Dyn Games Appl (2017) 7:507–529
DOI 10.1007/s13235-016-0194-2

Total Reward Semi-Markov Mean-Field Games with Complementarity Properties

Piotr Więcek[1]

Published online: 14 May 2016

Abstract We study a class of dynamic games with a continuum of atomless players where each player controls a semi-Markov process of individual states, while the global state of the game is the aggregation of individual states of all the players. The model differs from standard models of dynamic games with continuum of players known as mean field or anonymous games in that the moments when the decisions are made are discrete, but different for each of the players. As a result, the individual states of each player follow a continuous time Markov chain, but the global state follows an ordinary differential equation. Games of this type were introduced by Gomes et al. (Appl Math Optim 68:99–143, 2013) and received some attention in the literature in last few years. In our paper we introduce a novel model of this type where players maximize their cumulative payoffs over their lifetime. We show that the payoffs of the players using any stationary strategy of a certain class in a game with continuum of players are close to those obtained in n-person counterparts of this game for n large enough. This implies that equilibrium strategies in the anonymous model can well approximate equilibria in related games with large finite number of players. In the rest of the paper we concentrate on a subclass of games where the payoff and transition probability functions exhibit some strategic complementarities between players. In that case we prove that the game possesses a stationary equilibrium. Moreover, largest and smallest equilibrium strategies are nondecreasing in the states. It also turns out that these equilibria can be well approximated using a distributed iterative procedure.

Keywords Mean-field game · Anonymous game · Stochastic game · Total reward · Strategic complementarity · Stationary equilibrium

This work is supported by the NCN Grant No. DEC-2011/03/B/ST1/00325.

✉ Piotr Więcek
Piotr.Wiecek@pwr.edu.pl

[1] Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

1 Introduction

Games with infinitely many atomless players have since long ago been used both in engineering and economics to model strategic interaction between large number of players, when the influence of an individual on the outcome of the game becomes negligible. Since the pioneering papers of Schmeidler [31] and Wardrop [38], they have become an important tool in modelling competitive markets, stock exchange and exploitation of common resources on one side, and network congestion or power control on the other.

Dynamic games of this type have been introduced in a paper by Jovanovic and Rosenthal [19]. In their framework each of the players controls an individual discrete-time Markov chain, while the global state of the game, defined as a probability distribution of individual states of all the players, becomes deterministic. The reward of an individual is then computed as the expectation of discounted sum of utilities obtained by him in infinitely many stages of the game. Some generalizations of their model were provided in [1,5–7]. The extension of their model to cover other utility criteria such as expected average utility and expected total utility was provided in [39].

Another important class of dynamic games with continuum of players has been introduced independently by Lasry and Lions [23–26] and by Huang et al. [16–18]. In their model the time is continuous, and so the evolution of both the individual and the global state of the game are described by ordinary differential equations. One can view their model as a generalization of differential games to games with continuum of players, while that of [19] as an extension of Markov or stochastic games to games with infinitely many players. The papers of Lasry and Lions have made an important impact on the entire game-theoretic community, additionally providing the name which is now commonly used to describe games of both types—"mean-field games". An overview of the state of the art in mean-field game theory can be found in [11], [1] includes a review of applications of mean-field games in economics, while [35] takes a look at those in engineering.

In this paper we concentrate on an intermediate concept, linking some features of mean-field games à Lasry and Lions and anonymous games of Jovanovic and Rosenthal. In our model the moments when the decisions are made are discrete, but follow separate controlled continuous time Markov chains, each controlled by a different player. As a result, these moments are different for each of the players—the process of individual states for each is a continuous time Markov chain, but the global state is, as in other mean-field game models, deterministic—following an ordinary differential equation. Model of this type has first appeared in the literature in a seminal paper of Gomes et al. [10] where characterization in terms of differential equations and main properties of this model were provided, together with a result on the convergence of n-person counterparts of this game to mean field limit. Further results of this type were provided in [9]. Some particular cases or applications of this type of games were also studied in [13,14,20,21,40]. In the paper we introduce a novel model of games of this type where players, instead of maximizing some payoff accumulated over the entire game, maximize the reward obtained during their lifetime, which may be different for different players. We assume that a dead player can be replaced after some time by a newborn one, and thus after some time we can obtain stationary behavior of the system which is then used to define a mean-field-type equilibrium. In the first part of the paper we give some sufficient conditions for these games to possess equilibria. These are of strategic complementarity type and are inspired by a paper on Markov-type discounted mean-field games [1]. Further, we show that the payoffs of the players using any given stationary strategy of a certain class in a semi-Markov mean-field game are close to those obtained in

its n-person counterparts for n large enough. This implies that equilibrium strategies in the anonymous model can well approximate equilibria in related games with large finite number of players.

The organization of the paper is as follows: In Sect. 2 we present the general framework we are going to work with and define what kind of solutions we will be looking for. In Sect. 3 we present our main results about the existence of equilibrium in games with strategic complementarities and convergence to equilibrium of a simple learning procedure, followed by some examples of applications of our model. Section 4 contains results linking mean-field game model presented earlier with games with large finite number of players. It is followed by conclusions in Sect. 5.

2 The Model

In this section, we formally describe the game model and the solution we will analyze in the remainder of the paper.

The semi-Markov mean-field game with total reward is described by the following objects:

- The game is played by an infinite number (continuum) of players. Each player has his own private state $s \in S$, changing over time. We assume that S is a finite set. We assume that there exists an element[1] s_0 standing for "death" of a player. Any player in state s_0 has no choice of action to play and receives no rewards. Moreover, his reward is computed over his "lifetime", that is, from one visit in state s_0 to his next visit there.
- The global state of the system at time t, X_t is a probability measure over S. It describes the mass of the population, which is at time t in each of the individual states. The set of global states of the game is thus[2] $\Delta(S)$. We assume that any player α has an ability to observe the global state of the game, so from his point of view the state of the game at time t is $(s_t^\alpha, X_t) \in S \times \Delta(S)$.
- We assume that the time is continuous, but the individual state of player α can only change at specific times $T_0^\alpha, T_1^\alpha, \ldots$, where $T_0^\alpha = 0$. The time between successive transitions $\tau_k^\alpha = T_{k+1}^\alpha - T_k^\alpha$ is random exponentially distributed with intensity $\lambda(s_{T_{k-1}^\alpha}, X_{T_k^\alpha})$. τ_k^α are for different k and α independent random variables. λ is a positive, Lipschitz-continuous function of the global state of the game.
- The set of actions available to a player in state (s, X) is a nonempty set $A(s, X)$, with $A := \bigcup_{(s,X)\in S\times\Delta(S)} A(s, X)$—a finite set. We assume that the mapping A is an upper semicontinuous function. We also assume that any player in state s_0 plays some default action a_0, not available in any other individual state.
  Let D denote the set of feasible state-action vectors, that is
$$D := \{(s, X, a) \in S \times \Delta(S) \times A : a \in A(s, X)\}.$$
- The transition for player α at time T_{k-1}^α is according to the transition function $q : D \to \Delta(S)$ which is a Lipschitz-continuous function of the global state. $q(\cdot|s_{T_{k-1}^\alpha}, X_{T_k^\alpha}, a_{T_k^\alpha})$ denotes the distribution of the individual state of player α after jump he makes at time T_k^α, given his previous state $s_{T_{k-1}^\alpha}$, his action $a_{T_k^\alpha}$ and the state-action distribution of all the players at time T_k^α. In particular, a player in state s_0 can join the game (be reborn) at time T in state s with probability $q(s|s_0, X_T, a_0)$.

[1] We can assume there is a whole subset of such elements.

[2] Here and in the sequel for any finite set B $\Delta(B)$ denotes the set of all the probability measures over B.

– We assume that all the players use *stationary strategies*, that is, they choose their actions depending only on their current individual state and current global state. Thus any strategy f is a Borel-measurable function from $S \times \Delta(S)$ to A such that for any $s \in S$ and $X \in \Delta(S)$, $f(s, X) \in A(s, X)$. The set of all stationary strategies will be denoted by F.
– The changes in individual states are aggregated according to the Kurtz (see Theorem 5.3 in [32]) dynamics:

$$\dot{X}_t^s = \sum_{s' \in S} \sum_{a \in A} X_t^{s'} \lambda(s', X_t) q(s|s', X_t, a) \overline{f}_a(s', X_t) - X_t^s \lambda(s, X_t), \quad s \in S \tag{1}$$

with $X_0 \equiv x_0$, the initial global state, where $\overline{f}$ denotes the average stationary policy used by the players. This average can be defined if the function $f_\alpha(s, X)$ is jointly measurable in (α, X) by the following equality

$$\overline{f}_a(s, X) := \int_0^1 \mathbb{1}\{f^\alpha(s, X) = a\} \, \mathrm{d}\alpha,$$

where f^α is the stationary strategy of player α. As we will see, in all our considerations, this will be a.e. a constant function of α, so the joint measurability will be immediately implied by measurability w.r.t. X. In the sequel, we will write $X_t(\overline{f})$ for the global state satisfying (1) when average stationary strategy is $\overline{f}$.
– Given the evolution of the global state, which depends on the strategies of the players in a deterministic manner, we can define the individual history of player α as the sequence of his consecutive individual states, actions and sojourn times $h = \left(s_{T_0^\alpha}^\alpha, \tau_0^\alpha, a_{T_1^\alpha}^\alpha, s_{T_1^\alpha}^\alpha, \ldots\right)$. By the Ionescu-Tulcea theorem (see Chap. 7 in [4]), for any stationary strategy f of player α and any initial individual state distribution μ_0, there exists a unique probability measure P_{f,μ_0} on the set of all infinite histories of the game $H = (S \times \mathbb{R}^+ \times A)^\infty$ endowed with Borel σ-algebra consistent with f, q and μ_0. Then the individual α's *expected total reward* is defined as the integral of his immediate (per unit time) reward function $r : D \to \mathbb{R}$ over his lifetime, plus the sum of rewards received upon the change of state awarded according to the function $\widetilde{r} : D \to \mathbb{R}$, which can be written as

$$J(f, \overline{g}, \mu_0) = \mathbb{E}^{P_{f,\mu_0}} \left[\sum_{i=0}^{i_e - 1} \left(\widetilde{r}(s_{T_i^\alpha}^\alpha, X_{T_i^\alpha}(\overline{g}), a_{T_i^\alpha}^\alpha) + \int_{T_i^\alpha}^{T_{i+1}^\alpha} r(s_{T_i^\alpha}^\alpha, X_t(\overline{g}), a_{T_i^\alpha}^\alpha) \, \mathrm{d}t \right) \right], \tag{2}$$

where T_{i_e} is the moment of his first return to s_0 and μ_0 is the initial distribution of all the new-born players. We assume both r and $\widetilde{r}$ are continuous in the global state of the game.

Since the game is symmetric, the equilibrium can be defined in the following manner. A stationary strategy f and a measure $\mu \in \Delta(S)$ are in *equilibrium* in the semi-Markov mean-field game with total reward if $X_0 = \mu$ implies $X^t(\overline{f}) \equiv \mu$ for every $t \geq 0$ and for every other stationary strategy $g \in F$,

$$J(f, \overline{f}, \rho) \geq J(g, \overline{f}, \rho),$$

where $\rho = q(\cdot|s_0, \mu, a_0)$ is the distribution of individual states of new-born players when global state is μ.

3 Game with Strategic Complementarities

In this section, we present the results about the existence of and convergence to equilibrium in our game for the model under some lattice-theoretic assumptions. Since the reader may be unfamiliar with lattice theory, below we present a brief introduction to it with all the notions used in the remainder of the paper. Those interested in deepening their knowledge about this subject are referred to [36], where concepts of lattices and supermodularity together with their applications to decision and game theory are discussed in detail.

3.1 Lattice-Theoretic Preliminaries

Let B be a partially ordered set with order $\preceq$. An element $b \in B$ is called an *upper bound* of $C \subset B$ if $b \succeq c$ for every $c \in C$. Similarly, b is a *lower bound* of C if $b \preceq c$ for all $c \in C$. We say that b is a *supremum* or a *least upper bound* of C in B if it is an upper bound of C and $b \succeq b'$ for any other upper bound of C, b'. Similarly a *least lower bound* or an *infimum* is defined. We say that B is a *lattice* if for every $b, b' \in B$, $\sup\{b, b'\}$, $\inf\{b, b'\}$ exist in B. We say that it is a *complete lattice* if for every nonempty $C \subset B$, $\sup\{C\}$, $\inf\{C\}$ exist in B.

Many commonly used partially ordered sets are lattices. For example $\mathbb{R}$ is a lattice with usual ordering as well as any $\mathbb{R}^n$ with componentwise ordering.[3] None of them is a complete lattice though. Compact intervals of $\mathbb{R}^n$ are simple examples of complete lattices. A lattice which will be of particular interest to us is that of Borel probability measures on $\mathbb{R}$, $\Delta(\mathbb{R})$, with (first order) stochastic dominance ordering $\preceq_{\mathrm{SD}}$ defined as follows:

$$P \preceq_{\mathrm{SD}} Q \iff \int_{\mathbb{R}} g(x)P(\mathrm{d}x) \leq \int_{\mathbb{R}} g(x)Q(\mathrm{d}x)$$

for any nondecreasing bounded measurable function $g : \mathbb{R} \to \mathbb{R}$.[4] It is well known that $P \preceq_{\mathrm{SD}} Q$ is equivalent to $F_P(x) \geq F_Q(x)$ for any $x \in \mathbb{R}$, where F_P and F_Q are cumulative distribution functions corresponding to P and Q respectively. Again, $\Delta(\mathbb{R})$ is not a complete lattice, but for any compact subset B of $\mathbb{R}$, $\Delta(B)$ is complete. It has been shown in [29] that the same is not true already for $\mathbb{R}^2$. There, even the set of probability measures defined on the set $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$ with stochastic dominance ordering is not a lattice, so any results basing on the lattice structure of $\Delta(\mathbb{R})$ cannot be directly repeated for $\Delta(\mathbb{R}^n)$, $n \geq 2$.

Let B be a lattice. A function $f : B \to \mathbb{R}$ is *nondecreasing* if $b \preceq b'$ implies $f(b) \leq f(b')$. f is *supermodular* if $f(\sup\{b, b'\}) + f(\inf\{b, b'\}) \geq f(b) + f(b')$. If C is also a lattice, we say that a function $f : B \times C \to \mathbb{R}$ has *increasing differences* in b and c if $b \succeq b'$, $c \succeq c'$ implies $f(b, c) - f(b', c) \geq f(b', c) - f(b', c')$. Finally, a correspondence $T : B \to C$ is *nondecreasing* if for any $b \preceq b'$ and $c \in T(b)$, $c' \in T(b')$, $\inf\{c, c'\} \in T(b)$ and $\sup\{c, c'\} \in T(b')$. If, instead of real-valued functions f we consider a function whose values are probability measures on $\mathbb{R}$ (a parametrized measure) with stochastic dominance ordering, we use terms *stochastically nondecreasing* and *stochastically supermodular* for the counterparts of the above properties. We say that a parametrized measure $f(\cdot|b, c)$ has *stochastically increasing differences* if $\int_{\mathbb{R}} g(a) f(da|b, c)$ has increasing differences for any nondecreasing bounded measurable function g.

[3] In the remainder of the paper we will use symbol $\leq$ for ordering in $\mathbb{R}$, while $\preceq$ will be used to denote componentwise ordering in $\mathbb{R}^n$.

[4] The symbol $\preceq_{\mathrm{SD}}$ will be used throughout the paper to denote stochastic dominance ordering.

3.2 Assumptions

Below we present the set of assumptions for the model considered in our paper. These assumptions (except the first one) are not necessary for the model to make sense, but will be used either to prove the existence of equilibria there or in some further results.

(A1) There exists a $p_0 > 0$ such that for any fixed global state μ and under any stationary policy f the probability of getting from any state $s \in S \setminus \{s_0\}$ to s_0 in $|S| - 1$ steps is not smaller than p_0.

(A2) S and A are sublattices[5] of $\mathbb{R}$ with $s_0 = \min\{S\}$ and $a_0 = \min\{A\}$ and for any $s \in S$ and $X \in \Delta(S)$, $A(s, X)$ is a sublattice of A. Moreover, $A(s, X)$ is nondecreasing in (s, X).

(A3) $r(s, X, a)$ and $\widetilde{r}(s, X, a)$ are nonnegative nondecreasing in s and supermodular in (s, a). Moreover, they have increasing differences in (s, a) and X.

(A4) $q(\cdot|s, X, a)$ is stochastically supermodular in (s, a) and stochastically nondecreasing in s, a and X. Moreover, it has stochastically increasing differences in (s, a) and X.

(A5) $\lambda(s, X)$ does not depend on s and is nonincreasing in X.

(A6) The value of $A(s, X)$ does not depend on X.[6]

Remark 1 Note that some of the above assumptions can be slightly relaxed if, instead of considering each of the functions defining the game separately, some combinations of them were characterized. In particular, assumptions (A3) and (A5) could be relaxed, if we did not assume the positivity, monotonicity and supermodularity of each of r, $\widetilde{r}$ and λ^{-1}, but rather assumed that $\widetilde{r}(s, X, a) + \frac{r(s,X,a)}{\lambda(s,X)}$ is nonnegative nondecreasing in s, supermodular in (s, a) and having increasing differences in (s, a) and X.

Remark 2 The assumption (A3) can be slightly generalized by considering the reward functions r and $\widetilde{r}$ depending not only on the individual state s and the action a of a given player and the global state X but also on the global distribution of actions that we can denote as Z. Then we could assume the following:

(A3') $r(s, X, a, Y)$ and $\widetilde{r}(s, X, a, Y)$ are nonnegative nondecreasing in s, supermodular in (s, a). Moreover, they have increasing differences in (s, a) and (X, Y).

The proofs of Theorems 1 and 2 can be repeated when (A3) is replaced with (A3') in the assumptions, although they become more complex notationally.

Remark 3 Our supermodularity/increasing differences assumptions are closely related to the monotonicity assumptions used by Lasry and Lions [26] to establish the uniqueness of equilibrium solution in a mean-field game. The assumptions of this type have been extensively used in the mean-field game literature, also for games with finite state space [10]. The formulations of these assumptions may slightly differ depending on other assumptions that are made, but they all can be viewed as very close to requiring strictly increasing differences in individual and global states of some function related to the Hamiltonian corresponding to the immediate reward (cost) function (or the immediate reward itself, see e.g. [12]) as well as of the terminal reward (cost). In our assumptions we require weak supermodularity and weakly increasing (nondecreasing) differences of the functions defining our model. It is easy to see that in a degenerated case when each of the functions r, $\widetilde{r}$ and q is constant on $S \times A \times \Delta(S)$,

[5] Note that since they both are finite, they are clearly complete lattices.

[6] Alternatively we could write that the correspondence S is continuous, but, since the set A is finite, this reduces to this seemingly more restrictive assumption.

our assumptions will not be violated, while any of the monotonicity assumptions used in the literature will.[7] It is natural, as we do not expect uniqueness of equilibrium in our model, but rather a special structure of equilibrium strategy set. Similarly, monotonic mean-field game models typically require some convexity assumptions to hold. In our case no convexity in any variable is assumed. On the other hand, apart from assuming increasing differences in individual and global state we make additional (weak) monotonicity assumptions about functions defining our model.

Remark 4 There is no discounting in our model, as (A1) guarantees that the expected rewards for the players are bounded. Note however that adding discounting does not change our results, so if some real-life application requires adding it to the model (which is often the case in economics), one is free to do so.

The assumptions of the strategic complementarity type have been used in the game-theoretic literature for a long time. A review of results for one-step games can be found in [36]. Some results about dynamic games with strategic complementarities can be found in [2,3,8,15,30,33,37]. A model of discounted dynamic games with continuum of players satisfying similar assumptions can be found in [1]. A general intuition about this type of conditions is the following: Strategic complementarity between some two quantities describes a situation when they mutually reinforce one another, that is an increase in one of them implies that it is profitable to increase the other one and vice versa. In dynamic games with complementarities we usually assume that strategic complementarity takes place between individual states of players, so an increase in one's state makes increase in others' state profitable. In addition, we usually assume (as we do here) that there is a complementarity between player's actions and his states, so that an increase in the state makes higher actions more profitable. Finally, we also need to make some monotonicity assumptions about the immediate rewards and the transition law, which are crucial for the aggregate reward of a player to preserve the strategic complementarity of immediate reward functions. It turns out that many games possess this kind of properties, as seen in the example below. It should also be noted that many real-life applications can be modelled as total reward semi-Markov mean-field games with complementarities. Some of them are presented in Sect. 3.5.

Example 1 While some of the assumptions (A1–A6) are rather clear, it may be difficult for those not familiar with theory of supermodular functions to see what kind of functions satisfy assumptions (A3) and (A4). Below we present some examples. Functions r and $\widetilde{r}$ satisfying (A3) can be of any of the following forms:

$$\alpha(s)\beta(a)\mathbb{E}\left[\gamma(X)\right], \tag{3}$$

$$\min\{\alpha(s), \beta(a), \mathbb{E}\left[\gamma(X)\right]\}, \tag{4}$$

where $\alpha : S \to \mathbb{R}$, $\beta : A \to \mathbb{R}$, $\gamma : S \to \mathbb{R}$ are any nonnegative nondecreasing functions. They can also be of the form

$$c_1\alpha(s) + c_2\beta(a) + c_3\gamma(X) \tag{5}$$

where α is a nonnegative nondecreasing function, β and γ are any nonnegative functions of respective variables, while the constants $c_1, c_2, c_3 \geq 0$. Finally, they can be any conical

[7] Of course the same will be true if these functions are constant on some properly chosen subset of $S \times \Delta(S) \times A$.

combination of functions of forms (3–5), as well as of a quadratic function of the form[8]

$$-\mathbb{E}\left[(\beta(a)-\gamma(X))^2\right],$$

where β and γ are nondecreasing, provided it is nonnegative.

An example of the transition law satisfying (A4) was given by Nowak [30]:

$$q(\cdot|s,X,a)=f(s,X,a)q_1(\cdot|s,X,a)+(1-f(s,X,a))q_2(\cdot|s,X,a),$$

where $q_1 \succeq_{\mathrm{SD}} q_2$, while $f : S\times\Delta(S)\times A\to[0,1]$ is supermodular in (s,a) and nondecreasing in s,a and X. Moreover, it has increasing differences in (s,a) and X. Such a function can be constructed as a conic combination of functions (3–5) under additional condition that all the functions α,β,γ are nondecreasing.

3.3 Existence of Equilibrium

Now we can formulate the main result of this section.

Theorem 1 *A semi-Markov mean-field game with total reward satisfying assumptions (A1–A5) has an equilibrium (f^*,μ^*) such that f^* is nondecreasing in individual state and μ.*

Many of the arguments used in the proof are taken from [1] where discrete-time discounted mean-field games with strategic complementarities were considered. Whenever some results appearing there can be used here in an unchanged form, we refer the reader to some specific results in that paper. To start with, we need to introduce for any fixed global state X an auxiliary dynamic optimization model $\mathcal{M}(X)$. Suppose an individual controls a discrete-time Markov decision process with total cost, with

(a) the state space S and the action space A;
(b) the initial distribution of states μ_0;
(c) the transition probabilities[9]

$$Q_X(\cdot|s_t,a_t)=\begin{cases} q(\cdot|s_t,X,a_t) & \text{for any } s_t\neq s_0\\ \delta[s_0], & \text{for } s_t=s_0\end{cases},$$

so s_0 becomes now absorbing;
(d) the reward per stage given by the equality

$$R_X(s_t,a_t)=\widetilde{r}(s_t,X,a_t)+\frac{r(s_t,X,a_t)}{\lambda(s_t,X)}.$$

[8] If we assume that the reward functions depend also on the distribution of actions among the players Z (see Remark 2), then we can also add the quadratic function of the form

$$-\mathbb{E}\left[(\beta(a)-\gamma(Z))^2\right]$$

multiplied by a positive constant. In case of quadratic functions that depend only on individual and global state, function of the form

$$-\mathbb{E}\left[(\alpha(s)-\gamma(X))^2\right]$$

can only appear in a conic combination with some nonnegative nondecreasing function of s, such that the sum is also nonnegative and nondecreasing in s.

[9] Here and in the sequel $\delta[x]$ denotes a degenerate probability measure concentrated in point x.

Note that for any stationary strategy f, the reward received by the controller using f in this model equals the total reward (2) in case the global state induced by $\overline{g}$ is fixed and equal to X. Note also that this is a classic Markov decision process with total reward, as considered in the literature, and so standard dynamic programming arguments imply that:

(a) Since assumptions (A1) and (A2) hold, the optimal value in this model is finite.
(b) The optimal value in this model V_X^* has to satisfy for any $s \in S$ the following Bellman equation:

$$V_X^*(s) = \max_{a \in A(s,X)} \left[R_X(s,a) + \sum_{s' \in S} V_X^*(s') Q_X(s'|s,a) \right]. \tag{6}$$

(c) A is finite, and thus compact, which implies that 'sup' in (6) can be replaced by 'max', moreover, optimal stationary strategies in $\mathcal{M}(X)$ exist and can be identified as any strategies maximizing the RHS of (6).

In the first lemma we will show what are the main properties of V_X^*.

Lemma 1 *$V_X^*(s)$ is nondecreasing in s and has increasing differences in s and X.*

Proof The proof is for most part the repeat of the arguments used in [1]. It will be broken into three claims. Before we formulate the first one, we need to note two facts: First, that $R(s,X,a) = \frac{r(s,X,a)}{\lambda(s,X)}$ is by assumptions (A3) and (A5) a product of two functions that are nonnegative nondecreasing in s and supermodular in (s,a). As such, R preserves all these properties. Next, since $(\lambda(s,X))^{-1}$ in nonnegative, constant in s and nondecreasing in X while r has increasing differences in (s,a) and X,

$$\begin{aligned} R(s,X,a) - R(s',X,a') &= \frac{r(s,X,a)}{\lambda(s,X)} - \frac{r(s',X,a')}{\lambda(s',X)} \\ &= \frac{1}{\lambda(s,X)}(r(s,X,a) - r(s',X,a')), \end{aligned}$$

so for $(s,a) \succeq (s',a')$ it is a product of two nonnegative nondecreasing functions of X, thus a nondecreasing function itself. This means that R has increasing differences in (s,a) and X. Monotonicity, supermodularity, and increasing differences are preserved upon summation, so (by (A3)) $R_X(s,a) = \widetilde{r}(s,X,a) + \frac{r(s,X,a)}{\lambda(s,X)}$ also has all these properties.

Second, note that $Q_X(\cdot|s,a) = \begin{cases} q(\cdot|s,X,a) & \text{for any } s \neq s_0 \\ \delta[s_0], & \text{for } s = s_0 \end{cases}$ preserves all the properties of q, as:

(a) $\delta[s_0]$ is stochastically smaller than any other probability distribution over S, and so Q_X trivially stays stochastically nondecreasing in (s,a).
(b) Stochastically increasing differences in (s,a) and X are preserved because for $(s,a) \succ (s_0,a_0)$, $Q_X(\cdot|s,a) = q(\cdot|s,X,a)$ is stochastically nondecreasing in X, while $Q_X(\cdot|s_0,a_0)$ is constant.
(c) Supermodularity in (s,a) in D is trivial, as $(s_0,a_0) \prec (s,a)$ for any $(s,a) \neq (s_0,a_0)$, and so always $\sup\{(s_0,a_0),(s,a)\} = (s,a)$ and $\inf\{(s_0,a_0),(s,a)\} = (s_0,a_0)$.

Now we can pass to the main part of the proof.

Claim 1 *Let v be a bounded function of s and X, nondecreasing in s and having increasing differences in s and X. Then*

$$w(s,X,a) = \sum_{s' \in S} v(s',X) Q_X(s'|s,a)$$

is nondecreasing in s and a, supermodular in (s, a), and has increasing differences in (s, a) and X.

This claim has been shown in [1] as Lemma 3.

Claim 2 *Let v be a bounded function of s and X, nondecreasing in s and having increasing differences in s and X. Then*

$$T(s, X)(v) = \max_{a \in A(s,X)} \left[R_X(s, a) + \sum_{s' \in S} v(s', X) Q_X(s'|s, a) \right]$$

is nondecreasing in s and has increasing differences in s and X.

This claim has been shown in [1] as Lemma 4.

Claim 3 *$V_X^*(s)$ is nondecreasing in s and has increasing differences in s and X.*

By assumption (A1) we can write that for any two bounded functions of (s, X): v, w

$$\begin{aligned} &\max_{s \in S, X \in \Delta(S)} \left| T^{|S|}(s, X)(v) - T^{|S|}(s, X)(w) \right| \\ &\quad \leq (1 - \min_{s \in S, a \in A, X \in \Delta(S)} q^{|S|}(s_0|s, X, a)) \max_{s \in S, X \in \Delta(S)} |v(s, X) - w(s, X)| \\ &\quad \leq (1 - p_0) \max_{s \in S} |v(s) - w(s)| \end{aligned}$$

and so $T^{|S|}$ is a contraction. Since the set of bounded functions of (s, X) which are nondecreasing in s and have increasing differences in s and X is a closed subset of a complete metric space of bounded functions from $S \times \Delta(S)$ to $\mathbb{R}$, it is also a complete metric space, and consequently $T^{|S|}$ has a unique fixed point in this set.

Now take $V_X^0 \equiv 0$ and define for $k > 0$

$$V_X^k(s) = \max_{a \in A(s,X)} \left[R_X(s, a) + \sum_{s' \in S} V_X^{k-1}(s') Q_X(s'|s, a) \right].$$

It is clear that $V_X^k(s) = T^k(s, X)(V_X^0)$. Consequently, $V_X^*(s) = \lim_{k \to \infty} V_X^k(s) = \lim_{k \to \infty} T^{k|S|}(s, X)(V_X^0)$ which equals the fixed point of $T^{|S|}$. This proves that $V_X^*(s)$ has all the desired properties. □

Next, let us define a correspondence that can be viewed as a best response operator:

$$\mathcal{B}(X)(s) = \arg \max_{a \in A(s,X)} \left[R_X(s, a) + \sum_{s' \in S} V_X^*(s') Q_X(s'|s, a) \right].$$

Next, let $\underline{B}(X)$ and $\overline{B}(X)$ denote the smallest and the biggest best responses, that is

$$\underline{B}(X)(s) = \min \mathcal{B}(X)(s), \quad \overline{B}(X)(s) = \max \mathcal{B}(X)(s).$$

The fact that they are both well defined, as well as their crucial properties, are shown in the following lemma.

Lemma 2 *$\mathcal{B}(X)$ is nondecreasing in (s, X). Moreover, $\underline{B}(X)(s)$ and $\overline{B}(X)(s)$ are well defined, nondecreasing in X and, for a fixed X, also nondecreasing in s.*

Proof The proof is based on two results by Topkis. First, define

$$f(a, s, X) = R_X(s, a) + \sum_{s' \in S} V_X^*(s') Q_X(s'|s, a).$$

By Lemma 1 $V_X^*(s)$ is nondecreasing in s and has increasing differences in s and X. Next, we can use Claim 1 of this lemma to show that this implies that $\sum_{s' \in S} V_X^*(s') Q_X(s'|s, a)$ is nondecreasing in s, supermodular in (s, a), and has increasing differences in (s, a) and X. Since $R_X(s, a)$ also has these properties (which was shown at the beginning of the proof of Lemma 1) and as they are preserved under summation, $f(a, s, X)$ is also nondecreasing in s, supermodular in (s, a), and has increasing differences in (s, a) and X. Note also that by assumption (A2) $A(s, X)$ is nondecreasing in (s, X). Now we can apply Theorem 2.8.1 in [36] to obtain the first part of the lemma. The second statement follows from Theorem 2.8.3 (a) in [36]. □

In the next lemma we come back to the original game model and analyze the properties of stationary individual state distributions when a player applies a given stationary strategy.

Lemma 3 *Suppose that the global state of the game is constant and equal to X. Then the smallest stationary state distribution corresponding to a stationary strategy*

$$f \in F_0 := \{g \in F : g(s, X) \text{ is nondecreasing in } X \text{ and for any fixed } X \text{ in } s\},$$

$\underline{X}(f, X)$ *and the greatest stationary state distribution corresponding to* f, $\overline{X}(f, X)$, *are nondecreasing functions of* f *and* X *on* $F_0 \times \Delta(S)$.

Proof First, note that a stationary global state Y corresponding to the stationary strategy f used by all the players and the fixed global state of the game X must satisfy for every $s \in S$ the following equation:

$$\sum_{s' \in S} \sum_{a \in A} Y^{s'} \lambda(s', X) q(s|s', X, a) f_a(s', X) - Y^s \lambda(s, X) = 0.$$

Note however that by (A5) $\lambda(s, X)$ does not depend on s. As it is always nonzero, we can cancel out all the λ terms from the above equation, obtaining

$$Y^s = \sum_{s' \in S} \sum_{a \in A} Y^{s'} q(s|s', X, a) f_a(s', X). \tag{7}$$

Clearly, by (A4) and the fact that f is nondecreasing, $q(\cdot|s', X, f(s', X))$ is stochastically nondecreasing in s' and X, as well as in f, as long as strategies from F_0 are applied.

Now define $\phi : \Delta(S) \times \Delta(S) \times F_0 \to \Delta(S)$ with equality

$$\phi^s(Y, X, f) = \sum_{s' \in S} Y^{s'} q(s|s', X, f(s', X)).$$

We will show that this is a nondecreasing function. Let $Y \preceq_{\mathrm{SD}} \widetilde{Y}$, $f \preceq \widetilde{f}$ and $X \preceq_{\mathrm{SD}} \widetilde{X}$. As $q(\cdot|s'.X, f(s', X))$ is stochastically nondecreasing in X and f, clearly

$$\sum_{s \in S} w(s) q(s|s', X, f(s', X)) \leq \sum_{s \in S} w(s) q(s|s', \widetilde{X}, \widetilde{f}(s', \widetilde{X})) \tag{8}$$

for any $s' \in S$ and any bounded nondecreasing function $w : S \to S$. This implies that

$$\sum_{s \in S} w(s) \phi^s(\widetilde{Y}, \widetilde{X}, \widetilde{f}) = \sum_{s \in S} w(s) \sum_{s' \in S} \widetilde{Y}^{s'} q(s|s', \widetilde{X}, \widetilde{f}(s', \widetilde{X}))$$

$$= \sum_{s' \in S} \widetilde{Y}^{s'} \left[\sum_{s \in S} w(s) q(s|s', \widetilde{X}, \widetilde{f}(s', \widetilde{X})) \right] \geq \sum_{s' \in S} \widetilde{Y}^{s'} \left[\sum_{s \in S} w(s) q(s|s', X, f(s', X)) \right].$$

Now note that since $q(\cdot|s', X, f(s', X))$ is stochastically nondecreasing in s', the expression in brackets is a nondecreasing function of s', and so since $\widetilde{Y} \preceq_{\mathrm{SD}} Y$, the RHS of the last inequality is not smaller than

$$\sum_{s' \in S} Y^{s'} \left[\sum_{s \in S} w(s) q(s|s', X, f(s', X)) \right] = \sum_{s \in S} \phi^s(Y, X, f),$$

proving that ϕ is nondecreasing. Now we can apply Theorem 3 in [28] to show that for any $X \in \Delta(S)$, $f \in F_0$, there exists an $Y \in \Delta(S)$ such that

$$Y = \phi(Y, X, f). \tag{9}$$

Moreover, the greatest and the smallest Y satisfying (9), that is the greatest and the smallest stationary distributions corresponding to X and f, $\overline{X}(X, f)$ and $\underline{X}(X, f)$, are nondecreasing in f. □

Proof of Theorem 1 Define

$$\overline{\Psi}(X) = \overline{X}(\overline{B}, X) \quad \text{and} \quad \underline{\Psi}(X) = \underline{X}(\underline{B}, X).$$

Both functions are nondecreasing in X (as superpositions of functions that are nondecreasing by Lemmas 2 and 3 respectively) and defined on a nonempty complete lattice $\Delta(S)$. Thus by Tarski's theorem [34] each of them has a fixed point which clearly defines an equilibrium in the game. Note also, that by Lemma 2 equilibrium stationary strategies ($\overline{B}$ and $\underline{B}$ respectively) are nondecreasing in s and X.

3.4 Distributed Learning

In the next part of this section we present a distributed iterative algorithm allowing players to learn to play the game. This kind of algorithms are known to exist for some types of games, and games with strategic complementarities are known to be one of them. The very simple and intuitive algorithm presented below is an adaptation for our game of an algorithm presented in [1].

Algorithm 1 (Lower Myopic Learning) For each time moment $t \geq 0$ repeat the following steps:

1. Every player making his move at time t observes current population state X_t.
2. A player in the individual state s chooses action $a_t = \underline{B}(X_t)(s)$.

The following theorem summarizes main properties of the Lower Myopic Learning Algorithm.

Theorem 2 *Suppose assumptions (A1–A6) are satisfied. Additionally assume that the initial state of the game X_0 satisfies the inequality*

$$X_0 \preceq_{SD} \phi(X_0, X_0, \underline{B}(X_0)). \tag{10}$$

and that all the players adjust their strategies according to the Lower Myopic Learning Algorithm. Then:

(a) *For every* α $a^{\alpha}_{T^{\alpha}_{i+1}} \geq a^{\alpha}_{T^{\alpha}_{i}}$, $i = 0, 1, \ldots, i^{\alpha}_{e} - 1$.

(b) X_t *is an increasing function of* t *converging to some* $\mathcal{X}$ *as* $t \to \infty$, *such that* $(\underline{B}, \mathcal{X})$ *are an equilibrium in the game.*

One lemma will be used in the proof of the above theorem.

Lemma 4 *Suppose that assumptions (A1–A6) are satisfied and that* $\widehat{X}$, $X_t \in \Delta(S)$, $t \in \mathbb{R}^+$ *such that* $X_t \nearrow \widehat{X}$. *If* f *is a stationary strategy such that*

$$f(X_t, s) \to_{t\to\infty} f(\widehat{X}, s) \text{ for any } s \in S, \tag{11}$$

then for any $s \in S$ *the reward from using policy* f *in model* $\mathcal{M}(X_t)$, $J_{X_t}(f, s)$, *converges to the reward from using* f *in* $\mathcal{M}(\widehat{X})$, $J_{\widehat{X}}(f, s)$, *as* t *goes to infinity.*

Proof For any bounded function $v : s \times X \to \mathbb{R}$, nondecreasing in s, such that

$$v(s, X_t) \to v(s, \widehat{X}) \text{ for any } s \in S \tag{12}$$

let us define the operator

$$K_f(s, X)(v) = R_X(s, f(X, s)) + \sum_{s' \in S} v(s', X) Q_X(s'|s, f(X, s)).$$

It is clear that for any $s \in S$, (11) together with the continuity of r, $\widetilde{r}$ and λ implies the following:

$$\lim_{t\to\infty} R_{X_t}(s, f(X_t, s)) = \lim_{t\to\infty} \left[\widetilde{r}(s, X_t, f(X_t, s)) + \frac{r(s, X_t, f(X_t, s))}{\lambda(s, X_t)}\right]$$
$$= \widetilde{r}(s, \widehat{X}, f(\widehat{X}, s)) + \frac{r(s, \widehat{X}, f(\widehat{X}, s))}{\lambda(s, \widehat{X})} = R_{\widehat{X}}(s, f(\widehat{X}, s)).$$

Then, also

$$\lim_{t\to\infty} \sum_{s' \in S} v(s', X_t) Q_{X_t}(s'|s, f(X_t, s)) = \sum_{s' \in S} v(s', \widehat{X}) Q_{\widehat{X}}(s'|s, f(\widehat{X}, s)),$$

by (11), (12) and the continuity of Q. This obviously implies that

$$\lim_{t\to\infty} K_f(s, X_t)(v) = K_f(s, \widehat{X}).$$

Consequently, by induction the same is true for $K^k_f(s, X)(v)$ with $k \in \mathbb{N}$.

Next note that r, $\widetilde{r}$ and λ are continuous on a compact domain, hence bounded. Let L be such that $|r(s, X, a)| \leq L$, $|\widetilde{r}(s, X, a)| \leq L$ and $\lambda(s, X) \leq L$ for any $(s, X, a) \in D$. In addition λ is by assumption positive, so there also exists a $\underline{\lambda} > 0$ such that $\lambda(s, X) \geq \underline{\lambda}$ for any $(s, X, a) \in D$. Consequently, $|R_X(s, a)| < L + \frac{L}{\underline{\lambda}}$ for any $(s, X, a) \in D$. Further note that by (A1) for any X, s, f and v,

$$\left|\lim_{m\to\infty} K^m_f(s, X)(v) - K^k_f(s, X)(v)\right|$$
$$\leq \left(L + \frac{L}{\underline{\lambda}}\right)(1 - p_0)^{\left\lfloor \frac{k}{|S|-1} \right\rfloor} \sum_{i=0}^{\infty} (1 - p_0)^i = \frac{L(\underline{\lambda} + 1)}{\underline{\lambda} p_0}(1 - p_0)^{\left\lfloor \frac{k}{|S|-1} \right\rfloor}.$$

Thus, for any $\varepsilon > 0$, $\left|\lim_{m\to\infty} K^m_f(s, X)(v) - K^k_f(s, X)(v)\right| < \frac{\varepsilon}{2}$ for k big enough, say $k \geq k_0$. Consequently,

$$\left|\lim_{t\to\infty} \lim_{m\to\infty} K^m_f(s, X_t)(v) - \lim_{m\to\infty} K^m_f(s, \widehat{X})(v)\right|$$

$$< \left| \lim_{t\to\infty} K_f^{k_0}(s, X_t)(v) - K_f^{k_0}(s, \widehat{X})(v) \right| + \varepsilon = \varepsilon.$$

Note however that this, in view of the arbitrarity of ε and because both limits on the LHS of the above set of inequalities exist, implies

$$\lim_{t\to\infty} \lim_{m\to\infty} K_f^m(s, X_t)(v) = \lim_{m\to\infty} K_f^m(s, \widehat{X})(v).$$

Note however that by standard dynamic programming arguments, $\lim_{m\to\infty} K_f^m(s, X)(v)$ equals $J_X(f, s)$. Thus we have proved that for every $s \in S$ $J_{X_t}(f, s) \to_{t\to\infty} J_{\widehat{X}}(f, s)$. □

Proof of Theorem 2 First, note that

$$X_t \preceq_{\mathrm{SD}} \phi(X_t, X_t, \underline{B}(X_t)). \tag{13}$$

is by definition equivalent to

$$\sum_{s\in S} \left[(\phi(X_t, X_t, \underline{B}))^s - X_t^s \right] h(s) \geq 0$$

for any nondecreasing function $h : S \to \mathbb{R}$, and further by the definition of ϕ and (A5) to

$$\sum_{s\in S} \left[\sum_{s'\in S} X_t^{s'} \lambda(s', X_t) q(s|s', X_t, \underline{B}(X_t)(s')) - X_t^s \lambda(s, X_t) \right] h(s) \geq 0. \tag{14}$$

Next, define[10]

$$H(h)(s) := \sum_{s\in S} X_t^s(\underline{B}) h(s),$$

where (as before) h is an arbitrary function from S to $\mathbb{R}$. Then by (1) and (14)

$$\begin{aligned} \frac{\mathrm{d}H(h)}{\mathrm{d}t} &= \sum_{s\in S} \dot{X}_t^s h(s) \\ &= \sum_{s\in S} \left[\sum_{s'\in S} X_t^{s'} \lambda(s', X_t) q(s|s', X_t, \underline{B}(X_t)(s')) - X_t^s \lambda(s, X_t) \right] h(s) \\ &\geq 0 \end{aligned}$$

This however means that as long as (13) holds, the global state of the game is increasing as time increases. Of course, it also implies that $a^\alpha_{T^\alpha_{i+1}} \geq a^\alpha_{T^\alpha_i}$, $i = 0, 1, \ldots, i^\alpha_e - 1$ for any player α, as $\underline{B}$ is nondecreasing.

Next assume that at some time t (13) is violated. Then, since at the beginning of the game it was by assumption true, and because of the continuity of the trajectory of X_t, there must exist a function h_0, such that

$$\sum_{s\in S} \left[(\phi(X_t, X_t, \underline{B}))^s - X_t^s \right] h_0(s) = 0.$$

Then, it is easy to see that

$$\frac{\mathrm{d}H(h_0)}{\mathrm{d}t} = \sum_{s\in S} \dot{X}_t^s h_0(s)$$

[10] In the remainder of the proof we skip the information that the global state corresponds to all players using policy $\underline{B}$.

$$= \sum_{s\in S}\left[\sum_{s'\in S} X_t^{s'}\lambda(s', X_t)q(s|s', X_t, \underline{B}(X_t)(s')) - X_t^s\lambda(s, X_t)\right]h_0(s)$$
$$= 0,$$

which implies that when the boundary of the set where (13) is satisfied is reached, the trajectory cannot leave the set.

Next note that since at any time t $X_t \preceq_{\mathrm{SD}} \delta[\max\{S\}]$, the fact that X_t is increasing implies that it converges to some $\mathcal{X}$ (recall that the stochastic domination ordering is equivalent to ordering of CDFs, which, as S is finite, is in turn equivalent to componentwise ordering in $\mathbb{R}^{|S|}$).

Next, define

$$\widehat{\underline{B}}(X)(s) := \begin{cases} \underline{B}(X)(s) & \text{for } X \neq \mathcal{X} \\ \lim_{t\to\infty} \underline{B}(X_t)(s) & \text{for } X = \mathcal{X} \end{cases}$$

We will now show that $\mathcal{X}$ is a stationary distribution corresponding to $\widehat{\underline{B}}$. From the definition of $\widehat{\underline{B}}$ and the continuity of λ and q we can infer that

$$\lim_{t\to\infty}\left[\sum_{s'\in S} X_t^{s'}\lambda(s', X_t)q(s|s', X_t, \widehat{\underline{B}}(X_t)(s')) - X_t^s\lambda(s, X_t)\right]$$
$$= \sum_{s'\in S} \mathcal{X}^{s'}\lambda(s', \mathcal{X})q(s|s', \mathcal{X}, \widehat{\underline{B}}(\mathcal{X})(s')) - \mathcal{X}^s\lambda(s, \mathcal{X}). \tag{15}$$

On the other hand, since $X_t \to_{t\to\infty} \mathcal{X}$ monotonically, for any s $\dot{X}_t^s \to 0$, which is equivalent to

$$\lim_{t\to\infty}\left[\sum_{s'\in S} X_t^{s'}\lambda(s', X_t)q(s|s', X_t, \widehat{\underline{B}}(X_t)(s')) - X_t^s\lambda(s, X_t)\right] = 0.$$

Combining this with (15) we obtain

$$\sum_{s'\in S} \mathcal{X}^{s'}\lambda(s', \mathcal{X})q(s|s', \mathcal{X}, \widehat{\underline{B}}(\mathcal{X})(s')) - \mathcal{X}^s\lambda(s, \mathcal{X}) = 0.$$

But this, by the definition of ϕ and (A5), means that $\mathcal{X}$ is a fixed point of $\phi(\cdot, \mathcal{X}, \widehat{\underline{B}})$, and consequently $\mathcal{X}$ is a stationary distribution corresponding to $\widehat{\underline{B}}$.

Next, note that under (A6) any vector of actions $\overline{a} = (a_s)_{s\in S}$ from sets $A(s, \mathcal{X})$ can be obtained as a value of a global state-independent policy defined by

$$f_{\overline{a}}(X, s) = a_s, \quad s \in S.$$

Clearly, each of the policies $f_{\overline{a}}$ satisfies (11). So does $\widehat{\underline{B}}$ by its construction. Thus we can use Lemma 4 to show the following

$$J_{\mathcal{X}}(\widehat{\underline{B}}, s) = \lim_{t\to\infty} J_{X_t}(\widehat{\underline{B}}, s) \geq \lim_{t\to\infty} J_{X_t}(f_{\overline{a}}, s) = J_{\mathcal{X}}(f_{\overline{a}}, s),$$

where the inequality follows from the fact that $\widehat{\underline{B}}(X) = \underline{B}(X)$ for $X \neq \mathcal{X}$ and the fact that for each t, $\underline{B}(X_t)$ is a best response to X_t. But this proves that $\widehat{\underline{B}}(\mathcal{X})$ is a best response to $\mathcal{X}$, as strategies $f_{\overline{a}}$ cover all the possible actions that a player can use at the global state $\mathcal{X}$. To end the proof, note that by the monotonicity of $\underline{B}$,

$$\widehat{\underline{B}}(\mathcal{X}) = \lim_{t\to\infty} \underline{B}(X_t) \leq \underline{B}(\mathcal{X}).$$

This however implies that $\widehat{\underline{B}}(\mathcal{X}) = \underline{B}(\mathcal{X})$, as the latter is by definition the smallest best response to $\mathcal{X}$. Since the two strategies could only differ at $\mathcal{X}$, this means that they are equal and that $(\underline{B}, \mathcal{X})$ is an equilibrium in the game. □

Remark 5 The assumption (10) is both difficult to check and rather restrictive. In [1] to avoid this kind of problem the authors start the algorithm by setting the initial state of each player to $\min\{S\}$. This kind of solution seems doubtful. Note that the notion of state of the game or that of a player captures the properties of his environment, and as such depends only partially (and in a nondeterministic way) on his decisions. It cannot thus be set by a player at the beginning of the game. Note however that in our setting this kind of assumption could make more sense than in [1]. In our framework $\min\{S\} = s_0$, so assuming that the algorithm is initialized by setting individual states of all the players to s_0 would mean that the game is started before any players join it, which could make sense in many practical applications. On the other hand, for $X_0 = \delta[s_0]$ (10) is trivially satisfied, as it reduces to $\delta[s_0] \preceq_{\mathrm{SD}} q(\cdot|s_0, \delta[s_0], a_0))$, which is true for any transition probability defined on S.

Remark 6 In our setting the players join the game at different times. This naturally implies that those joining at later stages of the game hardly need any adjustment to their initial strategies, because the global state of the game is already very close to $\mathcal{X}$ when they appear. Consequently, the expected rewards they receive over their lifetime are very close to equilibrium payoffs corresponding to the smallest equilibrium in the game.

3.5 Examples of Application of the Model

In the remainder of this section we will briefly present some natural applications of our framework. Some further ones could be possible, if the sets of states and actions were multidimensional or the rewards could be negative. Generalizing to these situations is left however for further research.

Research and development race In this game the players are firms choosing their technological profile. Let s be the level of technological development of firm's products and a, its investment in research. The transition times for a player are technological breakthroughs for his firm. It is obvious that these moments do not come at the same time for each of the players, so this corresponds well to our framework. Next, a 'death' of a player is naturally interpreted as his firm's bankruptcy. Finally, let r describe his profit minus investment. We assume that there is no $\widetilde{r}$. It is natural to assume strategic complementarities between rewards for different firms—a higher level of technological development of the entire industry results in a higher demand for high-tech products. Also a higher investment in research is required if industry is at a higher level of development. Finally, one can argue that a firm with a higher technological profile is less likely to get bankrupt.

Corruption game This is a variant of the game presented in [20]. The players here are civil servants who can be in three states: corrupt, honest, excluded from the society. The last state can be naturally seen as a (civil) death of a player—in this state he is not able to receive any rewards. A player's transitions happen when he has to decide on some project. These moments are naturally different for different players. His actions describe his willingness to change his state. Obviously, a player who wants to be bribed is much more likely to become corrupt. Also the possibility of becoming corrupt increases as the society becomes more depraved. In corrupt state a player's rewards are the highest and naturally increase as the society becomes more corrupt. Finally, the possibility of death for a player decreases as the society becomes more corrupt, because the control is less stringent. Thus, we can argue this is a game with strategic complementarities as well.

Interdependent security A similar model has appeared in [22]. Let us consider a large number of computers in a cluster. Each of them is trying to avoid system failure due to viruses. Let s describe individual computer's security level, while a its investment in security. The transition times for an individual are moments of malicious attacks against him. A 'death' of a player is the time of system failure. We can assume that $r \equiv 1$ if the system is OK and zero otherwise (a number of different 'health' levels with different rewards is also possible). Further, let $\widetilde{r}$ be an individual's investment in security. As one can immediately see, this model fails to satisfy our assumptions, because $\widetilde{r}$ is negative. We can however argue that a weaker version of our assumptions presented in Remark 1 can be satisfied without making the model unrealistic. Note, that this game is a natural example of games with strategic complementarities, as higher level of security for other computers results in a lower probability of infecting any of them. It is also natural to assume that attacks on different machines are not coordinated, so the moves of different players are asynchronous, like in our framework.

Charging control for plug-in electric vehicles This model is inspired by the one presented in [27]. Let us consider a large population of plug-in electric vehicles. Each of them needs to load its battery regularly, but tries to do it as cheap as possible. The problem is that the cost of energy may depend on the hour of the day—from the electricity producers' point of view it is best if all the vehicles charge their batteries at the same time during the night when the overall energy consumption is relatively low, so they can incur some additional cost on the car owners for doing differently. On the other hand, the vehicle whose battery is empty needs to be recharged immediately, and otherwise it will decrease its owner's profits from using it. In our model each player tries to maximize his profits from use of the car minus the charging costs over the lifetime of the vehicle. There are two possible actions: $a = 1$ (not to charge) and $a = 2$ (to charge) and a number of states denoting the battery charge levels (plus artificial state $s_0 < 0$ and action $a_0 = 0$ denoting the breakdown of the car). The transition times can be viewed as moments when the battery of a given vehicle can be charged, so we can assume λ is constant. The battery state at each of transition times decreases by one with some positive probability, decreases to s_0 with some smaller positive probability and remains constant with the remaining one unless the user decides to charge the battery—then it increases to the maximum battery level $s_{\max}$. The immediate reward is of the form $r(s, X, a, Z) = R\mathbb{1}\{s > 0\}$, where R is the reward from the exploitation of the vehicle, while $\widetilde{r}$ is defined as

$$\widetilde{r}(s, X, a, Z) = \mathbb{1}\{a = 2\}[p(s - s_{\max}) - c\mathbb{E}\left[(a - Z)^2\right]],$$

where p is the nominal energy price and c is the additional cost for deviating from the average policy of the population. Again, this model fails to satisfy assumptions (A1–A5) ($\widetilde{r}$ is nonpositive and it depends on Z), but it can be directly checked that for R big enough it satisfies all the assumptions of the model combining its two generalizations described in Remarks 1 and 2.

Remark 7 It is worth noting that the last model is one of many models considered in the engineering literature where the so-called crowd-seeking behavior is beneficiary for the players. Strategic complementarity between states or actions of the players seems a perfect mathematical description of this kind of situation. It turns out however that engineering applications of our model are limited for several reasons. The first one is that typically engineering models consider costs, not rewards, so the positivity assumption appearing in (A3) (very important, since we consider a total reward model) fails. The second one is that

we also assume that r and $\widetilde{r}$ are nondecreasing in s, which often is not satisfied. One should however note that this monotonicity assumption is crucial in proving that the aggregate utility of each player preserves the strategic complementarity structure, so we cannot easily get rid of it. Finally, the problems can be caused by the fact that we assume that the state space is a sublattice of $\mathbb{R}$ (which is important because for $S \subset \mathbb{R}^n$, $n \geq 2$, the set $\Delta(S)$ does not preserve the lattice structure).

4 Relation to Games with Finitely Many Players

In this section we provide a result which links the model with a continuum of players studied above with related models with finite numbers of players. In turn, this result provides an explanation to the use of the Kurtz dynamics (1) for the global state of the game. To begin with, we need to introduce the finite models we will discuss below. Let Γ denote the game with continuum of players defined in Sect. 2. Then Γ_n will denote its counterpart with n players played in exactly the same way as game Γ and such that:

(a) The global state of the game at time t is denoted by $X_t[n]$ and defined by the formula

$$X_t^s[n] = \#\{\alpha \in \{1, \ldots, n\} : s_t^\alpha = s\}.$$

Next, the normalized global state of the game at time t is denoted by $\overline{X}_t[n]$ and defined as

$$\overline{X}_t^s[n] = \frac{1}{n} X_t^s[n].$$

(b) All the functions defining the model are defined with respect to the normalized state, and so:

$$r[n](s_t, X_t[n], a_t) := r(s_t, \overline{X}_t[n], a_t), \qquad \widetilde{r}[n](s_t, X_t[n], a_t) := \widetilde{r}(s_t, \overline{X}_t[n], a_t),$$
$$q[n](\cdot|s_t, X_t[n], a_t) := q(\cdot|s_t, \overline{X}_t[n], a_t), \qquad \lambda[n](s_t, X_t[n]) := \lambda(s_t, \overline{X}_t[n]).$$

Next define the subset of strategies we shall concentrate on in this section.

$$F_c = \{f \in F : f(s, X) \text{ does not depend on } X\}.$$

The following result will link the game Γ with 'sufficiently close' games Γ_n.

Theorem 3 *Suppose assumption (A1) holds and take some $\Theta, \varepsilon > 0$. Then there exists an $N \in \mathbb{N}$ such that for any $n \geq N$ the expected reward of player α from playing policy $g \in F_c$ against $f \in F_c$ played by all the other players in the game Γ_n differs from his expected reward when he plays g against f in game Γ by at most ε.*

Proof First recall that $r, \widetilde{r}$ and λ are continuous on a compact domain, hence bounded. Let L be such that $|r(s, X, a)| \leq L$, $|\widetilde{r}(s, X, a)| \leq L$ and $\lambda(s, X) \leq L$ for any $(s, X, a) \in D$. In addition, note that λ is by assumption positive, so there also exists a $\underline{\lambda} > 0$ such that $\lambda(s, X) \geq \underline{\lambda}$ for any $(s, X, a) \in D$.

Next, note that under assumption (A1) absolute value of the sum of rewards received by (any given) player α from his kth change of state on

$$\left| \mathbb{E}\left[\sum_{i=k}^{i_e - 1} \left(\widetilde{r}(s^\alpha_{T_i^\alpha}, X_{T_i^\alpha}, a^\alpha_{T_i^\alpha}) + \int_{T_i^\alpha}^{T_{i+1}^\alpha} r(s^\alpha_{T_i^\alpha}, X_t, a^\alpha_{T_i^\alpha}) \, \mathrm{d}t \right) \right] \right|$$

can be bounded by

$$\left(L+\frac{L}{\underline{\lambda}}\right)\mathbb{P}[i_e > l]\sum_{i=0}^{\infty}(|S|-1)\mathbb{P}[i_e > i(|S|-1)+k|i_e > k]$$
$$\leq \left(L+\frac{L}{\underline{\lambda}}\right)(1-p_0)^{\left\lfloor \frac{k}{|S|-1}\right\rfloor}\sum_{i=0}^{\infty}(1-p_0)^i = \frac{L(\underline{\lambda}+1)}{\underline{\lambda}p_0}(1-p_0)^{\left\lfloor \frac{k}{|S|-1}\right\rfloor}. \quad (16)$$

It is then immediate that there exists a k_ε such that

$$\left|\mathbb{E}\left[\sum_{i=k_\varepsilon}^{i_e-1}\left(\widetilde{r}(s^\alpha_{T^\alpha_i}, X_{T^\alpha_i}, a^\alpha_{T^\alpha_i})+\int_{T^\alpha_i}^{T^\alpha_{i+1}} r(s^\alpha_{T^\alpha_i}, X_t, a^\alpha_{T^\alpha_i})\,dt\right)\right]\right| < \frac{\varepsilon}{6}.$$

The same bound will apply to every Γ_n.

Then, since τ^α_i is for any α stochastically dominated by an exponentially distributed random variable with intensity $\underline{\lambda}$, $\underline{\tau}^\alpha_i$, for any $T > 0$ we can conclude as follows:

$$\mathbb{P}\left[\sum_{i=0}^{k_\varepsilon}\tau^\alpha_i > T\right] \leq \mathbb{P}\left[\sum_{i=0}^{k_\varepsilon}\underline{\tau}^\alpha_i > T\right].$$

Since τ^α_i are for different i independent, we can assume the same about $\underline{\tau}^\alpha_i$. Then $\sum_{i=0}^{k_\varepsilon}\underline{\tau}^\alpha_i$ is Gamma-distributed with fixed parameters k_ε and $\underline{\lambda}$, thus the probability it is greater than T converges to 0 as T goes to infinity. Thus, there exists a $T_\varepsilon > 0$ such that

$$\mathbb{P}\left[\sum_{i=0}^{k_\varepsilon}\tau^\alpha_i > T_\varepsilon\right] < \frac{\underline{\lambda}p_0}{6L(\underline{\lambda}+1)}\varepsilon. \quad (17)$$

Consequently, by (16) the expected reward received by any player either in model Γ or any of models Γ_n from time T_ε on can be bigger than that received until the k_εth jump of his individual state by no more than

$$\frac{L(\underline{\lambda}+1)}{\underline{\lambda}p_0}\frac{\underline{\lambda}p_0}{6L(\underline{\lambda}+1)}\varepsilon = \frac{\varepsilon}{6},$$

which implies that the expected reward received by player α until time $\min\{T_\varepsilon, T^\alpha_{k_\varepsilon}\}$ in any of these models differs from the expected reward over his lifetime by at most $\frac{\varepsilon}{3}$.

Now note that since $f \in F_c$ and by Lipschitz continuity and boundedness of q and λ, all of the intensities $\sum_{s'\in S} X^{s'}\lambda(s', X)q(s|s', X, f(s', X))$, $-X^s\lambda(s, X)$ are Lipschitz-continuous and bounded functions of X, and so by the Kurtz theorem (see Theorem 5.3 in [32]) if all the players except α are using policy f,

$$\mathbb{P}[\sup_{0\leq t\leq T_\varepsilon}|\overline{X}_t[n]-X_t| \geq \delta] \leq De^{-nF(\delta)}$$

for some positive constant D and a function F satisfying $\lim_{\eta\searrow 0}\frac{F(\eta)}{\eta^2} \in (0,\infty)$. By this last property, the probability bounded above converges to zero as n goes to infinity at rate of e^{-n}, so for n large enough, say $n > N_\delta$,

$$\mathbb{P}[\sup_{0\leq t\leq T_\varepsilon}|\overline{X}_t[n]-X_t| \geq \delta] \leq \frac{\underline{\lambda}p_0}{3L(\underline{\lambda}+1)}\varepsilon \quad (18)$$

for any given $\delta > 0$.

Further, notice that $r(s, X, g(s, X))$ and $\widetilde{r}(s, X, g(s, X))$ are continuous on a compact domain, which by Heine's theorem implies that they are uniformly continuous, so we can find a $\delta > 0$ such that for any $X, X' \in \Delta(S)$,

$$|X - X'| < \delta \Longrightarrow \sup_{s \in S} |r(s, X, g(s, X)) - r(s, X', g(s, X'))| < \frac{\varepsilon}{4T_\varepsilon} \tag{19}$$

and

$$|X - X'| < \delta \Longrightarrow \sup_{s \in S} |\widetilde{r}(s, X, g(s, X)) - \widetilde{r}(s, X', g(s, X'))| < \frac{\varepsilon}{4}. \tag{20}$$

Then, let us fix a trajectory of $\overline{X}_t[n]$ and define

$$\begin{aligned} R(s,t) &= \mathbb{1}[t \le T_\varepsilon] \Bigg[\widetilde{r}(s, X_t, g(s, X_t)) \\ &\quad + \int_0^{T_\varepsilon} \int_0^{\min(T,(T_\varepsilon - t))} r(s, X_{t+u}, g(s, X_t))\, du\, e^{-\lambda(s,X_t)T} \lambda(s, X_t)\, \mathrm{d}T \Bigg], \\ Q(s', B|s,t) &= \int_B q(s'|s, X_t, g(s, X_t)) e^{-\lambda(s,X_t)T} \lambda(s, X_t)\, \mathrm{d}T, \end{aligned}$$

where B is any Borel set on $\mathbb{R}^+$, and analogously $R[n](s,t)$ and $Q[n](s', B|s,t)$ by replacing X_t in the above formulas with $\overline{X}_t[n]$ whenever $\sup_{t \in [0,T_\varepsilon]} |\overline{X}_t[n] - X_t| < \delta$ (and doing nothing otherwise). Note that measurability of the functions integrated in the above formulas is guaranteed by their continuity. Also in the cases of $R[n](s,t)$ and $Q[n](s', B|s,t)$, even though the functions integrated there are not continuous, their domain $S \times \mathbb{R}^+$ can be divided into a countable number of subsets of form $S \times [\underline{t}, \bar{t})$ where they are continuous. These sets are obviously Borel, which guarantees the measurability of the functions.

If we combine (19–20) with the definitions of $R[n]$ and $Q[n]$, we obtain that for n large enough

$$|R(s,t) - R[n](s,t)| < 2\left(\frac{\varepsilon}{4} + \frac{\varepsilon}{4T_\varepsilon} T_\varepsilon\right) = \varepsilon, \tag{21}$$

which means that $R[n]$ converges uniformly to R as n goes to infinity. Moreover, uniformly both in state (s,t) and the trajectory $\overline{X}_t[n]$. Similarly, we can show that the density appearing in the definition of $Q[n]$ converges uniformly to that appearing in the definition of Q. Note however that uniform convergence of densities together with uniform convergence and boundedness of rewards implies that

$$\begin{aligned} &R[n](s_0, 0) + \sum_{k=1}^{k_\varepsilon} \sum_{s \in S} \int_{\mathbb{R}^+} R[n](s,u)\, Q[n]^k(s, \mathrm{d}u|s_0, 0) \\ &\rightrightarrows R(s_0, 0) + \sum_{k=1}^{k_\varepsilon} \sum_{s \in S} \int_{\mathbb{R}^+} R(s,u)\, Q^k(s, \mathrm{d}u|s_0, 0), \end{aligned} \tag{22}$$

where the convergence is uniform with respect to both s_0 and $\overline{X}_t[n]$. Clearly, R and Q were constructed in such a way that the RHS of the above equation equals the expected reward received by player α until time $\min\{T_\varepsilon, T^\alpha_{k_\varepsilon}\}$ in model Γ. On the other hand, if we take the expected value of the LHS over all trajectories of $\overline{X}_t[n]$, (16) and (18) imply that for $n > N_0$ it will differ by at most

$$\frac{L(\underline{\lambda} + 1)}{\underline{\lambda} p_0} \frac{\underline{\lambda} p_0}{3L(\underline{\lambda} + 1)} \varepsilon = \frac{\varepsilon}{3} \tag{23}$$

from the expected reward received by player α until time $\min\{T_\varepsilon, T^\alpha_{k_\varepsilon}\}$ in model Γ_n.

Now, if we take n big enough, say bigger than N_1, the supremum over all s_0 and all $\overline{X}_t[n]$ of the two sides of (22) will differ by at most $\frac{\varepsilon}{3}$. This however, together with (23) and the fact that the expected reward until time $\min\{T_\varepsilon, T^\alpha_{k_\varepsilon}\}$ differs from that over lifetime of a player by at most $\frac{\varepsilon}{3}$, will imply that the reward received by player α in model Γ differs from that received in models Γ_n for[11] $n > \max\{N_\delta, N_1\}$ by at most

$$\frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon,$$

which ends the proof. □

Remark 8 The restriction of strategies to the set F_c may seem quite strong but, since at any fixed global state X a stationary strategy reduces to a mapping from S to A, it is enough to show the existence of approximate equilibria defined in a way similar to that equilibria are defined for mean-field game, which is obviously much weaker than how Nash equilibria are defined. Just this is done in a corollary below. On the other hand, note that the result presented in Theorem 3 can be easily generalized (in the sense that the proof will follow along the same lines as here) to Lipschitz-continuous randomized stationary strategies. However, as we limited our considerations to pure strategies in most of the paper, we have decided to present this result in this weaker form.

To formulate the next result, which will link equilibria of mean-field game Γ with approximate equilibria of games Γ_n, we need to introduce the following concept.

Definition 1 A stationary strategy f and a measure $\mu \in \Delta(S)$ are in *ε-weak equilibrium* in the semi-Markov n-person counterpart of mean-field game with total reward Γ_n, if μ is a stationary global state corresponding to f and for every other stationary strategy $g \in F$,

$$J(f, \overline{f}, \rho) \geq \overline{J}(g, \overline{f}, \rho) - \varepsilon,$$

where $\rho = q(\cdot|s_0, \mu, a_0)$ is the distribution of individual states of new-born players when the global state is μ.

The following result is an immediate consequence of Theorems 1 and 3.

Corollary 1 *Suppose that the total reward mean-field game Γ satisfies assumptions (A1–A6). Then for n big enough $(\underline{B}(\underline{X}), \underline{X})$ and $(\overline{B}(\overline{X}), \overline{X})$ are ε-weak equilibria in n-person counterparts of Γ, Γ_n.*

5 Conclusions

In our paper we presented a model of mean-field game where each of infinitely many players controls his own continuous time Markov chain of private states, but the global state follows an ordinary differential equation. We have made two main contributions here: the first one is the generalization of this type of games to a novel model where players, instead of maximizing some payoff accumulated over the entire game, maximize the reward obtained during their lifetime, which may be different for different players. Since any dead player can be replaced after some time by a newborn one, after some time stationary behavior of the system is obtained, which is then used to define a mean-field-type equilibrium. We have provided an

[11] Also with δ selected in a way guaranteing that the difference between the sides of (22) is below $\frac{\varepsilon}{3}$.

approximation result linking this new model with its n-player counterparts for n approaching infinity under some very mild assumptions.

The second main contribution of the paper is an equilibrium-existence result for mean-field game model discussed in this paper under some strategic complementarity conditions. These assumptions differ significantly from those discussed in the mean-field game literature, as no conditions based on convexity or strict monotonicity of the functions defining our model are required. Instead, properties implying that an increase in states of most of the players makes increase in any individual's state profitable and that an increase in one's state makes higher actions more profitable are assumed. This allows us to obtain the existence of equilibria in strategies with some monotonicity properties as well as the convergence of a myopic learning procedure. What is important, it turns out that many real-life applications of mean-field models satisfy our strategic complementarity assumptions. However, the applications of our contributions are limited especially due to two of them: positivity of reward functions and one-dimensional state space. It will be very interesting to see a generalization of our results getting rid of these two assumptions.

References

1. Adlakha S, Johari R (2013) Mean field equilibrium in dynamic games with strategic complementarities. Oper Res 61(4):971–989
2. Amir R (2002) Complementarity and diagonal dominance in discounted stochastic games. Ann Oper Res 114:39–56
3. Balbus Ł, Reffett K, Woźny Ł (2014) A constructive study of Markov equilibria in stochastic games with strategic complementarities. J Econ Theory 150:815–840
4. Bertsekas DP, Shreve SE (1978) Stochastic optimal control: the discrete time case. Academic Press, New York
5. Bergin J, Bernhardt D (1992) Anonymous sequential games with aggregate uncertainty. J Math Econ 21:543–562
6. Bergin J, Bernhardt D (1995) Anonymous sequential games: existence and characterization of equilibria. Econ Theor 5(3):461–489
7. Chakrabarti SK (2003) Pure strategy Markov equilibrium in stochastic games with a continuum of players. J Math Econ 39(7):693–724
8. Curtat L (1996) Markov equilibria in stochastic games with complementarities. Games Econ Behav 17:177–199
9. Ferreira E, Gomes DA (2014) On the convergence of finite state mean-field games through Γ-convergence. J Math Anal Appl 418(1):211–230
10. Gomes DA, Mohr J, Souza RR (2013) Continuous time finite state mean field games. Appl Math Optim 68:99–143
11. Gomes DA, Saúde J (2014) Mean field games models—a brief survey. Dyn Games Appl 4(2):110–154
12. Gomes D, Voskanyan V (2013) Extended deterministic mean-field games. Preprint http://arxiv.org/abs/1305.2600
13. Gomes D, Velho RM, Wolfram M-T (2014) Socio-economic applications of finite state mean field games. Philos Trans A. doi:10.1098/rsta.2013.0405
14. Guéant O (2015) Existence and uniqueness result for mean field games with congestion effect on graphs. Appl Math Optim 72(2):291–303
15. Horst U (2005) Stationary equilibria in discounted stochastic games with weakly interacting players. Games Econ Behav 51(1):83–108

16. Huang M, Caines PE, Malhamé RP (2007) Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized ε-nash equilibria. IEEE Trans Autom Control 52(9):1560–1571
17. Huang M, Caines PE, Malhamé RP (2007) The Nash certainty equivalence principle and McKean–Vlasov systems: an invariance principle and entry adaptation. In: 46th IEEE conference on decision and control, pp 121–123
18. Huang M, Caines PE, Malhamé RP (2007) An invariance principle in large population stochastic dynamic games. J Syst Sci Complex 20(2):162–172
19. Jovanovic B, Rosenthal RW (1988) Anonymous sequential games. J Math Econ 17:77–87
20. Kolokoltsov VN, Malafeyev OA (2015) Mean-field-game model of corruption. Dyn Games Appl. doi:10.1007/s13235-015-0175-x
21. Kolokoltsov VN, Yang W (2015) Inspection games in a mean field setting. Preprint arXiv:1507.08339
22. Kunreuther H, Heal G (2003) Interdependent security. J Risk Uncertain 26(2):231–249
23. Lasry J-M, Lions P-L (2006) Jeux à champ moyen. I. Le cas stationnaire. C R Math Acad Sci Paris 343(9):619–625
24. Lasry J-M, Lions P-L (2006) Jeux à champ moyen. II. Horizon fini et contrôle optimal. C R Math Acad Sci Paris 343(10):679–684
25. Lasry J-M, Lions P-L (2007) Large investor trading impacts on volatility. Ann Inst H Poincaré Anal Non Linéaire 24(2):311–323
26. Lasry J-M, Lions P-L (2007) Mean field games. Jpn J Math 2(1):229–260
27. Ma Z, Callaway DS, Hiskens IA (2013) Decentralized charging control of large populations of plug-in electric vehicles. IEEE Trans Control Syst Technol 21(1):67–78
28. Milgrom P, Roberts J (1994) Comparing equilibria. Am Econ Rev 84:441–459
29. Müller A, Scarsini M (2006) Stochastic order relations and lattices of probability measures. SIAM J Optim 16(4):1024–1043
30. Nowak AS (2007) On stochastic games in economics. Math Methods Oper Res 66(3):513–530
31. Schmeidler D (1973) Equilibrium points of nonatomic games. J Stat Phys 17:295–300
32. Schwartz A, Weiss A (1995) Large deviations for performance analysis. Chapman & Hall, London
33. Sleet C (2001) Markov perfect equilibria in industries with complementarities. Econ Theor 17(2):371–397
34. Tarski A (1955) A lattice-theoretical fixpoint theorem and its applications. Pac J Math 5:285–309
35. Tembine H, Le Boudec Y-I, El-Azouzi R, Altman E (2009) Mean field asymptotics of Markov decision evolutionary games and teams. In: Gamenets, international conference on game theory for networks, 13–15 May, Istanbul, Turkey
36. Topkis DM (1998) Supermodularity and complementarity. Princeton University Press, Princeton
37. Vives X (2009) Strategic complementarity in multi-stage games. Econ Theor 40(1):151–171
38. Wardrop JG (1952) Some theoretical aspects of road traffic research. Proc Inst Civ Eng 2:325–378
39. Więcek P, Altman E (2015) Stationary anonymous sequential games with undiscounted rewards. J Optim Theory Appl 166(2):686–710
40. Więcek P, Altman E, Ghosh A (2015) Mean-field game approach to admission control of an M/M/∞ queue with shared service cost. Dyn Games Appl. doi:10.1007/s13235-015-0168-9