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

Inducing Social Optimality in Games via Adaptive Incentive Design

Chinmay Maheshwari, Kshitij Kulkarni, Manxi Wu, and S. Shankar Sastry

2022 IEEE 61st Conference on Decision and Control (CDC) | 978-1-6654-6761-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/CDC51059.2022.9992685

Abstract— **How can a social planner adaptively incentivize selfish agents who are learning in a strategic environment to induce a socially optimal outcome in the long run? We propose a two-timescale learning dynamics to answer this question in games. In our learning dynamics, players adopt a class of learning rules to update their strategies at a faster timescale, while a social planner updates the incentive mechanism at a slower timescale. In particular, the update of the incentive mechanism is based on each player's externality, which is evaluated as the difference between the player's marginal cost and the society's marginal cost in each time step. We show that any fixed point of our learning dynamics corresponds to the optimal incentive mechanism such that the corresponding Nash equilibrium also achieves social optimality. We also provide sufficient conditions for the learning dynamics to converge to a fixed point so that the adaptive incentive mechanism eventually induces a socially optimal outcome. Finally, as an example, we demonstrate that the sufficient conditions for convergence are satisfied in Cournot competition with finite players.**

I. Introduction

The design of incentive mechanisms plays a crucial role in many social-scale systems, where the system outcomes depend on the selfish behavior of a large number of interacting players (human users, service providers, and operators). The outcome arising from such strategic interaction – Nash equilibrium – often leads to a suboptimal societal outcome. This is due to the fact that individual players often ignore the externality of their actions (i.e. how their actions affect the cost of others) when minimizing their own cost. An important way to address the issue of externality is to provide players with incentives that align their individual goal of cost minimization with the goal of minimizing the total cost of the society ([1], [2], [3], [4]).

The problem of incentive design is further complicated when the design faces a set of learning agents who are repeatedly updating their strategies ([5], [6], [7]). Such a problem is particularly relevant when the physical system has experienced a random shock, and players are in the process of reaching a new equilibrium. Designing a socially optimal incentive mechanism directly based on the convergent strategy of the learning agents is challenging because such an equilibrium is typically difficult to compute in large-scale systems. The question that then arises is: how can a social planner design an adaptive incentive mechanism to influence players' learning dynamics so that the strategy learning under the adaptive mechanism leads to a socially beneficial outcome in the long run?

We propose a discrete-time learning dynamics that jointly captures the players' strategy updates and the designer's updates of incentive mechanisms. The incentive mechanism designed by the social planner sets a payment (tax or subsidy) for each player that is added to their cost function in the game. In each time step, players update their strategies based on the opponents' strategies and the incentive mechanism in the current step, and the social planner updates the incentive mechanism in response to players' current strategies. We assume that the incentive update proceeds at a slower timescale than the strategy update of players. The slower evolution of incentives, which allows the players to consider the incentives as static, is in-fact a desirable characteristic for any societal scale system, where frequent changes of incentives may lead to instability in the system and may hamper participation.

A key feature of our learning dynamics is that the incentive update in each time step is based on the externality created by each player with their current strategy. In particular, given any strategy profile, the externality of each player is evaluated as the difference between the marginal cost of their strategy on themselves and the marginal social cost. In a static incentive design problem, when all players are charged with their externality, the change of their total cost – original cost in game plus the payment – with respect to their strategy becomes identical to the change of social cost. Consequently, the induced Nash equilibrium is also socially optimal [8], [9], [10]. In our learning dynamics, the social planner accounts for the externality of each player evaluated at their current strategy, which evolves with players' strategy updates.

The externality-based incentive updates distinguish our adaptive incentive design from other recent studies on incentive mechanisms with learning agents. The paper [11] studies the problem of incentive design while learning the cost functions of players. The authors assume that both the cost functions and incentive policies are linearly parameterized, and the incentive updates rely on the knowledge of players' strategy update rules instead of just the current strategy as in our setting. Additionally, the paper [12] considers a two-timescale discrete-time learning dynamics, where players adopt a mirror descent-based strategy update, and the social planner updates an incentive parameter according to a gradient descent method. The convergence of such gradient-based learning dynamics relies on the assumption that the social cost when evaluated at players' equilibrium strategy is convex in the incentive parameter. However, the convexity

C. Maheshwari, K. Kulkarni, and S. Sastry are with EECS, University of California Berkeley, USA. M. Wu is with the school of Operations Research and Information Engineering at Cornell University, NY, USA. The work was done while M. Wu was at the Simons Institute for the Theory of Computing, Berkeley. chinmay_maheshwari@berkeley.edu, kshitijkulkarni@berkeley.edu manxiwu@cornell.edu, sastry@eecs.berkeley.edu.

assumption can be restrictive since the equilibrium strategy as a function of the incentive parameter is nonconvex even in simple games.

We show that our externality-based incentive updates ensure that any fixed point of our learning dynamics corresponds to a optimal incentive mechanism, such that the induced Nash equilibrium of the game is also socially optimal (Proposition 3.1). This result is built on the fact that at any fixed point of our learning dynamics, the strategy profile is a Nash equilibrium corresponding to the incentive mechanism, and each player's payment equals to the externality created by their equilibrium strategy. Therefore, the equilibrium strategy associated with this externality-based payment also minimizes the social cost. Additionally, we present the sufficient conditions on the game such that the fixed point set is a singleton set, and thus the socially optimal incentive mechanism is unique (Proposition 3.2).

Furthermore, we provide sufficient conditions on games that guarantee the convergence of strategies and incentives induced by our learning dynamics (Theorem 3.3). Since the convergent strategy profile and incentive mechanism corresponds to a fixed point that is also socially optimal, these sufficient conditions guarantee that the adaptive incentive mechanism eventually induces a socially optimal outcome in the long run.

In the proof of our convergence theorem, we exploit the timescale separation between the strategy update and the incentive updates. We use tools from the theory of two-timescale dynamical systems [13] to analyze the convergence of strategy updates and incentive updates separately after accounting of time separation. In particular, the convergence of strategy updates can be derived from the rich literature of learning in games ([14],[15],[16], etc.) since the incentive mechanism can be viewed as static in the strategy updates, thanks to the time separation. On the other hand, the convergence of incentive vectors can be analyzed via the associated continuous-time dynamical system where the value of the externality function is evaluated at the converged value of fast strategy update, which is the Nash equilibrium. Our sufficient conditions are based broadly on two main techniques of proving global stability of non-linear dynamical system: *(i)* cooperative dynamical systems theory [17] and *(ii)* Lyapunov based methods [18].

Finally, we apply our general results to a practically relevant class of game: Cournot competition. We provide sufficient conditions on the game parameters and social cost functions under which the adaptive incentive design eventually induces a socially optimal outcome. In the extended version of this article [19], we have also verified the effectiveness of the proposed adaptive incentive design framework on games such as: (i) networked quadratic games; (ii) routing games.

The article is organized as follows: in Sec. II we describe the setup considered here. In addition, we also provide the joint strategy and incentive update considered in this paper. We present the main results in Sec. III and the applications of those results to Cournot competition in Sec. IV. We conclude our work in Sec. V. Due to space limitations we only provide the main ideas of proofs, the detailed proofs are available in the extended version of this paper [19].

Notations

For any vector $x \in \mathbb{R}^n$, we use x_j or x^j to denote the $j-$th component of that vector. Given a function $f : \mathbb{R}^n \to \mathbb{R}$, we use $D_{x_i} f(x)$ to denote $\frac{\partial f}{\partial x_i}(x)$, the derivative of f with respect to x_i for any $i \in \{1, 2, ..., n\}$. We use k to denote the discrete-time index and t to denote the continuous-time index. We use ∇ to denote the gradient of a function, and $\langle \cdot, \cdot \rangle$ to denote inner product.

II. Model

We introduce the model of static games considered in this paper in Sec.II-A. In Sec. II-B, we present the two-timescale dynamics of strategy learning and incentive design.

A. Static games

Consider a game G with a finite set of players[1] $\mathcal{I}$. The strategy of each player $i \in \mathcal{I}$ is $x_i \in X_i$, where X_i is a non-empty and closed interval in $\mathbb{R}$. The strategy profile of all players is $x = (x_i)_{i \in \mathcal{I}}$, and the set of all strategy profiles is $X \coloneqq \prod_{i \in \mathcal{I}} X_i$. The cost function of each player $i \in \mathcal{I}$ is $\ell_i : X \to \mathbb{R}$. For any $x_{-i} = (x_j)_{j \in \mathcal{I} \setminus \{i\}}$, we assume that the cost function $\ell_i(x_i, x_{-i})$ is twice-continuously differentiable and strictly convex in x_i for all $i \in \mathcal{I}$.

A *social planner* designs incentives by setting a payment $p_i x_i$ for each player i that is linear in their strategy x_i.[2] Here, p_i represents the marginal payment for every unit increase in strategy of player i. The value of p_i can either be negative or positive, which represents a marginal subsidy or a marginal tax, respectively.

Given the incentive vector $p = (p_i)_{i \in I}$, the total cost of each player $i \in \mathcal{I}$ is:

$$c_i(x, p) = \ell_i(x) + p_i x_i, \quad \forall\ x \in X. \tag{1}$$

Note that here we assume that the cost of any players is independent of incentives of other players. However, it depends indirectly on the incentives of other players through their strategies.

Definition 2.1: A strategy profile $x^*(p) \in X$ is a *Nash equilibrium* in the game G with the incentive vector p if

$$c_i(x_i^*(p), x_{-i}^*(p), p) \leqslant c_i(x_i, x_{-i}^*(p), p),\ \forall\ x_i \in X_i,\ \forall i \in \mathcal{I}.$$

Recall that the cost $\ell_i(x_i, x_{-i})$ is a continuous function, and is strictly convex in x_i. Additionally, the strategy set X_i is convex and compact for each player i. Therefore, we know that Nash equilibrium must exist and must be unique in G. Moreover, we can equivalently represent a Nash equilibrium x^* as a strategy profile that satisfies the following variational inequality ([20]):

$$\langle J(x^*(p), p), x - x^*(p) \rangle \geqslant 0, \quad \forall\ x \in X, \tag{2}$$

[1]Due to space limitations we have presented the results in the setting of non-atomic games in [19].

[2]Considering a linear payment is sufficient to ensure optimal incentive design. This is a consequence of our results presented in Sec. III.

where $J(x^*(p),p) = (J_i(x^*(p),p))_{i\in\mathcal{I}}$, and $J_i(x^*(p),p) = D_{x_i}c_i(x^*(p),p) = D_{x_i}\ell_i(x^*) + p_i$.

Furthermore, a strategy profile $x^\dagger \in X$ is *socially optimal* if $x^\dagger$ minimizes the social cost function $\Phi : X \to \mathbb{R}$. We assume that the social cost function $\Phi(x)$ is strictly convex and twice continuously differentiable in x. Then, the optimal strategy profile $x^\dagger$ is unique. Additionally, from the first order conditions of optimality, we know that $x^\dagger$ minimizes the social cost function Φ if and only if:

$$\langle \nabla\Phi(x^\dagger), x - x^\dagger\rangle \geqslant 0, \quad \forall\ x \in X. \tag{3}$$

Finally, given a strategy profile $x \in X$, we define the *externality* caused by player i as the difference between the marginal social cost, and the marginal cost of player i with respect to x_i. That is,

$$e_i(x) = D_{x_i}\Phi(x) - D_{x_i}\ell_i(x). \tag{4}$$

B. Learning dynamics

We now introduce the discrete-time learning dynamics considered in this paper. For every time step $k = 1, 2, ...$, we denote the strategy profile in the game G as $x_k = (x_{i,k})_{i\in\mathcal{I}}$, where $x_{i,k}$ is the strategy of player i in step k. Additionally, we denote the incentive vector as $p_k = (p_{i,k})_{i\in\mathcal{I}}$. The strategy update and the incentive update are presented below:

Strategy update. In each step $k+1$, the updated strategy is a linear combination of the previous strategy in stage k (i.e. x_k), and a new strategy (i.e. $f(x_k,p_k) \in X$) that depends on the previous strategy and the incentive vector in stage k. The relative weight in the linear combination is determined by the step-size $\gamma_k \in (0,1)$.

$$x_{k+1} = (1-\gamma_k)x_k + \gamma_k f(x_k,p_k) \qquad (x\text{-update})$$

We consider generic strategy update (x-update) such that the new strategy profile $f(x_k,p_k) = (f_i(x_k,p_k))_{i\in\mathcal{I}}$ can incorporate a variety of strategy update rules. Two simple examples of such update include:

1) *Equilibrium update:* The strategy update incorporates a Nash equilibrium strategy profile with respect to the incentive vector in stage k, i.e., $f(x_k,p_k) = x^*(p_k)$.
2) *Best response update:* The strategy update incorporates a best response strategy with respect to the strategy and the incentive vector in the previous step, i.e. $f_i(x_k,p_k) = \mathrm{BR}_i(x_k,p_k) = \underset{y_i\in X_i}{\arg\min}\ c_i(y_i,x_{-i,k},p_k)$.

Incentive update. In each step $k+1$, the updated incentive vector is a linear combination of the previous vector in step k (i.e. p_k), and the externality (i.e. $e(x_k)$) based on the strategy profile in step k. The relative weight in the linear combination is determined by the step size $\beta_k \in (0,1)$.

$$p_{k+1} = (1-\beta_k)p_k + \beta_k e(x_k); \qquad (p\text{-update})$$

The incentive update (p-update) modify the incentives on the basis of the externality caused by the players. We emphasize that this update is adaptive to the evolution of players' strategies since the externality is evaluated based on players' current strategies. Moreover, the computation of each player's externality (4) only requires that the social planner knows the gradients of its own costs and those of the players, evaluated at the players' current strategy profile.

The joint evolution of strategy profiles and incentive vectors $(x_k,p_k)_{k=1}^\infty$ in game G is governed by the learning dynamics (x-update) – (p-update). The step-sizes $(\gamma_k)_{k=1}^\infty$ and $(\beta_k)_{k=1}^\infty$ determine the speed of strategy update and incentive update. We make the following assumption on step-sizes:

Assumption 2.1:

(i) $\sum_{k=1}^\infty \gamma_k = \sum_{k=1}^\infty \beta_k = +\infty$, $\sum_{k=1}^\infty \gamma_k^2 + \beta_k^2 < +\infty$.
(ii) $\lim_{k\to\infty} \frac{\beta_k}{\gamma_k} = 0$.

In Assumption 2.1, *(i)* is a standard assumption on step-sizes that allow us to analyze the convergence of the discrete-time learning dynamics. Additionally, *(ii)* assumes that the incentive update occurs at a slower timescale compared to the update of strategies. A class of candidate sequences which satisfy Assumption 2.1 is $\gamma_k = 1/k^\rho$ and $\beta_k = 1/k^\eta$, where $0.5 < \rho < \eta \leqslant 1$.

Since the assumption on stepsizes (Assumption 2.1 *(ii)*) ensures that the incentive evolves on a slower timescale than the strategies, players may view the incentive mechanism as approximately static (although not completely fixed) when updating their strategies. Note that with any fixed incentive mechanism, the equilibrium update always converges. On the other hand, although best response update, which we also consider, do not converge in all games, they converge in many practically-relevant games such as zero sum games [21], potential games [22], and dominance solvable games [23]. Additionally, our strategy update (x-update) can incorporate many other learning dynamics; their convergence properties in static game environments have been extensively studied in the literature [24], [15], [14], [16].

We emphasize that the convergence of strategy update with fixed incentive mechanism is *not* the focus of our paper. Instead, our goal is to characterize conditions under which the adaptive incentive update (p-update) converge to a socially optimal mechanism. We note that such convergence cannot be achieved in scenarios where the strategy update does not converge even with a completely fixed incentive vector. Therefore, we impose the following assumption that the strategy update considered in our dynamics converge to a Nash equilibrium with any fixed incentive vector.

Assumption 2.2: In G, the update (x-update) starting from any initial strategy x_1 with $(p_k) \equiv p$ for any p, satisfies $\lim_{k\to\infty} x_k = x^*(p)$, where $x^*(p)$ is the Nash equilibrium corresponding to p.

III. General results

In Sec III-A we characterize the set of fixed points of the dynamic updates (x-update)-(p-update), and show that any fixed point corresponds to a socially optimal incentive mechanism such that the induced Nash equilibrium strategy profile minimizes the social cost. In Sec. III-B, we provide a set of sufficient conditions that guarantee the convergence of strategies and incentives in our learning dynamics.

Consequently, under these conditions, the proposed adaptive incentive mechanism eventually induces a socially optimal outcome.

A. Fixed point analysis

We first characterize the set of fixed points of our learning dynamics (x-update)-(p-update) as follows:

$$\{(x,p)|f(x,p)=x,\ e(x)=p\}, \tag{5}$$

We can check that if the learning dynamics starts with a fixed point strategy and incentive vector, then the strategies and incentive vectors remain at that fixed point for all time steps. Moreover, under Assumption 2.2, we know that for any incentive vector p, a strategy profile that satisfies $f(x,p)=x$ must be a Nash equilibrium $x^*(p)$. Thus, from (5), we can write the set of incentive vectors at the fixed point as $P^\dagger = \{(p_i^\dagger)_{i\in\mathcal{I}}|e(x^*(p^\dagger)) = p^\dagger\}$. That is, at any fixed point, the incentive of each player is set to be equal to the externality evaluated at their equilibrium strategy profile.

Our next proposition shows that the fixed point set $P^\dagger$ is non-empty. Moreover, given any fixed point incentive parameter $p^\dagger \in P^\dagger$, the corresponding Nash equilibrium is socially optimal.

Proposition 3.1: The set $P^\dagger$ is non-empty. Additionally, any $p^\dagger \in P^\dagger$ is socially optimal in that $x^*(p^\dagger) = x^\dagger$.

The existence of the optimal incentives implies that there exists a *linear* incentive policy (as in (1)) which is optimal.

To provide a sketch of the proof: first, we use the boundedness of the strategy space to construct a convex compact set which maps to itself under the function $e(x^*(\cdot))$. Therefore, by Brouwer's fixed point theorem the set $P^\dagger$ is non-empty. Next, by the definition of the set $P^\dagger$ and (4) we have $D_{x_i}\Phi(x^*(p^\dagger)) = D_{x_i}\ell_i(x^*(p^\dagger)) + p^\dagger$ for every $i \in \mathcal{I}$. Using this along with the conditions of optimality (2)-(3) we establish that $x^*(p^\dagger) = x^\dagger$.

Next, we provide sufficient conditions under which the fixed point set $P^\dagger$ is singleton.

Proposition 3.2: The set $P^\dagger$ is singleton if any one of the following conditions holds:

(i) The equilibrium strategy profile $x^*(p)$ is in the interior of the strategy set X for any p

(ii) $\langle e(x) - e(x'), x - x'\rangle > 0$ for all $x \neq x'$

Under the sufficient condition in Proposition 3.2 there exists a unique incentive mechanism in $P^\dagger$ such that players pay for their externality at equilibrium. From Proposition 3.1, such a mechanism induces a socially optimal outcome.

B. Convergence to optimal incentive mechanism

The next result provides sufficient conditions for strategies and incentives updates (x-update)-(p-update) to converge to social optimality.

Theorem 3.3: Under Assumptions 2.1 and 2.2, the sequence of strategies and incentives induced by the discrete-time dynamics (x-update)-(p-update) satisfies

$$\lim_{k\to\infty}(x_k,p_k) = (x^\dagger,p^\dagger) \tag{6}$$

if at least one of the following conditions holds:

(C1) $\frac{\partial e_i(x^*(p))}{\partial p_j} > 0$ for all $p \in \mathbb{R}^{|\mathcal{I}|}$ and all $i \neq j$. Additionally, any one of the following holds:

(C1-a) If $e_i(x^*(0)) \geqslant 0$, then $\lim_{p\to+\infty} e_i(x^*(p)) - p_i < 0$ for all $i \in \mathcal{I}$.[3]

(C1-b) If $e_i(x^*(0)) \leqslant 0$, then $\lim_{p\to-\infty} e_i(x^*(p)) - p_i > 0$ for all $i \in \mathcal{I}$

(C2) There exists a continuously differentiable, positive definite and decrescent function [4] $V(p): \mathbb{R}^n \to \mathbb{R}_+$ such that $V(p^\dagger) = 0$ and $V(p) > 0$ for all $p \neq p^\dagger$. Moreover:

$$\nabla V(p)^\top (e(x^*(p)) - p) < -\omega(\|p - p^\dagger\|) \quad \forall\ p \neq p^\dagger,$$

where $\omega(\cdot)$ is strictly increasing, and satisfies $\omega(0) = 0$.

Owing to Assumption 2.1, we utilize the timescale separation between the strategy update (x-update) and the incentive update (p-update) to prove Theorem 3.3. Indeed, the two-timescale stochastic approximation theory [13] suggests that the strategy update (x-update) is a fast transient while the incentive update (p-update) is a slow component. Therefore while considering the fast strategy update the slow incentive updates are quasi-static. Consequently, Assumption 2.2 along with Assumption 2.1 ensures that the tuple (x_k, p_k) converges to the set $\{(x^*(p), p) : p \in \mathbb{R}^{|\mathcal{I}|}\}$ [13]. Thus for sufficiently large values of k, the update x_k closely tracks $x^*(p_k)$. Therefore, we consider the following update to analyze the convergence of the slow incentive update (p-update):

$$\mathbf{p}_{k+1} = (1-\beta_k)\mathbf{p}_k + \beta_k e(x^*(\mathbf{p}_k)). \tag{7}$$

As the step sizes $\{\beta_k\}$ are asymptotically going to zero and satisfy Assumption 2.1-(i), we can approximate the updates in (7) by the following continuous-time dynamical system:

$$\dot{\mathbf{p}}(t) = e(x^*(\mathbf{p}(t))) - \mathbf{p}(t), \tag{8}$$

Convergence of the discrete-time updates (x-update)-(p-update) then hold if the flow of (8) globally converges to $P^\dagger$.

Requirement (C1) in Theorem 3.3 is a sufficient condition for convergence of the trajectories of (8) to the set $P^\dagger$. This condition is based on cooperative dynamical systems theory [17]. Particularly, the condition: $\frac{\partial e_i(x^*(p))}{\partial p_j} > 0$ for all $p \in \mathbb{R}^{|\mathcal{I}|}$ and all $i \neq j$, ensures that the flow of (8) is strongly monotone (in the sense of [17]). Intuitively, this condition requires that the externality created by any player increases (resp. reduces) if the marginal payment (resp. subsidy) of other players is increased. This should typically be the case for games with resource constraints e.g. routing games [19]. Further, the condition (C1-a) (resp. (C1-b)) ensures that $\mathbb{R}_+^{|\mathcal{I}|}$ (resp. $\mathbb{R}_-^{|\mathcal{I}|}$) is invariant with respect of flow of (8) Additionally, these conditions also ensure that the forward orbit closure of (8) is compact. Intuitively,

[3] $p \to +\infty$ means $p_i \to +\infty$ for all i. Similarly, $p \to -\infty$ means $p_i \to -\infty$ for all i.

[4] A function $V: \mathbb{R}^n \to \mathbb{R}$ is positive definite if $V(x) \geqslant \alpha_1(\|x\|)$ for some continuous, strictly increasing function $\alpha_1(\cdot)$ such that $\alpha_1(0) = 0$, and $\alpha_1(t) \to \infty$ as $t \to \infty$. V is decrescent if $V(x) \leqslant \alpha_2(\|x\|)$ for some continuous, strictly increasing function $\alpha_2(\cdot)$ such that $\alpha_2(0) = 0$.

condition (C1-a) (resp. (C1-b)) requires than if marginal increment in social cost due to a players strategy at the equilibrium, corresponding to no payments, is larger (resp. smaller) than marginal increment in their own cost then there exists suitably high payments (resp. subsidies) which balances marginal increments in social cost and players' cost. Requirement (C2) in Theorem 3.3 on the other hand ensures convergence by positing existence of a Lyapunov function [18] that is strictly positive everywhere except at $P^\dagger$ and decreases along the flow of (8).

Note that either one of the conditions (C1) or (C2) guarantees the convergence of the flow of the slow system (8) to $P^\dagger$. This in addition to the convergence of the fast strategy update (Assumption 2.2) leads to the convergence of the discrete-time dynamics (x-update)-(p-update) [13].

Thus, we have shown that there exists an incentive which induces an equilibrium which is socially optimal and the externality based pricing update along with any strategy update, satisfying requirements of Theorem 3.3, converges to the optimal incentive.

IV. Application: Cournot Competition

In this section we apply our general results to Cournot competition between firms. We design an adaptive incentive mechanism to account for the environmental externality produced by the firms. We show that our mechanism asymptotically induces a socially optimal outcome via both theoretical results and numerical simulations. Additionally, we have studied several other practically relevant class of games such as networked aggregative games and routing games in the extended version of this article [19].

Consider a finite set of firms $\mathcal{I}$ that compete in a single market. The strategy of each firm $i \in \mathcal{I}$ is its production quantity x_i. Given any strategy profile $x = (x_i)_{i\in\mathcal{I}}$, the price of the good is $\xi(x) = \theta - \delta\sum_{i\in\mathcal{I}} x_i$ with $\theta, \delta > 0$. The per-unit production cost of the good is ν_i. Consequently, the cost function of firm $i \in \mathcal{I}$ (written as negative of the profit) is $\ell_i(x) = -x_i\xi(x) + \nu_i x_i$.

A social planner designs an incentive mechanism that charges each player i with payment $p_i x_i$. The total cost of firm $i \in \mathcal{I}$ given x and p is: $c_i(x,p) = -x_i\xi(x) + (\nu_i + p_i)x_i$ The game has a unique Nash equilibrium given by: [5]

$$x_i^*(p) = \frac{1}{\delta(|\mathcal{I}|+1)}\left(\theta - \nu_i - |\mathcal{I}|p_i + \sum_{j\neq i} p_j\right) \quad (9)$$

The goal of the social planner is to minimize the aggregate cost of players while also accounting for the environmental cost of good production, which is unpriced in equilibrium. We model the environmental cost to be a quadratic function of production following [25]. Thus, the social cost function is $\Phi(x) = \sum_{i=1}^n \ell_i(x) + \lambda\sum_{i=1}^n x_i^2$ where $\lambda > 0$ is a parameter that determines the relative weight between the firm costs and environmental cost. Finally, the externality (4) caused by of a firm $i \in \mathcal{I}$ is $e_i(x) = 2\lambda x_i + \delta\sum_{j\neq i} x_j$. We consider *best-response* based strategy updates. Given any x_{-i}, the best response of firm $i \in \mathcal{I}$ is:

$$BR_i(x_{-i}, p_i) = \frac{\theta - \delta\sum_{j\neq i} x_j - \nu - p_i}{2\delta}.$$

Following (x-update) – (p-update), we can write the updates of strategies and incentives as follows:

$$x_{i,k+1} = (1-\gamma_k)x_{i,k} + \gamma_k\left(\frac{\theta - \delta\sum_{j\neq i} x_{j,k} - \nu - p_{i,k}}{2\delta}\right), \quad (10a)$$

$$p_{i,k+1} = (1-\beta_k)p_{i,k} + \beta_k(\delta\sum_{j\neq i} x_{j,k} + 2\lambda x_{i,k}). \quad (10b)$$

and the step-sizes $\{\gamma_k\}_{k=1}^\infty, \{\beta_k\}_{k=1}^\infty$ satisfy Assumption 2.1.

We show that for any fixed p, the best response learning dynamics (10a) converge to a Nash equilibrium $x^*(p)$ associated with p. Indeed, we show that the strategy update (10a) with fixed incentives asymptotically track the flow of a continuous-time linear dynamical system whose flow asymptotically converges to the Nash equilibrium $x^*(p)$. Thus, Assumption 2.2 is satisfied.

The next proposition shows that the optimal incentive $p^\dagger$ is unique. Moreover, the incentive vectors induced by (10b) converge to the socially optimal incentive $p^\dagger$ if the weight of environmental cost, λ, is sufficiently high.

Proposition 4.1: There exists a unique socially optimal incentive mechanism $p^\dagger$ that satisfies $p^\dagger = e(x^*(p^\dagger))$. Given $p^\dagger$, the induced equilibrium strategy profile is socially optimal, i.e. $x^*(p^\dagger) = x^\dagger$. Moreover, the discrete-time learning dynamics (10a)-(10b) satisfy $\lim_{k\to\infty}(x_k, p_k) = (x^\dagger, p^\dagger)$ if $\lambda > \delta$.

The sufficient condition $\lambda > \delta$ states that if the social planner assigns higher weight to the environmental cost compared to the per-unit increase of firm cost, then the adaptive incentive mechanism can asymptotically induce a socially optimal outcome.

To prove Proposition 4.1, we show that there is a unique incentive $p^\dagger$ such that the corresponding Nash equilibrium as in (9) equals to the socially optimal strategy profile, and $p^\dagger$ is a fixed point of the discrete-time learning dynamics (10b). Moreover, we show that when $\lambda > \delta$, we can construct a Lyapnov function that satisfies (C2) in Assumption (2.2). Therefore, following Theorem (3.3), we can conclude that the discrete-time learning dynamics converges to a socially optimal outcome.

We now numerically illustrate the convergence results in Proposition 4.1. We study the above model with parameters $N = 6, \theta = 100, \delta = 5, \lambda = 50, \nu = [10, 18, 26, 34, 42, 50]$. The simulation results are presented in Fig. 1. We note that as the strategies and prices have converged, this shows that game has reached a Nash equilibrium. Furthermore, since the social cost goes to zero, the resulting Nash equilibrium is infact the socially optimal outcome. Furthermore, in the limit, the firm with the lowest production cost (P1) produces the highest quantity of the good but is also charged the

[5] We assume that θ is large enough such that $x^*(p) > 0$ for all p in a neighborhood of the socially optimal incentive $p^\dagger$. We have made this assumption to simplify exposition. For the general setting, refer to [19].

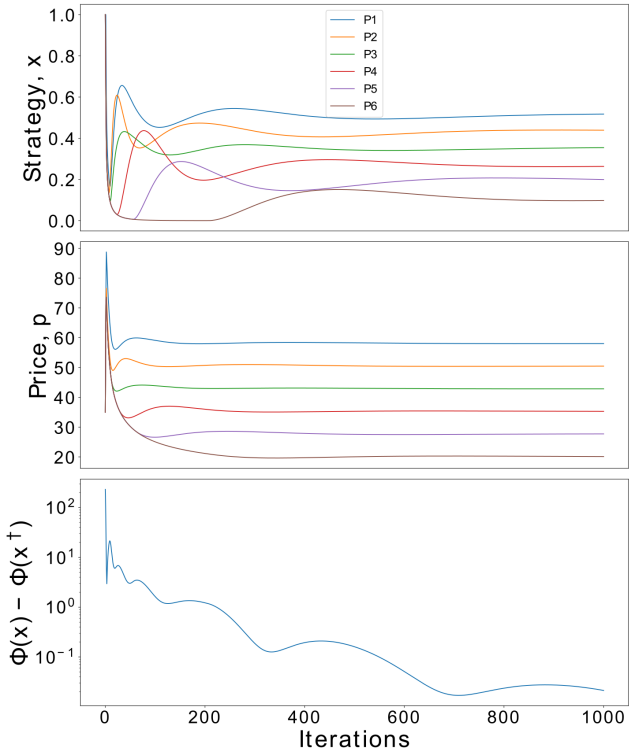

Fig. 1. The first subfigure depicts the evolution of the strategy (quantity produced) by different firms over time. The second subfigure depicts the evolution of price per unit production quantity for different firms. The last subfigure depicts the evolution of social cost, showing that asymptotically, the strategies are socially optimal.

highest in order to reduce its externality. Similarly, the firm with the highest production cost (P6) has the lowest production quantity and consequently lowest payment.

V. Conclusion

We propose an incentive design framework to steer selfish players, who are dynamically updating their strategies in a strategic environment, to socially optimal outcome. The incentives for players are dynamically updated based on the difference between the marginal impact of players' strategies on social cost and their own cost (the externality). Furthermore, the incentives are updated at a slower rate than the rate at which players update their strategies. We show that at the fixed point of joint strategy-incentive updates, the resulting Nash equilibrium coincides with a socially optimal outcome. Furthermore, we provide sufficient conditions for convergence of the joint strategy-incentive updates. Finally, we demonstrate the applicability of the proposed incentive design framework to Cournot competition.

Acknowledgements

Research is supported in part by the U.S. ONR MURI grant N00014-16-1-2710, NSF Partnership for International Research and Education Excellence (PIRE): Science of Design for Societal Scale Cyber-Physical Systems, NSF award number OISE-1743772, NSF grant DMS 2013985 "THEORINet: Transferable, Hierarchical, Expressive, Optimal, Robust and Interpretable Networks" and C3.ai Digital Transformation Institute. Manxi Wu acknowledges the fellowship support provided by the Simons Institute for the Theory of Computing.

References

[1] D. Paccagnan, R. Chandan, B. L. Ferguson, and J. R. Marden, "Incentivizing efficient use of shared infrastructure: Optimal tolls in congestion games," *arXiv preprint arXiv:1911.09806*, 2019.
[2] D. Levin, "Taxation within cournot oligopoly," *Journal of Public Economics*, vol. 27, no. 3, pp. 281–290, 1985.
[3] Y.-C. Ho, P. B. Luh, and G. J. Olsder, "A control-theoretic view on incentives," *Automatica*, vol. 18, no. 2, pp. 167–179, 1982.
[4] T. Başar, "Affine incentive schemes for stochastic systems with dynamic information," *SIAM Journal on Control and Optimization*, vol. 22, no. 2, pp. 199–210, 1984.
[5] J. Barrera and A. Garcia, "Dynamic incentives for congestion control," *IEEE Transactions on Automatic Control*, vol. 60, no. 2, pp. 299–310, 2014.
[6] G. Como and R. Maggistro, "Distributed dynamic pricing of multiscale transportation networks," *IEEE Transactions on Automatic Control*, 2021.
[7] C. Maheshwari, K. Kulkarni, M. Wu, and S. Sastry, "Dynamic tolling for inducing socially optimal traffic loads," *arXiv preprint arXiv:2110.08879*, 2021.
[8] H. R. Varian, "A solution to the problem of externalities when agents are well-informed," *The American Economic Review*, pp. 1278–1293, 1994.
[9] O. A. Davis and A. Whinston, "Externalities, welfare, and the theory of games," *Journal of Political Economy*, vol. 70, no. 3, pp. 241–262, 1962.
[10] A. C. Pigou and N. Aslanbeigui, *The economics of welfare*. Routledge, 2017.
[11] L. J. Ratliff and T. Fiez, "Adaptive incentive design," *IEEE Transactions on Automatic Control*, vol. 66, no. 8, pp. 3871–3878, 2020.
[12] 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 finds global optima," *arXiv preprint arXiv:2110.01212*, 2021.
[13] V. S. Borkar, "Stochastic approximation with two time scales," *Systems & Control Letters*, vol. 29, no. 5, pp. 291–294, 1997.
[14] D. Fudenberg and D. Levine, *The theory of learning in games*, vol. 2. MIT press, 1998.
[15] W. H. Sandholm, *Population games and evolutionary dynamics*. MIT press, 2010.
[16] D. Monderer and L. S. Shapley, "Fictitious play property for games with identical interests," *Journal of economic theory*, vol. 68, no. 1, pp. 258–265, 1996.
[17] M. W. Hirsch, "Systems of differential equations that are competitive or cooperative ii: Convergence almost everywhere," *SIAM Journal on Mathematical Analysis*, vol. 16, no. 3, pp. 423–439, 1985.
[18] S. Sastry, *Nonlinear systems: analysis, stability, and control*, vol. 10. Springer Science & Business Media, 2013.
[19] C. Maheshwari, K. Kulkarni, M. Wu, and S. Sastry, "Inducing social optimality in games via adaptive incentive design," *arXiv preprint arXiv:2204.05507*, 2022. (Extended version).
[20] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems*. Springer Science & Business Media, 2007.
[21] C. Harris, "On the rate of convergence of continuous-time fictitious play," *Games and Economic Behavior*, vol. 22, no. 2, pp. 238–259, 1998.
[22] B. Swenson, R. Murray, and S. Kar, "On best-response dynamics in potential games," *SIAM Journal on Control and Optimization*, vol. 56, no. 4, pp. 2734–2767, 2018.
[23] J. H. Nachbar, ""evolutionary" selection dynamics in games: Convergence and limit properties," *International journal of game theory*, vol. 19, no. 1, pp. 59–89, 1990.
[24] E. Mazumdar, L. J. Ratliff, and S. S. Sastry, "On gradient-based learning in continuous games," *SIAM Journal on Mathematics of Data Science*, vol. 2, no. 1, pp. 103–131, 2020.
[25] D. Cai, S. Bose, and A. Wierman, "On the role of a market maker in networked cournot competition," *Mathematics of Operations Research*, vol. 44, no. 3, pp. 1122–1144, 2019.