Learning Pure Nash Equilibrium in Smart Charging Games

B. Sohet and Y. Hayel and O. Beaude and A. Jeandin

Abstract—Reinforcement Learning Algorithms (RLA) are useful machine learning tools to understand how decision makers react to signals. It is known that RLA converge towards the pure Nash Equilibria (NE) of finite congestion games and more generally, finite potential games. For finite congestion games, only separable cost functions are considered. However, non-separable costs, which depend on the choices of all players instead of only those choosing the same resource, may be relevant in some circumstances, like in smart charging games. In this paper, finite congestion games with non-separable costs are shown to have an ordinal potential function, leading to the existence of an action-dependent continuous potential function. The convergence of a synchronous RLA towards the pure NE is then extended to this more general class of congestion games. Finally, a smart charging game is designed for illustrating convergence of such learning algorithms.

Index Terms—Reinforcement learning, Finite congestion games, Non-separable costs, Smart charging

I. INTRODUCTION

Understanding how decision makers react to signals is the basement of complex systems analysis. In such systems where decision makers are selfish and the outcome of each individual depends on the decision of others (typically a game theoretical framework), it is important to compute the equilibrium situation (i.e., Nash Equilibrium, NE) but also to understand how each player will adapt her strategy through time and therefore learn to play her Nash strategy. Learning algorithms in game theoretical settings have been known for several years [1]. Many techniques are based on the well known best response principle like fictitious play [2], which assumes that players choose a best reply to the observed empirical distribution of past actions of the other players. The main drawback of such techniques is the computation of the best action, which can be computationally complex and also needs specific information (utility functions, set of actions, ...). Other types of algorithms, based on reinforcement mechanisms, have also been employed in game theoretical problems. These types of decentralized learning techniques, based on experiments as in repeated games, proved to be efficient in particular games such as congestion games and more generally potential games, which have convergence properties, as illustrated recently in machine learning community with different feedback information [3].

Congestion games are a specific type of non-cooperative games in which the cost of a commodity depends on the number of players choosing it [4]. Some costs functions are said non-separable, like the energy cost for Electric Vehicle (EV) users, when it relies on a smart dynamic pricing. This means that the cost perceived by an individual choosing one commodity or route, depends on the choices of all players and not only on the one choosing the same commodity. Non-separability property is not present in standard congestion games where cost is only due to congestion. Few papers deal with congestion games with non-separable cost functions because of the non-existence of a Beckmann function [5], except in particular cases [6]. Most of them focus on the Price of Anarchy, which is a measure of the performance induced by a decentralized system compared to centralized optimization. Learning techniques in congestion games are not very studied because it mainly considers non-atomic games, i.e. a population of players like an infinite number of decision makers. In [7], the authors adapt the reinforcement learning algorithm (RLA) described in [8] to atomic congestion games, i.e. with a finite number of decision makers. In their adaptation of the RLA, costs functions are separable. In this paper, this assumption is relaxed and non-separable costs are considered, such as energy costs in smart charging games. In [14], it is shown that an asynchronous RLA converges for ordinal potential games.

Smart charging EV is a particularly interesting environment to implement online learning algorithms to deal with different degrees of uncertainty and randomness of future knowledge. Most papers related to smart charging consider machine learning techniques from a centralized point of view [9]. Deep learning techniques are suggested in [10] to study EV charging navigation and to minimize the total travel time and the charging cost at charging stations, with also a centralized point of view. In [11], the NE of the constrained energy trading game among players with incomplete information is found using a RLA. This game however does not consider a smart charging context. In our paper, we adapt RLA techniques to a finite smart charging game and show convergence to a pure NE. The contributions of this paper are the following:

- Proof of potentiality of an atomic congestion game with non-separable cost functions;
- Proof of RLA convergence for this game;
- Application to a smart charging game and numerical example.

The paper is organized as follows: finite congestion games with non-separable costs are introduced in Section [H] and an action-dependent continuous potential is defined. This leads to convergence results of RLA in such games in Section [I]. This methodology is applied to a smart charging game described in Section [V] in which numerical illustrations show its convergence. Finally Section [V] concludes the paper.
II. Potential functions in finite congestion games with non-separable costs

In the atomic game $G$ considered in this work, there are $N$ players. Each player $i \in N \triangleq \{1, \ldots, N\}$ chooses her resource $r_i$ among the same set $R$. This set $R$ of players is made of $M$ resources, so that this game is finite. Note that the following study applies directly to the case where each player $i$ chooses a set $R_i \subset R$ of resources. In a classical congestion game [12], a player $i$ choosing resource $r_i$ would have to pay for a cost $c_i(n_i(r_i))$ which depends on the number of players having chosen the same resource, $n_i(r_i)$, defined as:

$$\forall a \in R, \quad n_a(r) \triangleq \# \{i \in N \mid r_i = a\},$$

with $r \triangleq (r_1, \ldots, r_N) = (r_i, r_{-i})$ the vector of actions and $r_{-i}$ the vector of actions of the players other than $i$. Here, the costs are non-separable [13], meaning that the cost $c_a$ of a resource $a$ does not only depend on $n_a$, but also on the number of players $n_b$ choosing other resources $b \in R \setminus \{a\}$. More precisely, we can prove the existence of a potential function for specific non-separable costs:

**Definition 1.** Linearly non-separable congestion costs $c_a$, for all resources $a \in R$ and vectors of actions $r \in R^N$, are defined as:

$$c_a(r) \equiv \alpha_a \lambda(L(r)),$$

with:

$$\begin{align*}
\alpha_a & \geq 0, \text{ constant}, \\
\lambda : & \mathbb{R}_+ \to \mathbb{R}_+, \text{ an increasing function}, \\
L(r) & \equiv \sum_{i=1}^N \alpha_{r_i} = \sum_{b \in R} \alpha_b n_b(r).
\end{align*}$$

Note that we chose the term “linearly” because of function $L(\cdot)$, which is a linear combination of the number of players $n_a$ choosing each resource $a$, while $\lambda$ may be non-linear.

**Remark 1.** Here, the game is supposed symmetric (between the players), meaning that the cost of a player $i$ only depends on her choice, and not on the player herself. However, note that the following study can be extended to a non-symmetric case where player $i$ gets the cost $c_{i,a}$, with $c_{i,a}(r_i, \lambda(\sum_{j=1}^N \alpha_{j,r_j}))$, which is the same as [4] when $\alpha_{i,a} = \alpha_a$ for all $i, a$.

In this particular context of atomic congestion game with non-separable cost functions, we are able to find potential properties for the game, which is a powerful tool for the study of NE in pure strategy and convergence of learning procedures [14].

**A. Potential function of pure strategies**

Following Definition [1] of linearly non-separable congestion costs $(c_a)_{a \in R}$, it is possible to extend the ordinal potential property of separable congestion games [15] to the non-separable game $G \triangleq (N, R^N, (c_a))$.

**Definition 2.** An ordinal potential for game $G$ is a function $P : R^N \to \mathbb{R}$ verifying $\forall i \in N, \forall \alpha_{-i} \in R^{N-1}, \forall a, b \in R$,

$$c_a(a, r_{-i}) < c_b(b, r_{-i}) \iff P(a, r_{-i}) < P(b, r_{-i}).$$

This definition follows the idea that an ordinal potential function follows the sign of the difference of cost for any player that changes her action unilaterally. Even if our game $G$ is not a standard congestion game due to the non-separability of costs functions, it is possible to show the existence of an ordinal potential function.

**Proposition 1.** The finite non-cooperative congestion game $G$ with non-separable costs defined in (2) has the following ordinal potential function:

$$\forall r \in R^N, \quad P(r) \triangleq \lambda(L(r)).$$

**Proof.** Let $i \in N$ be any player and $a, b \in R, r_{-i} \in R^{N-1}$ any actions.

Firstly, note that $L(a, r_{-i}) = L(b, r_{-i}) + \alpha_a - \alpha_b$, by definition. Then, as function $\lambda$ is increasing:

$$\lambda(L(a, r_{-i})) < \lambda(L(b, r_{-i})) \iff \alpha_a - \alpha_b < 0,$$

i.e. $P(a, r_{-i}) < P(b, r_{-i}) \iff \alpha_a < \alpha_b$, by definition of $P = \lambda \circ L$.

Secondly, function $C_r : \alpha \mapsto \alpha \times \lambda(\alpha + \sum_{j \neq i} \alpha_{r_j})$ is increasing on $\mathbb{R}_+$ as a product of positive increasing functions, meaning that:

$$\alpha_a \lambda(\alpha_a + \sum_{j \neq i} \alpha_{r_j}) < \alpha_b \lambda(\alpha_b + \sum_{j \neq i} \alpha_{r_j}) \iff \alpha_a < \alpha_b,$$

i.e. $c_a(a, r_{-i}) < c_b(b, r_{-i}) \iff \alpha_a < \alpha_b$.

The existence of a potential implies the existence of a pure NE in such non-cooperative games, which is defined here:

**Definition 3.** A pure strategy Nash Equilibrium (NE) of game $G$ is a vector of actions $r^* \in R^N$ which verifies:

$$\forall i \in N, \forall a \in R, \quad c_{r^*_i}(r^*_i, r_{-i}^*) \leq c_a(a, r_{-i}^*).$$

In other words, a NE is a strategy vector such that no player can reduce her cost by changing her strategy unilaterally. Note that the existence of pure NE is not a standard result, but it is true for games with an ordinal potential function. Indeed, as the sets of actions are compact, the minimum of the potential exists and corresponds to a pure NE of the game [16]. In this particular non-separable atomic congestion game, pure NE can be fully characterized. Let us define the set of resources $R^+ \triangleq \{a \in R \mid \alpha_a > \min_{a \in R} \langle \alpha_a \rangle\}$.

**Proposition 2.** The NE of $G$ are the $r^* \in R^N$ such that:

$$\forall a \in R^+, \quad n_a(r^*) = 0.$$
Then, $\alpha_a > \alpha_b$ by definition of $R^+$, and $c_{r_i^a}(r_i^a, r_{-i}^a) > c_b(b, r_{-i})$ using again function $C_{r^+}$.

This proposition shows that in games following Definition 1 pure NE correspond to situations where all players choose the resource $a$ with the lowest coefficient $\alpha_a$.

This parameter may not be known by players in advance, hence the need of learning algorithms by players in order to optimally adapt their actions. Such RLA are fully decentralized and are based on updates, for each player, of probability distributions over pure strategies (called mixed strategies). It is shown in [8] that there is a link between the potential function in pure strategies and the one in mixed strategies for common payoff games. This result is extended to more general games and is fundamental in order to prove the convergence of RLA to pure NE.

### B. Action-dependent continuous potential

Let $\pi_{i,a}$ denote the probability with which player $i$ chooses pure strategy $a \in R$, $\pi_i \in \Delta_i$, the mixed strategy vector of player $i$ in a simplex $\Delta_i$ of $\mathbb{R}^n$ and $\pi \in \Delta = \prod_i \Delta_i$ the mixed strategies of all players. The mixed strategy notation of player $i$ playing pure strategy $a$ is $\pi_i = e_a$, with $e_a$ the null vector except for the $a$-th component, equal to 1. The expected cost $c_{i,a}$ for player $i$ playing pure strategy $a$ is:

$$c_{i,a}(\pi) \triangleq \mathbb{E}_\pi(c_a | \pi_i = e_a) = \sum_{r_{-i}} (c_a(a, r_{-i}) \prod_{j \neq i} \pi_j, r_j),$$

(7)

with $c_a$ the linearly non-separable congestion cost of $r_i$.

Considering mixed strategies, the strategy sets are topological spaces and the expected cost functions given in (7) are continuously differentiable. Moreover in such continuous games, there may exist continuous potential functions, defined in [16]: the gradient of these functions correspond to the expected costs. This type of potential is widely considered in population games [17], as it serves as a Lyapunov function for strategies’ dynamics, or in games with non-atomic players [18]. In our particular setting, a generalization of these potential functions is needed, and defined as follows:

**Definition 4.** An action-dependent continuous potential is a $C^1$ function $F$ over mixed strategies such that, for all resources $a \in R$, there exists a constant $\gamma_a$ verifying:

$$\forall i, \quad \frac{\partial F}{\partial \pi_{i,a}}(\pi) = \gamma_a c_{i,a}(\pi).$$

(8)

Note that as expected cost are $C^1$ functions, such potentials $F$ are then $C^2$ functions. Continuous potential functions verify (8) with $\gamma_a = 1$ for all resources $a$. Our atomic game $G$ considering continuous strategy sets of mixed strategies has an action-dependent continuous potential function. In fact, this function is the conditional expectation of the ordinal potential function $F$ when players choose pure strategies according to the mixed strategy vector $\pi$.

**Proposition 3.** Atomic games $G$ with linearly non-separable congestion costs have the following action-dependent continuous potential function (associated to $\gamma_a = \frac{1}{\alpha_a}$) for all $a$:

$$F(\pi) \triangleq \mathbb{E}_\pi[P],$$

(9)

with $P$ the ordinal potential of $G$.

**Proof.** By linearity of the expected value $\mathbb{E}_\pi[P]$:

$$F(\pi) = \sum_{i,a} \pi_{i,a} \mathbb{E}_\pi[P | \pi_i = e_a]$$

$$= \sum_{i,a} \pi_{i,a} \frac{1}{\alpha_a} \mathbb{E}_\pi[c_a | \pi_i = e_a],$$

using $c_a(\cdot) = \alpha_a \lambda(L(\cdot)) = \alpha_a P(\cdot)$. Then, (8) is found by differentiating by $\pi_{i,a}$ with $\gamma_a = \frac{1}{\alpha_a}$ ($\forall i, a$).

The previous proposition generalizes the particular case studied in [8] for games with common payoff, while a similar result is obtained in [14] for continuous potential games. The following proposition gives a more precise result and shows that only games with particular cost functions admit an action-dependent continuous potential function.

**Proposition 4.** Finite games with action-dependent continuous potential function correspond to finite games with cost functions defined as:

$$\forall i, a, \quad c_{i,a}(r) \triangleq \beta_a \mu(r),$$

(10)

with $\beta_a$ any constant which depends on the action $a$, and $\mu$ any function of pure strategies (not necessarily increasing or linearly non-separable).

**Proof.** Let $F$ be a $C^2$ perturbed continuous potential function, associated to constants $\gamma_a$ ($\forall a$). Then, by Definition (8) of $F$ and according to Clairaut-Schwarz theorem (symmetry of second derivatives):

$$\forall i, j, a, b, \quad \gamma_a \frac{\partial^2 c_{i,a}}{\partial r_{j,b}} = \gamma_b \frac{\partial^2 c_{j,b}}{\partial r_{i,a}},$$

which, using (7), leads to (for all mixed strategies $\pi$):

$$\sum_{r_{-ij}} \left( \prod_{k \neq i,j} \pi_k, r_k \right) [\gamma_a c_{i,a} - \gamma_b c_{j,b}] (a, b, r_{-ij}) = 0.$$

For all pure strategies $r_{-ij} \in \mathbb{R}^{N-2}$, last equation considered with $\pi_k = e_{r_k}$ ($\forall k \neq i, j$) becomes:

$$\gamma_a c_{i,a}(a, b, r_{-ij}) = \gamma_b c_{j,b}(a, b, r_{-ij}).$$

Let $\mu = c_{i,a}$ and $\beta_b = \frac{\gamma_a}{\gamma_b}$ ($\forall b$). Then, (10) is true.

Inversely, suppose a game with cost functions verifying (10). Then, $\pi_{i,a} = \beta_a \mathbb{E}_\pi[\mu | \pi_i = e_a]$, as seen in (7).

Let $F(\pi) = \mathbb{E}_\pi[\mu]$. By linearity of the expected value, $F = \sum_{i,a} \pi_{i,a} \mathbb{E}_\pi[\mu | \pi_i = e_a]$. Therefore, (8) is verified, with $\gamma_a = \beta_a$ ($\forall a$).

This type of games is a generalization of common payoff games with action-dependent cost.

The property of having an action-dependent continuous potential leads to convergence of simple RLA for which only local information is accessible for each player (basically her own perceived cost) in order to update her mixed strategy vector and find her best action. In fact, in most cases, players are not even aware that they are involved in a game.
with other players and interact through their actions. That
is why the framework of online learning in which players
make repeated decisions with a priori unknown rules and
outcomes is suitable. In the next section, a simple RLA
is described, whose convergence has already been proven
when the game has a continuous potential function. We prove
that there is still convergence in a case of action-dependent
continuous potential and non-separable congestion game as
in our setting.

III. REINFORCEMENT LEARNING ALGORITHM

In our framework, players possess incomplete information:
their only knowledge is the observation of their cost after
taking an action. Note that best response algorithms can
also be applied to game \( G \), but players require additional
information (the exact formulation of their own cost function).
Here, game \( G \) will be repeated so that players learn
what their best strategy is. More precisely, every iteration \( n \)
is split into two phases. In the first phase, each player \( i \)
chooses an action \( r_i^{(n)} \) in accordance with her mixed strategy
vector \( \pi_i^{(n)} \). Thus, a vector of actions \( r^{(n)} \) is induced by
the decisions of all players, which in turn implies a cost for each
player based on the cost functions defined in [2]. Then, in
the second phase of the iteration, each player updates her
strategy probability vector based on the noisy cost. This
update mechanism is a reinforcement mechanism. This type
of RLA, used in stochastic games, is a linear reward-in-action
scheme [8]. Each player \( i \) updates her mixed strategy vector
\( \pi_i \) as follows (for any iteration \( n \)):

\[
\pi_i^{(n+1)} = \pi_i^{(n)} + \delta \times \left( 1 - \frac{c_{i,a}(r^{(n)})}{c_{\max}} \right) \times \left( e_a - \pi_i^{(n)} \right),
\]

with:
- \( 0 < \delta < 1 \) the learning parameter, fixed;
- \( a = r_i^{(n)} \) the action taken by player \( i \) at iteration \( n \);
- \( c_{\max} \triangleq \max_{i,a} c_{i,a}(r) \).

The basic idea of the updating rule expressed by equation \((11)\) is to ensure that actions prompting small or high
costs are promoted or not. This update scheme is decentral-
ized, and the global algorithm is fully distributed. This is
an important property in order to deploy it in large scale
complex systems, typically congestion games with a large
number of players. The global algorithm works as follows:

\[
\text{Algorithm 1: RLA with synchronous global updates}
\]

\begin{algorithm}
\begin{algorithmic}[1]
\State \textbf{Input:} \( \pi^{(0)}, n = 0 \)
\While{not converged}
  \For{all players \( i \)}
    \State Actions \( r^{(n)} \) according to mixed strategies \( \pi^{(n)} \);
    \State Cost \( c_{i,a}(r^{(n)}) \) at \( a = r_i^{(n)} \) given by \((2)\);
    \State Update mixed strategy of \( i \) with \((11)\);
  \EndFor
  \State \( n \leftarrow n + 1 \);
\EndWhile
\end{algorithmic}
\end{algorithm}

This algorithm converges for games having an action-
dependent continuous potential function. The updating mecha-
nism \((11)\) can be written as:

\[
\pi^{(n+1)} = \pi^{(n)} + \delta G \left( \pi^{(n)}, r^{(n)} \right),
\]

with \( G \) the updating function. Let us define function \( f(\pi) \triangleq \mathbb{E}_\pi[G] \) and for any \( \delta \) the function \( \Pi_\delta : \mathbb{R}_+ \rightarrow \mathbb{R} \) as
the piecewise-constant interpolation of sequence \( \left( \pi^{(n)} \right) \)
with the mixed strategies of all players. A direct application of The-
orem 3.1 in [8] demonstrates that our learning algorithm \([1]\)
converges weakly, as \( \delta \) tends to 0, to the solution \( \Pi \) of the
following Ordinary Differential Equation (ODE):

\[
\frac{dX}{dt} = f(X), \quad X(0) = \pi^{(0)}.
\]

Considering \( \Pi = (\pi_{i,a}) \) and \( f(\Pi) = (f_{i,a}) \), the ODE \((13)\)
can be written element-wise as:

\[
\forall i, a, \quad \frac{d\pi_{i,a}}{dt} = f_{i,a} = \pi_{i,a}(1 - \pi_{i,a})(1 - \frac{c_{i,a}}{c_{\max}}) + \sum_{b \neq a} \pi_{i,b}(-\pi_{i,a})(1 - \frac{c_{i,a}}{c_{\max}}) = -\pi_{i,a} \sum_{b \neq a} \pi_{b} \frac{(c_{i,a} - c_{i,b})}{c_{\max}}.
\]

Thus, the convergence points of our algorithm are related to
the solutions of a particular ODE, which must be character-
ized in order to prove the convergence of the RLA to NE
of game \( G \). In fact, next proposition proves that having an
action-dependent continuous potential implies convergence
to pure NE of our RLA described in Algorithm \([1]\).

**Proposition 5.** If a finite game \( G \) has an action-dependent
continuous potential then, for any initial non-pure strategies
\( \pi^{(0)} \), function \( \Pi \triangleq \lim_{\delta \rightarrow 0} \Pi_\delta \) converges to a pure NE.

**Proof.** This proof is inspired by the one of Theorem 3.3
of [8]. Here, the continuous potential \( F \) of game \( G \) is action-
dependant and associated with constants \( \gamma_a \):

\[
\frac{dF}{dt}(\Pi) = - \sum_{i,k} \sum_{l \neq k} \gamma_{a_k} \pi_{i,a_k} \pi_{i,a_l} \frac{(c_{i,a_k} - c_{i,a_l})^2}{c_{\max}} < 0,
\]

with \( \mathcal{R} = \{a_1, \ldots, a_M\} \) the set of resources. Then, \( t \mapsto F(\Pi(t)) \) is non-increasing. Therefore, the RLA always converges
to a pure NE.

Note that Algorithm 1 works with synchronous global
updates, meaning that all players update their strategy si-
multaneously at each iteration. An asynchronous version of
Algorithm 1 can be considered, where only one player
updates her strategy at each iteration. A player \( i \) is chosen
for iteration \( n \) with uniform probability \( p_i \triangleq \frac{1}{n} \). The
convergence of such an asynchronous algorithm towards a
pure \( \varepsilon \)-NE has been proven in [14] for games having an
ordinal potential function. This result can be applied directly

\footnote{For any \( \varepsilon \geq 0 \), a pure strategy vector \( r^* \) is an \( \varepsilon \)-Nash Equilibrium if \( \forall i, a, \quad c_{i,a}(r^*) \leq c_{i,a}(a, r^*_{-a}) + \varepsilon \).}
to our framework of finite congestion games with linearly non-separable costs, as an ordinal potential function exists (see Prop. [1]). In next Section, we illustrate the convergence of both synchronous and asynchronous RLA in a finite smart charging congestion game with linearly non-separable costs.

IV. A SMART CHARGING GAME

A. Description

In this section, the synchronous reinforcement learning algorithm and its asynchronous version are illustrated on a real-life example of a finite congestion game with linearly non-separable costs. The players are Electric Vehicle (EV) users who choose at which Charging Station (EVCS) they charge their EV battery (here, the EVCS are the resources). These EVCS are part of an electrical grid network shown in Fig. [1]. For simplicity, both the EVCS and the grid are supposed to be managed by the same Electric Network Operator (ENO). Note that the learning algorithm of previous section can be applied to more general grid topologies. At each electric bus \( r \), there is some (fixed) power demand \( L_r \) corresponding to other usages than EV charging. Choices of EV users only depend on the observed cost \( c_r \) of their charging operation at EVCS \( r \), proportional to the charging unit price \( p_r \). All users are supposed to have the same charging need \( \rho \) (in kWh) to simplify notations, but this assumption (symmetric game) can be readily relaxed, in line with Remark [1]. Note that the charging operation is supposed to last exactly one hour, at constant power (then equal to \( \rho \), in kWh). Thus, the charging cost for any EV user at EVCS \( r \) is \( c_r \equiv \rho \times p_r \).

In our framework, the ENO chooses charging unit prices \( p_r \) as incentives to reduce its global cost \( H \). Here, marginal cost pricing schemes are considered: \( p_r \equiv \partial H/\partial L_r \) with \( L_r \equiv L_r^0 + n_e(r) \rho \) the total power demand at EVCS \( r \). The ENO global cost \( H \) is defined as the cost of the electricity generation needed to fulfill the total power need \( \mathcal{L} \equiv \{L_r\}_{r \in \mathcal{R}} \) in the grid network considered. Typically, this cost is a quadratic function [19] of the (apparent) power \( S_0 \) at the head of the grid:

\[
H(\mathcal{L}) \equiv \eta S_0^2(\mathcal{L}),
\]

with \( \eta \) a scaling constant. This power can be obtained by running a Power Flow algorithm, more precisely the Bus Injection Model [20]. This algorithm solves the power balance equation at each bus (between the given power production/load \( S_{0,k} \) at bus \( k \) and power flows \( S_k \) from/to the bus):

\[
S_{0,k} = U_k \sum_{m \in X_k} Y_{k,m} S_m \quad (\equiv S_k),
\]

with \( U_k \) the complex voltage at bus \( k \), \( X_k \) the set of buses connected to bus \( k \) and \( Y_{k,m} \) the admittance of the line between buses \( k \) and \( m \).

The charging unit price \( p_r = \partial H/\partial L_r(\mathcal{L}) \) at EVCS \( r \) is not necessarily a function of a linear combination of the action vector \( r \) as in (2), due to the complexity of power flow equations (15). Therefore, the grid is reduced to only the transformer bus \( d \), as if all the power demand came from it, and the grid topology is replaced by constants \( \alpha_r \equiv (\tilde{\alpha}_r) \):

\[
\partial H/\partial L_r(\mathcal{L}) \simeq \partial H/\partial L_d(\tilde{\mathcal{L}}) = \alpha_r \partial H/\partial L_d(\tilde{\mathcal{L}}), \quad \text{with (16)}
\]

\[
\mathcal{L} \equiv \{L_a, L_b, L_c, 0\} \quad \text{and} \quad \tilde{\mathcal{L}} \equiv \{0, 0, 0, \sum_r \alpha_r L_r\}.
\]

The meaning of constant \( \alpha_r \) is that a power demand \( L_r \) at EVCS \( r \) is equivalent (i.e., yields equal marginal costs) to a power demand \( \alpha_r L_r \) at the transformer bus \( d \). Coefficients \( \alpha_r \) are obtained by solving the following square minimization:

\[
\alpha_r \equiv \arg\min_{\alpha > 0} \int_{t_{\min}}^{t_{\max}} \left[ \frac{\partial H}{\partial L_a}(L_a^0 + L_a, L_b^0, L_c^0, 0) \right] \alpha_a L_a^0 + \alpha_d L_d^0 \right] \d L_a + \int (b) + \int (c),
\]

where \( \int (b) + \int (c) \) means that the first integral in (17) is repeated for EVCS \( b \) and \( c \).

Then, charging unit prices \( p_r \equiv \alpha_r \partial H/\partial L_d(\tilde{\mathcal{L}}) \) lead to the following charging cost \( c_r \) for EV users choosing EVCS \( r \):

\[
c_r = \rho \times \tilde{\alpha}_r \partial H/\partial L_d(\tilde{\mathcal{L}}) \left( \sum_s \rho \tilde{\alpha}_s n_s(r) + \sum_s \tilde{\alpha}_s L_s^0 \right) \quad (18).
\]

This cost function verifies Definition [1] with \( \alpha_r \equiv \rho \tilde{\alpha}_r \) and \( \lambda \equiv \partial H/\partial L_d \), an increasing function of \( L(r) = \sum_s \rho \tilde{\alpha}_s n_s(r) \).

B. Numerical illustrations

The parameters of the smart charging game are set as follows: a typical French 20kV electrical network has around 1500 customers, with a standard 6kVA contract power. The total number of EV users is set to \( N = 1500 \) and the energy need is \( \rho = 3 \)kWh, half of the daily mean individual EV consumption in France[2]. Based on bus lines characteristics and the grid network topology considered in Fig. [1] for each EVCS \( r \), linear coefficients \( \tilde{\alpha}_r \) are obtained solving the square minimization problem (17) with \( \tilde{\alpha}_a = 1.12, \tilde{\alpha}_b = 1.07 \) and \( \tilde{\alpha}_c = 1.18 \). Note that EVCS \( a \) and \( c \) have a greater impact on ENO global cost \( H \) than EVCS \( b \) because they are further away from the transformer (see Fig. [1]). The ENO global cost \( H \) is adjusted with \( \eta = 5.10^{-3} \) (€/MVA²) so that charging unit prices \( p_r \) remain between 10 and 20c/€/kWh.

[2]Enquête Nationale Transports et Déplacements: https://utp.fr/system/files/Publications/UTP_NoteInfo103_Enseignements_ENTD2008.pdf (in French).
Fig. 2: Evolution of mixed strategies $\pi_i$ (of only 50 users) throughout iterations of Algorithm 1. Thicker lines represent the average (over all users) mixed strategy.

Regarding learning characteristics, the learning parameter is set to $\delta = 0.5$ and the initial mixed strategies are equally distributed among resources: $\pi_{i,r} = 1/3$ for all users $i$ and EVCS $r = a, b, c$. Similarly, for the asynchronous version of Algorithm 1, each player $i$ is chosen with a probability $p_i = 1/N$ for the update phase. Figures 2 and 3 show the evolution of mixed strategies $\pi_i$ (of only 50 users, for visibility) throughout iterations respectively for the synchronous Algorithm 1 and for the asynchronous version. As mentioned in Section II-A, although the pure NE of this game (all users choosing EVCS with lowest impact on grid costs) may seem trivial, EV users need hundreds of iterations to learn it (see Fig. 2), as they have no information on the grid topology. Considering thicker lines (average mixed strategy over all users), it can be seen that, while the synchronous algorithm converges in less than 500 iterations (Fig. 2), it takes more than thousand times as many iterations for the asynchronous one (Fig. 3). This is understandable since in the asynchronous version, only one player updates her strategy at each iteration, instead of all players like in the original Algorithm 1. This also explains why larger number of players lead to slower convergence for the asynchronous version, while it has no effect on the original Algorithm 1. Note that the evolution of mixed strategies may slightly vary from one execution to another (of either algorithm), due to actions randomly chosen from mixed strategies. Similarly, the number of iterations until convergence depends on the initial mixed strategies $\pi_i(0)$ and decreases with the learning parameter $\delta$.

V. CONCLUSIONS AND PERSPECTIVES

In this paper, a reinforcement learning algorithm (RLA) has been applied to obtain in a fully decentralized manner a pure Nash equilibrium (NE) for a finite congestion game with non-separable cost functions. Non-separability of cost functions makes it difficult to prove the existence of pure NE and the convergence property of RLA in such particular congestion games. When cost functions are separable, all these results come from the existence of an exact potential. However, assuming costs are linearly non-separable, we are able to prove the existence of an ordinal potential function, which can serve as a Lyapunov function for proving the convergence of simple RLA. Moreover, these results were applied to a smart charging game in which EV users choose selfishly a charging station depending on the smart charging price, which is based on a Power flow solution of the impact of EV charging on the grid. In a future work, a rationality coefficient for players will be added to the RLA.

REFERENCES

[1] D. Fudenberg and D. Levine, “The Theory of Learning in Games”, Series on Economic Learning and Social Evolution, vol. 2. MIT Press, Cambridge, MA, 1998.
[2] G. Brown, “Iterative solution of games by fictitious play”, Activity Analysis of Production and Allocation, vol. 13. John Wiley & Sons, Inc., New York, pp. 374–376, 1951.
[3] J. Cohen, A. Heliou and P. Mertikopoulos, “Learning with Bandit Feedback in Potential Games”. Proceedings of NIPS, 2017.
[4] N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, Algorithmic Game Theory, Cambridge University Press, 2007.
[5] M. J. Beckmann, C. B. McGuire, and C. B. Winsten, Studies in the Economics of Transportation, Yale University Press, 1956.
[6] B. Sohet, O. Beaude, Y. Hayel and A. Jeandin, “Coupled charging and driving incentives design for electric vehicles in urban networks,” unpublished, arXiv:2001.11758, 2020.
[7] D. Barth, J. Cohen, O. Bournez and O. Boussaton, “Distributed learning of equilibria in a routing game”, Parallel Processing Letters, vol. 19, num. 02, pp. 189-204, 2009.
[8] PS Sastry, VV Phansalkar and M. Thathachar, “Decentralized learning of Nash equilibria in multi-person stochastic games with incomplete information”, IEEE Trans. on systems, man, and cybernetics, vol. 24 (5), pp. 769-777, 1994.
[9] W. Tang, S. Bi and Y. J. Zhang, “Online Charging Scheduling Algorithms of Electric Vehicles in Smart Grid: An Overview”, IEEE Communications Magazine, vol. 54, no. 12, pp. 76-83, 2016.
[10] T. Qian, C. Shao, X. Wang and M. Shahidehpour, “Deep Reinforcement Learning for EV Charging Navigation by Coordinating Smart Grid and Intelligent Transportation System”, IEEE Trans. on Smart Grid, vol. 11, no. 2, pp. 1714-1723, March 2020.
[11] H. W. Wang, T. W. Huang, X. F. Liao, H. Abu-Rub and G. Chen, “Reinforcement learning for constrained energy games with incomplete information”, IEEE Trans. on Cybernetics, vol. 47(10) 2017.
[12] R. Rosenthal, “A Class of Games Possessing Pure-Strategy Nash Equilibria”, International Journal of Game Theory, vol. 2, 1973.
[13] C. Chau and K. Sim, “The price of anarchy for non-atomic congestion games with symmetric cost maps and elastic demands”, Operations Research Letters, vol. 31, 2003.
[14] O. Bournez and C. Johanne, “Learning equilibria in games by stochastic distributed algorithms”, Computer and information sciences III, Springer, London, 31-38, 2013.
[15] I. Milchtaich, “Congestion Games with Player-Specific Payoff Functions”, Games and Economic Behavior, vol. 13, pp. 111–124, 1996.
[16] D. Monderer and L. Shapley, “Potential Games”, Games and Economic Behavior, vol. 14, pp. 124-143, 1996.
[17] W.H. Sandholm, “Potential games with continuous player sets”, Journal of Economic Theory, 97(1):81-108, 2001.
[18] M. Cheung and R. Lhakar, “Nonatomic potential games: the continuous strategy case”, Games and Economic Behavior, vol. 108, 2018.
[19] A. H. Mohsenian-Rad, V. W. Wong, J. Jatskevich, R. Schober and A. Leon-Garcia, “Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid”, IEEE Trans. on Smart Grid, 1.3, 320-331, 2010.
[20] J. Hu, Optimization of power system operation, J.Wiley & Sons, 2015