2022 IEEE 61st Conference on Decision and Control (CDC)
December 6-9, 2022. Cancún, Mexico

Data-Driven Robust Congestion Pricing

Yize Wang, Dario Paccagnan

***Abstract*— Whilst overloaded transportation systems bear a significant impact on everyone's welfare, governments strive to improve their performances. Amongst the many solutions proposed, congestion pricing is becoming increasingly popular as it has the potential to reduce congestion by indirectly influencing the drivers' routing choices. Commonly advocated for, the marginal cost mechanism ensures that self-interested decision-making results in optimal system performances. However, such a mechanism suffers from three important drawbacks in that i) it requires levying tolls on every road, ii) it does not allow for upper bounds on the magnitude of the tolls, and iii) it is flow-dependent. In response to these challenges, researchers have introduced the *restricted network tolling* problem, seeking constant tolls of bounded magnitude that induce equilibria with a small social cost. However, tolls designed through this approach are tailored to a specific traffic demand, resulting in a design that has the potential to exacerbate the very issue it set out to solve, if the demand changes. Our work addresses this issue and aims at infusing robustness guarantees to the restricted network tolling problem. We do so by seeking tolls that have good performance over past demand realizations, and leverage recent results in scenario optimization to equip our design with formal generalization guarantees.**

I. Introduction

Mobility systems are facing unprecedented challenges as urban areas suffer from overloaded transportation systems causing users' dissatisfaction, pollution, and health dangers [27], [33]. Given the societal impact of these issues, a portfolio of solutions is being developed, including vehicle electrification, autonomous mobility, congestion pricing and many more. These measures often operate on different time scales, with a long term goal of moving away from car ownership to more sustainable mobility paradigms, mid term goal of reducing emissions, e.g., through electrification, and short term goal of optimizing traffic flow and drivers' behaviour.

Within the latter class of measures, congestion pricing, consisting in levying tolls for travelling through certain roads, is becoming increasingly popular. At its core, congestion pricing aims at influencing the way drivers route themselves with the goal of minimizing overall congestion and corresponding emissions. Its rationale lies in the observation that, when drivers aim solely at minimizing their individual travel time, congestion could be much higher compared to that achievable with central coordination.

Dating back to the seminal works of Pigou and Beckmann [1], [25], congestion pricing schemes are most commonly studied within the realm of non-atomic congestion games, where flows of self-interested vehicles route themselves along origin-destination pairs, giving rise to a so-called Wardrop equilibrium [34]. In this context, congestion pricing schemes aim at aligning the resulting Wardrop equilibrium with a socially optimal solution. A mathematically elegant solution to this problem was proposed by Pigou [25] with the introduction of the marginal cost tolls, which charge each driver for the externality they create on the entire system.

While commonly advocated for, marginal cost tolls suffer from three important drawbacks that render them difficult to use in practice. First, such an approach requires tolling every road. Second, it is not possible to impose upper bounds on the tolls' magnitude. Third, marginal cost tolls are flow dependent, and thus difficult for the drivers to predict. Motivated by these issues, researchers in the Transportation and Operations Research communities introduced the restricted network tolling problem[1], where a central planner aims at designing tolls that minimize the equilibrium congestion while limiting both their magnitude and the set of roads where such tolls can be levied [15]. From a computational standpoint, the restricted network tolling problem is more challenging as it requires solving a bilevel program. However, its main drawback lies in the lack of robustness: restricted network tolling advocates for the design of a fixed congestion pricing scheme that relies on knowledge of the exact traffic demand. Nonetheless, traffic demand is, in practice, difficult to predict and often highly volatile. The result is a design that has the potential to exacerbate the very issue it set out to solve [2].

Our work tackles such a challenge and aims at infusing robustness guarantees to the restricted network tolling problem. To the best of our knowledge, no other work has pursued this direction. We do so by blending the bilevel formulation arising in the restricted network tolling with recent developments in scenario optimization. Our core contributions can be summarized as follows:

1) We introduce and formulate the probabilistically robust restricted network tolling problem, where we seek flow-independent tolls that provide good performances against past demand realizations;
2) We propose a gradient-based method for determining a local optimal solution of the resulting bilevel program.
3) We exploit recent results in scenario optimization to provide formal out-of-sample guarantees to the tolling mechanism determined through our approach. Interestingly, such guarantees hold regardless of whether our algorithm terminates with an optimal solution or not;
4) We test our algorithm on a set of commonly employed instances. Our results show that tolling few edges suffices to obtain performances similar to those of the marginal cost toll, and that such performances hold robustly against unseen demand realizations taken form the same distribution generating the past demands.

[1] Also referred to as "second-best tolling", in the sense of second-best after the marginal cost tolls [19].

2022 IEEE 61st Conference on Decision and Control (CDC) | 978-1-6654-6761-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/CDC51059.2022.9993033

Related Works

The problem of designing congestion pricing mechanisms ensuring a desirable system-level behaviour was initially studied in a landmark work by Pigou and continues to be of interest [25]. Since then, a large body of literature has explored the power and limitations of such mechanisms within different settings. Broadly speaking, we distinguish the literature in simulation-based and model-based approaches.

Simulation-based approaches, typically leveraging agents-based simulators, have developed significantly over the past years especially within the transportation literature. However, while such frameworks allow for a fine-grained analysis of the transportation network, they are also generally unfit for optimization as their complexity renders this task prohibitive both from a computational and a mathematical standpoint. Here, the majority of the literature focuses on the analysis of *given* congestion pricing mechanisms, e.g., [9], [29], with recent attempts to develop surrogate-models that are more optimization-friendly [22].

On the other hand, model-based approaches usually build upon the original framework of atomic congestion games or their continuous flow counterpart. For both settings the problem is well understood if no limitations on the tolls magnitude and on the tollable roads are imposed. In the non-atomic setting, the seminal work of [25] shows that marginal cost tolls induce Wardrop equilibrium with minimum social cost. The same question proved to be more difficult when posed within the original atomic setting, with [23] providing an answer only very recently. Specifically [23] derives tight hardness-of-approximation results for the problem of minimizing the social cost, and develops polynomially computable tolls with optimal approximation factor.

Designing congestion pricing mechanisms becomes significantly more challenging when limitations on the tolls magnitude or on the tollable roads are imposed. The corresponding problem is commonly referred to as the restricted network tolling problem. Such problem is known to be NP-hard [15] already within the continuous flow model, with typical solution approaches relying on single level reformulations of the original bilevel program. Within this domain, the literature focuses either on efficient heuristics for small-size problems [11], [32], or on exact algorithms without convergence-time guarantees [18]. However, none of the existing work addresses the following crucial point: *how does demand variability impact the quality of the resulting mechanism?* Our work takes a first step in this direction by designing congestion pricing schemes whose performance is certified *a priori* against future demand realizations taken from the same distribution generating the past demand.

Organization

The remainder of this work is organized as follows. Section II introduces the model we adopt for traffic routing with tolls and formulates the probabilistically robust version of the restricted network tolling problem. Section III introduces the gradient-based algorithm we use to solve the resulting optimization program. We test our algorithm on a commonly employed dataset and present the results in Section IV before moving to the Conclusions in Section V.

II. Model

A. Preliminaries

We describe the road traffic infrastructure through a connected directed network $G = (V, E)$, where V and E represent the sets of vertices and edges, with corresponding cardinality $|V|$ and $|E|$. As common in the literature, vertices model intersections while edges describe road links. We denote a directed edge with $e = (v_1, v_2)$, where v_1 is the starting node. We then introduce a commodity set C, which contains pairs of origin-destination vertices (O^i, D^i) with $i \in I$, where I is an index set. Commodity i is associated with a flow demand d^i which needs to be routed from O^i to D^i. We represent the traffic flow on edge e due to the i^{th} commodity as f_e^i. Throughout the manuscript, we collect all the components of f_e^i for $e \in E$, $i \in I$ in the vector $f \in \mathbb{R}^{|E|\times|I|}$, and let f_e accumulate the traffic flows on edge e due to all commodities, i.e.,

$$f = \{f_e^i\}_{e\in E, i\in I}, \qquad f_e = \sum_{i\in I} f_e^i. \tag{1}$$

In order to route all passengers from their corresponding origin to their destination while satisfying flow conservation at the vertices, f must satisfy the following set of conditions (one condition per each vertex and commodity)

$$\sum_{e:v_1=v} f_e^i - \sum_{e:v_2=v} f_e^i = \begin{cases} d^i & \text{if } v = O^i \\ -d^i & \text{if } v = D^i \\ 0 & \text{otherwise} \end{cases}, \forall v \in V, \forall i \in I. \tag{2}$$

Eq. (2) requires the sum of incoming and outgoing flows associated with each commodity i to equal zero in every vertex, except for the origin (resp. destination) where an amount of flow d_i is injected (resp. extracted) from the network. This set of linear constraints can be compactly expressed as $Af = b$ for suitably defined A and b.

In the following, we adopt the model of non-atomic congestion games, and consider edges as congestible resources. When travelling through each edge of a chosen path, users incur a cost depending on the fraction of them currently travelling on that very same edge. The specific cost incurred is determined through a so-called latency function $\ell_e : \mathbb{R} \to \mathbb{R}$, assumed to be non-negative, differentiable, increasing, semi-convex in its argument. In our work, we consider the commonly employed BPR-type latency function introduced in [21], although our results extend to any latency function satisfying the above-mentioned properties,

$$\ell_e(f_e) = t_e \left(1 + B_e \left(\frac{f_e}{C_e}\right)^P\right), \tag{3}$$

where t_e, B_e, C_e, P are positive constants depending on the geometric properties of edge e. Finally, when congestion pricing mechanisms are utilized, users are additionally requested to pay a monetary toll $\tau_e \in \mathbb{R}_{\geq 0}$ to travel on edge e. We collect all these tolls in the vector $\tau = \{\tau_e\}_{e \in E}$.

We assume all users are rational and aim at routing themselves on paths that minimize their individual cost, intended as the sum of latencies and tolls. This is a commonly studied setting corresponding to considering users with the same sensitivity, that is, users valuing 1 unit of time as s units of money. Here we set $s = 1$ without loss of generality.[2]

This model for the emergent behaviour is typically captured by the well-known notion of Nash equilibrium, describing a situation where no user can decrease their own cost by means of any unilateral deviation [20]. However, in real traffic systems, there is often a large number of users concurrently on the road, each with a small impact on the congestion levels. This fact motivates the use of the so-called Wardrop equilibrium [34] which postulates the presence of infinitely many players, each controlling only a negligible portion of the traffic flow. Informally, a Wardrop equilibrium is a set of flows for each origin-destination pair such that every path with non-zero flow has equal and minimal travel time. Such equilibrium notion formally arises as the limit of a sequence of Nash equilibria for users populations of increasing size and decreasing individual impact [7], [14], [24]. Interestingly, the seminal work [1] shows that, under very mild assumptions on the latency functions (satisfied in our context), a Wardrop equilibrium can be equivalently obtained by solving the following convex optimization problem

$$\begin{aligned} &\min_f \sum_e \int_0^{f_e} (\ell_e(t) + \tau_e)\ \mathrm{d}t \\ &\text{subject to:} \quad Af = b, \quad f \geq 0, \end{aligned} \tag{4}$$

where f_e is linked to the flow vector f thanks to (1). Remarkably, the solution to (4) always exists and produces the same aggregate flow f_e over all edges when the latency functions are increasing, despite the fact that the optimal solution needs not be unique [1], [8]. For this reason, an optimal solution is often referred to as "essentially unique".

As anticipated earlier, self-interested decision making is often inefficient, corresponding to the fact that the social cost at a Wardrop equilibrium could be much higher than the minimum social cost. In the current context, we measure the social cost of a set of feasible flows using the total congestion. This is the most common approach in the literature, which accounts for all users' travel time

$$SC(f) = \sum_e f_e \ell_e(f_e). \tag{5}$$

Correspondingly, a socially optimal solution is a set of feasible flows minimizing the cost in (5).

[2] All forthcoming results continue to holds if $s \neq 1$, $s > 0$, upon utilizing the *weighted* sum of latencies and tolls with weights 1 and $1/s$. On the other hand, the problem becomes significantly more challenging if the agents are not homogeneous in their valuation of time and money. While this situation is often modelled assuming the presence of a sensitivity distribution [6], discussing this further is beyond the scope of our work.

To quantitatively evaluate the inefficiency of equilibria, the seminal paper [16] introduced the notion of *price of anarchy* (PoA), which describes the ratio between the worst social cost at an equilibrium and the optimal social cost. However, owing to the "essential uniqueness" property discussed above, all Wardrop equilibria produce the same social cost. Therefore, we can equivalently define the price of anarchy in this context as

$$\mathrm{PoA}(\tau) = \frac{SC(f_\tau^{eq})}{SC(f^{opt})}, \tag{6}$$

where f_τ^{eq} denotes a Wardrop equilibrium for given tolls τ, i.e., a solutions of (4), while f^{opt} denotes an optimal solution, i.e., a set of feasible flows minimizing the social cost (5). We remark that the price of anarchy depends on the chosen set of tolls, as different toll vectors induce different equilibria, which result in different efficiency levels.

Here, the goal of the central planner is to derive tolls that minimize the price of anarchy. We now turn the attention to this problem, starting from the (possibly unrealistic) setting where the traffic demand is known a priori.

B. Deterministic Toll Design

When the traffic demand is known a priori, we can formulate the restricted network tolling problem as a bilevel program. In the upper level we seek tolls that minimize the price of anarchy, or equivalently that minimize the social cost at equilibrium, since the socially optimal cost does not depend on the choice of tolls. We then use the lower level to ensure that the flow is a Wardrop equilibrium with respect to the chosen tolls. We let $T \subseteq \mathbb{R}_{\geq 0}^{|E|}$ denote the set of feasible tolls, assumed to be convex and compact, and use it to model, e.g., upper bounds on the tolls or restrictions on the set of tollable roads. The resulting program reads

$$\begin{aligned} &\min_{f, \tau \in T} \sum_e f_e \ell_e(f_e) \\ &\text{subject to:} \quad f \in \arg\min_f \sum_e \int_0^{f_e} (\ell_e(t) + \tau_e)\ \mathrm{d}t \\ &\qquad\qquad\quad Af = b, \quad f \geq 0. \end{aligned} \tag{7}$$

We remark that, for given tolls τ, the lower level problem forms a convex optimization as both the objective function and the constraints are convex. However, the bilevel nature of the problem makes it NP-hard and practically difficult to tackle for large-scale networks [15], [32]. More worryingly, the set of tolls is optimized for a specific choice of traffic demand, assumed to be known a priori. As a result, natural fluctuations in the demand can compromise the quality of the proposed set of tolls to the point of worsening the very problem they set out to solve [2]. We tackle this issue in the ensuing section.

C. Robust Toll Design

In this section we aim to immunize the solution of the restricted network tolling problem against potential variations in the traffic demand. The idea is to leverage readily available demand data from past days, for example, all afternoons in a year. These demands may be regarded as samples from an unknown distribution. We then introduce the *probabilistically robust restricted network tolling problem* where we aim to design tolls that perform well with respect to these past demand samples. By doing so, we are able to exploit recent advances in scenario optimization [5] and provide formal generalization guarantees to future demand realizations taken from the same (unknown) distribution. In the following, we provide a brief overview of scenario optimization, before turning the attention to introducing the probabilistically robust restricted network tolling problem.

Scenario optimization is a general methodology for data-driven optimization that allows to take decisions based on past data and quantify their performance against unseen future realizations taken from the same data-generating distribution. This approach was pioneered in [3], [4] and has, by now, reached a remarkable level of maturity. Interestingly, the approach allows to provide tight evaluations of the future performance without requiring any knowledge on the distribution generating the data (distribution-free). Furthermore, in its most recent version, scenario optimization applies beyond the classical optimization framework to a more general decision-making setting, thus allowing to provide generalization guarantees even for feasible solutions that are *not* necessarily optimal. This is important in our setting, as the computational difficulties described above prohibit us from developing an algorithm that efficiently computes optimal tolls.

We follow this approach, and assume to have access to $j \in \{1, \dots, S\}$ independent past demand realizations, also referred to as scenarios. Note that each demand realization is associated with a vector b entering in the demand-satisfying constraint arising in (2) and compactly denoted by $Af = b$. Therefore, in the following we denote with $b^{(j)}$ the vector corresponding to demand realization j and let $\mathbb{P}$ denote the (unknown) probability distribution from which vectors b are generated, and Δ its support.

We then aim at designing tolls that behave well with respect to all past demand realizations. We formalize this requirement by seeking tolls that minimize the social cost at the worst-case empirical realization. This results in the following epigraphic reformulation of the problem

$$
\begin{aligned}
&\min_{\tau \in T} p \\
&\text{subject to:} \\
&\qquad p \geq SC(f^{(j)}) \,/\, SC^{opt(j)} && \forall j \\
&\qquad SC(f^{(j)}) = \sum_e f_e^{(j)} \ell_e(f_e^{(j)}) && \forall j \\
&\qquad f^{(j)} \in \arg\min \sum_e \int_0^{f_e^{(j)}} (\ell_e(t) + \tau_e) \, \mathrm{d}t && \forall j \\
&\qquad Af^{(j)} = b^{(j)}, \quad f^{(j)} \geq 0, && \forall j
\end{aligned} \tag{8}
$$

Note that, for each fixed demand realization, the socially optimal cost $SC^{opt(j)}$ does not depend on the choice of tolls. As such, it can be pre-computed ahead of time and treated as a fixed constant in the above program.

Working in the setting already introduced, the decision-based variant of scenario optimization wishes to quantify the performance of a decision-making algorithm that maps a set of independent and identically distributed (i.i.d.) realizations to a decision vector. Such decision vector is required to belong to the intersection of S feasible sets, each produced by a scenario. Within this framework, the result in [5] provides generalization guarantees on the likelihood that the decision vector remains feasible for the constraints produced by a newly generated scenario.

In our case, the set of i.i.d. realizations corresponds to the set of past traffic demands embedded in $\{b^{(j)}\}_{j=1}^S$, while the corresponding decision vector represents the pair $\theta^* = (p^*, \tau^*)$ produced by any user-specified algorithm providing a feasible solution to (8), which we denote by $\mathcal{A}_S : (b^{(1)}, \dots, b^{(S)}) \mapsto \theta^*$. As in the above, each scenario produces a new constraint on the decision vector, which we refer to as $\Theta^{(b_j)}$. In our context, we require the set of tolls τ^* to induce a price of anarchy at most equal to p^*, as encoded in the set of constraints in (8). The generalization guarantee provided will then translate to assessing the likelihood that the chosen design τ^* induces a price of anarchy no higher than p^*. In other words, we will obtain a probabilistic robustness guarantee on the price of anarchy produced by the chosen set of tolls τ^* against new traffic demands.

The core idea behind the decision-making variant of scenario optimization relies on the notion of support subsample of an algorithm, which quantifies the number of scenarios needed to reconstruct the solution originally obtained when all scenarios are employed. Formally, [5] defines a support subsample as any subset of the set of scenarios based on which the algorithm reproduces the exact output of $\mathcal{A}_S(b^{(1)}, \dots, b^{(S)})$. Leveraging this notion, they show that the probability that the decision vector θ^* violates a newly generated constraint can be upper-bounded as follows. Specifically, letting $V(\theta^\star) = \mathbb{P}\{b \in \Delta \text{ s.t. } \theta^* \notin \Theta^{(b)}\}$, they prove that

$$\mathbb{P}^S[V(\theta^*) \geq \varepsilon(s^*, \beta)] \leq \beta, \tag{9}$$

where s^* is the cardinality of any support subsample, and $\varepsilon(s, \beta)$ is explicitly given in [5]. Interestingly, this result imposes virtually no assumptions on the algorithm $\mathcal{A}_S$, so long as it produces a feasible decision. In the following sections, we will exploit this result and equip our formulation with strong generalization guarantees. Towards this goal we utilize a greedy algorithm to derive a support subsample, where we remove one by one realizations that are not binding for the first set of constraints in (8), and thus do not modify the output of the algorithm.

III. A GRADIENT-BASED ALGORITHM

A commonly employed approach to solve the bilevel program in (7) consists in reformulating it as a single level program by replacing the lower level with its Karush-Kuhn-Tucker conditions [17]. The resulting program is then formulated as a mixed-integer program. While already computationally challenging in the settings of (7), such an approach is intractable when considering (8) due to the presence of multiple lower level programs, one per each realization.

In light of this, we proceed along a different direction, and begin observing that the objective function of (8), equivalently defined as $J(\tau) = \max_j \{SC(f^{(j)}) \,/\, SC^{opt(j)}\}$, where $f^{(j)}$ depends on τ through the solution of the Wardrop equilibrium problem, is continuous in the toll vector τ, following a general result on the sensitivity of the solution to a parametric optimization problem [30] and the "essential uniqueness" property mentioned before. Under mild technical assumptions, such cost function can be further shown to be piecewise differentiable in the toll vector [26]. Additionally, each lower level program can be solved very efficiently even for large scale networks if the vector of tolls is fixed, as such program corresponds to a classical traffic assignment problem [12] for which recent developments, e.g., on contraction hierarchies provide remarkable performance [13].

For these reasons, we propose to use a gradient projection algorithm over the space of feasible tolls [10]

$$\tau^{k+1} = \text{Proj}_T[\tau^k - \gamma^k \nabla J(\tau^k)],$$

where $\text{Proj}_T(\tau)$ denotes the metric projection of τ onto the convex set T and γ^k is a diminishing step size. We utilize, at each step, a numerically computed gradient evaluated based where we approximate the directional derivative $\partial_{\tau_e} J$ with $(J(\tau+\delta\mathbb{1}_e)-J(\tau))/\delta$, thus requiring to solve only two traffic assignment problems.[3] We then proceed to solve each traffic assignment problem using the classic Frank-Wolfe algorithm, which is known to converge with a rate of $\mathcal{O}(1/k)$ [30]. This algorithm requires the calculation of multiple shortest paths, which we tackle using contraction hierarchies [13].[4]

As the problem is inherently non-convex, the projected gradient algorithm may converge to an unsatisfying local optimum. To ameliorate this issue, we utilize a multi-start strategy, where we run the algorithm multiple times with different initial conditions.

Each run returns a set of tolls with the best price of anarchy over the available scenarios, and a probabilistic robustness guarantee obtained as discussed in Section II-C. In other words, the multi-start strategy yields multiple set of tolls each associated with a performance-robustness pair. Some tolls may result in a better performance at the cost of robustness, or the opposite. The decision-maker can then prioritize the performance or the robustness based on their preference.

IV. Numerical Results

We illustrate our methodology and results on a commonly employed case study, known as the Sioux Falls network. The demand data and network structure are available from [31], which we depict in Fig 1.

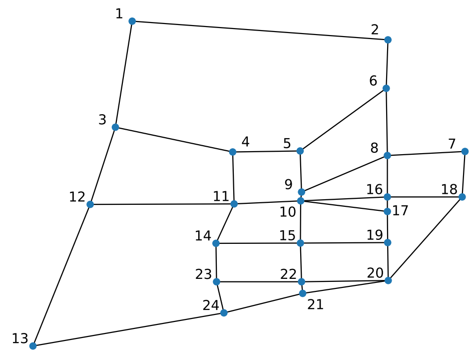
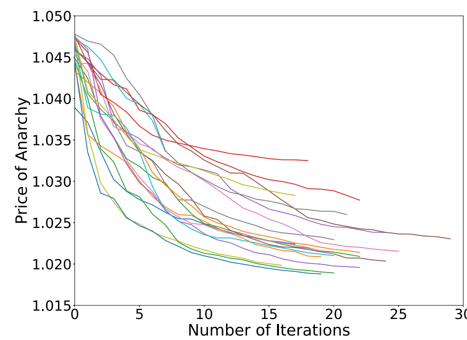


Fig. 1: Sioux Falls network. There are 24 vertices, 76 edges (38 two-ways roads), and 528 commodities. Each road is equipped with a BPR-type latency function.



(a) With Step Search

Fig. 2: Algorithm evolution for different initial conditions.

To produce different scenarios, we treat the traffic demand in [31] as the expectation of an underlying stochastic process, and generate 100 sets of origin-destination demand. These correspond to a central planner having access to traffic demand for, e.g., 100 past days. Each scenario is generated as follows. A set of demands, corresponding to a scenario, is produced by rescaling each entry d^i to $\alpha^i d^i$, where α^i is uniformly distributed in $[1-\alpha, 1+\alpha]$. We refer to α as the variation in the demand. For example, when $\alpha = 5\%$, each α^i is i.i.d. uniformly in $[0.95, 1.05]$. We then run the projected gradient algorithm starting from multiple randomly-generated initial conditions taken from the uniform distribution on $[0, 1]$, and terminate it when $||\tau^{k+1} - \tau^k|| \leq 10^{-4}$. An example is shown in Fig 2, where we constrain the toll vector to satisfy $0 \leq \tau \leq 2$. Finally, we evaluate the robustness of each of these terminal solutions by employing the result in (9).

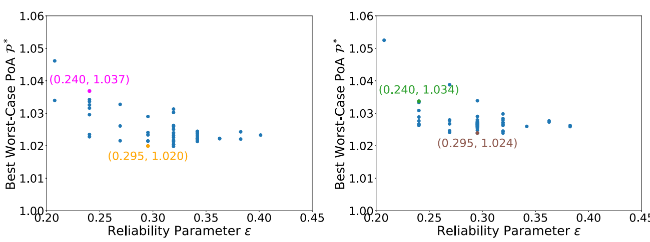
A. Upper Bounds on Tolls

In analyzing our results, we first consider the case where upper bounds on tolls are imposed, corresponding to a central planner requiring $0 \leq \tau \leq 2$. Fig 3 reports the performance and robustness for two different choices of

[3]Here $\mathbb{1}_e$ denotes the e-th canonical vector.

[4]We use the implementation of the traffic assignment problem from [28].

$\alpha = 5\%$, $\alpha = 20\%$ and 50 initial conditions. When the numerical gradient descent algorithm terminates, each initial set of tolls corresponds to a performance-robustness pair depicted with a point in Fig 3. The performance is measured by the worse-case price of anarchy encountered over the available scenarios while the robustness is quantified by the probability of violation against unseen scenarios $\varepsilon(s^*, \beta)$ entering (9), which we compute setting $\beta = 10^{-6}$. For example, the toll vector corresponding to the orange point in Fig 3a achieves an empirical price of anarchy of 1.02, and we can claim, with confidence $1 - \beta = 1 - 10^{-6}$, that this or a better price of anarchy value will be encountered on a large fraction of the demands, namely on $70.5\% = 100\% - 29.5\%$ of them. We believe these results to be particularly interesting, as they show that it is possible to design tolls with relatively small magnitude, good performance, and desirable robustness guarantees. For comparison, Fig 4 presents the values of the toll vectors corresponding to the purple and orange points from Fig 3a, which we refer to as TOLL1 and TOLL2, as well as that of the marginal cost mechanism. Although the marginal cost mechanism is known to induce a price of anarchy equal to one, it requires levying tolls that are between one and two orders of magnitude larger than those we design here, in addition to the fact that they are flow-dependent.



(a) Variation 5%　　(b) Variation 20%

Fig. 3: Performance and robustness. Fig 3a and 3b are obtained with 5% and 20% variation, respectively. A vector of tolls with good robustness (resp. performance) is marked in magenta (resp. orange) in Fig 3a. Similarly for the green and brown points in Fig 3b.

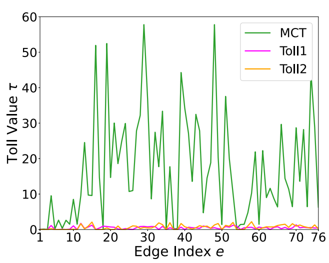
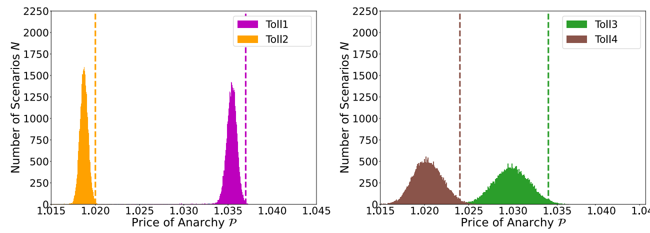


Fig. 4: Comparison between the toll magnitude for marginal cost tolls at equilibrium, and TOLL1 and TOLL2 corresponding to the purple and orange points in Fig 3.

As the guarantees obtained through (9) might be conservative, we conclude by empirically testing the robustness of the proposed toll design. We do so both for TOLL1, and TOLL2 previously defined, as well as for TOLL3, TOLL4 corresponding to the green and brown points in Fig 3b. Towards this goal, we generate 36,500 new demand realizations, taken from the same distribution, and compute the price of anarchy encountered for each realization, as defined in (6).



(a) Variation 5%　　(b) Variation 20%

Fig. 5: Performance of TOLL1-TOLL4 against $36,500$ new scenarios. The dashed line represent the values p^* obtained by approximately solving (8), for a given initial condition, when using only past demand data (100 scenarios).

An histogram of the resulting price of anarchy is presented in Fig 5. From these results we conclude that the robustness of the proposed tolls is significantly higher compared to that predicted by (9). For example, for TOLL1, there are 143 (0.39%) instances with a price of anarchy larger than the corresponding $p^* = 1.037$ as opposed to an upper bound of 24% obtained with (9). Similar results hold for TOLL2-TOLL4, where the more significant spread visible in Fig 5b is due to the significantly larger demand variation, here at 20%.

B. Partially-Tollable Networks

Finally, we consider the case of partially-tollable networks. Specifically, we assume that only half of roads in the Sioux Falls network can be charged, which we fix randomly, while still enforcing $0 \leq \tau \leq 2$. We run the numerical gradient descent with 50 different initial conditions, and select the toll vector with the best value of $p^* = 1.032$. We then test its performance against the same 36,500 scenarios generated before, and plot the results in Fig 6. Interestingly, this result show that it is possible to include significant limitations on the number of tollable edges, while still achieving satisfactory performance.

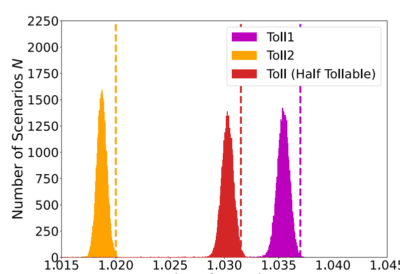


Fig. 6: Performance comparison with 5% variation. Here we compare the previously defined TOLL1 and TOLL2, which allow tolling the entire network, with a toll vector that is allowed to toll only half of the roads.

V. CONCLUSIONS

While marginal cost tolls do not allow to impose upper bounds on their magnitude or to limit the set of tollable roads, tolls designed solving the restricted network tolling problem lack robustness to demand variation. How to reconcile the two? In this work we introduced the *probabilistically robust restricted network tolling problem*, where we seek tolls that result in low social cost over a set of past demand realizations. We formulate the problem as a bilevel program and propose a gradient-based algorithm for computing an approximate solution, based on the observation that very efficient traffic assignment subroutines can be employed for this. We then exploit recent developments in scenario optimization to equip our approximate solution with formal generalization guarantees. Our experiments show that it is possible to design tolling mechanisms that are flow-independent and of limited magnitude, while ensuring remarkable performance and robustness.

REFERENCES

[1] M. Beckmann, C. B. McGuire, and C. B. Winsten. *Studies in the Economics of Transportation*. Rand Corporation, 1956.

[2] P. N. Brown and J. R. Marden. Optimal mechanisms for robust coordination in congestion games. *IEEE Transactions on Automatic Control*, 63(8):2437–2448, 2017.

[3] G. C. Calafiore and M. C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.

[4] M. C. Campi and S. Garatti. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3):1211–1230, 2008.

[5] M. C. Campi, S. Garatti, and F. A. Ramponi. A general scenario theory for nonconvex optimization and decision making. *IEEE Transactions on Automatic Control*, 63(12):4067–4078, 2018.

[6] R. Cole, Y. Dodis, and T. Roughgarden. Pricing network edges for heterogeneous selfish users. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 521–530, 2003.

[7] R. Cominetti, M. Scarsini, M. Schröder, and N. Stier-Moses. Approximation and convergence of large atomic congestion games. *arXiv preprint arXiv:2001.02797*, 2020.

[8] S. C. Dafermos and F. T. Sparrow. The traffic assignment problem for a general network. *Journal of Research of the National Bureau of Standards B*, 73(2):91–118, 1969.

[9] A. De Palma, M. Kilani, and R. Lindsey. Congestion pricing on a road network: A study using the dynamic equilibrium simulator metropolis. *Transportation Research Part A: Policy and Practice*, 39(7-9):588–611, 2005.

[10] F. Facchinei and J.-S. Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2003.

[11] P. Ferrari. Road network toll pricing and social welfare. *Transportation Research Part B: Methodological*, 36(5):471–483, 2002.

[12] M. Fukushima. A modified frank-wolfe algorithm for solving the traffic assignment problem. *Transportation Research Part B: Methodological*, 18(2):169–177, 1984.

[13] R. Geisberger, P. Sanders, D. Schultes, and D. Delling. Contraction hierarchies: Faster and simpler hierarchical routing in road networks. In *International Workshop on Experimental and Efficient Algorithms*, pages 319–333. Springer, 2008.

[14] A. Haurie and P. Marcotte. On the relationship between nash-cournot and wardrop equilibria. *Networks*, 15(3):295–308, 1985.

[15] M. Hoefer, L. Olbrich, and A. Skopalik. Taxing subnetworks. In *International workshop on internet and network economics*, pages 286–294. Springer, 2008.

[16] E. Koutsoupias and C. Papadimitriou. Worst-case equilibria. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 404–413. Springer, 1999.

[17] M. Labbé, P. Marcotte, and G. Savard. A bilevel model of taxation and its application to optimal highway pricing. *Management science*, 44(12-part-1):1608–1622, 1998.

[18] S. Lawphongpanich and D. W. Hearn. An MPEC approach to second-best toll pricing. *Mathematical programming*, 101(1):33–55, 2004.

[19] L. N. Liu and J. F. McDonald. Economic efficiency of second-best congestion pricing schemes in urban highway systems. *Transportation Research Part B: Methodological*, 33(3):157–188, 1999.

[20] J. F. Nash Jr. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49, 1950.

[21] U. B. of Public Roads. Office of Planning. Urban Planning Division. *Traffic Assignment Manual for Application with a Large, High Speed Computer*. US Department of Commerce, 1964.

[22] C. Osorio and B. Atasoy. Efficient simulation-based toll optimization for large-scale networks. *Transportation Science*, 55(5):1010–1024, 2021.

[23] D. Paccagnan and M. Gairing. In congestion games, taxes achieve optimal approximation. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 743–744, 2021.

[24] D. Paccagnan, B. Gentile, F. Parise, M. Kamgarpour, and J. Lygeros. Nash and wardrop equilibria in aggregative games with coupling constraints. *IEEE Transactions on Automatic Control*, 64(4):1373–1388, 2018.

[25] A. C. Pigou. *The Economics of Welfare*. London: Macmillan, 1920.

[26] D. Ralph and S. Dempe. Directional derivatives of the solution of a parametric nonlinear program. *Mathematical programming*, 70(1):159–172, 1995.

[27] T. Reed. Inrix global traffic scorecard. *INRIX research*, 2019.

[28] Routing Kit. https://github.com/vbuchhold/routing-framework, Accessed on Feb 4 2022.

[29] M. D. Simoni, K. M. Kockelman, K. M. Gurumurthy, and J. Bischoff. Congestion pricing in a world of self-driving vehicles: An analysis of different strategies in alternative future scenarios. *Transportation Research Part C: Emerging Technologies*, 98:167–185, 2019.

[30] R. K. Sundaram et al. *A first course in optimization theory*. Cambridge university press, 1996.

[31] Transportation Networks for Research Core Team. Transportation networks for research, https://github.com/bstabler/transportationnetworks, Accessed on Feb 4 2022.

[32] 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*, 36(8):707–729, 2002.

[33] H. Walton, D. Dajnak, S. Beevers, M. Williams, P. Watkiss, and A. Hunt. Understanding the health impacts of air pollution in london. *Transport for London and the Greater London Authority*, 1(1):6–14, 2015.

[34] J. G. Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the institution of civil engineers*, 1(3):325–362, 1952.