route choice models. We will also look at estimating the coefficient weighting the utility term associated with the path size correction. We would also like to leverage the use of GPS data to have a better prior of the path sets among O-D pairs and to improve the estimation of the utility function coefficients. Here the integration of our methodology with the nested recursive logit model (Mai et al., 2015) seems a promising avenue for further research.

This study chooses traffic flows for the response function of the non-linear least objective functions. However, other choices of response functions are also admissible, such as link travel times or traffic densities. In static traffic assignment, both quantities are monotonic functions of the traffic counts, and thus, they are suitable quantities to estimate the utility function coefficients. Finally, system-level data is less subject to sampling bias than data collected from individual surveys but is also more subject to measurement errors. Thus, we expect the joint use of travel surveys and system-level data can help leverage the strengths and weaknesses of each data source.

11. Model implementation and data

The Python code that implements our methodology and the data collected from the Fresno, CA network are found at: https://github.com/pabloguarda/isuelogit. The folder `notebooks` contains Jupyter notebooks that reproduce all the results presented in this paper.

CRediT authorship contribution statement

Pablo Guarda: Study conception and design, Data collection, Programming and experiments, Analysis and interpretation of results, Draft manuscript preparation. **Sean Qian:** Study conception and design, Data collection, Analysis and interpretation of results, Draft manuscript preparation.

Acknowledgment

This research is supported by a National Science Foundation, United States grant CMMI-1751448. All authors approved the version of the manuscript to be published.

Appendix A. Proofs and derivations

A.1. Notation

Tables 2–4 present the notation used throughout the paper.

Table 2
Variables and parameters of network model.

Notations	Definitions
$\mathcal{A}$	The set of all links
$\mathcal{V}$	The set of all nodes
$\mathcal{W}$	The set of O-D pairs
$\mathcal{H}$	The set of all paths
$\mathcal{H}_w$	The set of paths connecting O-D pair $w \in \mathcal{W}$
$\mathcal{A}^o, \mathcal{A}^u$	The sets of links with observed and unobserved traffic counts, respectively
$\mathbf{M} \in \mathbb{R}^{\lvert\mathcal{W}\rvert\times\lvert\mathcal{H}\rvert}$	The path-demand incidence matrix
$\mathbf{D} \in \mathbb{R}^{\lvert\mathcal{A}\rvert\times\lvert\mathcal{H}\rvert}$	The path-link incidence incidence matrix
$\mathbf{Q} \in \mathbb{R}^{\lvert\mathcal{V}\rvert\times\lvert\mathcal{V}\rvert}$	The O-D matrix
$\mathbf{q} \in \mathbb{R}^{\lvert\mathcal{W}\rvert}$	The dense vector associated to the O-D matrix
q_w	The demand in O-D pair $w \in \mathcal{W}$
$\mathbf{x} \in \mathbb{R}_{\geq 0}^{\lvert\mathcal{A}\rvert}$	The vector of link flows
$x_a \in \mathbb{R}_{\geq 0}$	Link flow in link $a \in \mathcal{A}$
$\bar{\mathbf{t}}^0 \in \mathbb{R}_{+}^{\lvert\mathcal{A}\rvert}$	The vector of links' free flow travel times
$\mathbf{f} \in \mathbb{R}_{\geq 0}^{\lvert\mathcal{H}\rvert}$	The vector of path flows
$f_h \in \mathbb{R}_{\geq 0}$	The path flow on path $h \in \mathcal{H}$
$\mathbf{p} \in \mathbb{R}_{]0,1[}^{\lvert\mathcal{H}\rvert}$	The vector of path choice probabilities
$p_h \in \mathbb{R}_{]0,1[}$	The choice probability of path $h \in \mathcal{H}$

Table 3
Variables and parameters for statistical inference.

Notations	Definitions
N	The sample of traffic counts
$\ell(\boldsymbol{\theta}) : \mathbb{R}^{\lvert\mathcal{K}\rvert} \to \mathbb{R}$	The objective function of the LUE problem
$\hat{\boldsymbol{\theta}} \in \mathbb{R}^{\lvert\mathcal{K}\rvert}$	The vector of estimated utility function coefficients
$\hat{\boldsymbol{\theta}}_{\mathbf{Z}} \in \mathbb{R}^{\lvert\mathcal{K}_Z\rvert}$	The vector of estimated utility function coefficients associated with the exogenous attributes
$\mathbf{p}(\hat{\boldsymbol{\theta}}) : \mathbb{R}^{\lvert\mathcal{K}\rvert} \to \mathbb{R}^{\lvert\mathcal{H}\rvert}$	The vector of path choice probability functions
$p_h(\hat{\boldsymbol{\theta}}) : \mathbb{R}^{\lvert\mathcal{K}\rvert} \to \mathbb{R}$	The path choice probability function associated to path $h \in \mathcal{H}$
$\mathbf{x}(\hat{\boldsymbol{\theta}}) : \mathbb{R}^{\lvert\mathcal{K}\rvert} \to \mathbb{R}^{\lvert\mathcal{A}\rvert}$	The vector of traffic flow (response) functions
$x_a(\hat{\boldsymbol{\theta}}) : \mathbb{R}^{\lvert\mathcal{K}\rvert} \to \mathbb{R}$	The traffic flow (response) function associated to link $a \in \mathcal{A}$
$\tilde{\mathbf{X}} = D_{\boldsymbol{\theta}}\ \mathbf{x}(\boldsymbol{\theta}) \in \mathbb{R}^{\lvert\mathcal{A}^o\rvert\times\lvert\mathcal{K}\rvert}$	The design matrix in NLLS
$\bar{\mathbf{x}} \in \mathbb{R}_{\geq 0}^{\lvert\mathcal{A}^o\rvert}$	The vector of observed traffic counts
$\bar{x}_a \in \mathbb{R}_{\geq 0}$	The traffic count measurement at link $a \in \mathcal{A}$
$\bar{T}_{d,H_0}$	The t-test associated to attribute $d \in \mathcal{K}$ and under null hypothesis H_0
$\bar{F}_{1,2} \in \mathbb{R}_{\geq 0}$	The F-test comparing models 1 and 2
$\sigma^2 \in \mathbb{R}_{\geq 0}$	The variance of the errors in the nonlinear regression model
$\hat{\sigma}^2 \in \mathbb{R}_{\geq 0}$	The estimated variance of the errors in the nonlinear regression model

Table 4
Variables and parameters of route choice model.

Notations	Definitions
$\mathcal{K}_{\mathbf{Z}}$	The set of exogenous attributes in the utility function
$\mathcal{K}$	The set of attributes in the utility function
$\mathcal{L}$	The set of travelers in the network
$\mathcal{J}_l$	Consideration set of traveler $l \in \mathcal{L}$
$\mathbf{t} \in \mathbb{R}_{\geq 0}^{\lvert\mathcal{A}\rvert}$	The vector of endogenous link travel times
$\mathbf{Z} \in \mathbb{R}^{\lvert\mathcal{A}\rvert \times \lvert\mathcal{K}_{\mathbf{Z}}\rvert}$	The matrix of exogenous link attributes' values
$\mathbf{z}_k \in \mathbb{R}^{\lvert\mathcal{A}\rvert}$	The vector of values for the exogenous link attribute $k \in \mathcal{K}_{\mathbf{Z}}$
Z_{ak}	The value of the exogenous attribute $k \in \mathcal{K}_{\mathbf{Z}}$ at link $a \in \mathcal{A}$
$\bar{\mathbf{t}} \in \mathbb{R}_{\geq 0}^{\lvert\mathcal{A}\rvert}$	The vector of exogenous link travel times
$\boldsymbol{\theta} \in \mathbb{R}^{\lvert\mathcal{K}\rvert}$	The vector of true utility function coefficients
$\boldsymbol{\theta}_{\mathbf{Z}} \in \mathbb{R}^{\lvert\mathcal{K}_{\mathbf{Z}}\rvert}$	The vector of true utility function coefficients associated with the exogenous attributes
θ_k	The utility function coefficient associated to attribute $k \in \mathcal{K}$
$\theta_t \in \mathbb{R}_{\leq 0}$	The travel time coefficient
$\mathbf{v} \in \mathbb{R}^{\lvert\mathcal{A}\rvert}$	The vector of link utilities
v_a	The link utility associated to link $a \in \mathcal{A}$
U_{jl}	The latent (unobservable) utility that traveler l attained to alternative (path) $j \in \mathcal{J}_l$
V_{jl}	The observable utility that traveler l attained to alternative (path) $j \in \mathcal{J}_l$
ϵ_{jl}	The latent (unobservable) error in the utility function associated with traveler l and alternative (path) $j \in \mathcal{J}_l$
$\mu \in \mathbb{R}_+$	Scale parameter of the logit model and the extreme value Type 1 distribution

A.2. Extension of SUELOGIT

A.2.1. Original travel time based formulation

The standard formulation of the SUE with logit assignment (SUELOGIT) problem assumes a utility function dependent on travel time only. Fisk (1980) proved that the first order necessary optimality condition of the following optimization problem gives a path flow solution that is logit distributed:

$$
\begin{aligned}
\underset{\{f_h\}_{h\in\mathcal{H}},\{x_a\}_{a\in\mathcal{A}}}{\text{minimize}} \quad & \sum_{a\in\mathcal{A}} \int_0^{x_a} t_a(u)du + \frac{1}{\theta} \sum_{w\in\mathcal{W}} \sum_{h\in\mathcal{H}_w} f_h \ln(f_h) \\
\text{subject to} \quad & \sum_{h\in\mathcal{H}_w} f_h = q_w \qquad \forall w \in \mathcal{W} \\
& x_a = \sum_{w\in\mathcal{W}} \sum_{h\in\mathcal{H}_w} f_h \delta_{ah} \qquad \forall a \in \mathcal{A} \\
& f_h \geq 0 \qquad \forall h \in \mathcal{H}
\end{aligned}
\tag{27}
$$

From this formulation is clear that if $|\theta| \rightarrow \infty^+$, the objective function reduces to the first term associated with the Beckmann transformation, and thus, the SUE and DUE optimization problems become equivalent. The LUE and ODLUE literature typically define $\theta \in \mathbb{R}_+$ as a dispersion parameter measuring the sensitivity of route choices to travel times (Yang et al., 2001) and interpret it as the accuracy of the travelers' perception about travel costs (Daganzo, 1977; Wang et al., 2016) or the level of information about travel costs (Lo and Chan, 2003; Liu and Fricker, 1996). Under a single attribute utility function, the interpretation of the parameter may be irrelevant or difficult to falsify. However, in the case of a multi-attribute utility function, these interpretations would rest on the relationship between the magnitude of the dispersion parameter and the amount of unobservable components of the utility function, which the modeler ignores. To understand this connection, it is critical to reformulate the problem into a utility-based representation.

A.2.2. Utility-based formulation with a single endogenous attribute

The objective function in Problem 27 is written in terms of the link performance functions $t_a : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_+$ instead of the observable component of the travelers' utility function $v_a : \mathbb{R} \rightarrow \mathbb{R}$ associated to each link $a \in \mathcal{A}$. Assume the observable component of the travelers' utility function is given by $v_a = \theta_t t_a$, where $\theta_t = \mu\tilde{\theta}_t$ is the coefficient measuring the preference of travelers for travel time. This coefficient is scaled by a factor $\mu \in \mathbb{R}_+$ proportional to the standard deviation of the unobservable component of the utility function, i.e., $\theta_t = \mu\tilde{\theta}_t$ where $\tilde{\theta}_t$ is the unscaled vector of logit coefficients and which is not identifiable. Then, the utility-based representation of Problem 27 can be written in a vectorized form as follows:

$$
\begin{aligned}
\min_{\mathbf{x},\mathbf{f}} \quad & \sum_{a\in\mathcal{A}} \int_0^{x_a} \frac{v_a(u)}{\theta_t} du - \frac{1}{\theta_t} \langle \mathbf{f}, \ln \mathbf{f} \rangle \\
\text{s.t.} \quad & \mathbf{M}\mathbf{f} = \mathbf{q} \\
& \mathbf{D}\mathbf{f} = \mathbf{x} \\
& \mathbf{x}, \mathbf{f} \geq \mathbf{0}
\end{aligned}
\tag{28}
$$

where $\mathbf{x} \in \mathbb{R}_{\geq 0}^{|\mathcal{A}|}, \mathbf{f} \in \mathbb{R}_{\geq 0}^{|\mathcal{H}|}, \mathbf{q} \in \mathbb{R}_{+}^{|\mathcal{V}\times\mathcal{V}|}, \mathbf{M} \in \mathbb{R}_{+}^{|\mathcal{V}\times\mathcal{V}|\times|\mathcal{H}|}$, $\mathbf{D} \in \mathbb{R}_{+}^{|\mathcal{A}|\times|\mathcal{H}|}$. A key observation with respect to Problem 27 is that $\theta = -\theta_t > 0$ and since $\theta_t = \mu\tilde{\theta}_t$ is now clear that the dispersion parameter is also scaled by the scale factor μ of the logit model.

A.2.3. Utility function with an endogenous attribute and multiple exogenous attributes

Problem 28 can written as:

$$\begin{aligned} \underset{\{f_h\}_{h\in\mathcal{H}}, \{x_a\}_{a\in\mathcal{A}}}{\text{minimize}} \quad & \sum_{a\in\mathcal{A}} \int_0^{x_a} v'_a(u)du - \frac{1}{\theta_t} \sum_{w\in\mathcal{W}} \sum_{h\in\mathcal{H}_w} f_h \ln f_h \\ \text{subject to} \quad & \sum_{h\in H_w} f_h = q_w \qquad \forall w \in \mathcal{W} \\ & x_a = \sum_{rs} \sum_{h\in\mathcal{H}_w} f_h \delta_{ah} \quad \forall a \in \mathcal{A} \\ & f_h \geq 0 \qquad \forall h \in \mathcal{H} \end{aligned} \tag{29}$$

where the link utility function $v_a(u)$ at a traffic flow level u was reparameterized as:

$$v'_a(u) = \frac{v_a(u)}{\theta_t} = \frac{1}{\theta_t}\left(\theta_t t_a(u) + \sum_{k\in\mathcal{K}_\mathbf{Z}} \theta_k \cdot Z_{ak}\right) \tag{30}$$

Z_{ak} represents the value of the exogenous attribute $k \in \mathcal{K}_\mathbf{Z}$ at link $a \in \mathcal{A}$ and $\theta_t \in \mathbb{R}_-, \boldsymbol{\theta}_Z \in \mathbb{R}^{|\mathcal{K}_\mathbf{Z}|}$ are the set of coefficients measuring the travelers' preferences for the endogenous attribute t and the exogenous attributes $z \in \mathcal{K}_\mathbf{Z}$. Note that if both terms of the objective function in Problem 29 are multiplied by $\theta_t < 0$, the problem becomes a maximization:

$$\begin{aligned} \underset{\{f_h\}_{h\in\mathcal{H}}, \{x_a\}_{a\in\mathcal{A}}}{\text{maximize}} \quad & \sum_{a\in\mathcal{A}} \int_0^{x_a} v_a(u)du - \sum_{w\in\mathcal{W}} \sum_{h\in\mathcal{H}_w} f_h \ln f_h \\ \text{subject to} \quad & \sum_{h\in\mathcal{H}_w} f_h = q_w \qquad \forall w \in \mathcal{W} \\ & x_a = \sum_{w\in\mathcal{W}} \sum_{h\in\mathcal{H}_w} f_h \delta_{ah} \quad \forall a \in \mathcal{A} \\ & f_h \geq 0 \qquad \forall h \in \mathcal{H} \end{aligned} \tag{31}$$

which, in compact form, can be written as:

$$\begin{aligned} \max_{\mathbf{x},\mathbf{f}} \quad & \sum_{a\in\mathcal{A}} \int_0^{x_a} v_a(u)du - \langle \mathbf{f}, \ln \mathbf{f} \rangle \\ \text{s.t.} \quad & \mathbf{Mf} = \mathbf{q} \\ & \mathbf{Df} = \mathbf{x} \\ & \mathbf{x}, \mathbf{f} \geq \mathbf{0} \end{aligned}$$

A.2.4. Logit assignment of path flows in utility-based formulation

Following the rationale of the proof in Fisk (1980), we can prove that the path flow solution of the optimization model presented in Problem 29 follows a logit assignment. The Lagrangian $\mathcal{L}$ of the problem is:

$$\mathcal{L} = \sum_{a\in\mathcal{A}} \int_0^{x_a} (t_a(w) + \sum_{k\in\mathcal{K}_\mathbf{Z}} (\theta_k/\theta_t) Z_{ak}) dw - \frac{1}{\theta_t} \sum_{w\in\mathcal{W}} \sum_{h\in\mathcal{H}_w} f_h \ln f_h + \sum_{w\in\mathcal{W}} \lambda_w (\sum_{h\in\mathcal{H}_w} f_h - q_w) \tag{32}$$

The set of first-order optimality conditions are:

$$\begin{aligned} \frac{\partial L}{\partial f_h} = \sum_{a\in\mathcal{A}} \frac{\partial x_a}{\partial f_h}\left(t_a(x_a) + \sum_{k\in\mathcal{K}_\mathbf{Z}} (\theta_k/\theta_t) Z_{ak}\right) - \frac{1}{\theta_t}\left(\ln f_h + f_h \frac{1}{f_h}\right) + \sum_{w\in\mathcal{W}} \lambda_w \delta_{hw}^w = 0 \\ \sum_{a\in\mathcal{A}} \delta_{ah}\left(t_a(x_a) + \sum_{k\in\mathcal{K}_\mathbf{Z}} (\theta_k/\theta_t) Z_{ak}\right) - \frac{1}{\theta_t}\left(\ln f_h + 1\right) + \lambda_w^h = 0 \quad \forall h \in \mathcal{H} \end{aligned} \tag{33}$$

where δ_{hw}^w takes the value 1 if path h belong to O-D pair $w \in \mathcal{W}$, and 0 otherwise, and $\lambda_w^h = \sum_{w\in\mathcal{W}} \lambda_w \delta_{hw}^w$. Let us define $V'_h = \sum_{a\in\mathcal{A}} \delta_{ah}(t_a(x_a)) + \sum_{k\in\mathcal{K}_\mathbf{Z}} (\theta_k/\theta_t) Z_{ak}$ as a reparameterized utility function associated to path $h \in \mathcal{H}$. Now we can find an expression for f_h using Eq. (33):

$$\begin{aligned} V'_h - \frac{1}{\theta_t}\left(\ln f_h + 1\right) = -\lambda_w^h \\ \ln f_h = \theta_t \lambda_w^h \delta_{hw}^w + \theta_t V'_h - 1 \\ f_h = \exp(\theta_t \lambda_w^h + \theta_t V'_h - 1) \end{aligned} \tag{34}$$

Using the conservation constraint of path flows and demand:

$$\sum_{h\in\mathcal{H}_w} f_w = \sum_{h\in\mathcal{H}_w} \exp(\theta_t \lambda_w^h + \theta_t V_h' - 1) = q_w \tag{35}$$

Noting that the value of λ_w^h is the same $\forall h \in \mathcal{H}_w$:

$$q_w = \exp(\theta_t \lambda_w^h) \sum_{h\in\mathcal{H}_w} \exp(\theta_t V_h' - 1)$$

$$\exp(\theta_t \lambda_w^h) = \frac{q_w}{\sum_{h\in\mathcal{H}_w} \exp(\theta_t V_h' - 1)} \tag{36}$$

Replacing Eq. (36) into Eq. (34):

$$f_h = \exp(\theta_t \lambda_w^h)\exp(\theta_t V_h' - 1) = q_w \frac{\exp(\theta_t V_h')}{\sum_{j\in\mathcal{H}_w} \exp(\theta_t V_j')} \tag{37}$$

Substituting by $V_h = V_h'\theta_t = \sum_{a\in\mathcal{A}} \delta_{ah}(\theta_t t_a(x_a)) + \sum_{k\in\mathcal{K}_Z}(\theta_k/\theta_t)Z_{ak})$:

$$f_h = q_w \frac{\exp(V_h)}{\sum_{j\in\mathcal{H}_w} \exp(V_h)} \tag{38}$$

where it is clear that the set of optimal path flows $\{f_h\}_{h\in\mathcal{H}}$ follows a logit assignment.

A.3. Monotonicity of path choice probabilities and traffic flow functions

Proposition 5 (*Monotone Path Choice Probabilities*). *Assume the travelers' utility function is a linear weight between a set of attributes and coefficients. If there are at most two paths to travel between every O-D pair, the path choice probabilities are coordinate-wise monotonic functions with respect to the utility function coefficients.*

Proof. Let us start from a general case where there are an arbitrary number of alternative paths and attributes and thus, where the choice probabilities are obtained from a softmax function. Define $\theta_t \in \mathbb{R}$ as the utility function coefficient associated with an attribute $t \in \mathcal{K}$ and τ_i as the utility component associated with the remaining set of attributes of the path $i \in \mathcal{H}_{rs}$ of an arbitrary O-D pair $w \in \mathcal{W}$. Then, the choice probability of path i in the O-D pair $w \in \mathcal{W}$ is:

$$p_i(\theta_t) = \frac{\exp(\theta_t t_i + \tau_i)}{\sum_{j\in\mathcal{H}_w} \exp(\theta_t t_j + \tau_j)} \tag{39}$$

The first derivative of the softmax function with respect to the utility function parameter θ is:

$$\begin{aligned}\frac{\partial p_i(\theta)}{\partial \theta} &= \frac{\exp(\theta_t t_i + \tau_i)t_i\left(\sum_{j\in\mathcal{H}_w}\exp(\theta_t t_j + \tau_j)\right) - \left(\sum_{j\in\mathcal{H}_w}\exp(\theta t_j + \tau_j)t_j\right)\exp(\theta_t t_i + \tau_i)}{\left(\sum_{j\in\mathcal{H}_w}\exp(\theta_t t_j + \tau_j)\right)^2} \\ &= \frac{\exp(\theta_t t_i + \tau_i)}{\sum_{j\in\mathcal{H}_w}\exp(\theta t_j + \tau_j)} \frac{\sum_{j\in\mathcal{H}_w}\exp(\theta_t t_j + \tau_j)(t_i - t_j)}{\sum_{j\in\mathcal{H}_w}\exp(\theta t_j + \tau_j)} \\ &= p_i \frac{\sum_{j\in\mathcal{H}_w}\exp(\theta_t t_j + \tau_j)(t_i - t_j)}{\sum_{j\in\mathcal{H}_w}\exp(\theta_t t_j + \tau_j)}\end{aligned}$$

The sign of the derivative is given by:

$$\operatorname{sign}\left(\frac{\partial p_i(\theta_t)}{\partial \theta}\right) = \operatorname{sign}\left(p_i \frac{\sum_{j\in\mathcal{H}_w}\exp(\theta_t t_j + \tau_j)(t_i - t_j)}{\sum_{j\in\mathcal{H}_w}\exp(\theta t_j + \tau_j)}\right) = \operatorname{sign}\left(\sum_{j\in\mathcal{H}_w}\exp(\theta t_j + \tau_j)(t_i - t_j)\right) \tag{40}$$

Given that exponential functions are always positive, Eq. (40) will be negative or positive for any $\theta_t \in \mathbb{R}$ when $t_i \neq \min\{t_j\}_{j\in\mathcal{H}_w}$ or $t_i \neq \max\{t_j\}_{j\in\mathcal{H}_w}$, respectively. Therefore, when there are at most two paths connecting the O-D pair, the path choice probability for any path $i \in \mathcal{H}_w$ will necessarily be a monotonic function with respect to θ_t. The same analysis can be extended to every utility function coefficient $\theta_k, \forall k \in \mathcal{K}$ and O-D pair $w \in \mathcal{W}$. This proves that the path choice probabilities are coordinate-wise monotonic functions with respect to the utility function coefficients when all path sets have at most two paths. □

Proposition 6 (*Monotone Traffic Flow Functions Under Dominant and Non Dominant Paths*). *Assume the travelers' utility function is a linear weight between a set of attributes and coefficients. Let us define dominated and dominating paths as those where an attribute of the utility function reaches its maximum or minimum value, respectively, within each O-D pair. If the set of paths traversing a link are all dominating or dominated paths, the traffic flow function at that link is coordinate-wise monotonic with respect to the utility function coefficients*

Proof. Consider an arbitrary attribute $t \in \mathcal{K}$ and define θ_t as the coefficient weighting that attribute in the travelers' utility function. Denote τ_i as the utility component associated with the remaining set of attributes in path $i \in \mathcal{H}$. The traffic flow function $x_a(\theta) : \mathbb{R}^{|\mathcal{K}|} \to \mathbb{R}$ associated to any link $a \in \mathcal{A}$ is given by:

$$x_a(\theta) = \sum_{i\in\mathcal{H}} f_i \delta_{ai} = \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} q_w p_i(\theta_t)\delta^w_{wi} = \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \frac{\exp(\theta_t t_i + \tau_i)}{\sum_{j\in\mathcal{H}_w} \exp(\theta_t t_j + \tau_j)} \tag{41}$$

where δ^a_{ai} takes the value 1 if path i traverses link a, and 0 otherwise, and δ^w_{wi} takes the value 1 if path i belong to O-D pair w, and 0 otherwise. If the set of paths traversing the link are all dominating or dominated, then $\forall w \in \mathcal{W},\ i \in \mathcal{H}_w,\ t_i = \min\{t_j\}_{j\in\mathcal{H}_w}$ or $t_i = \max\{t_j\}_{j\in\mathcal{H}_w}$, respectively. From Proposition 5, we observed that the softmax function associated with each path choice probability $p_i(\theta_t)$ will be either a monotonically decreasing or an increasing function when $t_i = \min\{t_j\}_{j\in\mathcal{H}_{rs}}$ or $t_i = \max\{t_j\}_{j\in\mathcal{H}_{rs}}$, respectively. By assumption, $x_a(\theta)$ is a positive weighted sum of choice probabilities associated with dominating or dominated paths, respectively, that is, a positive weighted sum of monotonically increasing or decreasing functions. Therefore, $x_a(\theta)$ it is either a monotonically increasing or decreasing function with respect to θ_t, and the same analysis can be applied to every coordinate $d \in \mathcal{K}$. Finally, $x_a(\theta)$ is a coordinate-wise monotonic function, which completes the proof. □

Proposition 7 (*Monotone Traffic Flow Functions Under Binary Attributes*)**.** *Suppose there are only two paths connecting every O-D pair and that the travelers' utility function only depends on a binary attribute that can take the values 0 or 1. Then, the traffic flow functions are monotonic with respect to the coefficient weighting the binary attribute.*

Proof. Consider the traffic flow function $x_a(\theta) : \mathbb{R}^{|\mathcal{K}|} \to \mathbb{R}$ associated to an arbitrary link $a \in \mathcal{A}$:

$$\begin{aligned} x_a(\theta) &= \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} q_w p_i(\theta_t)\delta^w_{wi} = \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \frac{\exp(\theta\mathbb{I}(z_i = 1))}{\exp(\theta\mathbb{I}(z_i = 1)) + \exp(\theta\mathbb{I}(z_{-i} = 1))} \\ &= \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \frac{1}{1 + \exp(\theta\mathbb{I}(z_{-i} = 1) - (\theta\mathbb{I}(z_i = 1)))} \end{aligned}$$

where $z_i \in \{0, 1\}, \forall i \in \mathcal{H}$, δ^a_{ai} takes the value 1 if path i traverses link $a \in \mathcal{A}$, and 0 otherwise, and δ^w_{wi} takes the value 1 if path i belong to O-D pair $w \in \mathcal{W}$, and 0 otherwise. Because of the existence of only two paths connecting each O-D pair, we can express $x_a(\theta)$ in terms of the sigmoid function $\sigma(\cdot)$:

$$\begin{aligned} x_a(\theta) &= \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \sigma\left(\theta(\mathbb{I}(z_{-i} = 1) - \mathbb{I}(z_i = 1))\right) \\ &= \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \mathbb{I}(z_i = z_{-i})\sigma(0) + \sigma(-\theta) \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \mathbb{I}(z_i > z_{-i}) \\ &+ \sigma(\theta) \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \mathbb{I}(z_i < z_j) \\ &= \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \mathbb{I}(z_i = z_{-i})\sigma(0) + (1 - \sigma(\theta)) \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \mathbb{I}(z_i > z_{-i}) \\ &+ \sigma(\theta) \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \mathbb{I}(z_i < z_j) \\ &= \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \mathbb{I}(z_i = z_{-i})\sigma(0) + \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \mathbb{I}(z_i > z_{-i}) \\ &+ \sigma(\theta)\left(\sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \mathbb{I}(z_i < z_j) - \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \mathbb{I}(z_i > z_{-i})\right) \end{aligned}$$

To analyze the monotonicity of $x_a(\theta)$ is convenient to compute the first derivative of $x_a(\theta)$:

$$\frac{\partial x_a(\theta)}{\partial\theta} = \left(\sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \mathbb{I}(z_i < z_{-i}) - \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \mathbb{I}(z_i > z_{-i})\right)\sigma(\theta)(1 - \sigma(\theta)) \tag{42}$$

and then analyzing its sign:

$$\text{sign}\left(\frac{\partial x_a(\theta)}{\partial\theta}\right) = \left(\sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \mathbb{I}(z_i < z_{-i}) - \sum_{i\in\mathcal{H}} \delta^a_{ai} \sum_{w\in\mathcal{W}} \delta^w_{wi}\ q_w \mathbb{I}(z_i > z_{-i})\right) \tag{43}$$

which does not depend on θ because $\sigma(\theta)(1 - \sigma(\theta)) > 0, \forall\theta \in \mathbb{R}$ in Eq. (42). Then, it is clear that the function $x(\theta)$ is monotonic with respect to θ, which completes the proof. □

Remark 10. Note that the left and right terms in Eq. (43) are the sums of demand associated to dominated and dominating paths, respectively. Therefore, the traffic flow function will be monotonically increasing or decreasing if the sum associated to the dominating paths is greater or lower, respectively.

A.4. Coordinate-wise properties of the LUE problem

For the analyses in this section, it will be helpful to introduce Definition 2 of coordinate-wise pseudo-convexity. A function f is coordinate-wise pseudo-convex if, when all coefficients of the utility function except for one coefficient are fixed, it is pseudo-convex with respect to the non-constant coefficient. Note that if the function is dependent on a single attribute, the definitions of coordinate-wise pseudo-convexity and pseudo-convexity are equivalent. Formally:

Definition 2 (*Coordinate Wise Pseudo-Convexity*).
$\ell : \mathbb{R}^{|\mathcal{K}|} \to \mathbb{R}$ is a coordinate-wise *pseudo-convex* function iff $\forall i \in \mathcal{K}, \forall \theta^1, \theta^2 \in \mathcal{S}_i$, such that $\mathcal{S}_i = \{\theta^1, \theta^2 \in \mathbb{R}^{|\mathcal{K}|} \mid \theta^1_j = \theta^2_j, \forall j \neq i\}$:

$$\ell(\theta^2) < \ell(\theta^1) \implies \frac{\partial \ell(\theta^1)}{\partial \theta_i}(\theta^2_i - \theta^1_i) < 0 \tag{44}$$

or equivalently,

$$\frac{\partial \ell(\theta^1)}{\partial \theta_i}(\theta^2_i - \theta^1_i) \geq 0 \implies \ell(\theta^2) \geq \ell(\theta^1) \tag{45}$$

Analogously to Definition 2, the definition of strict quasi-convexity can be also applied coordinate-wise:

Definition 3 (*Coordinate-Wise Strict Quasi-Convexity*). $\ell : \mathbb{R}^{|\mathcal{K}|} \to \mathbb{R}$ is said to be coordinate-wise strictly quasi-convex iff $\forall i \in \mathcal{K}, \forall \theta^1, \theta^2 \in \mathcal{S}_i$, such that $\mathcal{S}_i = \{\theta^1, \theta^2 \in \mathbb{R}^{|\mathcal{K}|} \mid \theta^1_j = \theta^2_j, \forall j \neq i\}$:

$$\ell(\theta_1) < \ell(\theta_2) \implies \ell(\lambda \theta_1 + (1 - \lambda_1)\theta_2) < \ell(\theta_1)$$

or equivalently:

$$\ell(\lambda \theta_1 + (1 - \lambda)\theta_2) < \max\left(\ell(\theta_1), \ell(\theta_2)\right)$$

Proposition 8, Appendix A.4 shows that coordinate-wise monotonicity of the traffic count (response) functions imply coordinate-wise pseudo-convexity of the objective function of the LUE problem under exogenous travel times. The definition of coordinate-wise monotonicity is the following:

Definition 4 (*Coordinate-Wise Monotonicity of Response Functions*). The traffic count (response) functions are monotonic with respect to each coefficient of the utility function if, when all coefficients of the utility function except for one coefficient are kept constant, the response functions are monotonic with respect to the non-constant coefficient.

Propositions 6 and 7, Appendix A.3 illustrate two cases where under mild assumptions the coordinate-wise monotonicity of the traffic flow functions holds in a general transportation network. While it is easy to create counter-examples where the coordinate-wise monotonicity of the traffic flow functions is violated, analysis of synthetic data suggests that the assumption holds for most links in practice (see Appendix D.3.1).

Proposition 8 (*Sufficient Condition for Coordinate-Wise Pseudo-Convexity of Objective Function*). *The objective function of the LUE problem under an uncongested network is coordinate-wise pseudo-convex if there is access to a single observation j associated with a traffic count $\bar{x}_j$ and the traffic flow functions are coordinate-wise monotonic with respect to each utility function coefficient.*

Proof. Let us define $\theta^1_d, \theta^2_d \in \mathbb{R}^{|\mathcal{K}|}$ as vectors with all coordinates set to 0, except for the coordinate $d \in \mathcal{K}$. Let θ^1_d, θ^2_d be the values of the non-zero coordinates in θ^1_d, θ^2_d. To prove coordinate-wise pseudo-convexity of the objective function $\ell : \mathbb{R}^{|\mathcal{K}|} \to \mathbb{R}$ when there is a single observation j available, it suffices to show that the following holds:

$$2\left(\mathbf{x}_j(\theta^1_d) - \bar{\mathbf{x}}_j\right)(\theta^2_d - \theta^1_d)\frac{\partial \mathbf{x}_j}{\partial \theta_d}\bigg|_{\theta=\theta^1_d} \geq 0 \implies (\mathbf{x}_j(\theta^2_d) - \bar{\mathbf{x}}_j)^2 \geq (\mathbf{x}_j(\theta^1_d) - \bar{\mathbf{x}}_j)^2 \tag{46}$$

By assumption, the traffic count (response) functions are coordinate-wise monotonic. Let us start considering the set $\mathcal{J}^+$ of functions that are monotonically increasing with respect to θ_d. Since $\frac{\partial \mathbf{x}_j}{\partial \theta_d}\Big|_{\theta=\theta^1_d} > 0, \forall j \in \mathcal{J}^+$, the LHS in Eq. (46) becomes non-negative in the following two cases:

• Case (i): $\mathbf{x}_j(\theta^1_d) - \bar{\mathbf{x}}_j \geq 0 \wedge \theta^2_d \geq \theta^1_d$. Because the increasing monotonicity of the traffic flow functions $j \in \mathcal{J}^+$ with respect to θ_d and given that $\theta^2_d \geq \theta^1_d$:

$$\begin{aligned} \mathbf{x}_j(\theta^2_d) &\geq \mathbf{x}_j(\theta^1_d) \\ \mathbf{x}_j(\theta^2_d) - \bar{\mathbf{x}}_j &\geq \mathbf{x}_j(\theta^1_d) - \bar{\mathbf{x}}_j \end{aligned}$$

Since $\mathbf{x}_j(\theta^1_d) - \bar{\mathbf{x}}_j \geq 0$:

$$\begin{aligned} \mathbf{x}_j(\theta^2_d) - \bar{\mathbf{x}}_j &\geq \mathbf{x}_j(\theta^1_d) - \bar{\mathbf{x}}_j \geq 0 \\ (\mathbf{x}_j(\theta^2_d) - \bar{\mathbf{x}}_j)^2 &\geq (\mathbf{x}_j(\theta^1_d) - \bar{\mathbf{x}}_j)^2 \end{aligned}$$

• Case (ii): $\mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j \leq 0 \wedge \theta_d^2 \leq \theta_d^1$. Because the increasing monotonicity of the traffic flow functions $j \in \mathcal{J}^+$ with respect to θ_d and given that $\theta_d^2 \leq \theta_d^1$:

$$\begin{aligned}\mathbf{x}_j(\theta_d^2) &\leq \mathbf{x}_j(\theta_d^1)\\ \mathbf{x}_j(\theta_d^2) - \bar{\mathbf{x}}_j &\leq \mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j\end{aligned}$$

Since $\mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j \leq 0$:

$$\begin{aligned}\mathbf{x}_j(\theta_d^2) - \bar{\mathbf{x}}_j &\leq \mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j \leq 0\\ (\mathbf{x}_j(\theta_d^2) - \bar{\mathbf{x}}_j)^2 &\geq (\mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j)^2\end{aligned}$$

Now consider the set $\mathcal{J}^-$ of traffic flow functions that are monotonically decreasing with respect to θ_d. Since $\left.\frac{\partial \mathbf{x}_j}{\partial \theta_d}\right|_{\theta=\theta_d^1} < 0, \forall j \in \mathcal{J}^-$, the LHS in Eq. (46) becomes non negative in the following two cases:

• Case (iii): $\mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j \geq 0 \wedge \theta_d^2 \leq \theta_d^1$. Because the decreasing monotonicity of the traffic flow functions $j \in \mathcal{J}^-$ with respect to θ_d and given that $\theta_d^2 \leq \theta_d^1$:

$$\begin{aligned}\mathbf{x}_j(\theta_d^2) &\geq \mathbf{x}_j(\theta_d^1)\\ \mathbf{x}_j(\theta_d^2) - \bar{\mathbf{x}}_j &\geq \mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j\end{aligned}$$

Since $\mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j \geq 0$:

$$\begin{aligned}\mathbf{x}_j(\theta_d^2) - \bar{\mathbf{x}}_j &\geq \mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j \geq 0\\ (\mathbf{x}_j(\theta_d^2) - \bar{\mathbf{x}}_j)^2 &\geq (\mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j)^2\end{aligned}$$

• Case (iv): $\mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j \leq 0 \wedge \theta_d^2 \geq \theta_d^1$. Because the decreasing monotonicity of the traffic flow functions $j \in \mathcal{J}^-$ with respect to θ_d and given that $\theta_d^2 \geq \theta_d^1$:

$$\begin{aligned}\mathbf{x}_j(\theta_d^2) &\leq \mathbf{x}_j(\theta_d^1)\\ \mathbf{x}_j(\theta_d^2) - \bar{\mathbf{x}}_j &\leq \mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j\end{aligned}$$

Since $\mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j \leq 0$:

$$\begin{aligned}\mathbf{x}_j(\theta_d^2) - \bar{\mathbf{x}}_j &\leq \mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j \leq 0\\ (\mathbf{x}_j(\theta_d^2) - \bar{\mathbf{x}}_j)^2 &\geq (\mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j)^2\end{aligned}$$

Finally, if the traffic flow functions are monotonic, the following condition holds $\forall j \in \mathcal{J}^- \cup \mathcal{J}^+$:

$$2\left(\mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j\right)(\theta_d^2 - \theta_d^1)\left.\frac{\partial \mathbf{x}_j}{\partial \theta_d}\right|_{\theta=\theta_d^1} \geq \mathbf{0} \implies (\mathbf{x}_j(\theta_d^2) - \bar{\mathbf{x}}_j)^2 \geq (\mathbf{x}_j(\theta_d^1) - \bar{\mathbf{x}}_j)^2$$

which proves the pseudo-convexity of the objective function of the LUE problem with respect to θ_d. The analysis conducted for θ_d can be applied to every coordinate $d \in \mathcal{K}$ of $\boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{K}|}$, which proves the coordinate-wise pseudo-convexity of the objective function ℓ with respect to an arbitrary vector of utility function coefficients $\boldsymbol{\theta}$. □

Remark 11. When there is access to multiple traffic count observations, the objective function becomes a sum of pseudo-convex functions. Since the sum of pseudo-convex functions is not necessarily pseudo-convex, extending the proof to multiple observations is not direct.

A.5. Existence and uniqueness of global minima of the LUE *problem*

A.5.1. No measurement error in traffic counts

Proofs of the existence and uniqueness of global optima typically rely on the convexity of the optimization problem. Despite the non-convexity of our problem, we can rely on Assumption 1, Section 3 of SUELOGIT and to assume the absence of measurement error to prove existence and uniqueness:

Proposition 9 (*Existence of Global Minima with No Noise in Traffic Counts*)**.** *The* LUE *problem under an uncongested network has a global optimum if traffic count measurements follow* SUELOGIT *and they have no measurement error*

Proof. By assumption, if $\bar{\mathbf{x}}$ follows SUELOGIT, $\exists \boldsymbol{\theta}^\star \in \mathbb{R}^{|\mathcal{K}|} : \mathbf{x}(\boldsymbol{\theta}^\star) = \bar{\mathbf{x}}$. Hence, the objective function reaches its lower bound when $f(\boldsymbol{\theta}^\star) = 0$ and thus $\boldsymbol{\theta}^\star$ is a global minimizer, which completes the proof. □

Proposition 10 (*Uniqueness of Global Minima with No Noise in Traffic Counts*)**.** *The* LUE *problem under an uncongested network has a unique global optimum at $\boldsymbol{\theta}^\star \in \mathbb{R}^{|\mathcal{K}|}$ if (i) the traffic count measurements $\bar{\mathbf{x}}$ follow* SUELOGIT *and they have no measurement error, (ii) the Jacobian matrix of the objective function has full rank at any feasible point $\boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{K}|}$, (iii) the response functions $\mathbf{x}(\boldsymbol{\theta})$ are strictly monotone*

Proof. By Proposition 9 and the assumption that $\bar{\mathbf{x}}$ follows SUELOGIT, the optimization problem has a global optima at $\boldsymbol{\theta}^\star$. To prove uniqueness, we first derive the first-order necessary optimality condition:

$$\nabla_{\boldsymbol{\theta}} \|\mathbf{x}(\boldsymbol{\theta}) - \bar{\mathbf{x}}\|_2^2 = -2 \left[D_{\boldsymbol{\theta}}\ \mathbf{x}(\boldsymbol{\theta})\right]^\top (\bar{\mathbf{x}} - \mathbf{x}(\boldsymbol{\theta})) = \mathbf{0} \tag{47}$$

By assumption, the Jacobian matrix $D_{\boldsymbol{\theta}}\ \mathbf{x}(\boldsymbol{\theta})$ is full rank $\forall \boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{K}|}$ and thus $[D_{\boldsymbol{\theta}}\ \mathbf{x}(\boldsymbol{\theta})]^\top (\bar{\mathbf{x}} - \mathbf{x}(\boldsymbol{\theta})) = \mathbf{0}$ iff $\mathbf{x}(\boldsymbol{\theta}) = \bar{\mathbf{x}}$. By the strict monotonicity of $\mathbf{x}(\boldsymbol{\theta})$, there exists a unique value of $\boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{K}|}$ such that $\mathbf{x}(\boldsymbol{\theta}) = \bar{\mathbf{x}}$. Therefore, this proved that $\boldsymbol{\theta} = \boldsymbol{\theta}^\star$ is the unique global optimum. □

A.5.2. Measurement error in traffic counts

The assumption of no measurement error in the traffic counts can be relaxed to prove the existence of a solution to the LUE problem. As shown in Proposition 11, this requires to leverage the pseudo-convexity of the LUE problem. Proposition 8, Appendix A.4 shows that the assumption on pseudo-convexity is plausible when the travelers' utility function is dependent on a single exogenous attribute. Proposition 12 proves uniqueness by leveraging the strict quasi-convexity of pseudo-convex functions (Bazaraa et al., 2006).

Proposition 11 (*Sufficient Condition for the Existence of a Global Minima*)**.** *Assume the objective function ℓ of the* LUE *problem is dependent on a single exogenous attribute and that it is pseudo-convex with respect to the parameter weighting that attributes in the travelers' utility function. Then, if the first derivative of ℓ vanishes at $\theta^\star$, $\theta^\star$ is a global minima.*

Proof. By assumption, $\frac{\partial \ell}{\partial \theta}(\theta^\star) = 0$. By Definition 1 of pseudo-convexity, $\frac{\partial \ell}{\partial \theta}(\theta^\star) = 0,\ \ell(\theta_1) \geq \ell(\theta^\star),\ \forall \theta_1 \in \mathbb{R}$, which completes the proof. □

Definition 5 (*Strict Quasi-Convexity*)**.** f is said to be strictly quasi-convex if for every $x_1, x_2 \in \mathcal{S}, x_1 \neq x_2, \lambda \in [0,1]$:

$$f(x_1) < f(x_2) \implies f(\lambda x_1 + (1 - \lambda_1)x_2) < f(x_1)$$

or equivalently:

$$f(\lambda x_1 + (1 - \lambda)x_2) < \max\left(f(x_1), f(x_2)\right)$$

Proposition 12 (*Sufficient Conditions for Uniqueness of the Global Minima*)**.** *Assume the objective function ℓ of the* LUE *problem is dependent on a single exogenous attribute and pseudo-convex with respect to the parameter weighting that attribute in the travelers' utility function. If global minima exist, then it is unique.*

Proof. Let us prove this by contradiction. Let $\ell : \mathbb{R} \to \mathbb{R}$ be the objective function of the problem and θ the only coefficient of the utility function. Suppose that there are two global minima $\theta_1^\star, \theta_2^\star : \ell(\theta^1) = \ell(\theta_2^\star) = \ell^\star$ and $\theta_1^\star \neq \theta_2^\star$. By the mean value theorem (MVT), for $\theta_3^\star$, $\exists \theta_3^\star \in (1-\lambda)\theta_1^\star + \lambda\theta_2^\star : \nabla \ell(\theta = \theta_3^\star) = 0,\ \lambda \in [0,1]$ and by Proposition 11, $\ell(\theta_3^\star) = f^\star$. By using MVT recursively, $\forall \theta_i^\star \in [\theta_1^\star, \theta_2^\star]$, $\ell(\theta_i^\star) = \ell^\star$, $\theta_i^\star = (1-\lambda)\theta_1^\star + \lambda\theta_2^\star$, $\lambda \in [0,1]$ and hence, $[\theta_1^\star, \theta_2^\star]$ forms a convex set. By Property 2, Mangasarian (1965), if ℓ is pseudo-convex, ℓ is also strictly quasi-convex. By applying Definition 5 of strict quasi-convexity on θ^1, θ^2 gives $\ell^\star = \ell(\lambda\theta^1 + (1-\lambda)\theta^2) < \max\left(\ell(\theta^1), \ell(\theta^2)\right) = \ell^\star$, which leads to a contradiction and it completes the proof. □

Remark 12. Extending the proofs of global optimality to the multivariate case is left for further research. This extension may require introducing additional assumptions or identifying problem settings where the pseudo-convexity is guaranteed.

A.6. Non-convexity of the LUE problem

Proposition 13 (*Non-Convexity of* LUE *Problem*). *Suppose the coefficients of the travelers' utility function in the* LUE *problem are identifiable. Then, the* LUE *problem is non-convex*

Proof. Let us prove this by contradiction. Assume that the LUE objective function $\ell(\boldsymbol{\theta}) = \|\mathbf{x}(\boldsymbol{\theta}) - \bar{\mathbf{x}}\|_2^2$, $\forall \boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{K}|}$ is convex. Then, let us use the definition of convexity at points $\mathbf{y}_1 = \frac{\theta_1 - (1-\lambda)\theta_2}{\lambda}$ and $\mathbf{y}_2 = \boldsymbol{\theta}_2$, with $\mathbf{y}_1, \mathbf{y}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathbb{R}^{|\mathcal{K}|}$ and $\lambda \in]0,1]$:

$$\ell(\lambda\mathbf{y}_1 + (1-\lambda)\mathbf{y}_2) \leq \lambda\ell(\mathbf{y}_1) + (1-\lambda)\ell(\mathbf{y}_2)$$
$$\ell\left(\lambda\left(\frac{\boldsymbol{\theta}_1 - (1-\lambda)\boldsymbol{\theta}_2}{\lambda}\right) + (1-\lambda)\boldsymbol{\theta}_2\right) \leq \lambda\ell\left(\frac{\boldsymbol{\theta}_1 - (1-\lambda)\boldsymbol{\theta}_2}{\lambda}\right) + (1-\lambda)\ell(\boldsymbol{\theta}_2)$$
$$\ell(\boldsymbol{\theta}_1) \leq \lambda\ell\left(\frac{\boldsymbol{\theta}_1 - (1-\lambda)\boldsymbol{\theta}_2}{\lambda}\right) + (1-\lambda)\ell(\boldsymbol{\theta}_2)$$

which implies that:

$$\ell\left(\frac{\boldsymbol{\theta}_1 - (1-\lambda)\boldsymbol{\theta}_2}{\lambda}\right) \geq \frac{\ell(\boldsymbol{\theta}_1) - (1-\lambda)\ell(\boldsymbol{\theta}_2)}{\lambda} = \frac{\ell(\boldsymbol{\theta}_1) - \ell(\boldsymbol{\theta}_2)}{\lambda} + \ell(\boldsymbol{\theta}_2) \tag{48}$$

If $\boldsymbol{\theta}$ is identifiable, $\exists \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 : \ell(\boldsymbol{\theta}_1) \neq \ell(\boldsymbol{\theta}_2)$. The latter implies that f is not a constant function and that $\exists \lambda \in [0,1] : \ell(\boldsymbol{\theta}_1) > \ell(\boldsymbol{\theta}_2)$. From Eq. (48), when $\ell(\boldsymbol{\theta}_1) > \ell(\boldsymbol{\theta}_2)$, the RHS grows with no bound as $\lambda \to 0^+$. Hence, the function f is not upper-bounded. However, by Proposition 1, the LUE problem is upper-bounded, which generates a contradiction and it completes the proof. □

A.7. Other properties of the LUE problem

Proposition 14. *Assume that the objective function ℓ of the* LUE *problem is dependent on a single exogenous attribute and that ℓ is pseudo-convex with respect to the coefficient weighting that attribute in the travelers' utility function. If a global optimum exists, the negative first derivative of ℓ with respect to the utility function coefficient always points out to the global descent direction*

Proof. The proof follows from leveraging the gradient property of pseudo-convex functions in the unidimensional case. Let $\theta^\star$, be the global minimum of the LUE problem, and $\bar{\theta}$ a feasible point that is not a global optimum. By the pseudo-convexity of $\ell : \mathbb{R} \to \mathbb{R}$:

$$\ell(\theta^\star) < \ell(\bar{\theta}) \implies \frac{\partial \ell(\bar{\theta})}{\partial \theta}(\theta^\star - \bar{\theta}) < 0 \tag{49}$$

Since $\ell(\theta^\star) < \ell(\bar{\theta})$, then $\frac{\partial \ell(\bar{\theta})}{\partial \theta}(\theta^\star - \bar{\theta}) < 0$, which implies that either $\bar{\theta} < \theta^\star \wedge \frac{\partial \ell(\bar{\theta})}{\partial \theta} < 0$ or $\bar{\theta} > \theta^\star \wedge \frac{\partial \ell(\bar{\theta})}{\partial \theta} > 0$. Therefore, if $\bar{\theta} - \theta^\star < 0$, $-\frac{\partial \ell(\bar{\theta})}{\partial \theta} > 0$ and if $\bar{\theta} - \theta^\star > 0$, $-\frac{\partial \ell(\bar{\theta})}{\partial \theta} < 0$, which proves that the negative first derivative of ℓ is always pointing to the global descent direction. □

Definition 6 (*Lipschitzness*). A unidimensional function $f : \mathbb{R} \to \mathbb{R}$ is Lipschitz iff $\forall x, y \in \mathbb{R}$, $G > 0$:

$$|f(x) - f(y)| \leq G\|x - y\| \tag{50}$$

Proposition 15. *Suppose that the travelers' utility function depends on a single exogenous attribute. Then, the objective function of the* LUE *problem is Lipschitz*

Proof. Let us define the objective function of the LUE problem as $\ell : \mathbb{R}^{|\mathcal{K}|} \to \mathbb{R}$. Starting from the LHS of Eq. (50) and using the reverse triangle inequality:

$$\begin{aligned} |\ell(\theta_t^1) - \ell(\theta_t^2)| &= \left| \|\mathbf{x}(\theta_t^1) - \bar{x}\|_2^2 - \|\mathbf{x}(\theta_t^2) - \bar{x}\|_2^2 \right| \\ &\leq \|(\mathbf{x}(\theta_t^1) - \bar{x}) - (\mathbf{x}(\theta_t^2) - \bar{x})\|_2^2 \\ &= \|(\mathbf{x}(\theta_t^1)) - \mathbf{x}(\theta_t^2)\|_2^2 \end{aligned} \tag{51}$$

We can now bound the range of each traffic flow function $x_i, \forall i \in N$ and the norm of the difference between the vector of traffic functions in Eq. (51) as follows:

$$\begin{aligned} 0 \leq x_i(\theta_t) \leq Q \implies -Q &\leq x_i(\theta_t^1) - x_i(\theta_t^2) \leq Q \\ 0 &\leq \left(x_i(\theta_t^1) - x_i(\theta_t^2)\right)^2 \leq Q^2 \\ 0 &\leq \|\mathbf{x}(\theta_t^1) - \mathbf{x}(\theta_t^2)\|_2^2 \leq NQ^2 \end{aligned} \tag{52}$$

where $Q \in \mathbb{R}_+$ is the total demand, namely, the sum of all cells in the O-D matrix $\mathbf{Q} \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$. Now define $|\theta_t^1 - \theta_t^2| = \delta > 0$ and set $G = NQ^2/\delta > 0$, and replace it into Eq. (52):

$$0 \leq \|\mathbf{x}(\theta_t^1) - \mathbf{x}(\theta_t^2)\|_2^2 \leq NQ^2 = G\delta = G|\theta_t^1 - \theta_t^2| \tag{53}$$

which proves that the objective function is G-Lipschitz. □

A.8. Connection between OLS and NLLS

Consider the vectorized version of the NLLS regression equation (Eq. (16), Section 6.1) and let us express it in terms of the response function $\mathbf{x}(\theta) : \mathbb{R}^{|\mathcal{K}|} \to \mathbb{R}^{|\mathcal{A}|}$ defined in Eq. (6), Section 3.3.1. Let us compute the first-order Taylor approximation of the response function with respect to $\theta \in \mathbb{R}^{|\mathcal{K}|}$ and around an arbitrary vector $\boldsymbol{\theta_0} \in \mathbb{R}^{|\mathcal{K}|}$:

$$\mathbf{x}(\theta) \approx \mathbf{x}(\boldsymbol{\theta_0}) + D_\theta\ x(\theta)\Big|_{\boldsymbol{\theta_0}} (\theta - \boldsymbol{\theta_0}) \tag{54}$$

where, for the ease of notation, $\mathbf{x}(\theta) = \mathbf{x}(\mathbf{Z}, \mathbf{t}, \theta)$. Replacing back into Eq. (16), Section 6.1:

$$\begin{aligned} \boldsymbol{x} &\approx x(\boldsymbol{\theta_0}) + D_\theta\ x(\theta)\Big|_{\theta_0} (\theta - \theta_0) + \boldsymbol{u} \\ \boldsymbol{x} - x(\boldsymbol{\theta_0}) + D_\theta\ x(\theta)\Big|_{\theta_0} \boldsymbol{\theta_0} &\approx D_\theta\ x(\theta)\Big|_{\theta_0} \boldsymbol{\theta} + \boldsymbol{u} \\ \tilde{\boldsymbol{x}} &\approx \tilde{\mathbf{X}}\boldsymbol{\theta} + \boldsymbol{u} \end{aligned} \tag{55}$$

From where is clear that Eq. (55) resembles the OLS regression equation, except that $\tilde{\mathbf{X}} = D_\theta\ x(\theta)\Big|_{\boldsymbol{\theta_0}}$.

Appendix B. Assumptions

B.1. Non-linear least squares (NLLS) estimator

Assumption 6 (*Model Specification*). The response function is well-specified, i.e. $\forall \theta_0 \in \mathbb{R}^{|\mathcal{K}|},\ \boldsymbol{x} = \mathbf{x}(\theta_0, \mathbf{Z}, \bar{\mathbf{t}}) + \mathbf{u}$

Assumption 7 (*Orthogonality Between Errors and Regressors*). In the data generating process, $\mathbb{E}[\mathbf{u}|\mathbf{Z}, \bar{\mathbf{t}}] = \mathbf{0}$ and $\mathbb{E}[\mathbf{u}\mathbf{u}^\top|\mathbf{Z}, \bar{\mathbf{t}}] = \boldsymbol{\Omega}_0$, where $\boldsymbol{\Omega}$ is the covariance matrix of the errors terms

Remark 13. Assumption 7 allows for heteroscedastic errors in the NLLS problem but requires knowing or estimating a functional form of the covariance matrix $\boldsymbol{\Omega}_0$ (Cameron and Trivedi, 2005). To ease the analysis, we introduce an additional assumption:

Assumption 8 (*Spherical Errors*). Errors are homoscedastic and non-autocorrelated, that is, their covariance matrix is diagonal $\boldsymbol{\Omega}_0 = \sigma^2\mathbf{I}$, for $\sigma \in \mathbb{R}_{\geq 0}$

Remark 14. Assumption 8 of spherical errors is satisfied if the errors are independent and identically distributed (*iid*) and with a constant variance

Assumption 9 (*Identifiability*). Each traffic flow function $x_i(\cdot)$ satisfies that: $\forall i \in N, \forall \theta_1, \theta_2 \in \mathbb{R}^{|\mathcal{K}|}$, $x_i(\theta_1, \mathbf{Z}, \bar{\mathbf{t}}) = x_i(\theta_2, \mathbf{Z}, \bar{\mathbf{t}})$ iff $\theta_1 = \theta_2$

Remark 15. Assumption 9 is directly satisfied when θ and the traffic flow functions are (coordinate-wise) monotonic (see Definition 4, Section 4.1.2).

Assumption 10 (*Rank of the Limit Distribution of the Hessian Matrix Approximation*).

The matrix

$$\mathbf{A}_0 = \operatorname{plim} \frac{1}{N} \left[D_\theta\ x(\theta)\Big|_{\theta_0} \right]^\top \left[D_\theta\ x(\theta)\Big|_{\theta_0} \right]$$

exists and is finite and nonsingular $\forall \theta_0 \in \mathbb{R}^{|\mathcal{K}|}$.

Remark 16. Assumption 10 formalizes the third rule of identifiability discussed in Appendix D.2. Proposition 16 shows that this assumption can be tested via the rank of the Jacobian of the traffic functions:

Proposition 16 (*Full Rank Jacobian and Non-Singular Hessian Matrix Approximation*). *Suppose the Jacobian matrix of the traffic flow functions given by* $D_\theta\ \mathbf{x}(\theta) \in \mathbb{R}^{n\times d}$*, with* $n > d$*, is full rank at* $\theta_0 \in \mathbb{R}^{|\mathcal{K}|}$*. Then, the square matrix* $\left[D_{\theta=\theta_0}\ \mathbf{x}(\theta)\right]^\top D_{\theta=\theta_0}\ \mathbf{x}(\theta) \in \mathbb{R}^{d\times d}$ *is positive definite.*

Proof. Consider $\forall \mathbf{u} \in \mathbb{R}^d \neq \mathbf{0}$:

$$\mathbf{u}^T \left(2\left[D_\theta\ \mathbf{x}(\theta)\right]^\top D_\theta\ \mathbf{x}(\hat{\theta})\right)\ \mathbf{u} = 2(D_{\theta=\theta_0}\ \mathbf{x}(\theta)\mathbf{u})^\top \left(D_{\theta=\theta_0}\ \mathbf{x}(\theta)\ \mathbf{u}\right) = \|D_{\theta=\theta_0}\mathbf{x}(\theta)\mathbf{u}\|_2^2$$

By assumption, the Jacobian matrix $D_{\theta=\theta_0}\ \mathbf{x}(\theta)$ is full rank at θ_0 and thus $D_{\theta=\theta_0}\ \mathbf{x}(\theta)\mathbf{u} = \mathbf{0}$ iff $\mathbf{u} = \mathbf{0}$. Since $\mathbf{u} \neq 0$, $\|D_{\theta=\theta_0}\mathbf{x}(\theta)\mathbf{u}\|_2^2 > 0$ which proves that the matrix $\left[D_{\theta=\theta^\star}\ \mathbf{x}(\theta)\right]^\top D_{\theta=\theta_0}\ \mathbf{x}(\theta)$ is positive definite and completes the proof. □

Thus, if the Jacobian of the traffic flow functions $D_\theta\ x(\theta)$ has full rank, by Proposition 16, the matrix specified in the argument of the probability limit defined in Assumption 10 is positive definite and hence non-singular.

Assumption 11 (*Central Limit Theorem*). $N^{-1/2}\sum_{i=1}^N [D_\theta\ x_i(\theta)]^\top u_i \xrightarrow{d} \mathcal{N}[0, \mathbf{B}_0]$, where

$$\mathbf{B}_0 = \operatorname{plim}\frac{1}{N}\left[D_\theta\ x(\theta)\Big|_{\theta_0}\right]^\top \boldsymbol{\Omega}_0 \left[D_\theta\ x(\theta)\Big|_{\theta_0}\right] = \operatorname{plim}\frac{\sigma^2}{N}\left[D_\theta\ x(\theta)\Big|_{\theta_0}\right]^\top \left[D_\theta\ x(\theta)\Big|_{\theta_0}\right]$$

and $\boldsymbol{\Omega}_0 = \sigma^2\mathbf{I}$ by Assumption 8 of homoscedasticity.

Remark 17. When the errors $\mathbf{u}$ are Gaussian and independently distributed, the linear combination of errors given by $[D_\theta\ x(\theta)]^\top u$ is a Multivariate Gaussian. As a result, Assumption 11 holds even in small samples. Alternatively, if each term $[D_\theta\ x_i(\theta)]^\top u_i$ follows an arbitrary distribution, the central limit theorem (CLT) must be invoked to ensure the asymptotic normality of the terms in the sequence $\{D_\theta\ x_i(\theta)\}_{i=1\ldots N}$. Since $D_\theta\ x_i(\theta)$ is assumed exogenous for statistical inference, the distribution of each term i of the previous sequence is determined by the distribution of u_i.

Appendix C. Algorithms

C.1. Stochastic network loading

Algorithm 2 Stochastic network loading (SNL)

Input: $\theta \in \mathbb{R}^{|\mathcal{K}|}$, incident matrices $\mathbf{M}, \mathbf{D}$, non-sparse O-D vector $\mathbf{q}$, vector of link travel times $\bar{\mathbf{t}}$, matrix of exogenous link attributes $\mathbf{Z}$

a: Travel time initialization: If $\bar{\mathbf{t}} = \emptyset$, then $\bar{\mathbf{t}} = \bar{\mathbf{t}}_f$, where $\bar{\mathbf{t}}_f$ is the vector of links' free flow travel times.

b: Computation of link utilities

$$\mathbf{v}_x \leftarrow \theta_t\bar{\mathbf{t}} + \theta_Z^\top\mathbf{Z}$$

c: Computation of path choice probabilities:

$$\mathbf{p} \leftarrow \exp\left(\mathbf{D}^\top\mathbf{v}_x\right) \oslash \left(\mathbf{M}^\top\mathbf{M}\exp(\mathbf{D}^\top\mathbf{v}_x)\right)$$

d: Computation of path flows:

$$\mathbf{f} \leftarrow (\mathbf{M}^\top\mathbf{q})\circ\mathbf{p}$$

e: Computation of link flows:

$$\mathbf{x} \leftarrow \mathbf{D}\mathbf{f}$$

return $\mathbf{x}, \mathbf{f}, \mathbf{p}$

where $\oslash$ is the operator for element-wise division, $\mathbf{v}_x \in \mathbb{R}^{|\mathcal{A}|}$ is the vector of link utilities, $\mathbf{p} \in \mathbb{R}^{|\mathcal{H}|}_{]0,1[}$ is a vector of path choice probabilities and $\bar{t}$ is the vector of travel times which is assumed to be exogenous during SNL. Note that the definition of $\mathbf{p}$ in the first step of SNL correspond to the vectorized form of the path flows at SUELOGIT presented in Eq. (3) and which is written as a function of link utilities and the network incidence matrices $\mathbf{D}, \mathbf{M}$.

C.2. Inner level optimization

Algorithm 3 `InnerLevelOptimization`

Input: # Iterations T, initial vector of estimated coefficients $\hat{\theta} \in \mathbb{R}^{|\mathcal{K}|}$, matrix of exogenous attributes $\mathbf{Z}$, vector of links' free flow travel times $\mathbf{t}^0$, vector of link capacities $\boldsymbol{\gamma}$, incidence matrices $\mathbf{M}, \mathbf{D}$, dense O-D vector $\mathbf{q}$, grid of values $\boldsymbol{\lambda}_{FW} \in \mathbb{R}^{|\mathcal{K}|}$ in Frank-Wolfe algorithm, the proportion of selected O-D pairs in column generation phase ρ_W, the proportion of generated and selected paths in column generation phase k_g, k_s:

Step 0: Initialization.
a: Compute initial vector of link utilities $\mathbf{v}_x \leftarrow \theta_t \bar{\mathbf{t}}_f + \hat{\theta}_Z^\top \mathbf{Z}$, where $\hat{\theta} = [\theta_t \ \theta_Z^\top]$
b: Generate k–shortest paths $\mathcal{S}_{pq}, \ \forall (p,q) \in \mathcal{W}$ based on link utilities $\mathbf{v}_x$
c: Compute incident matrices $\mathbf{M}, \mathbf{D}$ from $\mathcal{S}_{pq}$
d: Perform stochastic network loading: $\mathbf{x}^{(0)}, \mathbf{f}^{(0)} \leftarrow \mathtt{SNL}(\hat{\theta}, \mathbf{M}, \mathbf{D}, \bar{\mathbf{t}})$
e: $i = 0$.

Step 1: Column generation phase.
a: Select subset of O-D pairs $\mathcal{W}_s \in \mathcal{W}$ with the highest travel demand, such that $|\mathcal{W}_s| = \lceil |\mathcal{W}| \times \rho_W \rceil$[2]
b: Generate set of the k_g–shortest paths $\mathcal{C}_{pq}, \forall (p,q) \in \mathcal{W}_s$ based on current link utilities $\mathbf{v}_x$
c: Update path sets $\mathcal{S}_{pq} \leftarrow \mathcal{C}_{pq}, \forall (p,q) \in \mathcal{W}_s$
d: Update incidents matrices $\mathbf{M}, \mathbf{D}$ from $\mathcal{S}_{pq}, \forall p,q \in \mathcal{W}$

Step 2: `SUELOGIT`
a: Perform stochastic network loading `SNL` (algorithm 2):

$$\mathbf{x}, \mathbf{f}, \mathbf{p} \leftarrow \mathtt{SNL}(\hat{\theta}, \mathbf{M}, \mathbf{D}, \mathbf{q}, \bar{\mathbf{t}}, \mathbf{Z})$$

b: Update link travel times

$$\bar{\mathbf{t}} \leftarrow \bar{\mathbf{t}}^0 (1 + \alpha(\mathbf{x}/\boldsymbol{\gamma})^\beta)$$

c: Solve a linear search problem:

for $m = 1$ to $|\boldsymbol{\lambda}_{FW}|$ **do**
$\tilde{\mathbf{f}}^{(m)} \leftarrow \lambda_i \mathbf{f}^{(i-1)} + (1 - \lambda_i) \mathbf{f}^{(i)}$
$\tilde{\mathbf{x}}^{(m)} \leftarrow \mathbf{D} \tilde{\mathbf{f}}^{(m)}$
$\ell^{(m)} \leftarrow \sum_{a \in \mathcal{A}} \int_0^{\tilde{x}_a} v_a(u, \hat{\theta}) du - \left\langle \tilde{\mathbf{f}}, \ln \tilde{\mathbf{f}} \right\rangle$
$\lambda^{(m^\star)} = \arg\min_{\lambda \in \boldsymbol{\lambda}_{FW}} \ell(\lambda)$

d: Update path and link flow solutions, travel times, and path choice probabilities

$$\mathbf{f}^{(i)} \leftarrow \lambda^{(m^\star)} \mathbf{f}^{(m^\star - 1)} + (1 - \lambda^{m^\star}) \mathbf{f}^{(m^\star)}$$
$$\mathbf{x}^{(i)} \leftarrow \mathbf{D} \mathbf{f}^{(i)}$$
$$\bar{\mathbf{t}} \leftarrow \bar{\mathbf{t}}^0 (1 + \alpha(\mathbf{x}/\boldsymbol{\gamma})^\beta)$$
$$\mathbf{p} \leftarrow \exp\left(\mathbf{D}^\top \mathbf{v}_x\right) \oslash \left(\mathbf{M}^\top \mathbf{M} \exp(\mathbf{D}^\top \mathbf{v}_x)\right)$$

e: Go to step 3 if the desired level of accuracy has been achieved.[3] Otherwise, $i = i + 1$ and start again from Step 2

Step 3: Paths selection
a: Select the k_s–shortest paths $\mathcal{R}_{pq}, \forall p,q \in \mathcal{W}_s$
b: Update path sets $\mathcal{S}_{pq} \leftarrow \mathcal{R}_{pq}, \forall p,q \in \mathcal{W}_s$
c: Update incidents matrices $\mathbf{M}, \mathbf{D}$ from $\mathcal{R}_{pq}, \forall p,q \in \mathcal{W}$

return $\mathbf{x}, \mathbf{p}, \bar{\mathbf{t}}$

C.3. Outer level optimization

C.3.1. Algorithm

C.3.2. Gradients and Jacobian

The first derivative of the objective function $\ell(\cdot)$ with respect to a utility function coefficient $\hat{\theta}_k, \forall k \in \mathcal{K}$ can be written in vectorized form as:

$$\frac{\partial \ell(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_k} = \frac{\partial}{\partial \hat{\theta}_k} \left\| \mathbf{x}(\hat{\theta}) - \bar{\mathbf{x}} \right\|_2^2 = 2 \left(\frac{\partial \mathbf{x}(\hat{\theta})}{\partial \hat{\theta}_k} \right)^\top \left(\mathbf{x}(\hat{\theta}) - \bar{\mathbf{x}} \right) \tag{56}$$

[2] The O-D pairs selected at an iteration exclude the O-D pairs selected at the previous iteration
[3] For example, when the relative decrease of the objective function given by $1 - \ell^{(i)}/\ell^{(i-1)}$ is lower than a threshold

Algorithm 4 `OuterLevelOptimization`

Input: # Iterations in no-refined and refined stages T_1, T_2, initial vector of estimated coefficients $\hat{\boldsymbol{\theta}}_0 \in \mathbb{R}^{|\mathcal{K}|}$, link flows $\mathbf{x}$, link travel times $\tilde{\mathbf{t}}$ and path choice probabilities $\mathbf{p}$ from inner-level problem, vector of observed traffic counts $\bar{\mathbf{x}}$, learning rate η for `NGD` method, dumping parameter for `LM` method δ_{LM}

Step 1: no-refined stage
for $t = 1 \dots T_1$ **do**
  $\hat{\theta}_{t+1} \leftarrow$ `FirstOrderOptimization`($\hat{\theta}$, η, $\bar{\mathbf{x}}$, $\mathbf{x}$, $\mathbf{p}$, $\tilde{\mathbf{t}}$, $T = 1$)
end for

Step 2: refined stage
for $t = 1 \dots T_2$ **do**
  $\hat{\theta}_{T_1+t+1} \leftarrow$ `SecondOrderOptimization`($\hat{\theta}_{T_1+t}$, δ_{LM}, $\bar{\mathbf{x}}$, $\mathbf{x}$, $\mathbf{p}$, $\tilde{\mathbf{t}}$, $T = 1$)
end for

return $\bar{\hat{\theta}}_T = \operatorname{argmin}_{\{\hat{\theta}_1,\dots,\hat{\theta}_{T_1+T_2}\}} \ell_t(\hat{\theta}_t)$

where $\mathbf{x}(\cdot)$ is a vector-valued response function that receives as input a vector with the utility function coefficients and it returns a vector $\hat{\theta} \in \mathbb{R}^{|\mathcal{K}|}$ with the predicted traffic counts among all links in the transportation network:

$$\mathbf{x}(\hat{\boldsymbol{\theta}}) = \mathbf{D}\left((\mathbf{M}^\top \mathbf{q}) \circ \mathbf{p}(\hat{\boldsymbol{\theta}})\right) \tag{57}$$

and $\mathbf{p}(\hat{\theta})$ is a vector-valued function that receives as input a vector with the utility function coefficients and it returns a vector with the choice probabilities associated with all paths in the transportation network:

$$\mathbf{p}(\hat{\boldsymbol{\theta}}) = \exp\left(\mathbf{D}^\top \mathbf{v}_x(\hat{\boldsymbol{\theta}}, \mathbf{Z}, \bar{t})\right) \oslash \left(\mathbf{M}^\top \mathbf{M} \exp(\mathbf{D}^\top \mathbf{v}_x(\hat{\boldsymbol{\theta}}, \mathbf{Z}, \bar{\mathbf{t}}))\right) \tag{58}$$

The first derivative of $\mathbf{x}(\cdot)$ with respect to a utility function coefficient $\hat{\theta}_k$ can be written in vectorized form as:

$$\frac{\partial \mathbf{x}(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_k} = \mathbf{D}\left((\mathbf{M}^\top \mathbf{q}) \circ \frac{\partial \mathbf{p}(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_k}\right) \tag{59}$$

where

$$\frac{\partial \mathbf{p}(\hat{\theta})}{\partial \theta_k} = \left((\mathbf{M}^\top \mathbf{M}) \circ \left(\mathbf{p}(\hat{\theta})\mathbf{p}(\hat{\theta})^\top\right) \circ \left[\mathbf{z}_k \mathbf{1}^\top_{|\mathbf{z}_k|} - \mathbf{1}_{|\mathbf{Z}_k|} \mathbf{z}_k^\top\right]\right) \mathbf{1}_{|\mathbf{Z}_k|} \tag{60}$$

is a vector with the first derivatives of the path choice probabilities with respect to the utility function coefficient $\hat{\theta}_k$. $\mathbf{z}_k \in \mathbb{R}^{|\mathcal{H}|}$ is a column vector with the values of attribute $d \in \mathcal{K}$, including travel time, among all paths and $\mathbf{1}_{|\mathbf{Z}_k|} \in \mathbb{R}^{|\mathcal{H}|}$ is a column vector with ones. Then, the gradient $\nabla_{\hat{\theta}}\, \mathbf{x}(\hat{\theta})$ associated with a traffic flow function $i \in \mathcal{A}$ is obtained by stacking the first derivatives in a column vector as follows:

$$\nabla_{\hat{\theta}}\, x_i(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \dfrac{\partial x_i(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_1} & \dots & \dfrac{\partial x_i(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_{\mathcal{K}}} \end{bmatrix}^\top \tag{61}$$

and the Jacobian matrix $D_{\hat{\theta}}\, \mathbf{x}(\hat{\theta}) \in \mathbb{R}^{|N| \times |\mathcal{K}|}$ associated with the traffic flow functions correspond to the stacked gradient vectors of each observation $n \in N$:

$$D_{\hat{\theta}}\, x(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \nabla_{\hat{\theta}}\, x_1(\hat{\boldsymbol{\theta}})^\top \\ \vdots \\ \nabla_{\hat{\theta}}\, x_{|N|}(\hat{\boldsymbol{\theta}})^\top \end{bmatrix} = \begin{bmatrix} \dfrac{\partial x_1(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_1} & \cdots & \dfrac{\partial x_1(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_{|\mathcal{K}|}} \\ \vdots & & \\ \dfrac{\partial x_{|N|}(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_1} & \cdots & \dfrac{\partial x_{|N|}(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_{|\mathcal{K}|}} \end{bmatrix}^\top \tag{62}$$

Finally, the analytical gradient of the objective function with respect to the vector of utility function coefficients is :

$$\nabla_{\hat{\theta}}\, \ell(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \dfrac{\partial \ell(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_1} & \dots & \dfrac{\partial \ell(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_{|\mathcal{K}|}} \end{bmatrix}^\top \tag{63}$$

and where eack element $k \in \mathcal{K}$ is:

$$\frac{\partial \ell(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_k} = 2\left(\frac{\partial \mathbf{x}(\hat{\boldsymbol{\theta}})}{\partial \hat{\theta}_k}\right)^\top \left(\mathbf{x}(\hat{\boldsymbol{\theta}}) - \bar{\mathbf{x}}\right) \tag{64}$$

$$= 2\left(\mathbf{D}\left((\mathbf{M}^\top\mathbf{q})\circ\frac{\partial\mathbf{p}(\hat{\theta})}{\partial\hat{\theta}_k}\right)\right)^\top\left(\mathbf{x}(\hat{\theta})-\bar{\mathbf{x}}\right)$$

$$= 2\left(\mathbf{D}\left((\mathbf{M}^\top\mathbf{q})\circ\left((\mathbf{M}^\top\mathbf{M})\circ\left(\mathbf{p}(\hat{\theta})\mathbf{p}(\hat{\theta})^\top\right)\circ\left[\mathbf{z}_k\mathbf{1}_{|\mathbf{z}_k|}^\top-\mathbf{1}_{|\mathbf{z}_k|}\mathbf{z}_k^\top\right]\right)\mathbf{1}_{|\mathbf{z}_k|}\right)\right)^\top\left(\mathbf{x}(\hat{\theta})-\bar{\mathbf{x}}\right)$$

is a scalar.

C.3.3. Second derivatives

The second derivative of the objective function $\ell(\cdot)$ with respect to a utility function coefficient $\hat{\theta}_k, \forall k \in \mathcal{K}$ can be written in vectorized form as:

$$\frac{\partial^2\ell(\hat{\theta})}{\partial^2\hat{\theta}_k} = \frac{\partial}{\partial\hat{\theta}_k}\left(\frac{\partial\ell(\hat{\theta})}{\partial\hat{\theta}_k}\right) = 2\left(\left(\frac{\partial^2\mathbf{x}(\hat{\theta})}{\partial^2\hat{\theta}_k}\right)^\top\left(\mathbf{x}(\hat{\theta})-\bar{\mathbf{x}}\right)+\left(\frac{\partial\mathbf{x}(\hat{\theta})}{\partial\hat{\theta}_k}\right)^\top\frac{\partial\mathbf{x}(\hat{\theta})}{\partial\hat{\theta}_k}\right) \tag{65}$$

where

$$\frac{\partial^2\mathbf{x}(\hat{\theta})}{\partial^2\hat{\theta}_k} = \mathbf{D}\left((\mathbf{M}^\top\mathbf{q})\circ\frac{\partial^2\mathbf{p}(\hat{\theta})}{\partial^2\hat{\theta}_k}\right) \tag{66}$$

and

$$\frac{\partial^2\mathbf{p}(\hat{\theta})}{\partial^2\hat{\theta}_k} = (\mathbf{M}^\top\mathbf{M})\circ\left[\left(\frac{\partial\mathbf{p}(\hat{\theta})}{\partial\hat{\theta}_k}\right)\mathbf{p}(\hat{\theta})^\top+\mathbf{p}(\hat{\theta})\left(\frac{\partial\mathbf{p}(\hat{\theta})}{\partial\hat{\theta}_k}\right)^\top\right]\circ\left[\mathbf{z}_k\mathbf{1}_{|\mathbf{Z}_k|}^\top-\mathbf{1}_{|\mathbf{z}_k|}\mathbf{Z}_k^\top\right]\mathbf{1}_{|\mathbf{z}_k|} \tag{67}$$

is a vector of dimension $|\mathcal{H}|$ with the second derivatives of the path choice probabilities with respect to a utility function coefficient $\hat{\theta}_k$.

C.3.4. First-order and second-order optimization methods

Algorithms 5 and 6 show the pseudo code to implement the first-order and second-order optimization algorithms.

Algorithm 5 `FirstOrderOptimization`

Input: # Iterations T, initial vector of estimated coefficients $\hat{\theta}_0 \in \mathbb{R}^{|\mathcal{K}|}$, incident matrices $\mathbf{M}, \mathbf{D}$, non-sparse O-D vector $\mathbf{q}$, path choice probabilities $\mathbf{p}$, link flows $\mathbf{x}$, dense vector with O-D demand $\mathbf{q}$, matrix of exogenous link attributes $\mathbf{Z}$, vector of link travel times $\tilde{\mathbf{t}}$, observed traffic counts $\bar{\mathbf{x}}$, learning rate η

for $t = 1 \dots T$ **do**

Compute gradient of the objective function (Appendix C.3.2):

$$\mathbf{g}_t := \nabla_{\hat{\theta}}\ell(\hat{\theta}, \mathbf{x}, \mathbf{q}, \mathbf{p}, \bar{\mathbf{x}}, \tilde{\mathbf{t}}, \mathbf{Z}, \mathbf{M}, \mathbf{D})$$

Normalization of gradient

$$g_t \leftarrow \frac{g_t}{\|g_t\|}$$

Solution update:

$$\hat{\theta}_{t+1} = \hat{\theta}_t - \eta g_t$$

end for

return $\bar{\hat{\theta}}_T = \text{argmin}_{\{\hat{\theta}_1,\dots,\hat{\theta}_T\}}\,\ell_t(\hat{\theta}_t)$

Algorithm 6 `SecondOrderOptimization`

Input: # Iterations T, initial vector of estimated coefficients $\hat{\theta}_0 \in \mathbb{R}^{|\mathcal{K}|}$, incident matrices $\mathbf{M}, \mathbf{D}$, non-sparse O-D vector $\mathbf{q}$, path choice probabilities $\mathbf{p}$, link flows $\mathbf{x}$, dense vector with O-D demand $\mathbf{q}$, matrix of exogenous link attributes $\mathbf{Z}$, vector of link travel times $\tilde{\mathbf{t}}$, observed traffic counts $\bar{\mathbf{x}}$, dumping parameter δ_{LM} for `LM` method

for $t = 1 \dots T$ **do**

Compute Jacobian of the traffic flow functions (Appendix C.3.2):

$$J_t := D_{\hat{\theta}}\ \mathbf{x}(\hat{\theta}, \mathbf{x}, \mathbf{q}, \mathbf{p}, \tilde{\mathbf{t}}, \mathbf{Z}, \mathbf{M}, \mathbf{D})$$

if `method` = `GN` **then**

$$\Delta\hat{\theta}_t \leftarrow (J_t^\top J_t)^{-1} J_t^\top (\bar{\mathbf{x}} - \mathbf{x})$$

end if

if `method` = `LM` **then**

$$\Delta\hat{\theta}_t \leftarrow (J_t^\top J_t + \delta_{LM}\ I_{d\times d})^{-1} J_t^\top (\bar{\mathbf{x}} - \mathbf{x})$$

end if

Update:

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \Delta\hat{\theta}_t$$

end for

return $\bar{\hat{\theta}}_T = \text{argmin}_{\{\hat{\theta}_1,\dots,\hat{\theta}_T\}}\ \ell_t(\hat{\theta}_t)$

Appendix D. Analyses in small networks

D.1. Illustrative example

Consider a network with two parallel links and thus, with two alternative paths only (Fig. 14). The costs functions associated to each link/path are $t_1(x_1) = t_1^0(1 + \alpha(\frac{x_1}{\gamma_1})^\beta)$ and $t_2(x_2) = t_2^0(1 + \alpha(\frac{x_2}{\gamma_2})^\beta)$, where γ_1, γ_2 are the link capacities and $t_1^0 = t_2^0$ are the links' free flow travel times. The parameters of the link performance functions are $\alpha = 0.15$ and $\beta = 4$. Assume that the travelers' utility function is dependent on the travel time t and the monetary cost c of traversing each link of a path only. In addition, suppose that the coefficients θ_t, θ_c are linearly weighting each attribute of the utility function and that they are common among travelers. Thus, the deterministic component of the utility attained to each link/path can be expressed as $v_1(x_1) = \theta_t t_1(x_1) + \theta_c c_1$ and $v_2(x_2) = \theta_t t_2(x_2) + \theta_c c_2$.

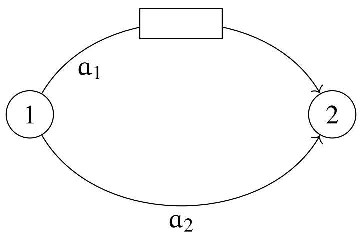


Fig. 14. Illustrative example.

Consider $\bar{q} \in \mathbb{R}_+$ individuals traveling between origin–destination 1–2 and making route choices consistent with a logit model. $\bar{q}$ is small enough such that links' travel times are approximately equal to their free flow travel times, i.e. $t_1(x_1) \approx t_1^0,\ t_2(x_2) \approx t_2^0$. Suppose link a_1 has a toll fee equal to 1 USD, i.e. $c_1 = 1, c_2 = 0$, and a sensor recording traffic counts. Then, path choice probabilities can be modeled with a sigmoid function and link flows can be obtained as follows:

$$x_1 = \frac{\exp(\theta_t t_1 + \theta_c c_1)}{\exp(\theta_t t_1 + \theta_c c_1) + \exp(\theta_t t_2 + \theta_c c_2)}\ \bar{q} = \sigma(\theta_c)\bar{q}, \quad x_2 = \bar{q} - x_1 = (1 - \sigma(\theta_c))\bar{q} = \sigma(-\theta_c)\bar{q} \tag{68}$$

Suppose the goal is to estimate the coefficients θ_t and θ_c of the travelers' utility function with the traffic count measurement $\bar{x}_1$. From the link flow solutions (Eq. (68)), we note first that θ_t is not identifiable because travel times are the same in the two alternative paths. In contrast, because the monetary costs are different between paths, θ_c is identifiable and equal to the solution $\theta_c^\star$ of the following nonlinear least square minimization problem:

$$\theta_c^\star = \arg\min_{\theta_c} (x_1(\theta_c) - \bar{x}_1)^2 = \arg\min_{\theta_t} \left(\sigma(\theta_c) - \frac{\bar{x}_1}{\bar{q}}\right)^2 \rightarrow \theta_c^\star = \ln(\bar{q}/\bar{x}_1 - 1) \tag{69}$$

Consider two scenarios for the solution of this problem (Eq. (69)). In the first scenario $\bar{x}_1 = 0.5\bar{q}$ meaning that $\bar{x}_1$ is deterministic and that $\theta_c^\star = 0$. In the second scenario, $\bar{x}_1 = x_1 + \epsilon$, $\epsilon \sim \mathcal{N}(0,1)$, i.e. $\bar{x}_1$ is not deterministic and $\bar{x}_1 \sim \mathcal{N}(0.5\bar{q}, 1)$. Note that a single realization of $\bar{x}_1$ in the second scenario could lead us to conclude with the same probability that the traveler's utility and the monetary cost are positively or negatively associated. However, in expectation and similar to the first scenario, this association does not exist.

This example illustrates important features of the problem studied in this paper. First, it shows that the non-linear least squares formulation of the problem is suitable to estimate the coefficients of the travelers' utility function with traffic count data. Second, the identification issue associated with θ_t illustrates the special considerations that must be made to specify the travelers' utility function and to prevent a naive estimation of coefficients. Third, the estimation of θ_c in the non-deterministic scenario illustrates the relevance of statistical inference tools to assess if a feature is a determinant of the travelers' route choices in the presence of randomness of the data-generating process.

In real-world transportation networks, the analytical derivation of the optimal solution for travelers' utility function parameters is intractable. In addition, the assumption that travel times at equilibrium are equal to free-flow travel times is unrealistic. Thus methods that can solve for network equilibrium in large-scale networks and that can iteratively approach the optimal solution for the utility function coefficients are needed. Besides, due to the class of non-linearity of the link flows with respect to the utility function coefficients, the objective function cannot be guaranteed to be convex, which imposes additional difficulties to search for an optimal solution. Another challenge is related to the impact of traffic congestion on travelers' choices and which is not addressed in the illustrative example.

D.2. Identifiability of utility function coefficients

The use of the MNL model in our problem imposes rules of identifiability that are similar to those used in discrete choice models. For the computation of the logit choice probabilities, differences in utility among the alternatives is what matters (Walker, 2002). Therefore, if an attribute of the paths' utility function is the same within each path set, then the coefficient weighting that attribute will be, by construction, not identifiable. From a behavioral point of view, travelers would perceive no difference in utilities when facing a path choice decision, and thus, their choices will be not informative about the strength of their preference for that attribute. From an optimization standpoint, there will be no gradient with respect to the attribute's coefficient because the choice probabilities are invariant to changes in the value of that coefficient.

A second rule of identifiability refers to the inclusion of alternative specific constants, which may lead to an over-specification of the utility function. The practical implication of our problem is that the specific constant in one of the paths connecting each O-D pair should be fixed, e.g. setting a constant to zero. A third rule of identifiability is related to the minimum amount of traffic counts measurements required to estimate a certain number of coefficients in the utility function, which can be referred to as empirical identification (Walker, 2002). Similar to the rank condition in OLS, the number of observations should be higher or equal to the number of parameters but, in practice, a larger amount of data may be needed due to the existence of collinearity between observations.

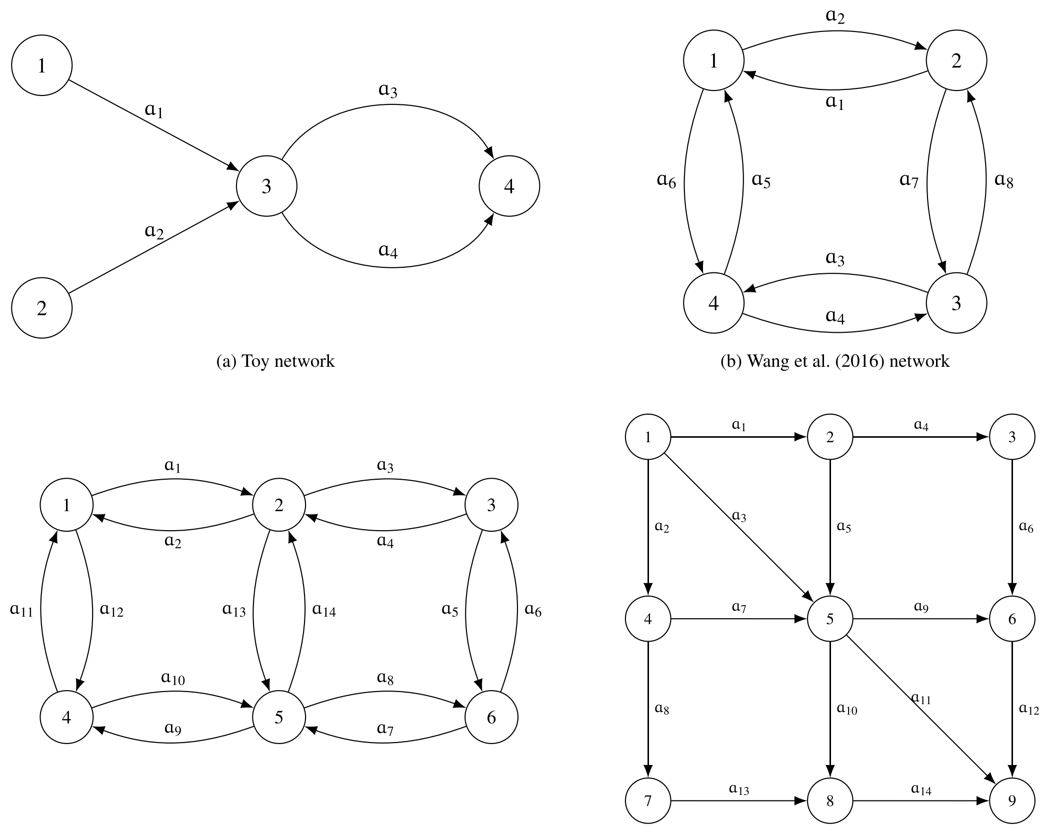
To show the relevance of the identification rules, let us consider the network in Fig. 14 and suppose that there are link flow measurements available for the two links. Assume the network is uncongested, $t_1(x_1) \approx t_1^0$, $t_2(x_2) \approx t_2^0$, and that the observed link counts perfectly matched the true link flow solution at equilibria, i.e. $(\bar{x}_1, \bar{x}_2) = (x_1^\star, x_2^\star)$, $\bar{x}_1 + \bar{x}_2 = \bar{q}$. From Eq. (68), Appendix D.1, we note that the link flow solution of the inner level problem can be written in closed form as:

$$x_1^\star = \frac{\bar{q}}{1 + \exp\left(\theta_t(t_1^0 - t_2^0) + \theta_c(c_1 - c_2)\right)}, \quad x_2^\star = \frac{\bar{q}}{1 + \exp\left(\theta_t(t_2^0 - t_1^0) + \theta_c(c_2 - c_1)\right)} \tag{70}$$

which satisfies both the conservation constraints between travel demand and path flows and the non-negativity constraints of link flows (Eq. 2, Appendix D.1) for $\forall \theta_t, \theta_c \in \mathbb{R}$. From the two link count measurements, we can derive a system of two equations that are dependent on the two unknown coefficients of the utility function:

$$\begin{aligned}
\theta_t(t_1(\bar{x}_1) - t_2(\bar{x}_2)) + \theta_c(c_1 - c_2) &= \ln\left(\frac{\bar{q}}{\bar{x}_1} - 1\right) = \ln\left(\frac{\bar{x}_1 + \bar{x}_2}{\bar{x}_1} - 1\right) = (\ln(\bar{x}_2) - \ln(\bar{x}_1)) \\
\theta_t(t_2(\bar{x}_1) - t_1(\bar{x}_2)) + \theta_c(c_2 - c_1) &= \ln\left(\frac{\bar{q}}{\bar{x}_2} - 1\right) = \ln\left(\frac{\bar{x}_1 + \bar{x}_2}{\bar{x}_2} - 1\right) = -(\ln(\bar{x}_1) - \ln(\bar{x}_2))
\end{aligned} \tag{71}$$

Note that in line with the third rule of empirical identification, the two traffic count measurements would suffice to identify the two parameters. However, the two equations in Eq. (71) are identical, hence linearly dependent, and thus, one equation is not providing additional information to identify the two coefficients of the utility function. Besides, from the first identification rule, Eq. (71) cannot be solved for θ_c or θ_t if the difference of travel time and the cost is zero (see Appendix D.1).



(a) Toy network (b) Wang et al. (2016) network (c) Lo and Chan (2003) network (d) Yang et al. (2001) network

Fig. 15. Topologies of small networks.

Remark 18. The objective function of the optimization problem may be globally minimized at a point that is not attainable or where the gradient does not vanish. Suppose that the traffic counts are consistent with UE, such that $(\bar{x}_1, \bar{x}_2) = (0, \bar{q})$ or $(\bar{x}_1, \bar{x}_2) = (\bar{q}, 0)$. With utility function dependent on travel time only, these link flow measurements are consistent with the limiting cases where $\theta_t \to -\infty$ or $\theta_t \to \infty$, respectively. While the identification rules are satisfied, the data-generating process does not follow SUELOGIT If this occurs in practice, the optimization algorithm chosen to fit the travel time coefficient is expected to improve the objective function over iterations even when the global optima is not attainable.

Remark 19. Suppose that the monetary cost and the free flow travel times are equal in the two links, i.e. $c_1 = c_2$, $t_1^0 = t_2^0$. Under UE, by definition, the link travel times will be the same for any level of demand $\bar{q}$. Note that the latter will be true even if the free flow travel time of the links is different, provided that the demand level surpasses a certain threshold such that both links are utilized. Because the travelers are uniformly distributed among paths and the path utilities are the same, SUELOGIT will reproduce the same equilibria. In line with the first rule of identifiability, the travel time coefficient θ_t is not identifiable but regardless, any value of θ_t will perfectly reproduce the observed link flows. An optimization library will not necessarily warn about these identification problems and it may return an arbitrary estimate for θ_t. A basic sanity check is to analyze if the objective function improves over iterations. Note, however, that this identifiability issue will not affect the goodness of fit but only statistical inference.

D.3. Small networks

All networks except for the toy network shown in Fig. 15(a) have been analyzed in prior literature (Wang et al., 2016; Yang et al., 2001; Lo and Chan, 2003). For the toy network, the number of trips between origin–destination pairs is assumed equal to $q_{1,4} = 50$, $q_{2,4} = 100$, $q_{3,4} = 150$ and $q_{i,j} = 0$, $\forall (i,j) \neq \{(1,4),(2,4),(3,4)\}$. For the remaining networks, the O-D matrices are equal to those provided by Wang et al. (2016), Yang et al. (2001) and Lo and Chan (2003). The travelers' consideration sets in the four networks include all acyclic paths connecting every O-D pair. The links' performance functions are BPR functions with parameters $\alpha = 0.15$ and $\beta = 4$ and the link capacities and free flow travel times are identical to those reported by Wang et al. (2016), Yang et al. (2001) and Lo and Chan (2003). The utility function is dependent on travel time only and with a ground truth coefficient $\theta_t^\star = -1 < 0$. Note that prior research defines $\theta = -\theta_t > 0$ as a positive dispersion coefficient that weights the link/path costs.

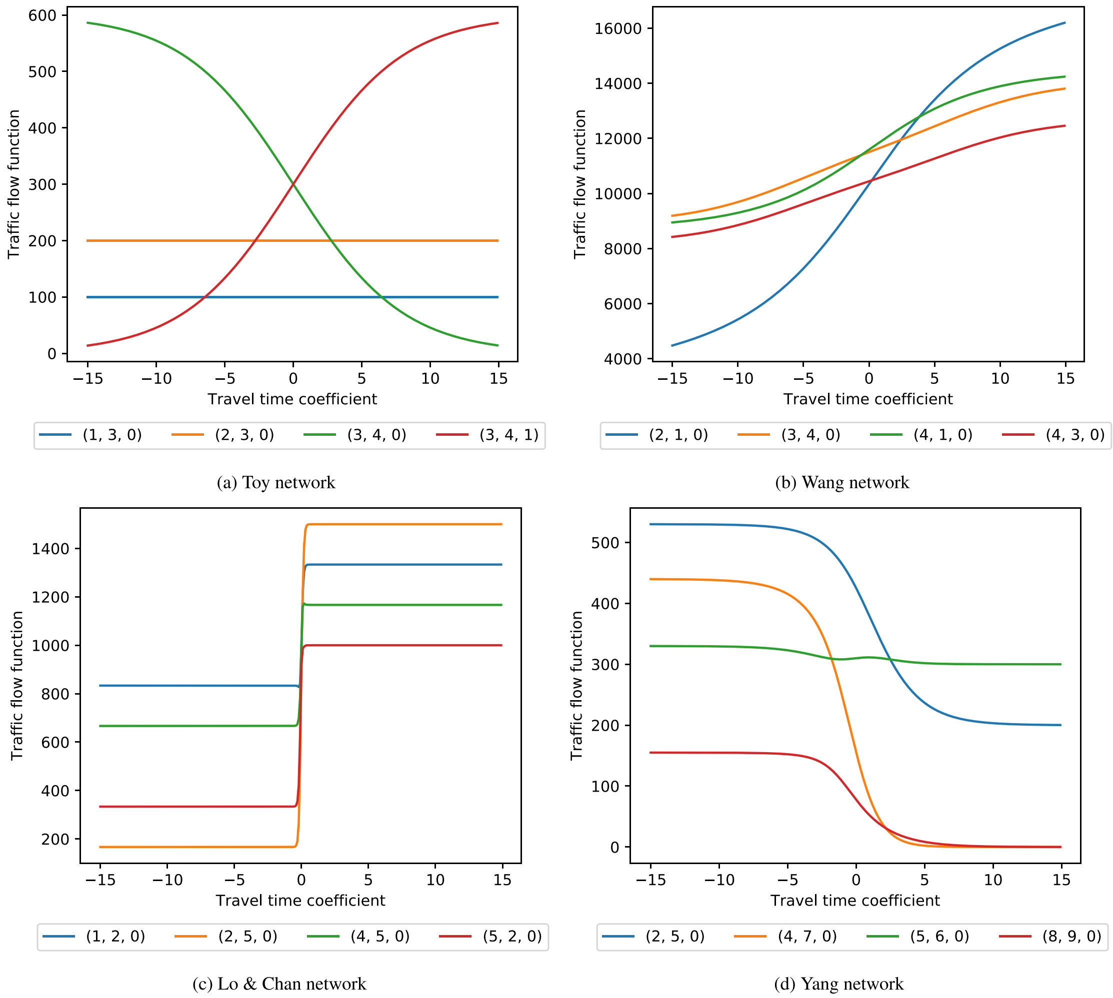


Fig. 16. Monotonicity of traffic flow functions in small networks.

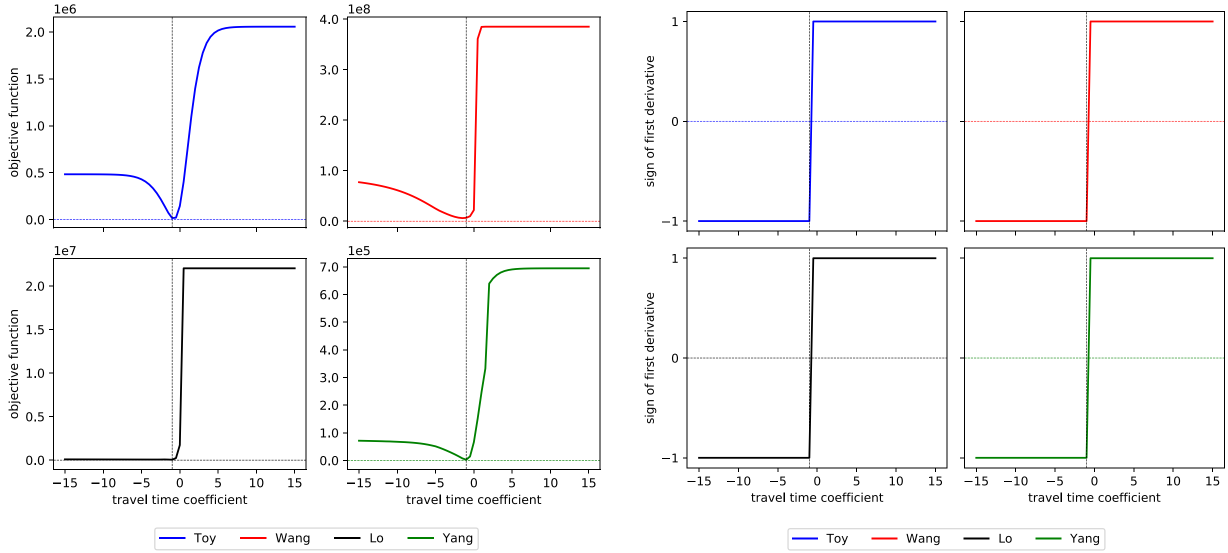
D.3.1. *Monotonicity of traffic flow functions*

Fig. 16 shows the output of the traffic flow functions for four arbitrary links in each network and for $\theta_t \in [-15, 15]$. The first two values of the tuples shown in the legend of each subfigure correspond to the origin and destination nodes of a link. The last element of the tuple is an index that distinguishes parallel links connecting the same O-D pair, e.g. a_3, a_4 in Fig. 16(a). Note that most traffic flow functions are monotonic and some exhibit convex and concave regions that resemble sigmoidal functions. In the toy network, the traffic flow functions of links $(2, 3)$ and $(1, 3)$ are constant because their link flows are invariant to the value of the travel time coefficient. Therefore, all individuals traveling from node 1 or 2 are forced to traverse those links and thus, the traffic count measurements are not contributing to the identifiability of additional coefficients of the travelers' utility function.

D.3.2. *Pseudo-convexity of the objective function*

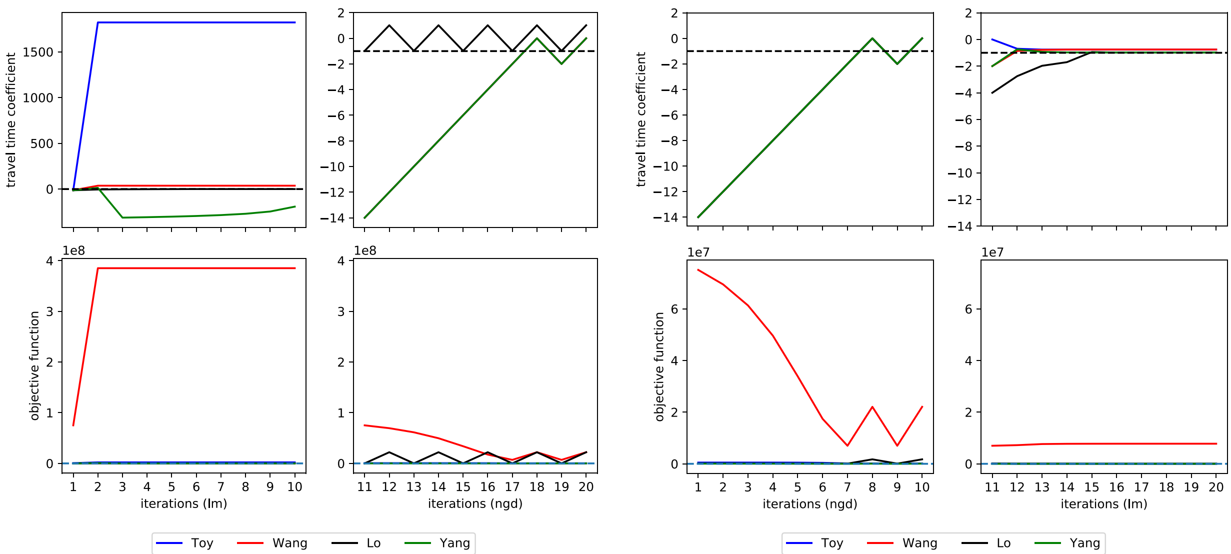
Fig. 17 illustrates the coordinate-wise pseudo-convexity of the LUE problem. The dashed vertical lines represent the true value of the travel time coefficient $\theta_t^\star$. Note that since the noise was introduced in the traffic count measurements, the error is not necessarily minimized at $\theta_t^\star$. In the four networks, the objective function ℓ is monotonically decreasing or increasing for the ranges of values that are lower or higher than $\theta^\star$, respectively, meaning that the negative slope of the first derivatives are pointing toward the region where the global optimum is located (Fig. 17(b)). As expected, the changes in the curvature of ℓ within the feasible domain suggest that ℓ is not convex but (coordinate-wise) pseudo-convex.

The proof of coordinate-wise pseudo-convexity of the LUE problem presented in Section 4 requires both coordinate-wise monotonicity of the traffic flow functions and a utility function with exogenous attributes only. Our empirical results suggest that the exogeneity of travel time and the monotonicity of the traffic flow function are not necessary but only sufficient conditions for the pseudo-convexity of the objective function.



(a) Objective function (b) Sign of the first derivative of objective function

Fig. 17. Pseudo-convexity of objective functions in small networks.



(a) No-refined stage: `LM`. Refined stage: `NGD` (b) No-refined stage: `NGD`. Refined stage: `LM`

Fig. 18. Convergence and consistency in parameter recovery with synthetic data from small networks.

D.3.3. Convergence and consistency in parameters' recovery

To study the advantages of using NGD instead of a second-order method in the non-refined stage of the optimization, we interchange the use of the methods between stages. Fig. 18(a) presents the convergence results obtained by applying LM and NGD in the no-refined and refined stages, respectively. Fig. 18(b) shows the converse case. To ensure that the global optima at $\theta_t = -1$ are attainable, no noise is introduced into the traffic count measurements. The learning rate for NGD was set to $\eta = 2$ which, for our unidimensional optimization problem on θ_t, equates to solution updates with a magnitude equal to 2. The starting point for optimization of the travel time coefficient is set to $\theta_t = -14$, which is the most challenging scenario of convergence analyzed by Yang et al. (2001).

We observe that the use of NGD in the no-refined stage makes the convergence faster and more stable in the four networks. In contrast, the use of LM in the no-refined stage results in unstable updates of θ_t. This behavior in second-order optimization methods is expected due to the constant change of curvature sign and the existence of flat regions within the range of pseudo-convex functions. In fact, for all networks, the objective function is concave and flat at the initial value for optimization $\theta_t = -14$ and it is convex and sharp around the global optima at $\theta_t = -1$ (Fig. 17).

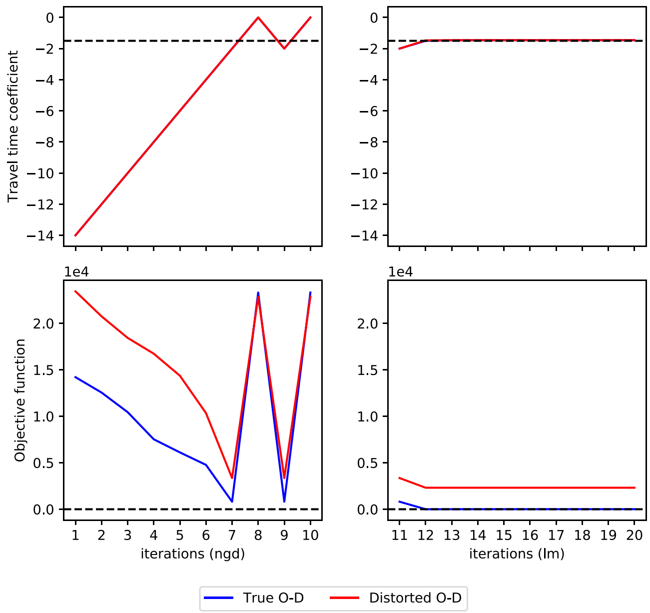


Fig. 19. Convergence under true and distorted reference OD matrix.

Table 5
Point estimates and t-tests for travel time coefficient.

Parameter (t-test)	Network			
	Toy	Wang et al. (2016)	Lo and Chan (2003)	Yang et al. (2001)
NGD	0.000 (0.0)	−2.000*** (−5.6)	−2.000*** (−6.9)	−4.000*** (−3.6)
LM	−14.000 (−0.0)	−14.000 (−0.6)	−0.984*** (−7.3)	−14.000 (−1.1)
NGD + LM	−0.761*** (−7.1)	−2.000*** (−5.6)	−0.976*** (−15.7)	−0.982*** (−7.3)
LM + NGD	0.000 (0.0)	−2.000*** (−5.6)	−0.984*** (−7.3)	−2.000*** (−6.8)
Links (coverage)	4 (100%)	8 (100%)	14 (100%)	14 (100%)
Paths	6	24	44	28
O-D pairs	3	12	12	9

Note: Significance levels: *p<0.1; **p<0.05; ***p<0.01.

D.3.4. Hypothesis testing

Table 5 shows the parameters estimates and t-tests obtained for the four small networks and under four scenarios; (i) only `NGD`, (ii) only `LM`, (iii) `NGD` in the no-refined stage and `LM` in the refined stage, (iv) `LM` in the no-refined stage and `NGD` in the refined stage. Overall, this evidence confirms the advantages of the integration of first and second-order optimization methods to perform statistical inference. The integration of `NGD` and `LM` in scenarios (iii) and (iv) generate solutions closer to the global optima than the standalone application of each optimization method. The hypothesis tests correctly reject the null at a 1% confidence level but only when `NGD` and `LM` are used in the no-refined and refined stages of the optimization, respectively.

D.3.5. Impact of error in reference OD matrix

To study the impact on convergence and inference of assuming an inaccurate reference O-D matrix, the experiment and testing network used by Yang et al. (2001) are used as a baseline. The ground truth and reference O-D matrices are defined as:

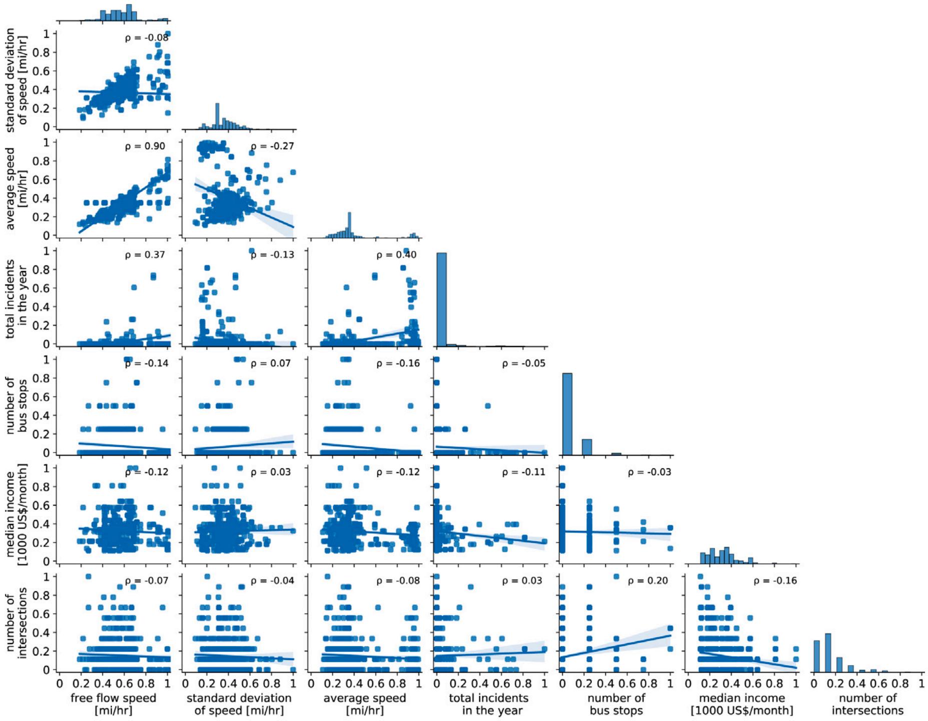
$$\mathbf{q}^{\text{true}} = (q_{1,6}, q_{1,8}, q_{1,9}, q_{2,6}, q_{2,8}, q_{2,9}, q_{4,6}, q_{4,8}, q_{4,9}) = (120, 150, 100, 130, 200, 90, 80, 180, 110)$$
$$\mathbf{q}^{\text{distorted}} = (100, 130, 120, 120, 170, 140, 110, 170, 105)$$

where $\mathbf{q}^{\text{distorted}}$ is the reference O-D matrix. In line with Yang et al. (2001), it is assumed that traffic counts are only observed in the following links: $(3,6), (5,6), (5,8), (5,9), (7,8)$. Fig. 19 shows the convergence toward the optimal solution when the reference O-D matrix is assumed equal to (i) the true O-D matrix $\mathbf{q}^{\text{true}}$ (blue curve) or to (ii) the distorted O-D matrix $\mathbf{q}^{\text{distorted}}$ (red curve). Note that in the two scenarios, the algorithm converges close to the global optima ($\theta_t^{\text{true O-D}} = -1.058$, $\theta_t^{\text{distorted O-D}} = -1.054$) and the objective function ($\ell_\star^{\text{true O-D}} = 758.8$, $\ell_\star^{\text{distorted O-D}} = 12301.6$) is minimized over iterations. Note that when $\mathbf{q} = \mathbf{q}^{\text{true}}$, a zero objective function is not attainable because of the introduction of random error in the traffic count measurements (see Section 7.2). Despite the bias in the reference O-D matrix, the null was correctly rejected under the two scenarios at a 95% level of confidence (true O-D: $T_{H_0:\theta_t=0} = -10.3$, $p < 0.01$, distorted O-D: $T_{H_0:\theta_t=0} = -4.1$, p< 0.05))

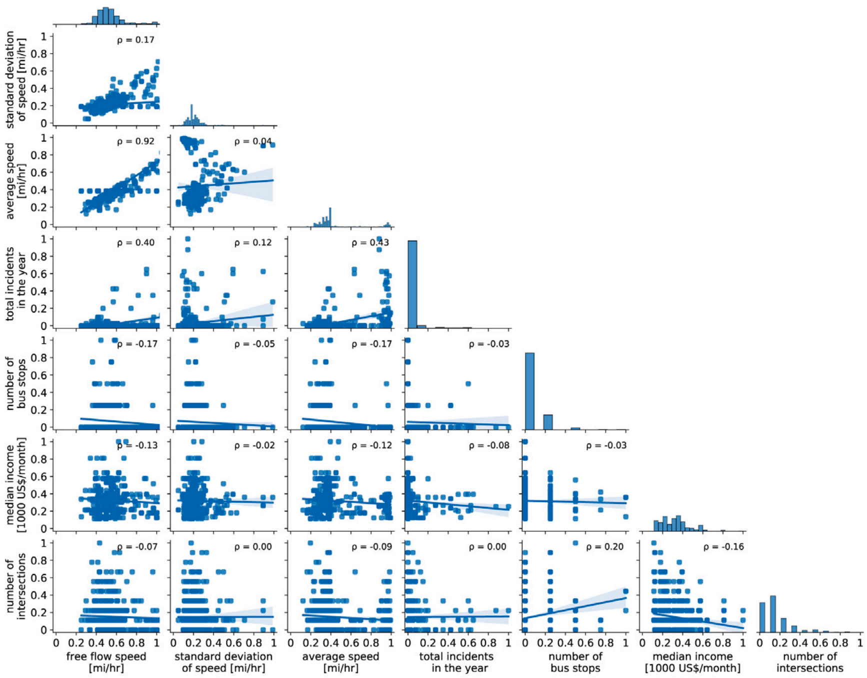
Appendix E. Additional analyses in Fresno, CA

E.1. Descriptive statistics

See Fig. 20.



(a) October, 2019



(b) October, 2020

Fig. 20. Correlation between normalized system-level attributes in Fresno, CA.

E.2. Estimation results

See Figs. 21 and 22.

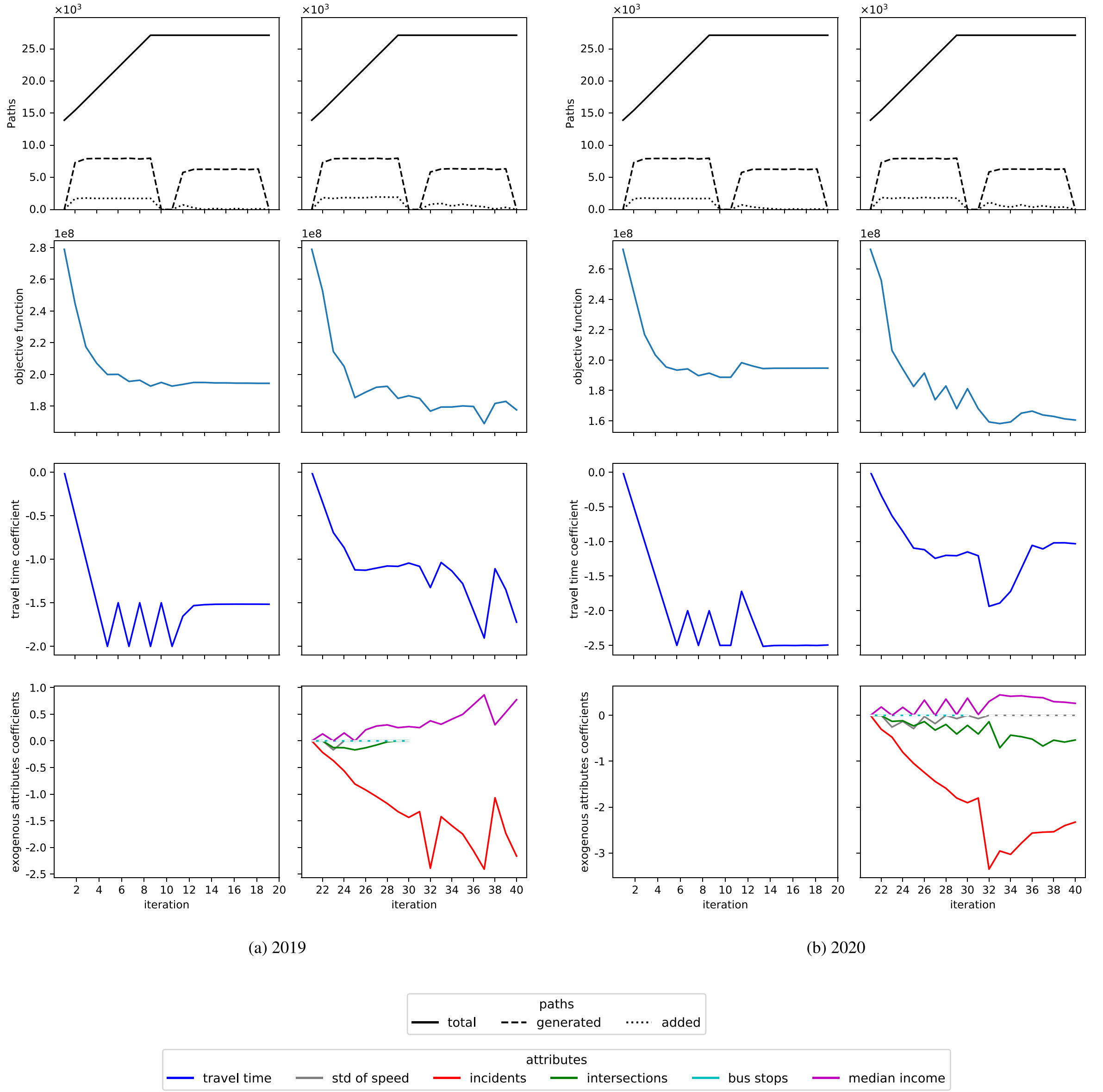


Fig. 21. Comparison of convergence in estimation of baseline and full models.

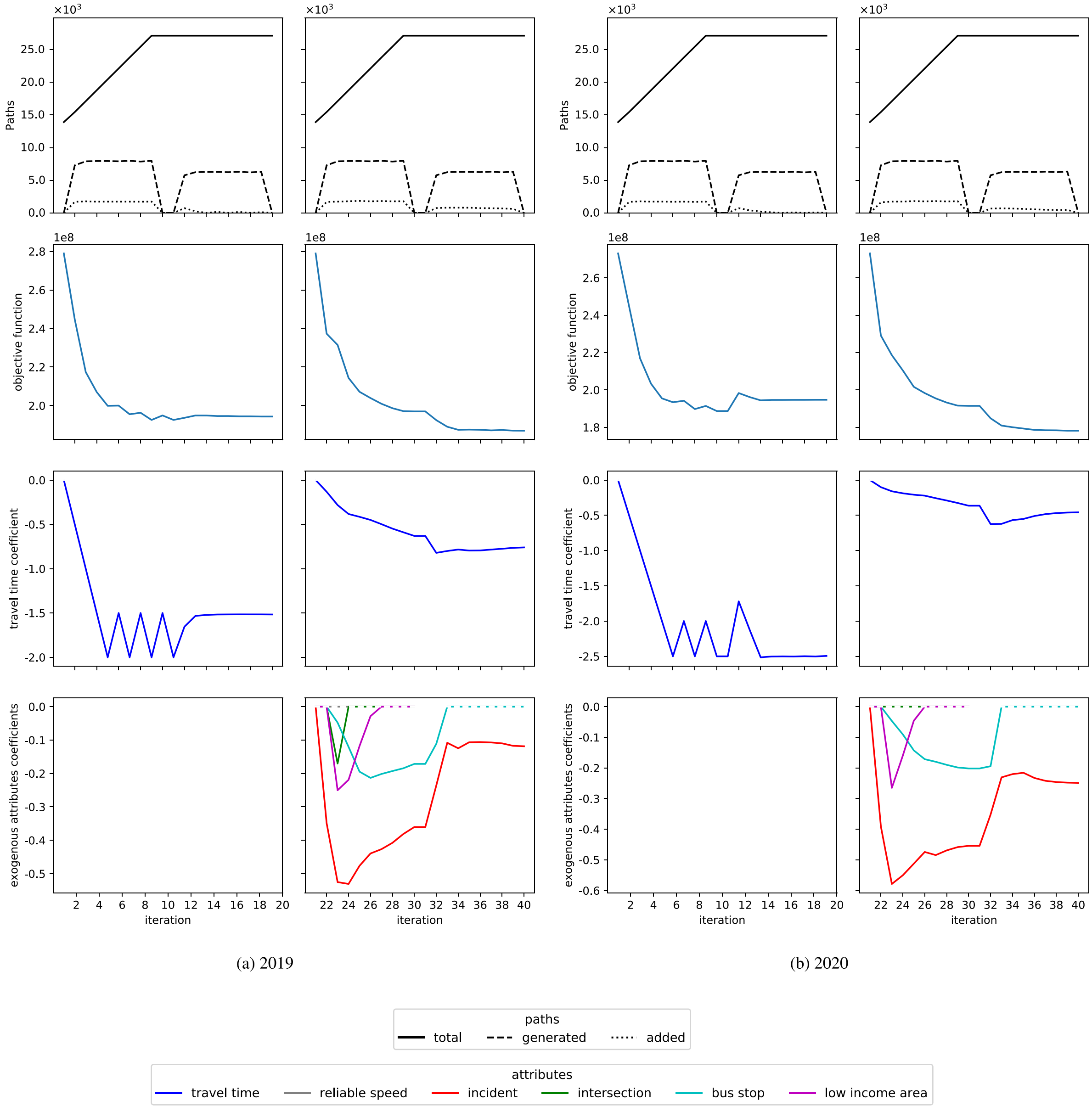


Fig. 22. Comparison of convergence in estimation of baseline and binarized models.

References

Amemiya, T., 1983. Chapter 6 Non-linear regression models. In: Handbook of Econometrics, Vol. 1. Elsevier, pp. 333–389. http://dx.doi.org/10.1016/S1573-4412(83)01010-7, URL: https://www.sciencedirect.com/science/article/pii/S1573441283010107.

Anas, A., Kim, I., 1990. Network loading versus equilibrium estimation of the stochastic route choice model maximum likelihood and least squares revisited. J. Reg. Sci. 30 (1), 89–103. http://dx.doi.org/10.1111/j.1467-9787.1990.tb00082.x.

AWS, 2022. Amazon EC2 instance types. URL: https://aws.amazon.com/ec2/instance-types/.

Bazaraa, M., Sherali, H., Shetty, C., 2006. Nonlinear Programming: Theory and Algorithms. John Wiley & Sons, New Jersey, URL: http://link.springer.com/10.1007/978-3-540-68651-4_4.

Bell, M.G., 1995. Alternatives to Dial's logit assignment algorithm. Transp. Res. B 29 (4), 287–295. http://dx.doi.org/10.1016/0191-2615(95)00005-X.

Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their applications to short term travel decisions. In: Handbook of Transportation Science. Springer, pp. 5–33.

Ben-Akiva, M.E., Lerman, S.R., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand, Vol. 9. MIT Press.

Bierlaire, M., Frejinger, E., 2008. Route choice modeling with network-free data. Transp. Res. C 16 (2), 187–198. http://dx.doi.org/10.1016/j.trc.2007.07.007, URL: https://linkinghub.elsevier.com/retrieve/pii/S0968090X07000551.

Bovy, P.H., Bekhor, S., Prato, C.G., 2008. The factor of revisited path size: Alternative derivation. Transp. Res. Rec. (2076), 132–140. http://dx.doi.org/10.3141/2076-15.

Boyles, S.D., Lownes, N.E., Unnikrishnan, A., 2020. Transportation Network Analysis, Vol. I. URL: https://sboyles.github.io/blubook.html.

Bwambale, A., Choudhury, C., Hess, S., 2019. Modelling long-distance route choice using mobile phone call detail record data: a case study of Senegal. Transport. A: Transp. Sci. 9935 (May), http://dx.doi.org/10.1080/23249935.2019.1611970.

Cameron, A.C., Trivedi, P.K., 2005. Microeconometrics: Methods and Applications. Cambridge University Press.

Cascetta, E., Postorino, M.N., 2001. Fixed point approaches to the estimation of O/D matrices using traffic counts on congested networks. Transp. Sci. 35 (2), 134–147. http://dx.doi.org/10.1287/trsc.35.2.134.10138.

Chen, M., Alfa, S.A., 1991. Algorithms for solving fisk's stochastic traffic assignment model. Transp. Res. B 25 (6), 405–412. http://dx.doi.org/10.1016/0191-2615(91)90033-F, URL: https://linkinghub.elsevier.com/retrieve/pii/019126159190033F.

Crouzeix, J.P., Ferland, J.A., 1982. Criteria for quasi-convexity and pseudo-convexity: Relationships and comparisons. Math. Program. 23 (1), 193–205. http://dx.doi.org/10.1007/BF01583788, URL: http://link.springer.com/10.1007/BF01583788.

Daganzo, C.F., 1977. Some statistical problems in connection with traffic assignment. Transp. Res. 11 (6), 385–389. http://dx.doi.org/10.1016/0041-1647(77)90002-8, URL: https://linkinghub.elsevier.com/retrieve/pii/0041164777900028.

Daganzo, C.F., Sheffi, Y., 1977. On stochastic models of traffic assignment. Transp. Sci. 11 (3), 253–274. http://dx.doi.org/10.1287/trsc.11.3.253, URL: http://pubsonline.informs.org/doi/abs/10.1287/trsc.11.3.253,

Damberg, O., Lundgren, J.T., Patriksson, M., 1996. An algorithm for the stochastic user equilibrium problem. Transp. Res. B 30 (2), 115–131. http://dx.doi.org/10.1016/0191-2615(95)00026-7.

Del Pino, G., 1989. The unifying role of iterative generalized least squares in statistical algorithms. Statist. Sci. 4 (4), 394–403. http://dx.doi.org/10.1214/ss/1177012408.

Dempe, S., Franke, S., 2016. On the solution of convex bilevel optimization problems. Comput. Optim. Appl. 63 (3), 685–703. http://dx.doi.org/10.1007/s10589-015-9795-8, Publisher: Springer US.

Dial, R.B., 1971. A probabilistic multipath traffic assignment model which obviates path enumeration. Transp. Res. 5 (2), 83–111. http://dx.doi.org/10.1016/0041-1647(71)90012-8.

Duchi, J.C., Hazan, E., Singer, Y., Bartlett, P.L., Wainwright, M.J., 2011. Adaptive subgradient methods for online learning and stochastic optimization. J. Mach. Learn. Res. 12, 2121–2159, Publisher: JMLR.org.

Evans, S.P., 1973. A relationship between the gravity model for trip distribution and the transportation problem in linear programming. Transp. Res. 7 (1), 39–61. http://dx.doi.org/10.1016/0041-1647(73)90005-1.

FHA, 2020. Traffic Trends Volume, October 2020. Technical Report, Federal Highway Administration, Washington, D.C., U.S..

Fisk, C., 1977. Note on the maximum likelihood calibration on Dial's assignment method. Transp. Res. 11 (1), 67–68. http://dx.doi.org/10.1016/0041-1647(77)90069-7, URL: https://linkinghub.elsevier.com/retrieve/pii/0041164777900697.

Fisk, C., 1980. Some developments in equilibrium traffic assignment. Transp. Res. B 14 (3), 243–255. http://dx.doi.org/10.1016/0191-2615(80)90004-1.

Fisk, C.S., 1989. Trip matrix estimation from link traffic counts: The congested network case. Transp. Res. B 23 (5), 331–336. http://dx.doi.org/10.1016/0191-2615(89)90009-X.

Gallant, A.R., 1975a. Nonlinear Regression. Amer. Statist. 29 (2), 73–81. http://dx.doi.org/10.1080/00031305.1975.10477374, URL: http://www.tandfonline.com/doi/abs/10.1080/00031305.1975.10477374.

Gallant, A.R., 1975b. The power of the likelihood ratio test of location in nonlinear regression models. J. Amer. Statist. Assoc. 70 (349), 198–203. http://dx.doi.org/10.1080/01621459.1975.10480288.

Gallant, A.R., Jorgenson, D.W., 1979. Statistical inference for a system of simultaneous, non-linear, implicit equations in the context of instrumental variable estimation. J. Econometrics 11 (2–3), 275–302. http://dx.doi.org/10.1016/0304-4076(79)90041-1.

García-Ródenas, R., Marín, Á., 2009. Simultaneous estimation of the origin-destination matrices and the parameters of a nested logit model in a combined network equilibrium model. European J. Oper. Res. 197 (1), 320–331. http://dx.doi.org/10.1016/j.ejor.2008.05.032, Publisher: Elsevier B.V.

GIS Division City of Fresno, 2021. City of Fresno GIS data hub. URL: http://gis4u.fresno.gov/downloads/.

Hazan, E., Levy, K.Y., 2015. Beyond convexity: Stochastic quasi-convex optimization. Adv. Neural Inf. Process. Syst. 1–9.

INRIX, 2021. Inrix analytics: Speed. URL: https://inrix.com/products/speed/.

Kaggle, 2021. California traffic collision data from SWITRS. URL: https://www.kaggle.com/alexgude/california-traffic-collision-data-from-switrs/.

Kingma, D.P., Ba, J.L., 2015. Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. pp. 1–15, arXiv:1412.6980.

Krishnakumari, P., van Lint, H., Djukic, T., Cats, O., 2020. A data driven method for OD matrix estimation. Transp. Res. C 113 (December 2018), 38–56. http://dx.doi.org/10.1016/j.trc.2019.05.014, Publisher: Elsevier.

Liu, H.X., Ban, J.X., Chen, A., 2006. A streamlined network calibration procedure for California SR41 corridor traffic simulation study. In: 85th Transportation Research Board Annual Meeting, No. May 2014.

Liu, S., Fricker, J.D., 1996. Estimation of a trip table and the θ parameter in a stochastic network. Transp. Res. A 30 (4 PART A), 287–305. http://dx.doi.org/10.1016/0965-8564(95)00031-3.

Lo, H.P., Chan, C.P., 2003. Simultaneous estimation of an origin-destination matrix and link choice proportions using traffic counts. Transp. Res. A 37 (9), 771–788. http://dx.doi.org/10.1016/S0965-8564(03)00048-X.

Ma, W., Qian, Z.S., 2017. On the variance of recurrent traffic flow for statistical traffic assignment. Transp. Res. C 81, 57–82. http://dx.doi.org/10.1016/j.trc.2017.05.009, Publisher: Elsevier Ltd.

Ma, W., Qian, S., 2018a. A generalized single-level formulation for origin–destination estimation under stochastic user equilibrium. Transp. Res. Rec. 2672 (48), 58–68. http://dx.doi.org/10.1177/0361198118782041.

Ma, W., Qian, S., 2018b. Statistical inference of probabilistic origin-destination demand using day-to-day traffic data. Transp. Res. C 88 (August 2017), 227–256. http://dx.doi.org/10.1016/j.trc.2017.12.015, Publisher: Elsevier.

Mai, T., Fosgerau, M., Frejinger, E., 2015. A nested recursive logit model for route choice analysis. Transp. Res. B 75, 100–112. http://dx.doi.org/10.1016/j.trb.2015.03.015, Publisher: Elsevier Ltd.

Mangasarian, O., 1965. Pseudo-convex functions. SIAM J. Control Optim. 3 (2), 281–290. http://dx.doi.org/10.1016/B978-0-12-780850-5.50009-5, URL: https://linkinghub.elsevier.com/retrieve/pii/B9780127808505500095, Publisher: Elsevier.

Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. 11 (2), 431–441. http://dx.doi.org/10.1137/0111030, URL: http://epubs.siam.org/doi/10.1137/0111030, arXiv:1011.1669v3,

McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. Front. Econom. 105–142. http://dx.doi.org/10.1108/eb028592,

McFadden, D., 1977. Quantitative Methods for Analysing Travel Behaviour of Individuals: Some Recent Developments. Cowles Foundation Discussion Paper, Technical Report, New Haven, Connecticut, ISBN: 9781000362978, pp. 279–318, Publication Title: Behavioural Travel Modelling.

Mereau, P., Paquet, J.G., 1974. Second order conditions for pseudo-convex functions. SIAM J. Appl. Math. 27 (1), 131–137. http://dx.doi.org/10.1137/0127011.

Parady, G., Ory, D., Walker, J., 2021. The overreliance on statistical goodness-of-fit and under-reliance on model validation in discrete choice models: A review of validation practices in the transportation academic literature. J. Choice Model. 38 (February 2020), 100257. http://dx.doi.org/10.1016/j.jocm.2020.100257, Publisher: Elsevier Ltd.

Patriksson, M., 2004. Sensitivity analysis of traffic equilibria. Transp. Sci. 38 (3), 258–281. http://dx.doi.org/10.1287/trsc.1030.0043.

PeMS, 2021. Performance measurement system. URL: https://pems.dot.ca.gov/.

Prashker, J.N., Bekhor, S., 2004. Route choice models used in the stochastic user equilibrium problem: A review. Transport Reviews 24 (4), 437–463. http://dx.doi.org/10.1080/0144164042000181707.

Qian, S., Zhang, H.M., 2012. On centroid connectors in static traffic assignment: Their effects on flow patterns and how to optimize their selections. Transp. Res. B 46 (10), 1489–1503. http://dx.doi.org/10.1016/j.trb.2012.07.006, Publisher: Elsevier Ltd.

Robbins, H., Monro, S., 1951. A stochastic approximation method. Ann. Math. Stat. 400–407, Publisher: JSTOR.

Robillard, P., 1974. Calibration of Dial's assignment method. Transp. Sci. 8 (2), 117–125. http://dx.doi.org/10.1287/trsc.8.2.117.

Russo, F., Vitetta, A., 2011. Reverse assignment: Calibrating link cost functions and updating demand from traffic counts and time measurements. Inverse Probl. Sci. Eng. 19 (7), 921–950. http://dx.doi.org/10.1080/17415977.2011.565339.

Sabach, S., Shtern, S., 2017. A first order method for solving convex bilevel optimization problems. SIAM J. Optim. 27 (2), 640–660. http://dx.doi.org/10.1137/16M105592X, URL: https://epubs.siam.org/doi/10.1137/16M105592X.

Shen, W., Wynter, L., 2012. A new one-level convex optimization approach for estimating origin-destination demand. Transp. Res. B 46 (10), 1535–1555. http://dx.doi.org/10.1016/j.trb.2012.07.005, Publisher: Elsevier Ltd.

Staib, M., Reddi, S., Kale, S., Kumar, S., Sra, S., 2019. Escaping saddle points with adaptive gradient methods. In: 36th International Conference on Machine Learning, Vol. 2019-June. ICML 2019, ISBN: 9781510886988, pp. 10420–10454, arXiv:1901.09149.

SWITRS, 2021. Codebook of statewide integrated traffic records system (SWITRS). URL: https://tims.berkeley.edu/help/SWITRS.php.

TNTP, 2021. Transportation networks for research core team. URL: https://github.com/bstabler/TransportationNetworks.

Train, K.E., 2003. Discrete Choice Methods with Simulation, Vol. 9780521816. Cambridge University Press, Cambridge, MA, http://dx.doi.org/10.1017/CBO9780511753930, Publication Title: Discrete Choice Methods with Simulation.

US Census Bureau, 2019. TIGER/line with selected demographic and economic data. URL: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-data.html.

Waetjen, D., Shilling, F., 2021. Leveraging the California Highway Incident Processing System for Traffic Safety Policy and Research. Technical Report.

Walker, J., 2002. Mixed logit (or logit kernel) model: Dispelling misconceptions of identification. Transp. Res. Rec. (1805), 86–98. http://dx.doi.org/10.3141/1805-11.

Wang, Y., Ma, X., Liu, Y., Gong, K., Henricakson, K.C., Xu, M., Wang, Y., 2016. A two-stage algorithm for origin-destination matrices estimation considering dynamic dispersion parameter for route choice. PLoS One 11 (1), 1–24. http://dx.doi.org/10.1371/journal.pone.0146850.

Wu, C.F., 1981. Asymptotic theory of nonlinear least squares estimation. Ann. Statist. 9 (3), 590–606. http://dx.doi.org/10.1214/aos/1176345455, URL: https://projecteuclid.org/journals/annals-of-statistics/volume-9/issue-3/Asymptotic-Theory-of-Nonlinear-Least-Squares-Estimation/10.1214/aos/1176345455.full.

Yang, H., Meng, Q., Bell, M.G., 2001. Simultaneous estimation of the origin-destination matrices and travel-cost coefficient for congested networks in a stochastic user equilibrium. Transp. Sci. 35 (2), 107–123. http://dx.doi.org/10.1287/trsc.35.2.107.10133.

Yang, H., Sasaki, T., Iida, Y., Asakura, Y., 1992. Estimation of origin-destination matrices from link traffic counts on congested networks. Transp. Res. B 26 (6), 417–434. http://dx.doi.org/10.1016/0191-2615(92)90008-K.

Yao, W., Qian, S., 2020. Learning to recommend signal plans under incidents with real-time traffic prediction. Transp. Res. Rec. 2674 (6), 45–59. http://dx.doi.org/10.1177/0361198120917668.

Zhang, M., Ma, J., Singh, S.P., Chu, L., 2008. Developing Calibration Tools for Microscopic Traffic Simulation Final Report Part III: Global Calibration - O-D Estimation, Traffic Signal Enhancements and a Case Study. Technical Report, California Path Program, Institute of Transportation Studies, University of California, Berkeley, p. 81, Publication Title: California PATH Research Report UCB-ITS-PRR-2008-8.