Probably Approximately Correct Nash Equilibrium Learning

Filiberto Fele and Kostas Margellos

***Abstract*—We consider a multiagent noncooperative game with agents' objective functions being affected by uncertainty. Following a data driven paradigm, we represent uncertainty by means of scenarios and seek a robust Nash equilibrium solution. We treat the Nash equilibrium computation problem within the realm of probably approximately correct learning. Building upon recent developments in scenario-based optimization, we accompany the computed Nash equilibrium with *a priori* and *a posteriori* probabilistic robustness certificates, providing confidence that the computed equilibrium remains unaffected (in probabilistic terms) when a new uncertainty realization is encountered. For a wide class of games, we also show that the computation of the so called compression set—which is at the core of scenario-based optimization—can be directly obtained as a byproduct of the proposed methodology. We demonstrate the efficacy of our approach on an electric vehicle charging control problem.**

***Index Terms*—Electric vehicles (EVs), Nash equilibria, robust game theory, scenario approach, variational inequalities.**

I. Introduction

Game theory has attracted significant attention and has found numerous applications from smart grid [2]–[4], and electricity markets [5], [6], to communication networks [7] and regulatory compliance [8], [9]. Nash equilibrium (NE) computation has been an important concept to characterize no-regret solutions for noncooperative agents [10], in multiagent distributed and decentralized control architectures [11]–[15]. Stochastic considerations were included in noncooperative games for risk-averse [16]–[18], and expected value settings [19]–[21], by imposing certain assumptions on the probability of the uncertainty realizations. Alternatively, worst-case approaches relied on assumptions on the geometry of the uncertainty set [16], [22], [23].

We consider a multiagent NE seeking problem with uncertainty affecting agents' objective functions. Here, we follow a data driven methodology, where we represent uncertainty by scenarios that could either be extracted from historical data, or by means of some prediction model [24]. However, this poses a major challenge, since NE are inherently random as they depend on the extracted scenarios. Therefore, our objective is to investigate the sensitivity of the resulting NE to the uncertainty, in a probabilistic sense. More specifically, our contributions can be summarized as follows.

1) We treat the NE computation problem in a probably approximately correct (PAC) learning framework [25]–[27], and employ the so called scenario approach [28]. Building on [29] we first provide an *a posteriori* certificate on the probability that a NE remains unaltered upon a new realization of the uncertainty. We then rely on [30] and provide an *a priori* probabilistic certificate on the equilibrium sensitivity, under an additional nondegeneracy assumption (see Section II for a definition). The obtained results are distribution-free, and as such the probability distribution of the uncertainty could be unknown and the only requirement is sample availability.

2) Under the additional assumption that the game under consideration admits a unique NE, or for aggregative games with multiple equilibria but a unique aggregate solution, we show that a *compression set* (see Section II for a definition) can be directly computed by inspection of the solution returned by the proposed algorithm. This feature has significant computational advantages as it prevents the use of greedy mechanisms (see, e.g., [29]), which would require running up to numerical convergence multiple times (possibly as many as the number of samples) a NE seeking iterative algorithm.

The results presented in this article do not contemplate constraints coupling agents' strategies. The latter give rise to *generalized* NE problems; we refer the reader to [11], [14], [31], and [32] for details.

It should be noted that similar results have recently and independently appeared in our preliminary work [1] and in [33]. In particular, Corollary 9 is directly related to Corollary 2 in [33] (as well as Theorem 5 in [1]). However, Theorems 7 and 8 introduce a characterization that does not appear in [33], which relies on a probabilistic sensitivity notion [see (4)] that serves as the game theoretic counterpart of the so called probability of violation (or equivalently cost deterioration) that appears in e.g., [28], [34], and is also employed in [33] (see Corollary 9 for further details). It should be noted that for all our probabilistic statements, unlike [33], we do not require the NE to be unique, and allow for degenerate problem instances in all our *a posteriori* results. The latter circumvents the need for checking whether the underlying game is nondegenerate which is in general a difficult task. As a result, the resulting bound is more conservative with respect to the one of [33]; we can retrieve that tighter bound by imposing a nondegeneracy assumption, as discussed below Theorem 7. Note also that the exposition of [33] allows for agent dependent uncertain terms in agents' objective functions; here we use the same term for all agents, however, our main results are directly applicable to the case of different terms (see Remark 10).

In Section II we present the main results of the article. Section III contains the proof of the main results, while in Section IV we provide for a wide class of games a methodology to determine a compression set. Section V provides an electric-vehicle (EV) charging control case study. Section VI concludes the article and provides some directions for future work.

II. Scenario-Based Multiagent Game

A. Gaming Set-Up

Let the set $\mathcal{N} = \{1, \ldots, N\}$ designate a finite population of agents. The decision vector, henceforth referred to as strategy, of agent $i \in \mathcal{N}$ is denoted by $x_i \in \mathbb{R}^n$ and satisfies the individual constraint set $\mathcal{X}_i \subset \mathbb{R}^n$.

Manuscript received April 14, 2020; revised September 4, 2020; accepted October 4, 2020. Date of publication October 13, 2020; date of current version August 30, 2021. This work was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/P03277X/1. Recommended by Associate Editor R. P. Malhame. *(Corresponding author: Filiberto Fele.)*

The authors are with the Department of Engineering Science, University of Oxford, OX1 3PJ Oxford, U.K. (e-mail: filiberto.fele@eng.ox.ac.uk; kostas.margellos@eng.ox.ac.uk).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2020.3030754

We denote by $x = (x_i)_{i\in\mathcal{N}} \in \mathcal{X} \subset \mathbb{R}^{nN}$ the collection of all agents' strategies, where $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$.

Let θ be an uncertain vector taking values in a set Θ, endowed with a σ-algebra, and let $\mathbb{P}$ denote the associated probability measure. Fix any $M \in \mathbb{N}$, and let $\{\theta_1, \ldots, \theta_M\} \in \Theta^M$ be a finite collection of independently and identically distributed (i.i.d.) scenarios/realizations of the uncertain vector θ, that we refer to as an M-multisample. For given strategies of the remaining agents x_{-i}, each agent $i \in \mathcal{N}$ aims at minimizing with respect to x_i the function

$$J_i(x_i, x_{-i}) = f_i(x_i, x_{-i}) + \max_{m\in\{1,\ldots,M\}} g(x_i, x_{-i}, \theta_m) \quad (1)$$

where $f_i : \mathbb{R}^{nN} \to \mathbb{R}$ expresses a deterministic objective, different for each agent i but still dependent on the strategies of all agents, while $g : \mathbb{R}^{nN} \times \Theta \to \mathbb{R}$ encodes a common component in the agents' objective function that depends on the uncertain vector. Agents are interested in minimizing their local objective f_i and the worst-case (maximum) value g can take among a finite set of scenarios. The EV charging control problem of Section V provides a natural interpretation of such a set-up, where EVs are selfish entities each one with a possibly different utility function f_i; however, they could be participating in the same aggregation plan or belonging to a centrally managed fleet, thus giving rise to a common g. Here, the fact that g is influenced by uncertainty accounts for price volatility.

We consider a noncooperative game among the N agents, described by the tuple $\mathcal{G} = \langle \mathcal{N}, (\mathcal{X}_i)_{i\in\mathcal{N}}, (J_i)_{i\in\mathcal{N}}, \{\theta_m\}_{m=1}^M \rangle$, where $\mathcal{N}$ is the set of agents/players, $\mathcal{X}_i$, J_i are, respectively, the strategy set and the cost function for each agent $i \in \mathcal{N}$, and $\{\theta_1, \ldots, \theta_M\}$ is a finite collection of samples.

Definition 1 (Nash equilibrium): Let $\Omega \subseteq \mathcal{X}$ denote the set of Nash equilibria of $\mathcal{G}$, defined as

$$\Omega = \{x^* = (x_i^*)_{i\in\mathcal{N}} \in \mathcal{X} : \\ x_i^* \in \arg\min_{x_i\in\mathcal{X}_i} J_i(x_i, x_{-i}^*),\ \forall i \in \mathcal{N}\}. \quad (2)$$

Assumption 2: 1) For any $\theta \in \Theta$, and any $x_{-i} \in \mathcal{X}_{-i} = \Pi_{j\neq i\in\mathcal{N}}\mathcal{X}_j$, $f_i(\cdot, x_{-i}) + g(\cdot, x_{-i}, \theta)$ is convex and continuous differentiable, while the local constraint set $\mathcal{X}_i$ is nonempty, compact and convex for all $i \in \mathcal{N}$.

2) For any $\theta \in \Theta$, and for all $i \in \mathcal{N}$, the functions g and f_i are twice differentiable on an open convex set containing $\mathcal{X}$.

3) The pseudogradient $(\nabla_{x_i} f_i(x))_{i=1}^N$ is monotone with constant $\chi^f \in \mathbb{R}$, while $\nabla_x g(x, \theta)$ is monotone with constant $\chi^g \in \mathbb{R}$ for any fixed θ, i.e., for any $u, v \in \mathbb{R}^{nN}$, and $\theta \in \Theta$

$$(u - v)^\top((\nabla_{u_i} f_i(u))_{i=1}^N - (\nabla_{v_i} f_i(v))_{i=1}^N) \geq \chi^f \|u - v\|^2$$
$$(u - v)^\top(\nabla_u g(u, \theta) - \nabla_v g(v, \theta)) \geq \chi^g \|u - v\|^2 \quad (3)$$

and $\chi^f + \chi^g \geq 0$.

Notice that we only need that $f_i(\cdot, x_{-i}) + g(\cdot, x_{-i}, \theta)$ is convex for any fixed x_{-i}, without requiring that both $f_i(\cdot, x_{-i})$ and $g(\cdot, x_{-i}, \theta)$ are simultaneously convex. Such monotonicity requirements have been also employed in [35] (see (4) therein), and do not require both χ^f, χ^g to be non-negative, but only $\chi^f + \chi^g \geq 0$.

B. Problem Statement

As every NE $x^* \in \Omega$ is a random vector due to its dependency on the M-multisample, a question that naturally arises is how sensitive a NE is against a new realization of the uncertainty. More formally, let Ω be the NE set of the game with M samples. Consider a new extraction $\theta \in \Theta$, and let $\mathcal{G}^+ = \langle \mathcal{N}, (\mathcal{X}_i)_{i\in\mathcal{N}}, (J_i)_{i\in\mathcal{N}}, \{\theta_m\}_{m=1}^M \cup \{\theta\} \rangle$ be a game defined over the $M + 1$ scenarios $\{\theta_1, \ldots, \theta_M, \theta\}$; denote by Ω^+ the set of the associated NE. Then, for all $x^* \in \Omega$, let

$$V(x^*) = \mathbb{P}\{\theta \in \Theta : x^* \notin \Omega^+\} \quad (4)$$

denote the probability that a NE of $\mathcal{G}$ does not remain a NE of $\mathcal{G}^+$, i.e., of the game characterized by the extraction of an additional sample. Note that $V(x^*)$ is in turn a random variable, as its argument depends on the multisample $\{\theta_1, \ldots, \theta_M\}$.

Within the realm of a PAC learning framework, with a given confidence/probability with respect to the product measure $\mathbb{P}^M$ (as the samples are extracted in an i.i.d. fashion), *we aim at quantifying* $V(x^*)$. To achieve such a characterization we provide some basic definitions. Let $\Phi : \Theta^M \to \Omega$ be a *single-valued* mapping from the set of M-multisamples to the set of equilibria of $\mathcal{G}$.

Remark 3: The game $\mathcal{G}$, the set of NE Ω, the mapping Φ (as well as of other associated quantities introduced in the sequel) depend on M via the M-multisample employed. Therefore, they are parameterized by M, giving rise to a family of games, NE sets and mappings. To ease notation we do not show this dependency explicitly. Also, the dimension of the domain of Φ is to be intended in accordance with M.

Definition 4 (Support sample [30]): Fix any i.i.d. M-multisample $(\theta_1, \ldots, \theta_M) \in \Theta^M$, and let $x^* = \Phi(\theta_1, \ldots, \theta_M)$ be a NE of $\mathcal{G}$. Let $x^\circ = \Phi(\theta_1, \ldots, \theta_{s-1}, \theta_{s+1}, \ldots, \theta_M)$ be the solution obtained by discarding the sample θ_s. We call the latter a *support sample* if $x^\circ \neq x^*$.

Definition 5 (Compression set — adapted from [29]): Fix any i.i.d. M-multisample $(\theta_1, \ldots, \theta_M) \in \Theta^M$, and let $x^* = \Phi(\theta_1, \ldots, \theta_M)$ be a NE of $\mathcal{G}$. Consider any subset $\mathcal{C} \subseteq \{\theta_1, \ldots, \theta_M\}$ and let $x^\circ = \Phi(\mathcal{C})$. We call $\mathcal{C}$ a compression set if $x^\circ = x^*$.

The properties of the compression set have been studied in detail in [29], where it is referred to as support subsample. Here we adopt the term compression set as in [25] and [27] to avoid confusion with Definition 4.

Let $\mathfrak{C}(\theta_1, \ldots, \theta_M)$ be the collection of all compression sets associated with $\{\theta_1, \ldots, \theta_M\}$. For a given compression set $\mathcal{C} \in \mathfrak{C}(\theta_1, \ldots, \theta_M)$ we refer to the compression cardinality as $d^* = |\mathcal{C}|$. Note d^* depends on $\mathcal{C}$; we do not show this dependence explicitly to ease notation. Also note $\mathfrak{C}$, hence d^*, is itself a random variable as it depends on the M-multisample.

Definition 6 (Nondegeneracy — adapted from [36]): For any $M \in \mathbb{N}$, with $\mathbb{P}^M$-probability equal to 1, the NE $x^* = \Phi(\theta_1, \ldots, \theta_M)$ coincides with the NE returned by Φ when the latter takes as argument *only* the support samples. The corresponding game is then said to be nondegenerate; otherwise it is called degenerate.

It follows that for nondegenerate problems the support samples form a compression set with $\mathbb{P}^M$-probability one. For degenerate problems the notions in Definitions 4 and 5 do not necessarily coincide (if only the support samples are used as argument to Φ, the returned solution might be different from x^*). In the latter case the support samples form a strict subset of any compression set in $\mathfrak{C}$ (see [34] and [36] for more details).

C. Main Results

Under Assumption 2, it is shown in [1, Section IV] that a single-valued mapping $\Phi : \Theta^M \to \Omega$ indeed exists, and can be computed in a decentralized manner, thus ensuring that $x^* = \Phi(\theta_1, \ldots, \theta_M)$ in Theorems 7 and 8 below is well defined.

1) A Posteriori Certificate: We provide an *a posteriori* quantification of an upper bound for $V(x^*)$.

Theorem 7: Consider Assumption 2. Fix $\beta \in (0,1)$ and let $\varepsilon : \{0,\ldots,M\} \to [0,1]$ be a function satisfying

$$\varepsilon(M) = 1$$
$$\text{and} \sum_{k=0}^{M-1} \binom{M}{k} (1-\varepsilon(k))^{M-k} = \beta. \tag{5}$$

Let $x^* = \Phi(\theta_1,\ldots,\theta_M)$. Consider any compression set and denote by $d^* \leq M$ its cardinality. We then have that

$$\mathbb{P}^M\{(\theta_1,\ldots,\theta_M) \in \Theta^M : V(x^*) \leq \varepsilon(d^*)\} \geq 1-\beta. \tag{6}$$

Theorem 7 shows that with confidence at least $1-\beta$ the probability that the NE $x^* \in \Omega$, computed on the basis of the randomly extracted samples $(\theta_1,\ldots,\theta_M) \in \Theta^M$, does not remain an equilibrium of the game $\mathcal{G}^+$ when an additional sample $\theta \in \Theta$ is considered, is at most $\varepsilon(d^*)$. Note that (6) captures the generalization properties of x^*, where $1-\beta$ accounts for the "probably" and $\varepsilon(d^*)$ for the "approximately correct" term used within a PAC learning framework.

The structure of $\varepsilon(\cdot)$ is determined in accordance with [29]. Its value depends on d^*, which in turn depends on $\{\theta_1,\ldots,\theta_M\}$ thus giving rise to the *a posteriori* nature of the result. Hence, the level of conservatism of the obtained certificate depends on d^*; the smaller the cardinality of the computed compression set, the tighter the bound (see Section IV for a detailed elaboration on the computation of d^*).

In the case of a nondegenerate game, the bound could be significantly improved by means of the *wait-and-judge* analysis of [36]: specifically, by Theorem 2 in [36], we can replace the expression for $\varepsilon(\cdot)$ in (5) with $\varepsilon(k) = 1 - t(k)$, where $t(k)$ is shown to be the unique solution in (0,1) of

$$\frac{\beta}{M+1} \sum_{m=k}^{M} \binom{m}{k} t^{m-k} - \binom{M}{k} t^{M-k} = 0. \tag{7}$$

We note that in [33, Cor. 2] a similar bound is derived on the related quantity of cost deterioration (see V_c in Corollary 9). We wish to emphasize, however, that nondegeneracy is a condition in general difficult to verify even in convex optimization settings, a challenge that becomes more prominent in games.

2) A Priori Certificate: We now provide an *a priori* quantification of an upper-bound of $V(x^*)$.

Theorem 8: Consider Assumption 2, and further assume that the game is nondegenerate as in Definition 6. Fix $\beta \in (0,1)$ and consider $\varepsilon : \{0,\ldots,M\} \to [0,1]$ satisfying (5). Let $x^* = \Phi(\theta_1,\ldots,\theta_M)$. We then have that

$$\mathbb{P}^M\{(\theta_1,\ldots,\theta_M) \in \Theta^M : V(x^*) \leq \varepsilon((n+1)N)\} \geq 1-\beta. \tag{8}$$

Although similar in form to Theorem 7, the bound on $V(x^*)$ provided by Theorem 8 is *a priori* and relies on the developments in [30] and [34]. In this way, $\varepsilon(\cdot)$ is evaluated on the sample-independent quantity $(n+1)N$, expressing the dimension nN of the agents' decision space plus N additional variables, explained by the epigraphic reformulation introduced in the proof of Theorem 8. If we further assume that for all $i \in \mathcal{N}$, for every fixed $x_{-i} \in \mathcal{X}_{-i}$ and $\theta \in \Theta$, both $f_i(\cdot, x_{-i})$ and $g(\cdot, x_{-i}, \theta)$ are convex, we would only need one epigraphic variable, hence the argument of $\varepsilon(\cdot)$ could be replaced by $nN+1$ (see Section III-C).

Since we strengthen here the assumptions of Theorem 8 by imposing a nondegeneracy condition, (5) could be directly replaced by the tighter expression in (7). We wish to emphasize that, even if the nondegeneracy assumption holds, it may still be preferable to calculate the cardinality d^* in an *a posteriori* fashion, as in certain problems the latter might be significantly lower compared to $(n+1)N$.

Corollary 9: Let $x^* = \Phi(\theta_1,\ldots,\theta_M)$ and consider

$$V_c(x^*) = \mathbb{P}\{\theta \in \Theta : g(x^*,\theta) > \max_{m \in \{1,\ldots,M\}} g(x^*,\theta_m)\}. \tag{9}$$

Under the assumptions of Theorems 7 and 8, respectively, (6) and (8) hold with $V_c(x^*)$ in place of $V(x^*)$.

Corollary 9 shows that with given confidence the probability that $g(x^*,\theta)$, and hence also each agent's objective function, deteriorates when a new realization of the uncertainty is encountered can be bounded as in Theorems 7 and 8, respectively. This statement is established within the proofs of Theorems 7 and 8.

Remark 10: The results of Theorems 7 and 8 remain valid in the case where the uncertain part of the objective function is different for each agent, i.e., if g is replaced by g_i, $i \in \mathcal{N}$ (such a setup is considered in [33]). We keep our presentation with a common g since for this case we are able to construct Φ in a decentralized manner; we refer the reader to [1] for implementation details (see also [7], [37], and [38]). We point out that a decentralized implementation of Φ for problems where the uncertain part of the objective function is different for each agent encompasses additional challenges (see [35, Rem. 1]) and is outside the scope of our article.

III. Proofs of *A Posteriori* and *A Priori* Certificates

A. Game Reformulation and Variational Inequalities

NE are commonly characterized as solutions to a variational inequality (VI) [39]. However, in $\mathcal{G}$ the presence of the $\max$ operator renders agents' objective functions (1) nondifferentiable. To circumvent the computation of sub-gradients and exploit the wide range of algorithms available to solve VIs in the differentiable case, we define the augmented game $\widehat{\mathcal{G}}$ between $N+1$ agents [35]. In $\widehat{\mathcal{G}}$ each player $i \in \mathcal{N}$, given x_{-i} and $y = (y_m)_{m=1}^M$, computes

$$x_i \in \arg\min_{\nu_i \in \mathcal{X}_i} f_i(\nu_i, x_{-i}) + \underbrace{\sum_{m=1}^{M} y_m g(\nu_i, x_{-i}, \theta_m)}_{\hat{g}(\nu_i, x_{-i}, y)} \tag{10}$$

where $\hat{g}(x,y)$ follows from the equivalence that for any x

$$\max_{m \in \{1,\ldots,M\}} g(x,\theta_m) = \max_{y \in \Delta} \sum_{m=1}^{M} y_m g(x,\theta_m) \tag{11}$$

where

$$\Delta = \left\{ y \in \mathbb{R}^M : y \geq 0, \sum_{m=1}^{M} y_m = 1 \right\} \tag{12}$$

is a continuous set that forms a simplex in $\mathbb{R}^M$ [40, Lemma 6.2.1], [35]; $y_m, m = 1,\ldots,M$, are continuous variables. Notice that their number increases with the number of samples M. The additional agent, given x, will act instead as a maximizing player for the uncertain component of J_i, $i \in \mathcal{N}$, i.e.,

$$y \in \arg\max_{\nu \in \Delta} \hat{g}(x,\nu). \tag{13}$$

We can now link the NE of the augmented game $\widehat{\mathcal{G}}$ to a VI. Following [7], we consider the optimization problem

$$(x^*,y^*) = \arg\min_{(x^*,y^*) \in \mathcal{X} \times \Delta} \frac{1}{2} \|(x^*,y^*)\|_2^2$$
$$\text{subject to } ((x,y) - (x^*,y^*))^\top F(x^*,y^*) \geq 0,$$
$$\forall (x,y) \in \mathcal{X} \times \Delta. \tag{14}$$

The constraint in (14) is a VI, where $F(x,y) : \mathcal{X} \times \Delta \to \mathbb{R}^{(nN+M)}$ is the pseudogradient [39, §1.4.1]

$$F(x,y) = \begin{bmatrix} (\nabla_{x_i} f_i(x) + \nabla_{x_i} \hat{g}(x,y))_{i\in\mathcal{N}} \\ -(\nabla_{y_m} \hat{g}(x,y))_{m=1}^M \end{bmatrix} \tag{15}$$

representing the first-order optimality conditions for the $N+1$ individual problems described by (10) and (13). Notice the slight abuse of notation in (14), where by (x^*, y^*) we denote both the decision vector and the resulting optimal solution.

Any algorithm that returns the optimal solution of (14) serves as a construction of Φ. The motivation of selecting the minimum norm NE in (14) stems from our requirement (see Section II-B) that $\Phi : \Theta^M \to \Omega$ is single-valued; any strictly convex objective function could be used instead (see [7, Th. 21]). Such a tie-break rule is needed even if only one NE is returned by the given algorithm, to prevent the case where different initial conditions produce different NE.

The following proposition establishes a link between the optimal solution of (14) (i.e., in fact this result holds for any feasible solution of the VI) and the set of equilibria of the original game; its proof follows from [39, Prop. 1.4.2] and [35, Th. 1], and can be found in [1, Prop. 7].

Proposition 11: Under Assumption 2, (14) is feasible. If (x^*, y^*) is the solution to (14), then x^* is a NE of $\mathcal{G}$.

Since F is monotone but not strongly monotone (see, e.g., [7] for a definition), a proximal decentralized algorithm based on [7, Algorithm 4] is employed in [1]. A direct consequence of the algorithm adopted in [1] is that x^* in (14) satisfies the following fixed point equation for $\tau > 0$ "big" enough (see [7, Lemma 20] for a lower bound on τ)

$$\begin{aligned} (x_i^*, \gamma_i^*)_{i\in\mathcal{N}} = &\underset{\{x_i\in\mathcal{X}_i, \gamma_i\in\mathbb{R}\}_{i\in\mathcal{N}}}{\arg\min} \sum_{i\in\mathcal{N}} \left(\gamma_i + \tau\|x_i - x_i^*\|_2^2\right) \\ &\text{subject to } f_i(x_i, x_{-i}^*) + g(x_i, x_{-i}^*, \theta_m) \le \gamma_i, \\ &\qquad \forall i \in \mathcal{N}, \forall m \in \{1, \cdots, M\} \end{aligned} \tag{16}$$

where γ_i, $i \in \mathcal{N}$, are epigraphic variables, and we have equality as the set of minimizers is a singleton due to the presence of the regularization term $\tau\|x_i - x_i^*\|_2^2$. Notice that at the NE the regularization term vanishes.

Note that Φ is single-valued, hence the returned solution is independent of the initial condition of the algorithm used. However, in the proof of Theorem 8 it becomes insightful to make this dependency explicit. Thus, for the analysis of Section III-C we will introduce the notation Φ_{x_0}, with $x_0 \in \mathcal{X}$ playing the role of the initial condition. Notice that y implicitly depends on the M-multisample, which is already an argument of Φ, hence we only include x_0 as a subscript. It follows from the fixed point equation (16) that $x^* = \Phi_{x^*}(\theta_1, \ldots, \theta_m)$.

B. Proof of Theorem 7

Fix $M \in \mathbb{N}$. Consider $(\theta_1, \ldots, \theta_M) \in \Theta^M$, and let $d^* \le M$ be the cardinality of any compression set of $\{\theta_1, \ldots, \theta_M\}$ (recall that it depends on the realization of the M-multisample). Let $x^* = \Phi(\theta_1, \ldots, \theta_M) \in \Omega$, and $\gamma^* = \max_{m\in\{1,\ldots,M\}} g(x^*, \theta_m)$. For any $\theta \in \Theta$, let

$$H_\theta = \{(x, \gamma) : g(x, \theta) \le \gamma\}. \tag{17}$$

Fix $\beta \in (0,1)$ and consider $\varepsilon(\cdot)$ defined as in (5). Under Assumption 2, Φ is single-valued as discussed in Section II-C. By [29, Th. 1] we then have that

$$\begin{aligned} &\mathbb{P}^M\{(\theta_1, \ldots, \theta_M) \in \Theta^M : \\ &\qquad \mathbb{P}\{\theta \in \Theta : (x^*, \gamma^*) \notin H_\theta\} \le \varepsilon(d^*)\} \ge 1 - \beta \end{aligned} \tag{18}$$

if $(x^*, \gamma^*) \in H_{\theta_m}$, $\forall m \in \{1, \ldots, M\}$ (consistency condition in [27]). To show this, notice that for each $i \in \mathcal{N}$, by the NE definition (Definition 1), (x_i^*, γ^*) will belong to the set of minimizers of the following epigraphic reformulation of (2)

$$\begin{aligned} (x_i^*, \gamma^*) \in &\underset{x_i\in\mathcal{X}_i, \gamma\in\mathbb{R}}{\arg\min} \; f_i(x_i, x_{-i}^*) + \gamma \\ &\text{subject to } g(x_i, x_{-i}^*, \theta_m) \le \gamma, \; \forall m \in \{1, \cdots, M\}. \end{aligned} \tag{19}$$

By (19) it follows that the consistency condition is satisfied, thus establishing (18). Note that for the result of [29] to be invoked, the aforementioned program is not required to be convex, hence the fact that for each $i \in \mathcal{N}$, for any $\theta \in \Theta$, only $f_i(\cdot, x_{-i}^*) + g(\cdot, x_{-i}^*, \theta)$ is assumed to be convex by Assumptions 2, is sufficient.

By the definition of γ^* and H_θ, (18) implies that with confidence at least $1 - \beta$

$$\mathbb{P}\{\theta \in \Theta : g(x^*, \theta) > \max_{m\in\{1,\ldots,M\}} g(x^*, \theta_m)\} \le \varepsilon(d^*) \tag{20}$$

thus establishing the statement related to (6) in Corollary 9.

We now proceed to demonstrate the claim in (6). Recall that, by (14) and (15), we can obtain $x^* \in \Omega$ as solution of the following optimization program:

$$\begin{aligned} &\min_{(x^*, y^*)\in\mathcal{X}\times\Delta} \frac{1}{2}\|(x^*, y^*)\|^2 \\ &\text{subject to} \\ &\sum_{i\in\mathcal{N}} (x_i - x_i^*)^\top \nabla_{x_i}(f_i(x^*) + \hat{g}(x^*, y^*)) \\ &- \sum_{m=1}^M (y_m - y_m^*) \nabla_{y_m} \hat{g}(x^*, y^*) \ge 0, \; \forall x \in \mathcal{X}, \; y \in \Delta \end{aligned} \tag{21}$$

where (x^*, y^*) is a NE of $\widehat{\mathcal{G}}$. By definition of $\hat{g}$ in (10), and since $\nabla_{y_m}(\sum_{m=1}^M y_m^* g(x^*, \theta_m)) = g(x^*, \theta_m)$, the constraint in (21) can be equivalently written as

$$\begin{aligned} &\sum_{i\in\mathcal{N}} (x_i - x_i^*)^\top \nabla_{x_i}(f_i(x^*) + \sum_{m=1}^M y_m^* g(x^*, \theta_m)) \\ &\quad + \sum_{m=1}^M y_m^* g(x^*, \theta_m) - \max_{y\in\Delta} \sum_{m=1}^M y_m g(x^*, \theta_m) \end{aligned} \tag{22}$$

where the presence of the maximum is due to the fact that (21) holds for any $y \in \Delta$. By (11) this is in turn equivalent to

$$\begin{aligned} &\sum_{i\in\mathcal{N}} (x_i - x_i^*)^\top \nabla_{x_i}(f_i(x^*) + \sum_{m=1}^M y_m^* g(x^*, \theta_m)) \\ &\quad + \sum_{m=1}^M y_m^* g(x^*, \theta_m) - \max_{m\in\{1,\ldots,M\}} g(x^*, \theta_m) \ge 0. \end{aligned} \tag{23}$$

For a given $\theta \in \Theta$, recall from Section II-C the definition of the game $\mathcal{G}^+$ associated with the samples $\{\theta_1, \ldots, \theta_M\} \cup \{\theta\}$, and the associated set of NE Ω^+. Moreover, let $\widehat{\mathcal{G}}^+$ denote the associated augmented game. Analogously to (23), any solution $(x^+, y^+) \in \mathcal{X} \times \Delta^+$ (where Δ^+ is the simplex in $\mathbb{R}^{M+1}$) of the augmented game $\widehat{\mathcal{G}}^+$ will satisfy the following VI:

$$\sum_{i\in\mathcal{N}}(x_i - x_i^+)^\top \nabla_{x_i}\left(f_i(x^+) + \sum_{m=1}^{M} y_m^+ g(x^+,\theta_m)\right)$$
$$+ \sum_{i\in\mathcal{N}}(x_i - x_i^+)^\top \nabla_{x_i}\left(y_{M+1}^+ g(x^+,\theta)\right)$$
$$+ \sum_{m=1}^{M} y_m^+ g(x^+,\theta_m) + y_{M+1}^+ g(x^+,\theta)$$
$$- \max\left\{\max_{m\in\{1,\ldots,M\}} g(x^+,\theta_m),\, g(x^+,\theta)\right\} \geq 0. \quad (24)$$

Note the analogy between (23) and (24), with the additional terms corresponding to the new sample θ (y_{M+1} is the additional decision variable corresponding to the new sample).

We are interested in quantifying the probability of $x^* \in \Omega^+$. To this end, notice that if $g(x^*,\theta) \leq \gamma^*$, then $x^+ = x^*$ and $y^+ = (y^{*\top}, 0)^\top$ constitute a feasible pair for (24). This is due to the fact that under this choice $y_{M+1}^+ = 0$ and

$$\max\left\{\max_{m\in\{1,\ldots,M\}} g(x^*,\theta_m),\, g(x^*,\theta)\right\}$$
$$= \max\{\gamma^*,\, g(x^*,\theta)\} = \gamma^* = \max_{m\in\{1,\ldots,M\}} g(x^*,\theta_m) \quad (25)$$

thus (24) reduces to (23). Applying Proposition 11 to $\mathcal{G}^+$ and $\widehat{\mathcal{G}}^+$, we have that if (x^+, y^+) satisfies (24) (i.e., it is a NE of the augmented game $\widehat{\mathcal{G}}^+$) then $x^+ \in \Omega^+$. Therefore, $x^* \in \Omega^+$ whenever $g(x^*,\theta) \leq \gamma^*$, or in other words

$$\mathbb{P}\{\theta \in \Theta : (x^*,\gamma^*) \in H_\theta\} = \mathbb{P}\{\theta \in \Theta : g(x^*,\theta) \leq \gamma^*\}$$
$$\leq \mathbb{P}\{\theta \in \Theta : x^* \in \Omega^+\}. \quad (26)$$

By (18) and (26), (6) follows, thus concluding the proof. ■

C. Proof of Theorem 8

Let $\mathcal{C}_0 \subseteq \{\theta_1,\ldots,\theta_M\}$ be the minimal cardinality compression set for the minimum norm NE x^* of (14); note that under the nondegeneracy assumption it will be unique and will coincide with the set of support samples. Following the discussion at the end of Section III-A, denote by Φ_{x_0} an algorithm that returns x^*, where we make explicit the dependence on the initial condition $x_0 \in \mathcal{X}$. As Φ is single-valued, by the definition of a compression set, we have that $x^* = \Phi_{x_0}(\theta_1,\ldots,\theta_M) = \Phi_{x_0}(\mathcal{C}_0)$, for all $x_0 \in \mathcal{X}$.

Consider now the fixed point characterization of x^* in (16), for which we have that $x^* = \Phi_{x^*}(\theta_1,\ldots,\theta_M)$. Also note that we have introduced one epigraphic variable per agent $i \in \mathcal{N}$; we will invoke in the sequel the fact that (16) is convex due to Assumption 2 . However, if we further assume that for all $i \in \mathcal{N}$, for every fixed $x_{-i} \in \mathcal{X}_{-i}$ and $\theta \in \Theta$, the individual functions $f_i(\cdot, x_{-i})$ and $g(\cdot, x_{-i}, \theta)$ are convex, we would only need one epigraphic variable, as we could perform an epigraphic reformulation only for g [this would give rise to the constraint in (19)], which is common to all agents.

Let $\mathcal{C}$ denote a minimal cardinality compression set for x^* in (16). We claim that $\mathcal{C}_0 \subseteq \mathcal{C}$. To show this, assume for the sake of contradiction that there exists $k \in \{1,\ldots,M\}$ such that $\theta_k \in \mathcal{C}_0$ but $\theta_k \notin \mathcal{C}$. Consider the set $\{\theta_1,\ldots,\theta_M\} \setminus \{\theta_k\} \supseteq \mathcal{C}$, and notice that this has to be a compression set for x^* in (16) as it is a superset of $\mathcal{C}$. By Definition 5, this implies that $x^* = \Phi_{x^*}(\{\theta_1,\ldots,\theta_M\} \setminus \{\theta_k\})$ (recall that the solution of (16) is given by $x^* = \Phi_{x^*}(\theta_1,\ldots,\theta_M)$). However, $\theta_k \in \mathcal{C}_0$, which due to the imposed nondegeneracy assumption implies that it belongs to the set of support samples (see Definition 4) for x^*, i.e., if removed then the solution alters. Hence, $x^* \neq \Phi_{x^*}(\{\theta_1,\ldots,\theta_M\} \setminus \{\theta_k\})$, thus establishing a contradiction, showing that $\mathcal{C}_0 \subseteq \mathcal{C}$ (hence $|\mathcal{C}_0| \leq |\mathcal{C}|$).

By Assumption 2, (16) is a convex scenario program, and admits a unique solution due to the fact that the objective function in (16) is strictly convex. Therefore, by [30], [34], we have that any minimal cardinality compression set $\mathcal{C}$ has cardinality upper-bounded by $(n+1)N$, i.e., the number of decision variables in (16). Therefore, $|\mathcal{C}_0| \leq |\mathcal{C}| \leq (n+1)N$. As a result, $|\mathcal{C}_0|$ can be upper-bounded by the *a priori* known quantity $(n+1)N$. As Theorem 7 holds for any compression cardinality $d^* \geq |\mathcal{C}_0|$, we can apply it with $d^* = (n+1)N$. Hence, Theorem 8 as well as the statement related to (8) in Corollary 9 directly follow, concluding the proof. ■

IV. Computation of the Compression Set Cardinality

The result of Theorem 7 relies on the computation of the compression cardinality d^*. In [29, §II] a greedy procedure is outlined to estimate (an upper bound to) the minimal compression cardinality. However, there are two associated drawbacks: first, the computational cost is generally high, as the algorithm $\Phi(\cdot)$ employed to determine a NE should be evaluated at least M times, and this may involve an asymptotic scheme; second, in practice, limited numerical accuracy makes the greedy procedure amenable to numerical errors.

To alleviate these, we provide a computationally efficient way to determine a compression set, and hence d^*, by direct inspection of the NE. To achieve this, we impose certain NE uniqueness requirements. However, it should be noted that for the wide class of aggregative games, the additional structure required in the proposition below implies only uniqueness of an aggregate strategy, and multiple equilibria may exist.

Proposition 12: Consider Assumption 2. Further assume that for all $M \in \mathbb{N}$, either

1) $\mathcal{G}$ admits a unique NE;
2) or, g depends on the aggregate strategy[1] $\sigma(x) : x \mapsto \sum_{i\in\mathcal{N}} x_i$, and $\mathcal{G}$ admits a unique NE aggregate $\sigma(x)$.

Then, $\mathcal{Y}^* \triangleq \{m \in \{1,\ldots,M\} : y_m^* > 0\}$ includes the indices of a compression set, i.e., $x^* = \Phi(\theta_1,\ldots,\theta_M) = \Phi(\{\theta_m\}_{m\in\mathcal{Y}^*})$.

Proof: Part 1: Uniqueness of NE. Fix $(\theta_1,\ldots,\theta_M) \in \Theta^M$ and notice that it forms a (trivial) compression set for x^*. Let (x^*, y^*) be a solution of $\widehat{\mathcal{G}}$, where $y^* = (y_m^*)_{m=1}^M$.

To prove that $x^* = \Phi(\{\theta_m\}_{m\in\mathcal{Y}^*})$ it suffices to show that the solution returned by Φ remains unaltered after removing all samples from $\{\theta_1,\ldots,\theta_M\}$ whose associated component of y^* is zero. To this end, suppose that at least one such sample exists: without loss of generality, assume $y_M^* = 0$ (i.e., that sample has index M). We will first show that $\{\theta_1,\ldots,\theta_{M-1}\}$ is a compression set, i.e., $x^* = \Phi(\theta_1,\ldots,\theta_{M-1})$. Let $\mathcal{G}^- = \langle \mathcal{N}, (\mathcal{X}_i)_{i\in\mathcal{N}}, (J_i)_{i\in\mathcal{N}}, \{\theta_j\}_{j=1}^{M-1}\rangle$ be the game with samples $\{\theta_1,\ldots,\theta_{M-1}\}$. Moreover, let $\widehat{\mathcal{G}}^-$ denote the associated augmented game, and Δ^- the simplex in $\mathbb{R}^{M-1}$. Since (x^*, y^*) is an NE of $\widehat{\mathcal{G}}$, it will satisfy the VI in (23). At the same time, every solution $(x^-, y^-) \in \mathcal{X} \times \Delta^-$ of the augmented game $\widehat{\mathcal{G}}^-$ satisfies the following VI:

$$\sum_{i\in\mathcal{N}}(x_i - x_i^-)^\top \nabla_{x_i}\left(f_i(x^-) + \sum_{m=1}^{M-1} y_m^- g(x^-,\theta_m)\right)$$
$$+ \sum_{m=1}^{M-1} y_m^- g(x^-,\theta_m) - \max_{m\in\{1,\ldots,M-1\}} g(x^-,\theta_m) \geq 0. \quad (27)$$

[1] With a slight abuse of notation, in the second part of the proposition it is to be understood that for all $i \in \mathcal{N}$ and for any given x_{-i}, Assumption 2 refers to the function $f_i(\cdot, x_{-i}) + g(\sigma(\cdot, x_{-i}), \theta)$.

Set $x^- = x^*$ and $y^- = (y_m^*)_{m=1}^{M-1}$. Under this choice (x^-, y^-) satisfies (27), as

$$\max_{m\in\{1,\ldots,M-1\}} g(x^*, \theta_m) \leq \max_{m\in\{1,\ldots,M\}} g(x^*, \theta_m). \tag{28}$$

Equivalently, $(x^*, (y_m^*)_{m=1}^{M-1})$ is an NE for $\widehat{\mathcal{G}}^-$, and by applying Proposition 11 to $\mathcal{G}^-$ and $\widehat{\mathcal{G}}^-$ we have that x^* is an NE for $\mathcal{G}^-$. However, due to the uniqueness assumption, x^* has to be the only NE of $\mathcal{G}^-$, showing that $x^* = \Phi(\theta_1, \ldots, \theta_{M-1})$.

Following the same procedure, removing one by one all samples for which the associated elements of y^* are zero, shows that $x^* = \Phi(\theta_1, \ldots, \theta_M) = \Phi(\{\theta_m\}_{m\in\mathcal{Y}^*})$, thus concluding the proof of the first part.

Part 2: Uniqueness of NE aggregate. The proof follows the same arguments as in Part 1 with the following modifications. The derivation until (28) remains unaltered, showing that $(x^*, (y_m^*)_{m=1}^{M-1})$ is a NE of $\widehat{\mathcal{G}}^-$. To prove that $x^* = \Phi(\theta_1, \ldots, \theta_{M-1})$ it suffices to show that $(x^*, (y^*)_{m=1}^{M-1})$ is the minimum norm NE of $\widehat{\mathcal{G}}^-$. We thus assume for the sake of contradiction that $(\hat{x}, \hat{y}) \in \mathcal{X} \times \Delta^-$ is the NE of $\widehat{\mathcal{G}}^-$ that achieves the minimum norm, i.e., $\|(\hat{x}, \hat{y})\|^2 < \|(x^*, (y^*)_{m=1}^{M-1})\|^2$. We distinguish two cases:

Case 1: $g(\sigma(\hat{x}), \theta_M) \leq \max_{m\in\{1,\ldots,M-1\}} g(\sigma(\hat{x}), \theta_m)$. Under this condition observe that $(\hat{x}, (\hat{y}^\top, 0)^\top)$ satisfies the VI in (23) for the game with M samples. However, as (x^*, y^*) is the minimum norm equilibrium for that game, we have that

$$\|(x^*, y^*)\|^2 \leq \|(\hat{x}, (\hat{y}^\top, 0)^\top)\|^2. \tag{29}$$

Recalling that $y_M^* = 0$

$$\begin{aligned}\|(x^*, (y^*)_{m=1}^{M-1})\|^2 &= \|(x^*, y^*)\|^2 \\ &\leq \|(\hat{x}, (\hat{y}^\top, 0)^\top)\|^2 = \|(\hat{x}, \hat{y})\|^2\end{aligned} \tag{30}$$

thus establishing a contradiction. We can then show that $x^* = \Phi(\theta_1, \ldots, \theta_M) = \Phi(\{\theta_m\}_{m\in\mathcal{Y}^*})$ as in the last paragraph of Part 1.

Case 2: $g(\sigma(\hat{x}), \theta_M) > \max_{m\in\{1,\ldots,M-1\}} g(\sigma(\hat{x}), \theta_m)$. We will show that, under our assumptions, this case cannot occur. By the uniqueness assumption we have that $\sigma(\hat{x}) = \sigma(x^*)$ for any equilibrium $\hat{x} \neq x^*$ (the NE is not necessarily unique, but all equilibria have the same aggregate). We then have

$$\begin{aligned}g(\sigma(x^*), \theta_M) &= g(\sigma(\hat{x}), \theta_M) \\ &> \max_{m\in\{1,\ldots,M-1\}} g(\sigma(\hat{x}), \theta_m) \\ &\geq g(\sigma(\hat{x}), \theta_m) = g(\sigma(x^*), \theta_m)\end{aligned} \tag{31}$$

for any $m \in 1, \ldots, M-1$. Since (31) holds for any m

$$g(\sigma(x^*), \theta_M) > \max_{m\in\{1,\ldots,M-1\}} g(\sigma(x^*), \theta_m). \tag{32}$$

Consider now (19). By direct computation of the Karush-Kuhn-Tucker (KKT) optimality conditions [40, §6.2.1] of (19) and (13), respectively, it can be verified that the decision variable $y \in \Delta$ introduced in (10)–(13) is a *shadow price* for the constraint (19). Then, by the complementary slackness condition

$$y_m^*\left(g(\sigma(x^*), \theta_m) - \gamma^*\right) = 0, \ \forall m \in \{1, \ldots, M\}. \tag{33}$$

Since $y_M = 0$ implies $g(\sigma(x^*), \theta_M) \leq \gamma^*$ we obtain

$$\max_{m\in\{1,\ldots,M\}} g(\sigma(x^*), \theta_m) = \max_{m\in\{1,\ldots,M-1\}} g(\sigma(x^*), \theta_m). \tag{34}$$

From (34) it follows that:

$$\begin{aligned}&\max_{m\in\{1,\ldots,M-1\}} g(\sigma(x^*), \theta_m) \\ &\quad = \max_{m\in\{1,\ldots,M\}} g(\sigma(x^*), \theta_m) \geq g(\sigma(x^*), \theta_M).\end{aligned} \tag{35}$$

This establishes a contradiction with (32), and concludes the proof. ■

Based on the shadow price interpretation of y notice that if $g(x^*, \theta_m) < \gamma^*$ (inactive constraint) then $y_m^* = 0$. Note that samples with $y_m^* = 0$ can be removed without altering x^* due to the imposed uniqueness requirements; otherwise, the feasibility region of the VI in (27) may enlarge, thus resulting in a different minimum norm NE. Moreover, it should be noted that Proposition 12 does not provide guarantees that a minimal cardinality compression set is determined; this can be obtained by the greedy algorithm of [29].

V. Case Study: EV Charging

We consider a stylised EV charging control game. Let $\{1, \ldots, N\}$ index a finite population of EV vehicles/agents. We denote by $x_i \in \mathbb{R}^n$ the demand each EV seeks to determine over n time slots, where for simplicity slots are taken to be of 1 h. Vehicles' strategy is in response to a pricing signal, which in turn depends on demand of all agents. We consider price to be an affine function of the aggregate strategy $\sigma(x) : x \mapsto \sum_{i\in\mathcal{N}} x_i$, but other choices are also supported by our analysis. However, price is subject to uncertainty, e.g., externalities acting on the energy spot market, encoded by $\theta \in \Theta$, which we model by means of scenarios. In particular, each scenario is a realization of prices along the considered n-slot interval. Note that these scenarios are i.i.d., however, each of them is a finite horizon path, whose entries can be correlated. Each agent $i = 1, \ldots, N$ aims at minimizing

$$\begin{aligned}&f_i(x_i, x_{-i}) + \max_{m\in\{1,\ldots,M\}} g(x_i, x_{-i}, \theta_m) \\ &= x_i^\top(A_0\sigma(x) + b_0) + \frac{1}{N}\max_{m\in\{1,\ldots,M\}} \sigma(x)^\top(A_m\sigma(x) + b_m)\end{aligned}$$

where $A_m \in \mathbb{R}^{n\times n}$, for $m = 0, 1, \ldots, M$, are diagonal matrices, and $b_m \in \mathbb{R}^n$. Moreover, we assume the charging operations are subject to

$$\begin{aligned}\mathcal{X}_i = \{x_i \in \mathbb{R}^n :\ &\mathbf{1}^\top x_i \geq E_i,\ 0 \leq x_{ij} \leq P_i, \\ &\forall j = 1, \ldots, n\}\end{aligned} \tag{36}$$

where $E_i, P_i \in \mathbb{R}$ denote the desired final state of charge (SoC) and the maximum power deliverable by the charger, respectively.

We analyse the results of randomly generated cases, differing in the parameters characterizing the EV constraints $\mathcal{X}_i$, selected from a uniform random distribution: specifically, $P_i \in [6, 15]$ kW, and E_i is chosen to be feasible in the specified time interval (∼0–35 kWh per 12 h interval). The entries of $\{A_m, b_m\}_{m=1}^M$ are i.i.d. extracted from a lognormal distribution for A_m, and a uniform distribution for b_m. The nominal electricity price, i.e., the diagonal entries $\{a_t\}_{t=1}^n$ of the matrix A_0, have been derived by rescaling a winter weekday demand profile in the UK [41], whereas $b_0 = 0$.

To determine a NE we use the decentralized algorithm based on regularization outlined in [1, Sec. IV]. Note that at each algorithm iteration agents need to broadcast their strategies to a common authority/aggregator, and receive updates about the aggregate agents' strategy. Even though this example fits in the class of aggregative games, it does not necessarily meet the uniqueness requirement of Proposition 12. However, we have empirically observed that the main conclusion of the proposition still holds. Informally, this is due to the fact that for any feasible instance the constraint $\mathbf{1}^\top x_i \geq E_i$ will always be binding at the optimum, and as result $\mathbf{1}^\top\sigma(x)$ will be constant for any NE x (see transparent plane in Fig. 1).

To validate the *a posteriori* result of Theorem 7, Table I shows the average robustness performance of several solutions (with $N = 20$, $n = 24$) obtained from different sets of $M = 500$ samples, grouped according to the *a posteriori* observed compression cardinality d^*; we have set $\beta = 10^{-6}$. The violation rate $V(x^*)$ of each solution is

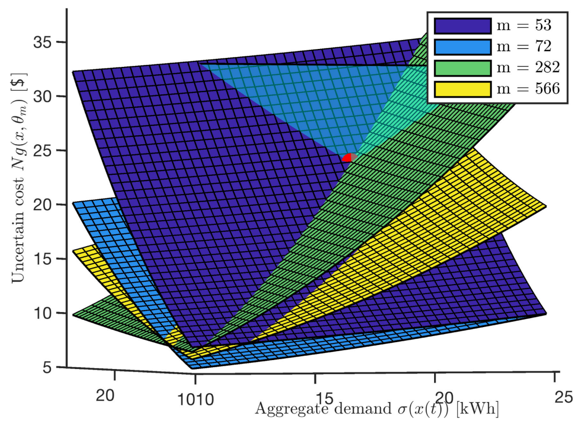

Fig. 1. Case with $N = 20$, $n = 2$: Uncertain cost component $Ng(x^*, \theta_m)$ for a subset $m \in \{53, 72, 282, 566\}$ of the $M = 1000$ samples used for the derivation of the NE x^*. In this case $d^* = 2$, with $\{\theta_{53}, \theta_{282}\}$ supporting the solution together with the SoC constraint which is binding in this case. This is the transparent plane, representing all aggregate strategies fulfilling $\sigma(x)|_{t=1} + \sigma(x)|_{t=2} = \sum_{i\in\mathcal{N}} E_i$ over the considered time interval. $\sigma(x^*)$, visible in red, lies at the intersection of the three surfaces.

TABLE I
EMPIRICAL VALIDATION OF THE A POSTERIORI RESULT OF THEOREM 7

d^* [-]	4	6	7	9
Empirical $V(x^*)$ [%]	0.98	1.09	1.26	1.33
$\varepsilon(d^*)$ - Thm 7 [%]	8.06	9.76	10.55	12.06
$\varepsilon(d^*)$ - bound (7) [%]	5.30	6.11	6.49	7.22

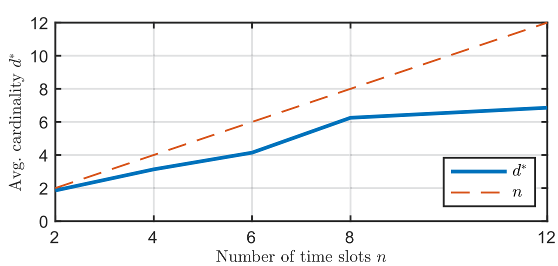

Fig. 2. Empirical validation of the *a priori* result of Theorem 8. The plot shows the average compression set cardinality d^* observed over 50 trials, corresponding to different randomly generated cases corresponding to different values of n and M. In all cases, d^* is bounded by n, and hence also by the theoretical bound $(n + 1)N$.

empirically computed using 10^6 newly extracted samples (according to the same aforementioned distributions) and counting the fraction of them that result in a change of the computed NE. Consistently with [29], we note that the observed value of d^* is indicative of the confidence level on the equilibrium robustness. The experimental results are compared with the theoretical bound provided by Theorem 7 (third row). For nondegenerate problems, the conservatism of the latter can be reduced by employing the tighter expression for $\varepsilon(\cdot)$ reported in (7), leading to the fourth row of Table I. However, note that in general it is difficult to verify whether a given problem is nondegenerate, thus preventing the use of (7).

A visual representation of the compression set concept is given in Fig. 1. The plot depicts the curves expressing the uncertain cost term $Ng(x^*, \theta_m)$ associated to a subset $m \in \{53, 72, 282, 566\}$ of the $M = 1000$ samples used for the derivation of the NE x^*. Values are plotted as a function of the aggregate demand $(\sigma(x)|_{t=1}, \sigma(x)|_{t=2})$ on an interval around $\sigma(x^*)$. In this case the compression cardinality is $d^* = 2$, with $\{\theta_{53}, \theta_{282}\}$ supporting the solution together with the constraint on the state of charge which is binding in this case (transparent plane). Note that in this instance the constraints on the power rate P_i are not active, and omitted from the plot for clarity.

Fig. 2 shows the average compression set cardinality d^* observed over 50 trials, corresponding to different randomly generated cases corresponding to different values of n and M. In all cases, d^* is bounded by $(n + 1)N$ as suggested by Theorem 8 (solid line). In fact, the empirically calculated cardinality d^* is significantly lower, suggesting that in this case study an *a posteriori* quantification is less conservative. Moreover, in all our numerical investigations we noticed that $d^* \leq n$ (dashed line), i.e., the empirical compression set cardinality is independent of the number of agents and is bounded by the individual number of decision variables. We conjecture that for this example, the so called support rank (see [42] for a definition) offers a tighter bound on the compression set cardinality compared to $(n + 1)N$.

VI. CONCLUSION

We considered the problem of NE computation in multiagent games in the presence of uncertainty, and accompanied them with *a priori* and *a posteriori* certificates regarding the probability that the NE equilibrium remains unchanged when a new uncertainty realization is encountered.

Current article is concentrated towards relaxing the uniqueness requirements underpinning the compression set quantification of Section IV for the class of aggregative games. Moreover, we aim at extending our results to the class of generalized games [11], [14], [31], [32], and at investigating the use of distributed NE seeking algorithms that require communication only among neighbouring agents [38].

ACKNOWLEDGMENT

Preliminary results related to this article can be found in [1]; here we complement these results by including an *a priori* analysis, by providing complete proofs for the *a posteriori* case and a more detailed numerical analysis, while we are not concerned with the issue of decentralized equilibrium computation.

REFERENCES

[1] F. Fele and K. Margellos, "Probabilistic sensitivity of Nash equilibria in multi-agent games: A wait-and-judge approach," in *Proc. IEEE 58th Conf. Decis. Control*, 2019, pp. 5026–5031.
[2] H. Le Cadre, I. Mezghani, and A. Papavasiliou, "A game-theoretic analysis of transmission-distribution system operator coordination," *Eur. J. Oper. Res.*, vol. 274, no. 1, pp. 317–339, 2019.
[3] A. De Paola, F. Fele, D. Angeli, and G. Strbac, "Distributed coordination of price-responsive electric loads: A receding horizon approach," in *Proc. IEEE Conf. Decis. Control*, Dec. 2018, pp. 6033–6040.
[4] I. Atzeni, L. G. Ordez, G. Scutari, D. P. Palomar, and J. R. Fonollosa, "Noncooperative day-ahead bidding strategies for demand-side expected cost minimization with real-time adjustments: A GNEP approach," *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2397–2412, May 2014.
[5] R. Kamat and S. Oren, "Two-settlement systems for electricity markets under network uncertainty and market," *Power. J. Regulatory Econ.*, vol. 25, pp. 5–37, Jan. 2004.
[6] J. Yao, I. Adler, and S. Oren, "Modelling and computation of two-settlement oligopolistic equilibrium in a congested electricity network," *Oper. Res.*, vol. 56, no. 1, pp. 34–47, Jan. 2008.
[7] G. Scutari, F. Facchinei, J. S. Pang, and D. P. Palomar, "Real and complex monotone communication games," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4197–4231, Jul. 2014.
[8] J. Krawczyk, "Numerical solutions to coupled-constraint (or generalised Nash) equilibrium problems," *Comput. Manage. Sci.*, vol. 4, pp. 183–204, Nov. 2007.

[9] J. Ohlin, "Nash equilibrium and international law," *Eur. J. Int. Law*, vol. 23, no. 4, pp. 915–940, Dec. 2012.
[10] T. Baar and G. Olsder, *Dynamic Noncooperative Game Theory, 2nd Edition*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1998.
[11] F. Facchinei and C. Kanzow, "Generalized Nash equilibrium problems," *4OR*, vol. 5, no. 3, pp. 173–210, Sep. 2007.
[12] S. Grammatico, "Dynamic control of agents playing aggregative games with coupling constraints," *IEEE Trans. Autom. Control*, vol. 62, no. 9, pp. 4537–4548, Sep. 2017.
[13] L. Deori, K. Margellos, and M. Prandini, "Price of anarchy in electric vehicle charging control games: When Nash equilibria achieve social welfare," *Automatica*, vol. 96, pp. 150–158, 2018.
[14] D. Paccagnan, B. Gentile, F. Parise, M. Kamgarpour, and J. Lygeros, "Nash and Wardrop equilibria in aggregative games with coupling constraints," *IEEE Trans. Autom. Control*, vol. 64, no. 4, pp. 1373–1388, Apr. 2019.
[15] F. Fele, A. De Paola, D. Angeli, and G. Strbac, "A framework for receding-horizon control in infinite-horizon aggregative games," *Annu. Rev. Control*, vol. 45, pp. 191–204, 2018.
[16] J.-S. Pang, S. Sen, and U. V. Shanbhag, "Two-stage non-cooperative games with risk-averse players," *Math. Program.*, vol. 165, no. 1, pp. 235–290, Sep. 2017.
[17] P. Couchman, B. Kouvaritakis, M. Cannon, and F. Prashad, "Gaming strategy for electric power with random demand," *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1283–1292, Aug. 2005.
[18] V. V. Singh, O. Jouini, and A. Lisser, "Existence of Nash equilibrium for chance-constrained games," *Oper. Res. Lett.*, vol. 44, no. 5, pp. 640–644, 2016.
[19] U. Ravat and U. Shanbhag, "On the characterization of solution sets of smooth and nonsmooth convex stochastic Nash games," *SIAM J. Optim.*, vol. 21, no. 3, pp. 1168–1199, 2011.
[20] J. Koshal, A. Nedic, and U. V. Shanbhag, "Regularized iterative stochastic approximation methods for stochastic variational inequality problems," *IEEE Trans. Autom. Control*, vol. 58, no. 3, pp. 594–609, Mar. 2013.
[21] J. Lei and U. V. Shanbhag, "A randomized inexact proximal best-response scheme for potential stochastic Nash games," in *Proc. IEEE 56th Annu. Conf. Decis. Control*, Dec. 2017, pp. 1646–1651.
[22] M. Aghassi and D. Bertsimas, "Robust game theory," *Math. Program.*, vol. 107, no. 1, pp. 231–273, Jun. 2006.
[23] J. Zazo, S. Zazo, and S. V. Macua, "Robust worst-case analysis of demand-side management in smart grids," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 662–673, Mar. 2017.
[24] A. J. Conejo, M. Carrión, and J. M. Morales, *Uncertainty Characterization via Scenarios*. Boston, MA, USA: Springer, 2010, pp. 63–119.
[25] S. Floyd and M. Warmuth, "Sample compression, learnability, and the Vapnik-Chervonenkis dimension," *Mach. Learn.*, vol. 21, no. 3, pp. 269–304, Dec. 1995.
[26] T. Alamo, R. Tempo, and E. F. Camacho, "Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems," *IEEE Trans. Autom. Control*, vol. 54, no. 11, pp. 2545–2559, Nov. 2009.
[27] K. Margellos, M. Prandini, and J. Lygeros, "On the connection between compression learning and scenario based single-stage and cascading optimization problems," *IEEE Trans. Autom. Control*, vol. 60, no. 10, pp. 2716–2721, Oct. 2015.
[28] G. C. Calafiore and M. C. Campi, "The scenario approach to robust control design," *IEEE Trans. Autom. Control*, vol. 51, no. 5, pp. 742–753, May 2006.
[29] M. C. Campi, S. Garatti, and F. A. Ramponi, "A general scenario theory for nonconvex optimization and decision making," *IEEE Trans. Autom. Control*, vol. 63, no. 12, pp. 4067–4078, Dec. 2018.
[30] M. Campi and S. Garatti, "The exact feasibility of randomized solutions of uncertain convex programs," *SIAM J. Optim.*, vol. 19, no. 3, pp. 1211–1230, 2008.
[31] M. A. Noor, "Quasi variational inequalities," *Appl. Math. Lett.*, vol. 1, no. 4, pp. 367–370, 1988.
[32] A. Kannan, U. Shanbhag, and H. Kim, "Addressing supply-side risk in uncertain power markets: Stochastic Nash models, scalable algorithms and error analysis," *Optim. Methods Softw.*, vol. 28, no. 5, pp. 1095–1138, 2013.
[33] D. Paccagnan and M. C. Campi, "The scenario approach meets uncertain game theory and variational inequalities," in *Proc. IEEE 58th Conf. Decis. Control*, 2019, pp. 6124–6129.
[34] G. Calafiore, "Random convex programs," *SIAM J. Optim.*, vol. 20, no. 6, pp. 3427–3464, 2010.
[35] F. Facchinei, J.-S. Pang, and G. Scutari, "Non-cooperative games with minmax objectives," *Comput. Optim. Appl.*, vol. 59, no. 1, pp. 85–112, Oct. 2014.
[36] M. C. Campi and S. Garatti, "Wait-and-judge scenario optimization," *Math. Program.*, vol. 167, no. 1, pp. 155–189, Jan. 2018.
[37] A. Kannan and U. Shanbhag, "Distributed computation of equilibria in monotone Nash games via iterative regularization techniques," *SIAM J. Optim.*, vol. 22, no. 4, pp. 1177–1205, 2012.
[38] P. Yi and L. Pavel, "Distributed generalized Nash equilibria computation of monotone games via double-layer preconditioned proximal-point algorithms," *IEEE Trans. Control Netw. Syst.*, vol. 6, no. 1, pp. 299–311, Mar. 2019.
[39] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*. New York, NY, USA: Springer-Verlag, 2003.
[40] B. Rustem and M. Howe, *Algorithms for Worst-Case Design and Applications to Risk Management*. Princeton, NJ, USA: Princeton University Press, 2002.
[41] National Grid. (2019, Mar.) Historical demand data. [Online]. Available: https://www.nationalgrideso.com/data-explorer
[42] G. Schildbach, L. Fagiano, and M. Morari, "Randomized solutions to convex programs with multiple chance constraints," *SIAM J. Optim.*, vol. 23, no. 4, pp. 2479–2501, 2013.