2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Inducing Desired Equilibrium in Taxi Repositioning Problem with Adaptive Incentive Design

Jianhui Li, Youcheng Niu, Shuang Li, Yuzhe Li, Jinming Xu, and Junfeng Wu

Abstract— **We study the problem of designing incentives to induce desired equilibrium in taxi repositioning problems. In this scenario, self-interested idle drivers will update their repositioning strategies with observed payoff. Meanwhile, the platform will adaptively design incentives to induce a better Nash equilibrium for global efficiency. We formulate the problem as a bi-level optimization problem where the incentive designer and idle drivers simultaneously update their decision variables. We prove that drivers' strategies will reach Nash equilibrium, and the incentive designer's objective function will reach optimality under Polyak Lojasiewicz (PL) condition. Furthermore, we derive a sufficient condition for the PL condition to hold for the upper-level objective function and lower-level agents' payoff function. Finally, we demonstrate the efficiency of the proposed method by numerical results.**

I. Introduction

In recent years, e-hailing taxis have become increasingly popular in cities, providing a convenient way for people to travel. However, as more and more drivers join the service, global efficiency becomes a significant concern. Since drivers are self-interested agents who optimize their payoff function, they may not act in the best interest in a global perspective [1]. For example, when there is high demand for taxis in certain areas, drivers may flock to these areas even if there are still unserved orders elsewhere. This motivates using incentives to induce desired behavior and balance distributions between drivers and demands.

The incentive design problem has been extensively studied in [2]–[7]. The effect of incentives is typically only observable after the agents' strategies reach an equilibrium. Reference [8] uses a double-loop algorithm that updates the incentive policy after the convergence of drivers' strategies. However, since an equilibrium in taxi repositioning setting does not admit a closed-form solution and is time-consuming to compute for iterative methods, the papers [9]–[11] update incentive policies simultaneously with drivers' strategies. They fall into the bi-level optimization setting, where the upper and lower agents update their decision variables in a single loop.

J. Li, Y. Niu, and J. Xu are with the College of Control Science and Engineering, Zhejiang University, Hangzhou, China; Email: {jianhuili, ycniu, jimmyxu}@zju.edu.cn. Y. Li is with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang, China; Email: yuzheli@mail.neu.edu.cn. S. Li and J. Wu are with the School of Data Science, The Chinese University of Hong Kong (Shenzhen), Shenzhen, China; Email: {lishuang, junfengwu}@cuhk.edu.cn. This work was supported in part by the National Natural Science Foundation of China under grant no. 62003303, in part by Shenzhen Science and Technology Program JCYJ20210324120011032.

Since the update occurs simultaneously, the impact of incentives on agents' strategies is hard to analyse. The algorithm in [9] applies Bayesian optimization, a black-box optimization method, to the taxi repositioning problem. While in [10], the authors use a heuristic method to approximate the incentive effect. However, both algorithms lack convergence proof. The paper [11] examines the setting in that the incentive designer uses sensitivity theory to approximately capture the impact of incentives and provides the convergence result, while it assumes that the objective function is strongly convex, which is less realistic in practice.

In the taxi-repositioning problem, we divide the city map into several regions and group drivers at the same region as a collective agent, instead of modeling each driver as a single agent in [9]. This formulation enables us to deal with large-scale problems more effectively. Our paper has two main contributions. First, we introduce a new formulation for the driver repositioning problem. Second, we prove that the upper-level objective function converges to an optimal value under the Polyak Lojasiewicz (PL) condition, and we propose a sufficient condition for PL condition to hold.

The proposed method differs from [11] in that our problem does not conform to its assumption on the upper-level objective function being strongly convex. In the taxi repositioning setting, the objective function $f_*(\theta) := f(x_*(\theta))$ is a composition function of $f(\cdot)$ and $x_*(\cdot)$, where $f(\cdot)$ is a global criterion on drivers' repositioning strategies, and $x_*(\theta)$ is the Nash equilibrium given an incentive θ. Since the equilibrium mapping $x_*(\cdot)$ is complicated, the function $f_*(\cdot)$ is possibly non-convex.

The article is organized as follows. In section II, we describe the basic setup of the taxi repositioning problem, and also introduce the update rule of both the drivers' strategies and the incentive policy. In section III, we derive the convergence result. We then apply the algorithm in a taxi repositioning simulation in Section IV and demonstrate its effectiveness. Finally, we conclude our work in Section V.

II. Basic Setup of Taxi Repositioning Problem

We divide a city into several regions and denote each region as a node. Then, we construct edges based on distances between regions. The graph contains self-loops since drivers can stay in the current region. Consider a graph with $N = \{1, 2, \cdots, n\}$ nodes, where the ith node has v^i number of idle drivers and d^i number of demands. We examine a single period problem within a specific time interval in a day, e.g., the peak hour, and formulate the problem as a non-cooperative game [9]. An extension to a sequential

decision game is left for future work. Drivers will choose the probability of driving to a neighbor node and update their strategies with the observed payoff. We assume drivers can only reposition to one-hop neighbor nodes. Furthermore, since the distribution of drivers after repositioning may not conform to the distribution of demands, an incentive $\theta \in \mathbb{R}^n$ is introduced to induce the desired distribution of idle drivers. Drivers repositioned to node i will receive a bonus θ^i for compensation or penalty. We formulate the problem as a bi-level optimization problem [11]:

$$\begin{aligned} \min_{\theta} \quad & f(x_*(\theta)) := f_*(\theta) \\ \text{s.t.} \quad & x_*^i(\theta) \in \arg\min_{x^i \in \mathscr{A}^i} \left\{ u^i(x^i, x_*^{-i}(\theta); \theta) \right\} \quad \forall i \in N \end{aligned} \tag{1}$$

where $f_*(\cdot)$ is the upper-level objective function, and $u^i(\cdot)$ is agent i's payoff function. We denote drivers' actions as strategies and incentive designer's actions as policy in this paper. Besides, we use the terms drivers, nodes, and agents interchangeably in the rest of the paper.

A. Drivers' Game at Lower Level

We assume that drivers are homogeneous and they share an identical strategy. Thus we group drivers at each region as a single agent and define the strategy as the distribution of idle drivers. From a macro perspective, agents' strategies are the distribution of drivers. From a micro perspective, drivers' strategies are the probability for him/her to drive to a specific neighboring node in the next time episode.

We denote the set of ith node's neighboring nodes as N^i, and thus the strategy set $\mathscr{A}^i = \Delta(N^i)$ is a simplex. Drivers will update their strategies $\{x^i\}_{i \in \mathscr{N}}$ while observing the payoff experienced at each episode. The ith agent's strategy is denoted as $x^i := [x^{ij}]_{j \in \mathscr{N}}$, where x^{ij} denotes the distribution of drivers reposition from node i to node j. We denote the payoff function of agent i as $u^i(x^i, x^{-i}; \theta)$ where $x^i \in \mathscr{A}^i$ is the reposition distribution of drivers at node i, x^{-i} is the strategy of other nodes, and $\theta \in \mathbb{R}^n$ is the incentives applied to nodes. In specific, we assume the payoff function has the following quadratic form:

$$u^i(x^i, x^{-i}; \theta) = (x^i)^\top Q^i x^i + \sum_{j \neq i} (x^j)^\top R^{ji} x^i + (b^i)^\top x^i - \theta^\top x^i \tag{2}$$

where Q^i is a positive semi-definite matrix and R^{ij} is certain payoff related matrix. With fixed incentive θ, we can obtain the gradient $\nabla_{x^i} u^i(x^i, x^{-i}; \theta)$ of payoff function u^i with respect to the decision variable x^i. We denote the stacked gradient as $v_{\theta_k}(x_k) := [\nabla_{x^i} u^i(x_k^i, x_k^{-i}; \theta_k)]_{i \in N}$. In practice, the gradient is hard to obtain, and thus we use an estimate $\hat{v}_k^i$ of the gradient from samples. we adopt a mirror descent update rule to update the drivers' strategies as

$$x_{k+1}^i = \arg\min_{x^i \in \mathscr{A}^i} \left\{ \langle \hat{v}_k^i, x^i \rangle + \frac{1}{\beta_{k,i}} D_\psi\left(x^i, x_k^i\right) \right\}$$

where $\beta_{k,i}$ is the step size at iteration k for agent i, $D_\psi(x, x')$ is a Bregman divergence induced by a strictly convex function $\psi(\cdot)$, which can be interpreted as a distance between strategy x and x'. Since the strategy space is a simplex $\Delta(N^i)$, we take the Kullback-Leibler (KL) divergence as the Bregman divergence, i.e.,

$$\begin{aligned} D_\psi(x^i, x^{i\prime}) &:= \psi(x^i) - \psi(x^{i\prime}) - \langle \nabla\psi(x^{i\prime}), x^i - x^{i\prime} \rangle \\ &= (x^i)^T \log(x^i / x^{i\prime}) \end{aligned} \tag{3}$$

with $\psi(x) := x^\top \log(x)$. Besides, to avoid reaching a boundary value, we add a mixing step, which adds $\gamma_k / |N^i|$ to strategy x_k^i at each iteration:

$$\tilde{x}_{k+1}^i = (1 - \gamma_k) x_{k+1}^i + \gamma_k / |N^i| \mathbf{1}_{|N^i|},$$

where $\gamma_k > 0$ is the mixing step size, N^i is the set of neighbour nodes of ith node.

B. Incentive Designer at Upper Level

The incentive designer uses incentives to affect agents' game strategies and indirectly optimize the criterion. We denote the optimization problem at the upper level as:

$$\min_{\theta} f(x_*(\theta)) := f_*(\theta)$$

As in most optimization methods, we require the gradient $\nabla_\theta f_*(\theta)$ to update the incentive policy. By definition of $f_*(\cdot)$, we have $\nabla_\theta f_*(\theta) = \nabla_\theta x_*(\theta) \nabla_x f(x)|_{x = x_*(\theta)}$. To account for the effect of incentives on agents' Nash equilibrium, we apply Lemma 4 in Section III-B, which derives a closed-form solution of $\nabla_\theta x_*(\theta)$ as a function of the strategies $x_*(\theta)$ and incentive θ.

In an ideal case, we require the convergence of drivers' strategies x to a Nash equilibrium $x_*(\theta)$ to evaluate the effect of incentive policy and calculate the gradient $\nabla_\theta f_*(\theta)$. Since the calculation of a Nash equilibrium strategy is prohibitively expensive, we apply a single loop algorithm as in [11], which uses current strategies rather than an exact Nash equilibrium x_* to calculate an approximate for $\nabla_\theta x_*(\theta)$ and $\nabla_x f(x)|_{x = x_*(\theta)}$ and thus obtain an approximate gradient of the incentive designer's objective function $\tilde{\nabla} f(\theta, x)$. Note that when drivers' strategies are the Nash equilibrium, we have $\tilde{\nabla} f(\theta, x_*(\theta)) = \nabla f_*(\theta)$.

Similar to the lower-level game, we also assume that the algorithm can only obtain an estimate $\hat{\nabla} f_k$ from samples of the approximate gradient $\tilde{\nabla} f(\theta_k, \tilde{x}_{k+1}))$. The update rule of incentive policy is as follows:

$$\theta_{k+1} = \theta_k - \alpha_k \hat{\nabla} f_k$$

where α_k is the step size. Putting these together, we obtain a two-timescale algorithm for lower-level agents and the upper-level incentive designer as is illustrated in Algorithm 1.

III. Main Results

We make following assumptions for the lower-level agents.

Assumption 1: The lower-level game between the drivers satisfies the following:

(1). The gradient is bounded, i.e., there exists $V_u > 0$ such that for all $i \in N$,

$$\|\nabla_{x^i} u^i(x^i, x^{-i}; \theta)\|_\infty \leq V_u$$

Algorithm 1 Incentive policy update

Input: Step size α_k (upper-level), $\{\beta_{k,i}\}_{i\in N}$ (lower level), mixing step size γ_k
Output: Incentive policy θ
Initialization: Set $\{x_0^i\}_{i\in N}$ to average distribution and θ to $0_{|N|}$
for episode $k = 0, \cdots, T-1$ **do**
  Each agent choose actions $\{x_k^i\}_{i\in N}$ and observe payoff
  Lower Level Agents Update
  for $i \in N$ **do**
    Update parameters:

$$x_{k+1}^i = \arg\min_{x^i\in\mathscr{A}^i}\left\{\langle \hat{v}_k^i, x^i\rangle + \frac{1}{\beta_{k,i}} D_\psi\left(x^i, x_k^i\right)\right\}$$

    Mixing strategies:

$$\tilde{x}_{k+1}^i = (1-\gamma_k)x_{k+1}^i + \gamma_k/|N^i|\mathbf{1}_{|N^i|}$$

  end for
  Incentive Designer Update
  Update incentive policy: $\theta_{k+1} = \theta_k - \alpha_k\hat{\nabla} f_k$
end for

(2). We define function $v_\theta(x) := [\nabla_{x^i} u^i(x^i, \mathbf{x}^{-i}; \theta)]_{i\in\mathscr{N}}$, then $v_\theta(x)$ is strongly monotone with respect to x with fixed θ, i.e., there exists $\mu_v > 0$ such that

$$\langle v_\theta(x) - v_\theta(x'), x - x'\rangle \geq \mu_v \|x - x'\|^2$$

(3). For each $i \in N$, the gradient $\nabla_{x^i} u^i(x^i, x^{-i}; \theta)$ is Lipschitz continuous with respect to $\overline{D}_\psi$, i.e., there exists $H_u > 0$ such that for $i \in N$

$$\|\nabla_{x^i} u^i(x^i, x^{-i}; \theta) - \nabla_{x^i} u^i(x^{i\prime}, x^{-i\prime}; \theta)\|_2^2 \leq H_u^2 \overline{D}_\psi(x, x')$$

(4). There exist constants $\rho_\theta, \rho_x > 0$ such that

$$\|\nabla_\theta v_\theta(x)\|_2 \leq \rho_\theta, \quad \|\left[\nabla_x v_\theta(x)\right]^{-1}\|_2 \leq 1/\rho_x$$

The strongly monotonicity of $v_\theta(\cdot)$ ensures existence and uniqueness of the Nash equilibrium [16]. Then we make the following assumption for the upper-level objective function.

Assumption 2: The upper-level objective function satisfies the following:

(1). The objective function $f_*(\theta)$ satisfies the PL condition with $\mu_f > 0$, i.e., for all θ:

$$\|\nabla f_*(\theta)\|_2 \geq \mu_f\left[f_*(\theta) - \min\{f_*(\theta)\}\right]$$

(2). The objective function $f_*(\theta)$ satisfies the L_f smoothness, i.e., for all θ and θ'

$$\|\nabla f_*(\theta) - \nabla f_*(\theta')\|_2 \leq L_f\|\theta - \theta'\|_2$$

(3). There exists $M > 0$ such that for all θ it holds that $\|\nabla f_*(\theta)\|_2 \leq M$

(4). The approximated gradient $\tilde{\nabla} f(\theta, x)$ is Lipschitz continuous with respect to $\overline{D}_\psi$, i.e., there exists $\tilde{H} > 0$ such that for all $x, x' \in \mathscr{A} := \{[x^i]_{i\in N} | x^i \in \mathscr{A}^i\}$ and all θ,

$$\|\tilde{\nabla} f(\theta, x) - \tilde{\nabla} f(\theta, x')\|_2^2 \leq \tilde{H}^2 \overline{D}_\psi(x, x')$$

The PL condition ensures that we can obtain global convergence result in non-convex optimization setting [17]. We also give an example in section IV that objective function in taxi-reposition problem could be designed to satisfy this assumption.

Since the gradient may be inaccurate, we take account of potential error in estimation. We make the following assumption on the gradient estimate $\hat{v}_k^i$ and $\hat{\nabla} f_k$, assuming that the estimates are unbiased and have bounded mean squared errors.

Assumption 3: Define the filtration by $\mathscr{F}_0^\theta = \{\theta_0\}, \mathscr{F}_0^x = \emptyset$ and

$$\mathscr{F}_k^\theta = \mathscr{F}_{k-1}^\theta \bigcup \{x_{k-1}, \theta_k\}, \mathscr{F}_k^x = \mathscr{F}_{k-1}^x \bigcup \{\theta_k, x_k\}$$

then the filtration satisfy:

(1). The gradient estimate $\hat{v}_k^i$ and $\hat{\nabla} f_k$ are unbiased estimates, i.e.,

$$\mathbb{E}\left[\hat{\nabla} f_k | \mathscr{F}_k^\theta\right] = \tilde{\nabla} f(\theta_k, \tilde{x}_{k+1})$$

$$\mathbb{E}\left[\hat{v}_k^i | \mathscr{F}_k^x\right] = \nabla_{x^i} u^i(\tilde{x}_k^i, \tilde{x}_k^{-i}; \theta_k) \quad \forall i \in N$$

(2). The estimates have bounded mean squared estimation error, i.e., there exist $\delta_f, \delta_u > 0$ such that $\forall i \in N$:

$$\mathbb{E}\left[\|\hat{\nabla} f_k - \tilde{\nabla} f(\theta_k, \tilde{x}_{k+1})\|_2^2 | \mathscr{F}_k^\theta\right] \leq \delta_f^2$$
$$\mathbb{E}\left[\|\hat{v}_k^i - \nabla_{x^i} u^i(\tilde{x}_k^i, \tilde{x}_k^{-i}; \theta_k)\|_2^2 | \mathscr{F}_k^x\right] \leq \delta_u^2$$

A. Convergence Result

Under the above assumptions, we establish the convergence result of the lower-level game and the upper-level objective function. We define the optimality criterion as:

$$\begin{aligned} \varepsilon_k^x &:= \overline{D}_\psi(\tilde{x}_*(\theta_{k-1}), \tilde{x}_k) = \sum_{i\in N} D_\psi(\tilde{x}_*^i(\theta_{k-1}), \tilde{x}_k^i) \\ \varepsilon_k^\theta &:= \mathbb{E}\left[f_*(\theta_k)\right] - \min_\theta\{f_*(\theta)\} \end{aligned} \tag{4}$$

where $\tilde{x}_*(\theta_{k-1}) := (1-\gamma_k)x_*(\theta_{k-1}) + \gamma_k \mathbf{1}_{|N^i|}/|N^i|$. When ε_k^x converge to zero, all drivers' mixed strategies have converged to the mixed Nash equilibrium strategies $\tilde{x}_*(\theta_{k-1})$, which means that $x_k = x_*(\theta_{k-1})$. Note that agents' strategies are influenced directly by θ_{k-1} but not θ_k (as is shown by Algorithm 1), thus the Bregman divergence is defined between $\tilde{x}_*(\theta_{k-1})$ and $\tilde{x}_k$. Similarly, when ε_k^θ equals 0, the value of the upper-level objective function reaches an optimal value.

We first cite a lemma from [11] which state the convergence of the lower-level agents. This lemma uses variationally stable assumption, which can be derived by the strongly monotone assumption [15].

Lemma 1 (Lemma C.1 [11]): Let the step sizes $\beta_k = \beta/(k+1)^{2/7}$, $\gamma_k = \gamma/(k+1)^{4/7}$, and $\alpha_k = \alpha/(k+1)^{1/2}$, with $\alpha, \beta > 0$ satisfying

$$\beta \leq 1/(6N^2H_u^2), \quad \alpha/\beta^{3/2} \leq 1/(7\tilde{H}\tilde{H}_*)$$

where $\tilde{H}_* := (1+d)\rho_\theta/\rho_x$ and $d := \dim(ALA^\top)$, where dim is the dimension of the rectangle matrix, A and L are defined

in lemma 4. Then, for all $k \geq 0$, we have

$$\begin{aligned}\varepsilon_{k+1}^x \leq & N\left(\sum_{l=0}^{k}(\beta_l\gamma_l\log(1/\gamma_l)+2\gamma_{l+1}+2\gamma_l^2)\prod_{j=l+1}^{k}(1-\beta_j/8)\right)\\ &+\left((\delta_u^2+3V^*)N+(\delta_f^2+2M^2+6N\tilde{H}^2)/8\tilde{H}^2\right)\\ &\cdot\left(\sum_{l=0}^{k}\beta_l^2\prod_{j=l+1}^{k}(1-\beta_j/8)\right)+\varepsilon_0^x\prod_{j=0}^{k}(1-\beta_j/8).\end{aligned}$$

Furthermore $\varepsilon_{k+1}^x \leq \tilde{c}\beta_k = O(k^{-2/7})$, where $\tilde{c} = 8\varepsilon_0^x + 8(\delta_u^2 + 3V_u^2)N + (\delta_f^2+2M^2+6N\tilde{H}^2)/\tilde{H}^2 + 32N(1/\beta^2+4/7\beta)$.

Above lemma illustrates the convergence of lower level agents' strategies to the Nash equilibrium. In the following lemma, we prove the convergence of the upper-level objective function

Lemma 2: For all $k \geq 0$, we have

$$\varepsilon_{k+1}^\theta \leq (1-\frac{\mu\alpha_k}{2})\varepsilon_k^\theta + \frac{\alpha_k^{3/2}}{2}\delta_f^2 + \alpha_k\tilde{H}^2(\varepsilon_{k+1}^x + 2\gamma_k\log(1/\gamma_k))$$

Proof: We start from the L_f smoothness of $f_*(\theta)$:

$$\begin{aligned}&f_*(\theta_{k+1}) - f_*(\theta_k)\\ \leq&\langle\nabla f_*(\theta_k), \theta_{k+1}-\theta_k\rangle + \frac{L_f}{2}\|\theta_{k+1}-\theta_k\|_2^2\\ =&\langle\nabla f_*(\theta_k), \alpha_k\hat{\nabla}f_k\rangle + \frac{\alpha_k^2 L_f}{2}\|\hat{\nabla}f_k\|_2^2\\ \leq&\langle\nabla f_*(\theta_k), \alpha_k\hat{\nabla}f_k\rangle + \frac{\alpha_k^{3/2}}{2}\|\hat{\nabla}f_k\|_2^2\end{aligned} \tag{5}$$

where the last line is due to the step size satisfying $\alpha_k \leq 1/L_f^2$. Taking expectation over both sides, we have:

$$\begin{aligned}&\mathbb{E}[f_*(\theta_{k+1}) - f_*(\theta_k)|\mathscr{F}_k^\theta]\\ \leq&\alpha_k\langle\nabla f_*(\theta_k), \tilde{\nabla}f(\theta_k,\tilde{x}_{k+1})\rangle + \frac{\alpha_k^{3/2}}{2}\mathbb{E}\left[\|\hat{\nabla}f_k\|_2^2|\mathscr{F}_k^\theta\right]\\ =&\alpha_k\langle\nabla f_*(\theta_k), \tilde{\nabla}f(\theta_k,\tilde{x}_{k+1})\rangle + \frac{\alpha_k^{3/2}}{2}\|\tilde{\nabla}f(\theta_k,\tilde{x}_{k+1})\|_2^2\\ &+\mathbb{E}[\frac{\alpha_k^{3/2}}{2}\|\hat{\nabla}f_k - \tilde{\nabla}f(\theta_k,\tilde{x}_{k+1})\|_2^2|\mathscr{F}_k^\theta]\\ \leq&\alpha_k\langle\nabla f_*(\theta_k), \tilde{\nabla}f(\theta_k,\tilde{x}_{k+1})\rangle + \frac{\alpha_k}{2}\|\tilde{\nabla}f(\theta_k,\tilde{x}_{k+1})\|_2^2 + \frac{\alpha_k^{3/2}}{2}\delta_f^2\\ =&\frac{\alpha_k}{2}\|\nabla f_*(\theta_k) - \tilde{\nabla}f(\theta_k,\tilde{x}_{k+1})\|_2^2 - \frac{\alpha_k}{2}\|\nabla f_*(\theta_k)\|_2^2 + \frac{\alpha_k^{3/2}}{2}\delta_f^2\\ \leq&\frac{\alpha_k}{2}\|\nabla f_*(\theta_k) - \tilde{\nabla}f(\theta_k,\tilde{x}_{k+1})\|_2^2 + \frac{\alpha_k^{3/2}}{2}\delta_f^2\\ &-\frac{\mu\alpha_k}{2}\left[f_*(\theta_k) - \min\{f_*(\theta)\}\right]\end{aligned} \tag{6}$$

where the first equality is due to the unbiased property of $\hat{\nabla}f_k$, the second inequality is due to $\alpha_k \leq 1$ and assumption (3), the last line is due to the PL condition. Then we deal with the first term in (6). Based on assumption 1, we have

$$\begin{aligned}&\|\nabla f_*(\theta_k) - \tilde{\nabla}f(\theta_k,\tilde{x}_{k+1})\|_2^2\\ \leq&\|\nabla f_*(\theta_k) - \tilde{\nabla}f(\theta_k,\tilde{x}_*(\theta_k))\|_2^2\\ &+\|\tilde{\nabla}f(\theta_k,\tilde{x}_*(\theta_k)) - \tilde{\nabla}f(\theta_k,\tilde{x}_{k+1})\|_2^2\\ \leq&\tilde{H}^2\left[\overline{D}_\psi(x_*(\theta_k),\tilde{x}_*(\theta_k)) + \overline{D}_\psi(\tilde{x}_*(\theta_k),\tilde{x}_{k+1})\right]\end{aligned}$$

where the last line is due to the assumption (2). Further by the definition of $D_\psi(x,x') := \psi(x) - \psi(x') - \langle\nabla\psi(x'), x-x'\rangle$ and $\psi(x) := x^\top\log(x)$ we have

$$\begin{aligned}&D_\psi(x_*^i(\theta_k),\tilde{x}_*^i(\theta_k))\\ =&\psi(x_*^i(\theta_k)) - \psi(\tilde{x}_*^i(\theta_k)) - \langle\nabla\psi(\tilde{x}_*^i(\theta_k)), x_*^i(\theta_k) - \tilde{x}_*^i(\theta_k)\rangle\\ =&-D_\psi(\tilde{x}_*^i(\theta_k), x_*^i(\theta_k))\\ &+\langle\nabla\psi(\tilde{x}_*^i(\theta_k)) - \nabla\psi(\tilde{x}_*^i(\theta_k)), \tilde{x}_*^i(\theta_k) - x_*^i(\theta_k)\rangle\\ \leq&\|\log(x_*^i(\theta_k)/\tilde{x}_*^i(\theta_k))\|_\infty\|x_*^i(\theta_k) - \tilde{x}_*^i(\theta_k)\|_1\\ \leq&2\gamma_k\log(1/\gamma_k)\end{aligned}$$

where the first inequality is due to the non-negativity of D_ψ, the last inequality is due to that $\tilde{x}_*^i(\theta_k) := (1-\gamma_k)x_*^i(\theta_k) + \gamma_k\mathbf{1}_{d_i}/d_i$. Combine above term and subtract $\min\{f_*(\theta)\}$ on both sides of (6), we have:

$$\begin{aligned}&\mathbb{E}[f_*(\theta_{k+1})|\mathscr{F}_k^\theta] - \min_\theta\{f_*(\theta)\}\\ \leq&(1-\frac{\mu\alpha_k}{2})[f_*(\theta_k) - \min_\theta\{f_*(\theta)\}] + \frac{\alpha_k^{3/2}}{2}\delta_f^2\\ &+\frac{\alpha_k}{2}\tilde{H}^2\left[2N\gamma_k\log(1/\gamma_k) + \overline{D}_\psi(\tilde{x}_*(\theta_k),\tilde{x}_{k+1})\right]\end{aligned} \tag{7}$$

By definition of ε_k^θ and ε_k^x, we have

$$\varepsilon_{k+1}^\theta \leq (1-\frac{\mu\alpha_k}{2})\varepsilon_k^\theta + \frac{\alpha_k^{3/2}}{2}\delta_f^2 + \frac{\alpha_k}{2}\tilde{H}^2\left[\varepsilon_{k+1}^x + 2N\gamma_k\log(1/\gamma_k)\right]$$

After both lemmas, we give the following convergence result of both the lower-level agents and the incentive designer as follows:

Theorem 3: Let the stepsizes $\alpha_k = \alpha/(k+1)^{1/2}, \beta_k = \beta/(k+1)^{2/7}$ and $\gamma_k = 1/(k+1)^{4/7}$. Suppose Assumption 1-3 hold. Then, we have

$$\varepsilon_k^x = O(k^{-2/7}), \quad \varepsilon_k^\theta = O(k^{-1/4}).$$

Proof: For the lower-level agents, invoking lemma 1 we have

$$\varepsilon_{k+1}^x \leq \tilde{c}\beta_k = O(k^{-2/7})$$

As for the incentive designer, we take $\alpha_k = \alpha/(k+1)^{1/2}$ and assume that $\alpha \leq \min\left\{1, 1/L_f^2\right\}$. By recursively applying the result of lemma 2, we have

$$\begin{aligned}\varepsilon_{k+1}^\theta \leq&\prod_{j=0}^{k}(1-\frac{\mu\alpha_k}{2})\varepsilon_0^\theta + \frac{\delta_f^2}{2}\sum_{l=0}^{k}\left[\alpha_l^{3/2}\prod_{j=l+1}^{k}(1-\frac{\mu\alpha_j}{2})\right]\\ &+\tilde{H}^2/2\sum_{l=0}^{k}\left[\alpha_l\varepsilon_{l+1}^x\prod_{j=l+1}^{k}(1-\frac{\mu\alpha_j}{2})\right]\\ &+N\tilde{H}^2\sum_{l=0}^{k}\left[\alpha_l\gamma_l\log(1/\gamma_l)\prod_{j=l+1}^{k}(1-\frac{\mu\alpha_j}{2})\right]\end{aligned} \tag{8}$$

For the second term, we have $\varepsilon_{l+1}^x = \tilde{c}\beta_k$. For the third term, we have $\gamma_l\log(1/\gamma_l) \leq 2\sqrt{\gamma_l}$. Besides, using Lemma 10 in Appendix E from [13], we have

$$\begin{aligned}\prod_{j=0}^{k}(1-\frac{\mu\alpha_k}{2}) \leq&\sum_{l=0}^{k}\left[\alpha_l^{11/7}\prod_{j=l+1}^{k}\left(1-\frac{\mu\alpha_j}{2}\right)\right]\\ \leq&\sum_{l=0}^{k}\left[\alpha_l^{3/2}\prod_{j=l+1}^{k}\left(1-\frac{\mu\alpha_j}{2}\right)\right] \leq \frac{1}{\mu}\alpha_k^{1/2}\end{aligned} \tag{9}$$

Therefore, we have:

$$\varepsilon_{k+1}^{\theta} \leq \frac{1}{\mu}\alpha_k^{1/2}\left(\varepsilon_0^{\theta}+\delta_f^2/2+2\tilde{H}^2\right) = O(k^{-1/4}) \tag{10}$$

This completes the overall proof.

B. Sufficiency for PL condition

We first refer to a game-sensitivity lemma that describes the relationship between an external parameter and the Nash equilibrium. We use this lemma to approximately evaluate the effect of incentives and calculate $\bar{\nabla} f(\theta, x)$.

***Lemma** 4 (Theorem 1 [12]):* Let $\mathscr{A}^i = \Delta(|N^i|)$ be the strategy set, $\nabla_x v_\theta(x)$ be non-singular and $\nabla_\theta x_*(\theta)$ be the Jacobian of $x_*(\theta)$. Then, we have

$$\nabla_\theta x_*(\theta) = -J_\theta \nabla_\theta v_\theta(x_*(\theta)),$$

where $J_\theta = L - LA^\top [ALA^\top]^{-1}AL$, $L = [\nabla_x v_\theta(x_*(\theta))]^{-1}$ with $A := blckdiag\{[\mathbf{1}_{|N^i|}]_{i\in N}\}$ representing the constraint $Ax = \mathbf{1}_n$.

Using the above lemma and combined with the quadratic payoff function formula $u^i(x^i, x^{-i}; \theta)$, we obtain that the sensitivity matrix is a constant matrix. Recall that with the definition of the payoff function

$$u^i(x^i, x^{-i}; \theta) = (x^i)^\top Q^i x^i + \sum_{j\neq i}(x^j)^\top R^{ji} x^i + (b^i)^\top x^i - \theta^\top x^i$$

and the definition of $v_\theta(x_*(\theta)) := \left[\nabla_{x^i} u^i(x^i, x^{-i}; \theta)\right]_{i\in N}$. We can obtain that the matrix L is a constant matrix. Moreover, since incentives influence drivers' strategies by term $\theta^\top x^i$ in the payoff function, thus the matrix $\nabla_\theta v_\theta(x_*(\theta))$ is also a constant matrix. Thus the sensitivity matrix $\nabla_\theta x_*(\theta)$, which is the multiplication of two constant matrices mentioned above, is also a constant matrix. We denote this constant sensitivity matrix as S.

The assumption on the objective function $f_*(\theta)$ satisfying PL condition is usually hard to verify since it is a composition function involving $f(\cdot)$ and $x_*(\theta)$. The equilibrium mapping $x_*(\cdot)$ does not admit a closed-form solution and thus makes the assumption hard to verify. Here we derive a sufficient condition for $f_*(\cdot)$ to satisfies the PL condition.

***Lemma** 5:* Consider the quadratic payoff function in (2). Suppose the upper-level objective function $f_*(\theta) = g(Bx_*(\theta))$, where $g(\cdot)$ is a μ_g strongly convex function and B is certain constant matrix. Then, we have

$$\|\nabla_\theta f_*(\theta)\|_2^2 \geq \mu_f\left[f_*(\theta) - f_*(\theta_p)\right],$$

where θ_p is the projection of θ to the optimality set Θ^*.

Proof: By the μ_g strongly convexity of $g(\cdot)$, we have, for any y, y':

$$g(y') \geq g(y) + \langle \nabla g(y), y' - y\rangle + \frac{\mu_g}{2}\|y - y'\|_2^2$$

By taking $y = Bx_*(\theta), y' = Bx_*(\theta_p)$, we have:

$$\begin{aligned} g(Bx_*(\theta_p)) \geq & g(Bx_*(\theta)) + \langle \nabla_y g(y)|_{y=Bx_*(\theta)}, B(x_*(\theta) - x_*(\theta_p))\rangle \\ & + \frac{\mu_g}{2}\|B(x_*(\theta) - x_*(\theta_p))\|_2^2 \\ \geq & g(Bx_*(\theta)) + \langle \nabla_x g(Bx)|_{x=x_*(\theta)}, x_*(\theta) - x_*(\theta_p)\rangle \\ & + \frac{\mu_g}{2}\sigma(B)^2\|x_*(\theta) - x_*(\theta_p)\|_2^2 \\ \geq & g(Bx_*(\theta)) - \frac{1}{2\mu_g\sigma(B)^2}\|\nabla_x g(Bx)|_{x=x_*(\theta)}\|_2^2 \end{aligned}$$

where the second inequality is due to $\nabla_x g(Bx) = B^\top \nabla_y g(y)|_{y=Ax}$, and $\sigma(B)$ is the smallest non-zero absolute singular value of B. Reorganizing the inequality, we have:

$$\|\nabla_x g(Bx)|_{x=x_*(\theta)}\|_2^2 \geq 2\mu_g\sigma(B)^2\left[f_*(\theta) - f_*(\theta_p)\right]$$

By lemma 4 and due to the definition of quadratic payoff function, we have derived above that $\nabla_\theta x_*(\theta)$ is a constant matrix and denote it by S. With the chain rule, we have

$$\begin{aligned} \|\nabla_\theta f_*(\theta)\|_2^2 &= \|\nabla_\theta x_*(\theta)\nabla_x g(Bx)|_{x=x_*(\theta)}\|_2^2 \\ &= \|S\nabla_x g(Bx)|_{x=x_*(\theta)}\|_2^2 \\ &\geq 2\mu_g\sigma(B)^2\sigma(S)[f_*(\theta) - f_*(\theta_p)] \end{aligned}$$

This completes the proof with $\mu_f = 2\mu_g\sigma(B)^2\sigma(S)$.

IV. Simulation

In this section, we apply our algorithm to a simple simulation setting. We show numerically that the upper-level objective function would eventually reach an optimal value and induce the desired distribution of idle drivers. We construct a graph consisting of five nodes, where ith node has v_i number of idle drivers and d_i number of demands. The topology is shown in Fig. 1.

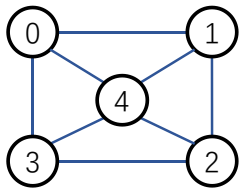

Fig. 1. Topology of the simulation settings

We exemplify the payoff function in (2) takes form as follows:

$$u^i(x^i, x^{-i}; \theta) := \sum_{j\in N^i} x^{ij}\frac{\sum_{k\in N_j} x^{kj} v_k}{d_j} + \sum_{j\in N^i} x^{ij}\frac{t_{ij} - t_{\min}}{t_{\max} - t_{\min}} - \sum_{j\in N^i}\theta_j x^{ij}$$

The first term in the payoff function penalize driving to nodes which have high drivers-demands ratio, the second term is the travelling time cost, and the third term is the received incentives. Drivers would update their strategies x^i to minimize the payoff function.

We also define the upper-level objective function $f(\cdot)$ to be the mean-squared error between the idle drivers distribution and the demand distribution, i.e.,

$$f_*(\theta) := \sum_{i\in N}\left(\frac{\sum_{j\in N^i} x_*^{ji}(\theta) v_j}{v_{sum}} - \frac{d_i}{d_{sum}}\right)^2$$

where $v_{sum} := \sum_{i \in N} v_i$ is the summation of all idle drivers, and $d_{sum} := \sum_{i \in N} d_i$ is the summation of demands.

Note that the distribution of idle drivers after repositioning can be represented as:

$$\bar{v} = \left[\sum_{j \in N^i} x_*^{ji}(\theta) v_j \right]_{i \in N}$$

could be represented by an affine function, i.e., Cx_* with some constant matrix C, of strategies x_*. Therefore, the upper-level objective function can be represented as $\|Ax - \bar{d}\|_2^2$ where $\bar{d} := [d_i/d_{sum}]_{i \in N}$ is the demands distribution. Note that the objective function is a L_2-norm and is obviously a strongly convex function with respect to Ax. We also add a zero mean Gaussian noise with 0.001 variance to the gradient of both lower-level payoff function and the upper-level objective function.

We randomly generate the number of idle drivers and demands at each node with the sum of drivers equals to the sum of demands. Therefore, the optimal distribution of repositioned idle drivers equals to the distribution of demands. In other words, the ratio of drivers and demands at each node is 1.

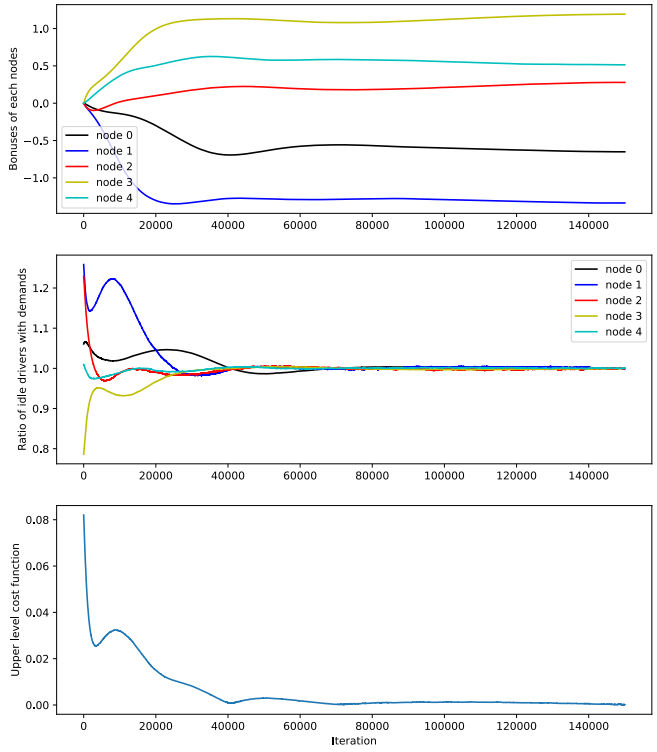

Fig. 2. Top figure shows the evolution of incentive policy. The middle figure shows the ratio of each nodes. Last figure shows the trajectory of upper-level objective function.

The simulation results are presented as in Fig 2. Note with increasing iteration, the ratio at each node all converge to the optimal ratio 1.0. Besides, the upper-level objective function decreases to 0. This demonstrate our result and prove that the upper-level objective function converge to the optimal value.

V. Conclusions

We have applied and analyzed a two time-scale bi-level optimization algorithm in solving the taxi repositioning problem. We proved that the incentive designer's objective function will converge to the optimal value under PL condition, implying that the social optimality is attained. Furthermore, under the setting of this work, we derive a sufficient condition for the PL condition to hold for both the upper-level objective function and lower-level agents' payoff function. Finally, we validated the convergence result of the algorithm with a numerical example.

References

[1] D. Mguni, J. Jennings, S. V. Macua, E. Sison, S. Ceppi, and E. M. De Cote, "Coordinating the crowd: Inducing desirable equilibria in non-cooperative systems," arXiv preprint arXiv:1901.10923, 2019.

[2] L. J. Ratliff, R. Dong, S. Sekar, and T. Fiez, "A perspective on incentive design: Challenges and opportunities," Annual Review of Control, Robotics, and Autonomous Systems, vol. 2, pp. 305–338, 2019.

[3] N. Adler, A. Brudner, and S. Proost, "A review of transport market modeling using game-theoretic principles," European Journal of Operational Research, vol. 291, no. 3, pp. 808–829, 2021.

[4] L. J. Ratliff and T. Fiez, "Adaptive incentive design," IEEE Transactions on Automatic Control, vol. 66, no. 8, pp. 3871–3878, 2020.

[5] D. Paccagnan, R. Chandan, and J. R. Marden, "Utility design for distributed resource allocation—part i: Characterizing and optimizing the exact price of anarchy," IEEE Transactions on Automatic Control, vol. 65, no. 11, pp. 4616–4631, 2019.

[6] J. Holler, R. Vuorio, Z. Qin, X. Tang, Y. Jiao, T. Jin, S. Singh, C. Wang, and J. Ye, "Deep reinforcement learning for multi-driver vehicle dispatching and repositioning problem," in 2019 IEEE International Conference on Data Mining (ICDM). IEEE, 2019, pp. 1090–1095.

[7] Y. Yue, B. L. Ferguson, and J. R. Marden, "Incentive design for congestion games with unincentivizable users," in 2021 60th IEEE Conference on Decision and Control (CDC). IEEE, 2021, pp. 4515–4520.

[8] C. Maheshwari, K. Kulkarni, M. Wu, and S. S. Sastry, "Inducing social optimality in games via adaptive incentive design," in 2022 IEEE 61st Conference on Decision and Control (CDC). IEEE, 2022, pp. 2864–2869.

[9] Z. Shou and X. Di, "Reward design for driver repositioning using multi-agent reinforcement learning," Transportation research part C: emerging technologies, vol. 119, p. 102738, 2020.

[10] J. Yang, E. Wang, R. Trivedi, T. Zhao, and H. Zha, "Adaptive incentive design with multi-agent meta-gradient reinforcement learning," arXiv preprint arXiv:2112.10859, 2021.

[11] B. Liu, J. Li, Z. Yang, H.-T. Wai, M. Hong, Y. M. Nie, and Z. Wang, "Inducing equilibria via incentives: Simultaneous design-and-play ensures global convergence," 2021. [Online]. Available: https://arxiv.org/abs/2110.01212

[12] F. Parise and A. Ozdaglar, "Sensitivity analysis for network aggregative games," in 2017 IEEE 56th Annual Conference on Decision and Control (CDC). IEEE, 2017, pp. 3200–3205.

[13] M. Hong, H.-T. Wai, Z. Wang, and Z. Yang, "A two-timescale framework for bilevel optimization: Complexity analysis and application to actor-critic," arXiv preprint arXiv:2007.05170, 2020.

[14] Facchinei, F. Pang, J. Finite-dimensional variational inequalities and complementarity problems. (Springe,2003)

[15] Mertikopoulos, P. & Zhou, Z. Learning in games with continuous action sets and unknown payoff functions. Mathematical Programming. **173** pp. 465-507 (2019)

[16] Facchinei, F. & Pang, J. Finite-dimensional variational inequalities and complementarity problems. (Springer,2003)

[17] H. Karimi, J. Nutini, and M. Schmidt, "Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition," in Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part I 16. Springer, 2016, pp. 795–811.