Tolling for MDP Congestion Games

authors

***Abstract*—We present a dual perspective on MDP routing games, from a social planner's perspective and individual game player's perspective. In addition, we show that state constraints on MDP problem can be imposed in an equivalent unconstrained version of the problem with auxiliary rewards. Finally, we apply the formulation on a ride share scenario where the social planner needs to satisfy certain driver density constraints by adjusting local rewards.**

I. Introduction

[1] Although a single population congestion game has been extensively analyzed(CITATION), we adopt the perspective of a global planner to enforce state space constraints without altering individual agent's behaviour.

Other's work: What have others done in this?

In order to enforce global constraints, we treat the routing game as a finite horizon, potential mean-field game that can be solved with a Markov Decision Process(MDP). Using the game reward as planner's control, it can be shown that state constraints can be reformulated as rewards, which in turn motivate game players to achieve a Wardrop Equilibrium that satisfies the constraints.

In particular, we reformulate state space constraints as a toll or flow based congestion control. Interestingly, it can be shown that a flow based control will not satisfy any state-space constraint by itself. The tolls are then applied to a ride-share scenario where drivers are attempting to maximize individual monetary return.

II. Notation

Behcet vs Yoyo vs Einsum

III. Problem Formulation

Brief explanation of relevant concepts from each of the following fields

A. Markov decision processes

Consider the following linear programming formulation of finite horizon MDP [2]

$$\begin{aligned}\min_{\substack{y_t\\ t\in\mathcal{T}}}.\ & \sum_{t\in\mathcal{T}}\sum_{s\in\mathcal{S}}\sum_{a\in\mathcal{A}} y_t(s,a)c_t(s,a)\\ \text{s.t.}\ & \sum_{a\in\mathcal{A}} y_{t+1}(s',a) = \sum_{s\in\mathcal{S}}\sum_{a\in\mathcal{A}}\Gamma(s'\mid s,a)y_t(s,a),\\ & \sum_{a\in\mathcal{A}} y_0(s,a) = p_0(s),\\ & y_t(s,a)\geq 0,\quad \forall s,s'\in\mathcal{S}, a\in\mathcal{A}, t\in\mathcal{T}\end{aligned} \tag{1}$$

where $\mathcal{S}$ and $\mathcal{A}$ denote respectively the set of states and actions, $\mathcal{T}=\{0,\dots,|\mathcal{T}|\}$ denote the time step, $y_t(s,a)$ and $c_t(s,a)$ denote respectively the probability and cost of state-action pair (s,a), $\Gamma(s'|s,a)$ denotes transition probability from state s to state s' given action a, $p_0(s)$ denotes the probability that the decision processes starts from state s.

B. Potential Games

Definition 1. *$\exists$ a function $F(z)\in\mathbf{C}^1$ such that*

$$\frac{\partial F}{\partial z_t[s,a]} = R_t[s,a](z_t) \tag{2}$$

In the standard case, $R_t[s,a](z_t) = R_t[s,a]\Big(z_t[s,a]\Big)$ and we can write a potential function similar to the Rosenthal potential from non-atomic routing games

$$F(z) = \sum_t\sum_s\sum_a\int_0^{z_t[s,a]} R_t[s,a](u)\ du \tag{3}$$

1) Potential Optimization Problem: Initial population distribution: $m_0\in\mathbb{R}^n$

$$\begin{aligned}\max_{\{z_t\}_{t=0}^N}\quad & F(z)\\ \text{s.t.}\quad & z_0 = m_0,\quad z_t\geq 0\qquad t=0,\dots,N\\ & \sum_k z_{t+1}[s',a] = \sum_i\sum_k G_{t,k}^T[s',s]z_t[s,a]\end{aligned} \tag{4}$$

C. Wardrop Equilibrium

Theorem 1. *A minimizer of* (4) *is a* Wardrop Equilibrium.

D. Exact Penalty

Theorem 2 ([3]). *Consider constrained optimization problem*

$$\begin{array}{ll} \textit{minimize}_{x\in\mathcal{X}} & f(x) \\ \textit{subject to} & g(x) \le 0, \end{array} \tag{5}$$

and its penalized form

$$\textit{minimize}_{x\in\mathcal{X}} \quad f(x) + \tau^\top [g(x)]_+ \tag{6}$$

where $f : \mathbf{R}^n \to \mathbf{R}$ *and* $g(x) = \begin{bmatrix} g_1(x) & \dots g_m(x) \end{bmatrix}^\top$, $g_i(x) : \mathbf{R}^m \to \mathbf{R}$ *are convex functions,* $\mathcal{X} \subseteq \mathbf{R}^n$ *is a convex set,* $[y]_+ = y$ *if* $y \in \mathbf{R}^m_+$ *and zero otherwise. Further, assume* $(x^\star, \tau^\star)$ *be a optimal primal-dual pair of problem* (5) *(satisfies the KKT conditions). Then*

- *problem* (5) *and* (6) *have the same optimal values if and only if* $\tau \ge \tau^\star$.
- *problem* (5) *and* (6) *have the same optimal solutions if* $\tau > \tau^\star$.

IV. Constrained MDP Congestion Games

A. Linear Toll

B. Quadratic Toll

C. Individual Game Player Perspective

Theorem 3. *Assume there exists a feasible, non-zero toll corresponding to constraint* $y_t(s,a) \le d_t(s,a)$ *for Eqn. 1. then any additional agents added to network will avoid the constrained state when toll is active and will not cause constraint violation.*

Proof: Consider an optimal solution to Eqn. 1, $y^*_t(s,a)$, and a constrained optimal solution, $x^*_t(s,a)$. From KKT conditions, toll is non-zero where $y^*_t(s,a) > d_t(s,a)$. Suppose for t_c, s_c, a_c, y^* violates the constraint. Then from exact penalty theorem, $x^*_{t_c}(s_c,a_c) = d_{t_c}(s_c,a_c)$.

At t_c, s_c, x^*, the value function of the tolled MDP problem is

$$V_{t_c}(s_c) = \max_a f(a) = c_{t_c}(s_c,a) + \sum_{s'} \Gamma(s'|s_c,a) V_{t_c+1}(s')$$

The policy is

$$\pi_{t_c}(s_c) = \operatorname*{argmax}_a c_{t_c}(s_c,a) + \sum_{s'} \Gamma(s'|s_c,a) V_{t_c+1}(s')$$

If $y^*_{t_c}(s_c,a_c) > d_{t_c}(s_c,a_c)$ and $x^*_{t_c}(s_c,a_c) = d_{t_c}(s_c,a_c)$, then there exists alternative a' such that $f(a') > f(a_c)$. Therefore any additional agent will not choose a_c, instead will choose a', and therefore avoid the constrained state action pair. ∎

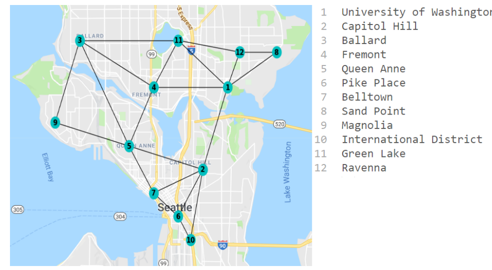


Fig. 1: (a) Illustration of Seattle Neighborhoods .

V. Numerical Example

SUMMARY: This is a thorough example where we are trying to capture as much of the ride sharing dynamics as possible.

A. Simulation: Ride-sharing game

To demonstrate the usage of exact penalty as a toll generation method, we consider a game scenario that ride-sharing drivers might play in Seattle, seeking to optimize their fares. The city is abtracted an undirected, connect graph, where neighbourhoods are nodes, and drivers may traverse adjacent nodes as specified by edges (shown in Figure 1).

At each node, the driver can choose from several actions. The first action, a_r, is to wait for a random rider and transition to whatever node that rider wants to go to. We assume that the driver will always pick up a rider, although it may take long time if there already exists many drivers doing the same. In addition, the passenger's destination will appear as random to the driver. At each node, a constant percentage of riders want to travel to each of the other nodes. This model would be useful for ride sharing services where drivers are simply assigned riders or taxi drivers who queue at transportation hubs.

The driver can also choose to transition without a rider. The action of transitioning to node s' without a rider is $a_{s'}$. In general, this would result in the driver paying the travel costs without receiving a fare (however, there could also be a small possibility that the driver will find a customer along the way).

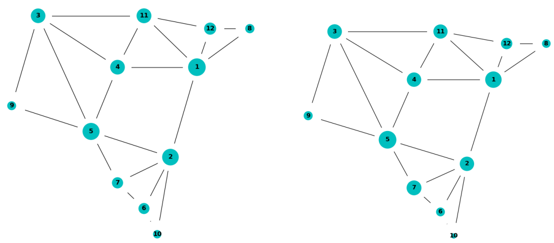
The rewards for taking each action (and then transitioning from state s to state s') is given by

$$R_t[s', s, a_r]\Big(z_t\Big) = M_t[s', s] \tag{7}$$

$$- C^{\text{trav}}_t[s', s] - C^{\text{wait}}_t \cdot z_t[s, a_r] \tag{8}$$

$$R_t[s', s, a_{s'}] = -C^{\text{trav}}_t[s', s] \tag{9}$$

where $M_t[s', s]$ is the monetary cost for transitioning from state s to s', $C^{\text{trav}}_t[s', s]$ is the travel cost from state



(a) (b)

Fig. 2: Steady state distribution of drivers at each node under the (a) equilibrium strategies and (b) socially optimal strategies showing the portion of drivers that take riders and the portion that do not take riders. Drivers transitioning between nodes without riders in the (c) equilibrium case and (d) under the socially optimal strategies.

s to s', and C_t^{wait} is the coefficient of the cost of waiting for a rider. We compute these various parameters as

$$M_t[s', s] = \big(\text{Rate}\big) \cdot \big(\text{Dist}\big) \tag{10}$$

$$C_t^{\text{trav}}[s', s] = \tau \underbrace{\big(\text{Dist}\big)}_{\text{mi}} \underbrace{\big(\text{Vel}\big)^{-1}}_{\text{hr/mi}} + \underbrace{\begin{pmatrix}\text{Fuel}\\\text{Price}\end{pmatrix}}_{\$/\text{gal}} \underbrace{\begin{pmatrix}\text{Fuel}\\\text{Eff}\end{pmatrix}^{-1}}_{\text{gal/mi}} \underbrace{\big(\text{Dist}\big)}_{\text{mi}} \tag{11}$$

$$C_t^{\text{wait}} = \tau \cdot \Bigg(\underbrace{\begin{matrix}\text{Customer}\\\text{Demand Rate}\end{matrix}}_{\text{rides/hr}}\Bigg)^{-1} \tag{12}$$

and τ is a time-money tradeoff parameter which we take to be \$27/hr.

The values independent of specific transitions are listed in the table below.

Rate	Velocity	Fuel Price	Fuel Eff
\$6 /mi	8 mph	\$2.5/gal	20 mi/gal

We compute both the equilibrium strategies and the socially optimal strategies in the infinite horizon game. Figure 2 shows the steady state distribution of drivers at the nodes in both cases including the portion that take riders and the portion that do not as well as the transitions that drivers make without riders.

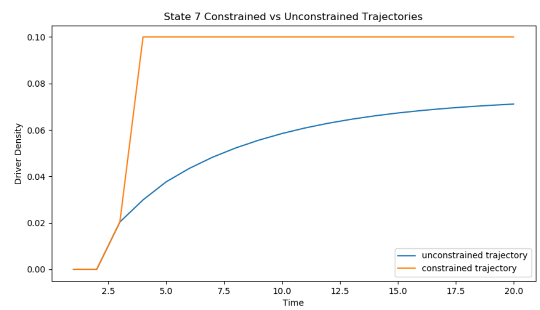
VI. Conclusions

References

[1] E. Altman, *Constrained Markov decision processes*. CRC Press, 1999, vol. 7.

[2] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[3] D. P. Bertsekas, *Nonlinear programming*. Athena scientific Belmont, 1999.



(a)

Fig. 3: Time profile of state 7 in unconstrained optimal solution and unconstrained optimal solution with exact penalty