Adaptive Incentive Design with Minimized Regret

Georgios Vasileiou, Lantian Zhang and Silun Zhang

Abstract— **Incentive problems represent a well-established mathematical framework to describe interactions between sequential decision makers. Recent developments have leveraged control-theoretic and identification-based methods to address the analytic complexity and uncertainty of the problem. In this work we extend this framework by investigating the persistence of excitation necessary for the convergence of the underlying parameter estimator and further examine the exploration vs. exploitation trade-off introduced by decision makers' need to learn others' preferences. We derive an explicit persistence of excitation condition on the Recursive Least Squares estimator proposed and further demonstrate that normally distributed incentives adequately explore the parameter space. We utilize the resulting estimator to propose a switching-based incentive law that alternates between exploratory perturbations and targeted incentives, yielding minimized regret.**

I. INTRODUCTION

Incentive design problems, also referred to in literature as principal-agent (PA) problems or reverse Stackelberg games, consider a class of sequential decision makers coupled via interdependent costs. In these games the principal designs a mapping from agents' action space to his own and then makes this information available to influence the agents' responses. While this framework historically has its roots in economic theory [1], [2], it has recently seen significant development in control and machine learning to model human-in-the-loop systems [3]–[6]. Applications include adaptive pricing of energy products in smart grids [7]–[9], congestion-aware road tolling [10], [11] and data crowdsourcing markets [12], [13].

A significant challenge in incentive design problems is that of adverse selection [14], i.e. information asymmetry arising from unknown agent cost functions. While this challenge is usually addressed via the design of mechanisms that induce truthful participation [15], [16], these approaches are restricted to static, one-shot games and cannot be easily generalized. For this work we consider an approach introduced by Ratliff et. al. [17] in which agents' preferences (types) are adaptively estimated by the principal through repeated PA games and information is derived from the resulting Nash equilibria (NE). Although very promising, this method necessitates a restrictive persistence of excitation assumption (see [17, Theorem 1] and numerical examples therein) that is difficult to satisfy or verify a priori.

Our work makes two contributions toward addressing this limitation. We investigate games with quadratic costs [18], [19] and define a linear regression that enables the estimation of agent types through repeated observation of NE. We derive explicit eigenvalue bounds for the design matrix under normally distributed incentives and therefore guarantee the excitation of the type identification portion of the algorithm. Moreover, we consider the impact parameter estimation has on the principals ability to control agents' response and therefore introduce an incentive rule that asymptotically minimizes tracking regret by switching between exploration of the parameter space and exploitation of current type estimates.

II. PROBLEM FORMULATION AND PERSISTENTLY EXCITING INCENTIVES

We consider a noncooperative game of n agents, each choosing a strategy $x_i \in \mathbb{R}$ [1]. Each agent is equipped with a cost function

$$c_i(x,p) = \frac{1}{2} m_{ii} x_i^2 + \sum_{j \neq i} m_{ij} x_i x_j + p_i x_i, \tag{1}$$

with $m_i \in \mathbb{R}^n$, where the first two terms represent a nominal cost dependent on the collective action $x = (x_i)_{i=1}^n$ and $p_i x_i$ is the incentive (reward or penalty) that the planner imposes on agent i. The principal's objective is to drive the agents' response to some desired point $x^d \in \mathbb{R}^n$ though agents' nominal costs are unknown. Following the method introduced in [17], we address the principal-agent problem as a control-theoretic estimation and tracking problem where the principal models agents' costs as $c_i(x,p) = \Phi_i^\top(x)\theta_i^* + p_i x_i$, where $\Phi_i : \mathbb{R}^n \to \mathbb{R}^m$ represents a collection of m monomial kernel functions and $\theta_i^* \in \mathbb{R}^m$ an unknown parameter to be estimated. We call θ_i^* the *agents' type*.

Assumption 1. *All agents' cost functions c_i are strongly convex and continuously differentiable.*

Under Assumption 1, the differentiable game $(c_i)_{i=1}^n$ enjoys a desirable property: for each incentive $p = (p_i)_{i=1}^n \in \mathbb{R}^n$, there exists a unique NE which we denote $x^*(p) \in \mathbb{R}^n$, and moreover the mapping $x^*(\cdot)$ is Lipschitz continuous [20]. Since each agent acts to minimize their own cost, the NE is determined by a first order optimality condition [19], [21] and therefore there exists an (unknown) nonsingular matrix $M \in \mathbb{R}^{n\times n}$, $M = [m_{ij}]_{i,j=1}^n$ such that $Mx^*(p) + p = 0$ for any $p \in \mathbb{R}^n$.

It is significant to note that the bijection $x^*(p)$ is linearly parametrized by $\theta^* = (\theta_i^*)_{i=1}^n \in \mathbb{R}^{nm}$ in the sense that there

Georgios Vasileiou, Lantian Zhang, and Silun Zhang are with the School of Engineering Sciences, KTH Royal Institute of Technology, Stockholm, Sweden {geovas, lantian, silunz}@kth.se

[1] We assume without loss of generality that agents' strategies are scalar to simplify notation. Results can be generalized to arbitrary finite-dimensional strategies with similar techniques.

exists a matrix-valued, nonlinear map $V : \mathbb{R}^n \to \mathbb{R}^{nm \times n}$ such that

$$V^\top(x^*(p))\theta^* + p = 0 \tag{2}$$

for any $p \in \mathbb{R}^n$, where

$$V(x) = \text{diag}\left(\frac{\partial \Phi_1(x)}{\partial x_1}, \dots, \frac{\partial \Phi_n(x)}{\partial x_n}\right).$$

Note that $V(x)$ is a polynomial mapping that may be computed from $\{\Phi_i\}$ given the principal's decision of kernels. The principal will utilize the linearity of agents' response, given in (2), by iteratively issuing incentives and observing the resulting NE.

Assumption 2. *For any incentive p_k issued at iteration k, the planner can observe the response $x_{k+1} = x^*(p_k)$.*

Remark. *Assumption 2 is restrictive in the sense that it necessitates agents' response to an issued incentive to be instantaneous or, vice versa, that "enough time" has passed until observation [17]. Recent works have attempted to address this limitation by considering the dynamics of agent's responses [19], [22] but usually must impose more structure onto agents' nominal costs.*

Given the collection of observations $\{(p_t, x_{t+1})\}_{t=0}^k$, the principal updates his estimate of types to minimize the squared mean error of regression (2), i.e.

$$\theta_{k+1} = \arg\min_{\theta \in \mathbb{R}^{nm}} \left\{\sum_{t=0}^{k} (V^\top(x_{t+1})\theta + p_t)^2\right\}.$$

A recursive least squares (RLS) estimator for the above loss is considered and therefore the x, θ-update iterations are

$$\begin{aligned} x_{k+1} &= x^*(p_k) = -M^{-1}p_k, && (3)\\ \theta_{k+1} &= \theta_k - L_k(V^\top(x_{k+1})\theta_k + p_k), && (4)\\ L_k &= \Delta_k V(x_{k+1})\left(\mathbb{I} + V^\top(x_{k+1})\Delta_k V(x_{k+1})\right)^{-1} && (5)\\ \Delta_{k+1} &= \Delta_k - \Delta_k V(x_{k+1}) \\ &\quad \cdot \left(\mathbb{I} + V^\top(x_{k+1})\Delta_k V(x_{k+1})\right)^{-1} V^\top(x_{k+1})\Delta_k && (6) \end{aligned}$$

where $\Delta_0 = \epsilon^{-1}\mathbb{I}$ for some $\epsilon > 0$ and p_k denotes the incentive rule to be designed.

Theorem 1. *Let Assumptions 1-2 be satisfied and NE response according to* (3)*, then for normally distributed incentives $p_k \sim \mathcal{N}(0, \Sigma_k)$, $\Sigma_k \succ 0$, the regressor covariance satisfies*

$$\delta_1 \mathbb{I} \preceq E\left[V(x_{k+1})V^\top(x_{k+1})\right] \preceq \delta_2 \mathbb{I},\ \forall k \geq 0, \tag{7}$$

where constants $\delta_1,\ \delta_2$ are

$$\begin{aligned} \delta_1 =& m^{-2}\left(\frac{\lambda_{min}(\Sigma_k^{\frac{1}{2}})}{\|M^{-1}\|_2}\right)^{2nK} \cdot \left(\max\{1, \|M\Sigma_k^{\frac{1}{2}}\|\}\right)^{2(1-m)} \\ &\cdot \left(\binom{d}{\lfloor d/2 \rfloor}(\lfloor d/2 \rfloor - 1)!!\right)^{-2n} > 0, \\ \delta_2 =& d!(1+d)^2 m^{2(m-1)} \cdot \left(\max\{1, \|M\Sigma_k^{\frac{1}{2}}\|\}\right)^2 \\ &\cdot \left(\binom{d}{\lfloor d/2 \rfloor}(\lfloor d/2 \rfloor - 1)!!\right)^{2n(m-1)} > 0, \end{aligned}$$

and $K \triangleq \sum_{a=0}^{d}\binom{n+a-1}{a}$.

By Theorem 1 it holds that $\limsup k^{-1}\sum_{t=1}^{k}\|V(x_{t+1})\|_F^2 < \infty$ a.s. and moreover $k^{-1}\sum_{t=1}^{k} V(x_{t+1})V^\top(x_{t+1}) \to \Gamma$ a.s. with the PSD matrix $\Gamma \succeq \delta_1\mathbb{I}$. Utilizing [23, Theorem 2] we can then conclude the strong consistency of the estimator θ_k.

Theorem 2. *Let Assumptions 1-2 hold and the NE response satisfy* (3)*. Then for $p_k \sim \mathcal{N}(0, \Sigma_k)$, $\Sigma_k \succ 0$, the type updates (4)-(6) guarantee $\theta_k \to \theta^*$ a.s. while $k \to \infty$.*

III. INCENTIVE DESIGN WITH REGRET MINIMIZATION

In this section, we consider the incentive design problem. The planner's objective is to design the incentive when the parameter θ^* is unknown, in such a way that it minimizes the average regret $\mathcal{R}_k$, defined as

$$\mathcal{R}_k = \frac{1}{k}\sum_{t=0}^{k-1}\|x_{t+1} - x^d\|_2^2.$$

To this end, we propose the following switching incentive law:

$$p_k = \begin{cases} -V^\top(x^d)\theta_k, & k \in [\tau_i, \sigma_i) \\ w_k, & k \in [\sigma_i, \tau_{i+1}) \end{cases} \tag{8}$$

where $w_k \sim \mathcal{N}(0, \mathbb{I})$, and $\{\tau_i\}_{i\geq 0}$, $\{\sigma_i\}_{i \geq 0}$ are switching times satisfying $0 = \sigma_0 \leq \tau_1 \leq \sigma_1 \leq \dots$ and

$$\sigma_i = \min\left\{t \geq \tau_i : \lambda_{min}(t) < C\log\left(\sum_{k=0}^{t-1}\|V(x_{k+1})\|_F^2\right)^{\delta_n}\right\}$$

$$\tau_{i+1} = \min\left\{t \geq \sigma_i : \lambda_{min}(t) \geq C\log\left(\sum_{k=0}^{t-1}\|V(x_{k+1})\|_F^2\right)^{\delta_c}\right\}$$

where δ_n, δ_c and C are fixed constants with $0 < \delta_n < \delta_c$ and $C > 0$, and $\lambda_{min}(t)$ is defined as $\lambda_{min}(t) = \lambda_{\min}\left\{\sum_{k=0}^{t-1} V(x_{k+1})V^\top(x_{k+1})\right\}$.

Theorem 3. *Let Assumptions 1-2 be satisfied, then the planner that imposes online type identification* (4)-(6) *and incentive design law* (8) *can minimize the average regret $\mathcal{R}_k$ asymptotically, i.e.,*

$$\lim_{k\to\infty}\mathcal{R}_k = 0. \tag{9}$$

IV. NUMERICAL EXAMPLE

Consider the two player quadratic game given by costs according to (1) where parameters $m_{ij},\ i,j \in \{1,2\}$ are such that M is nonsingular and Assumption 1 is satisfied. Define kernel functions for players to be $\Phi_i(x) = x_i\begin{bmatrix}1 & x_1 & x_2\end{bmatrix}^\top$, $i \in \{1,2\}$, and the corresponding types $\theta_1^* = \begin{bmatrix}0 & 0.5m_{11} & m_{12}\end{bmatrix}^\top$ and $\theta_2^* = \begin{bmatrix}0 & m_{21} & 0.5m_{22}\end{bmatrix}^\top$. At each iteration, the system planner uses the type identification given in (4)-(6) and incentives (8), and observes the agent response x_{k+1} that is the unique NE satisfying equilibrium condition (3). Switching times for (8) are selected with

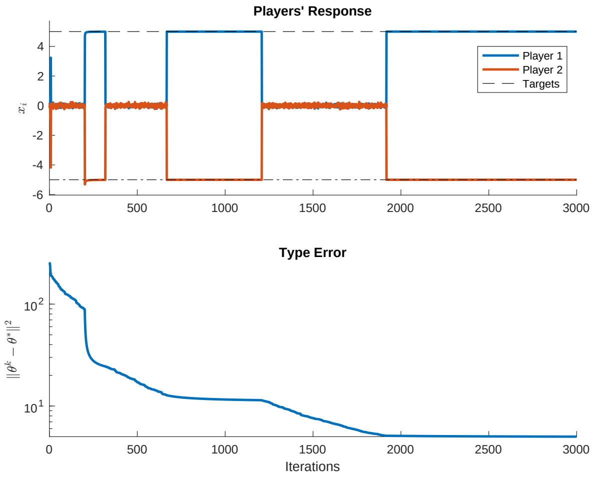

Fig. 1. (a) Players' response to issued incentives and (b) parameter estimation error. Notice that the first switch to exploitation occurs early in the learning cycle, so agents' responses exhibit some error. As estimation error converges, principal can precisely incentivize the agents.

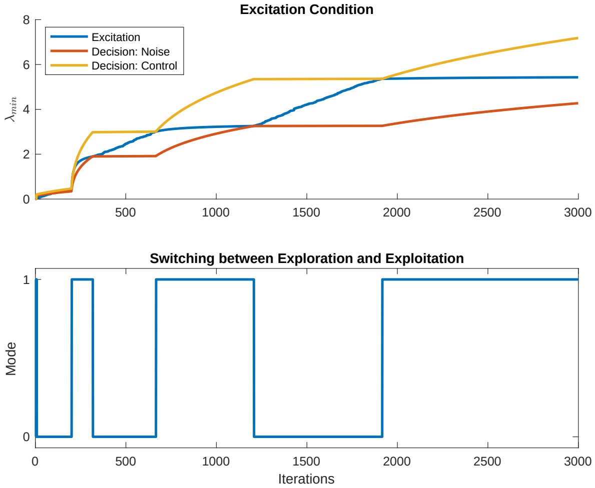

Fig. 2. (a) Switching boundaries used in (8). (b) Switching trajectory between exploration and exploitation: *Mode 1* represents the exploitation incentive, and *Mode 0* is exploration.

constants $C = 10^{-2}$, $\delta_n = 3.5$ and $\delta_c = 3.8$. Fig. 1 presents the trajectories of player response to issued incentives and shows the strong convergence of the type estimator. Fig. 2 presents the switching criteria and decision boundaries dependent on the excitation signal $\lambda_{min}(t)$. Finally, Fig. 3 demonstrates the per iteration regret accumulated by the proposed scheme.

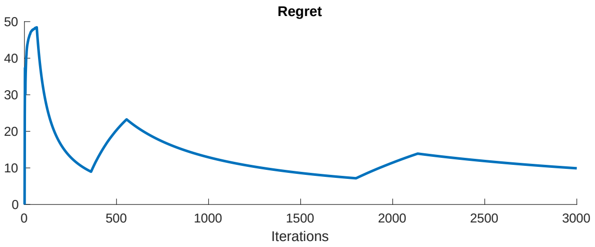

Fig. 3. Accumulated regret per iteration for the incentive (8).

References

[1] H. von Stackelberg, *The Theory of Market Economy*. Oxford University Press, 1952.

[2] J.-J. Laffont and D. Martimort, *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press, 2009. [Online]. Available: http://www.jstor.org/stable/10.2307/j.ctv7h0rwr

[3] Y.-c. Ho, P. Luh, and G. Olsder, "A control-theoretic view on incentives," in *1980 19th IEEE Conference on Decision and Control Including the Symposium on Adaptive Processes*. IEEE, 1980, pp. 1160–1170.

[4] T. Başar, "Affine Incentive Schemes for Stochastic Systems with Dynamic Information," vol. 22, no. 2, pp. 199–210, 1984.

[5] N. Groot, B. De Schutter, and H. Hellendoorn, "Reverse Stackelberg games, Part I: Basic framework," in *2012 IEEE International Conference on Control Applications*. IEEE, 2012, pp. 421–426. [Online]. Available: http://ieeexplore.ieee.org/document/6402334/

[6] L. J. Ratliff, R. Dong, S. Sekar, and T. Fiez, "A Perspective on Incentive Design: Challenges and Opportunities," vol. 2, no. 1, pp. 305–338, 2019.

[7] P. Li, H. Wang, and B. Zhang, "A distributed online pricing strategy for demand response programs," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 350–360, 2019.

[8] J. Li, M. Motoki, and B. Zhang, "Socially optimal energy usage via adaptive pricing," vol. 235, p. 110640. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0378779624005261

[9] C. Eid, P. Codani, Y. Perez, J. Reneses, and R. Hakvoort, "Managing electric flexibility from distributed energy resources: A review of incentives for market design," *Renewable and Sustainable Energy Reviews*, vol. 64, pp. 237–247, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032116302222

[10] J. I. Poveda, P. N. Brown, J. R. Marden, and A. R. Teel, "A class of distributed adaptive pricing mechanisms for societal systems with limited information," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017, pp. 1490–1495.

[11] J. Barrera and A. Garcia, "Dynamic incentives for congestion control," *IEEE Transactions on Automatic Control*, vol. 60, no. 2, pp. 299–310, 2015.

[12] C.-J. Ho, A. Slivkins, and J. W. Vaughan, "Adaptive contract design for crowdsourcing markets: bandit algorithms for repeated principal-agent problems," in *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, ser. EC '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 359–376. [Online]. Available: https://doi.org/10.1145/2600057.2602880

[13] A. Singla and A. Krause, "Truthful incentives in crowdsourcing tasks using regret minimization mechanisms," in *Proceedings of the 22nd International Conference on World Wide Web*, ser. WWW '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 1167–1178. [Online]. Available: https://doi.org/10.1145/2488388.2488490

[14] P. Picard, "On the design of incentive schemes under moral hazard and adverse selection," vol. 33, no. 3, pp. 305–331, 1987.

[15] N. Nisan, T. Roughgarden, E. Tardos, and Vazirani, *Algorithmic Game Theory*. Cambridge university press, 2007.

[16] J. D. Hartline. (2013) Mechanism Design and Approximation. [Online]. Available: http://jasonhartline.com/MDnA/

[17] L. J. Ratliff and T. Fiez, "Adaptive Incentive Design," vol. 66, no. 8, pp. 3871–3878, 2021.

[18] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory, 2nd Edition*. Society for Industrial and Applied Mathematics, 1998.

[19] C. Maheshwari, K. Kulkarni, M. Wu, and S. Sastry. (2024) Adaptive Incentive Design with Learning Agents.

[20] S. Dafermos, "Sensitivity Analysis in Variational Inequalities," vol. 13, no. 3, pp. 421–434, 1988.

[21] L. J. Ratliff, S. A. Burden, and S. S. Sastry, "On the Characterization of Local Nash Equilibria in Continuous Games," vol. 61, no. 8, pp. 2301–2307, 2016.

[22] J. Li, J. Wei, M. Motoki, Y. Jiang, and B. Zhang. Adaptive Pricing for Optimal Coordination in Networked Energy Systems with Nonsmooth Cost Functions. [Online]. Available: http://arxiv.org/abs/2504.00641

[23] T. L. Lai and C. Z. Wei, "Least Squares Estimates in Stochastic Regression Models with Applications to Identification and Control of Dynamic Systems," vol. 10, no. 1, 1982.