2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Improving Social Cost in Traffic Routing with Bounded Regret via Second-Best Tolls

Arwa Alanqary, Abdul Rahman Kreidieh, Samitha Samaranayake, and Alexandre M Bayen

Abstract— **In this work, we investigate algorithmic improvements that navigation services can implement to steer road networks toward a system-optimal state while retaining high levels of user compliance. We model the compliance of users using marginal regret, and we extend the definition of the social cost function to account for various traffic congestion externalities. We propose a routing algorithm for the static traffic assignment problem that improves the social cost with guarantees on the worst-case regret in the network. This algorithm leverages the connection we establish between this problem and that of second-best toll pricing. We present numerical experiments on different networks to illustrate the trade-off between regret and efficiency of the resulting assignment for arbitrary social costs.**

I. Introduction

Congestion in traffic networks is a long-standing challenge with widespread societal, economic, and environmental implications. Reducing congestion has been tackled through intelligent transportation systems, which utilize existing infrastructure and advancements in sensing and communication technologies for efficient traffic control [1]. One such technology is navigation services through mobile applications, which leverage real-time data and algorithms to provide users with optimized routes that account for current traffic conditions. Today, services like Google Maps, Apple Maps, Waze, and Moovit are widely used and have a strong influence on motorists' choices. Their growing adoption rates create new traffic patterns, making them a crucial aspect to consider when studying the traffic assignment problem [2], [3]. A particular concern involves the potential negative impact they may have on the efficiency of the road network [2], [4]–[6].

The core of the problem is that most navigation services are designed to provide users with shortest path recommendations without accounting for the potential externalities of their routing algorithms. While these recommendations minimize individual travel costs, they can lead to inefficient outcomes, likely a Nash equilibrium [7], [8] (also called user equilibrium (UE) or Wardrop equilibrium), which are often far from the system-optimal (SO) outcomes [9]–[11]. This problem becomes even more eminent when users' cost metrics, such as travel time, tolls, and fuel consumption, differ from the social cost metrics on which the system efficiency is measured, like emissions or road utilization.

Arwa Alanqary and Alexandre M Bayen are with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA. arwa@berkeley.edu, bayen@berkeley.edu

Abdul Rahman Kreidieh and Alexandre M Bayen are with Google Research, Mountain View, CA. aboudy@google.com

Samitha Samaranayake is with the School of Civil and Environmental Engineering, Cornell University, Ithaca, NY. samitha@cornell.edu

Addressing the impact of navigation services and leveraging their potential benefits in reducing congestion requires an algorithmic solution. Specifically, we need routing algorithms that steer users toward a system-optimal state. However, system-optimal recommendations might not be satisfactory to the users, as they can lead to much higher costs for some users compared to alternative routes [12]–[14]. So if users decide that these recommendations are not in their best interest – even though they might lead to system optimal outcomes – they can disregard them, abandon the platform, and become non-compliant. Thus, the question of compliance with the algorithm's recommendations is of utmost importance.

In this work, we design routing algorithms that steer users toward a system-optimal state without compromising users' compliance. To that end, we use the notion of *fairness* as a criterion for compliance. Specifically, we propose using marginal regret [7] as a measure of unfairness in the system, which captures the maximum difference between a user's travel cost and the minimum cost of any path between the same origin and destination pair. We then propose a game-theoretic framework to design better route recommendations with compliance as a core consideration. The proposed framework is formulated as a constrained system optimum (CSO) problem and utilizes the relaxed behavioral assumption of bounded rationality [15]. Our solution approach makes the connection between the CSO problem and optimal toll design, which addresses the design of toll schemes that induce system optimality in routing games [16]–[18].

A. Contributions

The key contributions of the article include

- **Formulating the problem of system-optimal recommendations.** We formulate it as a constrained system optimum problem with dual-criteria for the users and the planner and unfairness constraint as a criterion for compliance. We measure the unfairness in the system using marginal regret.
- **Formalizing the connection between this problem formulation and that of optimal toll design.** By introducing appropriate constraints on the tolls, we guarantee that the equilibrium assignment resulting from such tolls has bounded marginal regret.
- **Proposing a heuristic algorithm for solving the system-optimal recommendations problem.** This algorithm is based on the established connection with optimal toll design. It finds assignments with bounded regret and improved social cost by computing tolls

979-8-3503-1632-2/24/$31.00 ©2024 IEEE

with constraints on their maximum difference. As an extension to the algorithm, we propose an iterative procedure that relaxes these constraints to further improve the social cost. We evaluate the performance of the proposed approach on benchmark networks at different levels of allowable regret.

B. Related work

a) Constrained system optimum: The problem of system optimal recommendations we formulate in this work is part of the literature that studies the balance between the efficiency of traffic assignments and their level of "unfairness" to users. The first attempt to address this problem was presented by Jahn et al. [13], who introduced the *constrained system optimum (CSO)* model, which aims at finding system-optimal assignments under constraints on the ratio of the *normal length* of a given path to that of the shortest path. The normal length refers to any exogenous property of the path independent of the flow (e.g., length, free-flow travel time, etc). This unfairness criterion simplifies the problem significantly, and many studies have proposed algorithms for this CSO model for various social objectives and normal length choices [19], [20].

One drawback of the CSO model is that the experienced unfairness in the network can be far worse than the bound that the model guarantees. To overcome this limitation, the model needs to incorporate flow-dependent criteria of unfairness, and thus, flow-dependent constraints need to be imposed on the CSO problem. A common choice for such a criterion is the loaded unfairness defined as the ratio between the travel cost along a path to the minimum travel cost among all used paths, both flow-dependant. Using this criterion, Angelelli et al. [21] proposed a heuristic algorithm for the CSO problem that relies on linearizing the edge cost functions. More recently, Jalota et al. [14] proposed another algorithm to solve this problem by finding the optimal interpolation between the system optimal and the user equilibrium assignments.

In the present work, we propose marginal regret as an alternative measure of unfairness. Bounds on the marginal regret characterize the set of boundedly-rational user equilibrium assignments (BRUE) [15]. Thus, our formulation naturally connects this problem to that of finding the best-case BRUE [22]. Further, [14] proposed a pricing mechanism to enforce the resulting constrained system optimal assignment. Our approach simultaneously solves the assignment and pricing problems by translating the user fairness constraints into toll constraints.

b) Optimal toll design: Imposing tolls on road networks is one of the most popular mechanisms for reducing the inefficiency of selfish routing behavior, both in theory and practice. The classical approach to toll design is the marginal cost pricing, first introduced by Pigou [16]. Beckmann et al. [17] showed that such marginal pricing is optimal in the sense that it always induces system-optimal flows as equilibrium. A long line of research then focused on studying and characterizing the set of all optimal tolls [18], [23], [24], showing that it can be described by linear equations and inequalities. This enabled further development in designing toll vectors that optimize a secondary objective, such as minimizing the toll amount in the network or minimizing the number of tolled edges [25]–[28].

A closely related problem, often referred to as second-best pricing, is concerned with finding tolls that induce equilibrium assignments with minimum social cost under some constraints on the tolls. In this problem, achieving system optimality might not be possible. Most studies on second-best pricing consider support constraints to restrict tolls to certain links in the network. The problem is typically formulated as a mathematical program with equilibrium constraints (MPEC) and heuristic algorithms are proposed to approximate the optimal solution [29]–[34]. Hoefer et al. [35] proved the NP-hardness of this problem for general networks. Bonifaci et al. [36] studied bounds on the efficiency of second-best tolls in the more general setup of threshold constraints. In our work, we formulate the problem of optimal route recommendations under unfairness constraints as a second-best pricing problem where the constraints on the tolls are used to control the regret of the resulting assignment.

C. Organization

The remainder of this article is organized as follows. Section II gives formal definitions of our framework, including the network model, equilibrium and system optimal assignments, marginal regret as a measure of unfairness, and our problem formulation. Section III outlines our proposed algorithm and introduces the related problem of optimal toll design. In section IV, we illustrate the use of the proposed algorithm through numerical experiments.

II. PROBLEM SETUP

A. Non-cooperative routing games

a) Network: We represent the road network as a directed graph G with node set N and edge set E. On this graph, we let $\mathcal{Z} = \{(r_i, s_i, d_i)\}_{i=1}^{K}$ be a set of K origin-destination (OD) pairs (r_i, s_i) and their corresponding non-atomic demand d_i. Each element of $\mathcal{Z}$ corresponds to a different *commodity* (a subset of users) that utilizes the network. For each OD pair, there is a set of simple (non-cyclic) paths $\mathcal{P}_i$. We define the set of all simple paths in the graph as $\mathcal{P} = \bigcup_{i=1}^{K} \mathcal{P}_i$.

b) Flows: An assignment in graph G is a vector $\boldsymbol{h} \in \mathbb{R}_+^{|\mathcal{P}|}$ with non-negative entries h_p denoting the flow routed through each path $p \in \mathcal{P}$. Such assignment is feasible if $\sum_{p \in \mathcal{P}_i} h_p = d_i$ for all $i = 1, \ldots, K$. Intuitively, this means that the demand d_i between an OD pair has been fully routed. We denote by $\mathcal{H}_{\boldsymbol{d}}$ the set of feasible flows corresponding to the demand vector $\boldsymbol{d} \in \mathbb{R}_+^{K}$. The path flow vector $\boldsymbol{h}$ induces a flow on each edge. We define the edge flow vector $\boldsymbol{x} \in \mathbb{R}_+^{|E|}$ with non-negative entries x_e for each edge $e \in E$. The edge flows are related to the path flows as $\boldsymbol{x} = H\boldsymbol{h}$, where

$H \in \mathbb{R}^{|E| \times |\mathcal{P}|}$ is the edge-path incidence matrix with entries

$$H_{e,p} = \begin{cases} 1 & \text{if edge } e \text{ is in path } p \\ 0 & \text{otherwise.} \end{cases}$$

We denote by H_p the p^{th} column of H. We define the set of feasible edge flows as all such vectors that are induced by a feasible path flow

$$\mathcal{X}_{\boldsymbol{d}} = \{\boldsymbol{x} \in \mathbb{R}_+^{|E|} : \exists \boldsymbol{h} \in \mathcal{H}_{\boldsymbol{d}} \text{ s.t. } \boldsymbol{x} = H\boldsymbol{h}\}. \quad (1)$$

Note that the edge flows vector is uniquely defined by the assignment vector but the inverse is not necessarily true.

B. User behavior and social objectives

Central to the traffic assignment problem is the behavioral assumption that users aim to "selfishly" minimize travel cost. Such selfish behavior can lead to a *user equilibrium* assignment. In contrast, the navigation service (which we refer to as the *planner*) aims to achieve a *system-optimal* assignment. Further, the metrics on which users evaluate their travel costs and the planner evaluates the social cost need not be the same. These two assignment rules can lead to very different outcomes. We make this concrete in the following.

a) Edge cost functions: Let $\mathcal{L} = \{l_e(\cdot)\}_{e\in E}$ be a set of flow-dependent cost functions where each $l_e : \mathbb{R}_+ \mapsto \mathbb{R}_+$ maps x_e, the flow on edge e, to the travel cost incurred by users on that edge. Similarly, let $\mathcal{W} = \{w_e(\cdot)\}_{e\in E}$ be another set of flow-dependent cost functions where each $w_e : \mathbb{R}_+ \mapsto \mathbb{R}_+$ maps x_e to the cost incurred by the planner. Throughout, we assume that the functions of both sets are continuous, differentiable, and nondecreasing. Further, for each $w_e \in \mathcal{W}$, we assume the function $x_e w_e(x_e)$ is convex. We define the vectors of cost-functions $\boldsymbol{l}(\boldsymbol{x}) = [l_e(x_e)]_{e\in E}$ and $\boldsymbol{w}(\boldsymbol{x}) = [w_e(x_e)]_{e\in E}$. With a slight abuse of notation, we denote the cost incurred along a path $p \in \mathcal{P}$ as $l_p(\boldsymbol{x}) = \sum_{e\in p} l_e(x_e)$ or in matrix form $l_p(\boldsymbol{x}) = H_p^\top \boldsymbol{l}(\boldsymbol{x})$. Given the graph G, a set of "user" cost functions $\mathcal{L}$, a set of "planner" cost functions $\mathcal{W}$, and an origin-destination set $\mathcal{Z}$, we define the routing game instance as $\mathcal{G}(G, \mathcal{L}, \mathcal{W}, \mathcal{Z})$.

b) Assignment rules: Next, we define the user equilibrium (also referred to as Wardrop's first principle of route choice [11]) and system-optimal (also referred to as Wardrop's second principle [11]) assignments.

Definition II.1. *(User equilibrium (UE)) For a routing game $\mathcal{G}(G, \mathcal{L}, \mathcal{W}, \mathcal{Z})$ a feasible assignment vector $\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{d}}$ with induced edge flows $\boldsymbol{x}$ is at user equilibrium if and only if for every $i = 1, \dots, K$ and paths $p_1, p_2 \in \mathcal{P}_i$ with $h_{p_1} > 0$*

$$l_{p_1}(\boldsymbol{x}) \le l_{p_2}(\boldsymbol{x}).$$

This definition states that at UE, all used path between an OD pair have equal and minimal travel cost.

Definition II.2. *(System-optimal (SO) assignment) For a routing game $\mathcal{G}(G, \mathcal{L}, \mathcal{W}, \mathcal{Z})$ a feasible assignment vector $\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{d}}$ with induced edge flows $\boldsymbol{x}$ is a system-optimal assignment if*

$$\boldsymbol{x}^\top \boldsymbol{w}(\boldsymbol{x}) \le \tilde{\boldsymbol{x}}^\top \boldsymbol{w}(\tilde{\boldsymbol{x}}) \quad \forall \tilde{\boldsymbol{x}} \in \mathcal{X}_{\boldsymbol{d}} \quad (2)$$

The system-optimal assignment minimizes the social cost $C(\boldsymbol{x}) := \boldsymbol{x}^\top \boldsymbol{w}(\boldsymbol{x})$ with respect to the planner's cost functions $\mathcal{W}$. Here, we generalized the definition of the social cost [11] to not necessarily be based on the users' cost functions. This allows us to incorporate different factors in the social cost that the users might not consider, such as emissions and road utilization.

C. Marginal regret as a measure of unfairness

We consider the setting in which the planner aims to achieve system optimality by giving route recommendations to the users. A primary concern is ensuring compliance, as users may deviate from the recommendations if they are not in their best interest. To address this, we use fairness as a criterion for determining user compliance. Our modeling assumption is that users can tolerate a certain level of unfairness and will follow the planner's recommendations if the perceived unfairness remains within acceptable range. To quantify the unfairness of an assignment, we use the marginal regret concept introduced in [7]. For a given feasible assignment $\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{d}}$ and induced edge flows $\boldsymbol{x}$, we define the *marginal regret* along a path $p \in \mathcal{P}_i$ as

$$R_p(\boldsymbol{x}) = l_p(\boldsymbol{x}) - \min_{\tilde{p}\in\mathcal{P}_i} l_{\tilde{p}}(\boldsymbol{x}), \quad (3)$$

which captures how much users on path p could have reduced their travel cost had they chosen the shortest path from their origin to destination.

At the network level, we define the worst-case and average marginal regrets below.

Definition II.3 (Worst-case marginal regret)**.** *The worst-case marginal regret of a feasible assignment vector $\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{d}}$ with induced edge flows $\boldsymbol{x}$ is defined as*

$$R^{\max}(\boldsymbol{x}) = \max_{i=1,\dots,K} \max_{p\in\mathcal{P}_i : h_p > 0} R_p(\boldsymbol{x}) \quad (4)$$

It measures the worst-case regret across all used paths and commodities in the network.

Definition II.4 (Average marginal regret)**.** *The average marginal regret of a feasible assignment vector $\boldsymbol{h} \in \mathcal{H}_{\boldsymbol{d}}$ with induced edge flows $\boldsymbol{x}$ is defined as*

$$\bar{R}(\boldsymbol{x}) = \frac{1}{\|\boldsymbol{d}\|_1} \sum_i \sum_{p\in\mathcal{P}_i} h_p(l_p(\boldsymbol{x}) - \min_{\tilde{p}\in\mathcal{P}_i} l_{\tilde{p}}(\boldsymbol{x})). \quad (5)$$

It measures the weighted average of the regret on each path weighted by the flow on that path.

Remark 1. *Both notions of regret characterize the set of UE in the network when the regret is zero. They are both suitable measures of unfairness and indicators of user compliance. The average marginal regret is a less stringent condition and enjoys desirable computational properties, as discussed in [7]. The worst-case regret guarantees bounded levels of unfairness for all users and is aligned with the commonly studied assumption of bounded rationality [22], [37]. In our following problem statement, we will use the worst-case regret as it is motivated by our proposed algorithm. In the* *numerical experiments, we report both regret measures and show that the worst-case regret is much more conservative in most cases.*

D. Problem statement

Following the notion of regret introduced in definition II.3, we can now write the planner's recommendation as the solution to the following constrained system optimum (CSO) problem

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathcal{X}_{\boldsymbol{d}}} \quad & C(\boldsymbol{x}) := \boldsymbol{x}^\top \boldsymbol{w}(\boldsymbol{x}) \\ \text{s.t.} \quad & R^{\max}(\boldsymbol{x}) \le \epsilon \end{aligned} \tag{CSO}$$

The feasible set of problem CSO, is referred to as the set of $\epsilon-$Nash equilibria or *boundedly rational user equilibria* (BRUE) [22], [37]. Thus the planner's task is to find the best BRUE assignment that minimizes the social cost. Unlike UE, BRUE can not be characterized as a solution to some convex optimization problem. In fact, under additional mild conditions (strict monotonicity of the cost functions), the edge flows induced by the UE assignment are unique [17] while BRUE are not. They do not define a convex set, even for affine cost functions [22], [37].

III. Computational Approaches

Our goal in this work is to devise a computational algorithm to solve problem CSO. Our approach is based on linking the problem we formulate here to the well-studied optimal toll design problem. To set the stage, we first provide an overview of relevant results that will inform our methodology.

A. Optimal toll design

Optimal toll design tackles the problem of enforcing an SO assignment (definition II.2) via the use of additive tolls on the edges of the network. Let $\boldsymbol{x}^\star$ be the edge flows induced by an SO assignment of the game instance $\mathcal{G}(G,\mathcal{L},\mathcal{W},\mathcal{Z})$. Consider a related game instance $\mathcal{G}(G,\mathcal{L}^{\boldsymbol{t}},\mathcal{W},\mathcal{Z})$ defined by a toll vector $\boldsymbol{t}\in\mathbb{R}^{|E|}$ with user cost functions

$$l_e^{\boldsymbol{t}}(x_e) = l_e(x_e) + t_e \quad \forall e\in E \qquad \text{where } l_e(x_e)\in\mathcal{L}. \tag{6}$$

The goal is to find $\boldsymbol{t}$ such that the edge flows $\boldsymbol{x}^{\text{eq}}$ induced by a UE assignment in $\mathcal{G}(G,\mathcal{L}^{\boldsymbol{t}},\mathcal{W},\mathcal{Z})$ satisfy $\boldsymbol{x}^{\text{eq}} = \boldsymbol{x}^\star$.

Here, we present results by Fleischer et al. [18], which give a constructive proof of the existence of such toll vectors for any feasible assignment in multi-commodity routing games. In our proposed algorithm, we modify this construction to find toll vectors that guarantee an $\epsilon-$bound on the worst-case regret.

Definition III.1 (Enforceable assignment)**.** *An edge flow vector $\boldsymbol{x}$ is called enforceable if it is induced by a UE assignement $\boldsymbol{h}^{eq}$ in the game instance $\mathcal{G}(G,\mathcal{L}^{\boldsymbol{t}},\mathcal{W},\mathcal{Z})$ for some toll vector $\boldsymbol{t}$.*

Theorem III.1 (Enforcing a feasible $\boldsymbol{x}$ [18])**.** *Any feasible edge flow vector $\boldsymbol{x}\in\mathcal{X}_{\boldsymbol{d}}$ is enforceable in the sense of definition III.1.*

The proof of theorem III.1 utilizes a linear programming formulation of the optimal toll design problem and provides a polynomial time algorithm to compute these toll. Now for $\boldsymbol{x}^\star$, an SO assignment of $\mathcal{G}(G,\mathcal{L},\mathcal{W},\mathcal{Z})$, being the flow we want to enforce, define the following linear program

$$\begin{aligned} \min_{\boldsymbol{h}} \quad & \sum_{p\in\mathcal{P}} l_p(\boldsymbol{x}^\star) h_p \\ \text{s.t.} \quad & \boldsymbol{x}^\star = H\boldsymbol{h} \\ & \sum_{p\in\mathcal{P}_i} h_p = d_i \quad \forall i=1,\dots,K \\ & h_p \ge 0 \qquad\qquad \forall p\in\mathcal{P} \end{aligned} \tag{$P_{\boldsymbol{x}^\star}$}$$

And its dual problem

$$\begin{aligned} \max_{\boldsymbol{z},\boldsymbol{t}} \quad & \sum_i z_i d_i - \sum_{e\in E} t_e x_e^\star \\ \text{s.t.} \quad & z_i - \sum_{e\in p} t_e \le l_p(\boldsymbol{x}^\star) \end{aligned} \tag{$D_{\boldsymbol{x}^\star}$}$$

Let $\boldsymbol{h}^\star$ and $(\boldsymbol{z}^\star,\boldsymbol{t}^\star)$ be the primal and dual solutions, respectively. The proof of theorem III.1 in [18] shows that the vector $\boldsymbol{t}^\star$ is an optimal toll vector that enforces the SO assignment that induces $\boldsymbol{x}^\star$.

B. Translating tolls into regret

In this section, we aim to derive a relationship between tolls in the network and the worst-case regret. The goal is to impose linear constraints on the toll vector to control for regret. We present the following proposition that makes such a relation.

Proposition III.1. *Let $\boldsymbol{t}\in\mathbb{R}^{|E|}$ be a vector of tolls that characterizes a game instance $\mathcal{G}(G,\mathcal{L}^{\boldsymbol{t}},\mathcal{W},\mathcal{Z})$ and let $\boldsymbol{h}^{eq}$ be a UE in that game that induces edge flows $\boldsymbol{x}^{eq}$. If the vector $\boldsymbol{t}$ satisfies*

$$\delta_i(\boldsymbol{t}) := \max_{p\in\mathcal{P}_i} H_p^\top \boldsymbol{t} - \min_{p\in\mathcal{P}_i} H_p^\top \boldsymbol{t} \le \epsilon \quad \forall i=1,\dots,K, \tag{7}$$

then the UE assignment $\boldsymbol{h}^{eq}$ has a worst-case marginal regret

$$R^{\max}(\boldsymbol{x}^{eq}) \le \epsilon \tag{8}$$

with respect to the game instance $\mathcal{G}(G,\mathcal{L},\mathcal{W},\mathcal{Z})$.

Proof: Given a toll vector $\boldsymbol{t}\in\mathbb{R}^{|E|}$, let us consider the paths toll vector $\boldsymbol{\tau}\in\mathbb{R}^{|\mathcal{P}|}$ defined as

$$\boldsymbol{\tau} = H^\top \boldsymbol{t} \tag{9}$$

with entries τ_p for the total amount of tolls imposed on path p. For any UE assignment $\boldsymbol{h}^{\text{eq}}$ and induced $\boldsymbol{x}^{\text{eq}}$ in the tolled game instance $\mathcal{G}(G,\mathcal{L}^{\boldsymbol{t}},\mathcal{W},\mathcal{Z})$, and for any commodity $i=1,\dots,K$, define the set of minimal cost paths

$$\mathcal{P}_i^\star = \arg\min_{p\in\mathcal{P}_i} l_p(\boldsymbol{x}^{\text{eq}}). \tag{10}$$

From the UE conditions in the tolled game (see definition II.1), we have for any $p_i^\star\in\mathcal{P}_i^\star$:

$$l_p(\boldsymbol{x}^{\text{eq}}) + \tau_p \le l_{p_i^\star}(\boldsymbol{x}^{\text{eq}}) + \tau_{p_i^\star} \quad \forall p\in\mathcal{P}_i : h_p^{\text{eq}} > 0 \tag{11}$$

Thus

$$\begin{aligned} R_p(\boldsymbol{x}^{\text{eq}}) =& l_p(\boldsymbol{x}^{\text{eq}}) - l_{p_i^\star}(\boldsymbol{x}^{\text{eq}}) \\ \leq& \tau_{p_i^\star} - \tau_p \qquad \forall p \in \mathcal{P}_i : h_p^{\text{eq}} > 0 \end{aligned} \tag{12}$$

Taking the maximum of both sides with respect to all $p \in \mathcal{P}_i$ such that $h_p^{\text{eq}} > 0$ we get

$$\begin{aligned} \max_{p \in \mathcal{P}_i : h_p > 0} R_p(\boldsymbol{x}^{\text{eq}}) &\leq \max_{p \in \mathcal{P}_i : h_p > 0} \left[\tau_{p_i^\star} - \tau_p\right] \\ &= \tau_{p_i^\star} - \min_{p \in \mathcal{P}_i : h_p > 0} \tau_p \\ &\leq \tau_{p_i^\star} - \min_{p \in \mathcal{P}_i} \tau_p \\ &\leq \max_{p \in \mathcal{P}_i} \tau_p - \min_{p \in \mathcal{P}_i} \tau_p. \end{aligned} \tag{13}$$

By taking the maximum with respect to the commodities we complete the proof.

Remark 2. *The bound on the toll vector in proposition III.1 can be a conservative upper bound on the worst-case regret. This is due to two inequalities that might not be tight at the solution. The first inequality is in eq.* (13)*, which results from taking the minimum tolls over the set of all paths* $\mathcal{P}_i$ *between an OD pair rather than the set of utilized paths* $\{p \in \mathcal{P}_i | h_p^{eq} > 0\}$*. The second inequality is in eq.* (13)*, which results from replacing the tolls of the minimal cost path* $p_i^\star$ *with the maximum tolls over the set of all paths* $\mathcal{P}_i$ *between the OD pair. These steps enable us to derive a set of linear constraints on the tolls that guarantee bounded regret without knowing the set of utilized paths or the set of minimal cost paths* $\mathcal{P}_i^\star$*. If these sets are known prior to computing the tolls, a less stringent set of constraint of the form*

$$\min_{p \in \mathcal{P}_i^\star} \tau_p - \min_{p \in \mathcal{P}_i : h_p > 0} \tau_p \leq \epsilon \qquad \forall i = 1, \ldots, K$$

would suffice to guarantee an ϵ*-bound on the worst-case regret. These constraints will give a tight upper bound on the regret if there exists* $p_i^\star \in \mathcal{P}_i^\star$ *such that* $h_{p^\star}^{eq} > 0$ *for all commodities. We attempt to improve the constraints in proposition III.1 by proposing a procedure that alternates between: (1) finding optimal tolls for a given set of utilized and minimal cost paths, and (2) updating these sets based on the resulting UE assignment given the tolls.*

C. Proposed algorithm

Proposition III.1 gives us a guide on how to use linear constraints on tolls to control for the regret of the resulting UE assignment. Based on this we propose the following three-step algorithm as a heuristic to solving problem CSO. This approach is similar to the one proposed in [34] for solving second-best tolls with support constraints.

Step 1. Find the edge flow vector $\boldsymbol{x}^\star$ that is induced by an unconstrained SO assignment (as per Def II.2). This can be done using Frank-Wolfe (FW) algorithm as described in [38].

Step 2. Compute the toll vector $\tilde{\boldsymbol{t}}$ as the solution to the *constrained* dual linear program

$$\begin{aligned} \max_{\boldsymbol{z},\boldsymbol{t}} \quad & \sum_i z_i d_i - \sum_{e \in E} t_e x_e^\star \\ \text{s.t.} \quad & z_i - \sum_{e \in p} t_e \leq l_p(\boldsymbol{x}^\star) \\ & \delta_i(\boldsymbol{t}) \leq \epsilon \qquad \forall i = 1, \ldots, K. \end{aligned} \tag{$\tilde{D}_{\boldsymbol{x}^\star}$}$$

Step 3. Compute the resulting assignment $\tilde{\boldsymbol{h}}$ as a UE of the game instance $\mathcal{G}(G, \mathcal{L}^{\tilde{\boldsymbol{t}}}, \mathcal{W}, \mathcal{Z})$.

a) Refining the toll constraints: As discussed in remark 2, the bound on the worst-case regret guaranteed by the algorithm can be conservative. For this, we propose an extension to the three-step algorithm described above to refine the constraints on the toll vector in problem $\tilde{D}_{\boldsymbol{x}^\star}$ to achieve a tighter bound on the worst-case regret. Precisely, we define a set of *active paths* $\bar{\mathcal{P}}_i$ for each commodity $i = 1, \ldots, K$, which is the set of paths on which the toll constraints are imposed. Initially, we follow the three-step algorithm which results in assignment $\tilde{\boldsymbol{h}}$ and edge flows $\tilde{\boldsymbol{x}}$. We then initialize $\bar{\mathcal{P}}_i$ for each commodity $i = 1, \ldots, K$ as

$$\bar{\mathcal{P}}_i = \{p \in \mathcal{P}_i : \tilde{h}_p > 0\} \cup \arg\min_{p \in \mathcal{P}_i} l_p(\tilde{\boldsymbol{x}}).$$

Then in every iteration $n = 1, 2, \ldots$ we perform the following steps to update $\bar{\mathcal{P}}_i$ for all $i = 1, \ldots, K$:

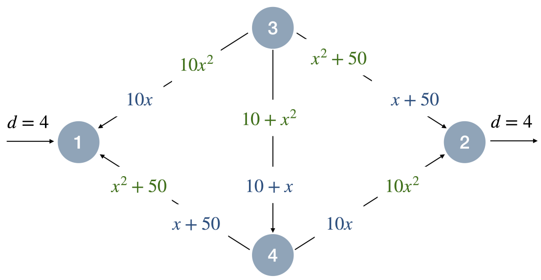
1) Solve for the toll vector

$$\begin{aligned} \tilde{\boldsymbol{t}}^{(n)} = \arg\max_{\boldsymbol{z},\boldsymbol{t}} \quad & \sum_i z_i d_i - \sum_{e \in E} t_e x_e^\star \\ \text{s.t.} \quad & z_i - \sum_{e \in p} t_e \leq l_p(\boldsymbol{x}^\star) \\ & \max_{p \in \bar{\mathcal{P}}_i} H_p^\top \boldsymbol{t} - \min_{p \in \bar{\mathcal{P}}_i} H_p^\top \boldsymbol{t} \leq \epsilon \quad \forall i \end{aligned}$$

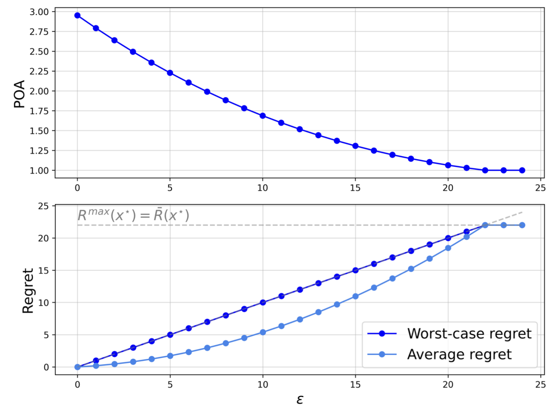
2) Compute the resulting equilibrium assignment $\tilde{\boldsymbol{h}}^{(n)}$ and $\tilde{\boldsymbol{x}}^{(n)}$ in the game instance $\mathcal{G}(G, \mathcal{L}^{\tilde{\boldsymbol{t}}^{(n)}}, \mathcal{W}, \mathcal{Z})$
3) Identify the set of utilized paths $\mathcal{P}_i^n = \{p \in \mathcal{P}_i : \tilde{h}_p^{(n)} > 0\}$ and the set of minimal travel cost paths $\mathcal{P}^{\star(n)} = \arg\min_{p \in \mathcal{P}_i} l_p(\tilde{\boldsymbol{x}})$
4) Terminate and output $\tilde{\boldsymbol{h}}^{(n)}$ and $\tilde{\boldsymbol{x}}^{(n)}$ if $\mathcal{P}_i^n \cup \mathcal{P}^{\star(n)} \subseteq \bar{\mathcal{P}}_i$. Otherwise, set $\bar{\mathcal{P}}_i = \bar{\mathcal{P}}_i \cup \mathcal{P}_i^n \cup \mathcal{P}^{\star(n)}$ and repeat.

Note that this procedure has to converge since the set of paths is finite. In our numerical experiments, we show that it actually converges in a few steps. We also note that the output of this procedure is guaranteed to have an ϵ-bound on the worst-case regret since in the last iteration the constraints were imposed on all utilized paths and minimal travel cost paths. This refinement offers significant improvement in the regret bound and the social cost as illustrated in the numerical experiments in section IV.

b) Computational Complexity: The proposed algorithm involves solving two convex optimization problems and one LP. The first convex problem in **Step 1** can be solved using the edge-based FW algorithm. The linear program $\tilde{D}_{\boldsymbol{x}^\star}$ in **Step 2**, however, imposes constraints on the path tolls so it can have an exponential-size. We note here that the



(a) Braess's network with a single commodity with demand $d = 4$. The users' cost functions (blue) are affine and the planner's cost functions (green) are quadratic.



(b) Results for Braess's network showing the decay in the POA (top) and values of the of worst-case and average regret (bottom) as ϵ increases.

Fig. 1: Braess's network.

original dual problem $D_{\boldsymbol{x}^\star}$ can be written in polynomial-size using the node-edge formulation as discussed in [18], [33]. Furthermore, path-based algorithms [38] can be used to compute the resulting UE assignment in **Step 3**. Such algorithms do not require the enumeration of all paths, and work only with the set of utilized paths for each commodity. Using the resulting assignment, along with a shortest path algorithm, one can identify the sets of active paths $\bar{\mathcal{P}}_i$ for each commodity and apply the proposed refinement which only imposes toll constraints on paths in the sets $\bar{\mathcal{P}}_i$ $\forall i = 1, \ldots, K$.

IV. Numerical Experiments

To illustrate the performance of the proposed algorithm, we perform a set of numerical experiments on benchmark networks. In these experiments, we vary the regret bound ϵ and study the trade-off between the regret and the efficiency of the resulting assignment.

a) Evaluation metrics: For assignment $\tilde{\boldsymbol{h}}$ (and induced edge flows $\tilde{\boldsymbol{x}}$) that is the outcome of the proposed algorithm, we measure the unfairness using the two notions of regret introduced in definitions II.3 and II.4. We measure the efficiency of the assignment using the price of anarchy defined as

$$\text{POA} = \frac{C(\tilde{\boldsymbol{x}})}{C(\boldsymbol{x}^\star)}. \tag{14}$$

In this definition, the numerator is the social cost of the equilibrium assignment in the *tolled network*. The social cost function $C(\cdot)$ is evaluated with respect to the planner's cost functions, which can be different from the costs that define the equilibrium. The theoretical bounds on the POA known in the literature [9] do not extend to this definition.

b) Networks: We perform experiments on two benchmark networks with varying complexities. We start with the well-studied Braess's network structure with a single commodity. In this network, the users' cost functions are affine in the flow while the planner's cost functions are quadratic as shown in fig. 1a.

The second network is a synthetic simplification of the Bay Area's road network with 13 commodities as shown in fig. 2a. For the users' cost functions we use the Bureau of Public Road (BPR) travel time function

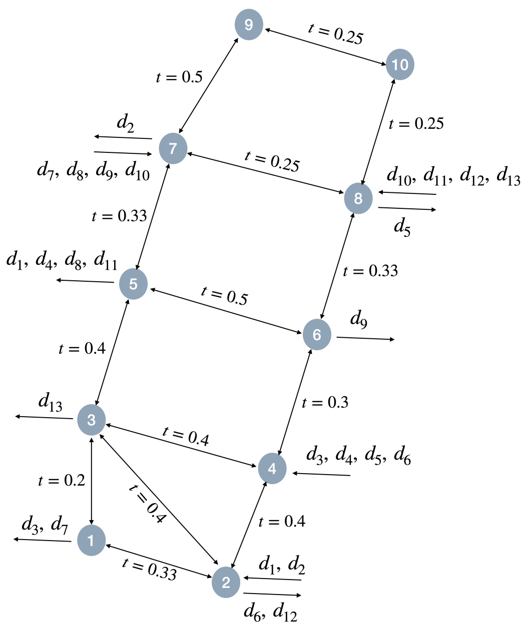
$$l_e(x_e) = t_e\left(1 + \alpha\left(\frac{x_e}{c_e}\right)^\beta\right), \tag{15}$$

where t_e, c_e are the edge's free-flow travel time (in hours) and capacity (in veh/hour), respectively, and $\alpha = 0.15$ and $\beta = 4$ are model parameters. The capacity is set to $c_e = 1000$ for all edges and the values of t_e are shown in fig. 2a. The total demand in the network is set to 12000 vehicles distributed among the 13 commodities. We first consider the case where the planner has the same cost functions as the users (i.e. $w_e(x_e) = l_e(x_e)$ for all $e \in E$). We then consider the case where the planner has affine cost functions that take the same form as the users' cost with $\beta = 1$.

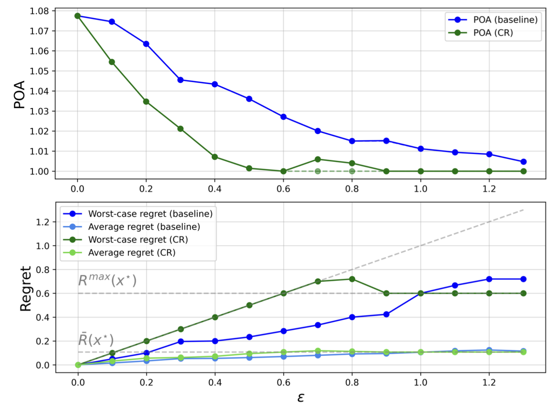
c) Results: In fig. 1b, we show the POA (top) and the regret (bottom) in Braess's network at different values of the bound ϵ. In this network, the unconstrained SO assignment has a worst-case regret of 22.0 (which is 0.44 times the travel cost of the shortest path in the network). The price of anarchy of the UE assignment (with zero regret) is POA $= 2.95$ and it drops rapidly to 1 as epsilon increases. Using the proposed algorithm, we reduce the POA by more than 50% (to 1.44) while maintaining a worst-case regret of 13 (which is 0.19 times the travel time of the shortest path).

In this network, the worst-case regret bound is consistently tight until the optimal assignment is reached. This is because all the paths are utilized in the resulting assignments. Consequently, it is not necessary to use the proposed constraints refinement. This, however, is seldom the case in more realistic networks where the set of utilized paths is typically a small subset of all paths.

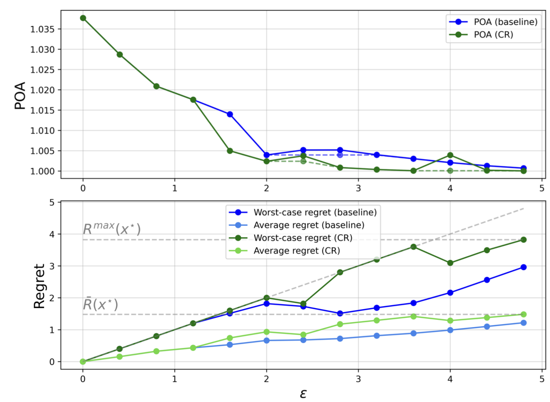
In fig. 2, we show the results for the synthetic Bay Area network. We note that the worst-case regret is computed based on paths that carry at least 1% of the demand of



(a) Synthetic Bay Area network with 10 nodes, 14 edges, and 13 commodities.



(b) Results for the synthetic Bay Area network with the same cost functions for the users and planner.



(c) Results for the synthetic Bay Area network with different cost functions for the users and planner.

Fig. 2: Synthetic Bay Area network. In (b) and (c) we report the results of the proposed algorithm with and without the constraint refinement (CR) procedure. The dashed line associated with each POA plot indicates the outcomes with the lowest POA achieved up to the current value of ϵ.

the corresponding commodity. This is done to avoid large regret values experienced by a negligible fraction of the population (on paths with very small flows). In the first case where the cost functions are aligned (fig. 2b), the unconstrained SO assignment has a worst-case regret of 0.6 hours (which is 1.17 times the travel cost of the shortest path for the commodity that suffers the worst-case regret). Using our algorithm, we reduce the worst-case regret by 50% while increasing the social cost by only 2.1% relative to the unconstrained SO assignment. With a 75% reduction in the worst-case regret, the social cost increases by 4.3%.

In the second case where the cost functions are different (fig. 2c), the unconstrained SO assignment has a worst-case regret of 3.8 hours (which is 2.85 times the travel cost of the shortest path for the commodity that suffers the worst-case regret). This is an example of severe unfairness induced by the SO assignment due to the misalignment between the users' and planner's objectives. While reducing the worst-case regret by 50% results in a minor increase in the social cost ($< 0.5\%$), this level of regret remains relatively high.

In fig. 2, we also compare the performance of the proposed algorithm with and without constraints refinement (CR). For both cases, the constraints refinement makes the worst-case regret bounds tighter and achieves more efficient assignments (with lower POA) for most values of ϵ. In all experiments, the procedure runs for at most 3 iterations before it converges.

In fig. 2, we observe that the algorithm's outputs do not always exhibit a monotonically decreasing POA as the value of ϵ increases, although a general trend is present. This is an artifact of the heuristic used to compute the toll vector at a given regret bound ϵ. To address this, one approach is to explore a neighborhood around the desired ϵ and identify the most efficient assignment that respects the regret bound. In fig. 2, the dashed lines represent the lowest POA values achieved up to each ϵ, indicating the most efficient outcomes observed so far while respecting the current regret bound ϵ. Finally, we note from fig. 2 that the average regret in the network is significantly lower than the worst-case regret, especially near the SO assignment.

V. Conclusion

In this work, we formulated the problem of optimal route recommendations as a constrained system optimum problem. We present an algorithm for finding toll schemes to improve an arbitrary social cost function while maintaining an upper bound on the users' regret. The proposed approach can be straightforwardly extended to the setting of heterogeneous users using the results in [18]. It also allows for defining different regret tolerances for different commodities in the network.

Our numerical results show a large gap between worst-case and average regrets. This motivates us to move beyond worst-case criteria for user compliance, and extend this work to study SO assignments with constraints on the average unfairness, which is not well-studied in the literature.

To gain a better understanding of the problem and to further investigate the performance of the algorithm there is a need for a more extensive numerical study with a larger variety of network structures and cost functions. We also aim to compare the performance of our proposed algorithm against other approaches [14], [20].

References

[1] Y. Wang, "Review of road traffic control strategies," *Proceedings of the IEEE*, vol. 91, no. 12, pp. 2043–2067, 2003.

[2] J. Macfarlane, "When apps rule the road: The proliferation of navigation apps is causing traffic chaos. it's time to restore order," *IEEE Spectrum*, vol. 56, no. 10, pp. 22–27, 2019.

[3] C. Chan, A. Kuncheria, and J. Macfarlane, "Simulating the impact of dynamic rerouting on metropolitan-scale traffic systems," *ACM Transactions on Modeling and Computer Simulation*, vol. 33, no. 1-2, pp. 1–29, 2023.

[4] T. Cabannes, M. A. S. Vincentelli, A. Sundt, H. Signargout, E. Porter, V. Fighiera, J. Ugirumurera, and A. M. Bayen, "The impact of gps-enabled shortest path routing on mobility: a game theoretic approach 2," *University of California, Berkeley*, vol. 29, 2017.

[5] J. Thai, N. Laurent-Brouty, and A. M. Bayen, "Negative externalities of gps-enabled routing applications: A game theoretical approach," in *2016 IEEE 19th international conference on intelligent transportation systems (ITSC)*. IEEE, 2016, pp. 595–601.

[6] A. Festa and P. Goatin, "Modeling the impact of on-line navigation devices in traffic flows," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 323–328.

[7] T. Cabannes, M. Sangiovanni, A. Keimer, and A. M. Bayen, "Regrets in routing networks: Measuring the impact of routing apps in traffic," *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, vol. 5, no. 2, pp. 1–19, 2019.

[8] D. Jalota, D. Paccagnan, M. Schiffer, and M. Pavone, "Online routing over parallel networks: Deterministic limits and data-driven enhancements," *INFORMS Journal on Computing*, vol. 35, no. 3, pp. 560–577, 2023.

[9] T. Roughgarden and É. Tardos, "How bad is selfish routing?" *Journal of the ACM (JACM)*, vol. 49, no. 2, pp. 236–259, 2002.

[10] M. Patriksson, *The traffic assignment problem: models and methods*. Courier Dover Publications, 2015.

[11] J. G. Wardrop and J. I. Whitehead, "Correspondence. some theoretical aspects of road traffic research." *Proceedings of the institution of civil engineers*, vol. 1, no. 5, pp. 767–768, 1952.

[12] T. Roughgarden, "How unfair is optimal routing?" in *Symposium on Discrete Algorithms: Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms*, vol. 6, no. 08, 2002, pp. 203–204.

[13] O. Jahn, R. H. Möhring, A. S. Schulz, and N. E. Stier-Moses, "System-optimal routing of traffic flows with user constraints in networks with congestion," *Operations research*, vol. 53, no. 4, pp. 600–616, 2005.

[14] D. Jalota, K. Solovey, M. Tsao, S. Zoepf, and M. Pavone, "Balancing fairness and efficiency in traffic routing via interpolated traffic assignment," *Autonomous Agents and Multi-Agent Systems*, vol. 37, no. 2, p. 32, 2023.

[15] H. S. Mahmassani and G.-L. Chang, "On boundedly rational user equilibrium in transportation systems," *Transportation science*, vol. 21, no. 2, pp. 89–99, 1987.

[16] A. Pigou, *The economics of welfare*. Routledge, 2017.

[17] M. Beckmann, C. B. McGuire, and C. B. Winsten, "Studies in the economics of transportation," Tech. Rep., 1956.

[18] L. Fleischer, K. Jain, and M. Mahdian, "Tolls for heterogeneous selfish users in multicommodity networks and generalized congestion games," in *45th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, 2004, pp. 277–285.

[19] E. Angelelli, I. Arsik, V. Morandi, M. Savelsbergh, and M. G. Speranza, "Proactive route guidance to avoid congestion," *Transportation Research Part B: Methodological*, vol. 94, pp. 1–21, 2016.

[20] E. Angelelli, V. Morandi, M. Savelsbergh, and M. G. Speranza, "System optimal routing of traffic flows with user constraints using linear programming," *European journal of operational research*, vol. 293, no. 3, pp. 863–879, 2021.

[21] E. Angelelli, V. Morandi, and M. G. Speranza, "Minimizing the total travel time with limited unfairness in traffic networks," *Computers & Operations Research*, vol. 123, p. 105016, 2020.

[22] Y. Lou, Y. Yin, and S. Lawphongpanich, "Robust congestion pricing under boundedly rational user equilibrium," *Transportation Research Part B: Methodological*, vol. 44, no. 1, pp. 15–28, 2010.

[23] P. Bergendorff, D. W. Hearn, and M. V. Ramana, *Congestion toll pricing of traffic networks*. Springer, 1997.

[24] M. B. Yildirim and D. W. Hearn, "A first best toll pricing framework for variable demand traffic assignment problems," *Transportation Research Part B: Methodological*, vol. 39, no. 8, pp. 659–678, 2005.

[25] D. W. Hearn and M. V. Ramana, "Solving congestion toll pricing models," in *Equilibrium and advanced transportation modelling*. Springer, 1998, pp. 109–124.

[26] L. Bai and P. A. Rubin, "Combinatorial benders cuts for the minimum tollbooth problem," *Operations research*, vol. 57, no. 6, pp. 1510–1522, 2009.

[27] L. Bai, D. W. Hearn, and S. Lawphongpanich, "Decomposition techniques for the minimum toll revenue problem," *Networks: An International Journal*, vol. 44, no. 2, pp. 142–150, 2004.

[28] R. B. Dial, "Minimal-revenue congestion pricing part ii: An efficient algorithm for the general case," *Transportation Research Part B: Methodological*, vol. 34, no. 8, pp. 645–665, 2000.

[29] S. Lawphongpanich and D. W. Hearn, "An mpec approach to second-best toll pricing," *Mathematical programming*, vol. 101, no. 1, pp. 33–55, 2004.

[30] E. T. Verhoef, "Second-best congestion pricing in general networks. heuristic algorithms for finding second-best optimal toll levels and toll points," *Transportation Research Part B: Methodological*, vol. 36, no. 8, pp. 707–729, 2002.

[31] V. V. Kalashnikov, R. C. Herrera Maldonado, J.-F. Camacho-Vallejo, and N. I. Kalashnykova, "A heuristic algorithm solving bilevel toll optimization problems," *The International Journal of Logistics Management*, vol. 27, no. 1, pp. 31–51, 2016.

[32] J. Ekström, L. Engelson, and C. Rydergren, "Heuristic algorithms for a second-best congestion pricing problem," *NETNOMICS: Economic Research and Electronic Networking*, vol. 10, pp. 85–102, 2009.

[33] M. Patriksson and R. T. Rockafellar, "A mathematical model and descent algorithm for bilevel traffic management," *Transportation Science*, vol. 36, no. 3, pp. 271–291, 2002.

[34] C. Maheshwari, K. Kulkarni, D. Pai, J. Yang, M. Wu, and S. Sastry, "Congestion pricing for efficiency and equity: Theory and applications to the san francisco bay area," *arXiv preprint arXiv:2401.16844*, 2024.

[35] M. Hoefer, L. Olbrich, and A. Skopalik, "Taxing subnetworks," in *International workshop on internet and network economics*. Springer, 2008, pp. 286–294.

[36] V. Bonifaci, M. Salek, and G. Schäfer, "Efficiency of restricted tolls in non-atomic network routing games," in *Algorithmic Game Theory: 4th International Symposium, SAGT 2011, Amalfi, Italy, October 17-19, 2011. Proceedings 4*. Springer, 2011, pp. 302–313.

[37] X. Di, H. X. Liu, J.-S. Pang, and X. J. Ban, "Boundedly rational user equilibria (brue): mathematical formulation and solution sets," *Procedia-Social and Behavioral Sciences*, vol. 80, pp. 231–248, 2013.

[38] S. D. Boyles, N. E. Lownes, and A. Unnikrishnan, "Transportation network analysis," *Volume I, Version 0.85*, 2020.